



TRIFACTA

Application Reference Guide

Version: 9.2

Doc Build Date: 07/29/2022

Copyright © Trifacta Inc. 2022 - All Rights Reserved. CONFIDENTIAL

These materials (the “Documentation”) are the confidential and proprietary information of Trifacta Inc. and may not be reproduced, modified, or distributed without the prior written permission of Trifacta Inc.

EXCEPT AS OTHERWISE PROVIDED IN AN EXPRESS WRITTEN AGREEMENT, TRIFACTA INC. PROVIDES THIS DOCUMENTATION AS-IS AND WITHOUT WARRANTY AND TRIFACTA INC. DISCLAIMS ALL EXPRESS AND IMPLIED WARRANTIES TO THE EXTENT PERMITTED, INCLUDING WITHOUT LIMITATION THE IMPLIED WARRANTIES OF MERCHANTABILITY, NON-INFRINGEMENT AND FITNESS FOR A PARTICULAR PURPOSE AND UNDER NO CIRCUMSTANCES WILL TRIFACTA INC. BE LIABLE FOR ANY AMOUNT GREATER THAN ONE HUNDRED DOLLARS (\$100) BASED ON ANY USE OF THE DOCUMENTATION.

For third-party license information, please select **About Trifacta** from the Help menu.

1. Reference	6
1.1 UI Reference	7
1.1.1 Home Page	8
1.1.1.1 Download Logs Dialog	12
1.1.2 Flows Page	13
1.1.2.1 Create Flow Page	17
1.1.2.2 Flow View Page	18
1.1.2.2.1 View for Connections	26
1.1.2.2.2 View for Imported Datasets	29
1.1.2.2.3 View for Dataset with Parameters	33
1.1.2.2.4 View for Recipes	36
1.1.2.2.5 View for Reference Datasets	41
1.1.2.2.6 View for Unstructured Datasets	45
1.1.2.2.7 View for Outputs	47
1.1.2.2.8 Share Flow Dialog	53
1.1.2.2.9 Change Dataset Dialog	56
1.1.2.2.10 Add Schedule Dialog	58
1.1.2.2.11 Manage Flow Notifications Dialog	60
1.1.2.2.12 Manage Parameters Dialog	62
1.1.2.2.13 Flow Search Panel	67
1.1.2.2.14 Flow Optimization Settings Dialog	69
1.1.3 Plans Page	72
1.1.3.1 Plan View Page	74
1.1.3.1.1 Plan Run Details Page	79
1.1.3.1.2 Share Plan Dialog	81
1.1.3.1.3 Manage Plan Notifications Dialog	82
1.1.3.1.4 Plan View for Flow Tasks	84
1.1.3.1.5 Plan View for HTTP Tasks	90
1.1.3.1.6 Plan View for Slack Tasks	93
1.1.3.1.7 Plan View for Delete Tasks	96
1.1.4 Library Page	99
1.1.4.1 Dataset Details Page	102
1.1.4.2 Import Data Page	105
1.1.4.2.1 Create Connection Window	112
1.1.4.2.2 Database Browser	117
1.1.4.2.3 File System Browser	120
1.1.4.2.4 HDFS Browser	122
1.1.4.2.5 S3 Browser	124
1.1.4.2.6 File Import Settings	126
1.1.4.2.7 Table Import Settings	128
1.1.4.3 Macros Page	129
1.1.4.3.1 Macro Details Page	131
1.1.5 Connections Page	133
1.1.5.1 Share Connection Dialog	136
1.1.6 Jobs Page	140
1.1.6.1 Job Details Page	145
1.1.6.1.1 Publishing Dialog	157
1.1.6.1.2 Schema Changes Dialog	159
1.1.6.2 Plan Runs Page	161
1.1.6.3 Sample Jobs Page	164
1.1.7 Transformer Page	166
1.1.7.1 Data Grid Panel	170
1.1.7.1.1 Column Histograms	176
1.1.7.1.2 Data Quality Bars	178
1.1.7.1.3 Lookup Wizard	179
1.1.7.1.4 Standardize Page	182
1.1.7.2 Recipe Navigator	185
1.1.7.3 Sample Indicator	186
1.1.7.4 Transformer Toolbar	187
1.1.7.5 Column Menus	191

1.1.7.5.1	<i>Choose Datetime Format Dialog</i>	195
1.1.7.5.2	<i>Transformation by Example Page</i>	197
1.1.7.6	<i>Column Browser Panel</i>	200
1.1.7.7	<i>Column Details Panel</i>	204
1.1.7.8	<i>Transform Preview</i>	207
1.1.7.9	<i>Context Panel</i>	210
1.1.7.9.1	<i>Recipe Panel</i>	211
1.1.7.9.2	<i>Transform Builder</i>	216
1.1.7.9.3	<i>Search Panel</i>	221
1.1.7.9.4	<i>Edit History Panel</i>	224
1.1.7.9.5	<i>Samples Panel</i>	226
1.1.7.9.6	<i>Selection Details Panel</i>	232
1.1.7.10	<i>Filter Panel</i>	238
1.1.7.11	<i>Visible Columns Panel</i>	240
1.1.7.12	<i>Join Window</i>	241
1.1.7.13	<i>Union Page</i>	247
1.1.7.14	<i>Run Job Page</i>	251
1.1.7.14.1	<i>Publishing Actions</i>	255
1.1.7.14.2	<i>SQL Scripts Panel</i>	266
1.1.7.14.3	<i>Redshift Table Settings</i>	268
1.1.7.14.4	<i>Hive Table Settings</i>	269
1.1.7.14.5	<i>Databricks Tables Table Settings</i>	270
1.1.7.14.6	<i>Tableau Server Datasource Settings</i>	272
1.1.7.14.7	<i>Spark Execution Properties Settings</i>	273
1.1.7.14.8	<i>Microsoft SQL Data Warehouse Table Settings</i>	275
1.1.8	<i>Preferences Page</i>	277
1.1.8.1	<i>User Profile Page</i>	279
1.1.8.1.1	<i>Storage Config Page</i>	281
1.1.8.2	<i>Account Settings Page</i>	283
1.1.8.3	<i>Email Notifications Page</i>	285
1.1.8.4	<i>Sessions Page</i>	287
1.1.8.5	<i>Access Tokens Page</i>	288
1.1.8.6	<i>Databricks Settings Page</i>	290
1.1.9	<i>Schedules Page</i>	292
1.1.10	<i>UI Index</i>	294
1.2	<i>Supported File Formats</i>	300
1.3	<i>Column Statistics Reference</i>	306
1.4	<i>Supported Data Types</i>	308
1.4.1	<i>String Data Type</i>	310
1.4.2	<i>Integer Data Type</i>	311
1.4.3	<i>Decimal Data Type</i>	312
1.4.4	<i>Boolean Data Type</i>	313
1.4.5	<i>Social Security Number Data Type</i>	314
1.4.6	<i>Phone Number Data Type</i>	315
1.4.7	<i>Email Address Data Type</i>	316
1.4.8	<i>Credit Card Data Type</i>	317
1.4.9	<i>Gender Data Type</i>	318
1.4.10	<i>Zip Code Data Type</i>	319
1.4.11	<i>State Data Type</i>	320
1.4.12	<i>Object Data Type</i>	321
1.4.13	<i>Array Data Type</i>	322
1.4.14	<i>IP Address Data Type</i>	323
1.4.15	<i>URL Data Type</i>	324
1.4.16	<i>HTTP Code Data Type</i>	325
1.4.17	<i>Datetime Data Type</i>	326
1.5	<i>Supported Numeric Formatting</i>	331
1.6	<i>Type Conversions</i>	334
1.6.1	<i>Avro Data Type Conversions</i>	336
1.6.2	<i>DB2 Data Type Conversions</i>	337

1.6.3	<i>Hive Data Type Conversions</i>	338
1.6.4	<i>Oracle Data Type Conversions</i>	341
1.6.5	<i>MySQL Data Type Conversions</i>	344
1.6.6	<i>Parquet Data Type Conversions</i>	346
1.6.7	<i>Postgres Data Type Conversions</i>	348
1.6.8	<i>Redshift Data Type Conversions</i>	351
1.6.9	<i>Snowflake Data Type Conversions</i>	353
1.6.10	<i>AWS Glue Data Type Conversions</i>	355
1.6.11	<i>Salesforce Data Type Conversions</i>	356
1.6.12	<i>SQL Server Data Type Conversions</i>	357
1.6.13	<i>SQL DW Data Type Conversions</i>	360
1.6.14	<i>Databricks Tables Data Type Conversions</i>	363
1.6.15	<i>Tableau Hyper Data Type Conversions</i>	366
1.6.16	<i>Teradata Data Type Conversions</i>	367
1.6.17	<i>SharePoint Data Type Conversions</i>	370
1.7	<i>Transformation Reference</i>	372
1.8	<i>Plan Metadata References</i>	378
1.9	<i>cron Schedule Syntax Reference</i>	385
1.10	<i>Supported Time Zone Values</i>	390
1.11	<i>Locale Settings</i>	406
1.12	<i>Character Encoding</i>	408
1.13	<i>Supported File Encoding Types</i>	409
1.14	<i>Sort Order</i>	410
1.15	<i>Join Types</i>	412
1.16	<i>Join Metrics</i>	416
1.17	<i>Supported SQL Syntax</i>	417
1.18	<i>Sample Types</i>	419
1.19	<i>Support Bundle Contents</i>	422

Reference

This section contains reference content on the interface and other aspects of Trifacta®.

UI Reference

Review reference information on each screen available in the Trifacta® application.

For more information on UI pages that apply to administrators, see *Admin Reference*.

Home Page

Contents:

- Recently Updated
- Recent Runs
- Resources
- Left Nav Bar
 - Flows
 - Plans
 - Library
 - Connections
 - Jobs
 - Schedules
 - Help menu
 - User menu

From the Home page, you can create or access your flows, datasets, and jobs, as well as configure settings and find additional resources.

Tip: Click the logo at the top of the menu bar to return to the Home page.

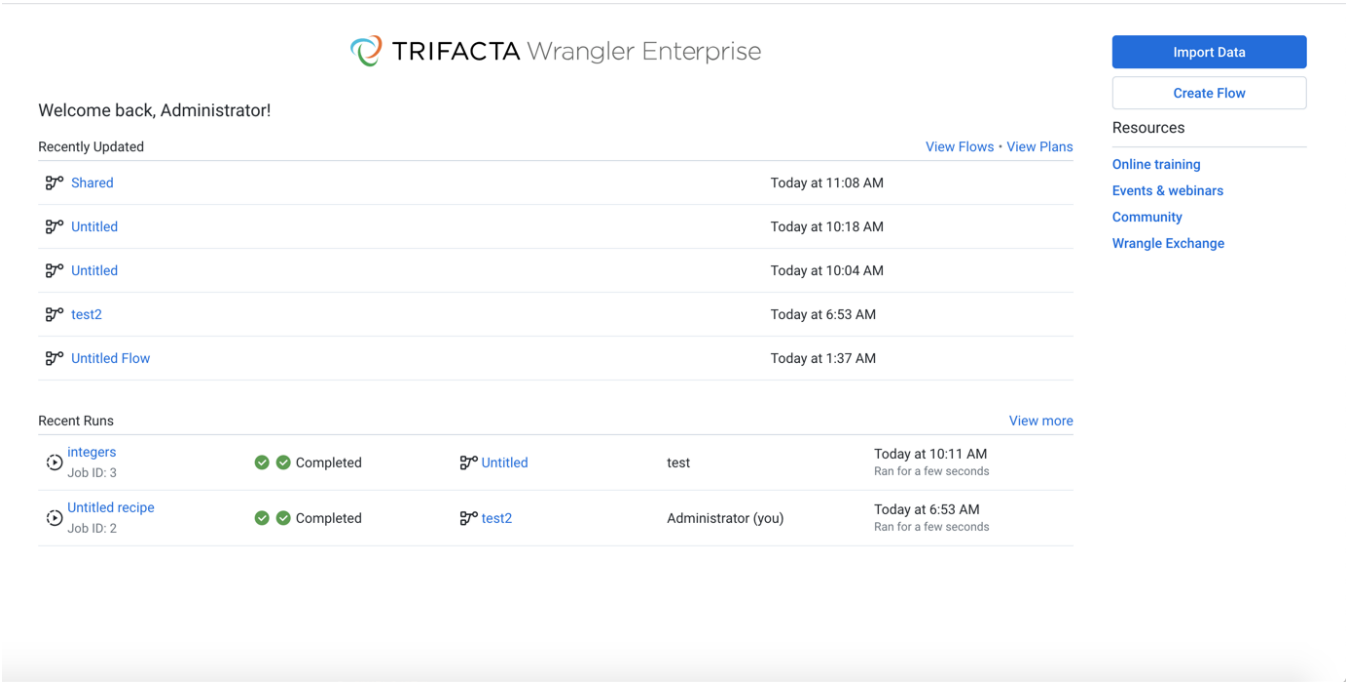


Figure: Home Page

From the Home page, you can quickly access your recent activities in Trifacta or jump to creating flows and importing datasets.

Tip: When keyboard shortcuts are enabled, press ? in the application to see the available shortcuts. Individual users must enable them. See *User Profile Page*.

- **Import Data:** Import new datasets into Trifacta. See *Import Data Page*.
- **Create Flow:** Create a new flow to hold your datasets. See *Create Flow Page*.

Tip: Use the controls on the left side of the screen to access other areas of the application. For more information, see Menu Bar below.

Recently Updated

Access the flows that have been recently changed. Click the flow name to open it. See *Flow View Page*.

Tip: When an object within a flow has been changed, its timestamp here is updated, so the Home page becomes an easy location where you can monitor changes to the flows to which you have access. Monitored changes include editing a recipe or adding or removing datasets.

Actions:

- **View more:** See all of your flows.
- For more information on these options, see *Flows Page*.

Recent Runs

Recent jobs:

Review jobs that you have been recently queued or completed in Trifacta.

- Click a job ID to view its details. See *Job Details Page*.
- Click the name of the flow to open it. See *Flow View Page*.
- Click the name of a recipe to select it in Flow View.

Recent Runs:

Use the links to explore recent plan runs that have been queued or completed. For more information, see *Plan Runs Page*.

Actions:

- **Cancel job:** If present, you can cancel a job in progress.

NOTE: Some jobs cannot be canceled through the Trifacta application.

- **View more:** See all of your jobs.
- For more information on these options, see *Jobs Page*.

Resources

Use the links on the right to explore available resources for Trifacta.

Left Nav Bar

From the left side of the screen, you can access the top-level pages of the Trifacta application.

Flows

Use the Flows page to create and manage your flows.

- A flow is a container for one or more datasets. See *Create Flow Page*.

Plans

Use the Plans page to build sequences of tasks that are executed based upon triggers that you create. See *Plans Page*.

- Plans enable the scheduling of execution of task sequences. For more information, see *Overview of Operationalization*.

Library

From the Library page, you create and manage your datasets. See *Library Page*.

- You can import datasets to begin creating your transformations. See *Import Data Page*.
- Click the linked name of a dataset to transform the data. See *Transformer Page*.

Connections

Connections enable you to import datasets for use in your flows. Depending on the type of connection, they can also be used for writing or publishing data back to the datastore.

NOTE: This option appears only if connectivity must be enabled in your environment.

For more information on creating and editing connections to your data, see *Connections Page*.

Jobs

After you finish building your transformation recipe, you can run jobs to execute the recipe against your dataset. The Jobs page shows status and history of your jobs. See *Jobs Page*.

You can also review any plan runs that you have executed. See *Plan Runs Page*.

Schedules

Administrators can review, toggle availability, and delete any schedule in the deployment.

NOTE: This page is available to administrators only.

See *Schedules Page*.

Help menu

Access help resources, including documentation, training, and more.

- **Community:** Explore the online community.
- **Wrangle Exchange:** Explore and download additional resources, such as macros, from the online Wrangle Exchange.
- **Documentation:** Access online product documentation.
- **API documentation:** Access reference documentation for available endpoints and methods.
 - For additional information on API workflows and other API-related content, see *API Reference*.
- **Trifacta Academy:** Explore the training programs and certifications available through Trifacta Academy. Wrangler certification is free.
- **Download logs:** Download logs from the current session. See *Download Logs Dialog*.
 - Admin users can download logs for jobs, sessions, or time periods. See *Admin Download Logs Dialog*.
- **About:** Review information about the product, build number, and licensing information.

Tip: You can acquire the name of your product edition from the About screen.

- **Keyboard shortcuts:** Review available keyboard shortcuts for the Trifacta application.
 - You must enable keyboard shortcuts. See *User Profile Page*.

User menu

Under the User icon, you can modify settings specific to your account.

Preferences

Review preferences for your account and other settings. See *Preferences Page*.

- **Profile:** Edit your profile. See *User Profile Page*.

Admin console

Review and modify settings and users for your workspace. See *Admin Console*.

Tip: If you have two open tabs for work on the same dataset, changes made in one browser tab may not be reflected in the other browser tab until you refresh the page. Overwriting results from one tab through another is certainly possible.

Download Logs Dialog

You can download logs for your current session in Trifacta®. From the Help menu, select **Download logs**.

NOTE: The data downloaded for end users from this dialog is encrypted by default.

NOTE: For more information on disabling this feature, see *Configure Support Bundling*.

Administrators have a separate interface for downloading log files, which provides access to a wider set of logging data. For more information, see *Admin Download Logs Dialog*.

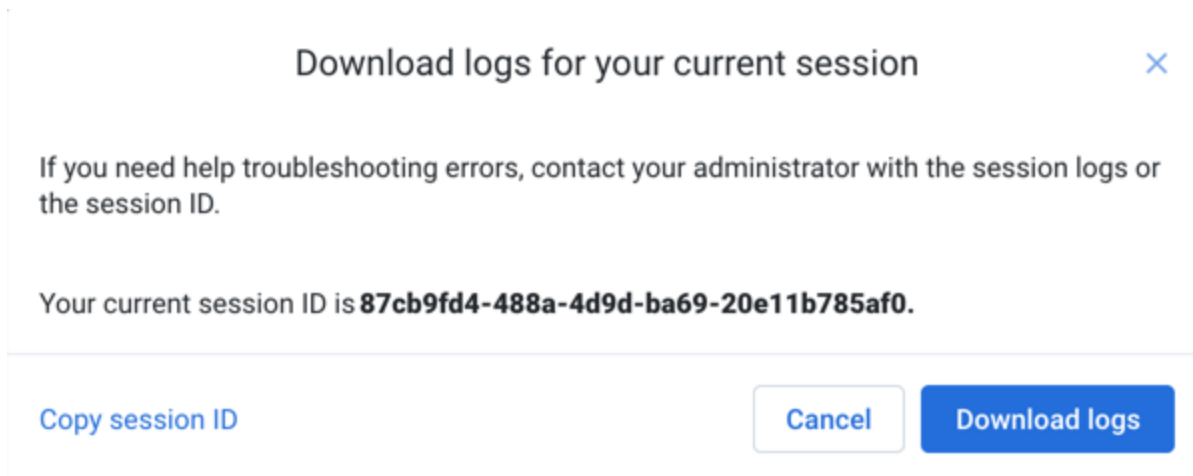


Figure: Download Logs Dialog

The above dialog contains your current session ID.

Tip: Using your session ID, your Trifacta administrator can download a larger set of logs, which can be useful for troubleshooting issues with the product.

- To copy your session identifier to the clipboard, click **Copy session ID**.
- To download logs for your current session, click **Download logs**. Logs pertaining to your current session are bundled in a ZIP and downloaded.

NOTE: There is a defined limit to the size of each log file. For more information, please contact your Trifacta administrator.

For more information on the contents of this download, see *Support Bundle Contents*.

Flows Page

Contents:

- All flows tab
- Owned by me tab
- Shared with me tab
- Folders

The Flows page displays the flows to which you have access and lets you create, review, and manage them. A **flow** is an object for bringing together and organizing the datasets, recipes, and other objects that you use to generate your results.

NOTE: Access to the Flows page in the application and privileges on flows is governed by roles in your workspace. For more information, please contact your workspace administrator.

- To create a new flow, click **Create Flow**. To rename the new flow, click the `Untitled` value at the top of the page. See *Create Flow Page*.
- You can also access the flows that have been shared with you.
- You can hover over the shared icon link next to the flow to view the name of shared users (up to three shared users) and the total number of shared users. Also, when you click the shared icon link, the share dialog is displayed.

You can organize your flows into folders. A **folder** is simply a container for your flows. To create a folder, click **Create**. Then, select **Create Folder**. For more information, see "Flows" below.

Flows

Create...Import Flow

Search flows

All flowsOwned by meShared with me

Name	Owner	Contains	Last updated
testflow	Administrator	6 Datasets, 0 Recipes	Today Share
[732ac110] implementing cube() function from sql	Administrator	3 Datasets, 6 Recipes	Today at 4:01 PM
[1e92ee00] [def96250] currencyFlow	Administrator	4 Datasets, 4 Recipes	Today at 2:58 PM
[1e92ee00] [def96250] currencyFlow	Administrator	4 Datasets, 4 Recipes	Today at 2:58 PM
Untitled Flow	Administrator	0 Datasets, 0 Recipes	Today at 1:50 PM

Figure: Flows Page

All flows tab

This tab includes all flows accessible to the user, either as owner or collaborator.

Owned by me tab

This tab contains the flows that you have created.

Columns:

- **Name:** The name of the flow.
 - Click the flow name to review the flow, its datasets, and its recipes. See *Flow View Page*.
- **Owner:** Indicates the user who is the owner of the flow.
- **Contains:** Count of datasets and recipes in the flow.
- **Last Updated:** Timestamp for the last time that the flow was modified.

Actions:

- **Create:** From the Create menu, choose to create a flow or a folder for holding flows.
 - For more information on creating a flow, see *Create Flow Page*.
 - For more information on folders, see "Folders" below.
- **Import:** From the context menu, select **Import Flow** to import a flow into this instance. See *Import Flow*.

Tip: You can import multiple flows (ZIP files) through the file browser or through drag-and-drop. Press **CTRL/COMMAND** + click or **SHIFT** + click to select multiple files for import.

- **Search:** To search flow names, enter a string in the search bar. Results are highlighted immediately in the Flows page.
- **Sort:** Click a column header to sort the display by the column's entries.

Flow options:

The following options are available on the right side of a flow's entry:

- **Share:** Enable other users to collaborate on your flows with you or create copies of your flow for their personal use. See *Share Flow Dialog*.
- **Rename:** Change the name and description of the flow.
- **Schedule:** To add a scheduled execution of the recipes in your flow:
 1. Define the scheduled time and interval of execution at the flow level. See *Add Schedule Dialog*.
 2. Define the scheduled destinations for each recipe through its output object. These destinations are targets for the scheduled job. See *Flow View Page*.
- **Email notifications:** Configure types of jobs that generate success or failure emails and who receives the messages. See *Manage Flow Notifications Dialog*.
- **Duplicate:** Create a copy of the flow. The copied flow is owned by the user who copied it.
- **Move:** Move the flow to a new or existing folder. See "Folders" below.
- **Export:** (Available to flow owner only) Export the flow from Trifacta . For more information, see *Export Flow*.
- **Delete:** Delete the flow.

Deleting a flow removes all recipes and related objects contained in the flow. If copies of these objects exist in other flows, they are not touched. Imported datasets are not deleted by this action.

For flows that have been shared with you, this command removes your access to them. To regain access, the owner of the flow must share it with you again.

Shared with me tab

If other users have shared flows with you, you can access them through the Shared with Me tab. Available options are very similar to the Owned by Me tab.

When a flow is shared with you, you are a collaborator in the flow. There are a few restrictions on how you can interact with a shared flow and its assets. See *Overview of Sharing*.

Folders

You can use folders to organize your flows. For example, you can use folders to group flows by project, by source of data, or by other meaningful grouping.

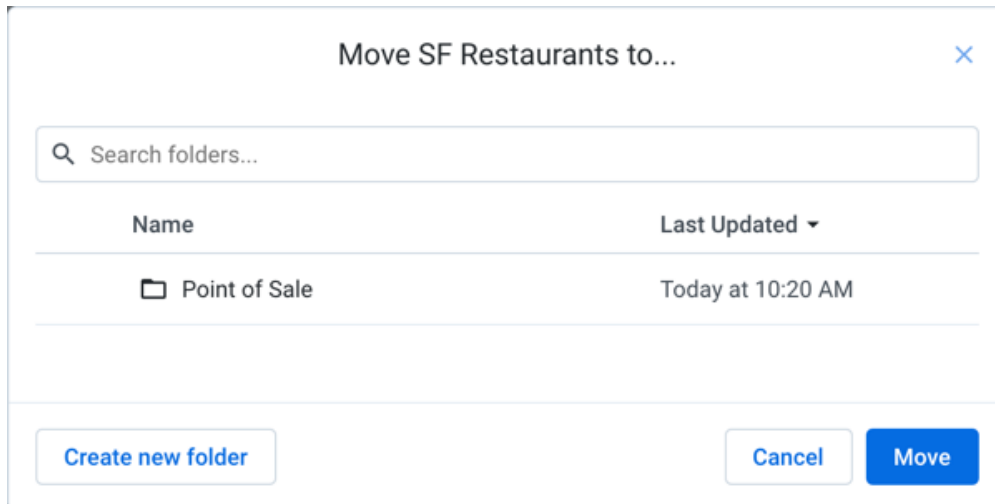


Figure: Moving a flow to a folder

Limitations:

- Each flow in a folder is an independent object. Permissions can vary between flows in a folder and should be reviewed after adding them.
- You can only move flows that you own.
- You cannot create nested folders.
- You cannot share folders or modify permissions at the folder level.
- Folders cannot be exported and imported.

Actions:

- To create a folder, click the **Create** button. Then, select **Create Folder**. Enter a name and description for the folder. These values appear in the application. Click **Create**.

Tip: When you move a flow, you can optionally choose to create a new folder for it.

- To move a flow to a folder, select **Move** from the context menu on the right side of the screen for the flow.
 - To move the flow, select the name of the flow. Click **Move**.
 - To search folder names, enter your search string in the Search textbox.
 - To move the flow into a new folder, click **Create new folder**. Enter a meaningful name and description for the folder. Select **Move to new folder**.
- To delete a folder, select **Delete Folder** from the context menu on the right side of the screen.

This step deletes the folder and all flows within it. This step cannot be undone.

Folder options:

For folders, the following options are available in the context menu.

- **Import flow:** Import the exported flow to Trifacta®
- **Edit Folder name and description:** Change the name and description of the folder.
- **Delete Folder:** Delete the folder and all flows within it.

Deleting a folder also removes any flows within it. This action cannot be undone.

Create Flow Page

You can use flows to organize your datasets and to track the jobs associated with them.

- A **flow** is a container for datasets, recipes, and related objects.
- To create a new flow click **Create Flow** in the *Flows Page*.

Tip: You can also create a flow while importing datasets.

Steps:

1. In the Flows page, click **Create Flow**. A new flow is created, with the name `Untitled - x`, where `x` is a number.
2. Click the `Untitled - x` to enter a flow name and description.
3. From the Flow View page, you can add datasets to your flow, or import new ones. You may add multiple datasets at this time and add more later.
 - a. **Add dataset:** You can browse or search for datasets to add to your flow from the available ones.
 - i. This list includes all imported and reference datasets to which you have access.
 - ii. Select a different search filter to display all, imported, or reference datasets.
 - iii. To add a selected dataset, click the checkbox next to it.
 - b. **Import Datasets:** Click this link to import a new dataset into the application. After it is imported, it is automatically added to your flow. See *Import Data Page*.
4. When finished, click **Add**.
5. The datasets are displayed in the flow. For more information, see *Flow View Page*.

Flow View Page

Contents:

- *Flow View Organization*
 - *Top Bar*
 - *Flow context menu*
 - *Add Datasets to Flow*
 - *Flow Canvas*
 - *Canvas context menu*
 - *Canvas notes*
 - *Flow objects*
 - *Context Panel*
 - *View for Imported Datasets*
 - *View for Dataset with Parameters*
 - *View for Unstructured Datasets*
 - *View for Recipes*
 - *View for Outputs*
 - *View for Reference Datasets*
-

In Flow View, you can access and manage the objects that you have added to or created in the selected flow. You can perform a variety of actions to effectively manage flow development and job execution through a single page in the Trifacta® application.

NOTE: Access to this page in the application and privileges on its related objects is governed by roles in your workspace. For more information, please contact your workspace administrator.

If you have enabled Deployment Manager, avoid making changes in Flow View on a Production instance of the platform.

- **Scheduling executions through Flow View in a Prod environment is not supported. Job executions must be executed through the APIs. See *API Workflow - Deploy a Flow*.**
- **Some Flow View options may not be available in a Prod environment.**
- **You should apply changes to your flow in the Dev instance and then re-deploy to Production. For more information, see *Overview of Deployment Manager*.**

NOTE: If the displayed flow has been shared with you, some options are not available.

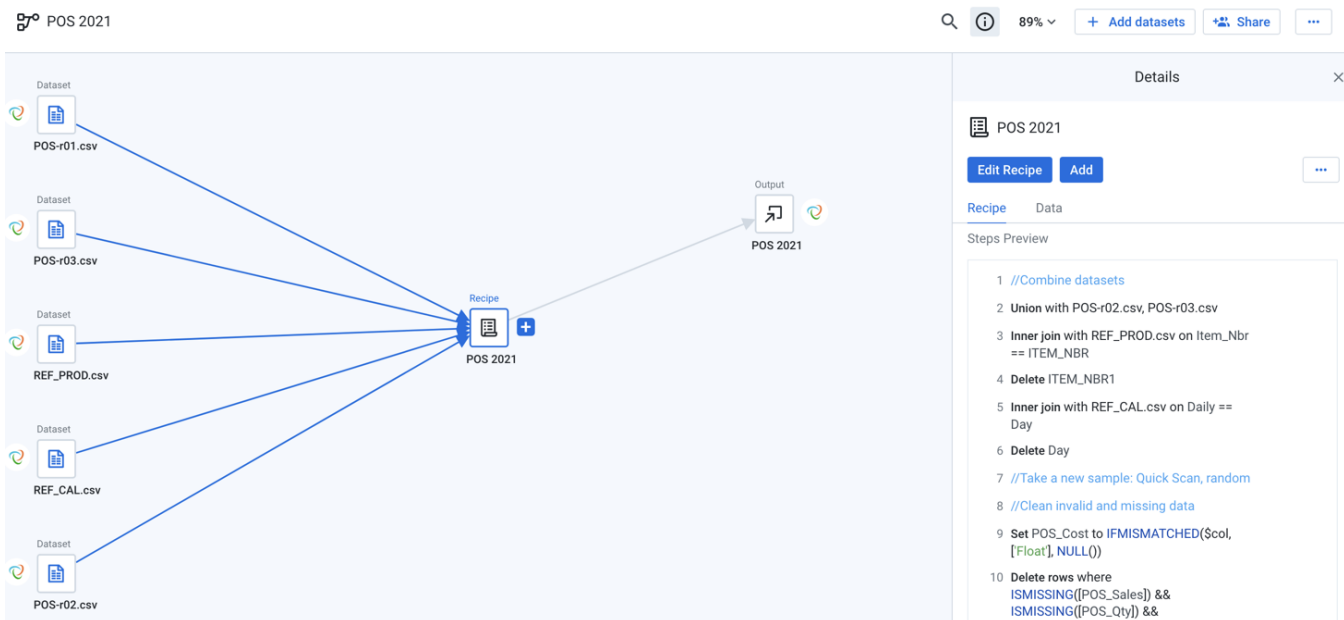


Figure: Flow View page

Flow View Organization

Flow View is organized into the following basic areas:

Area	Description
Top bar	At the top of the screen, you can access a menu of options, which enable adding objects to your flow, configuring aspects of your flow, and other flow management tasks. See "Top bar" below.
Flow canvas	The primary workspace of Flow View is the flow canvas, where you build and organize the objects in your flow. See "Flow Canvas" below.
Context panel	When you select one or more objects in your flow, the object details and relevant context menu options are available in the right-hand panel. See "Context Panel" below.

Top Bar

From the bar at the top of Flow View, the following options are available:

Tip: To rename the flow, click the flow name at the top of Flow View.

- **Search icon:** Click this icon to search for objects in your flow. See *Flow Search Panel*.
- **Details icon:** Click to toggle display of the details panel. This setting is saved for the individual user.
- **Zoom menu:** Flow View attempts to zoom the canvas to display as much of the flow as possible. As needed, you can change the level to zoom in or zoom out on the canvas.

Tip: You can access these zoom controls through the context menu for the flow canvas.

- **Zoom in:** Zoom in 10% on the canvas to focus on greater detail.
- **Zoom out:** Zoom out 10% from the canvas to see more of it.
- **Zoom to fit:** Change the zoom level to fit the objects of your flow.

- **Zoom to selection:** Zoom to center the selected object on the canvas.
- **25%, 50%, or 100%:** Change the zoom level to one of the preset levels.
- **Add Datasets:** Click to add new datasets to the flow. See "Add Datasets to Flow" below.
- **Share:** Click **Share** to collaborate with others on the same flow.

NOTE: When a flow containing one or more connections is shared, its connections are also shared. By default, credentials are included. If the sharing of credentials has been disabled, the new users must provide their own credentials for the shared connection. See *Configure Sharing*.

NOTE: You cannot share with users outside of your current workspace, including any account that you may have in a different workspace.

See *Share Flow Dialog*.

Flow context menu

- **Rename:** (Available to flow owner only) Change the name and description of the flow.
- **Schedule:** To add a scheduled execution of the recipes in your flow:
 - Flow owners can create schedules. Other users must have the appropriate privileges to create schedules for the flow.
 - Define the scheduled time and interval of execution at the flow level. See *Add Schedule Dialog*.
 - After the schedule has been created, you can review, edit, or delete the schedule through the Clock icon.
 - The time of the next scheduled run is displayed in the time zone of the local browser.
 - Define the scheduled destinations for each recipe through its output object. These destinations are targets for the scheduled job. See *View for Outputs* below.
- **Parameters:** Create and manage recipe parameters, as well as specify overrides for them. See *Manage Parameters Dialog*.
- **Webhooks:** You can define tasks to update third-party applications of the results of jobs executed from this flow. For more information, see *Create Flow Webhook Task*.

NOTE: Webhooks may need to be enabled in your environment. For more information, see *Workspace Settings Page*.

- **Email notifications:** Configure types of jobs that generate success or failure emails and who receives the messages. See *Manage Flow Notifications Dialog*.

NOTE: This feature uses an SMTP email server to send messages. For more information on configuring the server, see *Enable SMTP Email Server Integration*.

NOTE: This feature may need to be enabled in your environment. For more information, see *Workspace Settings Page*.

- **Optimization settings:** You can configure optimizations of your job executions of your flow. For more information, see *Flow Optimization Settings Dialog*.
- **Auto arrange:** Re-organize the custom node layout back to a fresh system generated layout. This option may be helpful for rearranging flow layouts that were originally created in classic Flow View.

The Auto arrange option cannot be undone.

- **Duplicate:** Create a copy of the flow.
 - The copied flow is independent of the source flow.
 - Optionally, you can duplicate the datasets from the original flow in the copy.
 - Duplicating datasets has some implications on shared flows. See *Overview of Sharing*.

NOTE: For flows using parameterized datasets, you should duplicate the datasets, which creates separate copies of parameters and their values in the new flow. If datasets are not copied, then parameter changes in the copied flow modify the values in the source flow.

- **Move:** Move the flow to a new or existing folder. See *Flows Page*.
- **Export:** Export the flow for archive or transfer. For more information, see *Export Flow*.
- **Delete:** Delete the flow.

Deleting a flow cannot be undone.

Deleting a flow removes all recipes that are contained in the flow. If copies of these objects exist in other flows, they are not touched. Imported datasets are not deleted by this action.

Add Datasets to Flow

From the Flow View page, you can add data through the following objects:

- Imported datasets - data sourced from outside the platform
- Reference datasets - dataset objects that are created from the output of a recipe in the current flow or another flow

NOTE: For long-loading relational sources, you can track progress of the load. While the data is loading, some recipe and flow options are disabled.

For more information on enabling long-loading, see *Configure JDBC Ingestion*.

These independent objects can be joined, unioned, or referenced by other datasets in the flow. For more information on these object types, see "View for Objects" below.

×

Add Datasets to Flow

All (20)

Imported (18)

Reference (2)

		NAME	SOURCE	LAST UPDATED
<input type="checkbox"/>		TESTDATA	oracle	Today at 8:48 AM
<input checked="" type="checkbox"/>		POS-PivotTable3.xlsx/Sheet1	HDFS	Today at 8:40 AM
<input type="checkbox"/>		POS-schema.csv	HDFS	Today at 8:30 AM
<input type="checkbox"/>		POS-r01.txt	HDFS	Today at 8:30 AM
<input type="checkbox"/>		POS-r02.txt	HDFS	Today at 8:30 AM
<input type="checkbox"/>		REF_CAL.txt	HDFS	Today at 8:30 AM

Import Datasets

Cancel

Add

Figure: Add datasets to current flow

- Search for or select the dataset to add.
 - Use the page view controls to browse for other datasets, or select the appropriate tab to filter the list to a specific type of object.
 - To import new datasets from external sources, click **Import Datasets**. See *Import Data Page*.
- When you have made your selections, click **Add**.
- The object or objects are added as a new object in flow view.

For large relational or Parquet datasets, you can monitor the import process through the Flow View page.

NOTE: This feature may require enablement in your deployment. See *Configure JDBC Ingestion*.

For more information, see *Overview of Job Monitoring*.

Flow Canvas

The central workspace of Flow View, the **canvas** is where you add and arrange the objects in your flow. In the canvas, you can select one or more objects at a time, drag them to reposition them on screen, and zoom in or out to focus on your current area of development.

When you add an object to the canvas, an icon representing it is added to the flow canvas. This object can be repositioned as needed.

Tip: The relative position of objects on the flow view canvas is preserved between screen updates. On refresh, the window on the canvas is repositioned based on the leftmost object on the canvas to focus on the flow to other objects from that one.

- If you create an object from another object, such as an output from a recipe, an arrow connects the recipe to the output.
- For any object, the objects on which it depends are displayed to the left of the original object, and there is a line from the preceding objects to the original object.

Tip: When you run a job for a recipe, all of the recipes steps for the preceding datasets are executed as part of the job, and only the results of the terminal dataset are generated.

NOTE: Objects marked with a red dot indicate a problem with the object's configuration. Please select the object to begin investigating the error. Error information may be displayed in the right panel.

Select:

- Click the icon for an object to select it.
 - Object details for the specific type of object are displayed in the context panel.
 - Right-click the object to open its context menu.
- To select multiple objects:
 - Click and drag over a set of objects.
 - To select a discrete set of objects, press **CTRL/COMMAND** + click the objects.
 - When you select multiple objects, the objects are listed in the context panel with options that are applicable to all of the objects. Some objects, like notes, do not have a context panel.
 - Right-click to display the context menu of options that are applicable to the selected objects.
- See "Context Panel" below.

Move:

- To move an object, click and drag it to a new location. Any arrows connecting to the object are repositioned as well.
- To move multiple objects, select them and then drag them to a new location.

Canvas context menu

When you right-click an empty part of the canvas, the following options are available:

- **Select All:** Select all objects on the canvas. Options that are relevant to all of the objects are displayed on the context panel.

Tip: You can drag these objects to reposition them together.

- **Add dataset:** Add new datasets to the flow. See "Add Datasets to Flow".
- **Auto arrange:** Re-organize the custom node layout back to a fresh system generated layout. This option may be helpful for rearranging flow layouts that were originally created in classic Flow View.

The Auto arrange option cannot be undone.

- **Add note:** Add a note with text or emojis to the canvas. See "Canvas notes" below.
- **Zoom:** Zoom in or zoom out on the canvas as needed. See "Top Bar" above for more details.

Canvas notes

As you develop your flow, you can add helpful notes on the canvas in various sections of the flow. For example, if you are collaborating with another user, you can leave status information about objects that are still in progress.

Tip: You can drag notes like other objects on the canvas, so they can be repositioned with the related flow object.

From a note's context menu, you can edit or delete the note.

Flow objects

In the flow canvas, you work with the following types of objects:

- Connections
- Imported datasets
 - Unstructured datasets
- Recipes
- Outputs
- Notes
- Reference datasets

- Datasets with parameters

Datasets:

- To begin working with data:
 - Click **Add datasets**. Locate your source data and import it into your flow. See "Add Datasets to Flow" above.
 - In Flow View, select the **imported dataset** on the flow canvas. Then, in the context panel, click **Add new recipe**. A new, empty recipe is associated with the dataset.

Tip: Double-click an imported dataset to see a preview of it. Some datasets cannot be previewed.

- To open in the Transformer page, click the recipe and select **Edit Recipe**. See *Transformer Page*.
- When created, these objects are connected together by lines flowing between them, which show the relationships between the objects in the flow.

Recipes:

A recipe is a set of steps to transform source data into the results you desire.

Tip: Double-click any recipe to edit it.

A recipe can be created from the following objects:

- An imported dataset, as described above.
- A **reference dataset** is an object that has been pulled into a flow from another flow. See "View for Reference Datasets" below.
- Another recipe. You can chain together recipes. For example, you may have a set of steps that you always apply at the beginning of transforming a specific type of feed. This recipe can be added into each flow as the first recipe chained to an imported dataset of that feed type.

Output objects:

The following objects can be created off of a recipe:

- An **output** object is a set of publishing targets for which you can execute jobs.
- A **reference** object is a reference to one of your flow's recipes that can be used in another flow. When a reference object is created, the target flow receives the output of the executed recipe.
 - In the target flow, this object appears as a **reference dataset**.
 - When a reference dataset is used in a flow, the target flow receives the output of the executed recipe.

For more information on these objects, see *Object Overview*.

Context Panel

Select an object from your flow to open an object-specific panel on the right side of the screen.

- When multiple objects are selected, the displayed details and options apply to all of the objects.
- If the selected objects are of different types, the available options are limited.

Tip: You can right-click any object in Flow View to see the list of available actions that appear when you select it and choose from the right panel.

Depending on the type of object or objects that you select in the canvas, the view in the context panel changes:

Tip: The object on the canvas and the context panel display the same set of context menu options. For more information on these options, see the links below.

View for Imported Datasets

See *View for Imported Datasets*.

View for Dataset with Parameters

For datasets with parameters, the context panel and available options are different from non-parameterized imported datasets. See *View for Dataset with Parameters*.

View for Unstructured Datasets

If detecting structure has been disabled for your imported dataset, the structure detecting steps are broken out into the first steps of a recipe that is auto-generated for you. You can review and modify them within the recipe. See *View for Unstructured Datasets*.

View for Recipes

See *View for Recipes*.

View for Outputs

Outputs are created from recipes. See *View for Outputs*.

View for Reference Datasets

A reference dataset is sourced from the outputs of another recipe. A reference dataset can be used in a flow other than the source flow. See *View for Reference Datasets*.

View for Connections

In Flow View, you can review the details of the connections used to access the flow's imported datasets, whether you created it or it was shared with you.

Select the imported dataset that uses the connection.

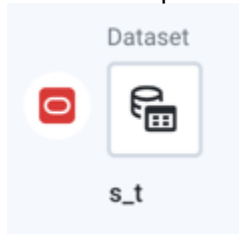


Figure: Connection icon

Icon context menu

Right-click the connection icon next to the imported dataset. The context menu items apply to the imported dataset. For more information, see *View for Imported Datasets*.

Details options


Select the connection to view the details options. Most options pertain to the dataset imported through the connection.

To review details about the connection, click the connection name in the Details panel.

< Dataset

Details

×

 oracle

Edit Connection

...

Connection Type

oracle

Shared

2 Users

Owner

Administrator

Created

Today at 8:48 AM

Updated

Today at 10:05 AM

Updated by

Administrator

Server Information

Host

oracle.ci.rds.trifacta.net

Port

1521

SSL

Disabled

Service Name

orcl

Username

trifacta

Password

.....

Figure: Connection details

Key Fields:**Connection Type:** For more information, see *Connection Types*.

Owner: User that owns the connection. This user can modify connection properties.

Server information: You can review information about the source to which the connection links.

Shared:

Private - Connection is available for use only for specified users of the platform.

Public - Connection is available for all users.

For more information, see *Share Connection Dialog*.

Details options

Edit Connection: If you have permissions, this option enables you to modify the connection. See *Create Connection Window*.

NOTE: For shared connections, you may only modify the username and password if they were not provided to you. All other fields are read-only.

Share: Click to share the connection with other users.

NOTE: You can share connections that have been shared with you. You cannot modify their properties.

Tip: If groups have been enabled in your instance of the Trifacta platform, you can share flows and connections to LDAP groups. For more information, see *Configure Users and Groups*.

See *Share Connection Dialog*.

Delete: If you are the connection owner, you can delete the connection.

Deleting a connection cannot be undone. All datasets that use the connection are no longer accessible.

View for Imported Datasets

When you select an imported dataset in Flow View, you can review its details in the context panel and select options from its context menu.



Figure: Imported Dataset icon

Icon context menu

The following menu options are available when you select the plus icon next to the imported dataset:

- **Add new recipe:** Add a new recipe extending from the current recipe. This new recipe operates on the outputs of the original recipe.
- **Add Join:** Add a join step as the new last step to the recipe. See *Join Window*.
- **Add Union:** Add a union step as the new last step to the recipe. See *Union Page*.

Details options

The following options are available in the details context menu when you select an imported dataset.

- **Add:**
 - **Recipe:** Add a recipe for this dataset.
 - **Join:** Join this dataset with another recipe or dataset. If this dataset does not have a recipe for it, a new recipe object is created to store this step.
 - **Union:** Union this dataset with one or more recipes or datasets. If this dataset does not have a recipe for it, a new recipe object is created to store this step.
- **View dataset details:** Explore details of the dataset. See *Dataset Details Page*.
- **Replace:** Replace the dataset with a different dataset or reference dataset.
- **Replace with dataset with Parameters:** For datasets that are not parameterized, you can choose to replace with datasets with parameters.

Tip: You may find it useful to create your recipes with a single static dataset and then later replace with a dataset with parameters.

- **Edit name and description:** (Available to flow owner only) Change the name and description for the object.
- **Edit custom SQL:** After you have created a dataset using custom SQL, you can modify the SQL used to construct the imported dataset. See *Create Dataset with Parameters*.
- **Edit parameters:** If your dataset contains parameters, you can change the parameters and their default values.
- **Remove structure:** (If applicable) Remove the initial parsing structure. When the structure is removed:

- The dataset is converted to an unstructured dataset. An **unstructured dataset** is the source data converted into a flat file format.
- All steps to shape the dataset are removed. You must break up columns in manual steps in any recipe created from the object.
- See *View for Unstructured Datasets*.
- **Remove from Flow:** Remove the dataset from the flow.
 - All dependent flows, outputs, and references are not removed from the flow. You can replace the source for these objects as needed.

NOTE: References to the deleted dataset in other flows remain broken until the dataset is replaced.

- **Refresh Dataset:** If available, this option refreshes the dataset's metadata with the latest source schema.

NOTE: If you attempt to refresh the schema of a parameterized dataset based on a set of files, only the schema for the first file is checked for changes. If changes are detected, the other files are contain those changes as well. This can lead to changes being assumed or undetected in later files and potential data corruption in the flow.


For more information, see *Overview of Schema Management*.

Tip: You can also right-click the imported dataset to view all the menu options.

When you select an imported dataset, you can preview the data contained in it, replace the source object, and more from the right-side panel.

Details

×

 POS-schema.csv

Add

View dataset details

...

Data Preview

# Item_Nbr	# Store_Nbr	# WM_Week
381000	1	201050
325000	2	201049
325000	2	201049
403000	2	201049
449000	2	201049
490000	2	201049
560000	2	201049

Type

HDFS

Location

hdfs:///trifacta/uploads/1/26149d74-6cf8-46a7-aa8d-a1f02cc3c242/POS-schema.csv

File Size

1.72MB

Size

29 columns · 5 types

Updated

Today at 8:30 AM

Figure: Imported Dataset view

Key Fields:

- **Data Preview:** In the Data Preview window, you can see a small section of the data that is contained in the imported dataset. This window can be useful for verifying that you are looking at the proper data.

Tip: Click the preview to open a larger dialog, where you can select and copy data.

- **Type:** Indicates where the data is sourced or the type of file.
- **Location:** Path to the location of the imported dataset.
- **File Size:** Size of the file. Units may vary.
- **Column Data Type Inference:**

NOTE: This field is only applicable to datasets imported from relational sources.

- **enabled** - Data types have been applied to the dataset during import.
- **disabled** - Data types were not globally applied to the dataset during import. However, some columns may have had overrides applied to them during the import process. See *Import Data Page*. For more information, see *Configure Type Inference*.
- **ConnectionName:** If the data is accessed through a connection, you can click this link to review connection details in the right-side panel.
- **More details:** Review details on the flows where the dataset is used.

View for Dataset with Parameters

When you select a dataset with parameters in Flow View, additional options are available in its context menu and the Details panel.

For more information on these objects, see *Overview of Parameterization*.

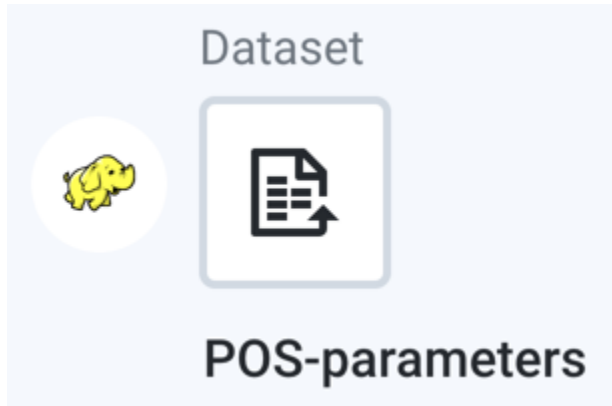


Figure: Dataset with Parameters icon

Icon context menu

The following menu options are available when you select the plus icon next to the dataset with parameters:

- **Add new recipe:** Add a new recipe extending from the current recipe. This new recipe operates on the outputs of the original recipe.
- **Add Join:** Add a join step as the new last step to the recipe. See *Join Window*.
- **Add Union:** Add a union step as the new last step to the recipe. See *Union Page*.

Details options

The following options are available in the Details context menu when you select a dataset with parameters:

- **Add:**
 - **Recipe:** Add a recipe for this dataset.
 - **Join:** Join this dataset with another recipe or dataset. If this dataset does not have a recipe for it, a new recipe object is created to store this step.
 - **Union:** Union this dataset with one or more recipes or datasets. If this dataset does not have a recipe for it, a new recipe object is created to store this step.
- **View dataset details:** Explore details of the dataset. See *Dataset Details Page*.
- **Edit name and description:** (Available to flow owner only) Change the name and description for the object.
- **Edit parameters:** If your dataset contains parameters, you can change the parameters and their default values.
- **Remove structure:** (If applicable) Remove the initial parsing structure. When the structure is removed:
 - The dataset is converted to an unstructured dataset. An **unstructured dataset** is the source data converted into a flat file format.
 - All steps to shape the dataset are removed. You must break up columns in manual steps in any recipe created from the object.
 - See *View for Unstructured Datasets*.
- **Remove from Flow:** Remove the dataset from the flow.
 - All dependent flows, outputs, and references are not removed from the flow. You can replace the source for these objects as needed.


NOTE: References to the deleted dataset in other flows remain broken until the dataset is replaced.

Tip: You can also right-click the dataset with parameters to view all the menu options.

When you select a dataset with parameters in Flow View, you can review the parameters that have been specified for the selected dataset in the right panel.

Details

×

 Dataset with Parameters

Add

View dataset details

...

Details

Parameters

Path

hdfs://
trifacta/uploads/1/26149d74-6cf8-46a7-aa8d-
a1f02cc3c242/POS-r .★ {digit}{digit} .txt

Parameters

Dataset has 1 parameters

.★ {digit}{digit}

Type	Pattern
Rule	matches against trifacta pattern: `{digit}{digit}`

Figure: Parameters tab in Flow View

Key Fields:

Details tab:

- **Type:** Indicates where the data is sourced or the type of file.
- **Size:** Size of the file. Units may vary.
- **More details:** Review details on the flows where the dataset is used.

Parameters tab:

- **Path:** Full path to the target location.
- **Parameters:** Indicates the number of parameters in the dataset.
- **Type:** Type of the pattern.
- **Rule:** Rule applied to the pattern.

View for Recipes

Contents:

- *Recipes in the Canvas*
 - *Icon context menu*
 - *Recipes flagged for review*
 - *Recipe Details*
 - *Details options*
 - *Recipe tab*
 - *Data tab*
 - *Target tab*
-

For each recipe in Flow View, you can review or edit its steps or create new recipes altogether. You can also create references to the recipe, modify outputs, and create new recipes off of the recipe.

Recipes in the Canvas



Figure: Recipe icon

Icon context menu

The following menu options are available when you select the plus icon next to the recipe:

- **Add new recipe:** Add a new recipe extending from the current recipe. This new recipe operates on the outputs of the original recipe.
- **Create Output to run:** Create an output for the recipe. See *Create Outputs*.
- **Create Reference Dataset:** Create a reference dataset of the recipe. For more information, see *View for Reference Datasets*.
- **Append Join:** Add a join step as the new last step to the recipe. See *Join Window*.
- **Append Union:** Add a union step as the new last step to the recipe. See *Union Page*.

Recipes flagged for review

You can flag a step or steps in a recipe that requires additional review. Whenever you flag steps or a reference dataset for review, the flow view page is highlighted with a warning icon on each recipe and reference dataset node.

NOTE: A recipe that has steps flagged for review cannot be executed. Additional limitations are placed on editing the recipe when some steps are flagged. For more information, see *Flag for Review*.

Actions:

If you hover the mouse over any recipe and click the warning indicator, the corresponding step that has pending for review icon is highlighted in the Details panel.

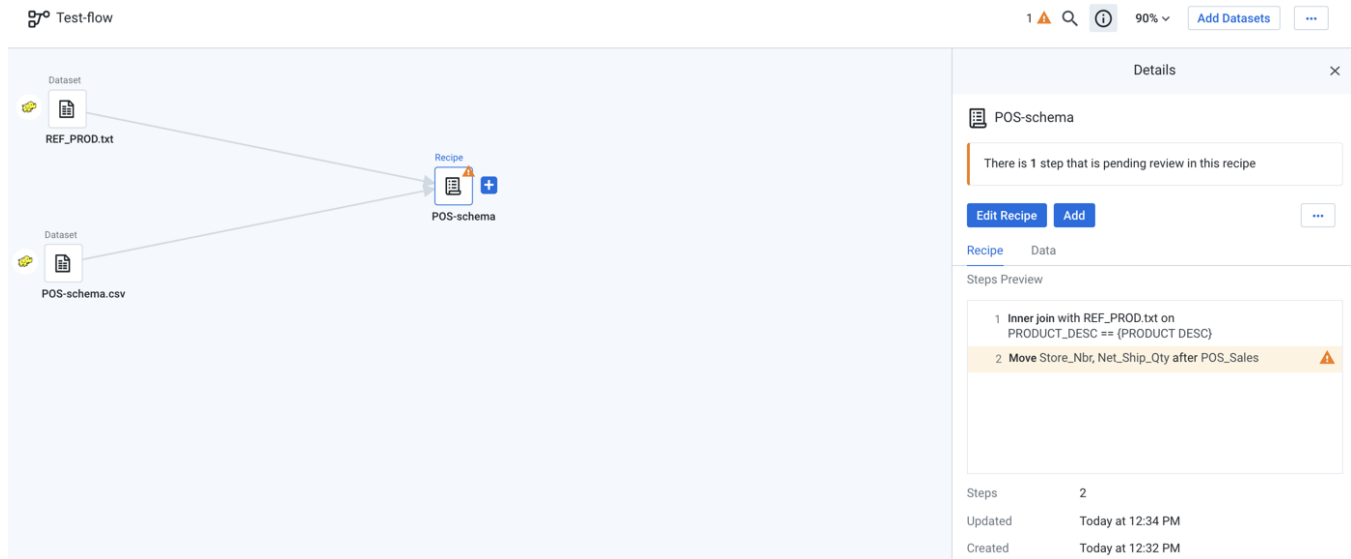


Figure: Recipe icon indicator

NOTE: The Flow View page header summarizes the total number of flagged steps and recipes that are pending for review.

Recipe Details

When you select a recipe:

- You can create an output object.
- You can create a reference object.
- The following options are available in the context panel.

Details

×

POS-r01

Edit Recipe

Add new Recipe

...

Recipe

Data

Steps Preview

1 Union with POS-r02.txt, POS-r03.txt

2 Inner join with REF_PROD.txt on Item_Nbr == ITEM_NBR

3 Delete ITEM_NBR1

4 Inner join with REF_CAL.txt on Daily == Day

5 Delete Day

Steps

5

Updated

Today at 12:18 PM

Created

Today at 12:01 PM

Figure: Recipe view

NOTE: If you flag a step for review, the corresponding step is highlighted with a warning icon and displayed in the Details panel. The warning icon in the Details panel header shows the total count of steps that are pending review.

Details options

The following options are available in the details context menu when you select a recipe.

- **Edit Recipe:** Open the recipe and begin editing. See *Transformer Page*.
- **Add:**

- **Recipe:** Add a recipe for this dataset.
 - **Branch Recipe:** Add a new recipe branching from the recipe. This new recipe operates on the outputs of the recipe from which it was branched.
 - **Output:** If the recipe does not have an output, you cannot run a job for it. Select this option to create a default output for your recipe.
 - **Reference:** Add a reference dataset from this recipe. A reference dataset can be used in other flows. For more information, see *View for Reference Datasets*.
- **Join:** Join this dataset with another recipe or dataset. If this dataset does not have a recipe for it, a new recipe object is created to store this step.
- **Union:** Union this dataset with one or more recipes or datasets. If this dataset does not have a recipe for it, a new recipe object is created to store this step.
- **Edit name and description:** (Available to flow owner only) Change the name and description for the object.
- **Append Join:** Add a join step as the new last step to the recipe.
- **Append Union:** Add a union step as the new last step to the recipe.
- **Assign Target to Recipe:** As needed, you can create a target and assign it to this recipe. For more information, see *Create Target*.
- **Remove Target:** Remove the currently assigned target from this recipe.

NOTE: You can toggle between **Assign Target to Recipe** and **Remove Target** to assign a target and to remove the target from the recipe.

- **Change input:** Change the input dataset associated with the recipe.

NOTE: This action substitutes only the primary input from a recipe, which does not include any datasets that are integrated from joins, unions, lookups, or other multi-dataset options.

Tip: You can swap in dynamic datasets for static datasets, if needed. This feature may not be enabled in your environment. See *Miscellaneous Configuration*.

- **Make a copy:** Create a copy of the recipe and its related objects. You can create the copy with the same inputs or without inputs at all. The copied recipe is owned by the user who copied it.

NOTE: The copied recipe is independent of the source recipe. Optionally, you can duplicate the datasets from the original recipe in the copy.

- **Move:** Move the recipe to a different flow, or create a new flow to contain it.
- **Download Recipe:** Download the recipe in Wrangle format to your local desktop.
- **Delete:** Delete the recipe.

Tip: When a recipe is deleted, all samples associated with the recipe are also removed, which may significantly reduce the total volume of storage you are using.

This step cannot be undone.

Tip: You can also right-click the recipe to view all the menu options.

Recipe tab

Preview the first steps in the recipe.

Key Fields:

- **Steps:** Total count of the steps in the recipe.

Data tab

Preview the data as reflected by the recipe.

NOTE: To render this data preview, some of the data must be loaded, and all steps in the recipe must be executed to generate the preview. Some delays may be expected.

Key Fields:

- **Size:** Total count of columns and data types in the dataset.

Target tab

When a target has been assigned for this recipe, you can review its schema information in the Target tab. This tab appears only after a target has been assigned to the recipe.

To remove the current target, select **Remove Target** from the context menu.

Columns:

- **Position:** Left-to-right position of the column in the target.
- **Name:** Name of the column in the target.
- **Type:** Trifacta data type of the column in the target.

For more information, see *Overview of RapidTarget*.

View for Reference Datasets

Contents:

- *Create reference dataset*
 - *Icon context menu*
 - *Details options*
 - *Add reference dataset in another flow*
 - *Icon context menu*
 - *Details options*
-

A **reference dataset** is a reference to a recipe's output, which can be added to a flow other than the one where the recipe is located.

Create reference dataset

Click the plus icon next to the recipe and select **Create Reference Dataset**.

NOTE: A reference dataset is a read-only object in the flow where it is referenced. A reference dataset must be created in the source flow from the recipe to use. For more information, see *View for Recipes*.



Figure: Reference Dataset icon in source flow

Icon context menu

The following options are available when you right-click the reference dataset:

- **View details:** Explore details of the reference dataset. See *Dataset Details Page*.
- **Add to Flow:** Click to add the reference dataset to a new or existing flow.
- **Edit name and description:** Change the name and description for the object.
- **Delete Reference Dataset:** Remove the reference dataset from the flow.

Deleting a reference dataset in the source flow causes all references to it to be broken in the flows where it is referenced. These broken references should be fixed by swapping in new sources.


Details options

The following options are available in the details context menu when you select a reference dataset.

- **Add to Flow:** Click to add the reference dataset to a new or existing flow.
- **Edit name and description:** (Available to flow owner only) Change the name and description for the object.
- **Delete Reference Dataset:** Remove the reference dataset from the flow.

Details


×

 POS-r01

Add to Flow...

...

Data Preview

 Daily	# Item_Nbr	# Store_Nbr	# WM_
2013/02/08	381000	1	201050
2013/02/07	325000	2	201049
2013/02/07	325000	2	201049
2013/02/07	403000	2	201049
2013/02/07	449000	2	201049
2013/02/07	490000	2	201049
2013/02/07	560000	2	201049

Updated

Today at 10:18 AM

Created

Today at 10:18 AM

Used in

0 Flows [More details](#)

Figure: Reference Dataset view

The following fields appear in the right panel.

Key Fields:

Used In: Indicates the number of flows where the reference appears. If this number is greater than one, click **More details** to review the flows. See *Dataset Details Page*.

More details: Review details on the flows where the dataset is used.

Add reference dataset in another flow

After you have created a reference dataset, you can use it other flows.

Options:

- In the source flow, select the reference dataset, and click **Add to Flow** in the Details options.
 - See above procedures on creating a reference dataset.
 - From the **Add - x to** dialog, where **x** is the name of the reference object, select the required flow or click **Create new flow** to add the reference dataset to the flow.
 - For more information on creating new flows, see *Create Flow Page*.
- In a target flow, you add a reference dataset like other datasets. In Flow View, when you add a dataset, reference datasets are listed under the Reference tab. For more information, see *Flow View Page*.

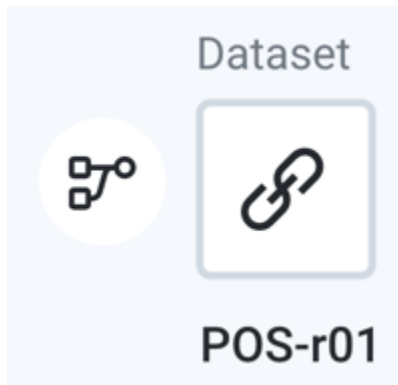


Figure: Reference Dataset icon in another flow

Icon context menu

The following menu options are available when you select the plus icon next to the dataset:

- **Add new recipe:** Add a new recipe extending from the reference dataset.
- **Add Join:** Add a join step as the new last step to the recipe. For more information, see *Join Window*.
- **Add Union:** Add a union step as the new last step to the recipe. For more information, see *Union Page*.

Details options

The following options are available in the details context menu when you select the reference dataset.


- **Add:**
 - **Recipe:** Add a recipe for this dataset.
 - **Join:** Join this dataset with another recipe or dataset. If this dataset does not have a recipe for it, a new recipe object is created to store this step.
 - **Union:** Union this dataset with one or more recipes or datasets. If this dataset does not have a recipe for it, a new recipe object is created to store this step.
- **View in library:** Review details on the flows where the dataset is used.
- **Go to original reference:** Open the flow containing the original reference for this dataset.
- **Remove from Flow:** Remove the reference dataset from the flow.
 - All dependent flows, outputs, and references are not removed from the flow. You can replace the source for these objects as needed.

NOTE: References to the deleted dataset in other flows remain broken until the dataset is replaced.

Tip: You can also right-click the reference dataset to view all the menu options.

Details

×

 POS-schema

Add

View in library

...

Data Preview

# Item_Nbr	# Store_Nbr	# WM_Week	⌚ Daily
381000	1	201050	2013/02/08
325000	2	201049	2013/02/07
325000	2	201049	2013/02/07
403000	2	201049	2013/02/07
449000	2	201049	2013/02/07
490000	2	201049	2013/02/07
560000	2	201049	2013/02/07
572000	2	201049	2013/02/07

Source Flow

[Test_flow](#)

Updated

Today at 2:53 PM

Created

Today at 2:52 PM

Figure: View for referenced dataset in a new flow

NOTE: Reference datasets marked with a red dot no longer have a source dataset for them in the other flow. These upstream dependencies should be fixed. See *Fix Dependency Issues*.

The following fields appear in the right panel.

Key Fields:

Source Flow: Flow that contains the dataset. Click the link to open the Flow View page for that dataset.

View for Unstructured Datasets

An **unstructured dataset** is an imported dataset that does not contain any initial parsing steps. All parsing steps must be added through recipes that are applied to the dataset. During the import process, you disable the initial steps that are applied to imported datasets. Instead, these steps are added as the first steps of the auto-generated recipe that appears with the dataset in Flow View.

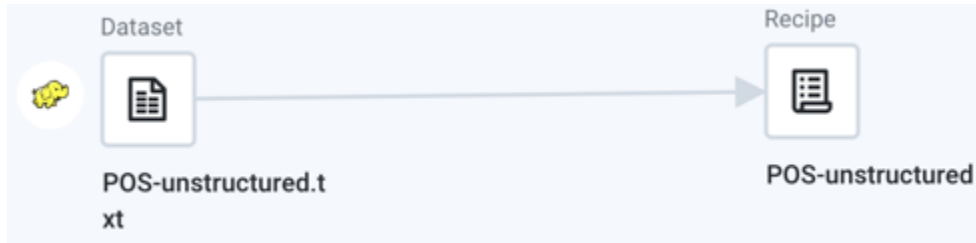


Figure: Unstructured Dataset icons

Icon context menu

The following menu options are available when you click plus icon next to the unstructured dataset:

- **Add new recipe:** Add a recipe for this dataset.
- **Add Join:** Join this dataset with another recipe or dataset. If this dataset does not have a recipe for it, a new recipe object is created to store this step. See *Join Window*.
- **Add Union:** Union this dataset with one or more recipes or datasets. If this dataset does not have a recipe for it, a new recipe object is created to store this step. See *Union Page*.

Details options


- **Add:**
 - **Recipe:** Add a recipe for this dataset.
 - **Join:** Join this dataset with another recipe or dataset. If this dataset does not have a recipe for it, a new recipe object is created to store this step.
 - **Union:** Union this dataset with one or more recipes or datasets. If this dataset does not have a recipe for it, a new recipe object is created to store this step.
- **View dataset details:** Explore details of the dataset. See *Dataset Details Page*.
- **Edit name and description:** (Available to flow owner only) Change the name and description for the object.
- **Remove from Flow:** Remove the dataset from the flow.
 - All dependent flows, outputs, and references are not removed from the flow. You can replace the source for these objects as needed.

NOTE: References to the deleted dataset in other flows remain broken until the dataset is replaced.

Tip: You can also right-click the unstructured dataset to view all the menu options.

Details

×

 POS-unstructured.txt

Add

View dataset details

...

Data Preview

```

Store_Nbr,Item_Nbr,WM_Week,Daily,Whse_Nbr,Whse_Name,
1,381000,201050,2013/02/08,0,Acme
Warehouse,7.00,7,4.97,0,0,Regular,24,.98,.71,.98
2,325000,201049,2013/02/07,0,Acme
Warehouse,.00,0,.00,0,7,Rollback,504,1.24,.93,1.24
2,325000,201049,2013/02/07,0,Acme
Warehouse,10.62,9,8.37,0,0,Regular,504,1.24,.93,1.24
2,403000,201049,2013/02/07,0,Acme
Warehouse,.00,0,.00,0,-1,n/a,432,1.24,.93,1.24
2,449000,201049,2013/02/07,0,Acme
Warehouse,7.00,6.5,5.58,72,0,Regular,456,1.24,.93,1.24

```

Type	HDFS
Location	hdfs:///trifacta/uploads/1/33ecce5d-b242-4025-ad07-e11c14520fb6/POS-r01.txt
File Size	285.95kB
Updated	Today at 2:54 PM
Created	Today at 2:54 PM
Used in	1 Flow More details

Figure: Unstructured Dataset view

Key Fields:

Data Preview: In the Data Preview window, you can see a small section of the data that is contained in the imported dataset. This window can be useful for verifying that you are looking at the proper data.

Tip: Click the preview to open a larger dialog, where you can select and copy data.

Type: Indicates where the data is sourced or the type of file.

Location: Path to the location of the imported dataset.

File Size: Size of the file. Units may vary.

View for Outputs

Contents:

- *Jobs tab*
 - *Manual Settings tab*
 - *Scheduled Settings tab*
-

Associated with each recipe is one or more outputs. These publishing destinations can be configured through the context panel in Flow View. Through outputs, you can execute and track jobs for the related recipe.

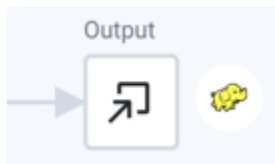


Figure: Output icon

In the context panel, the following options are available:


Run: Click **Run** to queue for immediate execution a job for the manual destinations. You can track the progress and results of this task through the Jobs tab.

In the context menu:

Delete Output: Remove this output from the flow. This operation cannot be undone. Removing an output does not remove the jobs associated with the output. You can continue working with those executed jobs. See *Jobs Page*.

Jobs tab

Details



POS 2021

Run


...

Jobs (4)

Manual settings

Scheduled settings

Latest job



Job 1897511 • Completed

Finished Yesterday at 1:09 AM

⌚ Daily	# Item_Nbr	# Store_Nbr	# WM_Wo
2/7/21	406000	4	201049
2/1/21	406000	14	201049
2/7/21	406000	21	201049
2/7/21	406000	21	201049
2/1/21	406000	28	201049
2/7/21	406000	42	201049
2/1/21	406000	46	201049
2/8/21	406000	46	201050
2/7/21	406000	50	201049
2/7/21	406000	50	201049

38 columns 7956 rows

Download

View details

The preview above shows the current data in the job destination. It is not a snapshot of the data as it exists in the source system.

Figure: Jobs tab

Each entry in the Jobs tab identifies a job that has been Queued, Completed, or In Progress for the selected output. You can track the progress, success, or failure of execution. If you have executed no jobs yet, the Jobs tab is empty.

For the latest job:

- You can preview the job results. Click the Preview pane to open the results in a separate window.

NOTE: The Preview pane reflects the state of the data at the location specified for the output. If other jobs are also writing to this location, the state of the data may not reflect the output for this specific job.

NOTE: This section is not displayed if the job fails. The Preview may not be available if errors occur.

- To download the results from the output location, click the **Download** button.

NOTE: This button may not be available for some successful jobs.

- To view job details, click **View details**. For more information, see *Job Details Page*.

You can also view the previous jobs that have been executed for the selected output.

Tip: When you hover the mouse over a job link, you can review details of the job in progress. For more information, see *Overview of Job Monitoring*.

When a job has finished execution, click the link to the job to view results.

Actions:

For a job, you can do the following:

Click the job link to view the results of your completed job. For more information, see *Job Details Page*.

Cancel job: Select to cancel a job that is currently being executed

Delete job: Delete the job from the platform.

Deleting a job cannot be undone.

NOTE: This feature may not be enabled in your instance of the platform. For more information, please contact your Trifacta Administrator. See *Miscellaneous Configuration*.

Download logs: Download the logs for the job. If the job is in progress, log information is likely to be incomplete.


Tip: When jobs fail, the downloaded package includes additional configuration files and service logs to assist in debugging job execution issues. For more information, see *Support Bundle Contents*.

Manual Settings tab

The Manual Settings tab contains configured outputs for manual execution of jobs through the application interface.

Details

×

 POS 2021

Run

...

Jobs (4)

Manual settings

Scheduled settings

Options

Edit

Profiling

yes

Ignore recipe errors

yes

Publishing actions

Create-CSV

Location

tfs://[redacted]/POS
2021.csv

Settings

no compression, single file, with headers, with quotes, with d...

Create-JSON

Location

tfs://[redacted]POS_202
1/POS. <> year json

Settings

no compression, multiple files

Figure: Manual Settings tab

Options:

Environment: The running environment where the job is configured to be executed.

Profiling: If profiling is enabled for this destination, this value is set to `yes`.

Ignore recipe errors: When enabled, non-fatal errors encountered in a recipe during job execution are ignored. These errors are available for review in the Job Details page.

- To create a new manual destination, click **Add**.
- To edit the current manual setting, click **Edit**.

For more information on these settings, see *Run Job Page*.

Publishing actions:

For the manual destination, this section outlines any additional publishing actions to be taken when generating the output.

Location: Full path to the target location. If output parameters have been created for the destination, you can review their names in the path. For more information, see *Overview of Parameterization*.

SQL scripts:

NOTE: The SQL Scripts feature may need to be enabled in your environment by an administrator.

Before or after a job, you can specify one or more SQL scripts to execute. These scripts can be used for tasks like staging data for job execution or to updating an audit table on job execution. For more information, see *SQL Scripts Panel*.


Scheduled Settings tab

If a schedule has been defined for the flow, you can define a separate set of destinations, which are populated whenever the schedule is triggered and the associated recipe is successfully executed. If any input datasets are missing, the job is not run.

NOTE: The Scheduling feature may need to be enabled in your environment by an administrator.

Details

×

 POS-r01

Run

...

Jobs (18)

Manual settings


Scheduled settings

Options

Edit

Environment	Dataflow
Profiling	no
Ignore recipe errors	yes

Publishing actions

 Create-CSV

Location

gs://[redacted]/jobrun/POS-r01.csv

Settings

no compression, multiple files, with quotes, with deli...

SQL scripts

No SQL scripts

Add script

Figure: Scheduled Settings tab

NOTE: Flow collaborators cannot modify publishing destinations.

See *Add Schedule Dialog*.

For more information, see *Overview of Automator*.

Share Flow Dialog

You can manage access to a flow for other users through the Share Flow dialog. In Flow View, select **Share** from the context menu.

Tip: If groups have been enabled in your instance of the Trifacta platform, you can share flows and connections to LDAP groups. For more information, see *Configure Users and Groups*.

When you grant another user access to one of your flows, you both can work on the objects of the flow. You can take turns editing the recipes, which allows the team to more rapidly complete the work.

NOTE: When a user is given access to a flow, that user is considered a **collaborator** on the flow and has a smaller set of permissions than the **owner** of the flow.

NOTE: Any user may be given access to a flow. However, this user must have access to the underlying data. If the imported dataset is accessed from a private location, the user cannot access datasets in the shared flow. For more information, see *Share Connection Dialog*.



Figure: Share Flow Dialog

Find users:

To add users as collaborators in your flow, start typing the name of a user with whom you'd like to collaborate. Select the user. Repeat this process to add multiple users.

Tip: You can paste a comma-separated list of email addresses to share to multiple users at the same time.

NOTE: For privacy reasons, search may not be available in some environments.

Set access level:

For each user, you can configure their level of access to the shared flow:

NOTE: You cannot set a user's access to a level that is higher than the limit set for the user at the workspace level. For example, if the user has Viewer access to flows at the workspace level, you cannot make the user an Editor on your flow.

NOTE: Workspace administrators have Owner-level access to all flows in the workspace. You do not need to share flows with them.

- Editor:
 - Modify the flow.
 - Share the flow.
 - All Viewer privileges.
- Viewer:
 - Read-only access to the flow.
 - Run jobs in the flow.

For more information, see *Overview of Sharing*.

To save your changes, click **Save**.

Each selected user now can access the flow through their flows page. See *Flows Page*.

Manage Access Tab

When you grant another user access to one of your flows, you both can work on the objects of the flow. You can take turns editing the recipes, which allows the team to more rapidly complete the work.

NOTE: When a user is given access to a flow, that user is considered a **collaborator** on the flow and has a smaller set of permissions than the **owner** of the flow.

NOTE: Any user may be given access to a flow. However, this user must have access to the underlying data. If the imported dataset is accessed from a private location, the user cannot access datasets in the shared flow. For more information, see *Share Connection Dialog*.

Through this tab, you can invite one or more collaborators to the flow, so that you may work together on the same

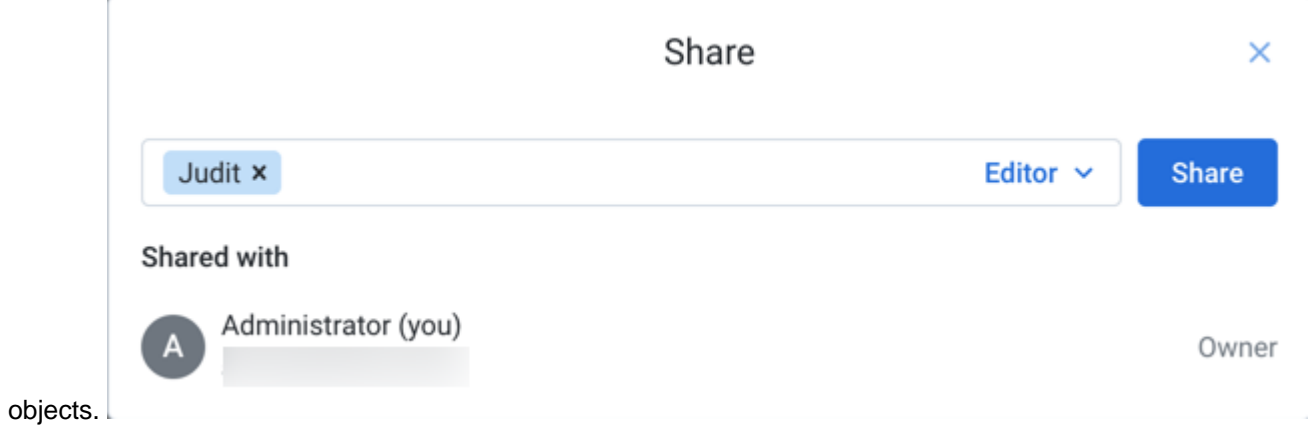


Figure: Manage Access Tab

- To add users as collaborators in your flow, start typing the name of a user with whom you'd like to collaborate. Select the user. Repeat this process to add multiple users.

Tip: You can paste a comma-separated list of email addresses to share to multiple users at the same time.

NOTE: For privacy reasons, search may not be available in some environments.

- To add a group of users as collaborators, select an entry that includes **(Group)**. Any user in the group has the same permissions as if you shared the flow with the user directly.

NOTE: This feature is in Beta release.

NOTE: This feature must be enabled. For more information, see *Configure Users and Groups*.

- Select the access level privilege from the drop-down on the right side of the textbox.

To save your changes, click **Save**.

Each selected user now can access the flow through their flows page. See *Flows Page*.

NOTE: Collaborators have a reduced set of privileges on the flow. For example, they cannot edit the flow name or description or delete it. See *Overview of Sharing*.

Change Dataset Dialog

Through the Flow View page, you can change the source that is used for your dataset. In this manner, you can apply the same recipe across datasets with the same schema. When the source dataset has been changed, a new sample is automatically generated for you.

For example, you build your recipe for a week's worth of sales data, which is sourced from an imported dataset based on a CSV called, `Week01-Sales.csv`. When the next week's source data is dropped in the appropriate directory, you can:

1. Import the new dataset,
2. Edit the recipe,
3. Change the source to the new file, and
4. Execute a job immediately to process the new week of data.

NOTE: A dataset source can be an imported dataset, a reference dataset, or a recipe. Subsequent changes to the source data affect your dataset in development.

Notes and Limitations:

- If there are differences between the schemas of the source and the new source, your recipe is likely to break on the dataset when the new dataset is selected.
- You can swap your original source dataset with an imported dataset, reference dataset, or a recipe. If needed, you can swap back to the original source at any time.
- If you have enabled relational connections, swapping relational sources may not work if they are from different database vendors.
- Data-dependent transforms, such as `header` and `valuestocols`, use the data that was present in the sample at the time that they were added to the recipe. This fact can cause unexpected changes or breakages when the recipe is applied to another source.
- You cannot undo or redo source swaps.

Steps:

1. To change a data source, open the flow containing it.
2. In Flow View, you can:
 - a. Click the imported dataset icon. Then, click **Replace**.

NOTE: This action removes the imported dataset and all links (edges) coming out of it. The replacement must be reconnected with any downstream objects.

- b. Click the recipe icon. Then, click **Change input**.

NOTE: This action substitutes only the primary input from a recipe, which does not include any datasets that are integrated from joins, unions, lookups, or other multi-dataset options.

3. Select the new source:

NOTE: You can select data from any flow to which you have access. Changes to the source are inherited.

Change input for POS-r01

×

Q Search...

All (17)

Imported (16)

Reference (0)

Recipe (1)

NAME	SOURCE	LAST UPDATED
<div></div> <div>POS-r01</div>	<div></div> <div>2013 POS (this flow)</div>	Today at 10:18 AM
<div></div> <div>twitter.json</div>	HDFS	Today at 12:12 PM
<div></div> <div>POS-r01.txt</div>	HDFS	Today at 12:11 PM
<div></div> <div>POS-schema.csv</div>	HDFS	Today at 10:17 AM
<div></div> <div>POS-r02.txt</div>	HDFS	Today at 10:17 AM
<div></div> <div>POS-r03.txt</div>	HDFS	Today at 10:17 AM

Cancel

Change

Figure: Change Dataset Dialog

- a. If replacing an imported dataset, you can import new data as the replacement. Click **Import Datasets**. For more information, see *Import Data Page*.
4. Click **Replace** or **Change**.
5. Your dataset is now using the selected dataset as its source, and the current recipe in the Transformer page is applied to the new source.

Add Schedule Dialog

To add a schedule to your flow, click the drop-down menu in the Flow View page, and select **Schedule**.

NOTE: Flow owners can create schedules. Other users must have the appropriate privileges to create schedules for the flow.

NOTE: By default, when scheduled or API jobs are executed, no validations are performed of any writesettings objects for file-based outputs. Issues with these objects may cause failures during transformation or publishing stages of job execution. Jobs of these types should be tested through the Trifacta application first. A workspace administrator can disable the skipping of these validations.

NOTE: For a dataset with parameters, scheduled times are used when resolving date range parameters.

NOTE: Do not schedule executions through Flow View in a Prod instance when Deployment Manager is enabled.

- Schedules defined in Flow View are applied to Active and Non-Active releases in Production environments.
- If the scheduled release is deactivated, the schedule still exists, and the jobs are executed on an flow that is now out-of-date.
- For more information, see *Overview of Deployment Manager*.

Add Schedule
×

Scheduling Options

Timezone

America/Los_Angeles ▼

Frequency

Weekly ▼

on

Sunday ▼

at

12:00

AM ▼

Add

Variables

Set variable values for this schedule's executions

< > regionNum

01

Cancel

Save

Figure: Add Schedule dialog

Scheduling Options

- **Timezone:** Select the timezone to apply to the schedule.
 - To use UTC time zone, select `UTC` in the drop-down. For more information, see *Supported Time Zone Values*.
- **Frequency:**
 - **Hourly, Daily, Weekly, Monthly:** Run the schedule at the specified moment for the interval.

Tip: For drop-downs showing days of the week or days of the month, you can click multiple values to select them.

- **cron:** Set the schedule according to cron syntax.
 - Time zone settings set in the drop-down are used with the cron schedule.
 - For more information, see *cron Schedule Syntax Reference*.

To add another trigger for the flow's schedule, click **Add**.

To create the schedule, click **Save**.

Variables

If your flow contains one or more variable parameters, you can apply overrides to any variables. When the scheduled job is executed, the variable value is applied to job at runtime.

For each listed variable, you can modify its value.

For more information, see *Overview of Parameterization*.

Manage Flow Notifications Dialog

When email notifications are enabled, flow owners and collaborators can configure the delivery of emails to interested stakeholders based on the success or failure of jobs executed within this flow. From the flow menu, select **Email notifications**.

NOTE: This feature requires access to an SMTP server to send emails. For more information, see *Enable SMTP Email Server Integration*.

Settings Tab

In the Settings tab, you configure the types of jobs that generate success or failure emails for jobs executed in this flow.

Manage notifications for 2013 POS

Settings

Watchers

Receive job failure emails

Only from scheduled jobs

Receive job success emails

Never

Cancel

Save

Figure: Manage Flow Notifications - Settings tab

These settings apply to jobs executed on the flow. Default settings are inherited from the workspace settings. For more information, see *Workspace Settings Page*.

- **Receive job failure emails:** Select the type of jobs that generate emails when they fail.

Setting	Description
From any job	Emails are generated for any type of job from this flow when it fails.
Only from scheduled jobs	Emails are generated when a scheduled job from this flow fails.
Only from manual jobs	Emails are generated when a manual job from this flow fails. <div>Tip: Jobs executed via API are manual jobs.</div>
Never	Emails are never generated when jobs from this flow fail.

- **Receive job success emails:** Select the type of jobs that generate emails when they succeed. See above for options.

Watchers Tab

In the Watchers tab, you can add or remove email addresses for interested stakeholders to receive email notifications.

Tip: Any flow collaborator can add or remove watchers from this list.

Manage notifications for 2013 POS

Settings

Watchers

Watchers are the people receiving email notifications about this flow activity.

Flow collaborators

Administrator ()

Watching

☒

Others

Non-collaborators receive notifications but are not able to access the flow or see job details

Enter a new email

Add

joe@example.com

Cancel

Save

Figure: Manage Flow Notifications - Settings tab

Flow collaborators:

By default, the flow owner and all collaborators receive any email notifications for any job executed for this flow.

Click the checkbox next to the name and email address to toggle whether that collaborator receives flow email notifications.

Others:

For non-collaborators, you can insert email addresses to receive email messages for jobs from the flow. Enter a valid email address and click **Add**.

To remove a non-collaborator, click the Trash icon next to the address.

Tip: Email recipients can remove themselves from receiving notifications on flow jobs using a link at the bottom of the email.

To apply your changes, click **Save**.

Manage Parameters Dialog

Contents:

- *Parameters Tab*
 - *Overrides Tab*
 - *A note on upgraded parameters*
 - *Manage Parameters for Plans*
-

Within a flow, you can create and manage flow parameters, including specifying override values. From the flow menu, select **Parameters**.

- A **flow parameter** is a reference token that can be invoked from within the flow. A flow parameter can be:
 - A string value
 - A Trifacta parameter
 - A regular expression
- Where it is invoked, the default value for the parameter is applied or, if an override value has been set, the override value is applied.
- For more information, see *Overview of Parameterization*.

You specify flow parameters at the flow level. They can be invoked in any recipe within the flow.

Tip: Override values apply to all parameters in the flow that share the same name, even if they are output object parameters.

Tip: Flow parameters can be inherited from upstream flows. For example, if you create a reference dataset that references flow parameters from its flow, those parameters are passed to the downstream flow. While you cannot change the default value for the downstream instance of the parameter, you can apply override values for all recipes in the downstream flow.

Parameters Tab

In the Parameters tab, you can manage the flow parameters in your flow.

NOTE: Non-flow parameters cannot be edited or deleted from the Parameters tab. These parameters must be modified in the interface for the object to which they apply. Overrides can be applied to these types of parameters through the Overrides tab.

Parameters for plans: When reviewing parameters for your plan, you may notice that a parameter has multiple values:

- Parameter values that do not conflict are ok. For example, if the parameter is applied to an imported dataset and to an output object which are attached to different recipes, these values are not in conflict.
- If the parameter values are in conflict, then a warning icon is displayed. For example, if a parameter is applied to an imported dataset and to an output object attached to the same recipe, then the conflicting parameter values must be overridden to fix them.

NOTE: When a parameter has conflicting values from objects in a plan task, you must fix them by applying a single override value. If these values are not fixed, then the plan fails to execute.

×

Manage parameters

Parameters (2)
Overrides (1)

Parameters used in this flow and upstream flows are displayed below. You can also add parameters to use in your recipes.

Name	Source	Value	Add parameter
<> colPrimaryKey	this flow	userId	
<> storeId	this flow	4	

[Learn more about parameters](#)
Close

Figure: Parameters tab

- To create a new parameter, click **Add parameter**.
 - Specify the Name parameter.

NOTE: Name values are case-sensitive. After saving a parameter, you cannot change its name.

- Specify the default Value for the parameter. Examples by parameter type:
 - String literal:

This is my string.

NOTE: Flow parameter values that are literal values are String values. You can convert them to other data types after they have been referenced in your recipe.

- Pattern : Patterns can be used to find matches in your recipes. The following pattern matches for two consecutive digits:

`{digit}{digit}`

For more information on Patterns , see *Text Matching*.

- Regular expression: These patterns also can be used for finding matches. The following pattern matches for two consecutive digits:

`/[0-9][0-9]/`

Regular expressions are a standard method for identifying patterns in data. The syntax is based on *RE2* and *PCRE* regular expressions.

- Click **Save**.
- To edit a parameter's default value, hover over its entry, and click the Pencil icon.

- To delete a parameter, hover over its entry and click the Trash icon.

Deleting a parameter cannot be undone. When you delete a parameter, all recipe steps that reference it are broken.

Tip: If you accidentally delete a flow parameter, you can recreate it with the same case-sensitive name. All references to it should work again.

After you create a parameter, you can insert it into your recipes using the following type of reference:

```
${myRecipeParameterName}
```

For more information, see *Create Flow Parameter*.

Overrides Tab

In the Overrides tab, you can apply override values to any parameters referenced in your flow, including:

- Flow parameters
- output object parameters
- Dataset parameters
- Parameters that are available through reference datasets

NOTE: The parameters listed in this tab have override values applied to them. Parameters that are using their default values in this flow are not listed.

Manage parameters

×

Parameters (2) Overrides (1)

Add an override to replace all variable parameters that share the same name.

Name	Value	Add override
< > storeId	3	

[Learn more about parameters](#)

Close

Figure: Overrides tab

All of the available parameter overrides are listed.

- To apply an override to all references of a parameter within the flow, click **Add override**.
 - Select the name of the parameter.
 - Enter its override value. Click **Save**.
- To edit a parameter override, hover over its entry and click the Pencil icon. Enter a new value.

NOTE: The override value is applied to all subsequent operations in the platform. When a job is submitted to the job queue, any overrides are applied at that time. Changes to override values do not affect jobs that are already in flight.

- To delete an override, hover over its entry and click the Trash icon.

A note on upgraded parameters

If you have upgraded from a version of the product before Release 7.1, any parameters that were defined in the previous version appear grayed out in the Parameters tab.

- The Parameters tab is used for defining new flow parameters, which is a new type of parameter. These parameters can be referenced inside your recipe steps. See above.


NOTE: Parameters that are not of flow parameter type cannot be edited in the Parameters tab. You can apply override values to these parameters through the Overrides tab.

- The Overrides tab is used to specify overrides to the default values for parameters, including parameters defined outside of Flow View.
 - If you add an override using the name of the parameter that was upgraded, the override is applied, based on the matching name.

Manage Parameters for Plans

You can also specify parameter overrides at the plan level. For the recipes in your plan, you can pass in parameter overrides values, which are used during plan runs.

Manage parameters

Parameter	From	Created in	Value
< > region	Run Datasets - WindowsFunctions Flow flowtask-32	 Datasets - Window...	<input type="text" value="02"/> <div><div>Cancel</div><div>Save</div></div>

Close

Figure: Manage Parameters Dialog for plans

In the Manage Parameters dialog, you can review the parameters that are defined for the flows that are part of your plan.

NOTE: You cannot create parameters through Plan View. You must create the parameters within the flows that are part of your plan. After they are created, they are available for overrides through Plan View.

NOTE: In a plan, parameters and their overrides apply only to the flow task from where they are sourced. They do not apply to other flow tasks. They do not apply back to the source flows.

Tip: To pass a flow parameter value from one recipe to another, you can insert the parameter value in a column in a recipe, export the results as a reference dataset, and then ingest that reference into another flow.

Columns:

- **Parameter:** The name of the parameter
- **From:** The source flow in your plan for the parameter
- **Created in:** The object in the source flow where the parameter is defined
- **Value:** The current value for the parameter

Tip: To override the inherited value for the parameter, click the Pencil icon. Then, click **Save**. Whenever the flow is executed as part of this current plan, the override value is used to replace the parameter value and any override that are defined within the flow.


Flow Search Panel


You can search for objects within your flow. In Flow View, click the Magnifying Glass icon at the top of the page.

Search (3) ×

Q POS ×

Datasets × × ▾

 POS-r01

 POS-r02


 POS-r03

Figure: Flow Search panel

Steps:

1. From Flow View, click the **Search** icon.
2. In the Search panel, enter a search term. As you type your search term, the search results are highlighted in yellow in the panel and in the flow canvas.
3. From the **Filter by type** drop-down, you can filter by flow object type:
 - a. Datasets
 - b. Recipes
 - c. Outputs
 - d. References

NOTE: If no object type is specified, then all object types are searched.

Tip: You may select one or more flow object types to search.

Context menu

For each object displayed in the search results, its context menu displays the same options that are available from the object in the canvas, including these additional options:

- **View details:** Open the context panel showing the details of the object.
- **Show in flow:** Center the viewport of Flow View on the object.

The context menu options vary for different types of objects for which you are searching. For more information:

- **Datasets:** See *View for Imported Datasets*.
 - This category includes reference datasets imported from other flows.
- **Recipes:** See *View for Recipes*.
- **Outputs:** See *View for Outputs*.
- **References:** This category covers reference datasets that are created from the recipes in your flow. See *View for Reference Datasets*.

Flow Optimization Settings Dialog

Contents:

- *Enable optimization for jobs from this flow*
 - *General Optimizations*
 - *Column pruning optimization*
 - *Filter optimization*
 - *Databases that Support Pushdown*
 - *Column pruning from source*
 - *Filter pushdown*
 - *Other Databases*
 - *Column pruning from source*
-

In the Flow Optimization Settings dialog, you can configure the following settings, which provide finer-grained control and performance tuning over your flow and its job executions. From the Flow View menu, select **Optimization settings**.

This feature must be enabled at the workspace level. When enabled, the settings in this dialog are applied to the current flow.

- See *Workspace Settings Page*.

These optimizations are designed to improve performance by pre-filtering the volume of data by reducing the columns and rows to the ones that are actually used.

When these filters are enabled, the number of filters successfully applied to a job execution is listed in the Optimization summary in the Job Details page. See *Job Details Page*.

Enable optimization for jobs from this flow

When enabled, the Trifacta application attempts to apply any of the listed optimizations that are enabled to jobs that are executed for this flow.

NOTE: When this option is disabled, then no optimization settings are available.

General Optimizations

The following optimizations can be enabled or disabled in general. For individual data sources, you may be able to enable or disable these settings based on your environment and its requirements .

Tip: These optimizations are applied at the recipe level. They can be applied on any flow and may improve performance within the Transformer page.

Column pruning optimization

When enabled, job execution performance is improved by removing any unused or redundant columns based on the recipe that is selected.

Filter optimization

When this setting is enabled, the Trifacta application optimizes job performance on this flow by pushing data filters to recipes.

Databases that Support Pushdown

Individual types of databases may support one or more of the following pushdowns. Additional restrictions may apply for your specific database.

Tip: These optimizations are applied to queries of your relational datasources that support pushdown. These optimizations are applied within the source, which limits the volume of data that is transferred during job execution.

NOTE: For each relational connection, you can enable the optimization capabilities to improve the flow and its job execution performance. The optimization settings may vary based on the type of relational connections.

Column pruning from source

When enabled, job execution performance is improved by removing any unused or redundant columns from the source database.

Limitations:

- Column pruning optimizations cannot be applied to imported datasets generated with custom SQL.

Filter pushdown

When this setting is enabled, the Trifacta application optimizes job performance on this flow by pushing data filters directly on the source database.

Limitations:

- Filter pushdown optimizations cannot be applied to imported datasets generated with custom SQL.
- Pushdown filters cannot be applied to dates in your relational sources.

NOTE: SQL-based filtering is performed on a best-effort basis. When these optimizations are enabled for your flow, there is no guarantee that they will be applied during job execution.

NOTE: The connection types may or may not be available in your product edition. For more information, see *Connection Types*.

Other Databases

Databases that do not support pushdown may support the following optimization settings.

Column pruning from source

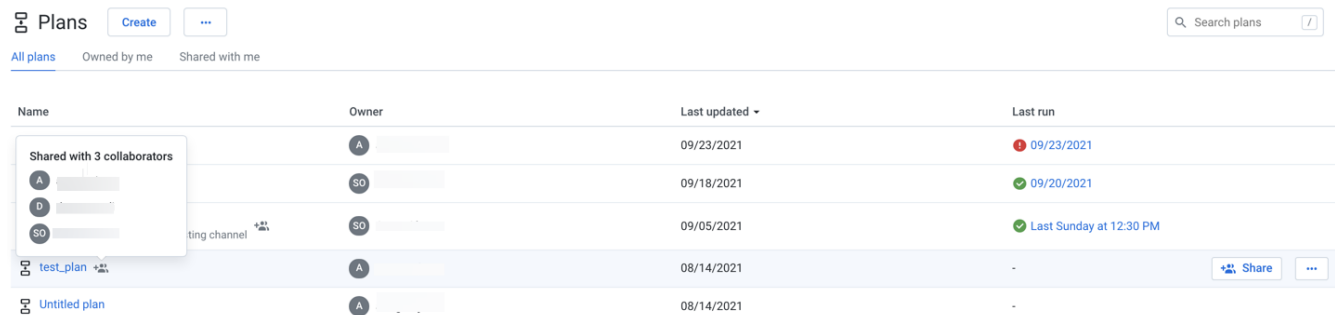
When enabled, job execution performance is improved by removing any unused or redundant columns from the source database.

Plans Page

The Plans page lets you create, review, and manage your plans. A **plan** is a sequence of tasks and the triggers that execute them. Plans can be applied across multiple flows in your workspace.

NOTE: Access to the Plans page in the application and privileges on plans is governed by roles in your workspace. For more information, please contact your workspace administrator.

- To create a new plan, click **Create....** To rename the new plan, click the `Untitled` value at the top of the page.
 - The maximum number of plan tasks may vary depending on the plan that you have licensed. For more information, please contact your Trifacta representative.
- Workspace admins can access all plans in the workspace.



Name	Owner	Last updated	Last run
Shared with 3 collaborators	A	09/23/2021	09/23/2021
	SO	09/18/2021	09/20/2021
	SO	09/05/2021	Last Sunday at 12:30 PM
test_plan	A	08/14/2021	-
Untitled plan	A	08/14/2021	-

Figure: Plans Page

Columns:

- **Name:** The name of the plan.

You can hover over the shared icon link next to the plan to view the name of shared users (up to three shared users) and the total number of shared users. Also, when you click the shared icon link, the share dialog is displayed.

- Click the plan name to review it. See *Plan View Page*.

- **Owner:** Owner of the plan.
- **Last Updated:** Timestamp for the last time that the flow was modified.
- **Last Run:** Timestamp for when the plan was last executed.
 - The displayed icon indicates whether the plan executed successfully or not.
 - Click the link to review details of the run.

Actions:

- **Create:** From the Create menu, choose to create a plan.
 - Enter a name and description for your plan. Click **Create**.
- **Import:** From the context menu, select **Import** to import a plan into this instance. See *Import Plan*.

Tip: You can import multiple plans (ZIP files) through the file browser or through drag-and-drop. Press **CTRL/COMMAND + click** or **SHIFT + click** to select multiple files for import.

- **Search:** To search plan names, enter a string in the search bar. Results are highlighted immediately in the Plans page.
- **Sort:** Some column headers can be selected to sort the display by the column's entries.

Plan options:

The following options are available on the right side of a plan's entry:

- **Share:** Share the plan with other users. See *Share a Plan*.
- **Rename:** Change the name and description of the plan.
- **Email notifications:** Send email notifications on the plan runs. See *Manage Plan Notifications Dialog*.
- **Export:** Export the plan from Trifacta. See *Export Plan*.
- **Delete Plan:** Delete the plan.

Deleting a plan removes all objects contained in the plan. Flows referenced in the tasks of the plan are not touched.

Plan View Page

Contents:

- Top Bar
 - Plan context menu
- Workflow
- Task Execution Rule
- Parallel Tasks
- View for Triggers
- View for Tasks

In Plan View, you design your plan, which includes the building and sequencing of tasks and the triggers that execute your sequence of tasks.

NOTE: Access to this page in the application and privileges on its related objects is governed by roles in your workspace. For more information, please contact your workspace administrator.

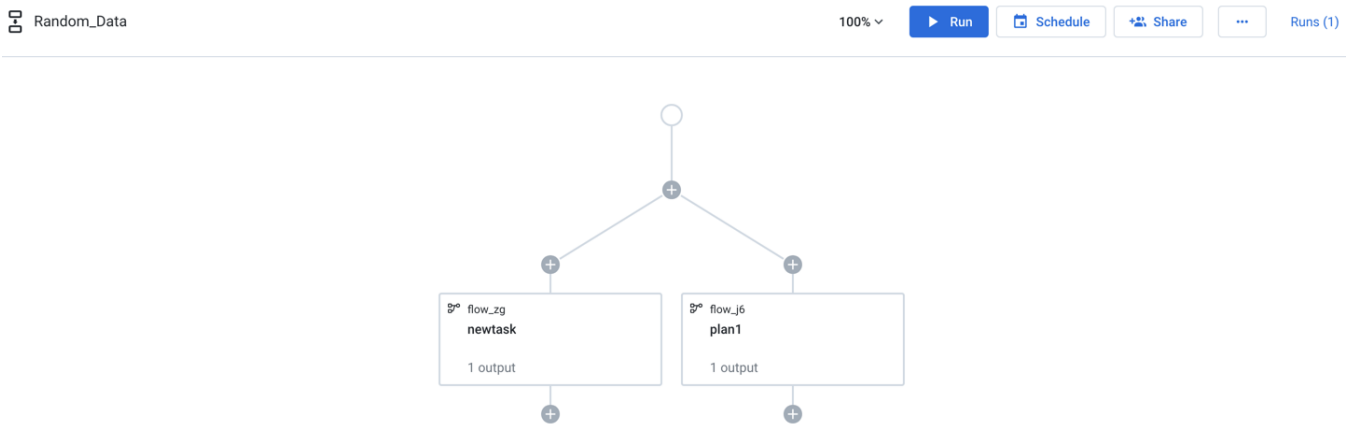


Figure: Plan View Page

The main panel is the plan canvas, where you build your plans.

On the right side is the context panel. Depending on what you select in the plan canvas, a different set of context options is displayed.

Top Bar

Tip: To rename the plan, click the plan name in the top bar.

Zoom Options:

You can zoom the plan canvas to display areas of interest in the plan graph.

The zoom control options are available at the right side of the canvas. The following are the available zoom options:

Tip: You can use the keyboard shortcuts listed in the zoom options menu to make quick adjustments to the zoom level.

- **Zoom in:** Zoom in 10% on the canvas to focus on greater detail.
- **Zoom out:** Zoom out 10% from the canvas to see more of it.
- **Zoom to fit:** Change the zoom level to fit all of the objects of your plan onto the screen.
- **25%, 50%, or 100%:** Change the zoom level to one of the preset levels.

NOTE: By default, the plan view page always opens in Zoom to fit option and it does not remember the previous zoom and position.

Other options:

- **Run:** Run the plan. You can track progress of your plan run. See *Plan Run Details Page*.

Tip: You can apply overrides to flow parameters through the Parameters tab. See *Plan View for Flow Tasks*.

See *Plan Runs Page*.

- **Schedule:** Create or edit the plan schedule with one or more triggers through the right context panel. See "View for Tasks."
- **Share:** Share the plan with other users. See *Share a Plan*.
- **Runs:** The Runs link tracks the current total number of runs that have been queued or executed for this plan. Click this link to track progress on your plan run.

Plan context menu

- **Rename:** Modify the name and description for your plan.
- **Parameters:** You can apply overrides to the recipe parameters for your plan tasks during plan job runs. See *Manage Parameters Dialog*.
- **Email notifications:** Send email notifications on the plan runs. See *Manage Plan Notifications Dialog*.
- **Export:** Export the plan from Trifacta. See *Export Plan*.
- **Delete:** Delete your plan.

Workflow

For more information, see *Create a Plan*.

Task Execution Rule

You can gate the execution of a task based on the completion status of its previous task. Click the line connecting the two tasks and select one of the following options:

- **On success:** Runs if the previous task in the node is successful.
- **On failure:** Runs if the previous task in the node failed.
- **On execution (any status):** Runs the task in the node irrespective of the previous task's status (success or failure).

Parallel Tasks

You can execute one or more tasks in parallel by clicking the plus node icon and selecting one of the following options:

- **Add a parallel node:** Adds a parallel node to the existing node.

- **Add a node:** Adds a node at the bottom of the existing node.

NOTE: Nodes added in a sequence are separated by a plus node. A node added in parallel has two plus nodes separating it from the parent node; one for adding nodes in parallel and other for adding a node in sequence.

View for Triggers

To create a trigger, you specify the time and frequency that the plan is to be executed.

NOTE: Plan owners can create triggers for their plans. Other users must have the appropriate privileges to create triggers for the plan.

< Trigger ×

Schedule

Enable ☒

Timezone

America/Los_Angeles ▼

Frequency

[Add another trigger](#)

Weekly ▼

On

Sunday, Wednesday ▼

×

At

12:10

AM ▼

Monthly ▼

On the

1 ▼

day(s) of the month ×

at

12:00

AM ▼

Save

Figure: Create plan trigger

- **Timezone:** Select the timezone for the time that you are specifying in the trigger's schedule.
 - To use UTC time zone, select `UTC` in the drop-down.
 - For a list of supported timezones, see *Supported Time Zone Values*.
- **Frequency:** Specify how frequently the plan is triggered:
 - **On:** You can specify multiple entries to trigger the plan more frequently.
 - **cron:** Set the schedule according to cron syntax.
 - Time zone settings set in the drop-down are used with the cron schedule.
 - For more information, see *cron Schedule Syntax Reference*.
- To add another trigger, click **Add another trigger**.
- To save your changes, click **Save**.

Trigger

Schedule

Enable ☒ Edit

Weekly: 12:10 AM
On Sunday and Wednesday

Monthly: 12:00 AM
On the 1st

Figure: Saved trigger

- To disable the trigger, click the slider.

NOTE: If you disable a trigger, no new scheduled executions of the tasks in the plan occur. You can still manually trigger plan runs.

- To make changes to the trigger, click **Edit**.

View for Tasks

When you create or select a task, you can modify its settings through the context panel on the right.

Tip: To rename the task, click the task name in the context panel.

Task context menu options:

- **View Flow:** Open the flow. See *Flow View Page*.
- **Edit name:** Modify the name of the task.
- **Delete:** Delete the task from your plan.

Task types

- **Flow task:** Generate all of the defined output objects for a flow. See *Plan View for Flow Tasks*.
- **HTTP task:** Execute a task over HTTP protocol. See *Plan View for HTTP Tasks*.

- **Slack task:** Send a message from the Trifacta application to a specified Slack channel. See *Plan View for Slack Tasks*.
- **Delete task:** Delete files or folders from backend data storage as a task. See *Plan View for Delete Tasks*.

To cancel a plan that is currently running, please do the following:

- In Plan View, click the **Runs (x)** link.
- In the Run Details page, click **Cancel plan run**.

Plan Run Details Page

You can review the details of individual executions of your plan. In Plan View, click the **Runs** link in the upper-right corner.

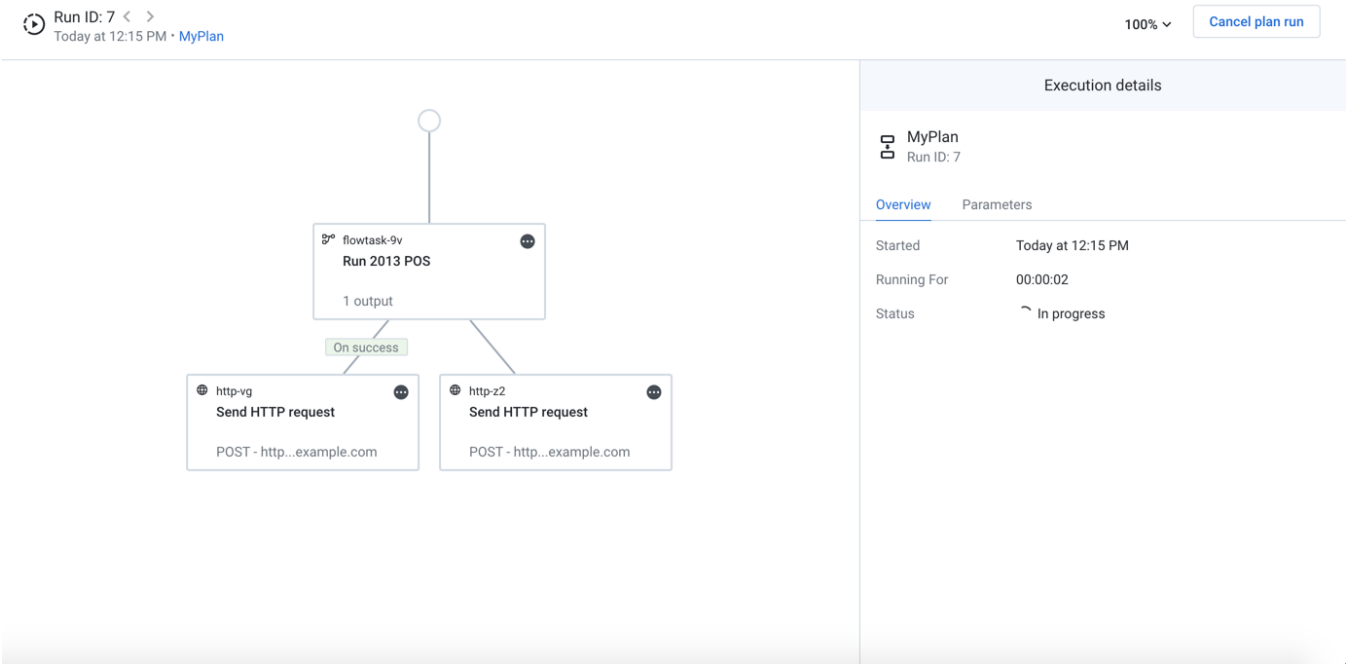


Figure: Run details

The latest plan run is displayed. You can review the progress of individual tasks throughout the plan run.

Tip: You can click the caret next to the Run ID to scroll back and forth to view multiple runs of the plan.

NOTE: When a plan run begins, a snapshot is taken of the plan. Subsequent changes to the underlying flows could impact the outcome of the flow tasks when they are later executed during the plan run.

- You can select individual triggers and tasks to review details of the plan run for that object in the context panel.
- To see other runs for the plan, use the angle brackets next to the timestamp at the top of the screen.

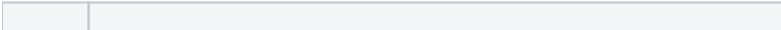
To cancel a plan that is currently running, please do the following:







- In Plan View, click the **Runs (x)** link.
- In the Run Details page, click **Cancel plan run**.

You can track all of the runs across all of your plans. See *Plan Runs Page*.

Task Execution

The following icons indicate the results of the execution of a task:



Icon	Task Status
	Task successful
	Task failed
	Task skipped. Task was not executed due to unmet conditions.
	Task canceled by user
	Task in progress
	Task pending

Share Plan Dialog

You can manage access to a plan for other users through the Share Plan dialog. From the context menu of the Plans page, select **Share**.

Tip: If groups have been enabled in your instance of the Trifacta platform, you can share flows and connections to LDAP groups. For more information, see *Configure Users and Groups*.

Permissions

You can grant permissions to other users to access the plan. When a user is given access to a plan, that user is considered a **collaborator** on the plan and has a smaller set of permissions than the **owner** of the plan. Users must have at least the Viewer role viewer permissions for plans.

For more information, see *Roles Page*.

NOTE: If the user does not have access to the underlying flows of the plan, the plan can still be shared and accessed, but the user cannot edit flows and run the plans.

Actions

Through this Share Plan dialog, you can invite one or more collaborators to the plan, so that you may work together on the same objects.

- **Add:** Add users or email addresses of users with whom you would like to share the plan.
 - To add users as collaborators in your flow, start typing the name of a user with whom you would like to share the plan.
 - Select the user.
 - Specify the privilege level of the user to whom you are sharing. For more information on sharing privileges, see *Overview of Sharing*.
 - Repeat this process to add multiple users.
- **Save:** Click **Save**.
- **Delete:** Select the user or email address in the search by name or email field and click delete using your keyboard.
- **Cancel:** To cancel sharing, click **Cancel**.

Each selected user now can access the plans through the plans page. See *Plans Page*.

Manage Plan Notifications Dialog

This section provides an overview of sending email notifications to plan owners and collaborators based on the results of plan runs.

When email notifications are enabled, plan owners and collaborators can specify the list of email recipients, based on the status of execution of plans. From the context menu of the Plans page, select **Email notifications**.

- Users who receive notifications for specific plans are considered plan watchers.
 - By default, all collaborators receive notifications about plan run failures.
 - If plan collaborators have only view permissions, they may not be able to edit the recipients.
 - You cannot enable or disable email notifications at the plan level. Workspace administrators can enable or disable email notifications for plans.
- For more information, see *Workspace Settings Page*.

NOTE: This feature requires access to an SMTP server to send emails. For more information, see *Enable SMTP Email Server Integration*.

Tip: Email recipients can remove themselves from receiving notifications on plan runs using a link at the bottom of the email.

Email notifications

Send email notifications when this plan runs

Add recipients

Add

Recipient	Notify
<div><div>TU</div><div>Test User 4147080222</div><div>4147080222@trifacta.com</div></div>	<div>On failure</div> <div></div> <div></div>
<div><div>A</div><div>Administrator</div><div>admin@trifacta.local</div></div>	<div>On failure</div> <div></div> <div></div>

Done

Figure: Manage Plan Notifications

NOTE: You can enable or disable the **Send email notifications when this plan runs** option to enable or disable plan email notifications. If this option is disabled, the below options are not available.

For more information, see *Email Notifications Page*.

In the Email notifications dialog, plan owner and collaborators can add the email addresses to stakeholders to receive notifications based on the plan run status:

- **On success:** Emails are generated if the plan run succeeded.
- **On failure:** Emails are generated if the plan run failed.
- **On execution (any status):** Emails are generated whether the plan run succeeded or failure.

Actions:

- **Add:** Add email addresses in the **Add recipients** field.

Tip: Users can send plan run emails to any valid email address or email alias, even if the user(s) do not have an account in Trifacta.

- **Delete:** To remove a user, click the Trash icon next to the email address.
- **Save:** Click **Done**.

Plan View for Flow Tasks

Contents:

- *Sources tab*
 - *Outputs tab*
 - *Parameters tab*
 - *Troubleshooting*
-

When you create a flow task in Plan View, you first search for the flow that you wish to run. Then, you specify the task to execute on the flow in the right context panel.

When a task is executed, the recipes in the flow that have been selected in the task are executed, and if successful, their selected outputs are generated.

NOTE: In a flow, all recipes that you wish to have executed by the corresponding task must have a defined output object. For each output object, you must create at least one write settings or publication object. During plan runs, these objects are not validated, and missing outputs are ignored.

Run flow

X

Q Search flows...

2013 POS

POS-r01 – 2.txt

Run-Job-Publishing-2020-04-16T20:28:40.141Z

allTypes

Run-Job-Publishing-2020-04-16T20:02:41.587Z

allTypes

Untitled Flow

No outputs

flowy

job – 2.log

Test-Notification

carriers_recipe, Flights_recipe

Test-Notification


carriers_recipe, Flights_recipe

Figure: Flow task - select flow

- Enter a search string in the search box.
- When you locate the flow to execute, select it.

Select the outputs that you wish to have generated for the flow.

< Run flow ×

 Run-Job-Publishing-2020-04-16T20:02:41.587Z

Choose outputs to run

☒ allTypes

Cancel Create task

Figure: Flow task - select outputs

Webhooks: If webhooks are configured for the underlying flow, you can optionally disable execution of the webhooks that are defined in the flow when the flow task in the plan is executed. Click the icon to enable or disable webhook execution.

To save the task, click **Create task**.

The saved task is displayed in the context panel.


Tip: To rename the task, click the task name.

Sources tab

Review the datasources for the task, including a full path to the source location.

Run flow

×


 Run Run-Job-Publishing-2020-04-16T20:02:41.587Z

Flow: Run-Job-Publishing-2020-04-16T20:02:41.587Z

...

Sources

Outputs

 allTypes.csv

hdfs://

trifacta/uploads/7/efc8632e-0400-4382-ba3f-88d4d2c6ac8a/allTypes.csv


Figure: Flow task - Sources tab

Outputs tab

Review the outputs that you have selected for the task to generate.

Run flow

×




Run Run-Job-Publishing-2020-04-16T20:02:41.587Z

Flow: Run-Job-Publishing-2020-04-16T20:02:41.587Z

...

Sources

Outputs



allTypes

Profile results

1 publishing action - [Show](#)

[Add/remove outputs](#)

Figure: Flow task - Outputs tab

- To review the publications configured for the output, click **Show**.
- To change the outputs generated by the task, click **Add/remove outputs**.


Parameters tab

If the flow from which the task was created contains parameters, you can review those parameters and apply overrides as necessary.

NOTE: Parameter overrides applied to a plan affect only plan execution. These overrides do not apply to any independent job executions of the underlying flows.

Flow task

×



Run 2013 POS

flowtask-9v • [2013 POS](#)

...

Sources

Outputs

Parameters

Parameter	Value	Manage
<> region	02	

Figure: Flow task - Parameters tab

To apply overrides:

1. Click **Manage**.
2. In the Manage Parameters dialog, mouse over the Value entry for the desired parameter. Click the Edit icon.

Tip: You can use the Search box to locate parameters by name.

3. Enter a new value and click **Save**.

Whenever the flow task is executed, this override value is applied to the execution of the flow task.

Troubleshooting

Flow tasks are essentially standard transformation jobs with additional metadata on them. When a flow task fails to execute for any reason, it's likely that the source of the problem resides in the flow definition, its objects, and its connections. For more information, see *Diagnose Failed Jobs*.


Plan View for HTTP Tasks

In Plan View, you can create HTTP tasks to send request to endpoints before or after the execution of other tasks. These tasks are specified in the right context panel.

<

HTTP task

×

 Send HTTP Request

Test

...

Method

GET


▼

URL

required

https://example.com/v4/connections

https://example.com/endpoint

Headers 

Add

Authorization

<my_authorization_key>

Remove

Key

Value

Remove

Secret Key

Secret Key

☒ Validate SSL certificate

Retry

3

times

Cancel

Save

Figure: HTTP task

Tip: To rename the task, click the task name.

Fields:

Field	Description
Method	Select the HTTP method to use to deliver the message. The appropriate method depends on the receiving application. Most use cases require the POST method.
URL	URL where the HTTP request is received by the other application.
Headers	<p>Insert HTTP content headers as key-value pairs. For example, if your body is in JSON format, you should include the following header:</p> <pre>key: Content-Type value: application/json</pre> <p>NOTE: You may be required to submit an authentication token as the value for the <code>Authorization</code> key.</p>
Body	<p>(POST , PUT , or PATCH methods only) The body of the request submitted to the receiving application. Request body is structured as follows:</p> <pre>{"text":"My text message to the receiving application."}</pre> <p>Tip: As part of the request body or header fields, you can insert metadata references to the plan definition, current run, tasks already executed in the run, and underlying flows, including column data and datasources. For more information on the available metadata, see <i>Plan Metadata References</i>.</p> <p>For examples of requests including metadata examples, see <i>Create HTTP Task</i>.</p>
Secret Key	<p>(Optional) A secret key can be used to verify the request payload. This secret value must be inserted in this location, and it must be included as part of the code used to process the requests in the receiving application. Insert the secret value here as a string without quotes.</p> <p>For more information on how this secret key is used to generate a signature, See <i>Create HTTP Task</i>.</p>
Validate SSL Certificate	<p>When set to <code>true</code>, HTTPS (SSL) communications are verified to be using a valid certificate before transmission.</p> <p>NOTE: If you must send a request to an endpoint that has an expired/invalid certificate, you must disable SSL verification.</p>
Retry	<p>If the returned status code is outside of the 200-299 range, then the HTTP task is considered to have failed. When this option is enabled, the request is retried.</p> <p>If the request fails, this value defined the number of times that the request should be retried. If this number of retries is reached without success, the task fails.</p>

Actions:

- To test if the specified endpoint is reachable, click **Test**.

Tip: A status code of 200 indicates that the test was successful.

Tip: You can use the GET method for testing purposes. A GET request does not change any data on the target platform but may permit you to specify elements in the request body.

- Edit task name:** Change the name of the task.

Tip: Good naming may include the target platform endpoint and method, as well as the purposes of the task in your plan.

- **Delete:** Delete the task.

This step cannot be undone.

For more information, see *Create HTTP Task*.

Plan View for Slack Tasks

In Plan View, you can create tasks to send messages to a Slack channel. These tasks are specified in the right context panel.

Tip: Slack tasks are a specialized form of HTTP tasks.

Requirements

The following requirements apply to the Slack app that receives the message. For more information on Slack apps, see <https://api.slack.com/apps>.

Please verify that your Slack app has the following:

- Create an OAuth Token that has `chat:write` scopes. This token is inserted into your task definition. There are two types of tokens:
 - **Bot Token:** These tokens post a Slack message from the name of the app.

NOTE: The Bot Token also requires the `chat:write.public` scope.

Tip: A Bot Token is required if you wish to send a direct message through the App category of messages.

Tip: To send a message to a private channel using a Bot Token, you must install the app in the channel through the Integrations window in the channel's settings.

- **User Token:** These tokens post a Slack message from the user who authorizes the message.

Tip: To send a message to a user or to a private channel using a User Token, additional configuration is required. Please see below.

- The OAuth Token that you create must be installed in your workspace.


NOTE: Copy the generated token to a text file and retain it for later. This token must be pasted into the definition of each Slack task where you wish to use it.

Create Slack Task

<

Slack task

×

 Post a message

Test

...

Request

Response

To integrate with Slack, you need a Slack app. See your apps or create a new one [here](#).

OAuth Token

required

xoxb

You will find it in [your app's](#) OAuth & Permissions section

Channel

required

testing

Paste the name of the channel, exactly as you can see it on Slack

Message

Plan {{plan.name}} is running.
Start time: {{plan.startTime}}

Add metadata to your message by simply pressing \$

Cancel

Save

Figure: Slack task

Tip: To rename the task, click the task name.

Fields:

Field	Description
OAuth Token	The OAuth token to use for posting the message.
Channel	Paste one of the following values from the Slack workspace for where to post the message: <ul style="list-style-type: none">Channel Name: Name of the channel as it appears in Slack.Channel ID: This value is available in the Settings page for the channel.

Copyright © 2022 Trifacta Inc.

Page #94

	<ul style="list-style-type: none"> • Member ID: You can post the message to a specific user instead of posting to a channel. <div> Tip: A user's member ID can be found in the user's Profile page in Slack. </div>
Message	The message to post. <div> Tip: Messages can include metadata information about the tasks in the current plan run. For more information, see <i>Plan Metadata References</i>. </div>

Actions:

- To test if the specified endpoint is reachable, click **Test**.

Tip: A status code of 200 indicates that the test was successful.

- **Edit task name:** Change the name of the task.
- **Delete:** Delete the task.

This step cannot be undone.

For more information, see *Create Slack Task*.

Plan View for Delete Tasks

A Delete task can be created in Plan View to delete existing files or folders from backend storage. These tasks are specified in the right context panel.

Tip: Delete tasks are useful for removing files that were generated as part of plan's execution but are not needed afterward.

Requirements

- You must have access to the connections through which a Delete task removes files or folders.
- You must have write permissions to:
 - Any connection through which you are removing files or folders.
 - Any bucket, folder, or file accessed through that connection.

NOTE: If you select to delete a bucket, the contents of the bucket are removed, but the bucket object remains.

- Delete tasks are supported for the following file systems:
 - S3
 - ADLS

Create Delete Task

<

Delete task

×

🗑️

Delete

⋮

Path

@trifacta.com/jobrun/POS-r01.json

Browse

Files to delete (1)

📄

@trifacta.com/jobrun/POS-r01.json

Cancel

Save

Figure: Delete task

Tip: To rename the task, click the task name.

Fields:

Field	Description
Connection	<p>If you have access to multiple file-based connections, you can select the connection to use from the Connection drop-down.</p> <p>If this drop-down is not present, you automatically connect through the base storage layer for your environment.</p>
Path	<p>For the selected connection, you must specify the path to the file or folder that you wish to remove. To navigate to the location, click Browse.</p>

Tip: If you edit the pathway while browsing, you can paste in paths, which you may copy from sources like the Dataset Details or Job Details pages.

Tip: You can insert plan metadata references in the path for tasks that have previously been executed in the plan. Enter \$ to begin exploring available references.

When the path is specified, a list of matching files is displayed.

To save your task, click **Save**.

Actions:

- **Edit name:** Change the name of the task.
- **Delete:** Delete the task.

This step cannot be undone.

For more information, see *Create Delete Task*.

Library Page

In the Library page, you can review your imported and reference datasets and any macros that you may have created. You can also import new data from this page.

NOTE: You can only see the imported datasets to which you have access in your currently selected project or workspace. If the data underlying the imported dataset is not available, the imported dataset is still listed in the Library page, since it is just a reference to the data.

Library

All Data

Imported Datasets

References

Macros

All

Owned by me

Shared with me

All Data

Import Data

Find Datasets... (/)

Name	In Flows	Source	Last Updated ▾
<div><div></div><div>Dataset with Parameters</div></div>	0	HDFS	Today at 11:33 AM
<div><div></div><div>POS-schema.csv</div></div>	1	HDFS	Today at 11:28 AM
<div><div></div><div>POS-r03.txt</div></div>	1	HDFS	Today at 11:28 AM
<div><div></div><div>POS-r01.txt</div></div>	1	HDFS	Today at 11:28 AM
<div><div></div><div>REF_CAL.txt</div></div>	1	HDFS	Today at 11:28 AM
<div><div></div><div>POS-r02.txt</div></div>	1	HDFS	Today at 11:28 AM
<div><div></div><div>REF_PROD.txt</div></div>	1	HDFS	Today at 11:28 AM

Figure: Library Page

To create a new imported dataset, click **Import Data**. See *Import Data Page*.

For large relational or Parquet datasets, you can monitor the import process through the Library page.

- During the import process, you can hover over the icon for a **pending dataset** to track status.
- Click the icon for additional details. See *Dataset Details Page*.

NOTE: This feature may require enablement in your deployment. See *Configure JDBC Ingestion*.

For more information, see *Overview of Job Monitoring*.

Filter by type:

Click one of the pre-defined filters to show datasets of the following types:

All Data: All imported datasets or references available to the current user.

Imported Datasets: Datasets that you have imported into Trifacta®.

- The Source column indicates where the original source data is located.
- You can also access datasets that were imported through a configured connection. For more information, see *Connections Page*.

References: Objects that you have created from your recipes that can be referenced in another flow as a dataset.

Macros: Sequences of steps that can be reused in other recipes. See *Macros Page*.

Filter by ownership:

For the selected object type, you can filter based on the ownership of the object:

- **All:** All objects of the selected type to which you have access.
- **Owned by me:** All objects of the selected type that you own.
- **Shared with me:** All objects of the type that have been shared with you.

Columns:

- **Name:** Name of the object.
- **In Flows:** Count of flows in which the object is in use.
- **Source:** Flow or datastore where the object is located.
- **Last Updated:** Timestamp of the last time that the object was modified.

Actions:

- **Browse:** If displayed, use the page browsing controls to explore the available objects.
- **Search:** To search object names, enter a string in the search bar. Results are highlighted immediately in the Library page.
- **Sort:** Click a column header to sort the display by the column's entries.

Object Actions:

Hover over an object to reveal these actions on the right side of the screen.

- **Details:** Review details about the dataset. See *Dataset Details Page*.
- **Preview:** Inspect a preview of the dataset.

NOTE: Preview is not available for binary format sources.

- **Use in new Flow:** (Imported dataset only) You can create a new flow and begin immediately wrangling the dataset. This step also creates a recipe in the flow.
- **Add to Flow:** Add the dataset to a new or existing flow.
- **Make a copy:** Create a copy of the imported dataset. This option is not available for reference datasets.
- **Edit name and description:** Change the name and description of the dataset.
- **Edit data settings:** If the source of the imported dataset required conversion to an internally supported format, you can modify settings related to that conversion process. For more information, see *File Import Settings*.

Tip: This setting applies primarily to binary file formats, such as PDF and Excel, or file formats that may require additional steps to convert into tabular data, such as JSON.

- **Delete Dataset:** Delete the dataset.

Deleting a dataset cannot be undone.

- **Refresh Dataset:** If available, this option refreshes the dataset's metadata with the latest source schema.

NOTE: If you attempt to refresh the schema of a parameterized dataset based on a set of files, only the schema for the first file is checked for changes. If changes are detected, the other files are contain those changes as well. This can lead to changes being assumed or undetected in later files and potential data corruption in the flow.

For more information, see *Overview of Schema Management*.

Dataset Details Page

Contents:

- *Imported Dataset*
- *Reference Dataset*
- *Dataset with Parameters*

Use the Dataset Details page to review a dataset's usage and to perform management tasks on it.


Imported Dataset

For datasets that have been imported into Trifacta®, you can review source location and current usage within flows to which you have access.

Status: For large relational datasets, you can track status of the import process. For more information, see *Overview of Job Monitoring*.

NOTE: This feature may require enablement in your deployment. See *Configure JDBC Ingestion*.

If the dataset is used in a flow, click the flow name to review its usage in the flow.

 allTypes_2.pqt

Use in new Flow

Preview

...

Last updated: Today at 5:37 PM

Created: Today at 5:37 PM

File size: 28.81kB

Size: 17 columns - 7 types

Location:

Column data type inference: Enabled

Used in 1 Flow


Name	Owner	Objects	Last updated
 \$USER_TEST20220119173730 API TestFramework - Flow	Administrator	1 Dataset, 1 Recipe	Today at 5:37 PM

Figure: Imported dataset details

Actions:

- **Use in new Flow:** Create a new flow for your dataset and begin wrangling.
- **Preview:** Review the first few rows of the dataset.
- **Edit custom SQL:** After you have created a dataset using custom SQL, you can modify the SQL used to construct the imported dataset.
- **Add to Flow:** Add imported dataset to a new or existing flow.
- **Make a copy:** Create a copy of the imported dataset.
- **Edit name and description:** Edit the name and description for the dataset.
- **Remove structure:** Remove initial steps applied to structure data.
- **Delete Dataset:** Delete the dataset.

Deleting a dataset cannot be undone.

- **Refresh Dataset:** If available, this option refreshes the dataset's metadata with the latest source schema. For more information, see *Overview of Schema Management*.

If the dataset was imported with a customized SQL statement, select **Edit custom SQL**. Modify the SQL statement(s) as needed.

Through the custom SQL interface, it is possible to enter SQL statements that can delete data, change table schemas, or otherwise corrupt the targeted database. Please use this feature with caution.

NOTE: If you modify the SQL statement for your imported dataset, any samples based on the old SQL statement are invalidated.

For more information, see *Create Dataset with SQL*.
For more information on the sources from which a dataset was created, see *Flow View Page*.

Reference Dataset

A **reference dataset** is a reference from one flow to the dataset that is sourced in another flow. When the source dataset is modified, the reference dataset automatically receives the changes. For more information on creating a reference dataset, see *Flow View Page*.

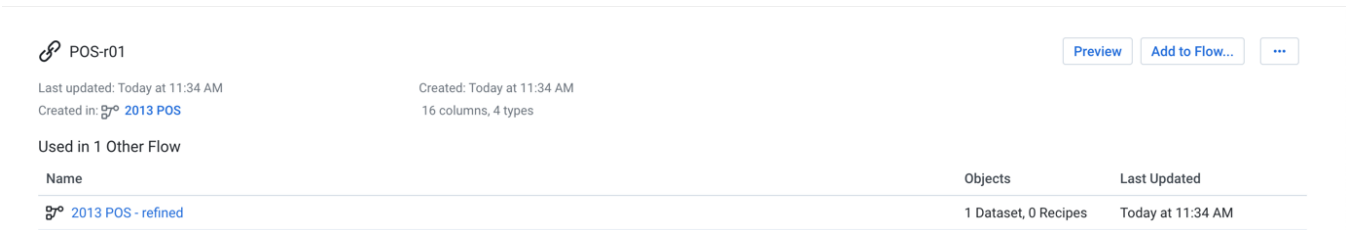


Figure: Reference Dataset details

Actions:


- **Preview:** Review a preview of the first few rows in the dataset.
- **Add to Flow:** Add the reference dataset to a new or existing flow.
- **Edit name and description:** Edit the name and description for the dataset.
- **Delete Reference Dataset:** Delete the reference dataset. The object on which the reference dataset is based is untouched.

Deleting a dataset cannot be undone.

Dataset with Parameters

If your dataset was created with parameters, you can review dataset and parameter information in the details.

- For more information on creating these datasets, see *Create Dataset with Parameters*.
- For more information, see *Overview of Parameterization*.

 Dataset with Parameters

Wrangle in new Flow
 Preview
 ...

Last updated: Today at 11:33 AM
 Created: Today at 11:33 AM
 Size: 16 columns · 4 types

Parameters

Path: eed8f72c-edcc-40f6-b067-797841e1cc1c/POS-r_* {digit}{digit}.txt

Parameters	Type	Default value	Name
* {digit}{digit}	Pattern	matches against trifacta pattern: '{digit}{digit}'	

Figure: Dataset with Parameters details

You can review the parameters and variables that have been defined for the dataset.

Action:

- **Wrangle in new Flow:** Create a new flow for your dataset and begin wrangling.
- **Preview:** Review the first few rows of the dataset.
- **Add to Flow:** Add imported dataset to a new or existing flow.
- **Make a copy:** Create a copy of the imported dataset.
- **Edit name and description:** Edit the name and description for the dataset.
- **Edit parameters:**Modify the parameters used to create the dataset. See *Create Dataset with Parameters*.
- **Remove structure:** Remove the initial parsing structure. When the structure is removed:
 - The dataset is converted to an unstructured dataset. An **unstructured dataset** is the source data converted into a flat file format.
 - All steps to shape the dataset are removed. You must break up columns in manual steps in any recipe created from the object. See *Flow View Page*.
- **Delete Dataset:** Delete the dataset.

Deleting a dataset cannot be undone.

Import Data Page

Contents:

- General Limitations
- Basic Workflow
 - 1. Connect to sources
 - 2. Add datasets
 - 3. Configure selections
 - 4. Import selections

Through the Import Data page, you can upload datasets or select datasets from sources that are stored on connected datastores. From the Library page, click **Import Data**.

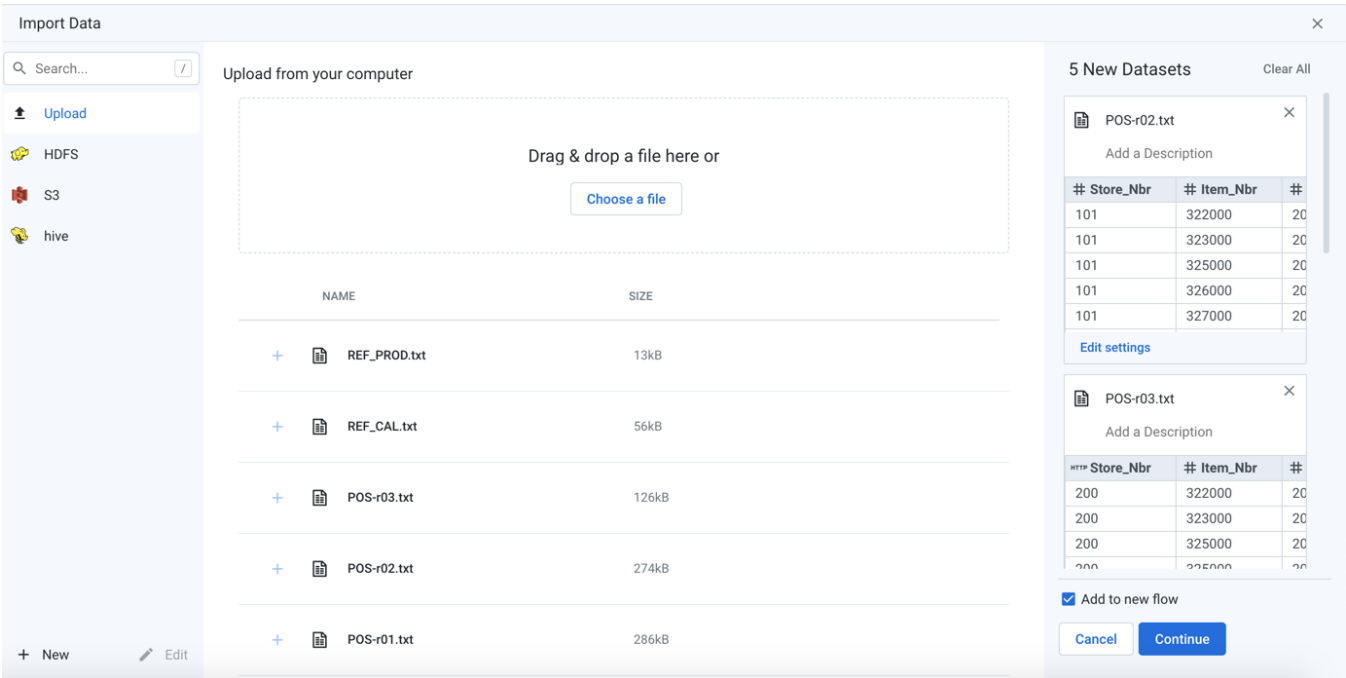


Figure: Import Data page

General Limitations

NOTE: For file-based sources, Trifacta® expects that each row of data in the import file is terminated with a consistent newline character, including the last one in the file.

- For single files lacking this final newline character, the final record may be dropped.
- For multi-file imports lacking a newline in the final record of a file, this final record may be merged with the first one in the next file and then dropped in the Trifacta Photon running environment.

NOTE: To be able to import datasets from the base storage layer, your user account must include the `dataAdmin` role.

NOTE: An imported dataset requires about 15 rows to properly infer column data types and the row, if any, to use for column headers.

File and path limitations:

- The colon character (`:`) cannot appear in a filename or a file path.
- Filenames cannot begin with special characters like dot (`.`) or underscore (`_`).
- Input file or table paths can have a maximum length of 1024 characters.

Basic Workflow

1. Connect to sources

During import, the Trifacta application identifies file formats based on the extension of the filename.

- Compressed files are recognized and can be imported based on their file extensions.
- Filenames that do not have an extension are treated as TXT files.

Upload: Trifacta® can also load files from your local file system.

Tip: You can drag and drop files from your desktop to to upload them.

NOTE: You can upload a file up to 1 GB in size.

NOTE: When you upload an updated version of a previously uploaded file, the new file is stored as a separate upload altogether. In your flow, you must swap out the old dataset to point to the new one.

HDFS: If connected to a Hadoop cluster, you can select file(s) or folders to import. See *HDFS Browser*.

Hive: If connected to a Hive instance, you can load datasets from individual tables within the set of Hive databases. See *Hive Connections*.

S3: If connected to an S3 instance, you can browse your S3 buckets to select source files.

Tip: For HDFS and S3, you can select folders, which selects each file within the directory as a separate dataset.

See *External S3 Connections*.

Redshift: If connected to an S3 data warehouse, you can import source from the connected database. See *Amazon Redshift Connections*.

WASB: If enabled, you can import data into your Azure deployment from WASB.

ADL: If enabled, you can import data into your Azure deployment from ADLS Gen1.

ADLS Gen2: If enabled, you can import data into your Azure deployment from ADLS Gen1.

Databases: If connected to a relational datastore, you can load tables or views from your database. See *Database Browser*.

NOTE: For long-loading relational sources, you can monitor progress through each stage of ingestion. After these sources are ingested, subsequent steps to import and wrangle the data may be faster.

For more information, see *Configure JDBC Ingestion*.

For more information, see *Overview of Job Monitoring*.

For more information on the supported input formats, see *Supported File Formats*.

New/Edit: Click to create or edit a connection. By default, the displayed connections support import.

Search: Enter a search term to locate a specific connection.

NOTE: This feature may be disabled in your environment. For more information, contact your Trifacta administrator.

See *Create Connection Window*.

2. Add datasets

When you have found your source directory or file:

- You can hover over the name of a file to preview its contents.

NOTE: Preview may not be available for some sources, such as Parquet.

- Click the Plus icon next to the directory or filename to add it as a dataset.

Tip: You can import multiple datasets at the same time. See below.

- **Excel files:** Click the Plus icon next to the parent workbook to add all of the worksheets as a single dataset, or you can add individual sheets as individual datasets. See *Import Excel Data*.

- If custom SQL query is enabled, you can click **Create Dataset with SQL** to enter a customized SQL statement to pre-filter the table within the database to include only the rows and columns of interest.

Through this interface, it is possible to enter SQL statements that can delete data, change table schemas, or otherwise corrupt the targeted database. Please use this feature with caution.

For more information, see *Create Dataset with SQL*.

If parameterization has been enabled, you can apply parameters to the source paths of your datasets to capture a wider set of sources. Click **Create Dataset with Parameters**. See *Create Dataset with Parameters*.

3. Configure selections

When a dataset has been selected, the following fields appear on the right side of the screen. Modify as needed:

- **Dataset Name:** This name appears in the interface.
- **Dataset Description:** You may add an optional description that provides additional detail about the dataset. This information is visible in some areas of the interface.

Tip: Click the Eye icon to inspect the contents of the dataset prior to importing.

Tip: You can select a single dataset or multiple datasets for import.

Edit settings

You can edit any additional or optional settings for an individual dataset. Perform the following:

Steps:

1. Click **Edit Settings** from the card for an individual dataset in the right panel. The dialog box is displayed.
2. In the dialog box, select the required options and modify the settings.
 - **File Import Settings:** For more information, see *File Import Settings*.
 - **Table Import Settings:** For more information, see *Table Import Settings*.

4. Import selections

Single dataset

If you have selected a single dataset for import:

Tip: If present, you can click the **Add to new flow** checkbox, which adds the imported datasets to an untitled flow.

- Click **Continue**. The dataset is imported.
- A recipe is created for it, added to a new flow, and loaded in the Transformer page for wrangling. See *Transformer Page*.

Multiple datasets

You can import multiple datasets from multiple sources at the same time. In the Import Data page, continue selecting sources, and additional dataset cards are added to the right panel.

NOTE: If you are importing from multiple files at the same time, the files are not necessarily read in a regular or predictable order.

NOTE: When you import a dataset with parameters from multiple files, only the first matching file is displayed in the right panel.

In the right panel, you can see a preview of each dataset and make changes as needed.

×

5 New Datasets

Clear All

POS-r02.txt

×

Add a Description

# Store_Nbr	# Item_Nbr	#
101	322000	20
101	323000	20
101	325000	20
101	326000	20
101	327000	20

Edit settings

POS-r01.txt

×

Add a Description

# Store_Nbr	# Item_Nbr	#
1	381000	20
2	325000	20
2	325000	20
2	402000	20

☐ Add to new flow

Cancel

Continue

Figure: Import Multiple Datasets

If you have selected multiple datasets for import:

Tip: If present, you can click the **Add to new flow** checkbox, which adds the imported datasets to an untitled flow. For more information, see *Flow View Page*.

- To import the selected datasets, click **Continue**.
 - To begin transforming one of these datasets in Flow View, select it. From its context menu, select **Add new recipe**. Select the recipe. In the context panel on the right, select **Edit Recipe**. See *Transformer Page*.
- To remove a dataset from import, click the X in the dataset card.

Create Connection Window

Through the Create Connection window, you can create and edit connections between Trifacta® and remote storage.

NOTE: Access to this page in the application and privileges on its related objects is governed by roles in your workspace. For more information, please contact your workspace administrator.

This window is available from the following locations:

- **From Import Data page:** By default, the window displays connections that support import. Deselect the checkbox to display all available connection types.
- **From Run Job page:** When you add a new connection as part of a publishing action, the window displays connections that support publishing by default.
- **From the Connections page:** All available connections are displayed.

NOTE: Some connection types may not be available for your environment.

NOTE: In your environment, creation of connections may be limited to administrators only. For more information, please contact your Trifacta administrator.

Tip: Administrators can edit any public connection.

General Connection Notes:

- After you create a connection, you cannot change its connection type. You must delete the connection and start again.
- Connections can be created, managed, shared, and deleted through the Connections page. See *Connections Page*.

Database Connection Notes:

- Database connections cannot be deleted if their databases host imported datasets that are in use by Trifacta. Remove these imported datasets before deleting the connection.
- Jobs created for datasets sourced from a database cannot be executed on a Spark-based running environment.

Connection Type

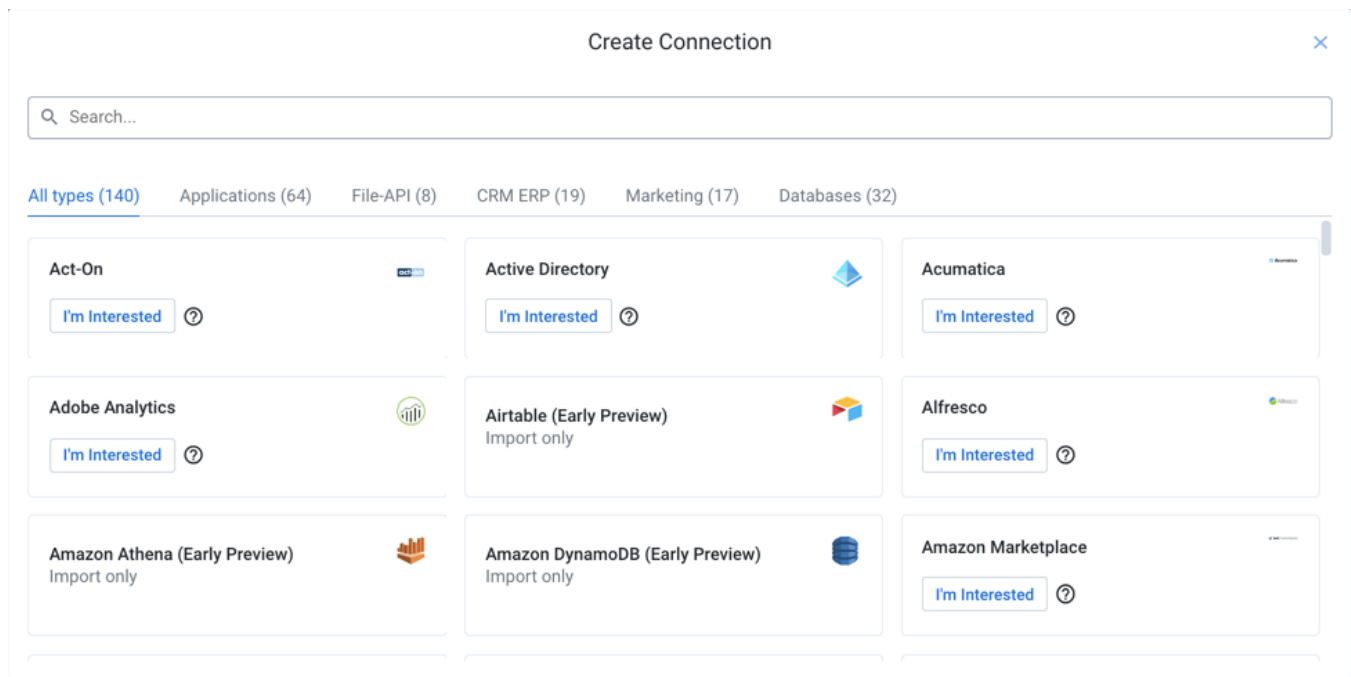


Figure: Connection Type window

In the Connection Type window, you search or browse for the type of connection to create.

Actions:

- Use the Search bar to perform real-time searches of connection types.
- Click one of the categories to browse for connection types that apply to the listed environment.
- Select the type of connection to continue:
 - `Import only` - Connection can be used only to import data into the platform.
 - `Publish only` - Connection can be used only to publish data from the platform to the connection target.
 - `Import and publish` - Connection can be used to import data and to publish your outputs.

For more information on these connections, See *Connection Types*.

Create Connection

Create Connection

PostgreSQL

Import and publish

SERVER INFORMATION

Host

myserver

Port

5432

Connect String Options (optional)

☒ Enable SSL

Database

myDB

User Name

dbuser

Password

Test Connection

< Back

Cancel

Create

Figure: Create Connection Window

Property	Description
Host	Host of the database.
Port	Port by which to access the database host. Default values are pre-populated based on the connection type you selected.
Connect String Options	(optional) If access to the database requires special connection string options, you may paste or enter them here. You only need to provide the parameter and string value. Example: <div>" ;transportMode=http;httpPath=cliservice "</div>
Enable SSL	To connect using SSL, click this checkbox.

	<p>If this checkbox is not present, SSL connections for this database type are not supported or are required:</p> <ul style="list-style-type: none"> • SSL connections are not supported for SQL Server or Hive. • SSL connections are required for Redshift and SQL DW. <p>No additional Connect String Options are required for supported database vendors.</p> <div> NOTE: The database must be configured to receive SSL connections. </div>
Service Name	(Oracle only) Name of the service. For example, enter <code>orcl</code> here.
Database	(PostgreSQL only) Name of the database to connect. The name of the default database is the username, so you should change this value in most cases.
Credential Type	<p>Depending on the type of datastore to which you are connecting, you may have multiple methods of providing credentials for authentication:</p> <ul style="list-style-type: none"> • <code>Basic</code> - Username and password credentials are provided as part of the connection definition. • <code>OAuth 2.0</code> - Connection accesses the datastore using OAuth 2.0 authentication. <div> NOTE: OAuth 2.0 authentication requires additional configuration. For more information, see <i>Enable OAuth 2.0 Authentication</i>. </div> <div> NOTE: For each type of connection that uses OAuth 2.0, you must create a client app and a client in the Trifacta application. See <i>Create OAuth2 Client</i>. </div> <div> NOTE: When you create the connection in this window for an OAuth 2.0 connection, you must click Authenticate, which uses the OAuth 2.0 client to connect to the app. This step is required. </div>
User Name	(basic credential type) Username to access the database. This value is encrypted for security.
Password	(basic credential type) Password for the specified user. This value is encrypted for security.
OAuth 2.0 Client	<p>(OAuth 2.0 credential type) Select the OAuth 2.0 client to use to connect to the datastore.</p> <div> NOTE: You must create a separate connection for each OAuth 2.0 client that is available in the drop-down list. </div>
Test Connection	When the above properties are specified, click Test Connection to validate that Trifacta can connect to the database. If the connection test fails, your administrator may need to install a keyfile. See <i>Relational Access</i> .
Advanced Options: Default Column Data Type Inference	You can choose to enable or disable type inferencing for individual connections, when the connection is created or edited. The default setting for this parameter is defined at the global level. For more information, see <i>Configure Type Inference</i> .
Advanced Options: Enable SSH Tunneling	This feature is not available.
Connection Name	<p>Display name of the connection, which appears in the application.</p> <div> NOTE: This value must be unique among all connections. </div>
Connection Description	User-friendly description for the connection, which appears in the application.

When you've finished, click **Ok** to save the connection.

After you have created your connection, run a simple job on data sourced from it.

NOTE: You can make the connection available to all users by sharing it. See *Connections Page*.

Database Browser

Contents:

- *Browse Databases*
- *Search List*
- *Preview Table Data*
- *Create Dataset with SQL*

The database browser enables you to interact with databases that are connected to Trifacta®.

The database browser appears when:

- You select one of the database tabs to create an imported dataset. See *Import Data Page*.
- You add a publishing action in the Run Job page and choose a database connection through which to write the job results. See *Run Job Page*.

For more information about interacting with databases through the product, see *Using Databases*.

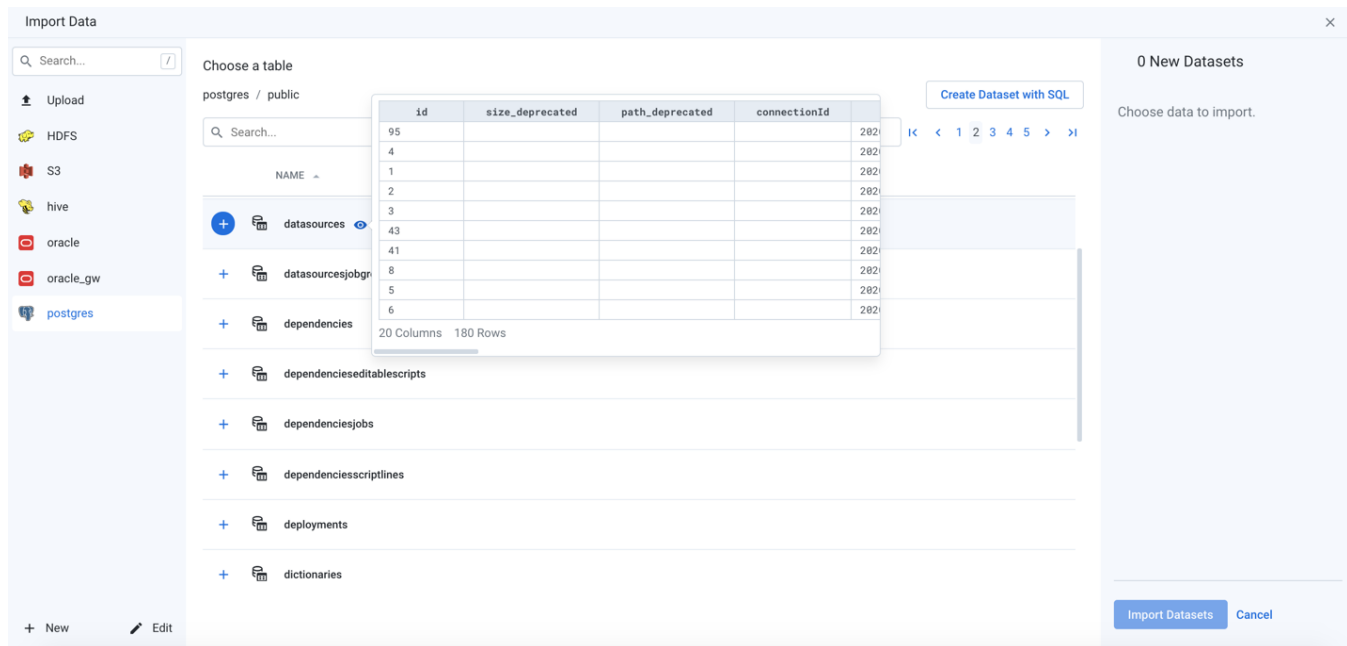






Figure: Database Browser

Browse Databases

Use the links and icons to browse databases and their tables and views.

NOTE: Avoid using the Back button on your browser, which exits the browser without applying changes to your configuration.

Identifier	Type	Description
	Datab	Click these links to open a database to reveal its tables and views.

	ase	
	Schema	(Postgres only) Click a schema link to display the tables and views that use the schema.
	Table	<p>Click the Plus icon to select this table.</p> <p>To preview its data, hover over the name of the table, and then click the Eye icon.</p> <div> <p>Tip: Sizes and update timestamps are calculated and displayed next to tables. They are not displayed next to databases.</p> <p>NOTE: Column count information is not available for nested tables.</p> </div>
	View	<p>Click the Plus icon to select this view as your source.</p> <p>To preview its data, click the Eye icon next to the view name.</p> <div> <p>NOTE: Previewing complex views may impact performance.</p> </div>
	Page navigation	<p>Use these links to navigate between pages of databases and tables and views.</p> <div> <p>NOTE: In some cases, subsequent pages of tables and views may be blank, and counts of tables and views may not match displayed figures. This is a known issue.</p> </div>
postgres / public	Breadcrumb	Click the links in the breadcrumb trail to navigate.

Search List

To filter the list, enter a string in the Search box. The filter is applied as you type and matches anywhere in the name of a currently displayed database, table, or view name.

Preview Table Data

Database tables are displayed by name only. To preview the data in the table, click the Eye icon next to the name of the table.

Tip: Table previews include available metadata information, such as column headers and column and row counts.

NOTE: Depending on the database type, rows may not be displayed in a specific order.

Create Dataset with SQL

As needed, you can pre-filter a selected table or view inside the database prior to import. By entering a custom SQL statement, you can remove unnecessary data from the dataset that is extracted from the database, which enables faster and more meaningful imports of your database data. See *Create Dataset with SQL*.

File System Browser

In Trifacta®, the file system browser lets you browse, select, and filter the sources that you can access through the datastore to which you are connected. You also use the browser to select targets for publishing job results.

Interactions with the connected file system may be determined base on:

- user permissions to specific directories
- features enabled in the product
- any impersonation or kerberos restrictions

For more information, please contact your administrator.

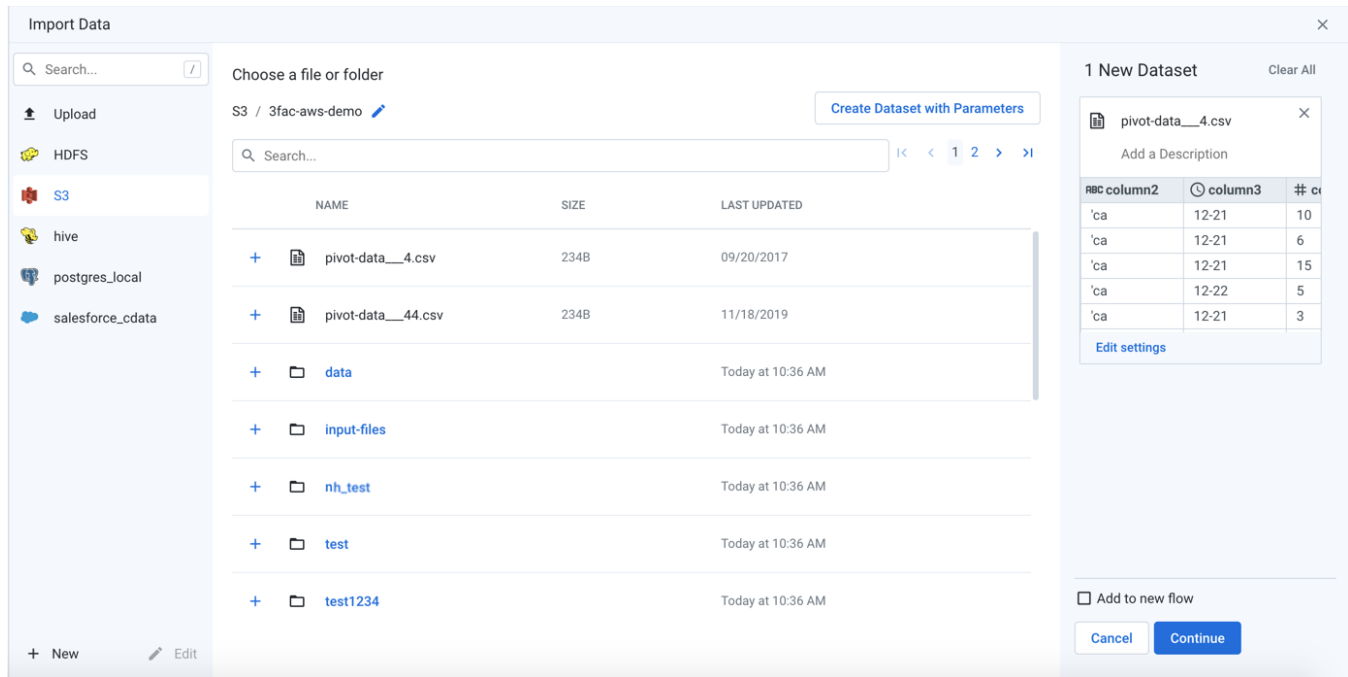


Figure: File System Browser

Through the file system browser, you navigate folders and select files through an easy-to-use interface. At a technical level, these objects are typically distributed across multiple servers and may be represented as part files of the whole virtual file.




Browse:

NOTE: Avoid using the Back button on your web browser, which exits the file system browser without applying changes to your configuration.

Use the links and icons to browse for files and folders in the file system tree structure.

NOTE: If you do not have the appropriate permissions, you may not be able to browse all of the folders of the directory. However, you may be able to paste in the full path to your location to gain access.

Tip: If the displayed file system is the base storage layer, then the path to your output home directory should be available through the browser.

Identifier	Type	Description
	Bucket	(not always present) In some file systems, the top level browsing object is called a bucket . NOTE: You cannot add an entire bucket as a source of data for your datasets.
	Folder	<ul style="list-style-type: none">Click the Plus icon to select all readable files in this folder.Click the text link to open the folder and browse further. You must have the appropriate permissions in your account. Tip: When you open a new folder, a reference to it is added to the Path value. You can modify the path value manually, which may be a faster way to navigate up a deep directory structure. Tip: Sizes are displayed next to files. They are not displayed next to folders.
	File	Click the Plus icon to select this file. The Last Updated column contains information only for files. It is not available for directories.

Specify Path:

In the browser, you can specify an explicit path to resources. Click the Pencil icon, paste the path value, and click **Go**.

For example, if your home input directory is the following:

```
/mydir/input/username@example.com
```

You should paste the following in the Path textbox:

```
<bucketname>/mydir/input/username@example.com
```

Tip: You can retrieve your home directory from your profile.

Search Files:

To display a subset of files, enter a string in the Search box. The filter is applied as you type and matches anywhere in the name of a currently displayed file or folder.

NOTE: If you have a folder and file with the same name, search may only retrieve the file. You can still navigate to locate the folder.

HDFS Browser

The HDFS browser enables you to browse, select, and filter the files to which you have access in the Hadoop cluster to which Trifacta® is connected.

The HDFS browser appears when you create a dataset in HDFS or in the HDFS tab when you import a dataset. See *Import Data Page*.

NOTE: Interactions with HDFS are determined by user permissions and features enabled in the Trifacta platform. For more information, see *Using HDFS*.

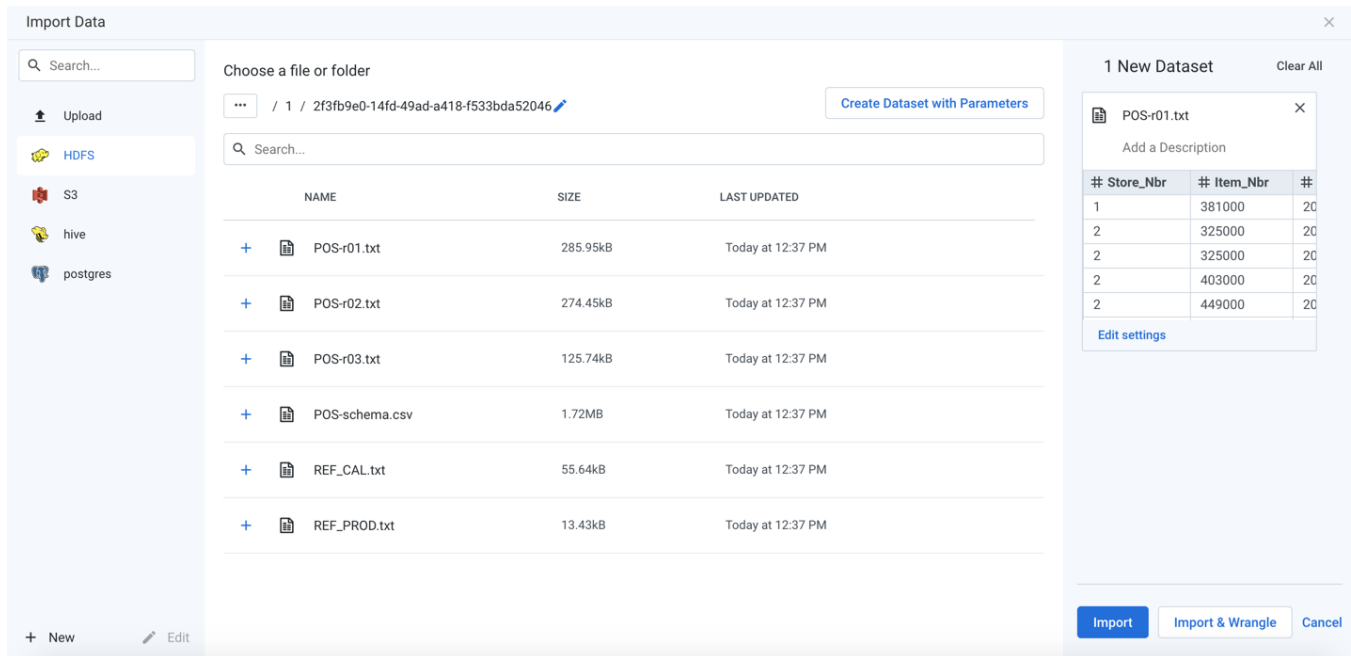



Figure: HDFS Browser

Browse HDFS:

Use the links and icons to browse for files and folders in the HDFS tree structure.

NOTE: Avoid using the Back button on your browser, which exits the HDFS browser without applying changes to your configuration.

Identifier	Type	Description
	Folder	<ul style="list-style-type: none">Click the Plus icon to select all readable files in this folder.Click the text link to open the folder and browse further. <div>Tip: When you open a new folder, a reference to it is added to the Path value. You can modify the path value manually, which may be a faster way to navigate up a deep directory structure.</div> <div>Tip: Sizes are displayed next to files. They are not displayed next to folders.</div>
	File	Click the Plus icon to select this file.



Specify HDFS Path:

In the HDFS browser, you can specify an explicit path to resources. Click the Pencil icon, paste the path value, and click **Go**.

```
/trifacta/input/username@example.com
```

You should paste the following in the Path textbox:

```
HDFS/trifacta/input/username@example.com
```

Tip: You can retrieve your home directory from your profile. See *Storage Config Page*.

Filter Files:

To display a subset of files, enter a string in the Search box. The filter is applied as you type and matches anywhere in the name of a currently displayed file or folder.

S3 Browser

In Trifacta®, the S3 browser lets you browse, select, and filter the sources that you can access through S3. You also use the browser to select targets for publishing job results.

NOTE: Interactions with S3 are determined by user permissions and features enabled in Trifacta. For more information, see *Using S3*.

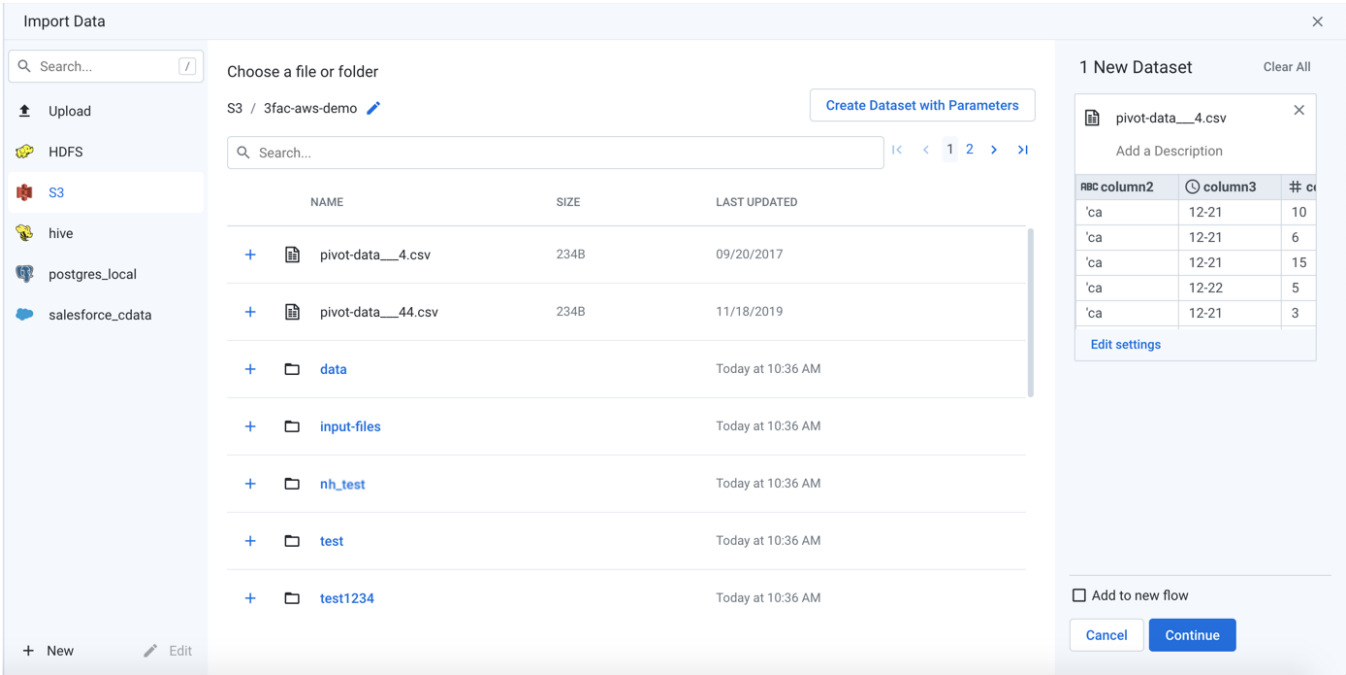


Figure: S3 Browser

Browse S3:

Use the links and icons to browse buckets for files and folders in the S3 tree structure.




NOTE: Permission to browse your buckets is determined by the permissions associated with your S3 credentials. If you do not have the appropriate permissions, you may not be able to browse the bucket. However, you may be able to paste in the full path to your location to gain access.

Tip: If S3 is the base storage layer, then the path to your output home directory should be available through the S3 browser. For more information on this path, see *Storage Config Page*.

- The Last Updated column contains information only for files. It is not available for directories.

NOTE: Avoid using the Back button on your browser, which exits the S3 browser without applying changes to your configuration.

Identifier	Type	Description
	Bucket	Indicates an S3 bucket.

		NOTE: You cannot add an entire S3 bucket as a source of data for your datasets.
	Folder	<ul style="list-style-type: none"> Click the Plus icon to select all readable files in this folder. Click the text link to open the folder and browse further. You must have the appropriate permissions in your S3 account. <div> Tip: When you open a new folder, a reference to it is added to the Path value. You can modify the path value manually, which may be a faster way to navigate up a deep directory structure. </div> <div> Tip: Sizes are displayed next to files. They are not displayed next to folders. </div>
	File	Click the Plus icon to select this file.

Specify S3 Path:

In the S3 browser, you can specify an explicit path to resources. Click the Pencil icon, paste the path value, and click **Go**.

For example, if your home input directory is the following:

```
/mydir/input/username@example.com
```

You should paste the following in the Path textbox:

```
<bucketname>/mydir/input/username@example.com
```

NOTE: The name of the bucket (<bucketname>) must appear at the beginning of the path. Do not add a backslash (/) as a prefix.

Tip: You can retrieve your home directory from your profile. See *Storage Config Page*.

Search Files:

To display a subset of files, enter a string in the Search box. The filter is applied as you type and matches anywhere in the name of a currently displayed file or folder.

NOTE: If you have a folder and file with the same name in S3, search only retrieves the file. You can still navigate to locate the folder.

File Import Settings

Contents:

- *Per-file encoding*
 - *Detect structure*
 - *Remove special characters from column names*
 - *Selecting column headers*
-

When you edit settings on a selected file in the Import Data page, the following settings are displayed.

You can edit any additional or optional settings for an individual dataset. Perform the following:

1. Click **Edit Settings** from the card for an individual dataset in the right panel. The dialog box is displayed.
2. In the dialog box, select the required options and modify the settings.

Per-file encoding

By default, Trifacta applies a specified encoding type on the imported the file. In some cases, the data preview panel may contain garbled data, due to a mismatch in encodings. In the Data Preview dialog, you can select a different encoding for the file. When the correct encoding is selected, the preview displays the data as expected.

NOTE: Assessing the file encoding type based on parsing an input file is not an accurate method. Instead, Trifacta assumes that the file is encoded in the default encoding. If it is not, you should change the encoding type for the file.

NOTE: In some cases, imported files are not properly parsed due to issues with encryption types or encryption keys in the source datastore. For more information, please contact your datastore administrator.

For a list of supported encoding types, see *Supported File Encoding Types*.

Detect structure

By default, Trifacta attempts to interpret the structure of your data during import. This structuring attempts to apply an initial tabular structure to the dataset.

- Unless you have specific problems with the initial structure, you should leave the Detect structure setting enabled. Recipes created from these imported datasets automatically include the structuring as the first, hidden steps. These steps are not available for editing, although you can remove them through the Recipe panel. See *Recipe Panel*.
- When detecting structure is disabled, imported datasets whose schema has not been detected are labeled, **unstructured datasets**. When recipes are created for these unstructured datasets, the structuring steps are added into the recipe and can be edited as needed.
- For more information, see *Initial Parsing Steps*.

Remove special characters from column names

When selected, characters that are not alphanumeric or underscores are stripped, and space characters are converted to underscores. For more information, see *Sanitize Column Names*.

Selecting column headers

You can apply the column headers to your datasets during import. Select the required option from the drop-down list:

- **Infer header:** (default) When selected, the Trifacta application infers the header based on the data in the import.
- **Use first row as header:** When selected, the first row is used as the column headers.
- **No header:** When selected, the inference is ignored and column headers are defined using generic names with no headers.

If replacing a file:

- If you replace a dataset in a flow and select the **Use first row as header** option, then the existing header row labels are updated with the new headers.
- Subsequent steps in a pre-existing recipe may be broken if the headers are changed by a replaced file.

Tip: After the dataset is imported, you can rename columns manually or using any row in the dataset. For more information, see *Rename Columns*.

Table Import Settings

When you edit settings for a selected table in the Import Data page, the following settings are displayed.

You can edit any additional or optional settings for an individual dataset. Perform the following:

1. Click **Edit Settings** from the card for an individual dataset in the right panel. The dialog box is displayed.
2. In the dialog box, select the required options and modify the settings.

Infer column data types

You can choose whether or not to apply Trifacta type inference to table data imported from a database.

- In the preview panel, you can see the data type that is to be applied after the dataset is imported. This data type may change depending on whether column data type inference is enabled or disabled for the dataset.
- To enable Trifacta type inference, select the Infer column data types checkbox.

Tip: To see the effects of Trifacta type inference, you can toggle the checkbox and review data type listed at the top of individual columns. To override an individual column's data type, click the data type name and select a new value.

For more information, see *Disable Type Inference*.

You can configure the default use of type inference at the individual connection level. For more information, see *Create Connection Window*.

Macros Page

In the Macros page, you can review and manage the macros to which you have access.

A **macro** is a saved sequence of one or more recipe steps that can be reused in other recipes. Values in your macros can be parameterized. For more information, see *Overview of Macros*.




Macros			Find Macros...
Name	Used in	Last Updated ▾	
 InitialCleanv3 Initial steps cleanup v3	1 Flow • 1 Recipe	Today at 5:36 PM	
 InitialCleanV2 Initial cleanup steps	0 Flows • 0 Recipes	Today at 5:34 PM	
 InitialClean Initial recipe cleanup steps	0 Flows • 0 Recipes	Today at 5:31 PM	

Figure: Macros Page

To review specifics about the macro, click its name. See *Macro Details Page*.

Columns:

- **Name:** Name of the macro.
- **Used in:** Count of flows and recipes in which the macro is used.
- **Last Updated:** Timestamp for when the macro was last modified.

Actions:

- **Import Macro:** Click to import a macro that has been exported to your local desktop. For more information, see *Import Macro*.
- **Get Macros:** Explore and download macros from Wrangle Exchange.
- **Search:** Enter a string in the search box. The list of macros is updated in real-time.
- **Sort:** Click the caret next to any column head to sort the list based on the column.
- **Context Menu:** See below.

Context Menu Options:

- **Edit:** Modify the name and description for the macro. You can also modify the name, description, and default values for the macro's inputs.
- **Inspect:** Review the recipe steps in the macro.
- **Export:** Export the macro to your local desktop. For more information, see *Export Macro*.
- **Replace:** Replace an existing macro definition with a macro that you have exported to your local desktop.

NOTE: If your imported macro contains macro inputs that are not in the macro that you are replacing, the existing instances of the replaced macro contain a broken step where the macro is referenced but has no data. These references must be fixed in each macro instance.

Tip: If you must add more macro inputs or steps to a macro that you have imported, you must convert the macro to steps, modify them, and then perform a replacement. For more information, see *Create or Replace Macro*.

- **Delete:** Delete the macro.

NOTE: When a macro is deleted, in any recipe that references it, the macro's steps are expanded into regular recipe steps. Any macro inputs are applied as static values in the expanded recipe steps.

Deleting a macro cannot be undone.

Macro Details Page

Through the Macro Details Page, you can review details about an individual macro. In the Macros page, click the name of the macro to review.

Actions:

To modify the macro name and description, click **Edit**.

Tip: You can also modify the name, description, and default values for the macro's inputs.

Context menu:

- **Export:** Export the macro to your local desktop. See *Export Macro*.
- **Delete:** Delete the macro.

NOTE: When a macro is deleted, in any recipe that references it, the macro's steps are expanded into regular recipe steps. Any macro inputs are applied as static values in the expanded recipe steps.

Deleting a macro cannot be undone.

Overview Tab

In the Overview tab, you can review the steps in the macro.

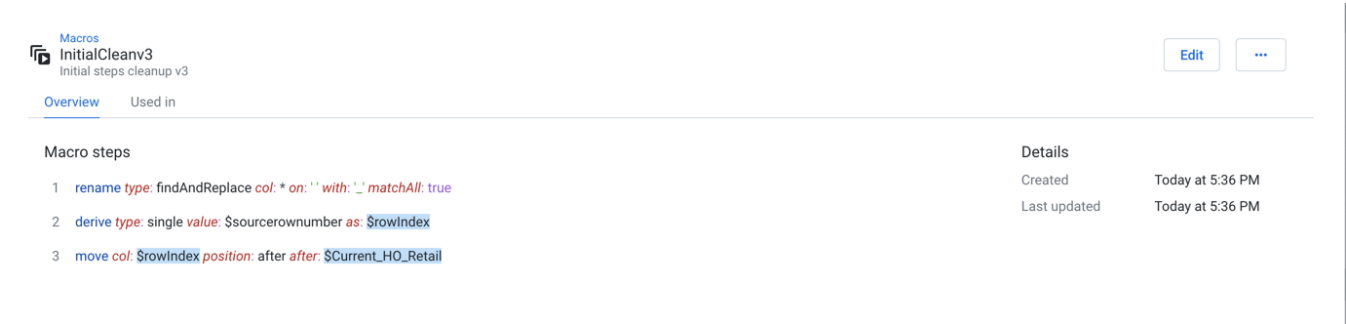


Figure: Macro Detail Page - Overview tab

- Steps are displayed in raw Wrangle .
- You can review creation and update timestamps.

Used In Tab

In the Used In tab, you can review all of the recipes and flows where the macro is referenced.

<div> <div> <div>Macros</div> <div>InitialCleanv3</div> <div>Initial steps cleanup v3</div> </div> <div> <div>Edit</div> <div>...</div> </div> </div>		
<div> <div>Overview</div> <div>Used in</div> </div>		
Name	Flow	Last Updated ▾
<div> <div></div> <div>POS-r01</div> </div>	<div> <div></div> <div>POS-r01 Flow</div> </div>	Today at 5:36 PM

Figure: Macro Detail Page - Used In tab

- Click the name of the recipe or flow to open the flow. See *Flow View Page*.

Connections Page

Contents:

- *Top Bar*
- *Connection Context Menu*
- *Connection Details Panel*

Through the Connections page, you can add new connections or modify the connections that you have already created. From the left nav bar, click the Connections icon.

NOTE: Access to the Connections page in the application and privileges on connections is governed by roles in your workspace. For more information, please contact your workspace administrator.

Connections

Create Connection

Search connections

All connections

Owned by me

Shared with me


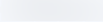
Name	Type	Shared	Owner	Last updated
 s3user	s3user	Private		04/15/2021
				<div>Share</div> <div></div>

Figure: Connections page

Fields:

- **Name:** Display name for the connection.

You can hover over the shared icon link next to the connection name to view the name of shared users (up to three shared user) and the total number of shared users. Also, when you click the shared icon link, the share dialog is displayed.

NOTE: If the connection has been shared, you can review whether its credentials have also been shared.

- **Type:** The type of connection.

NOTE: After you create a connection, you cannot modify its type.

For more information, see *Connection Types*.

- **Shared:** Review the sharing status of the connection:
- Global - connection has been shared with all users of the workspace.

NOTE: To make a global connection private, you must delete the connection and recreate it.

- **X Users:**
 - If this value is 1, the connection is private.
 - If this value is greater than 1, the connection has been shared. Click the link in this column to review sharing status. See *Share Connection Dialog*.

Top Bar

- **Create:** Click **Create Connection** to create a new connection. See *Create Connection Window*.
- **Filter:** In separate tabs, you can review connections that you own, that are shared with you, or all connections to which you have access.
- **Search:** Search connections by name.
- **Review details:** Select a connection or click the icon to review details through the right-side panel.

Connection Context Menu

- **Share:** For connections that you own, you can modify the sharing status of them. See *Share Connection Dialog*.
- **View Details:** Open the details of the connection in the side panel. See below.
- **Edit:** If you own the connection, you can review and modify the connection.
 - If the connection has been shared with you, you can edit it to modify the credentials.
 - Administrators can edit public connections.
 - See *Create Connection Window*.
- **Delete:** Delete the connection.


NOTE: This option is only available to the connection owner if the connection is not used for any datasets.

Connection Details Panel

When a connection is selected, you can review its details and make modifications as needed through the panel on the right.

Connection Details

×

 postgres

Edit Connection

...

Connection Type

postgres

Shared

Private

Owner

SteveO

Created

Today at 5:28 PM

Updated

Today at 5:28 PM

Updated by

SteveO

Server Information

Host

myHost

Port

5432

SSL

Disabled

Database

myDB

Username

myUserId

Password

.....

Figure: Connection Details panel

Key Fields:

- **Shared:** Number of users sharing the connection. If a link is present, click it to modify sharing of the connection. See *Share Connection Dialog*.
- **Server Information:** For server-based connections, you can review the connection properties.

Actions:

- **Edit Connection:** If you own the connection, you can review and modify the connection.
 - If the connection has been shared with you, connection properties are read-only.
 - See *Create Connection Window*.
- **Share:** You can share connections that you own or that are shared with you. See *Share Connection Dialog*.
- **Delete:** Delete the connection.

Deleting a connection cannot be undone.

Share Connection Dialog

Contents:

- *Actions*
- *Find Users*
- *Set Access Level*
- *Privileges*
- *Privacy*
- *Credentials*

Through the Share Connection dialog, users of the selected connection with appropriate privileges can modify who has access to the connection.

Tip: A workspace administrator has owner-level access to all connections in the workspace. However, a workspace admin cannot access or use a connection's credentials if those credentials were not shared by the owner of the connection. For more information, see *Workspace Admin Permissions*.

Share

×

Invite users

Editor

▼

Share

Shared with (2 users)

A

Administrator (you)

A

Abarnan

Editor

Can read, update, and share

Viewer

Can read and share

Owner

Editor

▼

🔒

Invite only

▼

Only users who have been invited can access this connection

🔑

Do not share credentials

▼

Users will need to provide their own credentials when accessing this connection

Figure: Share Connection Dialog

Actions

Find Users

Start typing names or email addresses of users to see matches.

NOTE: You may not be permitted to share objects with users who have not yet logged into the product.

Tip: You can paste a comma-separated list of email addresses to share to multiple users at the same time.

You may be able to browse a list of all usernames.

NOTE: This feature may need to be enabled in your environment. For more information, see *Workspace Settings Page*.

Set Access Level

As needed, you can configure the level of access to the connection for users with whom the connection is shared.

NOTE: You cannot set a user's access to a level that is higher than the limit set for the user at the workspace level. For example, if the user has Viewer access to connections at the workspace level, you cannot make the user an Editor on your connection.

NOTE: Administrators have owner-level access to all connections in the workspace or project. You do not need to share connections with them.

Privileges

- **Editor:**
 - User can use the connection to read data, update the connection, and share the connection.
 - User has all Viewer privileges.

NOTE: Editors cannot delete connections. Only the owner or an admin can delete a connection.

- **Viewer:**
 - User can use the connection to read data.
 - User can share connection.

For more information, see *Overview of Sharing*.

Privacy

- **Invite only:** Connection can be made available to other users only by invitation through this window.
- **Public :** Connection is available to all workspace users who can access connections.

NOTE: Only an administrator can make a connection public.

NOTE: After a connection is made public, it cannot be made private again. It must be deleted and recreated.

Credentials

By default, a connection is shared with credentials. Optionally, sharing of credentials can be disabled when sharing.

NOTE: The choice to share credentials or not is applied to all users with whom the connection has been shared, including users with whom the connection has been shared previously.

NOTE: Connections that use OAuth 2.0 authentication cannot be shared with credentials.

- **Share credentials:** When selected, the credentials that are specified in the owner's connection definition are shared with other users.

NOTE: Password values are always masked in the interface.

- **Do not share credentials:** When this option is selected, users of the shared connection must provide their own credentials.

NOTE: To use datasets previously imported through the shared connection, these credentials must provide access the source data. If shared credentials are removed from a connection, then any datasets imported through the connection are not accessible until credentials are provided. This may also apply to re-publishing previously generated results.

When credentials are shared:

NOTE: Users to whom credentials are shared cannot see any passwords in the Trifacta application.

- The credentials specified in the connection are shared to the users who are specified in the Share dialog for connections.
 - Users of a shared connection with credentials cannot insert their own credentials. They must create a new connection.
- Sharing of credentials may not guarantee access to the same locations as available to the owner.
Examples:
 - If your deployment uses Single Sign On, your enterprise login may provide access controls to the same resource that are different from the connection owner.
 - Network infrastructure may whitelist IP addresses for some users and block the same addresses for others.
 - Depending on the datastore, folder or directory permissions may limit access.
 - For more information, please contact your IT administrator.
- The owner of the connection can specify whether credentials are shared or not.
 - A workspace administrator has owner-level access to all connections in the workspace. However, a workspace admin cannot access or use a connection's credentials if those credentials were not shared by the owner of the connection.
- Shared users of the connection can share the connection if they have Editor privileges.

When credentials are not shared:

- Each user must provide credentials to use the connection.
 - A user's individual credentials may not provide read access to datasources, which may mean that imported datasets appear to be broken.
 - Individual credentials may not provide write access to the same output locations, which may cause jobs to fail.
- When sharing of credentials is disabled, shared users who share with other users cannot include credentials as part of the share.

Jobs Page

In the Jobs page, you can track the status of all of your jobs and plan runs.

- **Sample jobs:** For more information on sampling jobs, see *Sample Jobs Page*.
- **Plan runs:** For more information on plan runs, see *Plan Runs Page*.

You can only see jobs for the flows to which you have access in your current environment.

Jobs can be initiated from:

- Flow View: See *Flow View Page*.
- Transformer page: See *Transformer Page*.

Flow jobs

<<

<

1

2

>

>>

⌵

Search jobs

/

All jobs

Completed

Failed

Canceled

Running

Queued

Job	Status	Flow	User	Started
<div><div>⌚</div><div>POS-r01</div><div>Job ID: 2135347</div></div>	<div><div>⌚</div><div>In progress</div></div>	<div><div>🔗</div><div>2013 POS</div></div>	Abarna Nagarajan (you)	Today at 9:51 AM <div><div>Cancel job</div><div>⋮</div></div>
<div><div>⌚</div><div>POS-r01</div><div>Job ID: 2135265</div></div>	<div><div>✅</div><div>Completed</div></div>	<div><div>🔗</div><div>2013 POS</div></div>	Abarna Nagarajan (you)	Today at 9:38 AM Ran for 5 minutes
<div><div>⌚</div><div>POS-r01</div><div>Job ID: 2130018</div></div>	<div><div>⚠️✅</div><div>Completed with warnings</div></div>	<div><div>🔗</div><div>2013 POS</div></div>	Abarna Nagarajan (you)	Yesterday at 3:05 PM Ran for 7 minutes
<div><div>⌚</div><div>POS-r01</div><div>Job ID: 2129615</div></div>	<div><div>⚠️✅</div><div>Completed with warnings</div></div>	<div><div>🔗</div><div>2013 POS</div></div>	Abarna Nagarajan (you)	Yesterday at 1:15 PM Ran for 8 minutes
<div><div>⌚</div><div>POS-r01</div><div>Job ID: 2129611</div></div>	<div><div>⚠️✅</div><div>Completed with warnings</div></div>	<div><div>🔗</div><div>2013 POS</div></div>	Abarna Nagarajan (you)	Yesterday at 1:06 PM Ran for 7 minutes
<div><div>⌚</div><div>POS-r01</div><div>Job ID: 2127857</div></div>	<div><div>⚠️✅</div><div>Completed with warnings</div></div>	<div><div>🔗</div><div>2013 POS</div></div>	Steve Olson	Yesterday at 4:04 AM Ran for 7 minutes
<div><div>⌚</div><div>Untitled recipe</div><div>Job ID: 2105451</div></div>	<div><div>❌❌</div><div>Canceled</div></div>	<div><div>🔗</div><div>[cf8d8640] test1</div></div>	Abarna Nagarajan (you)	Last Friday at 5:40 PM Ran for 5 minutes
<div><div>⌚</div><div>POS-r01</div><div>Job ID: 1936242</div></div>	<div><div>✅</div><div>Completed</div></div>	<div><div>🔗</div><div>2013 POS</div></div>	Steve Olson	04/24/2021 Ran for 8 minutes

Figure: Jobs page

Job Types:

Each job listed in the Jobs page is a grouping of related jobs acting on the same recipe and dataset(s). Each of these **jobgroups** breaks down into one or more of the following job types.

Tip: To review the status of individual jobs within a jobgroup, hover over the icons in the Status column for the jobgroup.

- **Pre-ingest SQL:** These jobs are SQL scripts that execute before the source data is ingested to the platform.
 - For additional details on these jobs, see the SQL scripts tab in the Job Details page. See *Job Details Page*.
 - For more information on these types of SQL scripts, see *Create Output SQL Scripts*.
- **Transform:** These jobs perform transformations on imported datasets based on the recipe from which the job was launched.

- **Profile:** If enabled as part of the job definition, a Profile job generates a visual summary of the results of your transformation job.
 - Profiling jobs may take longer than transformation jobs.
 - Even when selected, profiling jobs may not appear in the Jobs page. In some cases, a profiling job may be folded into a transform job for optimization reasons.

NOTE: When the profiling job is run as part of the transform job, there is no listing for profiling in the mouse-over popup.

- See *Job Details Page*.
- **Publish:** Depending on multiple factors, your job may include a second Publish job that occurs after the Transform job. For example, job groups can include internal Publish jobs for writing results to the designated location in the base storage layer.

Publishing can also be executed as a separate, post-execution job. As needed, job results can be published from their target location to another location or data store. These jobs are tracked separately as Publish jobs and can be launched from the Job Details page.

- **Ingest:** For larger datasets from some relational connections, the platform transfers the data from the source to the default storage layer for faster processing. These ingest jobs occur before any transform or profiling takes place.
- **Post-ingest SQL:** These jobs are SQL scripts that execute after the job results have been published.
 - For additional details on these jobs, see the SQL scripts tab in the Job Details page. See *Job Details Page*.
 - For more information on these types of SQL scripts, see *Create Output SQL Scripts*.

Tabs and Statuses:

Each of the available tabs corresponds to a possible status for jobs that have been initiated on the platform.

- **All jobs:** All jobs that you have initiated are listed here.
- **Completed:** Job has successfully executed.

NOTE: Invalid steps in a recipe are skipped, and it's still possible for the job to be executed successfully.

NOTE: A warning icon may indicate that recipe errors were detected during the Transform phase. You can hover over the icon for more information.

- **Failed:** job failed to complete.

NOTE: You can re-run a failed job from the Transformer page. If you have since modified the recipe, those changes are applied during the second run. See *Transformer Page*.

- **Publish Failed:** Some failed jobs may be listed under this status, which means that the publishing step of the configured job failed to complete.
- **Canceled:** Job was canceled.
- **Running:** Job is in progress.
- **Queued:** Job has been queued for execution.

Access:

- You can review and drill into any job that you initiated.
- You can also drill into any job that was initiated from a flow that has been shared with you.

Columns:

- **Job:** Internal identifier for the job. This value is unique for all jobs in your Trifacta® instance.
 - Click the ID number to explore details about the job.
- **User:** The Trifacta user that initiated the job.
- **Run from:**
 - Location where the job was launched. Click the link to view details.
- **Status:**
 - See Tabs above.
- **Started:** Start timestamp for the job.
 - Scheduled jobs are indicated with an icon.

Actions:

- **Filter by status:** Click one of the tabs to filter the display to show only the listings for the selected job status.
- **Filter by type and date:** Click the Funnel icon to filter the list of jobs by source of execution, date range, or both. See below.
- **Search:** Enter text in the search field to filter the listed jobs by job ID, flow name, or dataset name.

Context menu:

Next to the job listing, click the options menu to see the following:

- **Cancel Job:** Select to cancel a job that is currently being executed.
- **Delete job:** Delete the job from the platform.

Deleting a job cannot be undone.

NOTE: This feature may not be enabled in your instance of the platform. For more information, please contact your Trifacta Administrator. See *Miscellaneous Configuration*.

- **Download Logs:** Download the logs for the job. If the job is in progress, log information is likely to be incomplete.

Tip: When jobs fail, the downloaded package includes additional configuration files and service logs to assist in debugging job execution issues. For more information, see *Support Bundle Contents*.

Additional options are available for each job. See *Job Details Page*.

Filter Jobs

To filter the list of jobs based on dates or source of execution, click the Funnel icon. You can use the following dialog to filter the display of jobs.

Filter Jobs

×

Job type

Only show manual jobs

▼

Date/Time

Started

Started After

▼

MM/DD/YYYY

📅

HH:MM

AM

▼

Ended

Ended Before

▼

MM/DD/YYYY

📅

HH:MM

AM

▼

Clear Filters

Cancel

Apply

Figure: Filter Jobs dialog

Job type:

Show jobs based on the following available options:

- Show all jobs
- Only show manual jobs
- Only show scheduled jobs

Started:

- Specify the date and time when the jobs to display began.
- If needed, you can specify the start time as a range. Select `Start Between` from the drop-down list and populate both date-time rows.

Ended:

- Specify the date and time when the jobs to display ended.
- If needed, you can specify the end time as a range. Select `Ended Between` from the drop-down list and populate both date-time rows.

Actions:

- To clear the time period values, click **Clear Filters**.
- To apply the specified time filter to the Jobs page, click **Apply**.

Job Details Page

Contents:

- *Overview Tab*
 - *Output Destinations Tab*
 - *Direct file download*
 - *Create imported dataset*
 - *Publish*
 - *SQL scripts Tab*
 - *Profile Tab*
 - *Dependency graph Tab*
 - *Data sources Tab*
 - *Parameters Tab*
 - *Webhooks Tab*
-

You can use the Job Details page to explore details about successful or failed jobs, including outputs, dependency graph, and other metadata. Download results to your local desktop or, if enabled, explore a visual profile of the data in the results for further iteration on your recipe.

Page options:

- **Cancel job:** Click this button to cancel your job while it is still in progress.

NOTE: This option may not be available for all running environments. Job cancellation is not supported in high availability deployments.

- **Publish results:** Publish your results to an external system. For more information, see *Publishing Dialog*.
- **Delete job:** Delete the job and its results.

Deleting a job cannot be undone.

NOTE: This feature may not be enabled in your environment. For more information, see *Miscellaneous Configuration*.

- **Download logs:** Download the log files associated with this job.

Tip: When jobs fail, the downloaded package includes additional configuration files and service logs to assist in debugging job execution issues. For more information, see *Support Bundle Contents*.

- **Download profile as PDF:** If visual profiling was enabled for the job, you can download the profile in PDF format.
- **Download profile as JSON:** If visual profiling was enabled for the job, you can download a JSON representation of the profile to your desktop.

Overview Tab

In the Overview tab, you can review the job status, its sources, and the details of the job run.

NOTE: If your job failed, you may be prompted with an error message indicating a job ID that differs from the listed one. This job ID refers to the sub-job that is part of the job listed in the Job summary.

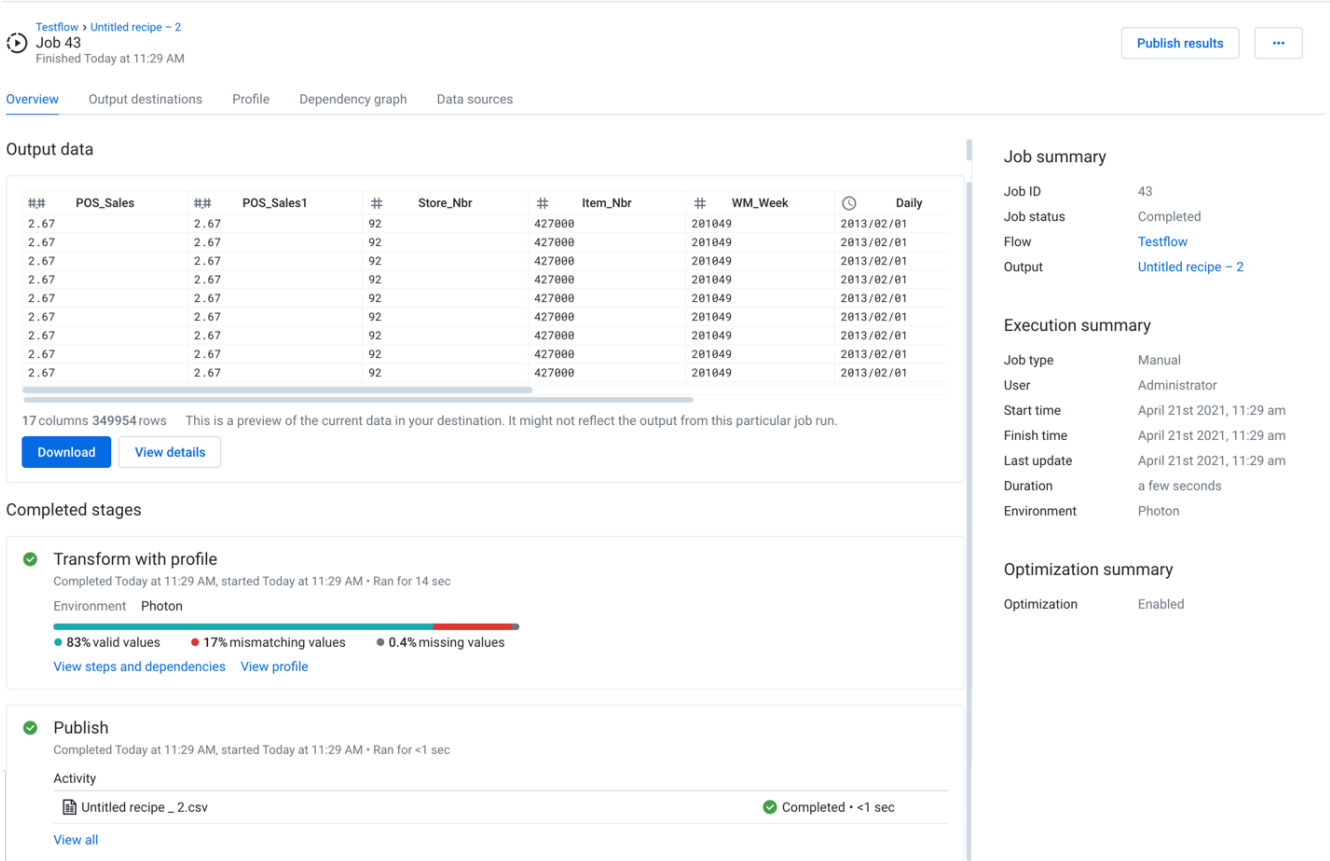


Figure: Overview tab

You can review a snapshot of the results of your job.

Output Data:

The output data section displays a preview of the generated output of your job.

NOTE: This section is not displayed if the job fails.

You can also perform the following:

- View:** If it is present, you can click the View link to view the job results in the datastore where they were written.

NOTE: The View link may not be available for all jobs.

- **Download** : If it is present, click the Download link to download the generated job results to your local desktop.
- **View details**: Click **View details** to view the generated results in the side bar. See the Output Destinations below.

Completed Stages:

This panel provides information on the progress and completion status of each stage of the job execution.

Tip: Depending on the operation, you may be able to monitor transfer rate performance for larger datasets.

- **Connect:** The platform is attempting to connect to the datastore hosting the asset sources for the datasets.
- **Schema validation:** When enabled, the schemas of a job's datasources are checked as the first step of job execution.
 - Datasets with changes in them are reported at the top of the list. Click **View all** to see schema validation for all of the datasets used in the job in the Data sources tab.
 - Optionally, the job can be halted if there are differences between the schema that is read and the schema that has been stored from the previous job run. This option can prevent data corruption. For more information, see *Run Job Page*.
 - If no errors are detected, then the job is completed as normal.
 - For more information on schema validation, see *Overview of Schema Management*.
- **Request:** The platform is requesting the set of assets to deliver.
- **Ingest:** Depending on the type of source data, some jobs ingest data to the base storage layer in a converted format before processing begins. This ingested data is purged after job completion.
- **Prepare:** (Publishing only) Depending on the destination, the Prepare phase includes the creation of temporary tables, generation of manifest files, and the fetching of extra connections for parallel data transfer.
- **Transfer:** Assets are transferred to the target, which can be the platform or to the output datastore.
- **Transform:** This stage covers the execution of your recipe steps in order to transform the source data.
- **Profile:** If you chose to profile your output data, this stage is completed after transformation is complete. Results are available in the Profile tab.

NOTE: If you chose to generate a profile of your job results, the transformation and profiling tasks may be combined into a single task, depending on your environment. If they are combined and profiling fails, any publishing tasks defined in the job are not launched. You may be able to ad-hoc publish the generated results. See below.

- **Publish:** This stage covers the writing of the outputs of the transformed data. These outputs are available through the Output destinations tab.
- **Process:** Cleanup after data transfer, including the dropping of temporary tables or copying data within the instance.

For more information, see *Overview of Job Monitoring*.

If present, you can click the **Show Warnings** link to see any warnings pertaining to recipe errors, including the relevant step number. To review the recipe and dependencies in your job, click **View steps and dependencies**. See the Dependencies tab below.

- If you chose to profile results of your job, click **View profile** to review. See Profile tab below.
 - A visual profile provides a graphical snapshot of the results of a successful transformation job for the entire dataset and individual columns in the dataset.
 - For more information on enabling a visual profile job, see *Run Job Page*.
 - For more information, see *Overview of Visual Profiling*.

- If your job output specified SQL scripts to run before or after job execution, you can track their progress in the following stages:
 - **Pre-ingest SQL:** Script that is configured to run before the source data is ingested to the platform.
 - **Post-publish SQL:** Script that is configured to run after the output data has been published.
 - For additional details, see the SQL scripts tab below.
 - For more information on SQL scripts in job execution, see *Create Output SQL Scripts*.

Publish:

You can also review the outputs generated as a result of your job. To review and export any of the generated results, click **View all**. See Outputs Destinations tab below.

Job summary:

- **Job ID:** Unique identifier for the job

Tip: If you are using the REST APIs, this value can be used to retrieve and modify specifics related to this job. For more information, see *API Reference*.

- **Job status:** Current status of the job:
 - **Queued:** Job has been queued for execution.
 - **Running:** Job is in progress.
 - **Completed:** Job has successfully executed.

NOTE: Invalid steps in a recipe are skipped, and it's still possible for the job to be executed successfully.

- **Failed:** Job failed to complete.

NOTE: You can re-run a failed job from the Transformer page. If you have since modified the recipe, those changes are applied during the second run. See *Transformer Page*.

- **Canceled:** Job was canceled.

- **Flow:** Name of the flow from which the job was executed. Click the link to open the flow. See *Flow View Page*.
- **Output:** Name of the output object that was used to define the generated results. Click the link to open the output. See *Flow View Page*.

Execution summary:

- **Job type:** The method by which the job was executed:
 - **Manual** - Job was executed through the application interface.
 - **Scheduled** - Job was executed according to a predefined schedule. See *Add Schedule Dialog*.

User: The user who launched the job

Environment: Where applicable, the running environment where the job was executed is displayed.

Start time: Timestamp for when processing began on the job. This value may not correspond to when the job was queued for execution.

Finish time: Timestamp for when processing ended on the job, successful or not

Last update: Timestamp for when the job was last updated

Duration: Elapsed time of job execution

Optimization summary:

For jobs sourced from relational datasets, you can optionally enable SQL-based optimizations, which apply some of the steps specified in your recipe back in the datasource, where they can be executed before the data is transferred to the running environment for execution. Using these optimizations means faster performance based on a lower volume of data transfer.

Workspace administrators must enable the optimization feature for the workspace. For more information, see *Workspace Settings Page*.

When the feature is enabled, optimizations must be enabled for each flow. You can also select the optimizations to apply. For more information, see *Flow Optimization Settings Dialog*.

When optimizations have been applied to your flow, they are listed on the Overview tab:

Optimization: This setting is displayed if flow optimizations have been enabled for this flow.

Columns pruned: If one or more unused columns have been pruned in the datasource via SQL, the count of columns is listed here.

Filters pushed down: If one or more row filters has been applied in the datasource via SQL, the count of filters is listed here.

If an optimization is disabled or was not applied to the job run, it is not listed.

Output Destinations Tab

If the job has successfully completed, you can review the set of generated outputs and export results.

test_flow > REF_PROD - 3

Job 60

Finished Today at 4:16 PM

Publish results

...

Overview **Output destinations** Profile Dependency graph Data sources

You can download the generated results locally or **publish** to another storage location.

Name	Location	Status
REF_PROD _ 3.csv	hdfs://hadoop:8020/trifacta/queryResults/admin@trifacta.local/REF_PROD _ 3.csv	Completed • 13 sec

Figure: Output Destinations tab

Actions:

For each output, you can do the following:

View details: View details about the generated output in the side bar.

Tip: The View details panel contains breakdowns for each phase of a job. If the job fails, you can review error messages, which correspond to entries in the Data Service log file.

Download result: Download the generated output to your local desktop.

NOTE: Some file formats may not be downloadable to your desktop. See below.

Create imported dataset: Use the generated output to create a new imported dataset for use in your flows. See below.

NOTE: This option is not available for all file formats.

Direct file download

Click one of the provided links to download the file through your browser to your local desktop.

NOTE: If these options are not available, data download may have been disabled by an administrator.

HYPER: You can download HYPER formatted outputs to your desktop.

If you have generated output in a Tableau format and have configured a connection to Tableau Server, you can publish directly to the server. See *Publishing Dialog*.

Create imported dataset

Optionally, you can turn your generated results into new datasets for immediate use in Trifacta. For the generated output, select **Create imported dataset** from its context menu.

NOTE: If you generated results in Parquet format only, you cannot create a dataset from it, even if the Create button is present. This is a known issue.

NOTE: When you create a new dataset from your job results, the file or files that were written to the designated output location are used as the source. Depending on your backend datastore permissions are configured, this location may not be accessible to other users.

After the new output has been written, you can create new recipes from it. See *Build Sequence of Datasets*.

Publish

If Trifacta is connected to an external storage system, you may publish your job results to it. Requirements:

Your version of the product supports publishing.

Your connection to the storage system includes write permissions.

Your results are generated in a format that the target system supports for writing.

All sub-jobs, including profiling, successfully completed.


For more information, see *Publishing Dialog*.

SQL scripts Tab

If the output for your job included one or more pre- or post-job SQL script executions, you can review the status of their execution during the job.

NOTE: If a SQL script fails to execute, all downstream phases of the job fail to execute.

Tip: If the SQL script execution for this job encountered errors, you can review those errors through this tab. For more detailed information, click **Download logs**.



The screenshot shows a web interface for a job named 'Job 19' (2013 POS > Untitled recipe), finished today at 5:48 PM. It features a 'Publish results' button and a menu icon. Below is a navigation bar with tabs: Overview, Output destinations, SQL scripts (selected), Dependency graph, and Data sources. The main content is a table with four columns: Connection, SQL statement, Settings, and Status. It lists two SQL scripts executed on a 'postgres' connection. Both scripts are 'CREATE TABLE IF NOT EXISTS' statements for a table named 'spotable' with columns for timestamp, jobType, and jobStatus. The first script is set to 'Run before data ingest' and is 'Completed' in 3 seconds. The second script is set to 'Run after data publish' and is 'Completed' in less than 1 second.

Connection	SQL statement	Settings	Status
postgres	CREATE TABLE IF NOT EXISTS "public"."spotable" (timestamp date, jobType varchar(255), jobStatus varchar(255)); INSERT INTO "public"."spotable"(timestamp, jobType, jobStatus) VALUES ('2021-06-22','Trifacta','started');	Run before data ingest	Completed • 3 sec
postgres	CREATE TABLE IF NOT EXISTS "public"."spotable" (timestamp date, jobType varchar(255), jobStatus varchar(255)); INSERT INTO "public"."spotable"(timestamp, jobType, jobStatus) VALUES ('2021-06-22','Trifacta','completed');	Run after data publish	Completed • <1 sec

Figure: SQL scripts tab

Columns:

Connection: Name of the connection through which the script was executed.

SQL statement: The first part of the SQL script that was executed.

Settings:

- Run before data ingest - script was executed pre-job.
- Run after data publish - script was executed post-job, after the job results had been written.

Status: Current status and execution duration of the SQL script.

NOTE: If you have multiple SQL scripts for each settings, they may execute in parallel. For example, if you created three pre-job SQL scripts, there is no guarantee that they executed in the order in which they are listed.

View details:

Hover over a SQL script entry and click **View details**.

In the SQL script details window, you can review:

Connection and SQL of the executed script.

Any error messages that occurred during execution.

Tip: To review log information for any error messages, click **Download logs**.

For more information on these types of SQL scripts, see *Create Output SQL Scripts*.

Profile Tab

Review the visual profile of your generated results in the Profile tab. Visual profiling can assist in identifying issues in your dataset that require further attention, including outlier values.

NOTE: This tab appears only if you selected to profile results in your job definition. See *Run Job Page*.

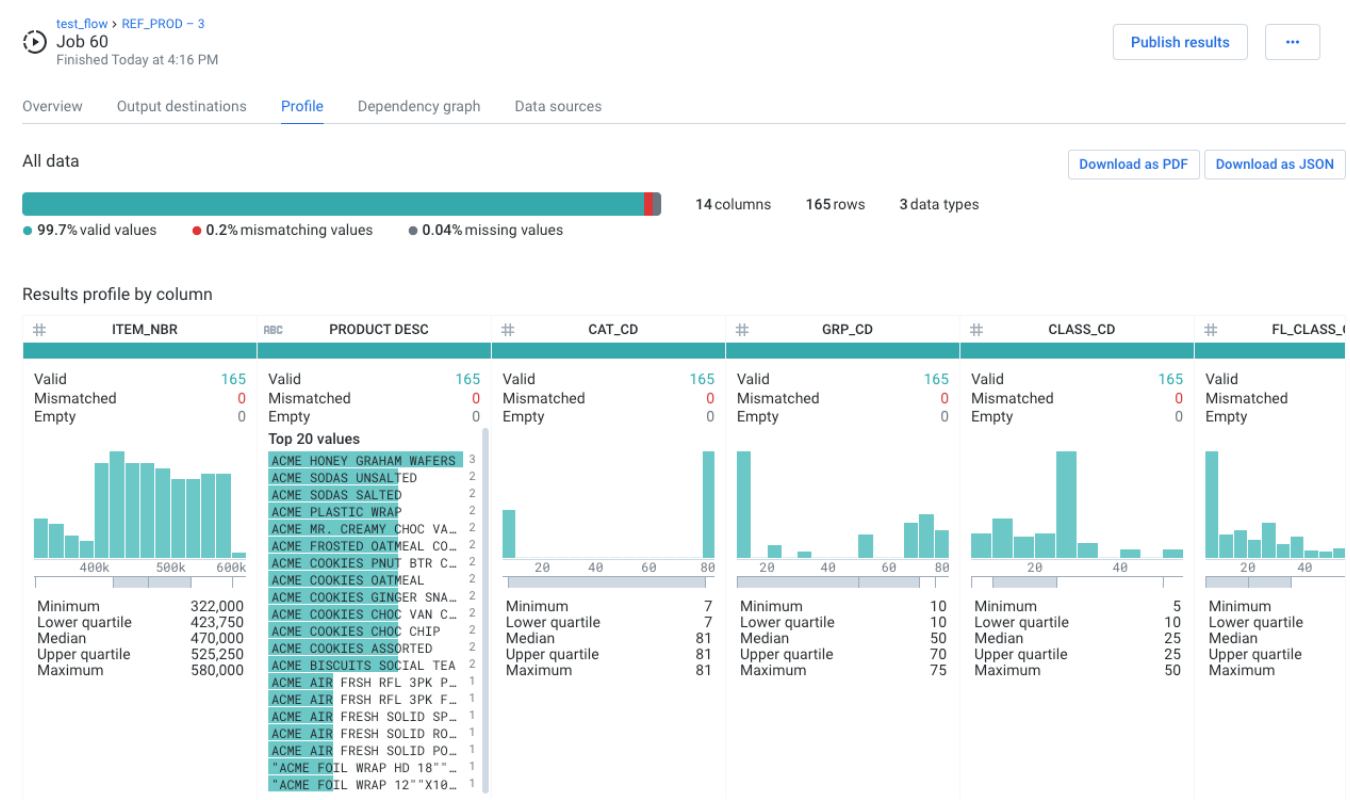


Figure: Profile tab

Download as PDF: Download your visual profile and results of your data quality rules on the entire dataset as a PDF file. For more information, see *Overview of Data Quality*.

Download as JSON: Download your visual profile as a JSON file.

In particular, you should pay attention to the mismatched values and missing values counts, which identify the approximate percentage of affected values across the entire dataset. For more information, see *Overview of Visual Profiling*.

NOTE: The computational cost of generating exact visual profiling measurements on large datasets in interactive visual profiles severely impacts performance. As a result, visual profiles across an entire dataset represent statistically significant approximations.

NOTE: Trifacta treats null values as missing values. Imported values that are null are generated as missing values in job results (represented in the gray bar). See *Manage Null Values*.

Tip: Mouse over the color bars to see counts of values in the category.

Tip: Use the horizontal scroll bar to see profiles of all columns in wide datasets.

In the lower section, you can explore details of the transformations of individual columns. Use this area to explore mismatched or missing data elements in individual columns.

Depending on the data type of the column, varying information is displayed. For more information, see *Column Statistics Reference*.

Tip: You should review the type information for each column, which is indicated by the icon to the left of the column.

Dependency graph Tab

In this tab, you can review a simplified representation of the flow from which the job was executed. This flow view displays only the recipes and datasets that contributed to the generated results.

Tip: To open the full flow, you can click its name in the upper-left corner.

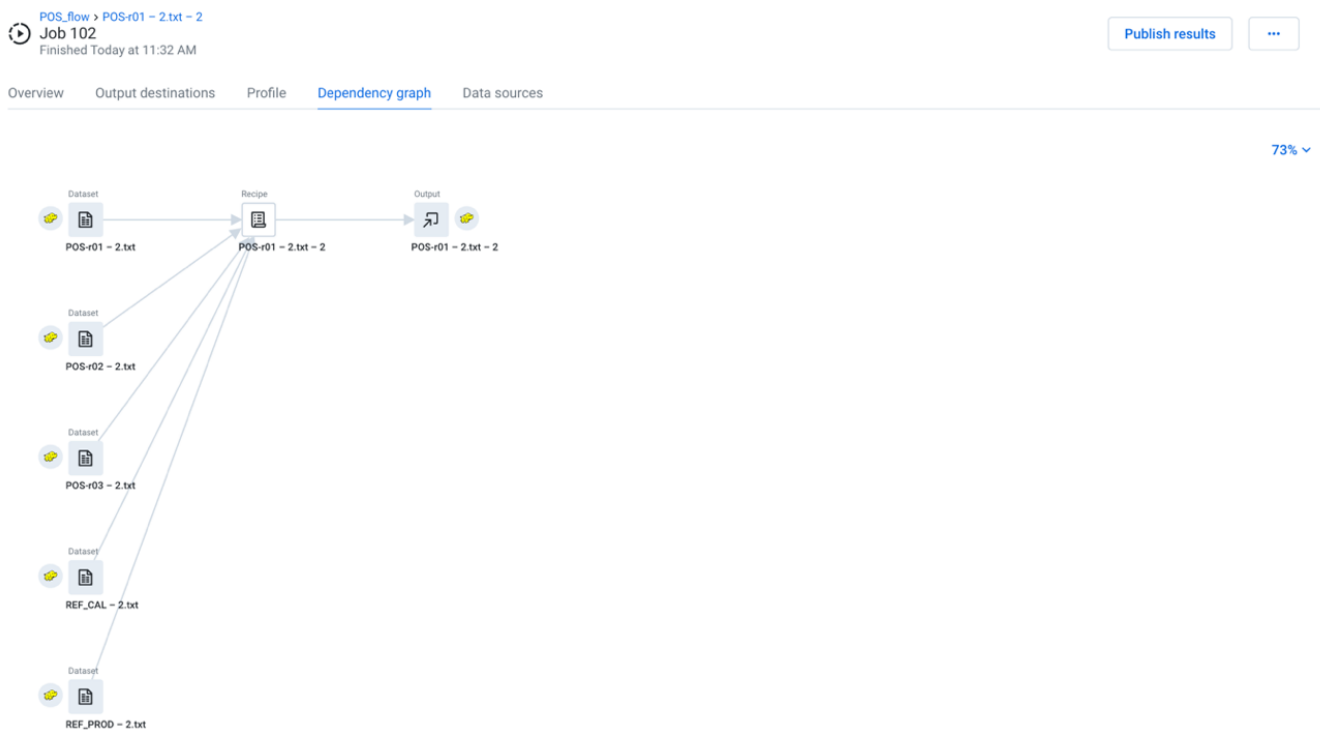


Figure: Dependency graph tab

Zoom menu:

You can zoom the dependency graph canvas to display areas of interest in the flow graph.

The zoom control options are available at the top-right corner of the dependency graph canvas. The following are the available zoom options:

Tip: You can use the keyboard shortcuts listed in the zoom options menu to make quick adjustments to the zoom level.

Zoom in: Zoom in 10% on the canvas to focus on greater detail.

Zoom out: Zoom out 10% from the canvas to see more of it.

Zoom to fit: Change the zoom level to fit all of the objects of your flow onto the screen.

25%, 50%, or 100%: Change the zoom level to one of the preset levels.

Recipe actions:

Download recipe: Download the text of the recipe in Wrangle .

Display Wrangle /natural language: Toggle display of the recipe in raw language or in readable language.

Limitations:

You can select only recipes in the flow graph.

Context controls and menus are not available.

Data sources Tab

In the Data sources tab, you can review all of the sources of data for the executing recipe.

Schema_drift > Untitled recipe

Job 2705

Finished Last Friday at 3:46 AM

Overview

Output destinations

Profile

Dependency graph

Data sources











Name	Location	Schema validation	Ingestion
 unload_all_data_types_del17	public/unload_all_data_types_del17	✔ No schema changes found	✔ Completed • 50 sec
 unload_all_data_types_del3	public/unload_all_data_types_del3	✔ No schema changes found	✘ Canceled
 unload_all_data_types_del10	public/unload_all_data_types_del10	✔ No schema changes found	✘ Canceled
 unload_all_data_types_del1	public/unload_all_data_types_del1	✔ No schema changes found	✔ Completed • 47 sec
 unload_all_data_types_del14	public/unload_all_data_types_del14	✔ No schema changes found	⚠ Completed • 2 min
 unload_all_data_types_del16	public/unload_all_data_types_del16	✔ No schema changes found	✘ Canceled
 unload_all_data_types_del18	public/unload_all_data_types_del18	✔ No schema changes found	✘ Canceled
 unload_all_data_types_del19	public/unload_all_data_types_del19	✔ No schema changes found	✔ Completed • 57 sec
 unload_all_data_types_del11	public/unload_all_data_types_del11	✔ No schema changes found	✘ Canceled
 unload_all_data_types_del15	public/unload_all_data_types_del15	✔ No schema changes found	✔ Completed • 56 sec
unload_all_data_types_del13	public/unload_all_data_types_del13	✔ No schema changes found	⚠ Completed • 2 min

Figure: Data sources tab

NOTE: If a flow is unshared with you, you cannot see or access the datasources for any jobs that you have already run on the flow, including any PDF profiles that you generated. You can still access the job results. This is a known issue.

Schema validation:

If schema validation has been enabled, you can review validation errors for individual datasets. For more information, see *Schema Changes Dialog*.

Datasets with parameters:

If your source is a dataset with parameters, you can review and count the individual files that were matched and imported.


For the imported dataset, click **View details**. Then, click the Files tab in the context panel.

This tab can be a good check to ensure that you have specified your dataset parameters correctly.

Parameters Tab

If your flow references parameters, you can review the state of the parameters at the time of job execution.

NOTE: This tab appears only if the job is sourced from a flow that references parameters. For more information, see *Overview of Parameterization*.

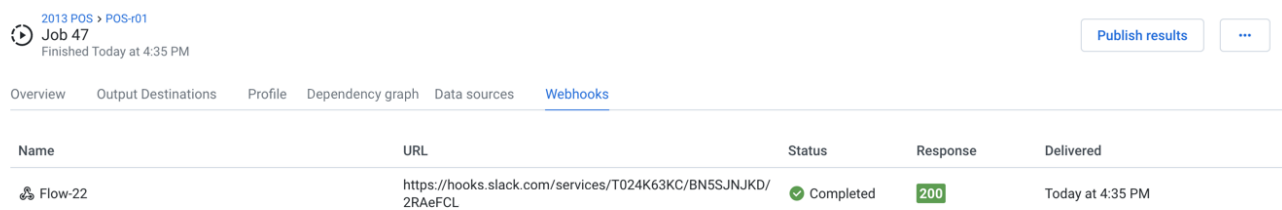


The screenshot shows the 'Parameters' tab for a job named 'Job 78' (2013 POS - params > Dataset with Parameters - 2). The job finished today at 4:22 PM. The interface includes tabs for Overview, Output Destinations, Profile, Dependency graph, Data sources, and Parameters. The Parameters tab is active, displaying a table with the following data:

Type	Source	Resolved value	Name
✱ Pattern	Dataset with Parameters - 2	{digit}{digit}	

Figure: Parameters tab

Webhooks Tab



The screenshot shows the 'Webhooks' tab for a job named 'Job 47' (2013 POS > POS-r01). The job finished today at 4:35 PM. The interface includes tabs for Overview, Output Destinations, Profile, Dependency graph, Data sources, and Webhooks. The Webhooks tab is active, displaying a table with the following data:

Name	URL	Status	Response	Delivered
Flow-22	https://hooks.slack.com/services/T024K63KC/BN5SJNJKD/2RAeFCL	Completed	200	Today at 4:35 PM

Figure: Webhooks Tab

When a webhook task has been triggered for this job, you can review the status of its delivery to the target system.

Webhooks are defined on a per-flow basis. For more information, see *Create Flow Webhook Task*.

NOTE: Webhook notifications may need to be enabled in your environment. See *Workspace Settings Page*.

Columns:

Name: Display name for the webhook task.

URL: Target URL where the webhook notification is delivered.

Status: HTTP status code returned from the delivery of the message.

- 200 - message was delivered successfully.

Delivered: Timestamp for when the webhook was delivered.

Publishing Dialog

Contents:

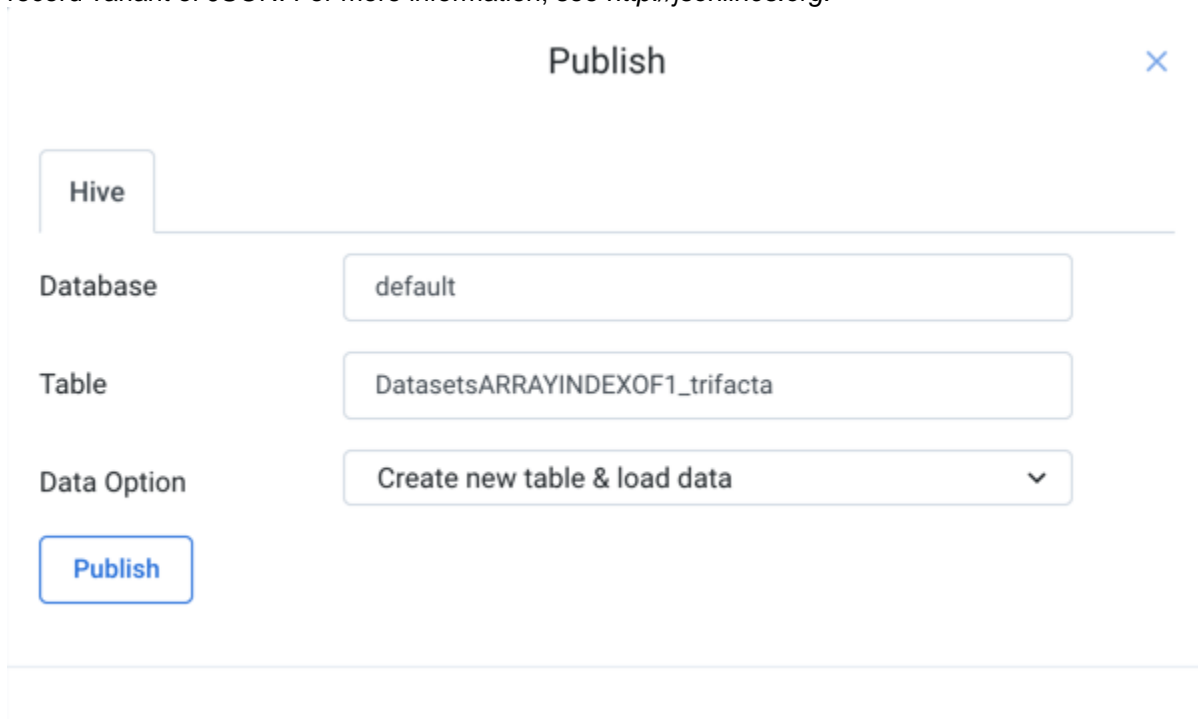
- *Limitations*
- *Publish to Tableau Server*

When a job has successfully completed, you can publish your job results to one of your connected datastores. In the Job Details page, click the Output Destinations tab. Then, click **Publish**.

Limitations

- You cannot publish ad-hoc results for a job when another publishing job is in progress for the same job through the application. Please wait until the previous job has been published before retrying to publish the failing job. This is a known issue.

JSON-formatted files that are generated by Trifacta are rendered in JSON Lines format, which is a single line per-record variant of JSON. For more information, see <http://jsonlines.org>.



The screenshot shows a 'Publish' dialog box. At the top, there's a title bar with the word 'Publish' and a close button (X). Below the title bar, there's a tab labeled 'Hive'. Under the tab, there are three input fields: 'Database' with the value 'default', 'Table' with the value 'DatasetsARRAYINDEXOF1_trifacta', and 'Data Option' with a dropdown menu showing 'Create new table & load data'. At the bottom left, there is a blue 'Publish' button.

Publish to Tableau Server

If you have created a Tableau Server connection, you can export results that have been generated to the connected server.

Supported Formats:

- **Hyper:** Results are written to your Tableau Server in Hyper format.

Options:

- **Connection:** If you have created multiple connections to Tableau Server, please select the connection to use from the list.
 - The Site name is specified as part of the connection. See *Tableau Server Connections*.
- **Project Name:** Name of the Tableau Server project.
- **Datasource Name:** Name of the Tableau Server datasource. This value is displayed for selection in Tableau Server.

Data Option:

If you are publishing to a pre-existing table, schema validation is automatically performed.

- **Create new datasource:** The platform creates the datasource and then loads it with the results from this job. If you attempt to use this option on a source that already exists, the publishing job fails, and an error is generated in the log.
- **Append data to existing datasource:** The results from this job are appended to the data that is already stored in Tableau Server. If you attempt to append to a source that does not exist, the publishing job fails, and an error is generated in the log. Append operations also fail if you publish to a target with a different schema.
- **Replace contents of existing datasource:** Target datasource is dropped. A new datasource is created using the schema of the generated output and filled with the job results.

Troubleshooting - Request timeout exception

When publishing to Tableau Server, you may encounter an error similar to the following for a PUT operation in the job log:

```
com.trifacta.clients.http.exceptions.RequestTimeoutException: PUT request to ...
```

Solution:

In this case, the size of individual chunks submitted to Tableau Server is too large. The PUT operation did not complete before a server timeout was encountered, and the operation failed.

To address this issue, you should lower the size of each chunk that is submitted to Tableau Server for publication. For more information, see *Configure Data Service*.

Schema Changes Dialog

When schema validation is enabled, the schemas of the job's datasets are checked against stored versions of the schema from the previous job run and reported in the Job Details page. Through the Schema Changes dialog, you can explore differences (findings) between the stored and read schemas for each dataset.

For more information, see *Overview of Schema Management*.

Schema changes

×

The following schema changes have been found during this job run in dataset "SCHEMA_DRIFT". To prevent changes in consequent jobs, the flow dataset must match the most up-to-date version of the source schema.

All findings (2)

Added columns (1)

Missing columns (1)

Moved columns (0)

Column	Finding
COL7	+ Added to position 0
COLNEW7n	× Missing from position 0

Close

Open flow

Figure: Schema Changes Dialog

Columns:

- **Column:** Name of the column in the dataset.
- **Finding:** Description on the change between the column in the stored schema and the column in the schema read during this job execution. Additional details on these finds is below.

Actions:

Use the tabs at the top of the screen to filter the list of findings.

Tip: In all cases, these issues should be explored.

NOTE: Renamed columns appear as a deletion of the original column and an addition of the renamed column.

Finding Type	Description	Risk
Added column	The indicated number of columns have been added to the dataset.	Risk: Moderate If these columns are added at the end of the dataset, they may not cause breakages. However, they may be omitted from some dataset reshaping transformation steps. If these columns are inserted in the middle of the dataset, they have caused some columns to be moved. See below.
Missing column	The indicated number of columns are missing in the new dataset.	Risk: High

		Depending on the use of the missing columns, these could be directly referenced in the steps of your recipe. Additionally, downstream consumers of your data may be counting on these columns.
Moved column	The indicated number of columns have been moved to new locations in the dataset.	<p>Risk: Moderate</p> <p>Named references to the columns should still function. However, if your recipe steps use column ranges to specify a selection of columns, then these steps may no longer include the moved columns.</p>

- Use the controls at the bottom to move to additional pages of findings.
- To open the flow from which the job and these findings were detected, click **Open flow**.
- To close the dialog, click **Close**.

Limits:

Maximum number of findings per type:

- Columns added: 125
- Columns missing: 125
- Columns moved: 50
- Total number of findings: 300

Plan Runs Page

In the Plan Runs page, you can track the status of all runs of your plans.

You can only see runs for the plans to which you have access in your current environment.

Plan runs are executed from the Plans page. See *Plans Page*.








Plan runs				
All runs Completed Failed Canceled Running				<input type="text" value="Search plan runs"/>
		Status	User	Started ▾
✓  p1 Run ID: 5		✓ Completed	Admin (you)	Today at 2:13 PM Ran for 2 minutes
Run parameters_flow				
⌚  customer - 2 Job ID: 107		✓ ✓ Completed		Today at 2:13 PM Ran for a minute
Run POS_flow				
⌚  POS-r01 - 2.txt - 2 Job ID: 108		✓ ✓ Completed		Today at 2:14 PM Ran for a minute
⌚  POS-r01 - 2.txt - 2 Job ID: 108		✓ ✓ Completed		Today at 2:14 PM Ran for a minute
>  new_plan Run ID: 4		✓ Completed	Admin (you)	Today at 1:39 PM Ran for 2 minutes
>  new_plan Run ID: 3		✓ Completed	Admin (you)	Today at 12:57 PM Ran for 2 minutes
>  new_plan Run ID: 2		✓ Completed	Admin (you)	Yesterday at 4:04 PM Ran for a minute

Figure: Plan Runs page

Tabs:

Each of the available tabs corresponds to a possible status for plan runs that have been initiated on the platform.

- **All runs:** All runs that you have initiated are listed here.
- **Completed:** Run has successfully executed.
- **Failed:** Run failed to complete.

NOTE: You can re-run a failed run from the Plans page.

To download the logs of a failed plan run, select **Download logs** from the run's context menu.

- **Canceled:** Run was canceled.
- **Running:** Run is in progress.

Access:

- You can review and drill into any run that you initiated.
 - For each plan run, you can click the caret to reveal details on the individual flow tasks that were part of the plan run.
 - Logs are available for individual flow tasks. Click **Download logs** from the flow task context menu.

- You can also drill into any run that was initiated from a flow that has been shared with you.

Columns:

- **Run:** Internal identifier for the plan run. This value is unique for all runs in your Trifacta® instance.
 - Click the ID number to explore details about the plan run. See *Plan View Page*.
- **Status:** The current status of the run. See above.
- **User:** The Trifacta user that initiated the run.
- **Started:** Start timestamp for the run.

Actions:

- **Filter by status:** Click one of the tabs to filter the display to show only the listings for the selected status.
- **Filter by type and date:** Click the Funnel icon to filter the list of plans by source of execution, date range, or both. See below.
- **Search:** Enter text in the search field to filter the listed runs by run ID, flow name, or recipe name.

Filter Plan Runs

To filter the list of plans, click the Funnel icon next to the search bar. You can use the following dialog to filter the display of plans based on dates or source of execution:

Filter plan runs
×

Type

Show all plan runs
▼

Date/Time

Started

Started After
▼

MM/DD/YYYY

HH:MM

AM
▼

Ended

Ended Before
▼

MM/DD/YYYY

HH:MM

AM
▼

Clear Filters

Cancel

Apply

Figure: Filter plan runs dialog

Plan type:

Show plans based on the following available options:

- **Show all plan runs**
- **Only show manual plan runs**

- **Only show scheduled plan runs**
- **Only show api plan runs**

Started:

- Specify the date and time when the plan runs to display started.
- Select **Started Between** or **Started After** from the drop-down list and populate both date-time rows.

Ended:

- Specify the date and time when the plan runs to display ended.
- Select **Ended Between** or **Ended After** from the drop-down list and populate both date-time rows.

Actions:

- To apply the specified time filter to the Plan Runs page, click **Apply**.
- To clear the time period values, click **Clear Filters**.

Sample Jobs Page

In the Sample Jobs page, you can track the status of all sample jobs to which you have access.

Sample jobs

Search samples

/

All jobsCompletedFailedCanceledRunningQueued

Job	Status	Flow	User	Started
Anomaly-based Job ID: 4	In progress	test POS-r02	Administrator (you)	Today at 4:27 <div>Cancel job</div> <div>...</div>
Filter-based Job ID: 2	Completed	test POS-r03	Administrator (you)	Today at 4:25 PM Ran for a few seconds
Random Job ID: 1	Completed	test POS-r01	Administrator (you)	Today at 4:24 PM Ran for a few seconds

Figure: Sample jobs page

Tabs:

Each tab corresponds to a possible status for Sample jobs that have been initiated in the Trifacta application.

Tip: Select **Download logs** from the sample job's context menu to download the log file for that sample job.

- **All jobs:** All sample jobs that have started.
- **Completed:** Sample jobs that have successfully completed.
- **Failed:** Sample jobs that have failed.
- **Canceled:** Sample jobs that were canceled.

Tip: To cancel an in-progress sample job, select the **Cancel job** option while the sample job is running.

- **Running:** Sample jobs that are currently in progress.

Access:

- You can review and drill into any available sample jobs.
 - For each sample job, you can click the caret to reveal details on the sample jobs.

Columns:

- **Jobs:** Sample jobs to which you have access.

Tip: Click the link for the sample type to load the sample in the Transformer page.

- **Rows:** Number of rows that have been used to generate a sample.
- **Status:** The current status of the sample jobs.
- **Flow:** Name of the flow in which the samples have been collected. You can click the flow link to go the Flow View Page.
- **User:** The Trifacta user who initiated the sample job.
- **Started:** Start timestamp for the sample job.

Actions:

- **Filter by status:** Click one of the tabs to filter the display to show only the listings for the selected status.
- **Filter by type and date:** Click the Funnel icon to filter the list of sample jobs by date range, time of execution, or both. For more information, see below.
- **Search:** Enter text in the search field to filter the listed sample jobs by job ID, sample type, status, or flow.

Filter Sample Jobs

To filter the list of sample jobs, click the Funnel icon next to the search bar. Use the dialog to filter the display of samples based on date ranges:

Filter sample jobs

Date/Time

Started

Started After

MM/DD/YYYY

HH:MM

AM

Ended

Ended Before

MM/DD/YYYY

HH:MM

AM

Clear Filters

Cancel

Apply

Figure: Sample Job dialog

Started:

- Specify the date and time when the sample jobs have started.
- Select **Started Between** or **Started After** from the drop-down list and populate both date-time rows.

Ended:

- Specify the date and time when the sample jobs have ended.
- Select **Ended Between** or **Ended Before** from the drop-down list and populate both date-time rows.

Actions:

- To clear the time period values, click **Clear Filters**.

Transformer Page

Contents:

- *Page Uses*
 - *Identify and Select Data*
 - *Get Statistics*
- *Context Panel*
 - *Build Recipes*
 - *Generate Samples*
- *Launch Jobs*
- *Transformer Bar*

In the Transformer page, you identify the data that you need to transform and build your transformation recipes on samples taken from your currently selected dataset.

When you make changes to your transformation recipe, those changes are immediately applied to your sample, so that you can preview the results of your recipe before you run it against the entire dataset. In this manner, you can quickly build and iterate on the transformations applied to your data.

- In the Library page, click the name of the dataset. See *Library Page*.
- By default, Trifacta® selects the first N of row data as the **head sample**. The number of rows depends on the number of columns, data density, and other factors. Depending on the size, this sample may be the full dataset.
- For more information, see *Overview of Sampling*.

The screenshot displays the Transformer Page interface. At the top, there's a header bar with 'USDA FARMERS MARKET 2014 FLOW' and a 'Full Data' button. Below this is a toolbar with various icons for data manipulation. The main area is divided into two panels: 'Source' and 'Preview'. The 'Source' panel shows a table with columns: #, FMID, RBC, MarketName, RBC, and MarketName. The 'Preview' panel shows a similar table with the same columns. On the right side, there's a 'Suggestions' panel with a search bar and a list of transformation suggestions. The suggestions include 'Replace', 'Split on values matching', 'Extract values matching', 'Count values matching', and 'Extract list of values'. Each suggestion has a 'See all' link and an 'Add' button.

#	FMID	RBC	MarketName	RBC	MarketName
20k - 1.01M		7,474 Categories		7,474 Categories	
-	1005969	Not Wednesday Farmers Market at Town Center	Not Wednesday Farmers Market at Town Center		
-	1008844	10:10 Farmers Market	10:10 Farmers Market		
-	1000618	100-Mile Market	100-Mile Market		
-	1002454	112st Madison Avenue	112st Madison Avenue		
-	1005586	12th & Brandywine Urban Farm Market	12th & Brandywine Urban Farm Market		
-	1008071	14&U Farmers' Market	14&U Farmers' Market		
-	1007098	17 on the Square	17 on the Square		
-	1000059	175th Street Greenmarket	175th Street Greenmarket		
-	1003877	17th Ave Market	17th Ave Market		
-	1002619	17th Street Farmers Market	17th Street Farmers Market		
-	1000709	19th Annual Highlands Business Partnership Farmers Market	19th Annual Highlands Business Partnership Farmers Market		
-	1003233	2012 Wood County Farmers' Market	2012 Wood County Farmers' Market		
-	1005309	22nd and Tasker Farmers' Market	22nd and Tasker Farmers' Market		
-	1006576	25th Avenue Farmers' Market	25th Avenue Farmers' Market		
-	1005299	29th and Wharton Farmers' Market	29th and Wharton Farmers' Market		
-	1000380	2nd Street Public Market	2nd Street Public Market		
-	1004958	3 French Hens French Country Market	3 French Hens French Country Market		
-	1005636	32nd Street/Waverly Farmers Market	32nd Street/Waverly Farmers Market		
-	1005310	33rd and Diamond Farmers' Market	33rd and Diamond Farmers' Market		
-	1003161	38th & Meridian Farmers Market	38th & Meridian Farmers Market		
-	1004414	3rd & Curry St. Farmers Market	3rd & Curry St. Farmers Market		
-	1002868	3rd Street N (hwy 11) beside Jack's Restaurant	3rd Street N (hwy 11) beside Jack's Restaurant		
-	1006234	4th Street Farmers Market	4th Street Farmers Market		
-	1006494	52 & Shadeland Avenue Farmers Market	52 & Shadeland Avenue Farmers Market		
-	1005504	62nd and University Farmers Market	62nd and University Farmers Market		

Figure: Transformer Page

Page Uses

Tip: When keyboard shortcuts are enabled, press ? in the application to see the available shortcuts. Individual users must enable them. See *User Profile Page*.

Identify and Select Data

In the main panel of the Transformer page, you can select one or more elements of sampled data, which prompts suggestions for steps that you can apply to transform them. Each of these views provides a different perspective on your data, and the results of any subsequent steps that you select or configure are previewed by default in the data grid:

NOTE: Before your job is run, profiling information such as column statistics are exact counts of the sample that is currently loaded. After the job is run, profiled results in the Job Results page might include estimates for some metrics and counts, depending on the scale of the dataset.

Tip: Click one or more column headings to be prompted for suggestions that apply to the selected column or columns.

Panel	Description	Recommended Uses
Transformer Toolbar	A toolbar of common transformations, filters, and other operations.	Use the tools in these drop-downs to quickly build common recipe steps.
Data Grid	By default, the Transformer page displays previews in a columnar grid, which is the default view. Click Grid .	Use for examining values in a column with appropriate surrounding context. How do missing values in one column compare to values in another column?
Column Details	For additional statistical information on individual columns, select Column Details from the drop-down next to the column title.	Explore values in an individual column, when their context in other rows is not necessary. Useful for managing outliers, reviewing mean, min, and max values.
Column Browser	Use the Column Browser to select the columns to display and review data across columns. Click Columns .	Navigate between columns and toggle their display in the data grid. Good for high-level perspective. Use histograms for selection of ranges of values.
Context Panel	Depending on the state or the current selection of the data grid, the right side of the page displays one of several contextual panels. These panels cover recipes, suggestions, steps, and more. See below.	Review recipe and edit, create, or delete recipe steps. Review and create samples.

Get Statistics

You can use the following methods for acquiring statistics on your dataset sample or individual columns in your sample:

- **Sample Indicator:** At the top of the data grid, you can see the name of the sample currently displayed in the grid. For smaller datasets, this sample is the entire dataset. Click this name to display statistics about the currently loaded sample. See *Sample Indicator*.
- **Status bar:** At the bottom of the page, you can review the number of data types and rows and column information for the sample currently displayed in the data grid. These metrics are updated based on the recipe steps that you apply to the sample.
 - Click the Eye icon to toggle display of individual columns. See *Visible Columns Panel*.
- **Column statistics:** You can review basic statistics on individual columns.
 - Select a column in the data grid. Column information is displayed in the context panel.

- You can also click the Columns icon at the top of the data grid to select your column to review detailed statistics. See *Column Browser Panel*.
- **Profile your data:** When you run a job on your dataset, you can optionally generate a visual profile of the resulting output. A visual profile can be useful for identifying key metrics on individual columns. See *Job Details Page*.
- **Computed statistical functions:** As needed, you can generate aggregated statistics as part of your recipe. See *Aggregate Functions*.

Context Panel

The following actions are applied through the context panel on the right side of the screen. See *Context Panel*.

Build Recipes

Use the following methods to add or modify recipe steps in the Transformer page:

Tip: To add a new recipe step, press `CTRL/COMMAND + K`. Enter a search string for your transformation step.

- **Suggestion Cards:** When you select data in the Transformer page, a set of suggested transformations is displayed in cards. Select the appropriate one to preview the results in the data grid. Then, add or edit the selected transformation. See *Selection Details Panel*.
- **Transformer Toolbar:** Select data in the data grid or column browser and then choose your transformation from the Transformer toolbar. The Transform Builder is pre-populated with your transformation. See *Transformer Toolbar*.
- **Search panel and Transform Builder:** Click the + icon in the Transformer page and use the Search panel to locate your preferred transformation. See *Search Panel*.
 - Complete the transform definition in the Transform Builder. See *Transform Builder*.
- **Recipe Panel:** After recipe steps have been created, you can review and edit them through the Recipe panel. See *Recipe Panel*.
- **Transform Preview:** Before a transform step has been added to the recipe, a preview of the transform is displayed in the data grid. See *Transform Preview*.

Generate Samples

For larger datasets, the Transformer page displays a sample of them, which you use as representative data to build your recipe. As needed, you can generate a new sample, which is useful for polishing your recipe.

The data that is displayed in the data grid is based on all of the upstream samples after which all subsequent steps in each upstream recipe are performed in the browser. If you have a large number of steps or complex steps between the recipe locations for your samples in use and your current recipe location, you may experience performance slow-downs or crashes in the data grid. For more information on sampling best practices, see <https://community.trifacta.com/s/article/Best-Practices-Managing-Samples-in-Complex-Flows>.

For more information, see *Samples Panel*.

Launch Jobs

Run jobs: To run a job that executes the transform recipe currently in the Transformer page across the entire dataset, click **Run**. See *Run Job Page*.

Transformer Bar

The Transformer page contains menus that are different from the standard Trifacta menu bar.

Figure: Transformer page toolbar

- **Flow name:** Click to review flow details. See *Flow View Page*.
- **Dataset menu:** Click the caret next to the flow name to open.
 - Review the datasets in the flow or open a different wrangled one.
 - See a mini-map of flow view for the flow.
 - See *Recipe Navigator*.
- **Samples:** Click the description of the current sample to review and create new samples from your dataset. See *Samples Panel*.
- **Search icon:** Search for transformations to add to your recipe.
- **Recipe icon:** Display the current recipe. See
- **Flow Parameters icon:** Review, create, and modify the parameters of your flow. See *Manage Parameters Dialog*.
- **Samples icon:** Click the dropper icon to review and create new samples. See *Samples Panel*.
- **Run:** Runs the currently specified recipe on the dataset. See *Run Job Page*.

Data Grid Panel

Contents:

- *Transformer Toolbar*
- *Status Bar*
- *Find Column*
- *Column Information*
 - *Selecting columns*
 - *Selecting values*
 - *Row Information*
- *Filter Data Grid*
- *Transformation Preview*
- *Target Matching Bar*

The data grid in the Transformer Page displays how your current recipe applies to the data in your currently selected sample.

- The grid is the default view in the Transformer page of Trifacta®.
- To open the data grid, click the Grid View icon in the Transformer bar at the top of the page.

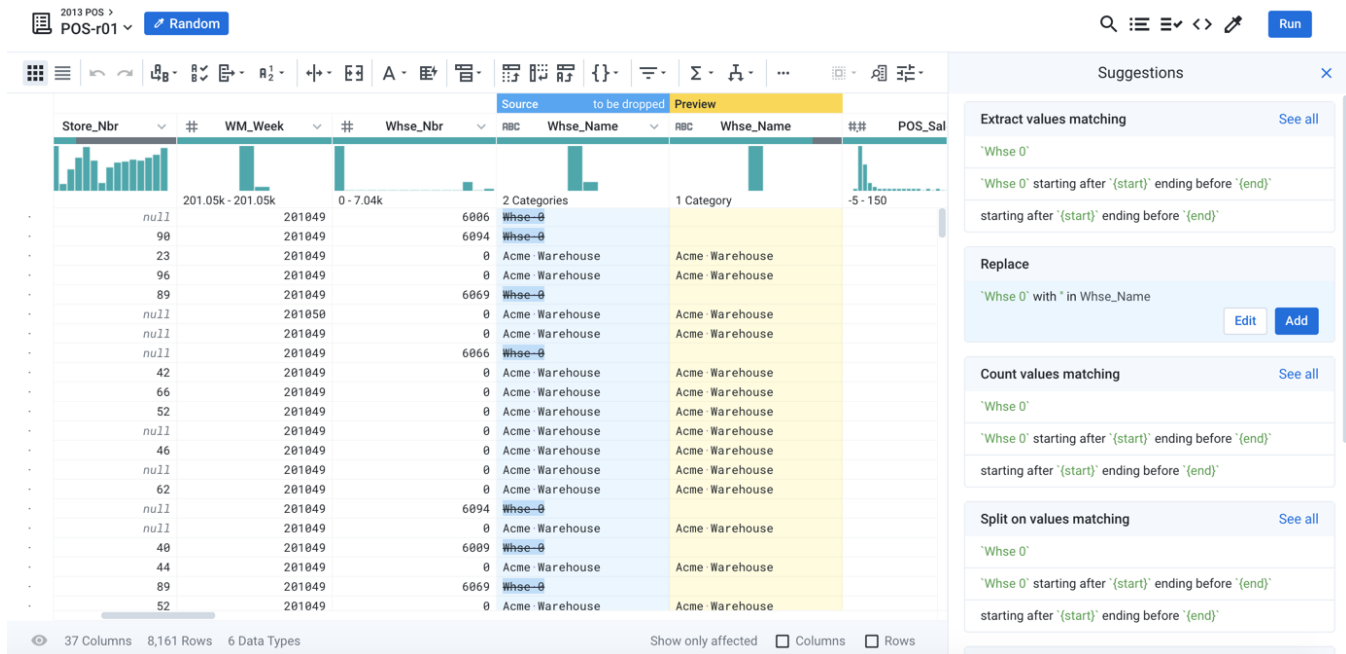


Figure: Data Grid Panel

Select:

- Click column headings to review a visual profile of the column's data and a set of suggestions for transformations to apply to the column.
 - These columns appear in the context panel on the right side of the screen.

Tip: Keep clicking columns. You can select multiple columns to prompt for another set of applicable suggestions.

- Suggestions are also generated when you select one or more values in the data histogram for a column or individual values in the displayed rows of the sample.
- See *Selection Details Panel*.
- Select specific values in a column for suggestions on those strings.

NOTE: Values in a cell cannot exceed 25,000 characters in length.

Tip: If you select a single value in the data grid, the suggestion cards suggest operations specific to that string. If you multi-select multiple values, the suggestions can apply any pattern shared between the values. For example, selecting ", CA" and ", NY" results in suggestions for how to handle state abbreviations in a column.

Scroll:

- Use the vertical scroll bar to the right of the displayed rows of data to show other rows in the sample. To review rows of the sample data that are not displayed, you may click values in a column and then scroll down through the sampled data.
- Use horizontal scrolling to review additional columns that are off-screen.

Tip: If the contents of a cell are too large for the display, you can click the Caret (>) icon to the right of the cell value in the data grid to display the entire contents of the cell.

Add or Edit:

- To add a selected suggestion card to your recipe, select the card. Then, click **Add**.
 - To modify a suggested recipe step, select its suggestion card and click **Edit**. See *Transform Builder*.
- To review details about an individual column, select **Column Details** from the column drop-down. See *Column Details Panel*.
- To review details about a selection of columns, click the Column View icon in the Transformer bar. See *Column Browser Panel*.

Ordering:

You can reorder the rows based on the values in a column. From the Column menu, select **Edit Column > Sort**. For more information, see *Column Menus*.

NOTE: Transforms that use the `group` parameter can result in non-deterministic re-ordering in the data grid. However, you should apply the `group` parameter, particularly on larger datasets, or your job may run out of memory and fail. To enforce row ordering, you can use the `sort` transform. For more information, see *Sort Transform*.

Transformer Toolbar

At the top of the data grid, you can use the toolbar to quickly build common transformations, filter the display, and other operations. See *Transformer Toolbar*.

Status Bar

Below the data grid, you can review summary information about the data in your currently selected sample.



Figure: Sample Status Bar

Click the Eye icon to open the Visible Columns panel, where you can toggle the display of individual columns. For more information, see *Visible Columns Panel*.

The status bar contains metrics about the current dataset sample for the currently selected recipe step.

- For example, if your first recipe step removes 100 rows of data, when you create your next recipe step, the status bar should indicate a row count that is 100 less than the row count at the start of the recipe. The other counts may be affected as well.
- The number of columns reflects the count that is currently displayed in the data grid. Toggling visibility of columns or applying column-based filters changes this value.

NOTE: Counts of data types may reflect that varying formats of Datetime columns are considered different types for this computation.

Tip: Before you begin transforming your data, you might want to verify the columns and count of data types against the data before it was imported. If there are discrepancies, you might want to investigate the differences. Unless your sample includes the entire dataset, row counts should differ.

NOTE: In the Trifacta Photon running environment, results can differ between executions of the same recipe due to its parallel execution and data limiting within the Transformer page. In particular, joins with multiple matches per key can sometimes cause a difference in the number of reported rows when the job is re-executed.

Show only affected:

When transformation steps are previewed, you can use these checkboxes to display only the previewed changes for affected rows, columns, or both.

Tip: These options assist in narrowing the data grid display to only the steps affected by the current recipe step.

Find Column

In a wide dataset, click the Find icon in the Transformer toolbar to locate the column of interest.

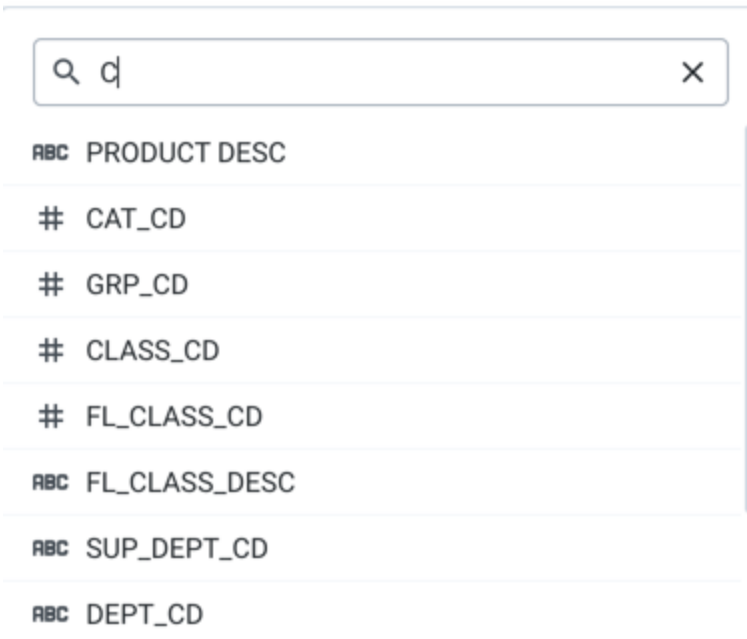


Figure: Find column search bar

- Use the up and down arrows to view the list of the columns in the dataset.
- You can start typing a column name to filter the list.

NOTE: An imported dataset requires about 15 rows to properly infer column data types and the row, if any, to use for column headers.

Column Information

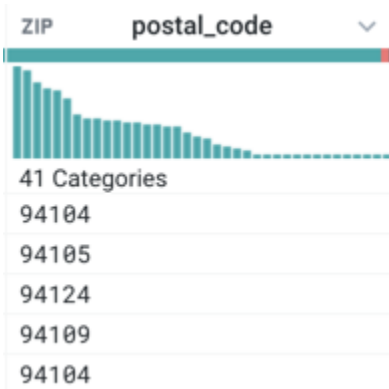


Figure: Column header, data quality bar, and histogram

- In the column header, counts reflect only the counts in the currently loaded sample. They do not reflect counts across the entire dataset, unless the entire dataset is the sample.
- There are some limitations on column names. For more information, see *Rename Columns*.

Item	Description
Data type	Identifies the selected data type, which can be inferred by the application based on the contents of the column. Click the icon to change the data type.

	<p>Tip: Before you start performing transformations on your data based on mismatched values, you should check the data type for these columns to ensure that they are correct. For more information, see <i>Supported Data Types</i>.</p> <p>See <i>Supported Data Types</i>.</p>
Column name	To change the column name, select Rename... from the column menu.
Column menu	Depending on the column data type, you can select from a set of predefined recipe steps in the column menu under the caret on the right side of the menu. See <i>Column Menus</i> .
Data quality bar	<p>The horizontal line shows valid, missing, and mismatched values in the column compared to the column's data type.</p> <p>Tip: You can click these colored bars to generate suggestion cards for transformations to act on these types of values.</p> <p>See <i>Data Quality Bars</i>.</p>
Column histogram	<p>For each column, you can see the range and frequency of values in the column.</p> <p>Tip: You can select one or more values a histogram to generate suggestion cards.</p> <p>See <i>Column Histograms</i>.</p>

Selecting columns

Through the Column Browser, you can use data quality bars and data type information to perform basic review of data across many columns. You can use these tools to select data of interest for display in the data grid or Column Details views or to prompt for suggestions of recipe steps.

Selecting values

You can click and drag to select values in a column:

- Select a single value in the column to prompt a set of suggestions.
 - Select multiple values in a single column to receive a different set of suggestions.
 - See *Selection Details Panel*.
- Double-click to select an individual word, and triple-click to select an entire cell value.
- When you select values, some values in other columns may be highlighted in a darker color, which provides some indication of correlation between values.

Row Information

On the left side of the screen, you can see a column of black dots. If you hover over one of these, you can see the current row number and, if the information is still available, the row number for the row from the original source data. These values apply only to the sample in the current dataset.

Tip: To review the original row number for a row, hover over the black dot in the data grid. These values can be referenced using the `$sourcerownumber` reference in your recipe steps. Some transformation steps, such as `pivot` and `union`, may make the original row information invalid or otherwise unavailable, which disables this option. See *Source Metadata References*.

Filter Data Grid

From the Filters drop-down, you can define filters to apply to columns, rows, or both in the data grid. See *Filter Panel*.

Transformation Preview

Before a transformation in development has been added to the recipe, a preview of the results is generated in the data grid. See *Transform Preview*.

Target Matching Bar

When a target has been assigned to your recipe, you can review the column names and data types that are expected for the target in the Target Matching bar above the column histograms.

- You can assign a dataset to be the target for the recipe you are constructing. This imported dataset, reference dataset, or recipe output contains the set of columns to which you are targeting your wrangling activities. When a target has been assigned, it is displayed in the data grid and column browser to assist you in defining your wrangling steps to match the target.
- For more information, see *Overview of Target Matching*.

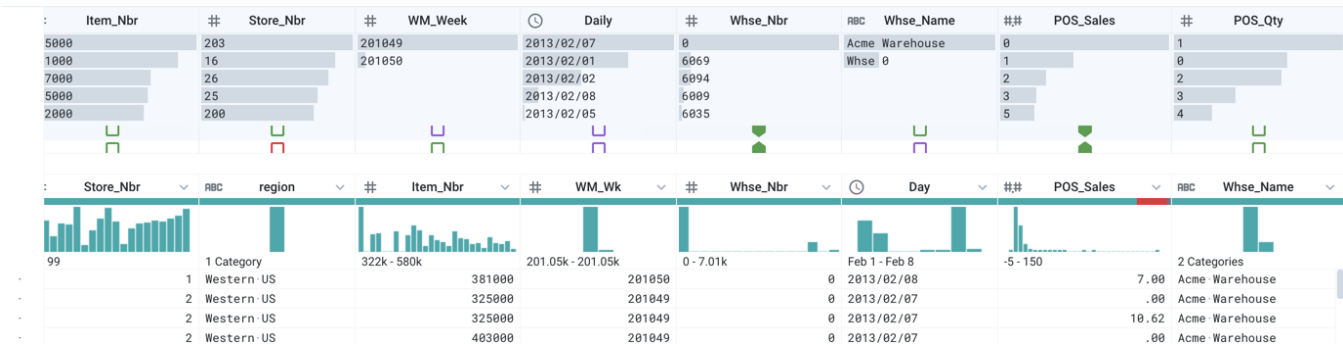


Figure: Target Matching Bar

In the Target Matching bar, you can review how the target above matches the current recipe below. For each column, matching assesses:

- Current column name vs. target column name
- Current column data type vs. target column data type
- Current column position vs. target column position
- Current column values vs. target column values

Tip: Two solid green schema tags indicate a perfect match based on the above conditions.

Actions:

- If you hover over the schema tags between a column and the target above it, you can review the detected differences between the target and the current column and select actions to fix any differences.
- Click the schema tag to auto-fix a mismatch or to select the column with which to match. These actions add a recipe step to create a match between the two columns.

For more information on the schema tags, see *Column Browser Panel*.

Column Histograms

The bar chart at the top of each column in the Transformer page, called a **histogram**, characterizes the data in that column. Each column histogram displays the count of each detected value in the column (for string data) or the count of values within a numeric range (for number data).

You can use this histogram to identify unusual values or outlier values, which should be removed or corrected.

NOTE: Counts in a column histogram reflect only the data in the sample in the data grid. Counts in the entire dataset may differ.

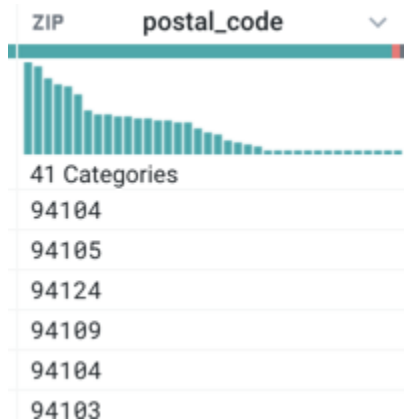


Figure: Column Histogram

Tip: When you resize the width of a column, the number of bars displayed in the column histogram changes accordingly. You can use this dynamic resizing to change the granularity displayed in histograms.

The contents of the column histogram vary depending on the data type for the column. For example:

- For numerical types (Integer or Decimal type), each bar covers a range of values, and the bars are sorted in numerical order.
 - For a numeric range bar that overlaps values in another bar, values are inclusive on the lower bound and exclusive on the upper bound. For example, if a histogram bar represents the values 0-10, it includes the count of instances of 0 and does not include the count of instances of 10. The count of instances of 10 is part of the adjacent bar in the histogram.
 - The above applies only when there are overlapping values between data ranges. If there are no overlapping values, then the range includes the values of the lower and upper boundaries.
- For non-numerical (i.e., “categorical”) types, each vertical bar covers a single value, ordered from most frequently-occurring values.

Tip: If you hover over a bar in the histogram, you can review specific values, the count of that value, and the percentage that value represents of the total count of values in the column.

When you select values:

- For the values represented by the bar(s) you selected, rows containing them are highlighted, and suggestion cards are presented for handling those values.
- Bars in other columns may partially change color. This feature, known as **brushing and linking**, illustrates the fraction of the bar values in other columns that correspond to your selected values. Brushing and linking is useful for identifying correlations in your data.

To select values:

- Use `CTRL` - click to select multiple discrete values.
- Click and drag across a range of values.

Data Quality Bars

Just below the column name in the data grid is a horizontal band, which identifies data quality issues among the sample values in the column.

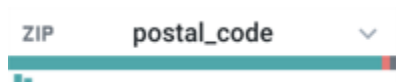


Figure: Data quality bars

Each color band identifies the relative number of records that fit the following data quality definitions:

Color	Type	Description
Green	Valid	Valid value for the currently selected data type.
Red	Mismatched	A value that does not match the listed data type. For example, if a column of Zip type contains, MISSING, it is considered a mismatched value. For more information, see <i>Find Bad Data</i> .
Gray	Missing	Value is empty or null. For more information, see <i>Manage Null Values</i> .

You can use a column's data quality bar to build a recipe step to address selected data. For example, click the red set of values in the data quality bar to generate a set of suggestion cards to address mismatched values in the column.

Tip: The histogram may also show you unwanted variation in your values. For example, if the column stores latitude data, the precision may be too fine (e.g. 37.764013 and 37.76022 versus 37.76). You can use the ROUND function to round your data to a more usable level of precision and thereby reduce the number of unique values in the column to a more manageable count.

For more information, see *Supported Data Types*.

Lookup Wizard

Contents:

- *Lookup Wizard - Step 1*
 - *Lookup Wizard - Step 2*
 - *Column Cleanup*
 - *Auto-updating Lookups*
-

Through the Transformer page, you can perform lookups from one set of values in your dataset into another set of values in another dataset using a simple wizard. To perform a lookup, select the caret next to a column title, and then select **Lookup....**

A **lookup** compares each value in the selected column against the values in a selected column of the target dataset. Where a match is found, the values in other columns of the target dataset are inserted as new columns in the dataset from which the lookup was executed.

For example, your enterprise is changing the names of all of your products. Instead of performing a complex set of replace transforms, you can perform a lookup from your productName column into a two-column dataset, which contains the original name and the new name in separate columns. When the new name is inserted into your source dataset via lookup, you can delete the source column and continue transforming your data with the new names.

- You cannot perform lookups on columns of Object or Array data type.
- A lookup essentially performs a left join between the first dataset and the second one. However, lookups are less flexible in terms of defining and editing them.








NOTE: If column values are non-unique, the resulting dataset can be significantly larger than the original dataset.

This workflow is best demonstrated by example. In this case, your raw sales data records product information in internal numeric identifiers. For analysis, you may want to integrate data from your products master data based on the internal identifier, so that you have a product description and other useful information as part of your dataset.

Lookup Wizard - Step 1

In the first step, you select the dataset against which you would like to perform your lookup for matching data for the Item_Nbr column. In this example, the products dataset is selected, since it contains the list of recognized products:

Tip: You can search your available flows and datasets. When you search for flows, all datasets in the flow are matched.

Q Search...	1 2 > >	All (29)	Imported (27)	Reference (2)	Recipe (0)
NAME	SOURCE	LAST UPDATED			
 POS-r02.txt	HDFS	Today at 10:17 AM			
 POS-r03.txt	HDFS	Today at 10:17 AM			
  REF_PROD.txt	HDFS	Today at 10:17 AM			
 REF_CAL.txt	HDFS	Today at 10:17 AM			
 USDA Farmers Market 2014	HDFS	Today at 10:15 AM			
 BOH August.csv	HDFS	Today at 9:49 AM			

Cancel

Select

Figure: Lookup Wizard - Step 1

Lookup Wizard - Step 2

After you select the dataset against which to perform the lookup, you select the field in the target dataset to use as the lookup key. The **lookup key** provides the set of identifiers for which you are trying to find a match for each value in the source column. In this case, the lookup key column has the same name as the source column: `ITEM_NBR`.

Step 2 of 2 Select Lookup Key ✕

ITEM_NBR ▼

[< Back](#)
[Cancel](#)
[Execute Lookup](#)

Figure: Lookup Wizard - Step 2

Column Cleanup

When the lookup is executed, for each value in the source `item_nbr` column that can be found in the target dataset's `ITEM_NBR` column, all of the other columns in the corresponding row of the second dataset are inserted as separate columns in the first dataset. These columns are inserted to the immediate right of the column that was used for the lookup:

#	Store_Nbr	#	Item_Nbr	ASC	PRODUCT DESC	#	CAT_CD	#	GRP_CD	#	CLASS_CD	#	FL_CLASS_CD
1 - 250		322k - 580k		70 Categories		7 - 81		10 - 75		5 - 50		5 - 75	
	1	381000	ACME LAWN GARDEN BAG CLEAR	07	75	05							
	2	325000	ACME COOKIES CHOC CHIP	81	10	25							
	2	325000	ACME COOKIES CHOC CHIP	81	10	25							
	2	403000	ACME SANDWICH BAG	07	70	05							
	2	449000	ACME SODAS SALTED	81	30	15							
	2	490000	ACME SCENTED OIL REFILL-CTRY SUN	07	65	20							
	2	560000	ACME LARGE FUDGE GRAHAMS COOKIES	81	10	25							
	2	573000	ACME SUGAR ICE WAFERS VANILLA	81	10	25							
	3	486000	ACME ZOO ANIMAL FRUIT SNACKS 6'S	81	70	30							
	3	488000	ACME WAFERS SUGER ICE	81	10	25							
	3	490000	ACME SCENTED OIL REFILL-CTRY SUN	07	65	20							
	3	498000	ACME RICE CRACKERS ONION	81	20	30							
	3	503000	ACME GARBAGE BAG BLACK	07	75	15							
	3	530000	ACME FUDGE DIP CHOC CHIP COOKIE	81	10	25							
	3	560000	ACME LARGE FUDGE GRAHAMS COOKIES	81	10	25							
	3	573000	ACME SUGAR ICE WAFERS VANILLA	81	10	25							
	4	325000	ACME COOKIES CHOC CHIP	81	10	25							
	4	325000	ACME COOKIES CHOC CHIP	81	10	25							
	4	326000	ACME DIGESTIVE RICH TEA BISCUITS	81	10	25							
	4	327000	ACME ASSORTED COOKIES DRP	81	10	25							
	4	328000	ACME KITCHEN BAG	07	75	10							

Figure: Lookup Wizard - Results

NOTE: If the second dataset contains multiple matching entries for individual lookup key values from the first dataset, rows from the first dataset are duplicated in the results.

NOTE: You may need to delete some of the columns that have been imported into your dataset.

Auto-updating Lookups

After you have added a lookup to your recipe, subsequent changes to that reference data are automatically reflected in the dataset.

Tip: If you must freeze the data in the dataset that you are using for a lookup, you should create a copy of the dataset as a snapshot. See *Dataset Details Page*.

To use the copy, delete the lookup and rebuild it using the copied version. See *Fix Dependency Issues*.

Standardize Page

Through the Standardize page, you can review similar column values and standardize them to values that you specify.

For example, master data on customers and products may use different names for the same product. For the Web team, the product may be called, "ACME Cookies Chocolate Chip," while the data from the Accounting team refers to this product as, "Cookies - Choc Chip." Through the Standardize page, you can normalize these values to a single consistent value for easier consumption downstream.

- Standardization can be applied to a single column at a time.
- For more information on how Trifacta® standardizes values, see *Overview of Standardization*.

Limitations:

- Standardizations applied through this page are stored in a connected database. These standardizations cannot be migrated between instances or workspaces.
- Standardizations cannot be published from a Dev instance to a Prod instance through Deployment Manager.

To open the Standardize page:

- From a specific column, click the column drop-down and then select **Standardize...**
- In the Search panel, enter `standardize` column. Then, select the column whose values you wish to standardize and click **Next**. See *Search Panel*.

USDA FARMERS MARKET 2014 FLOW > Full Data

Clustering options Search values...

Standardize

New value Revert to source

ACME ZOO ANIMAL FRUIT SNACKS 6'S Apply

Source value Multiple values

Row count 5

Summary

Source column ProdName

Unique new values 86

Source values updated 44 / 129 (34.11%)

Rows updated 54 / 300 (18.00%)

Cancel Add to Recipe

Figure: Standardize page

In the above image, Trifacta® groups the various references to product names into its interpretation of meaningful clusters. This clustering is based on pattern-matching between values in the column.

NOTE: The Standardize page displays column values from the currently selected sample only. If the sample does not span the entire dataset, column values that are not captured in the display are not affected by standardization changes. You may need to take additional samples to capture column values outside the current sample.

In the left side of the screen, you can review the clusters of values that have been detected in the column. In the above image, you can see that the platform has identified a number of clusters based on simple differences in capitalization.

- For each cluster, you can review the number of unique values and the total number of rows where the values appear in the column.
- At the bottom of the left pane, you can review the total number of unique values in the source column and the total number of rows in the displayed sample.

Toolbar



Figure: Standardize toolbar

- **Undo:** Undo the last action in the Standardize page

NOTE: This action does not undo recipe steps that have already been added.

- **Redo:** Redo the last undo action.
- **Auto Standardize:** The Wand tool automatically standardizes values based in a cluster to the most common value, as long as the most common value occurs in 25% of the clustered rows.

NOTE: For auto-standardization, the most common value is determined based on the cluster of values that are displayed in the current sample.

Tip: The Wand tool is recommended for beginning the standardization process. If values within a cluster have been modified, the cluster is not affected by the Wand tool. You can also apply the Wand tool on selected values in a cluster.

- **Clustering options:** By default, values are clustered based on similar spellings. To change the algorithm by which values are clustered, click **Clustering options**.
 - **None:** Do not cluster values. Individual values must be matched.
 - **Similar strings:** Cluster values based on similarities between the text of each value.
 - **Pronunciation:** Cluster values based on phonetic pronunciation of the values.
 - For more information on these options and their variations, see *Overview of Standardization*.
- **Search values:** To locate specific values in the column, enter a search string in the Search textbox.
- To reverse the sort order within your clusters, click Row Count.
 - To sort cluster values alphabetically, click the Source Value header.

Steps to Standardize

To standardize values:

1. If needed, change the clustering algorithm to apply to the values.
2. From the left panel, select the set of values that you wish to standardize.

Tip: Unclustered values are listed at the bottom of the panel. You should review these values when you are selecting clustered values.

- a. To select multiple values, press **COMMAND / CTRL + click** or select multiple values in the left column.

Tip: To select all values in a cluster, click the cluster header.

- b. To select a range of sequentially listed values, use **SHIFT + click**, which works across clusters of values.
3. After you have selected all values to standardize, you specify the new value to apply to these selected values in the right panel. This new value applies to all instances of the selected value or values. Changes are previewed next to the source values.

Tip: When you have values selected in a cluster, you can use one of the source values as your standardized value. Hover over the value in the left panel, and then click the icon that appears.

- a. Below the value you enter, you can review the number of rows in the sample that are affected by this change.
- b. At the bottom of the right panel, you can review the total effects of standardization on the dataset after this change is applied.
- c. If the new value is empty, then the values are kept as-is. No change is applied.
- d. To apply the standardized value to the affected clustered values, click **Apply**.
- e. At any time, you can revert the changes to the cluster values. Click **Revert to source**.
4. Repeat the previous steps as needed.

Tip: You can perform multiple replacements in a single recipe step. So, you can configure all of your standardization steps before adding the single step to your recipe. For debugging purposes, you may want to separate some or all standardization into separate steps.

5. To add the standardizations, click **Add to Recipe**.

NOTE: You cannot copy and paste standardization steps in the Recipe panel.

Recipe Navigator

Through the Recipe Navigator, you can locate and open any recipe from the current flow in the Transformer page. To open the Recipe Navigator, select the drop-down next to the name of the current recipe in the Transformer page.


This navigator can be helpful in fixing dependency issues between your current recipe and the recipes and datasets that it integrates from the flow. See *Fix Dependency Issues*.


NOTE: The listed recipes are constrained only to those that appear in the same flow as the currently loaded recipe.


NOTE: You can only open recipes in the Transformer page. To open an imported dataset, you must first create a recipe for it and then edit the recipe. See *Flow View Page*.


Navigate to...

Q Search...

✓  POS-r01

 REF_PROD

 POS-r03 – 2.txt

 REF_CAL – 2.txt


 POS-r02 – 2.txt

Figure: Recipe tab

To locate a recipe to load, use the Search box or browse the list. Then, select the recipe to load.

Sample Indicator

At the top of the Transformer page, you can see the Sample Indicator showing the name of the currently displayed sample. If you have not taken a sample, this value is **Initial data**. Click this button to show more information about the sample.

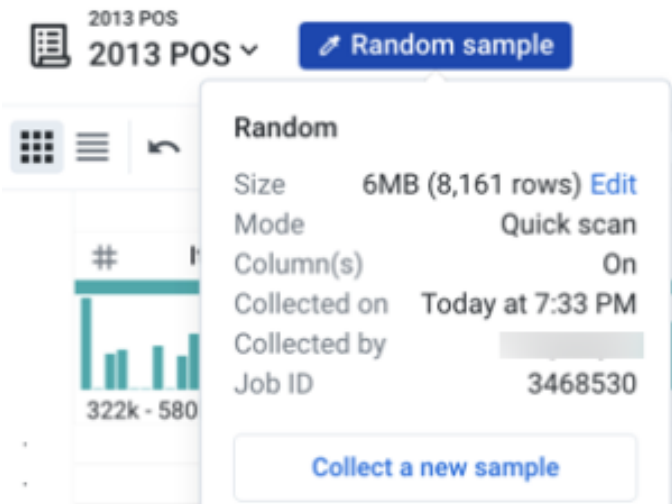


Figure: Sample Indicator

In the Sample Indicator, you can review the number of rows collected in the same and information about the user and time when the sample was collected.

- As needed, you can change the size of samples loaded into the browser for your current recipe, which may address performance issues. Click **Edit**. For more information, see *Change Recipe Sample Size*.
- To generate a new sample, click **Collect a new sample**. The Samples panel is displayed. For more information, see *Samples Panel*.

Transformer Toolbar

Contents:

- *Grid/Columns*
- *Undo/Redo*
- *Replace*
- *Extract*
- *Count Matches*
- *Split Column*
- *Merge Columns*
- *Format*
- *Create column by examples*
- *Group by*
- *Pivot*
- *Unpivot*
- *Convert values to columns*
- *Objects and Arrays*
- *Filter Rows*
- *Functions*
- *Conditions*
- *Join*
- *Union*
- *Comment*
- *Target*
- *Select*
- *Macros*
- *Find Column*
- *Filters*

At the top of the data grid and the column browser, the Transformer toolbar provides quick access to common transformations.

When a tool is selected from one of the drop-downs in the toolbar, a new step is inserted into the current location in your recipe with some of the transformation's parameters pre-specified for you. You can then finish specifying the parameters of the transformation in the Transform Builder.



Figure: Transformer Toolbar

Grid/Columns

Use these buttons to toggle between the data grid and column browser views of the Transformer page.

- See *Data Grid Panel*.
- See *Column Browser Panel*.

Undo/Redo

Undo or redo previous actions.

Replace

Replace cell values based on data type validation, string literals, or patterns. See *Replace Groups of Values*.

Extract

Extract string or numeric values using data validation, literal matches, or patterns. See *Extract Values*.

Count Matches

Count the number of matching literals or patterns in a column.

Split Column

Split a single column into multiple columns based on a common delimiter or a pattern-based expression.

Merge Columns

Merge two or more columns together.

Format

Format String and Datetime values using a variety of functions.

Create column by examples

Define transformations by mapping output values from a set of input values. After you specify a few values, Trifacta may be able to interpret the other ones for you. See *Create Column by Example*.

Group by

Compute aggregation functions for values grouped by columns. Output is new table or one or more columns in the current dataset. For more information, see *Create Aggregations*.

Pivot

Generate a pivot table of your data based on specified row and column labels. See *Pivot Data*.

Unpivot

Collapse columns into row data.

Convert values to columns

Reshape your data by converting unique values into columns or columns into rows.

Objects and Arrays

Use these tools to manipulate your dataset based on columns of Object or Array type.

Filter Rows

Remove or keep rows of data based on data type validation, conditionals, source row numbers, or other factors.

- See *Remove Data*.
- See *Deduplicate Data*.

Functions

Generate a new column of data based on a specified function.

For a list of available functions, see *Language Index*.

Conditions

Use if/then/else or case logic to apply conditionals to your dataset.

See *Apply Conditional Transformations*.

Join

Join the current dataset with another dataset using a pair of matching keys.

See *Join Window*.

Union

Add the rows of data from one or more datasets to your current dataset.

See *Union Page*.

Comment

Insert a comment in your recipe at the current step. Comments do not perform any modification to the data.

See *Add Comments to Your Recipe*.

Target

Attach a target schema to the current recipe or remove it. If a target is attached, you can attempt to auto-match all source and target columns.

- For more information on these options, see *Column Browser Panel*.
- For more information on this feature, see *Overview of RapidTarget*.

Select

NOTE: This menu is only available in the Column Browser.

Select all or no columns, or invert the currently selected columns. See *Column Browser Panel*.

Macros

Click the Macros icon to open the menu:

- **Insert macro:** Search for or select the name of the macro to apply. When a macro is selected, it is opened in the Transformer Builder, where you may specify any parameter values to apply to the macro. See *Apply a Macro*.

- **Manage in Library:** Review available macros. For more information, see *Macros Page*.
- **Create macro:** Enter a name and description for the macro to create. See *Create or Replace Macro*.

For more information, see *Overview of Macros*.

Find Column

Locate a column in your dataset by typing in the textbox. For more information, see *Data Grid Panel*.

Filters

Filter the displayed rows and columns based on selection or condition.

NOTE: Filtered data is hidden from display. It is not removed from the sample or the dataset during execution.

See *Filter Data*.

Column Menus

Contents:

- *Data Type Menu*
 - *Action Menu*
 - *Show in Grid*
 - *Automatically update*
 - *Rename*
 - *Change type*
 - *Move*
 - *Hide*
 - *Sort*
 - *Edit with formula*
 - *Format*
 - *Calculate*
 - *Create column from examples*
 - *Group by*
 - *Pivot*
 - *Restructure*
 - *Filter rows*
 - *Replace*
 - *Standardize*
 - *Extract*
 - *Split column*
 - *Column Details*
 - *Show related Steps in Recipe*
 - *Lookup*
 - *Delete*
 - *Delete others*
 - *Copy, Cut, and Paste*
-

You can use the menus available in a column context menu to perform a variety of actions on the column or columns.

Column menus are available in the following pages:

- *Data Grid Panel*
- *Column Browser Panel*
- *Column Details Panel*

Depending on the data type of the column, the available options may vary.

NOTE: In the Column Browser panel, you can select multiple columns at the same time. The displayed column menu items are only the ones that apply to columns of all data types.

Data Type Menu

On the left side of the column header, in the Data Grid you can click the data type icon to review and select a different data type for the column.

- When you select the Datetime data type, you can choose the format against which values in the column are validated. See *Choose Datetime Format Dialog*.
- For more information, see *Supported Data Types*.

Action Menu

On the right side of the column header, select the drop-down caret to choose one of the following actions.

- Some options are available only for specific data types.
- In most cases, these actions result in a new step being inserted at the current location in your recipe. Before you apply one, you should verify that you are at the proper position in the recipe to apply this step.

Show in Grid

When a single column is selected in the Column Browser panel, you can use this option to select the data grid and change its focus to display the selected column.

Tip: Use the Column Browser panel to quickly locate and select columns for display or other actions. See *Column Browser Panel*.

Automatically update

Unlock the column's data type. For more information, see *Change Column Data Type*.

Rename

Rename the column.

For more information on column name requirements, see *Rename Columns*.

Change type

Change the data type of the column. See *Supported Data Types*.

Move

Move the column to the beginning or end or to a specified location in the dataset.

Hide

Hide this column from display in the application. To redisplay this column, you must select it in the Visible Columns panel. See *Visible Columns Panel*.

NOTE: Hide/Show commands do not create transformation steps. Hidden columns still exist in the output.

NOTE: When a column is hidden from a dataset, it is hidden for all users. You should check the column browser for hidden columns in shared datasets.

Sort

Sort: Sort the values in the column by ascending or descending value. This command does not apply to some data types.

NOTE: Sort is intended primarily for display purposes in the Transformer page.

When a column is sorted in ascending order, the bottom values are mismatched and null values, in that order. These values appear at the top of the column when sorted in descending order.

Edit with formula

Modify the column values based on a formula that you input. For more information, see *Transform Builder*.

Format

Format the values in the column according to the selection.

NOTE: This menu is available only for columns of String and Datetime type.

Calculate

Calculate a new column containing the values computed from the source column based on the selected function. This command does not apply to all data types.

Create column from examples

Using a source column, you can create a new column using the source values as examples.

Tip: The transformation-by-example feature can be used in place of standardization or extraction transformations and may be simpler to use.

NOTE: This feature may need to be enabled in your environment.

See *Create Column by Example*.

For more information on transformation-by-example, see *Overview of TBE*.

Group by

Group by minimum or maximum value, counts, or arbitrary function. See *Create Aggregations*.

Pivot

Generate summary table based on computations aggregated across groups.

Restructure

Restructure your dataset by converting column values to columns, nesting values into objects or arrays, and unpivoting your data.

Filter rows

Filter the rows in your dataset based on column values, duplicate values in the column, or other condition. See *Filter Data*.

Replace

Replace values in the selected column with based on text values, patterns, or delimiters. See *Replace Groups of Values*.

Standardize

Standardize individual and clustered row values for consistency. See *Standardize Page*.

Extract

Extract values from the selected column based on text values, specific types of data, or patterns in the data. Optionally, extracted data can be removed from the source. See *Extract Values*.

Split column

Split the column into separate columns based on patterns, delimiters, positions in the value, or regular intervals throughout the value.

Column Details

Explore the interactive profile of the column details. See *Column Details Panel*.

Show related Steps in Recipe

Highlight steps in the recipe panel where the selected column is referenced.

- If another column is dependent on the selected column, all steps pertaining to that column are highlighted as well.
- To dismiss the highlighting, click **Clear** at the top of the recipe panel.
- For more information on disabling this feature, see *Miscellaneous Configuration*.

Lookup

Perform a lookup of the column values against a set of values in another column of another dataset. See *Lookup Wizard*.

Delete

Remove the column from the dataset.

Tip: You can also select **Hide** to remove the column from display in the data grid. The column does remain as part of the dataset and is included in any generated results.

Delete others

Delete the columns other than the selected column(s) from the dataset.

Copy, Cut, and Paste

You can copy and paste columns and column values into other areas of your dataset. For more information, see *Copy and Paste Columns*.

Choose Datetime Format Dialog

In the Datetime Format dialog, you can specify or locate the format that you'd like to use for validation of the values in the current column against the Datetime data type.

For example, if you choose `mmddyyyy` as your format, all values in the column are considered valid for the Datetime data type if they fit the `mmddyyyy` pattern. Otherwise, they are considered invalid values for the Datetime column.

Choose Date/Time Format

Search for formats or type an example

✓

yyyy*mm*dd

2019/11/7

mmddyyyy

1172019

mm*dd*yy

11/7/19

mm*dd*yyyy

11/7/2019

month*dd*yy

November 7 19

month*dd*yyyy

November 7 2019

Learn about date/time formats

Cancel

Select Format

Figure: Choose Datetime Format dialog

NOTE: The list of available Datetime formats is factored based on your current locale settings. Some formats from other locales may be available for selection. For more information, see *Locale Settings*.

To locate a format, you can:

- Search:** You can start entering your preferred format as tokens. For example, if you enter `yyyy`, you can narrow the list to the formats that support four-digit values for year.
- Browse:** You can scroll the list to see the available formats.
- Enter:** Type or paste in an example Datetime value that is in your preferred format.

For more information on the supported tokens for the Datetime data format, see *Datetime Data Type*.

To apply your selected format, click **Select Format**. The values in the column are validated against the selected format for the Datetime data type.

Transformation by Example Page

Contents:

- *Transform Builder*
 - *Grid View*
 - *Keyboard shortcuts*
 - *Pattern View*
 - *Toolbar*
-

In the Transform by Example page, you can build new columns of data by specifying values to map from the selected source column. For the column to transform, select **Create column by examples** from the column menu.

NOTE: Transformations by example are developed using the current sample. When the finished transformation is applied across the entire dataset, some source values may not be matched by the patterns you specified using the current sample.

Tip: Transformation by example works best for text-based inputs. For non-text inputs, you can convert the column data type to String.

To transform by example, you can use either of the following:

- **Grid View:** Displays Source and Preview column values as they appear in the sample in the data grid.
- **Pattern View:** Displays the Source and Preview values in groups of similar values based on pattern matching by Trifacta®.

For more information on these types of transformations, see *Overview of TBE*.

Transform Builder

In the Transform Builder panel on the right, you can review and change the Source and Preview columns to transform.

Tip: To transform from a different source column, select a new column from the Example column drop-down. This step clears all examples from the transformation you are building. Some columns may not be available for selection. To use such a column, change its type to String first.

Grid View

In grid view, the Source values for each row in the sample are listed next to an empty Preview column. You can create mappings by selecting a cell in the Preview column and manually entering a value.

DATASETS - TBE FLOW >
Datasets - TBE Full Data

🔍 ☰ ☑ <> ✎ Run Job

☰

✎

☑

☒

🗑

Source	Preview
Phone Number	Phone_Number_clean
1 510-221-2244	510 221 2244
2 510-221-2245	510 221 2245
3 510-221-2246	510 221 2246
4 510-221-2247	510 221 2247
5 510-221-2248	510 221 2248
6 510-221-2249	510 221 2249
7 510-221-2250	510 221 2250
8 510-221-2251	510 221 2251
9 510-221-2252	510 221 2252
10 510-221-2253	510 221 2253
11 510-221-2254	510 221 2254
12 (510) 434 4404	510 434 4404
13 (510) 434 4405	510 434 4405
14 (510) 434 4406	510 434 4406
15 (510) 434 4407	510 434 4407
16 (510) 434 4408	510 434 4408
17 (510) 434 4409	510 434 4409
18 (510) 434 4410	510 434 4410
19 (510) 434 4411	510 434 4411
20 (510) 434 4412	510 434 4412
21 (510) 434 4413	510 434 4413
22 (510) 434 4414	510 434 4414
...	...

195 rows 60 unique source values

Create column from examples

Create a new column by providing example values in the grid.

Column

Phone Number

required

New column name

Phone_Number_clean

Cancel

Add to Recipe

Figure: Transformation by Example - Grid View

After you enter a value, Trifacta attempts to match other values from the Source using the same pattern to generate additional values in the Preview column.

- Values that you manually enter are listed in dark text.
- Values that are inferred by the product are in lighter text.
- Null values indicate that no pattern has been identified to match the value.

Tip: Values that have been inferred can be replaced by manual entries for further refinement.

For more information on how to use, see *Create Column by Example*.

Keyboard shortcuts

- Use the arrow keys to navigate up and down the rows in the Preview column.
- CTRL + up arrow or CTRL + down arrow to jump to the first or last row of the sample
 - In Pattern View, the above shortcuts navigate between groups of values.
- ESC cancels your current edit.
- RETURN submits your current entry as a new example.

Tip: You can copy and paste values from the clipboard into the Preview column.

Pattern View

In Pattern view, Trifacta performs some preliminary pattern detection to group Source values together. Transformations are processed using Patterns .

DATASETS - TBE FLOW >
Datasets - TBE
Full Data

Source	Preview
Phone Number	Phone_Number_clean
Pattern 1	1 of 1 unique value
5104289900	510 428 9900
Pattern 2	5 of 25 unique values
(510) 434 4404	510 434 4404
(510) 434 4405	510 434 4405
(510) 434 4406	510 434 4406
(510) 434 4407	510 434 4407
(510) 434 4408	510 434 4408
Pattern 3	5 of 23 unique values
1-510-122-3300	510 122 3300
1-510-122-3301	510 122 3301
1-510-122-3302	510 122 3302
1-510-122-3303	510 122 3303
1-510-122-3304	510 122 3304
Pattern 4	5 of 11 unique values
510-221-2244	510 221 2244
510-221-2245	510 221 2245
510-221-2246	510 221 2246
510-221-2247	510 221 2247

Create column from examples

Create a new column by providing example values in the grid.

Column required

Phone Number

New column name

Phone_Number_clean

195 rows 60 unique source values 4 patterns

Figure: Transformation by Example - Pattern View

- This view displays a maximum of five example values per pattern group.
- **Other group:** Values that do not map to any identifiable pattern are placed in a value group labeled, `Other`.
- For more information on patterns, see *Text Matching*.

Toolbar



Figure: Transformation by Example toolbar

- **Grid View:** See previous.
- **Pattern View:** See previous.
- **Undo:** Undo the last change you made to the Preview values.
- **Redo:** Redo the most recently undo change you made to the Preview values.
- **Trash:** Remove all example values from the Preview column and start over.

Column Browser Panel

Contents:

- *Locate Columns*
- *Select Columns*
- *Transformer Toolbar*
- *Column Actions*
- *RapidTarget Matching*

Through the Column Browser, you can use data quality bars and data type information to perform basic review of data across many columns. You can use these tools to select data of interest for display in the data grid or Column Details views or to prompt for suggestions of recipe steps.

- You can also use the Column Browser to toggle the display of individual columns.
- To open the Column Browser, click the Column View icon in the Transformer bar for the Transformer page.

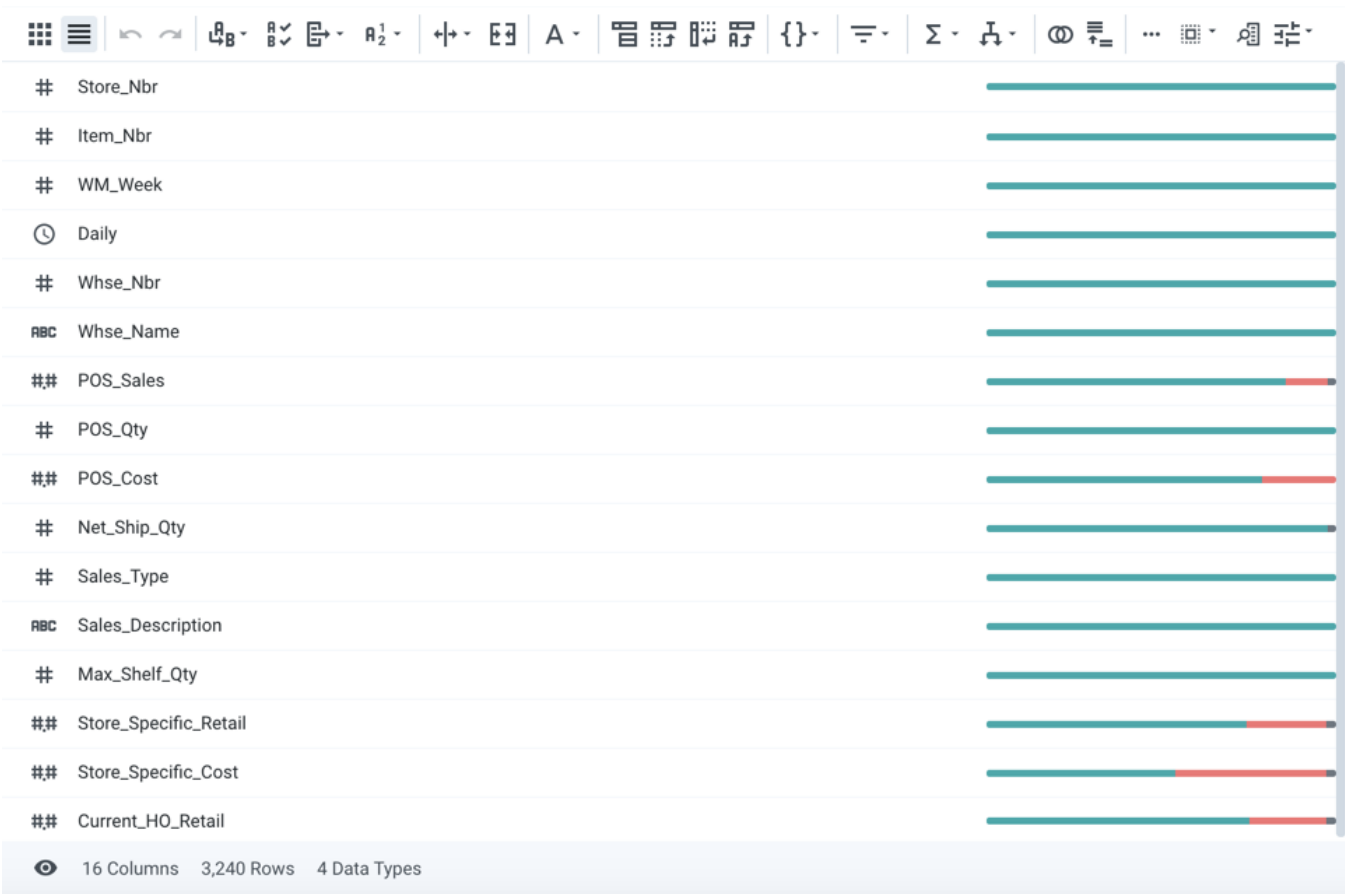


Figure: Column Browser

You can select one or more columns in the browser and then perform actions on them.

Locate Columns

You can apply one or more filters to limit the set of columns displayed in the browser. Click the Filter icon in the Transformer toolbar.

NOTE: Filters are additive and persist between the column browser and the data grid.

For more information, see *Filter Panel*.

Select Columns

You can manually select one or more columns or apply one of the predefined selections.

- To select a range of columns, click a column, press **SHIFT** and then click the ending column.
- To select multiple discrete columns, press **CTRL/COMMAND** and click additional columns.
- To toggle selection of a column, click it again.
- You can copy and paste columns and column values. For more information, see *Copy and Paste Columns*.

Transformer Toolbar

At the top of the column browser, you can use the toolbar to quickly build common transformations, filter the display, and other operations. See *Transformer Toolbar*.

Column Actions

For any individual column:

- Click the Eye icon to hide/show of the column in the Transformer page. See *Visible Columns Panel*.

NOTE: Hidden columns are only removed from view in the Transformer page. They still appear in any generated output.

- Hover over the color bars in the data quality bar to review counts. See *Data Quality Bars*.
- Right-click a column to display a list of actions in the context menu. Column actions apply only to the selected column and depend on its data type.
 - For multiple selected columns, you can choose an action from the Action menu an option that apply to all of the selected columns.
 - See *Column Menus*.

RapidTarget Matching

You can associate a target schema with a recipe. A **target schema** is information about the column names, data types, and order of the target dataset for which you are trying to build your recipe. For more information, see *Overview of RapidTarget*.

Actions taken based on the target schema are rendered as a new step in the location in your recipe.

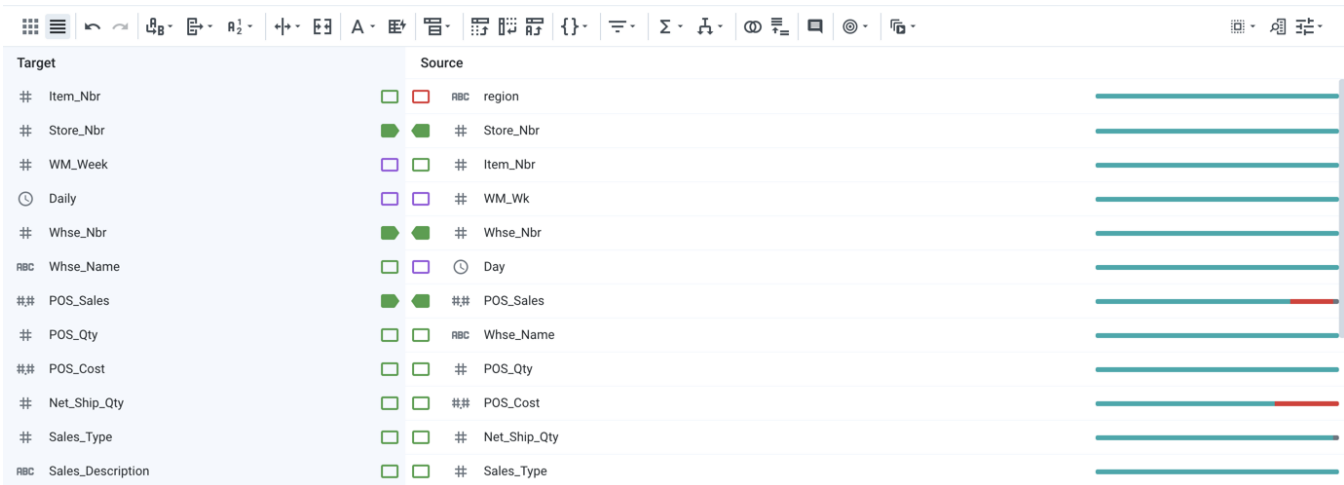


Figure: RapidTarget Panel in Column Browser

In the RapidTarget panel, the columns of the target schema assigned to your recipe are displayed on the left side. Columns are listed side-by-side in order:

Side	Description
Left	Target column
Right	Source column (current recipe)

Schema Tags:

Between the two sets of fields are a set of tags that describe the matching between the two columns in the order listed in each dataset. Each tag describes the match found for the corresponding target or source column.

- Matches by position are arrows.
- Matches by data type are green in color.
- Partial name matches are indicated by the color purple. In the cases, the platform has found a match when column names are inexact matches.
- Red color indicates no match at all.

Tip: Mouse over tags that are not solid green to learn the nature of the column mismatch. You can then click the tag to match the two columns.

Match by Name	Match by Data Type	Match by Position	Icon	Description
No	-	-	Square - red outline	Target column has no identifiable name match in source. Column appears to be missing from source. Tip: Click the source or target rectangle to manually assign a matching column.
Yes	No	No	Square - blue outline	Source column name has a matching target column name, but it is in the incorrect location.
Partial	No	No	Square - purple outline	Source column name has a partial match with a target column name, but it is in a different location.
Yes	Yes	No	Square - green	Source column name and data type have a match in the target, but it is

			outline	in a different position.
Yes	No	Yes	Arrow - blue outline	Source column and target column match in name and position, but have different types.
Yes	Yes	Yes	Arrow - green solid	Source and target columns match in name, position, and type.

Target menu actions:

From the Transformer toolbar, you can select the following options from the Target menu:

- **Attach a new Target:** Assign a target schema to the dataset.
- **Remove Target:** Remove the target from assignment to the source. Target schema dataset is not deleted.
- **Hide/Show Target data:** Toggle display of example rows from the target schema dataset.
- **Align on column name match:** Columns are automatically matched when column names match.
- **Align on fuzzy match:** Match columns based on the data contained in them.

NOTE: This method of matching is available if global fuzzy matching has been enabled. For more information, see *Overview of RapidTarget*.

When you have selected one or more source columns, you can additionally perform the following actions:

- **Align selected on column name:** Fix the selected columns to match target columns using the column names.
- **Align selected on fuzzy match:** Fix the selected columns to match target columns using the data contained in them.

NOTE: This method of matching is available if global fuzzy matching has been enabled. For more information, see *Overview of RapidTarget*.

Column Details Panel

Contents:

- Overview tab
- Patterns tab
 - Pattern reuse

In the Column Details panel, you can review additional details about a column of your dataset. Select **Column Details** from any column menu or the Action menu in the column browser.

Tip: Use the Column Details panel to explore values in an individual column, when the context of the value is not important for your current exploration. For example, you can identify outlier values for the column or compare the number of unique values to number of rows to determine whether the column could be a key value.

Overview tab

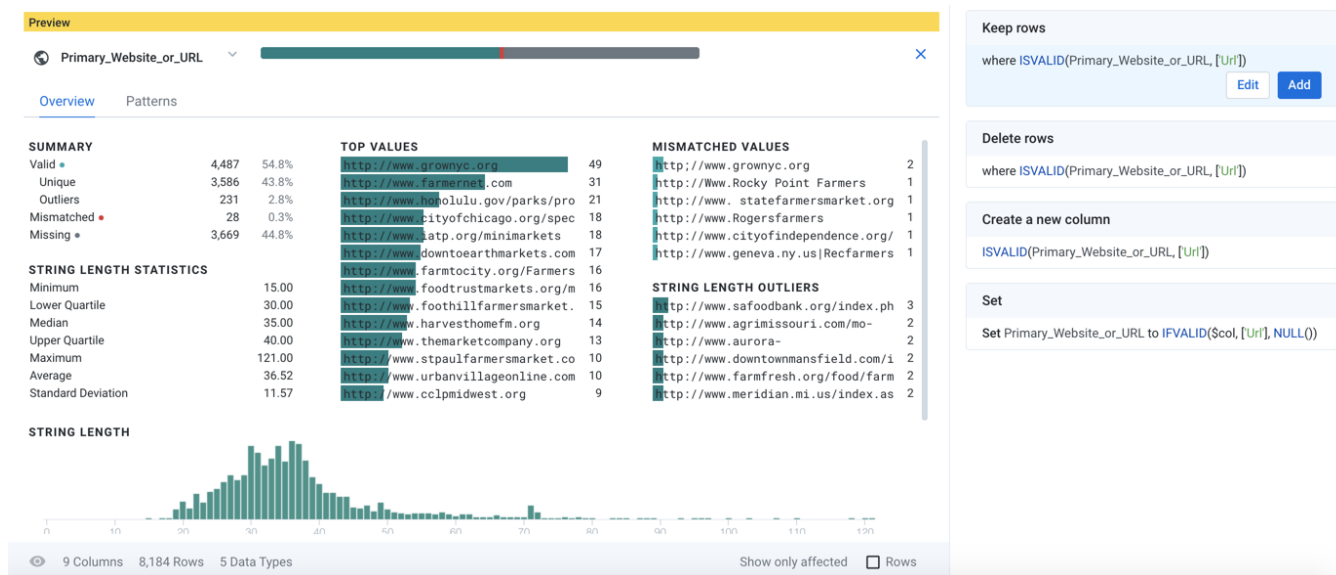


Figure: Column Details panel - Overview tab

Column statistics:

You can use this view to review basic counts and percentages of the values in the currently selected column. In addition to basic computations on valid, mismatched, and missing values, you can see breakdowns for the most frequent values and outlier values.

NOTE: Before your job is run, profiling information such as column statistics are exact counts of the sample that is currently loaded. After the job is run, profiled results in the Job Results page might include estimates for some metrics and counts, depending on the scale of the dataset.

Depending on the data type of the column, additional statistics provide information on data quality and variation. For more information, see *Column Statistics Reference*.

Actions:

- To change the data type, click the type indicator next to the column title in the Column Details panel.
- To perform commands on the column, select from the drop-down next to the column title. For more information, see *Column Menus*.
- Use the data quality bar to select categories of values: valid, mismatched, or missing. The context panel is updated based on your selection with recommended recipe steps. See *Selection Details Panel*.

Patterns tab

In the Patterns tab, you can review patterns identified by the platform in the selected column's data and then create steps based on patterns that you select. Pattern profiling automatically finds and groups clusters of the column's values based on similarities in format and structure, such as differently formatted phone numbers, addresses, log entries, and name fields. For example, if some of your dataset's address values include apartment numbers, you can create a Split transformation based on a pattern that includes the apartment numbers.

NOTE: In this tab, the count of values and the `all patterns` category do not include missing values.

NOTE: Wide columns, such as Arrays, Objects, or freeform text, might take a while to profile.

- Each non-blank value in the column is represented by one of the displayed patterns. Patterns are specified as a combination of literal values and `Patterns`. For more information on these patterns, see *Text Matching*.
- Patterns might be more generalized than the constraints of the column's data type.
- Token values are `Patterns` without braces.

Source

Primary_Website_or_URL

Overview Patterns

All patterns

url

31

27

1

1

1

Extract values matching

{url} starting after {start} ending before {end}

Replace

first occurrence of {url} starting after {start} ending before {end} with ' in Primary_Website_or_URL

Split on values matching

{url} starting after {start} ending before {end}

Figure: Column Details panel - Patterns tab

All non-blank values are captured in the `all patterns` category, which you can expand to display the patterns that capture subsets of all values. Patterns are displayed in a tree structure, with each lower level describing a subset of the parent pattern.

Tip: Hover over a pattern or sub-pattern to see the affected values in the example data beneath it.

Tip: When you select a pattern group, you may be presented with suggestions for standardizing the values in the column to a single format. In some cases, you might want to remove unnecessary data first. For example, standardization of phone numbers is easier if any +1 country codes are removed from the beginning of values.

Tip: Pattern suggestions are created based on the first few thousand rows of data in your sample. For best results, you should generate a random sample with a representative set of patterns in the first rows in the column.

Below the top level, patterns are displayed in order of decreasing frequency in the column, allowing you to choose the level of granularity for which you wish to address data issues in the column. For each pattern, you can review the counts of values matching the pattern.

In the above example, all values that have been identified as matching the `url` Pattern are contained in the first category.

- Select a pattern to trigger a set of suggestion cards to apply to the represented data.
 - When you select values from a pattern's histogram, all suggestions match the pattern. You cannot select the values that do not match the pattern from the histogram.
 - For more information, see *Explore Suggestions*.
- Select a token within a pattern or a highlighted block of text among the example values to trigger suggestion cards that apply the token within the pattern.
- You can modify the selected suggestion in the Transform Builder. See *Transform Builder*.
 - When you apply the transformation to your recipe, the Patterns tab is updated automatically.

Tip: When you see a pattern that you wish to reuse, select the pattern and one of its suggestion cards and then modify the step.

- Expand the caret next to any pattern to explore its sub-patterns, which identify subsets of values within the broader pattern.

NOTE: The `Other` pattern is a special category that contains values and counts not recognized by the currently selected pattern or sub-pattern. For example, when you select `url` pattern, the `Other` pattern captures the non-URL values. When you explore a sub-pattern of URLs, the `Other` category captures the values not recognized within the sub-pattern.

For more information on pattern standardization, see *Standardize Using Patterns*.

For more information on standardizing numeric values, see *Normalize Numeric Values*.

Pattern reuse

After patterns have been selected, they can be reused through the Transform Builder.

Transform Preview

When you create or edit a transform, the data grid displays a preview of results of the transform. **Transform previews** assist in specifying and validating the transformation steps before they are applied.

Although final results may not be exactly represented, a preview gives a good indication of the changes. When you review a preview, keep the following in mind:

- Previewed columns cannot be filtered or hidden.
- You cannot rename or change the data type of a previewed column until you add the change to your recipe.
- Selection of column headers is disabled.

For example, you want to remove leading and trailing quotation marks from all columns. You highlight the quotation marks in one cell and select a suggestion card for Replace:

Source

to be dropped

Preview

column2	ABC	column3	ABC	column4	ABC	column5	#	column5
ABC	column2	ABC	column3	ABC	column4	ABC	column5	
4208107376653317		2014-01-03		2007-01-03		7		
4208106585213952		2014-10-27		2004-10-27		9		
4208109505943761		2012-02-18		2010-02-18		2		
4208113686690378		2015-06-30		2007-06-30		6		
4208104023796892		2015-08-04		2006-08-04		7		
4208102203151829		2015-06-01		2003-06-01		10		
4208112003532067		2014-04-23		2004-04-23		9		
4208106783584469		2015-06-17		2006-06-17		7		
4208112393379296		2014-11-19		2003-11-19		10		

Affects 17 columns, all rows

393379296	"2014-11-19"	"2003-11-19"	10	10
840756415	"2014-01-02"	"2002-01-02"	12	12
829594575	"2014-06-16"	"2006-06-16"	7	7
424063352	"2014-09-18"	"2005-09-18"	8	8
992553499	"2014-11-14"	"2010-11-14"	3	3
312445947	"2015-04-30"	"2004-04-30"	9	9
660745200	"2014-02-24"	"2007-02-24"	6	6
568375941	"2015-04-07"	"2006-04-07"	7	7

nns 3,532 Rows 2 Data Types

Show only affected ☐ Columns ☐ Rows

Replace

" " with ' ' in column5

Edit Add

{start}"|"{end}` with ' ' in column5

{start}"|"{end}` with ' ' in all columns

Split on values matching

See all

" \" 2 times

" "

" " starting after {start} ending before {alphanum-underscore}+

Extract values matching

See all

" " starting after {start} ending before

Cancel

Figure: Replace Suggestion Card

To replace in all columns, you click the third option. To verify that it is correct for your needs, you click **Edit** .

The transform preview displays in the data grid for each column to be modified:

Tip: The preview also contains updated data quality bars and column histograms. You can use them to test for changes on counts or column values.

Source to be dropped Preview Source to be dropped Preview Source to be dropped Preview

ABC column2 column2 ABC column3 column3 ABC column4 column4

3,532 Categories 3,530 Categories 105 Categories Nov 2009 - Dec 2015 109 Categories

ACCOUNT ACCOUNT END_DATE END_DATE START_DATE

4208107376653317 4208107376653317 2014-01-03 2014-01-03 2007-01-03

4208106585213952 4208106585213952 2014-10-27 2014-10-27 2004-10-27

4208109505943761 4208109505943761 2012-02-18 2012-02-18 2010-02-18

4208113686690378 4208113686690378 2015-06-30 2015-06-30 2007-06-30

4208104023796892 4208104023796892 2015-08-04 2015-08-04 2006-08-04

4208102203151829 4208102203151829 2015-06-01 2015-06-01 2003-06-01

4208112003532067 4208112003532067 2014-04-23 2014-04-23 2004-04-23

4208106783584469 4208106783584469 2015-06-17 2015-06-17 2006-06-17

4208112393379296 4208112393379296 2014-11-19 2014-11-19 2003-11-19

4208111840756415 4208111840756415 2014-01-02 2014-01-02 2002-01-02

4208113829594575 4208113829594575 2014-06-16 2014-06-16 2006-06-16

4208106424063352 4208106424063352 2014-09-18 2014-09-18 2005-09-18

4208112992553499 4208112992553499 2014-11-14 2014-11-14 2010-11-14

4208104312445947 4208104312445947 2015-04-30 2015-04-30 2004-04-30

4208108660745200 4208108660745200 2014-02-24 2014-02-24 2007-02-24

4208106568375941 4208106568375941 2015-04-07 2015-04-07 2006-04-07

34 Columns 3,532 Rows 7 Data Types Show only affected Columns Rows

Replace

with in column5

{start}|{end} with in column5

{start}|{end} with in all columns

Edit Add

Split on values matching See all

2 times

starting after {start} ending before {alphanumeric-underscore}+

Extract values matching See all

starting after {start} ending before

Cancel

Figure: Transform Preview for Replace

The transform preview displays for all valid transforms.

Tip: Press ESC to cancel a preview.

When you modify the transform in the Transform Builder, the preview is updated as you type.

- You can review the preview results using independent scroll bars in the transform preview.
- Previews of transforms are displayed next to each source column.
- You can apply filters to the preview through the filter drop-down. These filters persist after the preview is completed.

Preview color scheme:

Color	Meaning
blue	matching data
yellow	updated data
green	unchanged data
red	deletions
blurred	transform is invalid

An error might appear while you are typing the transform. Because the transform is not valid, the transform preview appears blurred and does not display the results of the transform.

Variations:

Previews vary depending on the type of transform. For example, if you create a transform to replace values in a column, the preview displays the results of that transform. If you create a transform to extract values from a column, the preview displays the values that will be extracted.

Split pane previews: For transforms that change the number of rows and columns, the results of the transform are previewed in a split pane view in the data grid and in the column browser. The following transform types are previewed in a split pane:

- pivot
- deduplicate
- unpivot

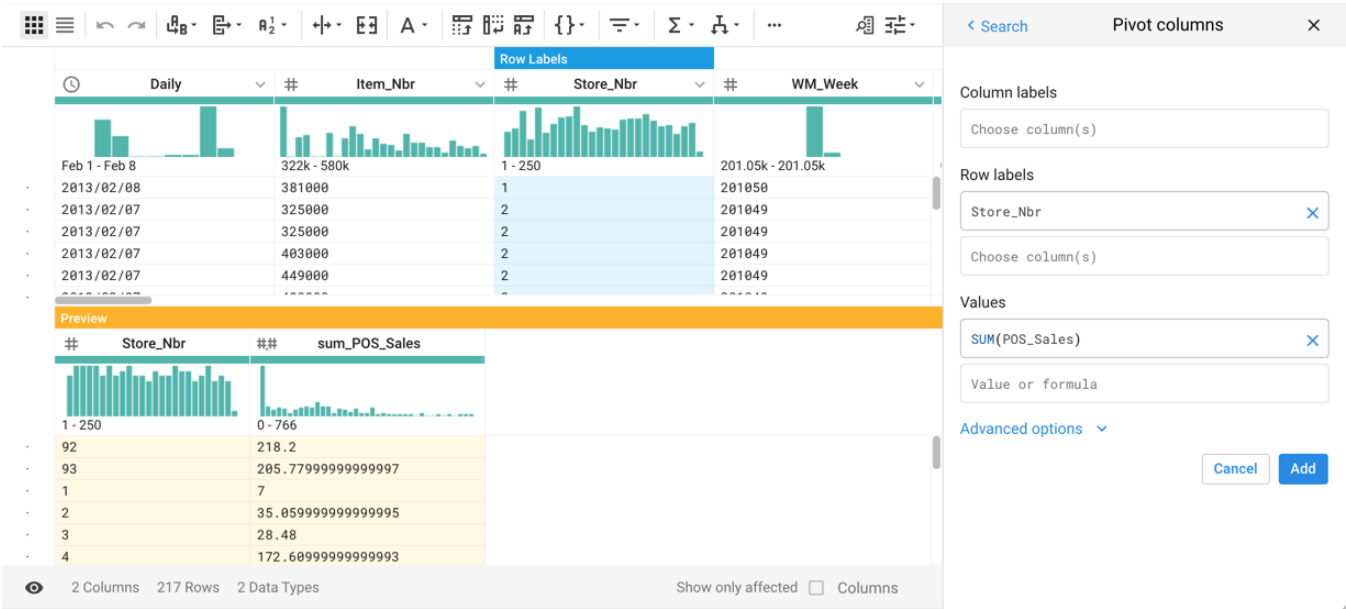


Figure: Split Pane Preview

Context Panel

On the right side of the Transformer page, the context panel displays one of multiple panels, depending on the current state or selection of the data grid.

The following panels may be displayed within the context panel.

To close the context panel, click the X icon in the upper-right corner.

Recipe Panel

Contents:

- *Recipe Toolbar*
 - *Toolbar context menu*
 - *Recipe Options*
- *Step Options*
 - *Single-step options*
 - *Multi-step options*

Through the Recipe panel, you can review and modify the steps of the recipe that you have already created and add new steps to your recipe at the current location. You can also flag a step that requires review without which jobs cannot be run.

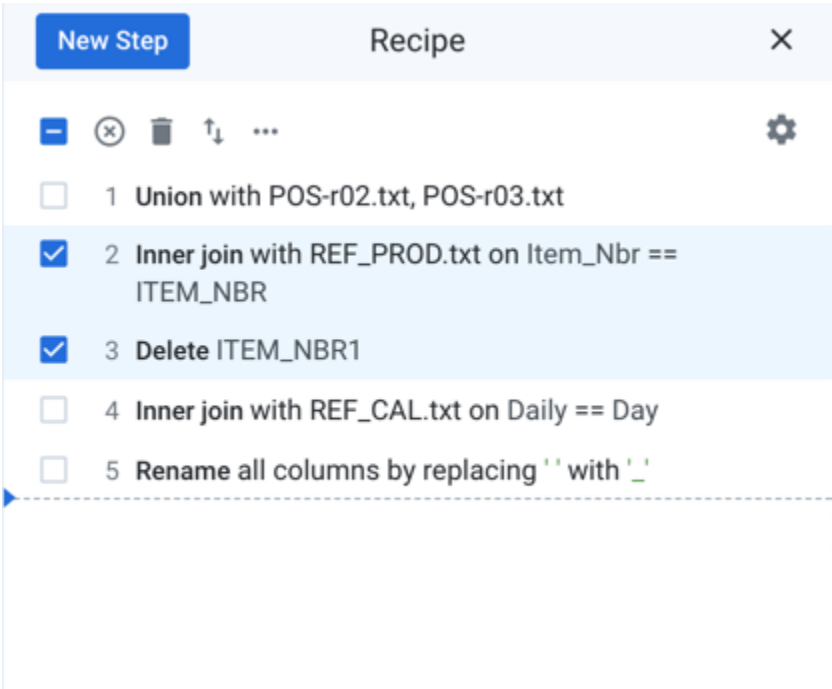


Figure: Recipe panel

In the Recipe panel, the dotted horizontal line indicates the state of what is displayed in the data grid.

- To add a new step at the cursor location, click **New Step**. See *Search Panel*.
- This cursor can be moved. Select where you'd like to move to display in the data grid. Your selection can be a step or between two steps. Then, from the context menu in the Recipe panel, select **Go to selected**. Details are below.

Tip: You can undo and redo changes to your recipe through the transformer toolbar. See *Transformer Toolbar*.

Select Steps:

You can also select one or more steps to move or modify.

- To select a step, click the step or its checkbox to the left.
- Select other step checkboxes to add to your selection.
- You can toggle selecting all steps at the top of your recipe.

When you select one or more steps, the recipe toolbar is displayed. See below.

Keyboard Shortcuts:

Shortcut	Description
COMMAND/CTRL + click	Toggle selection of a step without changing status of other selected steps.
SHIFT + click	Select the range of steps between the current one and the last selected one.

Recipe Toolbar

When you select one or more steps in your recipe, the recipe toolbar is displayed above your steps.

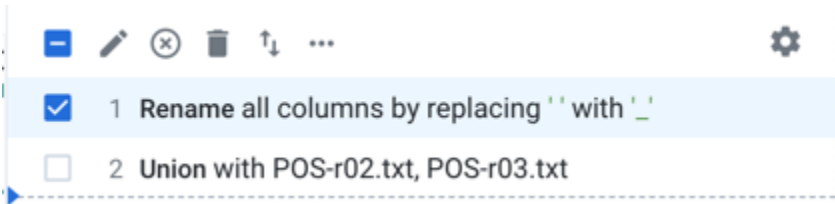


Figure: Recipe toolbar

Tools:

- **Select/Deselect checkbox:** Select or deselect all steps in your recipe.
- **Edit:** (Single step selected only) Edit the recipe step.
- **Disable/Enable:** Disable the selected step or steps.
- **Delete:** Delete the selected step or steps.
- **Move:** Move the selected steps to the start or end of the recipe or up or down one step in the recipe.

Moving a recipe can cause steps to break. Some fixups may be required.

- **Recipe toolbar context menu:** See below.
- **Recipe options:** See below.

Toolbar context menu

The following options are available in the toolbar context menu.

NOTE: Some of these options may not be available depending on the selected step or whether you have selected multiple steps.

- **Go to:**
 - **Selected:** Move the recipe cursor to the selected step. The data grid is updated to display the dataset up to the selected step in the recipe.

- **Start/End:** Go to the first or last step of the recipe.
- **Insert after step:** Insert a new step after the selected one. See *Search Panel*.
- **Duplicate:** Create a copy of the selected step and insert it after the selected one.

Tip: You can modify a duplicated step to create variations of your steps.

NOTE: Some steps, such as union or join operations, cannot be duplicated.

- **Cut:** Cut the selected step to the clipboard.
- **Copy:** Copy the selected step to the clipboard.

NOTE: Some steps, such as union or join operations, cannot be copied.

- **Paste after step:** Paste the step in the clipboard after the selected step. See Notes below.
- **Create macro:** You can create an independent object called a macro out of a selection of steps, which can be applied in other recipes. See *Create or Replace Macro*.

NOTE: Some types of steps cannot be included in macros. For more information, see *Overview of Macros*.

Notes on pasting:

- If you are accessing the application over secure HTTPS, you may be prompted to authorize permission to paste. Some browsers store copied data in the global clipboard for the computer, which causes this security warning to be displayed.
- If you are accessing the application over less-secure HTTP, pasting in Wrangle text that was modified outside of the application is not supported.

Recipe Options

The following options are available from the Gear menu for any number of selected steps.

- **Display Wrangle /natural language.** Toggle between displaying recipe steps in native Wrangle or in more readable language (default).
- **See Edit History.** Display history of recipe edits by user. See *Edit History Panel*.
- **Download Recipe as Wrangle .** You can download the recipe as text for offline review and storage.

NOTE: Wrangle recipes are stored in the Trifacta® database. You cannot upload new or modified recipes to the platform.

- **Download Sample data as CSV.** Download the sample currently displayed in the Transformer page in CSV format.

Tip: The downloaded CSV reflects the sample modified up to the currently selected recipe step, so you can use this to acquire and review data transformation in progress. For more information, see *Take a Snapshot*.

Step Options

NOTE: Some of these options may not be available depending on the selected step or whether you have selected multiple steps.

Single-step options

When you mouse over a step, you can choose to edit or remove the step or perform other operations from the context menu on the right side of the step. You can also right-click the step to open the menu.

NOTE: If you go to a step before the latest one in your recipe, the data grid is updated to reflect the state of the sample at that time. All subsequent steps are grayed out in the Recipe panel. When you run a job, all steps in the entire recipe are executed.

NOTE: In a dataset that is shared, multiple users cannot make changes to the recipe at the same time.

- **Go to selected:** Move the recipe cursor to the selected step.
- **Edit:** Edit the step.
- **Disable/Enable:** Disable or enable the selected step.
- **Delete:** Delete the selected step.
- **Move:** Move the step to the start or to the end of your recipe. Or, you can move it up or down one step at a time in the recipe.

Moving or deleting a recipe or disabling/enabling steps a recipe can cause steps to break. Some fixups may be required.

- **Insert new step after:** Insert a new step after the selected step. See *Search Panel*.
- **Duplicate:** Create a copy of the selected step and insert it after the selected one.

Tip: You can modify a duplicated step to create variations of your steps.

NOTE: Some steps, such as union or join operations, cannot be duplicated.

- **Cut:** Cut the selected step to the clipboard.
- **Copy:** Copy the selected step to the clipboard.

NOTE: Some steps, such as union or join operations, cannot be cut or copied.

- **Paste after:** Paste the step in the clipboard after the selected step.
- **Create macro:** See previous.

Flag for review

The Flag for review option enables you to flag a step in the recipe for review. When you flag a step for review, a warning icon is displayed for the corresponding step, alerting flow users that the step must be reviewed and cleared of the flag before running the job.

NOTE: This feature may need to be enabled in your environment.

NOTE: When a step is flagged for review, all downstream steps are disabled, and you cannot run the job until all flagged steps are reviewed. Steps must be reviewed in descending, top-to-bottom order.

Actions:

- **Flag for review:** Marks the current step as pending review.
 - Select the Flag for review option, add a name and description, and click **Flag**.
 - A warning icon is displayed against the corresponding step.
- **Unflag for review:** Removes the flag from the step, unblocking review.
- **Rename review step:** Edit the name and description of the flag.
- **Mark as reviewed /Mark as pending review:** You can toggle between these options to mark the review as complete or to mark the step as pending review. After you select Mark as reviewed, a tick mark is displayed against the reviewed step.

For more information, see *Flag for Review*.

Multi-step options

If you select multiple steps in your recipe, the following options are available:

- **Disable/Enable:** Disable or enable the selected steps.
- **Delete:** Delete the selected steps.
- **Move:** Move the steps to the start or to the end of your recipe. Or, you can move it up or down one step at a time in the recipe.

Moving a recipe can cause steps to break. Some fixups may be required.

- **Duplicate:** Create copies of the selected steps and insert them after the selected one.
- **Cut:** Cut the selected steps to the clipboard.
- **Copy:** Copy the selected steps to the clipboard.
- **Paste after:** Paste the steps in the clipboard after the selected step.
- **Create macro:** See previous.

Transform Builder

Contents:

- *Step 1 - Select transformation in the Search Panel*
- *Step 2 - Specify the column(s), formula, or condition*
 - *Columns*
 - *Patterns*
 - *Delimiter Groups*
 - *Condition*
- *Step 3 - Grouping, Ordering, and Naming*
- *Step 3 - Specify other parameters*
- *Step 4 - Add the step*
- *Edit a transform*

The Transform Builder enables you to search for transformations and to rapidly assemble complete transform steps through a simple menu-driven interface.

After you select the transformation to apply, all relevant parameters can be configured through selection or type-ahead fields, so that you can choose from only the elements that are appropriate for the selected transformation.

To open the Transform Builder, begin creating a step through one of the following methods:

- Select a transformation from the Transformer toolbar. See *Transformer Toolbar*.
 - Click the Macros icon in the toolbar to apply a macro as your next step. See *Apply a Macro*.
- Select a transformation from a column menu. See *Column Menus*.
- Search for and select a transformation in the Search panel. See *Search Panel*.
- Click **New Step** in the Recipe panel. See *Recipe Panel*.
- Edit an existing step.

[< Recipe](#)

New formula

×

Formula type

required

Single row formula

Create a new column from a single row formula

Formula

required

SUM(POS_Sales)

New column name

sumSales

Cancel

Add

Figure: Transform Builder

Keyboard shortcuts:

Tip: When keyboard shortcuts are enabled, press ? in the application to see the available shortcuts. Individual users must enable them. See *User Profile Page*.

Step 1 - Select transformation in the Search Panel

From the Search panel, begin typing to see the list of available transformations. Select your preferred one.

Join and union transformations have dedicated pages for configuring this transformations. You can enter `join` `datasets` or `union` as the search term to open the corresponding tool:

- See *Join Window*.
- See *Union Page*.

For a list of available transformations, see *Transformation Reference*.

Step 2 - Specify the column(s), formula, or condition

Depending on the transform that you have selected, you must specify one or more of the following in the Transform Builder.

- Some transforms support combinations of the following.
- Some transforms, like `deduplicate`, require no parameters.

The following are general categories of object types:

- **Literal values.** A literal, or constant, value is a fixed numeric, string, Boolean, or other type of value, which does not change depending on the row under evaluation.
- **Functions.** Trifacta® supports a wide variety of numerical, statistical, and other function types. For a list of available transforms and functions, see *Wrangle Language*.
- **Columns.** When a column name is used in a formula, the transform uses the value in the named column for the currently evaluated row.
- **Operators.** You can apply logical, numeric, or comparison operators as part of your formula.
 - See *Logical Operators*.
 - See *Numeric Operators*.
 - See *Comparison Operators*.
- **Parameters:** Add a reference to a flow parameter in your transformation. See *Manage Parameters Dialog*.
- **Metadata.** You can insert special strings that evaluate to references of your dataset's metadata. For more information, see *Source Metadata References*.

Columns

Using the Columns parameter, you can select or specify the column or columns to which to apply the transform.

The following options are available when specifying one or more columns in a transformation:

- **Multiple:** Select one or more discrete columns from the drop-down list.
- **All:** Select all columns in the dataset.
- **Range:** Specify a start column and ending column. All columns inclusive are selected.
- **Advanced:** Specify the columns using a comma-separated list. You can combine multiple and range options under Advanced. Ranges of columns can be specified using the tilde (~) character. Example:

Store_Nbr, Item_Nbr, WM_Week~POS_Cost

Patterns

For some transforms, you can specify patterns to identify conditions or elements of the data on which to take action. These matching patterns can be specified using one of the following types.

Pattern Type	Description	Example
Literal value	An exact string or value.	The following matches on the exact value between the quotes: 'This is what I want to match.'
Pattern	Trifacta supports a variety of macro-like pattern identifiers, which can be used in place of more complex regular expressions.	The following matches when two digits appear at the beginning of a value: {start}{digit}{digit}
Regular expression	Regular expressions are a standard method of describing matching patterns. NOTE: The syntax of regular expressions can be complex and can lead to unexpected results if they are improperly specified. Regex is considered a developer-level skill.	The following matches on all numerical values from 0 to 99: /^d\$ ^d\d\$/

For more information on pattern-based matching, see *Text Matching*.

Flow parameter: You can also insert a flow parameter into your pattern-based inputs in the Transform Builder. To reference a flow parameter, click the Parameterize icon above any field that accepts pattern-based inputs.

- The parameter values or any overrides applied to those values are applied to the results displayed in the data grid, as well as during job execution.
- For more information on creating flow parameters, see *Manage Parameters Dialog*.
- For more information on parameterization, see *Overview of Parameterization*.

Delimiter Groups

In the Transform Builder, transforms that require delimiter are organized into delimiter groups, so that you specify only the elements of a pattern that work together. Delimiter groups apply to the following transforms:

- *Countpattern Transform*
- *Extract Transform*
- *Replace Transform*
- *Set Transform*
- *Split Transform*

Delimiter groups are listed below.

Delimiter group	Description
On delimiter	Transformation is applied based on a specific literal or pattern.
Between delimiters	Transformation is applied on database between two literal or pattern-based delimiters. Details are below.
On multiple delimiters	Transformation is applied based on a sequence of delimiters. An individual pattern can be a string literal, Pattern , or regular expression, and the sequence can contain combinations of these pattern types.

Between positions	Transformation is applied based on a starting index position and an ending index position. Index positions start from 0 on the left side of any cell value.
On positions	Transformation is applied based on a sequence of listed index positions. Index positions start from 0 on the left side of any cell value.
At regular interval	Transformation is applied at every <i>nth</i> position. Index positions start from 0 on the left side of any cell value.

For more information on the underlying syntax for delimiter groups, see *Pattern Clause Position Matching*.

Between two delimiters

Matches any values that appear between two delimiters. One delimiter describes the beginning of the match, and the other delimiter describes the end of the match.

Each delimiter can either include or exclude the matching value:

Transform Builder option	Include as part of transform	Include/Exclude
Start delimiter	false	Excludes sub-pattern
Start delimiter	true	Includes sub-pattern
End delimiter	false	Excludes sub-pattern
End delimiter	true	Includes sub-pattern

Condition

A **condition** is an expression that yields a `true` or `false` value. A condition may include all of the elements of a formula. This value determines whether the transformation is applied to the evaluated row.

Step 3 - Grouping, Ordering, and Naming

A number of transforms support the following parameters.

NOTE: Transforms that use the `group` parameter can result in non-deterministic re-ordering in the data grid. However, you should apply the `group` parameter, particularly on larger datasets, or your job may run out of memory and fail. To enforce row ordering, you can use the `sort` transform. For more information, see *Sort Transform*.

Group parameter: For transforms that aggregate data, such as `pivot` or `window`, you can specify the column by which you wish to group the computed aggregations. In the following example, all values in the `POS_Sales` column are summed up for each value in the `Store_Nbr` column.

Transformation Name	Pivot columns
Parameter: Row labels	Store_Nbr
Parameter: Values	sum(POS_Sales)

Assuming that there are entries in the `Store_Nbr` column, the resulting transform step has 50 rows, each of which contains the total sales for the listed store number.

Order parameter: Some transforms support the `order` parameter, which allows you to specify the column of values that are used to sort the output. In the following example, all aggregates `Sales` values are ordered by the contract date and grouped by `State`:

Transformation Name	Pivot columns
Parameter: Row labels	Store_Nbr
Parameter: Column labels	contractDate
Parameter: Values	sum(Sales)

New Column Name parameter: For transforms that generate new columns, such as `derive` and `extract`, you can optionally specify the name of the new column, which saves adding a step to rename it. In the following example, the values of `colA` and `colB` are summed and written to the new column `colC`:

Transformation Name	New formula
Parameter: Formula type	Single row formula
Parameter: Formula	colA + colB
Parameter: New column name	colC

Step 3 - Specify other parameters

Depending on the transform, you may be presented with other required or optional parameters to specify. See *Transforms*.

Step 4 - Add the step

When you have finished your transform step, review the preview in the data grid.

If the results look ok, click **Add**.

The step is added to your recipe and applied to the data grid.

- See *Data Grid Panel*.
- See *Transform Preview*.

Edit a transform

After you have added a step, you can modify it as needed. In the Recipe panel, select the Pencil icon next to the recipe step. The step is displayed for editing in the Transform Builder.

Search Panel

Through the search context panel, you can locate transformations to specify and add at the current location in your recipe.

You can search for transformations to add in any of the following ways:

- At the top of the Transformer page, click the Magnifying Glass icon.
- When you choose to add a new step to your recipe, the Search panel opens in the context panel.
- To add a new recipe from anywhere in the Transformer page, press `CTRL/COMMAND + K`. Enter a search string for your transformation step.

Enter text or browse the available transformations to begin building your next step.

Tip: When you enter a search term, you can choose to use that term to search the product documentation. Select the **Search documentation** entry.

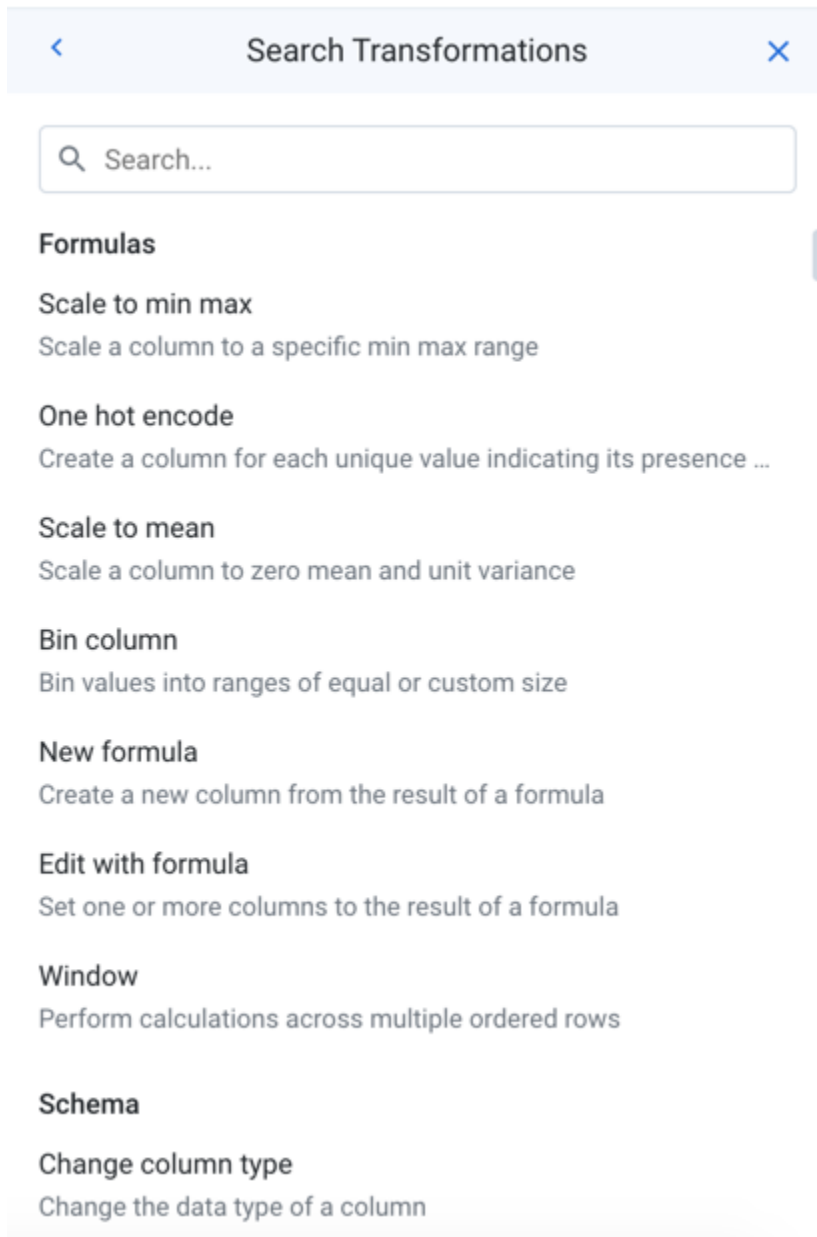


Figure: Search Panel

To locate transformations, you can browse or search.

Browse:

- Headings like **Formulas** indicate categories of transformations.
- For each transformation, you can review a brief description of it.

Search:

- Enter a few characters of a transformation, function or other object such as a metadata reference for which you are looking. Matches are underlined in the panel.
 - To see a list of all available functions, enter `function`. When selected, a New formula transform is pre-specified using the selected function.

- A copied step can be pasted back into the Search panel and modified from the Transform Builder. Copy and paste may not be supported across different releases of the product.
- If you are familiar with Wrangle transforms, you can enter the transform name in the search bar. For example, type `window`.

After you have selected the transformation to build, the Transform Builder is pre-populated with some configuration done for you, so you can begin specifying the transformation. For more information, see *Transform Builder*.

Edit History Panel

Through the Edit History panel, you can review the sequence of edits to the current recipe by individual contributors. This panel assists in determining who made which changes and when they were made.

- If the dataset is part of a shared flow, edits appear created by each user.
- If the dataset has been sent as a copy, all steps made by the user who shared the flow appear as a single edit.
- If the flow containing the dataset has not been shared, the only user listed in the Edit History is the owner of the flow.
- For more information on sharing, see *Overview of Sharing*.
- To open the Edit History panel, select **See Edit History** in the Recipe panel. See *Recipe Panel*.

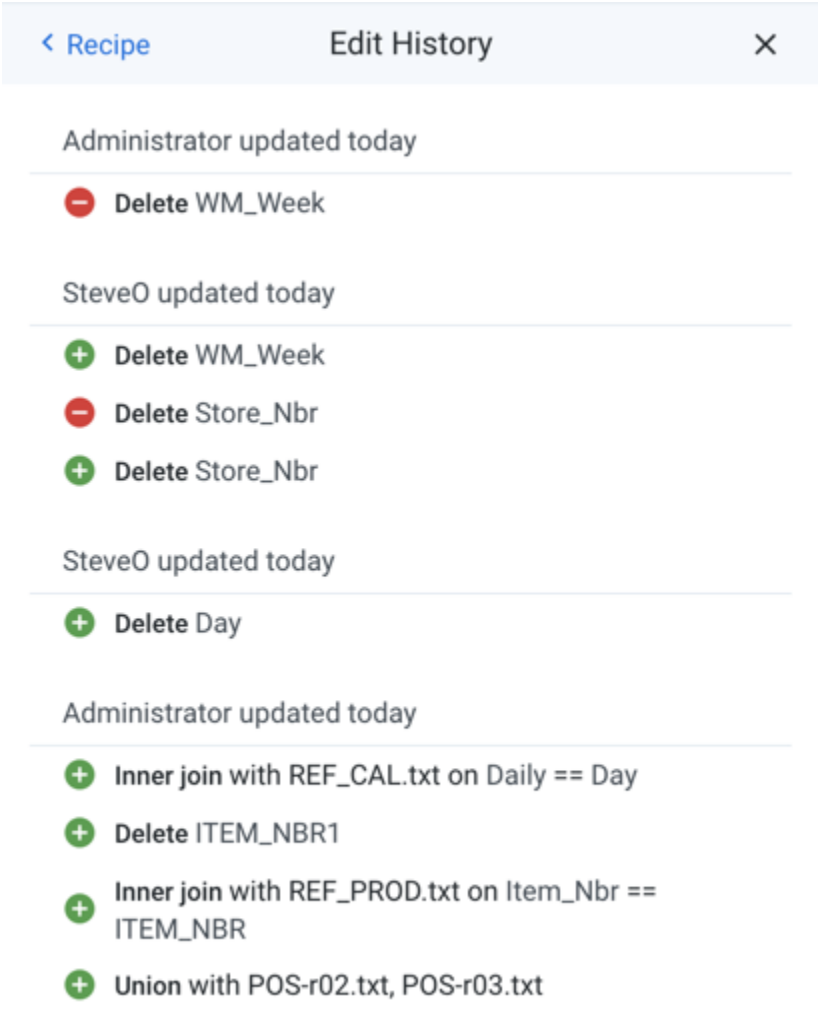


Figure: Edit History panel

Sets of changes are grouped by the user who performed them and listed with most recent changes at the top.

Changes to individual transforms are listed within a set:

Icon	Description
+	The listed step was added (add operation).
-	The listed step was deleted (delete operation).

Edit operations are represented by a delete operation and then an add operation.

Samples Panel

Contents:

- *Collect new sample*
- *Collected samples*
- *Cancel sample jobs*

For smaller datasets, the Transformer page displays the entire dataset. For larger ones, the source data is sampled for use in the Transformer page. Through the Samples panel, you can create new samples and select them for display in the Transformer page.

At the top of the Transformer page, the type of the current sample is displayed next to the dataset name. To open the Samples panel, click the current sample indicator:

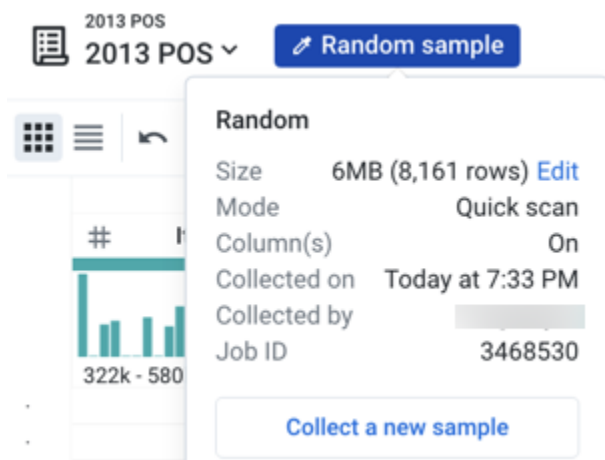


Figure: Click the current sample button.

In the example above, you can see that the current sample is a Random sample.

Initial Data: The sample is taken from the first set of rows in the first file or table that is part of the dataset.

- In some cases, the Initial Data sample is the entire dataset.

Tip: For purposes of loading the data, the initial data sample is generated and displayed at first. For a better representation of the entire dataset, you should create a new sample.

- In other cases, the Initial Data sample is generated from a collection of files. For more information on this special sampling type, see *Overview of Sampling*.

To create a new sample, click **Collect a new sample**.

The Samples panel is displayed on the right side of the screen:

Tip: You can also open the Samples panel by clicking the Eyedropper icon at the top of the page.

To review all samples that you have created, see *Sample Jobs Page*.

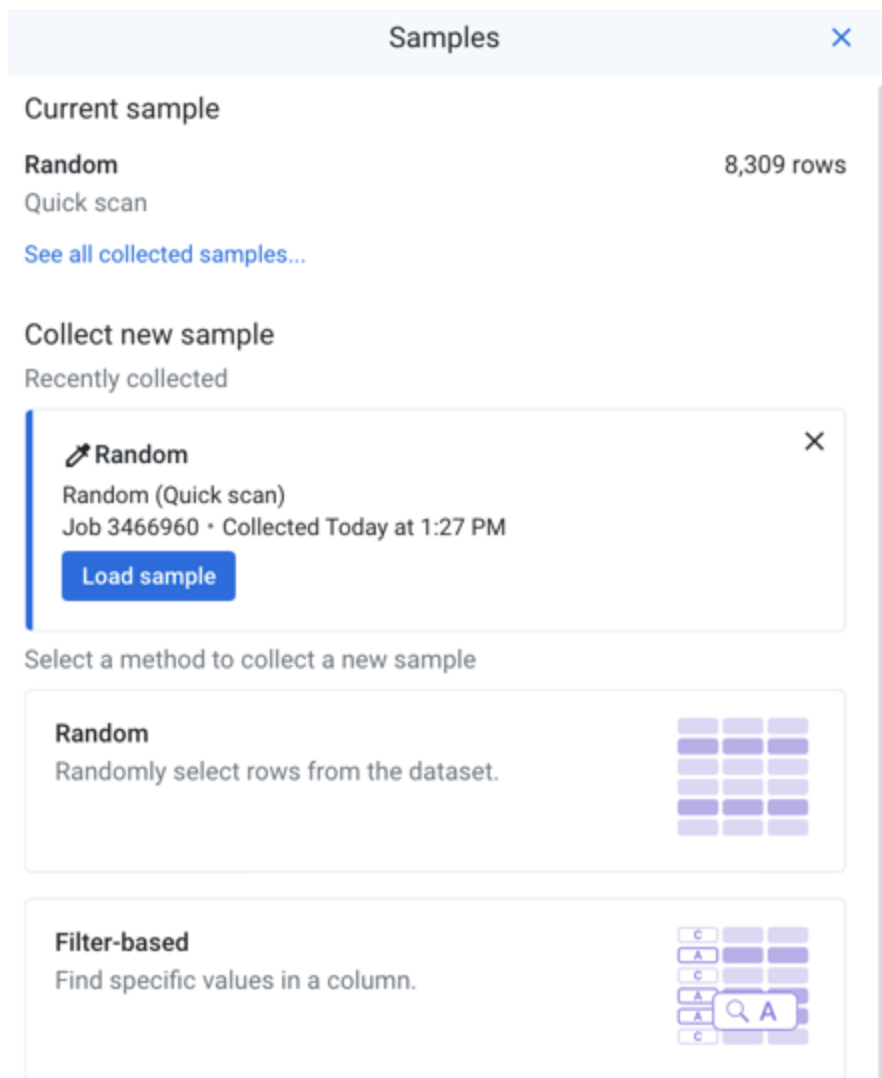


Figure: Samples Panel

Current sample:

At the top of the panel, you can review the currently loaded sample. Each user has his own active sample on a dataset.

NOTE: When a new sample is generated, any Sort transformations that have been applied previously must be re-applied. Depending on the type of output, sort order may not be preserved.

- **Initial Data:** By default, the application loads the first N rows of the dataset as the initial data sample when the Transformer page is opened. The number of rows depends on column count, data density, and other factors. If the dataset is small enough, the full dataset is used.

NOTE: By default, samples may be up to 10 MB in size or may be limited based on the maximum number of files that can be scanned. For datasets smaller than this limit, the entire dataset is loaded. See *Overview of Sampling*.

- Click the link in the current sample card to see the list of all available samples.

Tip: To change the name of a sample, click its card in the list of all available. Then, click the Edit icon.

New samples:

Below the current sample, you can review the available options for creating new samples. Each type of sample reflects a different method of collection.

The data that is displayed in the data grid is based on all of the upstream samples after which all subsequent steps in each upstream recipe are performed in the browser. If you have a large number of steps or complex steps between the recipe locations for your samples in use and your current recipe location, you may experience performance slow-downs or crashes in the data grid. For more information on sampling best practices, see <https://community.trifacta.com/s/article/Best-Practices-Managing-Samples-in-Complex-Flows>.

- To collect a new sample, click the appropriate sample card. See below.

Tip: A sample execution is a type of job. Any issues related to the execution of a sampling can be reviewed through the job logs.

- To cancel a sample collection, click the X next to the progress bar. The interrupted sample is listed as unavailable. You can download the logs from the unfinished sample collection.
- After a sample is created, you can load it at any time, as long as it is still valid. Next to a collected sample, click **Load sample**.
- For more information on sampling methods, see *Overview of Sampling*.

Status bar:

At the bottom of the Transformer page, you can review the number of rows and columns and count of data types in the currently displayed sample.

NOTE: As you add transformation steps to your recipe, the values in the status bar change to reflect the current state of the loaded sample.

NOTE: Some operations, such as `union`, may change the row counts without invalidating the sample. If the operation increases the size of the dataset beyond the sample size limit enforced by the application, then a subset of those rows is displayed. This is a known issue.

Collect new sample

When a new sample is collected, it is gathered based on the current location in the recipe when the sample is gathered. So, if the recipe contains steps that join in other datasets, those joins are performed to bring together the data from which the sample is executed.

Figure: Collect new sample panel

NOTE: Except for the initial data sample, all samples are generated based on the steps leading up to the location of the cursor in the recipe. If earlier steps are deleted or modified, the collected sample can be invalidated.

NOTE: When sampling from compressed data, the source is uncompressed, and a new sample of it is loaded into the data grid. As a result, the sample size you see in the grid corresponds to the uncompressed data.

Steps:

- In the Samples panel, select the type of sample to create. For more information on sample types, see *Overview of Sampling*.
- In the Collect new sample panel, specify the following parameters, some of which may not be required for your sampling method:
- **Choose a sampling method:** Select or enter the type of sample. If you already selected a sampling method, this value is pre-populated for you.
- **Name:** You can enter a new name of the sample as needed.

Tip: Naming your samples can assist in tracking them later. For example, you might choose to add a date stamp to the name to track when you captured the sample.

- **Scan Type:** (Does not apply to all sampling methods) Types of scans:
 - **Quick** - performs a random scan of the dataset to extract the appropriate number of rows for the sample.
 - **Full** - gathers the sample from the entire dataset. Depending on the size of the dataset, this method can take a while.
- **Use latest data:** When collecting a Full Scan sample from a JDBC source and performance ingest caching has been enabled, you can choose to override the cached data and to gather all of your data from the original sources.

NOTE: If the cached data has expired, the sample is always collected from the original sources, even if this option is not selected.

Click **more details** to review the list of datasets whose cached data will be overridden.

Ingest caching applies to non-native relational (JDBC) sources. For more information, see *Configure JDBC Ingestion*.

- **Column or columns:** (Stratified, Cluster-based) Name of the column from which to gather values to evaluate (Anomaly-based) Specify the name or names of one or more columns containing the anomalies to include in your sample. Multiple columns can be specified by comma-separated values. A column range can be specified using the tilde (~) character.
- **Condition:** (Filter-based, Stratified, Cluster-based, Anomaly-based) Filter the sample based on a specified condition. For example:

```
invoiceDate > 90
```

- **Anomaly type:**(Anomaly-based) Select the type of anomalous values to include in your sample: invalid, missing, or both types.
- **Variable overrides:** If one or more variables is associated with your dataset, you can define the value overrides to be applied when the sample is executed.
 - You can use these overrides to sample data from different source files in your dataset with parameters.
 - A variable can have an empty value.
 - For more information, see *Overview of Parameterization*.
- To begin collecting the sample, click **Collect**.
- You can continue working while the sample is collected. When the sample is available, a status message is displayed in the Transformer page.
- You can click **Load Sample** in the Samples panel to begin using it.

Collected samples

In the Collected samples panel, you can review the available and unavailable samples. If applicable, you can review the variable override values that were applied during the sampling.

To use one of the available samples, click **Load**. The sample is loaded in the data grid. For more information, see *Generate a Sample*.

NOTE: If you add recipe steps that change the number of rows in your dataset (or a few other edge case steps), some of your existing samples may no longer be valid. When you execute a join, union, or delete action or edit steps before this action, you may be prompted with the Change Recipe dialog, which includes the following message:

Your change will invalidate some of the currently available samples for this source. The invalid samples will be deactivated.

For more information on the types of transformations that can invalidate samples, see *Reshaping Steps*.

Cancel sample jobs

You can cancel a sample job that is currently being executed.

- In the Samples panel, locate the job in-progress. Click the X.
- You can also review and cancel sample jobs through a page in the Trifacta application. For more information, see *Sample Jobs Page*.

Selection Details Panel

Contents:

- *Select Column*
 - *Data Quality Bar*
 - *Unique Values*
 - *Patterns*
- *Other Selections*
 - *Select Multiple Columns*
 - *Select Column Values*
 - *Select Values in the Data Grid*
- *Suggestions*

In the Selection Details panel, you can review an active profile of your current selection or selections in the data grid or column browser and review patterns and suggestions for transformations.

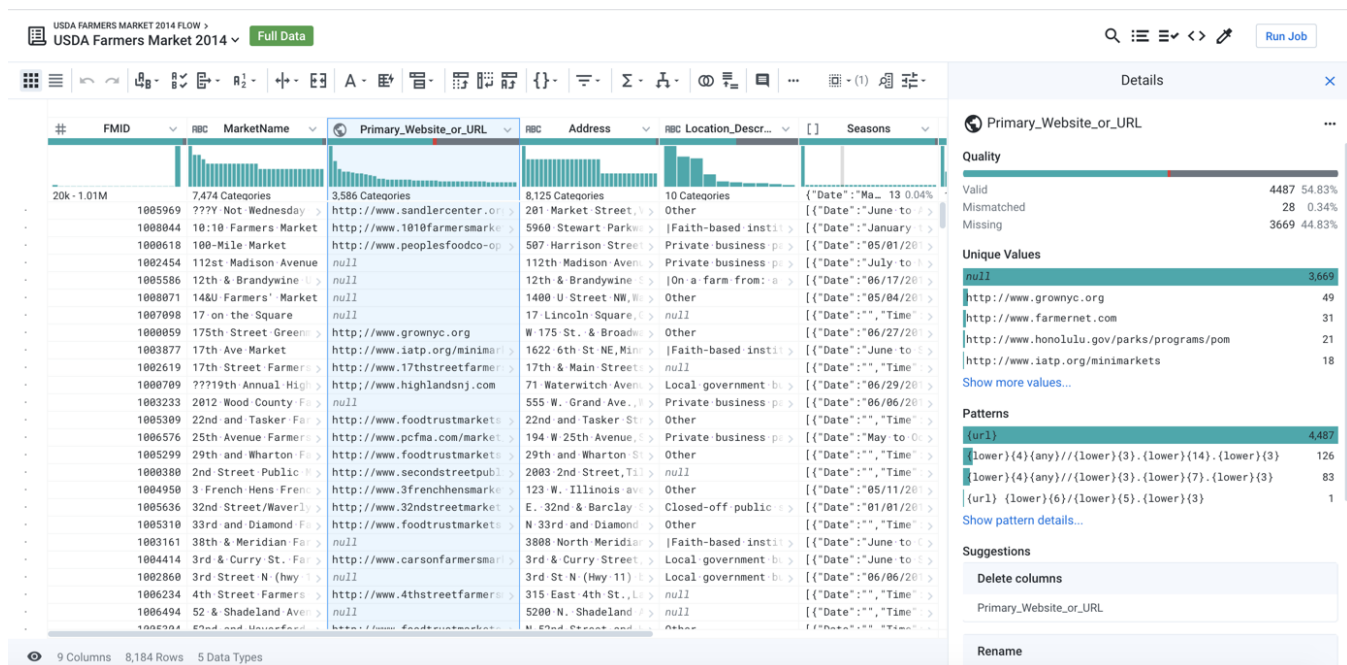


Figure: Selection Details Panel for selected column

Based on what you have selected, the panel is updated with context-specific information about your selection(s) and a set of actions that you can take on the data.

Select Column

When you select a column, the following sections appear in the Selection Details Panel.

- To change the data type of the column, click the menu to the left of the column name in the panel.
- From the caret to the right, you can make selections from the column menu. Available selections may vary by column data type.
- For more information, see *Column Menus*.

Data Quality Bar

Review the counts and percentages of valid, mismatched, and missing values in the column.

- Click one of the colored bars to select only the matching values in the column. See [Select Column Values](#) below.

Context menu:

Right-click the data quality bar to see a set of possible transformations:

- **Keep rows** - Keep rows that match the data quality bar you selected.
- **Delete rows** - Delete rows that match the data quality bar you selected.
- **Create new column flagging** - Create a new column containing `true` for each row that matches the data quality bar you selected. Otherwise, the row value in the new column is `false`.
- **Clear values** - For mismatched or empty rows, you can set the value to be empty.
- **Replace values** - For mismatched or empty rows, you can replace with a specific value. See [Replace Cell Values](#).

Unique Values

You can review the counts of the most frequently occurring values in the column.

To see all unique values, click **Show more values**.

- Use the Search bar to locate specific values among the list of unique ones.
- Click the Back button to return to the Selection Details panel.

Context menu:

Right-click any value bar to be prompted for a set of transformations specific to the applicable rows and values.

- **Keep rows with selected values** - Keep rows where the selected value appears. Delete the other rows.
- **Delete rows with selected values** - Delete rows where the selected value appears. Keep the other rows.
- **Create new column flagging** - Create a new column containing `true` for each row that matches the selected value(s). Otherwise, the row value in the new column is `false`.
- **Clear values** - You can set the row value to be empty for the selected value(s).
- **Clear others** - You can set the row value to be empty for all rows that do not match the selected value(s).
- **Replace values** - You can replace the selected value(s) with a specific value. See [Replace Cell Values](#).

Patterns

Based on your selected column or column values, Trifacta attempts to find patterns that match your selections. These patterns are represented as Trifacta patterns. For more information, see [Text Matching](#).

To see all patterns, click **Show more patterns**. For more information, see [Column Details Panel](#).

Context menu:

Right-click any pattern bar to be prompted for a set of transformations specific to the rows and values that match the pattern.

- **Keep rows with selected patterns** - Keep rows with values that match the selected pattern(s). Delete the other rows.
- **Delete rows with selected patterns** - Delete rows with values that match the selected pattern(s). Keep the other rows.
- **Create new column flagging** - Create a new column containing `true` for each row that matches the selected pattern(s). Otherwise, the row value in the new column is `false`.
- **Clear values matching patterns** - You can set the row value to be empty for the selected pattern(s).

- **Clear others** - You can set the row value to be empty for all rows that do not match the selected pattern(s).
- **Replace values with matching patterns** - You can replace the selected pattern(s) with a specific value.
See *Replace Groups of Values*.

Other Selections

Select Multiple Columns

When you select multiple columns, the following changes to the panel apply:

- Profiling of the data is not available, which also means that you cannot take action on individual values within your selection.
- The data type menu is no longer available.

You can access a column menu and review suggestion cards that are applicable to all of your columns.

Select Column Values

After you have selected a column, you can use the Selection Details panel to select individual values within the column, which updates the panel. Below, the mismatched values in the previously selected column have been selected from the data quality bar:

In the following example, your dataset contains a column of addresses. Within one of the values, you can select a zip code, which then triggers an appropriate set of suggestion cards:

The screenshot shows a data table titled 'USDA Farmers Market 2014' with columns: site_or_URL, RBC, Address, #, Address1, RBC, and Location_Description. The 'Address' column contains various addresses. To the right, a 'Suggestions' panel is open, displaying several suggestion cards. The first card, 'Extract values matching', has the suggestion '{zip}' selected. Other cards include 'Replace', 'Count values matching', and 'Split on values matching'. Each card has a 'See all' link and buttons for 'Edit' and 'Add'.

Figure: Suggestion Cards

In the above image, the first suggestion in the Extract values matching suggestion card has been selected by default, and its parameters have been specified to extract all zip codes from the source column (Address). While useful, this selection may not be your intention. Options:

1. Hover over different suggestions. A mini-preview appears to the left of the suggestion.
2. Click a different suggestion in the same card. In this case, you might click the first suggestion in the Replace card, so that you can remove zip code.

Tip: Optionally, suggestions can be provided to you based on your prior transformations or the transformations of other users in your workspace. These suggestions appear under the Recently used heading. For more information on enabling collaborative suggestions, see [Workspace Settings Page](#).

Tip: Among the suggestion cards, scroll down to see other suggestion cards that are off-screen. If it is present, a See all link displays more suggestions in that card. Variants further down in the suggestion card typically become more specific in their changes to the dataset or rarer in their usage.

3. Modify the selected transform. To edit the step, click **Edit**. You can fine-tune parameters of the transformation. See [Transform Builder](#).

Actions:

- To apply a suggestion, select the data, and then choose the suggestion to apply from the list of cards. Click **Add**.
- If needed, you can customize the selected suggestion. Click **Edit**.
- To generate a new set of suggestion cards, click **Cancel**. Then, select a different set of columns or values within a column.

Filter Panel

As needed, you can filter the rows and columns displayed in the data grid. To review and apply filters, click the Filter icon in the Transformer toolbar.

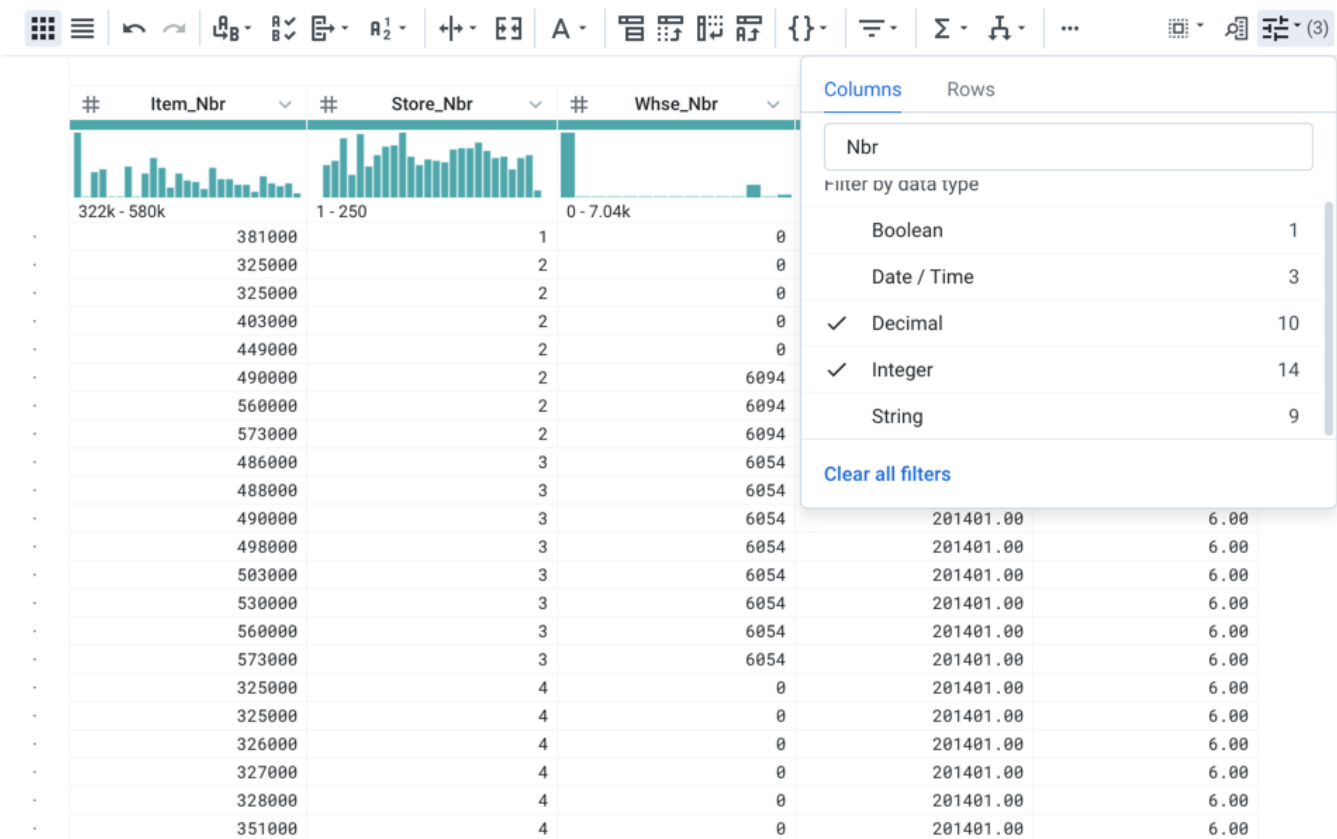


Figure: Data Grid Filters

Tip: Use data grid filters to assist in locating and selecting columns or values for selection and selection cards. After you have applied the suggested step to your recipe, remember to remove the filter on the data grid.

NOTE: Wildcards are not supported.

Columns: You can filter by column name values and data types.

- Enter a text string to immediately filter the display to show only columns with matching values.
- Click next to a data type to show columns of that type.
- Text and data type filters are additive. Both filters are applied to the display.

NOTE: If a column is hidden in the Visible Columns panel, it cannot be surfaced using a filter. You must toggle its display first. See *Visible Columns Panel*.

NOTE: Previewed columns are always displayed.

Rows: Enter values to highlight the rows where the values are present. Only rows where the values are present are displayed.

To remove all row and column filters, click **Clear all filters**.

For more information on the toolbar, see *Transformer Toolbar*.

Visible Columns Panel

In the status bar of the Transformer page, click the Eye icon to review the list of visible and hidden columns.

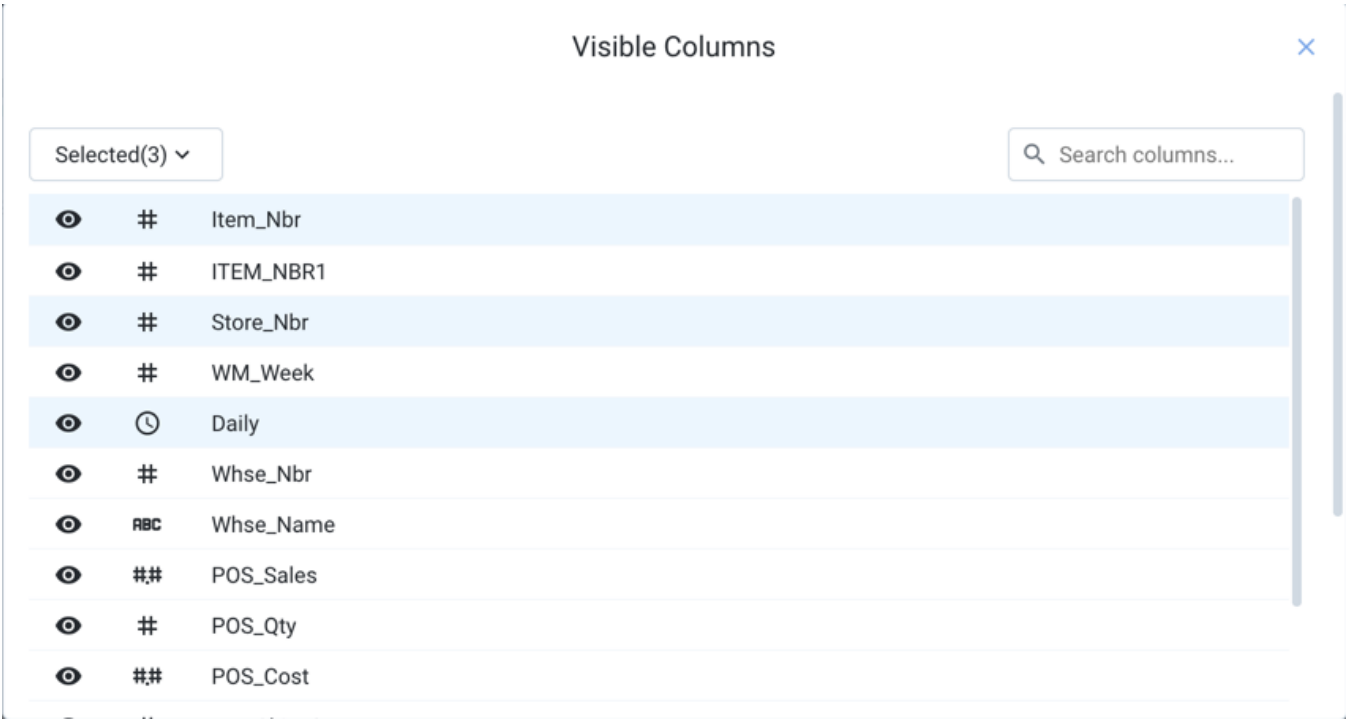


Figure: Visible Columns Panel

- Click the Eye icon next to a column name to toggle its visibility in the Transformer page.

NOTE: Columns that are not visible in the Transformer page are still generated in the output file. Before you run a job, you should review the Visible Columns dialog.

NOTE: Filters applied to the data grid or column browser are also applied in this panel. For more information, see *Filter Panel*.

- To toggle display of multiple columns at the same time, use CTRL or SHIFT to select columns. Then, click the Selected link and choose to show or hide them.
- Use the Search box to find matches for column names.
- To close the dialog, click the X icon.

Join Window

Contents:

- *Before You Begin*
 - *Step 1 - Select Dataset or Recipe*
 - *Step 2 - Select Join Conditions*
 - *Step 3 - Select Output Columns*
 - *Advanced options*
 - *Step 4 - Review Join*
-

In the Join window of the Trifacta® application, you can join your current dataset with another dataset or recipe based upon information that is common to both datasets.

For example, you could join together two sets of regional sales data based upon the product identifiers that they both use to track sales. In the Search panel, enter `join datasets` or select the Join icon from the toolbar.

- A **join** is a standard operation for merging the data from two different datasets. For more information, see *Join Types*.
- You cannot perform joins on columns of Object or Array data type.
- A join operation is different from a union operation. In a **union** operation, data from one or more datasets is appended to the current dataset, assuming that the columns are identical or very similar. For more information, see *Union Page*.

Tip: Depending on the types of operations you need to perform, you may need to perform joins earlier or later in your recipe. For more information, see *Optimize Job Processing*.

NOTE: Unnest, union, or join transforms may significantly increase the number of rows or columns in your dataset. To prevent overloading the browser's memory, the application may apply a limit function to the results to artificially limit the number of rows displayed in your sample. You can generate a new sample if desired. This limitation is not applied during the job execution.

Before You Begin

- **Review your record counts.** Before you specify the join, you should review your record counts and the uniqueness of your keys, which should provide an idea of the number of records you may see in the output. Note that the number of output records depends on the type of join and the matches between join keys.
- **Review your join key values.** If there are variations in the values in your join keys, you may end up with duplicate records in your joined dataset. Look for mismatched or missing values in your join keys, and correct if possible.
- **Review the granularity of your data.** If you bring together data at a lower fidelity than the source, you can end up with record matches that are not actually matching data. For example, if your timestamps are down-sampled from milliseconds to seconds as part of the join, you may have "matching" timestamps in seconds that were not matches at the millisecond level in the source data.

Step 1 - Select Dataset or Recipe

In the Search panel, enter `join datasets`. Then, select the dataset or recipe that you wish to join with your current dataset.

Choose dataset or recipe to join with POS-r01 – 2.txt

Search...

Recipes in current flow

Datasets in current flow

All datasets

Name	Last Updated	Source	Data
POS-r03.txt	Today at 11:05 AM	HDFS	
POS-r02.txt	Today at 11:05 AM	HDFS	
POS-r01.txt	Today at 11:05 AM	HDFS	
REF_CAL.txt	Today at 11:05 AM	HDFS	
REF_PROD.txt	Today at 11:05 AM	HDFS	

# ITEM_NBR	RBC	PRODUCT DESC
491000		ACME RICE CRACKERS CHEESE
474000		ACME RICE CRACKERS SESAME
498000		ACME RICE CRACKERS ONION
555000		ACME RICE CRACKERS BBQ
562000		ACME RICE CRACKERS ORIGINAL
352000		ACME RICE CRACKERS TERIYAKI
528000		ACME SODAS UNSALTED
528000		ACME SODAS SALTED

Browse current flow

Cancel

Accept

Figure: Select dataset or recipe to join

You can use the Data tab to preview the data in the selected object.

NOTE: You must have read access to the object to join it to your dataset.

- Use the Search bar to locate specific objects.
- Click **Accept**.

Step 2 - Select Join Conditions

In the next step, you specify the type of join and one or more join keys (columns).

Dataset samples

Join - Keys & Conditions

Join Key

Item_Nbr

ITEM_NBR

322k - 580k

322k - 580k

-	381000	381000
-	325000	325000
-	325000	325000
-	403000	403000
-	449000	449000
-	490000	490000
-	560000	560000
-	573000	573000
-	486000	486000
-	488000	488000
-	490000	490000
-	498000	498000
-	503000	503000
-	530000	530000
-	560000	560000
-	573000	573000
-	325000	325000
-	325000	325000

8,161 Rows in 165 Rows in 8,161 Rows in Output

Search row values...

Join type

required

Inner

Join keys

Add

Item_Nbr

ITEM_NBR

= (Equal to)

Suggested Q

99% match

Results summary

Based on current samples

Rows in Current

8161

Rows in Joined-in

165

Rows in Output

8161

Back

Next

Show only: ☒ Included Rows ☐ Excluded Rows

Figure: Specify join type and join keys

Dataset samples:

Mouse over the Dataset samples indicator to see the current samples from the datasets that are part of the join. For more information, see *Samples Panel*.

Join type:

From the drop-down, select the type of join to apply.

Join keys:

In the above image, the platform has determined that the item number (`Item_Nbr`) field of Region 1 data and the item number (`ITEM_NBR`) field from `REF_PROD` should be used as the keys for performing the join.

NOTE: By default, Trifacta displays a maximum of three rows of data for each join key value in your sample. So, when you specify your join, it may seem like there are joined values that are missing from the data grid panel. When the job is run across the entire dataset, however, the join generates the appropriate number of rows. For more information on changing the maximum number of rows that are previewed in the join, see *Miscellaneous Configuration*.

- To make changes to the two join keys, mouse over the specified keys:
 - To remove the two columns as join keys, click the X icon.
 - To edit the keys to use and other key options, click the Pencil icon. See below.
 - To add more join keys, click **Add**.

NOTE: Be careful applying multiple join keys. Depending on the join type, this type of join can greatly expand the size of the generated data.

Edit keys:

By default, matches between join keys are performed on a strict, case-sensitive matching between key values in the selected columns. In some cases, it may be useful to loosen the conditions under which matches are found. Depending on the type of join, you can specify a range of matching values for your join conditions. For more information, see *Configure Range Join*. The following options are applied to the join key columns in both sources to attempt to find matches. After the join is executed, no data in either column is changed based on these selections.

Option	Description
fuzzy match	<p>Use a fuzzy matching algorithm for key value matching.</p> <p>Tip: Use this option to perform fuzzy join matching of primary keys between datasets.</p> <p>NOTE: Fuzzy joins can only be applied to String data types. Other data types cannot be fuzzy-matched using the algorithm.</p> <p>Fuzzy matching uses the doublemetaphone algorithm for matching strings (keys). Both primary encodings of each key value must match. See <i>DOUBLEMETAPHONEEQUALS Function</i>.</p>
Ignore case	Ignore case differences between the join key values for matching purposes.
Ignore special	Ignore all characters that are not alphanumeric, accented Latin characters, or whitespace, prior to testing for a match.

characters	
Ignore whitespace	Ignore all whitespace characters, including spaces, tabs, carriage returns, and newlines.

Summary:

You can use these metrics to identify the likelihood of accurate matching between the join keys and the row count generated in the output.

Click **Next**.

Step 3 - Select Output Columns

From the selected datasets, you can specify the columns to include in the output.

The screenshot displays the 'Join - Output Columns' configuration window. The central 'Join Output Preview' section features histograms for six columns: Item_Nbr, ITEM_NBR1, Store_Nbr, WM_Week, Daily, and Whse_Nbr. Below these, a table lists 165 rows of data. The right-hand 'Columns' panel allows for selecting columns from a list of 30. The 'Current' tab is selected, showing 16 columns. The 'Advanced options' section is located at the bottom right of the column list.

Figure: Select output columns

Select columns:

Review the list of available columns, which are displayed for both sources.

- Use the search panel to search for specific columns.
- To include all columns:
 - Click the All, Current, or Join-In tabs.
 - Click the checkbox at the top of the list.

Advanced options

Name prefixes

You can apply prefixes to column names in the joined dataset, which can be helpful for tracking the source of a column in complex datasets. For example, you may wish to prepend each column from a dataset called, salesRegion01 with the prefix: sR01.

- **Name prefix for columns in Current data:** Enter a prefix to apply to the names of columns sourced from your current dataset that appear in the joined output.

- **Name prefix for columns in Joined_in data:** Enter a prefix to apply to the names of columns sourced from the joined-in dataset that appear in the joined output.

Dynamically updating Joins

After you have joined in another set of data, subsequent changes to that data can be automatically reflected in the output of the join:

- **Include all columns from Current data:** Dynamic updates always include the latest data from your current dataset.
- **Include all columns from Joined-In data:** Dynamic updates always include the latest data from the dataset that you are joining in.

NOTE: After you add your join to the recipe, if the data grid is empty, then the keys that you specified in the join may not have a match in the currently selected sample. You should revisit the keys used in your join. If the join still generates an empty grid on the current sample, you should collect a new sample. See *Samples Panel*.

Tip: If you must freeze the data in the dataset that you are joining in, you should create a copy of the dataset as a snapshot and join in the copy. See *Dataset Details Page*.

To join in the copy, edit the join and change the source that is being joined. See *Fix Dependency Issues*.

Click **Save and Continue**.

After you have selected your columns and any advanced settings, click **Review**.

Step 4 - Review Join

Review the join that you have specified. To modify any aspect of it, click the appropriate **Edit** link.

The screenshot displays the 'Join - Edit Step' interface. At the top, there's a search bar labeled 'Search row values.' Below it, a 'Join Output Preview' section shows a data grid with columns: Item_Nbr, ITEM_NBR1, Store_Nbr, WM_Week, Daily, and Whse_Nbr. The grid contains 165 rows of data. To the right of the grid, there's a 'Joined-in data' section showing 'REF_PROD.txt' with an 'Edit' link. Below that, the 'Join type' is set to 'Inner' with an 'Edit' link. The 'Join keys' section shows two keys: 'Item_Nbr' and 'ITEM_NBR' with an 'Equal to' relationship. The 'Output columns (30)' section shows '16 columns from Current' and '14 columns from Joined-in'. At the bottom right, there's a 'Save to Recipe' button.

Figure: Review join

To add the specified join to your recipe, click **Add to Recipe**.

Union Page

Contents:

- Mapping Schema
 - Custom column mappings
- Output Panel
- Updates

In the Union page, you can append data from one or more datasets to an existing dataset.

For example, if you have multiple datasets containing transactional data, such as log files, you can use the union operation to join daily or weekly slices of this data into a single dataset.

In a **union** operation, the Trifacta® application attempts to match columns between multiple datasets. As needed, you can perform manual tweaks to the matching and decide which columns to include or exclude in the resulting dataset.

- A union operation is different from a join operation. In a **join** operation, data from two datasets is brought together based on a defined primary key. The type of join determines the columns included in the output. For more information, see *Join Window*.

Tip: Depending on the types of operations you need to perform, you should perform your union steps earlier or later in the recipe. See *Optimize Job Processing*.

In the Search panel, enter `union` in the textbox.

Union

CancelAdd to Recipe

Match columnsAdd data

UNION DATA (2)

Union Output

4 Columns in Union

ABCKey2

ABCAntimal2

ABCVegetable2

ABCElement2

1 Dropped columnInclude all

+ABCMineral1

Dataset01 - 3

4 of 4 Columns in Union

ABCKey

ABCAntimal

ABCVegetable

ABCElement

Add column

No Dropped columns

Dataset02 - 3

4 of 5 Columns in Union

ABCKey

ABCAntimal

ABCVegetable

ABCElement

Add column

1 Dropped column

ABCMineral

Figure: Union Page

Dataset Actions:

- To add data from a dataset, recipe, or reference to the union, click **Add data**.
- Select one or more objects to add to the union and choose one of the following methods to match columns:
 - **Auto Align**. When this option is selected, Trifacta performs intelligent mapping of the columns of the new dataset(s) to the dataset already loaded in the Transformer page. Auto alignment uses the following to map:
 - Edit distance between column names
 - Column data types
 - Similarity between sampled data in the datasets

NOTE: Auto align is not available after you have selected the dataset to union. Auto align may add a few seconds to the union operation.

Add Datasets and Align by Name. Matches are made based on the name of each column. Partial matches might be identified as matches, as well.

- **Add Datasets and Align by Position.** Matches are made based on horizontal position of each column in each dataset. Extra columns will be dropped. This method might be useful if column names have changed between datasets.
- To remove data from the union, click the X next to its name in the right panel.
 - You cannot remove the original dataset from which the Union page was opened.

Mapping Schema

The schema of the output that is to be generated by the union operation is displayed in the left panel.

- The column names of the original dataset are used to populate the column names of the output dataset, where applicable.
- Each object that has been added to the union is displayed in the right panel.

Panel	Left Side	Right Side 1	Right Side 2
Upper	Output dataset - included cols.	Dataset 1 - included cols.	Dataset 2 - included cols.
Lower	Output dataset - excluded cols.	Dataset 1 - excluded cols.	Dataset 2 - excluded cols.

Custom column mappings

As needed, you can modify the default column mappings in your dataset. To remap a column, hover over the column entry in the right panel, Then, click the Plus icon:

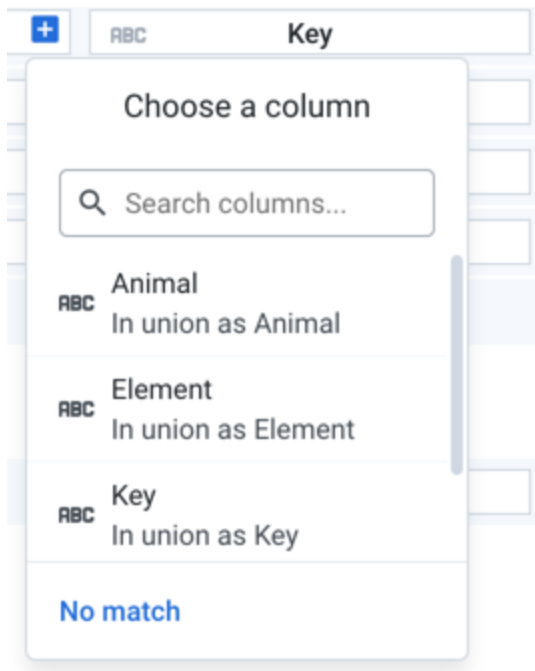


Figure: Custom Column Mapping

In the window, you can select the column in the current dataset that should appear in that location. Use this dialog to remap column order in each dataset.

- Click the Search columns field and begin typing to locate other columns.
- You can also specify that no match should be performed, which results in no data being imported from this column into the unioned dataset.

Tip: To map one of the dropped columns in your additional data to one of the source columns, hover over the empty No Match area next to the source column entry. Click the Plus icon to open the above mapping. Then, select the column from your additional data to slot into that location.

Output Panel

In the left panel, you can review and modify the columns to be included in and excluded from the output. By default, all matching columns are included in the output; if there are no initial matching columns, all columns from the original dataset are included in the output by default. You can see the columns that are sources for the union output column on the same line in the right panel.

- Each column entry contains a data type identifier for the source column. Data types may be re-inferred as part of the union. You can change the data type after the union is completed.
- To the right of the column name, you can see the number of datasets in the union where the column occurs.

Column Actions:

- To review the top five values for any column, click the Expand icon. You can see the count of each value across all included data.
- To remove a column from the union output, click the X icon to the left of the column entry in the upper panel.
- To add a column to the union output, click the + icon next to the left of the column entry in the lower panel.
- To include all available columns in the output, click **include all**.
- To add the union as specified, click **Add to Recipe**.

NOTE: Unnest, union, or join transforms may significantly increase the number of rows or columns in your dataset. To prevent overloading the browser's memory, the application may apply a limit function to the results to artificially limit the number of rows displayed in your sample. You can generate a new sample if desired. This limitation is not applied during the job execution.

Updates

To modify a union after it has been created, click the Edit icon for the entry in the Recipe panel. See *Recipe Panel*.

After you have added the union to your recipe, changes to the underlying data should automatically propagate to the dataset into which they have been unioned. No refreshing of the data is necessary.

However, it is possible that subsequent changes to your sources can cause problems in the output and downstream references. You can fix these dependency issues.

Tip: If you must freeze the data that you are adding in, you should create a copy of it as a snapshot and union in the copy. See *Dataset Details Page*.

To use the copy, edit the `union` transform in the copy and switch the data that is in use. See *Fix Dependency Issues*.

Run Job Page

Contents:

- *Running Environment*
 - *Options*
 - *Publishing Actions*
 - *Run Job*
 - *Automation*
 - *Run jobs via API*
-

In the Run Job page, you can specify transformation and profiling jobs for the currently loaded recipe. Available options include output formats and output destinations.

You can also configure the environment where the job is to be executed.

NOTE: If the job is executed in an environment other than Trifacta Photon, the job is queued for execution in the environment. During job execution, Trifacta observes the job in progress and reports progress as needed back into the application. Trifacta does not control the execution of the job.

Tip: Jobs can be scheduled for periodic execution through Flow View page. For more information, see *Add Schedule Dialog*.

Tip: Columns that have been hidden in the Transformer page still appear in the generated output. Before you run a job, you should verify that all currently hidden columns are ok to include in the output.

Run Job

Running Environment

☒ Trifacta Photon
Run job on Trifacta Photon (best for small and medium-sized jobs, up to approximately 1 GB of data)

☐ Spark
Run job on Spark

Options

☒ Profile results and assess data quality rules
Generate a statistical profile of the published data and evaluate data quality rules

☒ Validate Schema
Analyse input schema and compare with the flow dataset schema. [Learn more.](#)

☒ Fail job if dataset schemas change
Fail the job if the input data schema differs from the schema of the flow dataset.

☐ Ignore recipe errors
Allow jobs to be run even if there are errors present in recipe steps

Publishing Actions

Actions

Location

Create CSV

Settings

no compression, single file, with headers, with quotes, with mismatches, with delimiter.

Add Action

SQL Scripts

Connection

SQL statement

Settings

Add Script

No SQL scripts yet.

[Add new SQL script](#)

Cancel

Run

Figure: Run Job Page

Running Environment

Select the environment where you wish to execute the job. Some of the following environments may not be available to you. These options appear only if there are multiple accessible running environments.

NOTE: Running a job executes the transformations on the entire dataset and saves the transformed data to the specified location. Depending on the size of the dataset and available processing resources, this process can take a while.

Tip: The application attempts to identify the best running environment for you. You should choose the default option, which factors in the available environments and the size of your dataset to identify the most efficient processing environment.

Photon: Executes the job in Photon, an embedded running environment hosted on the same server as the Trifacta®.

Spark: Executes the job using the Spark running environment.

Advanced Execution Options:

- If Spark job overrides have been enabled in your environment, you can apply overrides to the specified job. See *Spark Execution Properties Settings*.
- This setting must be enabled. For more information, see *Enable Spark Job Overrides*.

Spark (Databricks): Executes the job on the Databricks cluster with which the platform is integrated.

NOTE: Trifacta platform can integrate with AWS Databricks or Azure Databricks, but not both at the same time.

For more information, see *Configure for AWS Databricks*.

For more information, see *Configure for Azure Databricks*.

NOTE: Use of Databricks is not supported on Marketplace installs.

Options

Profile results: Optionally, you can enable this option to generate a visual profile of your job results.

When the profiling job finishes, details are available through the Job Details page, including links to download results.

- Disabling profiling of your output can improve the speed of overall job execution.
- See *Job Details Page*.

NOTE: Percentages for valid, missing, or mismatched column values may not add up to 100% due to rounding. This issue applies to the Photon running environment.

Validate Schema: When enabled, the schemas of the datasources for this job are checked for any changes since the last time that the datasets were loaded. Differences are reported in the Job Details page as a Schema validation stage.

Tip: A **schema** defines the column names, data types, and ordering in a dataset.

Fail job if dataset schemas change: When Validate Schema is enabled, you can set this flag to automatically fail the job if there are differences between the stored schemas for your datasets and the schemas that are detected when the job is launched.

NOTE: If you attempt to refresh the schema of a parameterized dataset based on a set of files, only the schema for the first file is checked for changes. If changes are detected, the other files are contain those changes as well. This can lead to changes being assumed or undetected in later files and potential data corruption in the flow.

Tip: This setting prevents data corruption for downstream consumers of your executed jobs.

Tip: The default for validate schema is set at the workspace level. In the Run Job page, these settings are overrides for individual jobs.

For more information, see *Overview of Schema Management*.

Ignore recipe errors: Optionally, you can choose to ignore errors in your recipes and proceed with the job execution.

NOTE: When this option is selected, the job may be completed with warning errors. For notification purposes, these jobs with errors are treated as successful jobs, although you may be notified that the job completed with warnings.

Details are available in the Job Details page. For more information, see *Job Details Page*.

Publishing Actions

You can add, remove, or edit the outputs that are generated from this job. For more information, see *Publishing Actions*.

Run Job

To execute the job as configured, click **Run**. The job is queued for execution. After a job has been queued, you can track its progress toward completion. See *Job Details Page*.

Automation

Run jobs via API

You can use the available REST APIs to execute jobs for known datasets. For more information, see *API Reference*.

Publishing Actions

Contents:

- *Add Publishing Action*
- *Variables*
- *Output Settings*

You can add, remove, or edit the outputs generated from this job. By default, a CSV output for your home directory on the selected datastore is included in the list of destinations, which can be removed if needed. You must include at least one output destination.

Columns:

- **Actions:** Lists the action and the format for the output.
- **Location:** The directory and filename or table information where the output is to be written.
- **Settings:** Identifies the output format and any compression, if applicable, for the publication.

Actions:

- To change format, location, and settings of an output, click the Edit icon.
- To delete an output, click the X icon.

Add Publishing Action

From the available datastores in the left column, select the target for your publication.

Publishing Action

Search... (/)

HDFS

hive

Choose a file or folder

... / queryResults /admin

Create Folder

Search...

NAME	SIZE	LAST UPDATED
.trifacta		Today at 10:26 AM
POS-schema.csv		Today at 10:26 AM

Create a new file [Parameterize destination](#)

POS-r01

Output Directory
/trifacta/queryResults/admin@trifacta.local

Data Storage Format
CSV

☒ Create new file every run
Create a new file with an incremental number appended to the name (e.g. POS-r01_2.csv)

☐ Append to this file every run
Create it if it doesn't exist.

☐ Replace this file every run
Create it if it doesn't exist.

[More options](#)

+ New Edit

Cancel Add

NOTE: Do not create separate publishing actions that apply to the same file or database table.

New/Edit: You can create new or modify existing connections. By default, the displayed connections support publishing. See *Create Connection Window*.

Steps:

1. **Select the publishing target.** Click an icon in the left column.
 - a. If Hive publishing is enabled, you must select or specify a database table to which to publish.

Depending on the running environment, results are generated in Avro or Parquet format. See below for details on specifying the action and the target table.

If you are publishing a wide dataset to Hive, you should generate results using Parquet.

For more information on how data is written to Hive, see *Hive Data Type Conversions*.

2. **Locate a publishing destination:** Do one of the following.

- a. **Explore:**

NOTE: The publishing location must already exist before you can publish to it. The publishing user must have write permissions to the location.

NOTE: If your HDFS environment is encrypted, the default output home directory for your user and the output directory where you choose to generate results must be in the same encryption zone. Otherwise, writing the job results fails with a `Publish Job Failed` error. For more information on your default output home directory, see *Storage Config Page*.

- i. To sort the listings in the current directory, click the carets next to any column name.
 - ii. For larger directories, browse using the paging controls.
 - iii. Use the breadcrumb trail to explore the target datastore. Navigate folders as needed.
- b. **Search:** Use the search bar to search for specific locations in the current folder only.
 - c. **Manual entry:** Click the Edit icon to manually edit or paste in a destination.
3. **Choose an existing file or folder:** When the location is found, select the file to overwrite or the folder into which to write the results.

NOTE: You must have write permissions to the folder or file that you select.

- a. To write to a new file, click **Create a new file**.

Create a new file: See below.

4. **Create Folder:** Depending on the storage destination, you can click it to create a new folder for the job inside the currently selected one. Do not include spaces in your folder name.

5. **Create a new file:** Enter the filename under which to save the dataset.

- a. Select the Data Storage Format.
- b. For more information, see Output Settings below.

6. As needed, you can parameterize the outputs that you are creating. Click **Parameterize destination** in the right panel. See Parameterize destination settings below.
7. To save the publishing destination, click **Add**.

To update a publishing action, hover over its entry. Then, click **Edit**.

To delete a publishing action, select **Delete** from its context menu.

Variables

If any variable parameters have been specified for the datasets or outputs of the flow, you can apply overrides to their default values. Click the listed default value and insert a new value. A variable can have an empty value.

NOTE: Override values applied to a job are not validated. Invalid overrides may cause your job to fail.

NOTE: Unless this output is a scheduled destination, variable overrides apply only to this job. Subsequent jobs use the default variable values, unless specified again. No data validation is performed on entries for override values.

Tip: At the flow level, you can specify overrides at the flow level. Override values are applied to parameters of all types that are a case-sensitive match. However, values that are specified at runtime override flow-level overrides. For more information, see *Manage Parameters Dialog*.

For more information on variables, see *Overview of Parameterization*.

Output Settings

Depending on the type of output that you are generating, you must specify additional settings to define location, format, and other settings.

- See *File Settings*.
- See *Relational Table Settings*.

File Settings

When you generate file-based results, you can configure the filename, storage format, compression, number of files, and the updating actions in the right-hand panel.

NOTE: By default, when scheduled or API jobs are executed, no validations are performed of any write settings objects for file-based outputs. Issues with these objects may cause failures during transformation or publishing stages of job execution. Jobs of these types should be tested through the Trifacta application first. A workspace administrator can disable the skipping of these validations.

Create a new file [Parameterize destination](#)

POS-2021

Output Directory

/2227

Data Storage Format

CSV

- ☒ **Create new file every run**
Create a new file with an incremental number appended to the name (e.g. POS-2021_2.csv)

Append to this file every run
Append action is not available when publishing as single file to tfs

Replace this file every run
Create it if it doesn't exist.

[Less options](#) ^

Output options

- ☒ Include headers as first row on creation
- ☒ Include quotes
- ☒ Include mismatched values

Delimiter

,

Multi-part options

- ☒ **Single File**
Result is a single file in the output location.

Multiple Files
Result is split across multiple files placed in a new folder in the output location.

Compression

None

[Cancel](#)

[Update](#)



Figure: Output File Settings

Configure the following settings.

1. **Create a new file:** Enter the filename to create. A filename extension is automatically added for you, so you should omit the extension from the filename.
 - a. File output paths can have a maximum length of 2048 characters.
2. **Output directory:** Read-only value for the current directory. To change it, navigate to the proper directory.

NOTE: During job execution, a canary file is written for each set of results to validate the path. For datasets with parameters, if the path includes folder-level parameterization, a separate folder is created for each parameterized path. During cleanup, only the the canary files and the original folder path are removed. The parameterized folders are not removed. This is a known issue.

3. **Data Storage Format:** Select the output format you want to generate for the job.
 - a. **Avro:** This open source format is used widely for data serialization and data exchange between systems.
 - b. **CSV and JSON:** These formats are supported for all types of imported datasets and all running environments.

NOTE: JSON-formatted files that are generated by Trifacta are rendered in JSON Lines format, which is a single line per-record variant of JSON. For more information, see <http://jsonlines.org>.

- c. **Parquet:** This format is a columnar storage format.
- d. **HYPER:** Choose HYPER to generate results that can be imported into Tableau.

If you have created a Tableau Server connection, you can write results to Tableau Server or publish them after they have been generated in Hyper format.

NOTE: If you encounter errors generating results in Hyper format, additional configuration may be required. See *Supported File Formats*.

- e. For more information, see *Supported File Formats*.
4. **Publishing action:** Select one of the following:

NOTE: If multiple jobs are attempting to publish to the same filename, a numeric suffix (**_N**) is added to the end of subsequent filenames (e.g. `filename_1.csv`).

NOTE: If a single user executes two jobs with the same output settings except for different methods (e.g. create vs. replace) on the same output destination, the generated results and potential error conditions are unpredictable. Please wait for the first job to complete execution before changing the configuration for the second job.

- a. **Create new file every run:** For each job run with the selected publishing destination, a new file is created with the same base name with the job number appended to it (e.g. `myOutput_2.csv`, `myOutput_3.csv`, and so on).
- b. **Append to this file every run:** For each job run with the selected publishing destination, the same file is appended, which means that the file grows until it is purged or trimmed.

NOTE: The `append` action is not supported when publishing to S3.

NOTE: When publishing single files to WASB, the `append` action is not supported.

NOTE: When appending data into a Hive table, the columns displayed in the Transformer page must match the order and data type of the columns in the Hive table.

NOTE: Compression of published files is not supported for an `append` action.

- c. **Replace this file every run:** For each job run with the selected publishing destination, the existing file is overwritten by the contents of the new results.

5. More Options:

- a. **Include headers as first row on creation:** For CSV outputs, you can choose to include the column headers as the first row in the output. For other formats, these headers are included automatically.

NOTE: Headers cannot be applied to compressed outputs.

- b. **Include quotes:** For CSV outputs, you can choose to include double quote marks around all values, including headers.
- c. **Include mismatched values:** For CSV outputs, you can choose to include any value that is mismatched for its column data type. When disabled, mismatched values are written as null values.
- d. **Delimiter:** For CSV outputs, you can enter the delimiter that is used to separate fields in the output. The default value is the global delimiter, which you can override on a per-job basis in this field.

Tip: If needed for your job, you can enter Unicode characters in the following format: `\uXXXX`.

NOTE: The Spark running environment does not support use of multi-character delimiters for CSV outputs. You can switch your job to a different running environment or use single-character delimiters. For more information on this issue, see <https://issues.apache.org/jira/browse/SPARK-24540>.

- e. **Single File:** Output is written to a single file. Default setting for smaller, file-based jobs.
 - f. **Multiple Files:** Output is written to multiple files. Default setting for larger file-based jobs.
6. **Compression:** For text-based outputs, compression can be applied to significantly reduce the size of the output. Select a preferred compression format for each format you want to compress.

NOTE: If you encounter errors generating results using Snappy, additional configuration may be required. See *Supported File Formats*.

7. To save the publishing action, click **Add**.

Parameterize Destination Settings

For file- or table-based publishing actions, you can parameterize elements of the output path. Whenever you execute a job, you can pass in parameter values through the Run Job page.

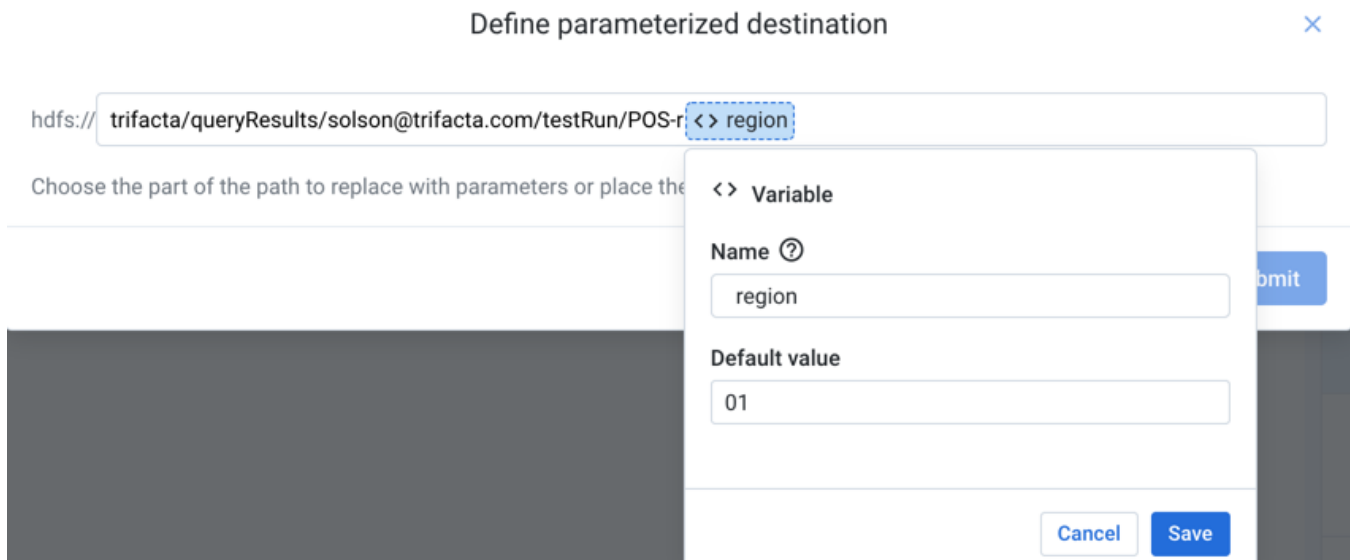
NOTE: Output parameters are independent of dataset parameters. However, two variables of different types with the same name should resolve to the same value.

NOTE: During job execution, a canary file is written for each set of results to validate the path. For datasets with parameters, if the path includes folder-level parameterization, a separate folder is created for each parameterized path. During cleanup, only the canary files and the original folder path are removed. The parameterized folders are not removed. This is a known issue.

Supported parameter types:

- Timestamp
- Variable

For more information, see *Overview of Parameterization*.



Define parameterized destination

hdfs://trifacta/queryResults/solson@trifacta.com/testRun/POS-r <> region

Choose the part of the path to replace with parameters or place the

<> Variable

Name ?

region

Default value

01

Cancel Save

Figure: Define destination parameter

Steps:

1. When you add or edit a publishing action, click **Parameterize destination** in the right panel.
2. On the listed output path, highlight the part that you wish to parameterize. Then, choose the type of parameter.
3. For Timestamp parameters:
 - a. Timestamp format: Specify the format for the timestamp value.
 - b. Timestamp value: You can choose to record the exact job start time or the time when the results are written relative to the job start time.
 - c. Timezone: To change the timezone recorded in the timestamp, click **Change**.
4. For Variable parameters:

- a. Name: Enter a display name for the variable.

NOTE: Variable names do not have to be unique. Two variables with the same name should resolve to the same value.

- b. Default value: Enter a default value for the parameter.

5. To save your output parameter, click **Save**.
6. You can create multiple output parameters for the same output.
7. To save all of your parameters for the output path, click **Submit**.
8. The parameter or parameters that you have created are displayed at the bottom of the screen. You can change the value for each parameter whenever you run the job.

Tip: At the flow level, you can specify overrides at the flow level. Override values are applied to parameters of all types that are a case-sensitive match. However, values that are specified at runtime override flow-level overrides. For more information, see *Manage Parameters Dialog*.

Relational Table Settings

For relational connections that support writing directly to the database, you can configure the following settings to specify the output table in the Run Job page.

NOTE: You cannot write to multiple relational outputs from the same job.

Steps:

1. **Select location:** Navigate the database browser to select the database and table to which to publish.
 - a. To create a new table, click **Create a new table**.
2. **Select table options:**
 - a. **Table name:**
 - i. **New table:** Enter a name for it. You may use a pre-existing table name, and schema checks are performed against it.
 - ii. **Existing table:** You cannot modify the name.
 - b. **Output database:** To change the database to which you are publishing, click the database icon in the sidebar. Select a different database.
 - c. **Publish actions:** Select one of the following.
 - i. **Create new table every run:** Each run generates a new table with a timestamp appended to the name.
 - ii. **Append to this table every run:** Each run adds any new results to the end of the table.
 - iii. **Truncate the table every run:** With each run, all data in the table is truncated and replaced with any new results.
 - iv. **Drop the table every run:** With each run, the table is dropped (deleted), and all data is deleted. A new table with the same name is created, and any new results are added to it.
3. To save the publishing action, click **Add** or **Update**.

SQL Scripts Panel

When specifying on-demand or scheduled outputs, you can define SQL scripts to execute before data ingestion, after output publication, or both. These scripts can be executed through any database connection to which you have write access.

NOTE: This feature may need to be enabled by a workspace administrator. For more information, see *Workspace Settings Page*.

When specifying your output, you can choose to add SQL scripts.

SQL Scripts			+ Add Script
Connection	SQL statement	Settings	
 teradata	INSERT INTO "dataprep"."joblog" ("TIMESTAMP","JOBTYPE","STATUS") VALUES ("210617-150422","orders","begin");	Run before data ingest	

Figure: SQL Scripts panel

Columns:

- **Connection:** The name of the connection where the script is to be executed.
- **SQL statement:** The first part of the SQL statement to be executed.
- **Settings:**
 - Run before data ingest: during job execution, script is to be run before data is ingested for job execution.
 - Run after data publish: during job execution, script is to be run after job has been executed and data is published.

NOTE: If publishing job fails, then all downstream tasks also fail, including the SQL script, which is not executed and is recorded as a failed phase of the job execution.

Actions:

- **Add Script:** To add a new SQL script for this output, click **Add Script**. See below.
- **Edit:** To modify the SQL script, highlight the entry and click **Edit**.
- **Delete:** To removal the SQL script, highlight the entry. Then, click **More menu > Delete**.

Add SQL Script Window

Enter your SQL statement in the window.

teradata

```
1 INSERT INTO "dataprep"."joblog" ("TIMESTAMP","JOBTYPE","STATUS") VALUES ("210617-150422","orders","begin");
```

Run before data ingest

Validate SQL Cancel Add

Figure: Add SQL Script window

Steps:

1. Select the connection through which to apply the SQL statement.
2. Enter your SQL statements in the window:

NOTE: Each line must end with a semi-colon (;). Validation fails if otherwise.

- a. You may enter multi-statement SQL scripts.
- b. SQL lines in an individual script are executed in the order listed in the script.
- c. Your SQL statements must comply with the expected syntax of the target system. For more information, see *Supported SQL Syntax*.
3. Choose when to run the SQL script:
 - a. Run before data ingest: SQL script is executed before the data is ingested for a job run.
 - b. Run after data publish: SQL script is executed after that data has been published from a job run.
4. To validate your SQL, click **Validate SQL**.
5. To add the SQL script, click **Add**.

If you have defined multiple scripts of the same type (before data ingest, for example), those scripts may be executed in parallel.

NOTE: The order of listing of scripts in the Trifacta application does not affect the order of execution of those scripts.

For more information on managing SQL scripts, see *Create Output SQL Scripts*.

Redshift Table Settings

If you are creating a publishing action for a Redshift database table in the Run Job page, you must provide the following information.

NOTE: Some Trifacta data types may be exported to Redshift using different data types. For more information, see *Redshift Data Type Conversions*.

Steps:

1. **Select location:** Navigate the Redshift browser to select the schema and table to which to publish.
 - a. To create a new table, click **Create a new table**.
2. **Select table options:**
 - a. **Table name:**
 - i. New table: enter a name for it. You may use a pre-existing table name, and schema checks are performed against it.
 - ii. Existing table: you cannot modify the name.
 - b. **Output database:** To change the database to which you are publishing, click the Redshift icon in the sidebar. Select a different database.
 - c. **Publish actions:** Select one of the following.
 - i. **Create new table every run:** Each run generates a new table with a timestamp appended to the name.
 - ii. **Append to this table every run:** Each run adds any new results to the end of the table.
 - iii. **Truncate the table every run:** With each run, all data in the table is truncated and replaced with any new results.
 - iv. **Drop the table every run:** With each run, the table is dropped (deleted), and all data is deleted. A new table with the same name is created, and any new results are added to it.
3. To save the publishing action, click **Add**.

Hive Table Settings

When publishing to Hive, please complete the following steps to configure the table and settings to apply to the publish action through the Run Job page.

NOTE: Some Trifacta data types may be exported to Hive using different data types. For more information on how types are exported to Hive, see *Hive Data Type Conversions*.

Steps:

1. **Select location:** Navigate the Hive browser to select the database and table to which to publish.
 - a. To create a new table, click **Create a new table**.
2. **Select table options:**
 - a. **Table name:**
 - i. New table: enter a name for it. You may use a pre-existing table name, and schema checks are performed against it.
 - ii. Existing table: you cannot modify the name.
 - b. **Output database:** To change the database to which you are publishing, click the Hive icon in the sidebar. Select a different database.

NOTE: You cannot publish to a Hive database that is empty. The database must contain at least one table.

- c. **Publish actions:** Select one of the following.

NOTE: If you are writing to unmanaged tables in Hive, create and drop & load actions are not supported.

- i. **Create new table every run:** Each run generates a new table with a timestamp appended to the name.
- ii. **Append to this table every run:** Each run adds any new results to the end of the table.

Tip: Optionally, users can be permitted to publish to Hive staging schemas to which they do not have full create and drop permissions. This feature must be enabled. For more information, see *Configure for Hive*.

When enabled, the name of the staging DB must be inserted into your user profile. See *User Profile Page*.

- iii. **Truncate the table every run:** With each run, all data in the table is truncated and replaced with any new results.
 - iv. **Drop the table every run:** With each run, the table is dropped (deleted), and all data is deleted. A new table with the same name is created, and any new results are added to it.
3. To save the publishing action, click **Add**.

Databricks Tables Table Settings

When you select a Databricks Tables database to store your job results in the Run Job page, you can configure the following options for the generated table.

NOTE: Access to Databricks Tables requires integration with Databricks, a Databricks Tables connection, and a Databricks personal access token.

- For more information, see *Configure for Azure Databricks*.
- For more information, see *Configure for AWS Databricks*.

Create a new table

Parameterize destination

POS_r01__2_txt

Output Database

default

☒ Use Delta table

☐ Publish as external table

☒ Create new table every run

Create a new table with a timestamp appended to the name (e.g. POS_r01__2_txt_20200421_113931)

Append to this table every run

Create it if it doesn't exist.

Truncate the table every run

Truncate existing data in the table and append new data.

Drop the table every run

Drop the table and create a new table of the same name.

Cancel

Add

Figure: Databricks Tables table settings

Steps:

1. **Select location:** Navigate the Databricks Tables browser to select the database and table to which to publish.
 - a. To create a new table, click **Create a new table**.
2. **Select table options:**
 - a. **Table name:**
 - i. New table: enter a name for it. You may use a pre-existing table name, and schema checks are performed against it.
 - ii. Existing table: you cannot modify the name.

NOTE: Writing to partitioned tables is not supported.

- b. **Output database:** To change the database to which you are publishing, click the Databricks icon in the sidebar. Select a different database.
 - c. **Optional table types:** Select one or more table types to publish as well:
 - i. **Use Delta table:** Output is stored as a Parquet-based Delta table.

NOTE: Versioning and rollback of Delta tables is not supported within the Trifacta platform. The latest version is always used. You must use external tools to manage versioning and rollback.

- ii. **Publish as external table:** Output is published as an external table to the specified location in your bucket.
 - d. **Publish actions:** Depending on your selection or selections above, the following publishing actions on the table are supported:
 - i. **Create new table every run:** Each run generates a new table with a timestamp appended to the name.
 - ii. **Append to this table every run:** Each run adds any new results to the end of the table.
 - iii. **Truncate the table every run:** With each run, all data in the table is truncated and replaced with any new results.

NOTE: Truncating the table is not supported for external tables.

- iv. **Drop the table every run:** With each run, the table is dropped (deleted), and all data is deleted. A new table with the same name is created, and any new results are added to it.

NOTE: Dropping the table is not supported for external tables.

3. To save the publishing action, click **Add**.

Tableau Server Datasource Settings

When publishing to Tableau Server through the Run Job page, please complete the following steps to configure the datasource and settings to apply to the publish action.

- A **datasource** is a table in your Tableau Server datastore that can be used as an input for your Tableau Server projects. For more information, see <https://onlinehelp.tableau.com/current/server/en-us/datasource.htm>.
 - For more information on creating a connection, see *Tableau Server Connections*.
 - For more information on how types are written to Tableau, see *Tableau Hyper Data Type Conversions*.
1. **Select location:** Navigate the Tableau Server browser to select the project and datasource to use for your publication.
 - a. For more information on projects, see <https://onlinehelp.tableau.com/current/server/en-us/projects.htm>.
 - b. To create a new datasource, click **Create a new datasource**.
 - i. For more information, see <https://onlinehelp.tableau.com/current/server/en-us/datasource.htm>.
 2. **Datasource options:**
 - a. **Datasource name:**
 - i. New datasource: enter a datasource for it. You may use a pre-existing datasource name.
 - ii. Existing datasource: you cannot modify the name.
 - b. **Output project:** To change the project to which you are publishing, click the Tableau icon in the sidebar. Select a different project.
 - c. **Publish actions:** Select one of the following.
 - i. **Create new datasource every run:** Each run generates a new datasource with a timestamp appended to the name.
 - ii. **Append to this datasource every run:** Each run adds any new results to the end of the datasource.
 - iii. **Drop the datasource every run:** With each run, the datasource is dropped (deleted), and all data is deleted. A new datasource with the same name is created, and any new results are added to it.
 3. To save the publishing action, click **Add**.

Tip: If you generate a Tableau format file as part of your output, you can choose to download and later publish it to Tableau Server. For more information, see *Publishing Dialog*.

Spark Execution Properties Settings

When you specify a job in the Run Job page, you may pass to the Spark running environment a set of Spark property values to apply to the execution of the job. These property values override the global Spark settings for your deployment.

NOTE: A workspace administrator must enable Spark job overrides and configure the set of available parameters. For more information, see *Enable Spark Job Overrides*.

Spark overrides are applied to individual output objects.

- You can specify overrides for ad-hoc jobs through the Run Job page.
- You can specify overrides when you configure a scheduled job execution.

User-specific Spark overrides: If you have enabled user-specific overrides for Spark jobs, those settings take precedence over the settings that are applied through this feature. For more information, see *Configure User-Specific Props for Cluster Jobs*.

In the Run Job page, click the Advanced Execution Settings caret.

Run Job

Running Environment

Photon
Run job on Trifacta Photon (best for small and medium-sized jobs, up to approximately 1 GB of data)

✓ Spark
Run job on Spark

Advanced environment options ^

Spark Execution Properties ?

Spark Driver Memory

e.g.: 8G

Spark Executor Memory

e.g.: 8G

Spark Executor Cores

e.g.: 4

Transformer Dataframe Checkpoint Threshold

e.g.: 500

Figure: Spark Execution Properties

Default Spark overrides:

The first four properties are available for all Spark job overrides:

Before you modify these parameters, you should review with your cluster administrator what are appropriate settings for each parameter. In some cases, you can set these values to cause failures on the cluster. No validation is performed for inputted values.

Spark parameter	Description
Spark Driver Memory	<p>Amount of physical RAM in GB on each Spark node that is made available for the Spark drivers.</p> <p>By raising this number:</p> <ul style="list-style-type: none">• The drivers for your job are allocated more memory on each Spark node.• There is less memory available for other uses on the node.
Spark Executor Memory	<p>Amount of physical RAM in GB on each Spark node that is made available for the Spark executors.</p> <p>By raising this number:</p> <ul style="list-style-type: none">• The Spark executors for your job are allocated more memory.• There is less memory available for other uses on the node.
Spark Executor Cores	<p>Number of cores on each Spark executor that is made available to Spark.</p> <p>By raising this number:</p> <ul style="list-style-type: none">• The maximum number of cores available for your job is raised on each Spark executor.• There are fewer cores for other uses on the node.
Transformer Dataframe Checkpoint Threshold	<p>When checkpointing is enabled, the Spark DAG is checkpointed when the approximate number of expressions in this parameter has been added to the DAG. Checkpointing assists in managing the volume of work that is processed through Spark at one time; by checkpointing after a set of steps, Trifacta can reduce the chances of execution errors for your jobs.</p> <p>By raising this number:</p> <ul style="list-style-type: none">• You increase the upper limit of steps between checkpoints.• You may reduce processing time.• It may result in a higher number of job failures.

For more details on setting these parameters, see *Tune Cluster Performance*.

Other Spark overrides:

Your workspace administrator may have enabled other Spark properties to be overridden. These parameters appear at the bottom of the list.

Please check with your administrator for appropriate settings for these properties.

Microsoft SQL Data Warehouse Table Settings

When you select a Azure® Synapse Analytics (Formerly Microsoft® SQL DW)® connection to publish your job results, you can configure the following options for the generated table.

For more information on creating these connections, see *Microsoft SQL Data Warehouse Connections*.

Create a new table [Parameterize destination](#)

MyExternalTable

Output Schema

INFORMATION_SCHEMA

☒ Publish as external table ?

Location required

/ testing_New_Folder

Browse

☒ **Create new table every run**
Create a new table with a timestamp appended to the name (e.g. MyExternalTable_20220420_165501)

☐ **Append to this table every run**
This action is not available for external tables

☐ **Truncate the table every run**
This action is not available for external tables

☐ **Drop the table every run**

Cancel

Update

Figure: Microsoft SQL Data Warehouse table settings

Steps:

1. **Select location:** Navigate the Azure Synapse Analytics (Formerly Microsoft SQL DW) browser to select the schema to which to publish.
 - a. To create a new table, click **Create a new table**.
2. **Select table options:**
 - a. **Table name:**

- i. **New table:** enter a name for it. You may use a pre-existing table name, and schema checks are performed against it.
 - ii. **Existing table:** you cannot modify the name.
- b. **Output schema:** To change the schema to which you are publishing, click the connection icon in the sidebar. Select a different schema.
- c. **Optional table types:** The following options may be available:
 - i. **Publish as external table:** Output is published as an external table to the specified location in your storage layer. Otherwise, the table is published as a managed table.

NOTE: When publishing to an external table, the output file type is Parquet.

- d. **Publish actions:** Depending on your selection or selections above, the following publishing actions on the table are supported. For more information, see "Publish Actions" below.
 - i. **Create new table every run:** Each run generates a new table with a timestamp appended to the name.
 - ii. **Append to this table every run:** Each run adds any new results to the end of the table.

NOTE: Appending the table is not supported for external tables.

- iii. **Truncate the table every run:** With each run, all data in the table is truncated and replaced with any new results.

NOTE: Truncating the table is not supported for external tables.

- iv. **Drop the table every run:** With each run, the table is dropped (deleted), and all data is deleted. A new table with the same name is created, and any new results are added to it.

NOTE: If the target is an external table, you can only drop the table when you first re-run a job to the target, after which you can choose to recreate the target as a managed or external table.

- 3. To save the publishing action, click **Add**.

Preferences Page

Contents:

- *User Preferences*
 - *Profile*
 - *Account*
 - *Email notifications*
 - *Sessions*
 - *Storage*
 - *Databricks settings*
-

From the Preferences page, you can manage aspects of your user account and other settings. Select **Help menu** > **Preferences**.

User Preferences

Profile

Review and manage your user profile. For more information, see *User Profile Page*.

Account

Change your password and set your locale among other preferences. See *Account Settings Page*.

Email notifications

Configure your personal preferences on receiving email notifications about activities in the product.

NOTE: This feature requires access to an SMTP server to send emails. For more information, see *Enable SMTP Email Server Integration*.

NOTE: This feature may need to be enabled in your workspace. See *Workspace Settings Page*.

For more information, see *Email Notifications Page*.

Sessions

Review the list of devices that have logged into your account. You can revoke any unrecognized and unauthorized sessions. For more information, see *Sessions Page*.

Storage

Review and modify the locations where you store data.

See *Storage Config Page*.

Databricks settings

Each user must save a Databricks Personal Access Token to their user account. See *Databricks Settings Page*.

User Profile Page

In your user profile, you can review your personal information and update your photo. Select **User menu > Preferences**.

NOTE: After saving changes to your user profile and exiting, please refresh the page.

Profile

Photo



Upload photo

Email

Name

Figure: User Profile Page

Profile settings

Name: Display name for your Trifacta® account.

Email: Email address associated with your account

NOTE: This value is the user ID. It must be a valid email address and cannot be modified after registration.

Upload photo

NOTE: This feature may need to be enabled. See *Miscellaneous Configuration*.

You can upload a preferred image associated with your user account. This image appears wherever the application contains a personal identifier, such as the icon for the User menu.

Image requirements:

- Format: JPG (JPEG), PNG, GIF, SVG, BMP, WEBP
- Dimensions: square dimensions work best. If you are using a non-square image, you should center the image details along the shorter edge of the image.

Steps:

1. Below the icon at the top of the User Profile page, click **Upload photo**.
2. Navigate your local desktop.
3. Select the file and click **Open**.
4. The icon is replaced by the image from the file you uploaded.

Storage Config Page

The Storage Config page allows you to configure storage locations for where you upload datasets and generate results.

Tip: When editing a directory location, click the Pencil icon. You can paste URLs to storage locations into the textbox.

The following options are available for configuring your storage environment.

NOTE: If you cannot modify the values in this page, you may need to enable the feature to modify user paths. For more information, see *Workspace Settings Page*.

Output home directory

Relative path to the directory where your results are stored by default and where your samples are stored. Click **Edit** to modify the home directory where results are stored.

Full path concatenates Output Protocol/Host value and this value.

Do not modify this value unless directed to do so. This path is not validated. If you specify a path to a directory to which you do not have appropriate permissions, all job exports will fail.

NOTE: Multiple users cannot share the same home directory.

NOTE: If your HDFS environment is encrypted, any location that you specify to write results for a job must be in the same encryption zone as this directory. For more information, please contact your HDFS administrator.

Upload directory

Relative path to the directory where your uploads are stored. To modify this value, click **Edit**.

NOTE: This setting only applies if Trifacta is connected to a backend datastore.

NOTE: You cannot upload to locations to which you do not have write access.

AWS credentials

If per-user mode is enabled, this option allows workspace members to apply individual key-secret values and roles to their accounts and modify other personal storage settings. For more information, see *Configure Your Access to S3*.

Account Settings Page

Contents:

- *Change Password*
 - *Locale*
 - *Other Settings*
-

In your Account Settings page, you can change your locale and modify other settings related to your account.

NOTE: After saving changes to your account settings and exiting, please refresh the page.

Change Password

NOTE: Users can reset their own passwords if a workspace administrator has enabled self service password reset. See *Workspace Settings Page*.

NOTE: If Single Sign-On (SSO) has been enabled, then these options are not available.

Tip: You can modify your password directly through the `/change-password` URL.

Old password: To change your password, enter your current password here.

New password: To change your password, enter a new password here.

To save the password, click **Update password**.

Tip: Forgot your password? Click **Reset password via email** to send a reset password to the email address registered for your account.

Locale

Select the locale to use when validating data types in the application.

NOTE: After saving changes to your locale, refresh your page. Subsequent executions of the data inference service use the new locale settings.

NOTE: When locale is changed, data type validation is affected only on subsequent executions of data type inference. If you are using structured datasets, such as schematized files, data types may be attached to the datasets that you have already imported. These data types are not affected.

For more information, see *Locale Settings*.

Other Settings

Enable keyboard shortcuts: When enabled, you can use keyboard shortcuts in the workspace or Transformer page.

Tip: When keyboard shortcuts are enabled, press ? in the application to see the available shortcuts.

Share usage data to improve product intelligence: When collaborative suggestions are enabled, anonymized data on how you use the product is aggregated with other workspace users' data to improve the suggestions provided by the product to all workspace users. You can use this setting to opt-out of sharing your data.

NOTE: This data is not shared with Alteryx or other users.

NOTE: If this setting is not present, the feature is disabled in your workspace. For more information, see *Workspace Settings Page*.

Email Notifications Page

Contents:

- *Receive emails about job activity*
 - *Receive emails about plan runs*
 - *Receive emails when a flow or plan is shared with you*
 - *Delivery address*
 - *Timezone*
-

You can configure the notifications that are sent to your email address based on job success or failure. Notifications can be sent for jobs executed from flows where you are the owner or a collaborator.

NOTE: Email notifications requires that you configure the Trifacta platform to use an available SMTP server. For more information, see *Enable SMTP Email Server Integration*.

NOTE: Email notifications may need to be enabled in your environment. You can also configure the type of jobs that generate success or failure emails. For more information, see *Workspace Settings Page*.

Tip: If visual profiling has been enabled, a PDF version of the profile of the job results is included as an attachment to the email.

Receive emails about job activity

When enabled, you can receive email notifications about job activity. You can receive emails from:

- flows where you are an owner or collaborator
- flows where someone has added you as a watcher

Email notifications are configured on a per-flow basis. For more information, see *Manage Flow Notifications Dialog*.

Receive emails about plan runs

When enabled, you can receive email notifications about plan runs if you are an owner or collaborator.

Email notifications are configured on a per-plan basis. For more information, see *Manage Plan Notifications Dialog*.

Receive emails when a flow or plan is shared with you

When enabled, you automatically receive notifications to your registered email address when a flow or plan is shared with you.

Delivery address

If needed, you can send notification emails to a different email address.

Uses:

- Your login email address cannot receive email.
- You wish to deliver email to an alias within your enterprise.
- You wish to deliver your email to a third-party application, such as Slack.

Tip: You can configure the sender email address and sender name that is used for all emails generated by the Trifacta node. For more information, see *Enable SMTP Email Server Integration*.

Timezone

Select the timezone in which email dates are displayed.

Sessions Page

The Sessions page enables you to view the number of devices signed into your account. You can use this information to make sure no one else has signed in to your account. You can review the devices that are authorized and revoke any unfamiliar devices.

Sessions

These are the devices that have logged into your account. You can revoke any sessions that you don't recognize.


Device	IP address	Last activity	First log in	
Chrome on Mac	76. [REDACTED]	Current session	Nov 15, 2021, 9:14 AM	
Chrome on Mac	76. [REDACTED]	Nov 15th 2021, 9:10 AM	Nov 15, 2021, 9:10 AM	

Figure: Sessions page

Columns:

- **Device:** Name of the device that is connected to your account.
- **IP address:** Unique address that identifies the device on the internet or a local network.
- **Last activity :** Displays the recent activity you performed on the device.
- **First log in:** Timestamp details at which the first session was logged.

Removing unrecognized devices:

Click the Delete icon next to any unrecognized device.

Access Tokens Page

From the Access Token page, you can generate and manage access tokens that apply to your account. Access tokens can be used when accessing the REST APIs.

Tip: Access token usage is the preferred method of authenticating from API. See <https://api.trifacta.com/ee/es.t/index.html#section/Authentication>

NOTE: If this page is not visible, the feature has not been enabled in your instance of the platform. See *Enable API Access Tokens*.

NOTE: Workspace administrators can choose to enable creation and use of access tokens to individual workspace users. For more information, see *Workspace Settings Page*.

NOTE: Workspace administrators can delete the access tokens of other users.

Access Tokens

Generate New Token

Token ID	Description	Status	Created	Expires	Last used
0bc1d49f-5475-4c62-a0ba-6ad269389ada	new token	Active	2019-01-15 12:58:28	2020-01-15 12:58:28	Never Used

Figure: Access Tokens Page

Actions:

- **Generate New Token:** Click to generate a new access token for your user account. See below.

Columns:

- **Token ID:** Internal identifier for the token

NOTE: This is not the token itself. That value is exposed when you create the token and must be retained for any use outside of the Trifacta application.

- **Description:** User-provided description of the token
- **Status:** Current status of the token:
 - Active - Token is active and can be used for access.
 - Expired - Token has expired after its expiration timestamp has been reached.
- **Created:** Timestamp for when the token was created.
- **Expires:** Timestamp for when the token expires.
- **Last Used:** Timestamp for when the token was last used.

Context menu:

- **Delete Token:** Click the delete the token.

Deleting a token cannot be undone.

NOTE: If you delete an active token, any API usage that references the token no longer works.

Generate Token

When you generate a token, you can provide the following information.

Generate Token

Lifetime (days)

365

Set lifetime to -1 to have the token never expire.

Description (optional)

Cancel

Generate

Figure: Generate Token Dialog

- **Lifetime:** Enter the number of days that you would like to use this token without renewal.
 - If the token expires, a new one must be created. You can generate a new token at any time.
 - You can generate any number of tokens.

Tip: Depending on your environment, you may be able to set this value to -1 to never expire the token.

- **Description:** (Optional) you can provide a user-friendly description of the purpose of the token. This value is for display purposes only.

To create the token, click **Generate**.

After the token is generated, it is automatically activated. You can have multiple active tokens.

You must copy the token out of the application. Click **Copy Token to clipboard** to copy the text value of the token.

For security purposes, after you close the My Token screen, the token is no longer accessible. You cannot reopen this dialog. You must copy this value and store it in a secure place for later use.

Databricks Settings Page

To access a Databricks cluster for running jobs, each user of the Trifacta® platform must insert their personal access token into their profile. This configuration enables the user to authenticate to a connected Databricks cluster.

NOTE: Each user must insert a personal access token into their profile. Users that do not provide a personal authentication token cannot run jobs on Databricks, including transformation, sampling, and profiling jobs.

NOTE: When you reset your personal access token (PAT), a new cluster is created if your new token does not have access to your current cluster. If you are resetting an expired personal access token, no new cluster is created. This new cluster is created when you first request access to the Databricks cluster. When you next use an interface that require access to the cluster, such as the relational browser, it may take some time to load.

Prerequisites

- Acquire your Databricks personal access token.

NOTE: Your Databricks personal access token must be acquired from the same region as your Databricks deployment. This region name is available through the Trifacta application.

For more information, see <https://docs.azuredatabricks.net/api/latest/authentication.html#requirements>.

Steps

1. Login to the application. From the menu bar, select **User menu > Preferences > Databricks**.
2. **Configure URL:**
 - a. **For Azure developments:** Acquire the Azure Databricks personal access token from the same region as your Azure Databricks deployment. The region name is available through the Trifacta application. For more information, see *Configure for Azure Databricks*.
 - b. **For AWS developments:** Edit the workspace URL, as required and click **Save**.
 - The existing property `databricks.serviceUrl` is used to configure the URL to the Databricks Service to run Spark jobs. For more information, see *Configure for AWS Databricks*.
 - The `databricks.serviceUrl` defines the default Databricks workspace for all user in the Trifacta workspace.
 - You can override the default settings in this page.
3. **Personal access token:** In the Personal Access Token field, paste your token.
 - a. To use a different token, click **Change**.
4. **Databricks table cluster name:** Each user can specify the name of a cluster to use to browse a Databricks Tables deployment.

NOTE: This cluster must be created and maintained by your Databricks administrator. This cluster can be shared among multiple users.

5. **Databricks policy name:** To select the cluster policy to use when you are executing jobs on the cluster, click **Edit**. The available policies are listed in the drop-down.

NOTE: Cluster policies determine characteristics of Databricks clusters that are used or created for job execution. This feature requires additional configuration.

- a. For more information, see *Configure for AWS Databricks*.
 - b. For more information, see *Configure for Azure Databricks*.
6. Click **Save**.

Schedules Page


In the Schedules page, administrators can be review the current set of schedules.

NOTE: This page is available to project owners and workspace administrators only.

A **schedule** is automated execution of an output in a flow on a regular basis. Schedules are composed of the following:

- A schedule
- A set of one or more scheduled destinations
- These objects are created from the flow. For more information, see *Flow View Page*.
 - See *Add Schedule Dialog*.

NOTE: Schedules owned by users that have been disabled continue to execute. An admin can disable them through the Schedules page.

 Schedules

Owner	Frequency	Flow	State	Last Updated ▾
Administrator (you)	Weekly: 12:00 AM - On Sunday	census data	Enabled	Today at 11:16 AM
Administrator (you)	Cron: 0 0 ** TUE * - At 12:00 AM, only on...	2013 POS	Enabled	Today at 11:14 AM

Figure: Schedules page

Columns:

- **Owner:** The user that owns the schedule.
- **Frequency:** The frequency of occurrence of the schedule.
 - **Cron:** Cron jobs utilize a modified form of cron scheduling syntax to define execution time. For more information, see *cron Schedule Syntax Reference*.
 - **Weekly/Monthly/Daily:** You can also schedule jobs to execute according to a regular calendar period.
- **Flow:** Name of the flow to which the schedule applies.
 - If available, you can click the link to open the flow.
- **State:** The current state of the schedule:
 - **Enabled** - The schedule is active and will execute at the next occurrence according to the frequency.
 - **Disabled** - The schedule is inactive and cannot be executed until it is enabled.
- **Last Updated:** Timestamp for the when the schedule was last modified.

Actions:

- **Filter by update:** Click the caret next to Last Updated to sort the list of schedules.
- **Search:** Enter text in the search field to filter the listed jobs by flow name.

Context menu:

Next to the schedule listing, click the options menu to see the following:

- **Enable/Disable Schedule:** Select this option to toggle availability of the schedule.
- **Delete Schedule:** Delete the schedule.

Deleting a schedule is permanent and cannot be undone.

UI Index

Use the links below to access the reference pages for the Trifacta application user interface. For more information on UI pages that apply to administrators, see *Admin Reference*.

Home

Item	Description
<i>Home Page</i>	From the Home page, you can create or access your flows, datasets, and jobs, as well as configure settings and find additional resources.
<i>Flows Page</i>	The Flows page displays the flows to which you have access and lets you create, review, and manage them. A flow is an object for bringing together and organizing the datasets, recipes, and other objects that you use to generate your results.
<i>Plans Page</i>	The Plans page lets you create, review, and manage your plans. A plan is a sequence of tasks and the triggers that execute them. Plans can be applied across multiple flows in your workspace.
<i>Library Page</i>	In the Library page, you can review your imported and reference datasets and any macros that you may have created. You can also import new data from this page.
<i>Connections Page</i>	Through the Connections page, you can add new connections or modify the connections that you have already created. From the left nav bar, click the Connections icon.
<i>Jobs Page</i>	In the Jobs page, you can track the status of all of your jobs and plan runs.
<i>Transformer Page</i>	In the Transformer page, you identify the data that you need to transform and build your transformation recipes on samples taken from your currently selected dataset.
<i>Preferences Page</i>	From the Preferences page, you can manage aspects of your user account and other settings. Select Help menu > Preferences .
<i>Schedules Page</i>	In the Schedules page, administrators can review the current set of schedules.

Item	Description
<i>Download Logs Dialog</i>	You can download logs for your current session in Trifacta®. From the Help menu, select Download logs .

Flows

Flows Page

Item	Description
<i>Create Flow Page</i>	You can use flows to organize your datasets and to track the jobs associated with them.
<i>Flow View Page</i>	In Flow View, you can access and manage the objects that you have added to or created in the selected flow. You can perform a variety of actions to effectively manage flow development and job execution through a single page in the Trifacta® application.

Flow View

Flow View Page

Item	Description
<i>View for Connections</i>	In Flow View, you can review the details of the connections used to access the flow's imported datasets, whether you created it or it was shared with you.

<i>View for Imported Datasets</i>	When you select an imported dataset in Flow View, you can review its details in the context panel and select options from its context menu.
<i>View for Dataset with Parameters</i>	When you select a dataset with parameters in Flow View, additional options are available in its context menu and the Details panel.
<i>View for Recipes</i>	For each recipe in Flow View, you can review or edit its steps or create new recipes altogether. You can also create references to the recipe, modify outputs, and create new recipes off of the recipe.
<i>View for Reference Datasets</i>	A reference dataset is a reference to a recipe's output, which can be added to a flow other than the one where the recipe is located.
<i>View for Unstructured Datasets</i>	An unstructured dataset is an imported dataset that does not contain any initial parsing steps. All parsing steps must be added through recipes that are applied to the dataset. During the import process, you disable the initial steps that are applied to imported datasets. Instead, these steps are added as the first steps of the auto-generated recipe that appears with the dataset in Flow View.
<i>View for Outputs</i>	Associated with each recipe is one or more outputs. These publishing destinations can be configured through the context panel in Flow View. Through outputs, you can execute and track jobs for the related recipe.
<i>Share Flow Dialog</i>	You can manage access to a flow for other users through the Share Flow dialog. In Flow View, select Share from the context menu.
<i>Change Dataset Dialog</i>	Through the Flow View page, you can change the source that is used for your dataset. In this manner, you can apply the same recipe across datasets with the same schema. When the source dataset has been changed, a new sample is automatically generated for you.
<i>Add Schedule Dialog</i>	To add a schedule to your flow, click the drop-down menu in the Flow View page, and select Schedule .
<i>Manage Flow Notifications Dialog</i>	When email notifications are enabled, flow owners and collaborators can configure the delivery of emails to interested stakeholders based on the success or failure of jobs executed within this flow. From the flow menu, select Email notifications .
<i>Manage Parameters Dialog</i>	Within a flow, you can create and manage flow parameters, including specifying override values. From the flow menu, select Parameters .
<i>Flow Search Panel</i>	You can search for objects within your flow. In Flow View, click the Magnifying Glass icon at the top of the page.
<i>Flow Optimization Settings Dialog</i>	In the Flow Optimization Settings dialog, you can configure the following settings, which provide finer-grained control and performance tuning over your flow and its job executions. From the Flow View menu, select Optimization settings .

Plans

Item	Description
<i>Plan View Page</i>	In Plan View, you design your plan, which includes the building and sequencing of tasks and the triggers that execute your sequence of tasks.

Library

Library Page

Item	Description
<i>Dataset Details Page</i>	Use the Dataset Details page to review a dataset's usage and to perform management tasks on it.
<i>Import Data Page</i>	Through the Import Data page, you can upload datasets or select datasets from sources that are stored on connected datastores. From the Library page, click Import Data .
<i>Macros Page</i>	In the Macros page, you can review and manage the macros to which you have access.

Import Data Page

Item	Description
<i>Create Connection Window</i>	Through the Create Connection window, you can create and edit connections between Trifacta® and remote storage.
<i>Database Browser</i>	The database browser enables you to interact with databases that are connected to Trifacta®.
<i>File System Browser</i>	In Trifacta®, the file system browser lets you browse, select, and filter the sources that you can access through the datastore to which you are connected. You also use the browser to select targets for publishing job results.
<i>HDFS Browser</i>	The HDFS browser enables you to browse, select, and filter the files to which you have access in the Hadoop cluster to which Trifacta® is connected.
<i>S3 Browser</i>	In Trifacta®, the S3 browser lets you browse, select, and filter the sources that you can access through S3. You also use the browser to select targets for publishing job results.
<i>File Import Settings</i>	When you edit settings on a selected file in the Import Data page, the following settings are displayed.
<i>Table Import Settings</i>	When you edit settings for a selected table in the Import Data page, the following settings are displayed.

Macros Page

Item	Description
<i>Macro Details Page</i>	Through the Macro Details Page, you can review details about an individual macro. In the Macros page, click the name of the macro to review.

Jobs

Item	Description
<i>Job Details Page</i>	You can use the Job Details page to explore details about successful or failed jobs, including outputs, dependency graph, and other metadata. Download results to your local desktop or, if enabled, explore a visual profile of the data in the results for further iteration on your recipe.
<i>Plan Runs Page</i>	In the Plan Runs page, you can track the status of all runs of your plans.
<i>Sample Jobs Page</i>	In the Sample Jobs page, you can track the status of all sample jobs to which you have access.

Transformer

Transformer Page

Item	Description
------	-------------

<i>Data Grid Panel</i>	The data grid in the Transformer Page displays how your current recipe applies to the data in your currently selected sample.
<i>Recipe Navigator</i>	Through the Recipe Navigator, you can locate and open any recipe from the current flow in the Transformer page. To open the Recipe Navigator, select the drop-down next to the name of the current recipe in the Transformer page.
<i>Sample Indicator</i>	At the top of the Transformer page, you can see the Sample Indicator showing the name of the currently displayed sample. If you have not taken a sample, this value is Initial data . Click this button to show more information about the sample.
<i>Transformer Toolbar</i>	At the top of the data grid and the column browser, the Transformer toolbar provides quick access to common transformations.
<i>Column Menus</i>	You can use the menus available in a column context menu to perform a variety of actions on the column or columns.
<i>Column Browser Panel</i>	Through the Column Browser, you can use data quality bars and data type information to perform basic review of data across many columns. You can use these tools to select data of interest for display in the data grid or Column Details views or to prompt for suggestions of recipe steps.
<i>Column Details Panel</i>	In the Column Details panel, you can review additional details about a column of your dataset. Select Column Details from any column menu or the Action menu in the column browser.
<i>Transform Preview</i>	When you create or edit a transform, the data grid displays a preview of results of the transform. Transform previews assist in specifying and validating the transformation steps before they are applied.
<i>Context Panel</i>	On the right side of the Transformer page, the context panel displays one of multiple panels, depending on the current state or selection of the data grid.
<i>Filter Panel</i>	As needed, you can filter the rows and columns displayed in the data grid. To review and apply filters, click the Filter icon in the Transformer toolbar.
<i>Visible Columns Panel</i>	In the status bar of the Transformer page, click the Eye icon to review the list of visible and hidden columns.
<i>Join Window</i>	In the Join window of the Trifacta® application, you can join your current dataset with another dataset or recipe based upon information that is common to both datasets.
<i>Union Page</i>	In the Union page, you can append data from one or more datasets to an existing dataset.
<i>Run Job Page</i>	In the Run Job page, you can specify transformation and profiling jobs for the currently loaded recipe. Available options include output formats and output destinations.

Data Grid Panel

Item	Description
<i>Column Histograms</i>	The bar chart at the top of each column in the Transformer page, called a histogram , characterizes the data in that column. Each column histogram displays the count of each detected value in the column (for string data) or the count of values within a numeric range (for number data).
<i>Data Quality Bars</i>	Just below the column name in the data grid is a horizontal band, which identifies data quality issues among the sample values in the column.
<i>Lookup Wizard</i>	Through the Transformer page, you can perform lookups from one set of values in your dataset into another set of values in another dataset using a simple wizard. To perform a lookup, select the caret next to a column title, and then select Lookup....
<i>Standardize Page</i>	Through the Standardize page, you can review similar column values and standardize them to values that you specify.

Column Menus

Item	Description
	In the Datetime Format dialog, you can specify or locate the format that you'd like to use for validation of the values in

Choose Datetime Format Dialog	the current column against the Datetime data type.
Transformation by Example Page	In the Transform by Example page, you can build new columns of data by specifying values to map from the selected source column. For the column to transform, select Create column by examples from the column menu.

Context Panel

Item	Description
Recipe Panel	Through the Recipe panel, you can review and modify the steps of the recipe that you have already created and add new steps to your recipe at the current location. You can also flag a step that requires review without which jobs cannot be run.
Transform Builder	The Transform Builder enables you to search for transformations and to rapidly assemble complete transform steps through a simple menu-driven interface.
Search Panel	Through the search context panel, you can locate transformations to specify and add at the current location in your recipe.
Edit History Panel	Through the Edit History panel, you can review the sequence of edits to the current recipe by individual contributors. This panel assists in determining who made which changes and when they were made.
Samples Panel	For smaller datasets, the Transformer page displays the entire dataset. For larger ones, the source data is sampled for use in the Transformer page. Through the Samples panel, you can create new samples and select them for display in the Transformer page.
Selection Details Panel	In the Selection Details panel, you can review an active profile of your current selection or selections in the data grid or column browser and review patterns and suggestions for transformations.

Preferences

Preferences Page

Item	Description
User Profile Page	In your user profile, you can review your personal information and update your photo. Select User menu > Preferences .
Account Settings Page	In your Account Settings page, you can change your locale and modify other settings related to your account.
Email Notification s Page	You can configure the notifications that are sent to your email address based on job success or failure. Notifications can be sent for jobs executed from flows where you are the owner or a collaborator.
Sessions Page	The Sessions page enables you to view the number of devices signed into your account. You can use this information to make sure no one else has signed in to your account. You can review the devices that are authorized and revoke any unfamiliar devices.
Access Tokens Page	From the Access Token page, you can generate and manage access tokens that apply to your account. Access tokens can be used when accessing the REST APIs.
Databricks Settings Page	To access a Databricks cluster for running jobs, each user of the Trifacta® platform must insert their personal access token into their profile. This configuration enables the user to authenticate to a connected Databricks cluster.

User Profile Page

Item	Description
------	-------------

<i>Storage Config Page</i>	The Storage Config page allows you to configure storage locations for where you upload datasets and generate results.
----------------------------	---

Connections Page

Item	Description
<i>Share Connection Dialog</i>	Through the Share Connection dialog, users of the selected connection with appropriate privileges can modify who has access to the connection.

Supported File Formats

Contents:

- *Filenames*
 - *File path length limits*
 - *Forbidden characters in import filenames*
- *Native Input File Formats*
 - *Converted file formats*
- *Native Output File Formats*
- *Compression Algorithms*
 - *Read Native File Formats*
 - *Write Native File Formats*
 - *Snappy compression formats*
- *Additional Configuration for File Format Support*
 - *Publication of some formats requires execute permissions*

This section contains information on the file formats and compression schemes that are supported for input to and output of Trifacta®.

NOTE: To work with formats that are proprietary to a desktop application, such as Microsoft Excel, you do not need the supporting application installed on your desktop.

Filenames

NOTE: During import, the Trifacta application identifies file formats based on the extension of the filename. If no extension is provided, the Trifacta application assumes that the submitted file is a text file of some kind. Non-text file formats, such as Avro and Parquet, require filename extensions.

NOTE: Filenames that include special characters can cause problems during import or when publishing to a file-based datastore.

File path length limits

Maximum character limits for file paths:

- **File paths to sources for imported datasets:** 1024

Tip: This limit (`storagelocations`) applies to both files and tables.

- **File paths to output files:** 2048

Tip: This limit (`writesettings`) applies to files stored on any file-based storage location.

Forbidden characters in import filenames

The following list of characters present issues in the listed area of the product. If you encounter issues, the following listings may provide some guidance on where the issue occurred.

Tip: You should avoid using any of these characters in your import filenames. This list may not be complete for all available running environments.

- **General:**

```
" / "
```

- **Web browser:**

```
" \"
```

- **Excel filenames:**

```
"# ", "{ ", " }
```

- **Spark-based running environment:**

```
"{ ", " * ", " \"
```

Native Input File Formats

Trifacta® can read and import directly these file formats:

- CSV
- JSON v1, including nested

NOTE: JSON files can be read natively but often require additional work to properly structure into tabular format. Depending on how the Trifacta application is configured (v1 or v2), JSON files may require conversion before they are available for use in the application. See "Converted file formats" below.

NOTE: Trifacta requires that JSON files be submitted with one valid JSON object per line. Consistently malformed JSON objects or objects that overlap linebreaks might cause import to fail.

- Plain Text
- LOG
- TSV
- Parquet

NOTE: When working with datasets sourced from Parquet files, lineage information and the `$source_rownumber` reference are not supported.

- Avro

NOTE: When working with datasets sourced from Avro files, lineage information and the `$source` `rownumber` reference are not supported.

- Google Sheets

NOTE: Individual users must enable access to their Google Drive. No data other than Google Sheets is read from Google Drive.

- XML

Tip: XML files can be ingested as unstructured text.

Converted file formats

Files of the following type are not read into the product in their native format. Instead, these file types are converted using the Conversion Service into a file format that is natively supported, stored in the base storage layer, and then ingested for use in the product.

NOTE: Compressed files that require conversion of the underlying file format are not supported for use in the product.

Converted file formats:

- Excel (XLS/XLSX)

NOTE: Other Excel-related formats, such as XLSM format, are not supported.

Tip: You may import multiple worksheets from a single workbook at one time. See *Import Excel Data* in the User Guide.

- Google Sheets

Tip: You may import multiple sheets from a single Google Sheet at one time.

- PDF

NOTE: PDF support may need to be enabled in your environment. See *Import PDF Data*.

- PDF is supported for import only.
- See *Import PDF Data* in the User Guide.

- JSON v2

Notes on JSON:

There are two methods of ingesting JSON files for use in the product.

- JSON v2 - This newer version reads the JSON source file through the Conversion Service, which stores a restructured version of the data in tabular format on the base storage layer for quick and simple use within the application.

Tip: This method is enabled by default and is recommended. For more information, see *Working with JSON v2*.

- JSON v1 - This older version reads JSON files directly into the platform as text files. However, this method often requires additional work to restructure the data into tabular format. For more information, see *Working with JSON v1*.

Native Output File Formats

Trifacta can write to these file formats:

- CSV
- JSON
- Tableau Hyper

NOTE: Publication of results in Hyper format may require additional configuration. See below.

- Avro

NOTE: The Trifacta Photon and Spark running environments apply Snappy compression to this format.

- Parquet

NOTE: The Trifacta Photon and Spark running environments apply Snappy compression to this format.

Compression Algorithms

When a file is imported, the Trifacta application attempts to infer the compression algorithm in use based on the filename extension. For example, `.gz` files are assumed to be compressed with GZIP.

NOTE: Import of a compressed file whose underlying format requires conversion through the Conversion Service is not supported.

NOTE: Importing a compressed file with a high compression ratio can overload the available memory for the application. In such cases, you can decompress the file before uploading. If decompression fails, you should contact your administrator about increasing the Java Heap Size memory.

NOTE: Publication of results in Snappy format may require additional configuration. See below.

NOTE: GZIP files on Hadoop are not split across multiple nodes. As a result, jobs can crash when processing it through a single Hadoop task. This is a known issue with GZIP on Hadoop.

Where possible, limit the size of your GZIP files to 100 MB or less, or use BZIP2 as an alternative compression method. As a workaround, you can try to run the job on the unzipped file. You may also disable profiling for the job. See *Run Job Page* in the User Guide.

Tip: If preferred, you can configure the Trifacta application to infer the compression scheme based on first few bytes of the file. For more information, see *Miscellaneous Configuration*.

Read Native File Formats

	GZIP	BZIP	Snappy	Notes
CSV	Supported	Supported	Supported	
JSON v2	Not supported	Not supported	Not supported	A converted file format. See above.
JSON v1	Supported	Supported	Supported	Not a converted file format. See above.
Avro			Supported	
Hive			Supported	

Write Native File Formats

	GZIP	BZIP	Snappy
CSV	Supported	Supported	Supported
JSON	Supported	Supported	Supported
Avro			Supported; always on
Hive			Supported; always on

Snappy compression formats

Trifacta supports the following variants of Snappy compression format:

File extension	Format name	Notes
.sz	Framing2 format	See: https://github.com/google/snappy/blob/master/framing_format.txt
.snappy	Hadoop-snappy format	See: https://code.google.com/p/hadoop-snappy/ <div>NOTE: Xerial's snappy-java format, which is also written with a .snappy file extension by default, is not supported.</div>

Additional Configuration for File Format Support

Publication of some formats requires execute permissions

When job results are generated and published in the following formats, the Trifacta platform includes a JAR, from which is extracted a binary executable into a temporary directory. From this directory, the binary is then executed to generate the results in the proper format. By default, this directory is set to `/tmp` on the Trifacta node.

In many environments, execute permissions are disabled on `/tmp` for security reasons. Use the steps below to specify the temporary directory where this binary can be moved and executed.

Steps:

- 1. Login to the application as an administrator.
- 2. From the menu, select **User menu > Admin console > Admin settings**.
- 3. For each of the following file formats, locate the listed parameter, where the related binary code can be executed:

File Format	Parameter	Setting to Add
Snappy	"data-service.jvmOptions"	-Dorg.xerial.snappy.tmpdir=<some executable directory>
Hyper	See previous.	See previous.

- 4. Save your changes and restart the platform.
- 5. Run a job configured for direct publication of the modified file format.

Column Statistics Reference

This page describes the statistical information available for individual columns of data.

- Statistics may vary depending on the column's data type. For example, the statistics retained for states may be different from the statistics for strings.
- Most of these statistics are available in the Column Details panel, which can be opened from the left side of the Transformer page.

Below, you can review general statistics maintained for each data type, followed by breakdowns of statistics for each specific type of data.

NOTE: Before your job is run, profiling information such as column statistics are exact counts of the sample that is currently loaded. After the job is run, profiled results in the Job Results page might include estimates for some metrics and counts, depending on the scale of the dataset.

General Column Counts

For any selection of values in a column, the following counts are generally available.

Count Name	Description
Valid Values	Count of values that are valid for the column's data type
Unique Values	Count of unique values. Duplicate values are not counted.
Outlier Values	Count of values that qualify as outliers. An outlier value is either: <ul style="list-style-type: none">• $< (25\text{th percentile}) - (2 * \text{IQR})$• $> (75\text{th percentile}) + (2 * \text{IQR})$• IQR (interquartile range) is the range of values between the two middle quarters, which is equivalent to the range between the 25th and 75th percentiles. Thus, in the above computations, the IQR factor ensures that the outliers are at the extremes of the entire range.
Mismatched Values	Count of values that do not confirm to the column's data type. For example, an Integer column with a value of "MISSING" results in a mismatched value.
Missing Values	Count of values that are not populated

General Column Statistics

These statistics are available for most types of data through the Column Browser.

- For string types (String, Phone Number, Social Security Number, Boolean, Email Address, Credit Card Number, Gender, IP Address, URL, HTTP Code, Date/Time), these stats measure string length.
 - For structured string types (Phone Number, Social Security Number, Boolean, Gender, IP Address, HTTP Code, Date/Time), any variation in these numbers indicates data problems.
- Does not apply to: State

Statistic Name	Description
Minimum	Lowest value in the column
Lower Quartile	The median of the lower half of values (25th percentile)
Median	

	<p>The middle value of the selected set. For example, in a set of 21 values, the median value is the 11th value in ascending order.</p> <ul style="list-style-type: none"> • For datasets with an even number of values, the median is the mean of the two middle values.
Upper Quartile	The median of the upper half of values (75th percentile)
Maximum	Highest value in the column
Average	Average value in the column
Standard Deviation	The computed standard deviation for the selected values.

Supported Data Types

Trifacta® supports the following data types.

Tip: Transforms that include functions work only if the inputs are of a data type and format valid for the function. You should clean up the type and format of your columns before you apply transformations to them.

Supported Data Types

Item	Description
<i>String Data Type</i>	Any non-null value can be typed as String. A String can be anything.
<i>Integer Data Type</i>	The Integer data type applies to positive and negative numeric values that have no decimal point.
<i>Decimal Data Type</i>	Decimal data type applies to floating points up to 15 digits in length. <ul style="list-style-type: none">• In the Trifacta application, this data type is referenced as <code>Decimal</code>.• In storage, this data type is written as <code>Double</code>.
<i>Boolean Data Type</i>	The Boolean data type expresses true or false values.
<i>Social Security Number Data Type</i>	This data type is applied to numeric data following the pattern for United States Social Security numbers.
<i>Phone Number Data Type</i>	This data type is applied to numeric data following common patterns that express telephone numbers and known valid phone numbers in the United States.
<i>Email Address Data Type</i>	This data type matches text values that are properly formatted email addresses.
<i>Credit Card Data Type</i>	Credit card numbers are numeric data that follow the 14-digit or 16-digit patterns for credit cards.
<i>Gender Data Type</i>	This data type matches a variety of text patterns for expressing male/female distinctions.
<i>Zip Code Data Type</i>	This data type matches five- and nine-digit U.S. zipcode patterns.
<i>State Data Type</i>	State data type is applied to data that uses the full names or the two-letter abbreviations for states in the United States.
<i>Object Data Type</i>	An Object data type is a method for encoding key-value pairs. A single field value may contain one or more sets of key-value pairs.
<i>Array Data Type</i>	An array is a list of values grouped into a single value. An array may be of variable length; in one record the array field may contain two elements, while in the next record, it contains six elements.
<i>IP Address Data Type</i>	The IP Address data type supports IPv4 address.
<i>URL Data Type</i>	URL data type is applied to data that follows generalized patterns of URLs.
<i>HTTP Code Data Type</i>	Values of these data types are three-digit numeric values, which correspond to recognized HTTP Status Codes.
<i>Datetime Data Type</i>	Trifacta® supports a variety of Datetime formats, each of which has additional variations to it.

Custom Types

If you have created a custom type, it is available for selection from the column type drop-down.

NOTE: After a custom type has been created, a platform restart is required. Please contact your Trifacta administrator.

Developers may also define custom data types using regular expressions. See *Create Custom Data Types Using RegEx*.

String Data Type

Any non-null value can be typed as String. A String can be anything.

NOTE: If a column of values fails to match another data type, the column is typed as String data.

The length of an individual string value is limited only by the limit applied to an entire row of data, which is approximately 1 MB of characters.

NOTE: On export to relational systems, string lengths are limited to 256 for performance reasons.

Integer Data Type

The Integer data type applies to positive and negative numeric values that have no decimal point.

- Punctuation such as commas and dollar signs (\$) are not supported. These markers must be removed from numeric values through transform steps before you can change the type to Integer.
- The following range is considered safe for values of this type. There may be inconsistencies in output for values outside this range:
 - **Safe minimum:** -9007199254740991 ($-2^{53} + 1$)
 - **Safe maximum:** 9007199254740991 ($2^{53} - 1$)

NOTE: Scientific notation is not supported for Integer data type. Please use Decimal data type instead.

Examples:

- 4
- -23
- 1234567890123456
- -1234567890123456

Decimal Data Type

Decimal data type applies to floating points up to 15 digits in length.

- In the Trifacta application, this data type is referenced as `Decimal`.
- In storage, this data type is written as `Double`.

Punctuation such as commas and dollar signs (\$) are not supported. These markers must be removed from numeric values through transformation steps before you can change the type to `Decimal`.

The following range is considered safe for values of this type. There may be inconsistencies in output for values outside this range:

- **Safe minimum:** `4.9406564584124654e-324`
- **Safe maximum:** `1.7976931348623157e+308`

Notes:

- Decimal values that are longer than 15 digits are treated as String values and may appear as mismatched values in a `Decimal` column.
- Scientific notation is supported.
- Trifacta utilizes Java's `Float` data type for its `Decimal` data validation, which may result in some loss of precision in rare cases.

Examples:

- `123.45`
- `3000.00`
- `1.2345678E+22`
- `7.4423e-12`
- `-4.123e+12`

Boolean Data Type

The Boolean data type expresses true or false values.

Supported Values:

Values can also be expressed as combinations of the following (e.g. `True` / `f`).

- `true` / `false`
- `True` / `False`
- `t`/`f`
- `T`/`F`
- `yes` / `no`
- `Yes` / `No`
- `y`/`n`
- `Y` / `N`
- `on` / `off`
- `1` / `0`

Social Security Number Data Type

This data type is applied to numeric data following the pattern for United States Social Security numbers.

Supported Patterns and Delimiters:

Delimiter	Example
dash	###-##-####
space	### ## ####
no space	#####

Data Validation

When values are validated against the Social Security Number data type, the Trifacta® application validates the values using regular expressions. This regular expression method checks for general Social Security Number patterns and is fast to evaluate.

However, some values may follow the regular expression validation pattern but are not accurate social security numbers. These values may be detected as valid values.

Phone Number Data Type

This data type is applied to numeric data following common patterns that express telephone numbers and known valid phone numbers in the United States.

- Parentheses are optional around area code values: (###) ###-####
- Phone numbers may or may not include the 1 at their beginning.
- Dashes and spaces are optional between groups of numbers.

NOTE: Spaces within a group of values are not supported and will result in a mismatched data type entry.

Standard Domestic U.S.:

```
###-###-####  
(###)-###-####  
#####  
1#####
```

Standard Domestic U.S. with extensions:

```
(###)-###-#### x ###
```

- The dash in ###-#### is required when extension is present.
- The extension indicator can be one of the following: x, X, ext, or EXT.

International U.S.:

```
+1(###)-###-####
```

Not supported:

- Local U.S. format not supported:

```
###-####
```

- International format:

```
+##-##-###-####
```

- Internal international format not supported:

```
##-##-##-###-####
```

Data Validation

In addition to validation against common phone patterns, values are checked against permitted U.S. phone numbers. For example, an area code value of 123 is invalid, as this area code value does not exist in the U.S. phone system.

Email Address Data Type

This data type matches text values that are properly formatted email addresses.

Data in the following format is likely to be typed as email addresses:

```
String@String.aaa  
String@String.aaaa
```

where:

- `String` - any set of readable characters of 1 or more characters in length
- `aaa` or `aaaa` - three-letter or four-letter suffix

Credit Card Data Type

Credit card numbers are numeric data that follow the 14-digit or 16-digit patterns for credit cards.

Pattern:

```
#### #### #### ####  
####-####-####-####  
#### #### #### ##  
####-####-####-##
```

Data Validation

When values are validated against the Credit Card data type, the Trifacta® application validates the values using regular expressions. This regular expression method checks for general Credit Card patterns and is fast to evaluate.

Some numerical combinations in the above patterns may be:

- widely known test values
- invalid for credit card numbers

However, some values may follow the regular expression validation pattern but are not accurate credit card numbers. These values may be detected as valid values.

Gender Data Type

This data type matches a variety of text patterns for expressing male/female distinctions.

The following gender identifiers are supported:

Supported Values:

- m / f
- M / F
- male / female
- Male / Female

Zip Code Data Type

This data type matches five- and nine-digit U.S. zipcode patterns.

Zip code data follows the following possible patterns:

```
#####  
#####-####
```

Some numerical combinations of the above are not valid zipcodes.

State Data Type

State data type is applied to data that uses the full names or the two-letter abbreviations for states in the United States.

Supported Patterns and Examples:

Pattern	Examples
Full names	<ul style="list-style-type: none">• California• Arizona• New York
Abbreviations	<ul style="list-style-type: none">• CA• AZ• NY

Also supported:

Abbreviation	Full Name
DC	District of Columbia
PR	Puerto Rico
VI	Virgin Islands

Object Data Type

An **Object** data type is a method for encoding key-value pairs. A single field value may contain one or more sets of key-value pairs.

An Object data type is identified as a set of nested objects in the following format:

```
{ "key": "value" }  
{ "New York": "NY" }  
{ "California": "CA" }
```

Notes:

- Individual values of an Object type must have unique keys. Values, however, may be repeated.
- The order of key-value pairs is not guaranteed.
- The \n and \t escaped characters are supported in inputs for this data type.

NOTE: The Trifacta application recognizes up to 250 unique keys in a column of Object data type.

Array Data Type

An **array** is a list of values grouped into a single value. An array may be of variable length; in one record the array field may contain two elements, while in the next record, it contains six elements.

- Arrays must be wrapped in square brackets; parentheses are not supported.
- The `\n` and `\t` escaped characters are supported in inputs for this data type.

Examples:

```
[ "123" , "456" , "789" ]  
[ "Hello" , "Goodbye" ]  
[ "abc" , "2" ]  
[ "abc" , "3" ]  
[ "A" , [ "B" , "C" ] , "D" ]
```

Ragged arrays: The above arrays are collectively a set of ragged arrays. The number of elements in each array varies. When arrays are ragged, some array functions may return null or empty values.

Nested arrays: The last example above is a nested array, in which one array is nested inside of another.

IP Address Data Type

The IP Address data type supports IPv4 address.

Supported format:

```
###.###.###.###
```

URL Data Type

URL data type is applied to data that follows generalized patterns of URLs.

The domain and suffix of the URL are the only required elements (e.g. `example.com`).

Supported URL Elements:

```
scheme:[//[user:password@]sub-domain.domain.sfx[:port]][/]path[?query][#fragment]
```

- scheme
- username
- password
- full-domain (sub-domain + domain above)
- sub-domain
- domain (required)
- suffix (sfx above) (required)
- port
- path
- query
- fragment

Examples:

```
example.com  
http://example.com  
https://docs.example.com/dosearchsite.action?where=HOME&spaceSearch=true&queryString=Support
```

HTTP Code Data Type

Values of these data types are three-digit numeric values, which correspond to recognized HTTP Status Codes.

If your column contains three-digit numbers in the following ranges, it may be typed as Http Code.

Supported Values:

- 100 - 102
- 200 - 208
- 300 - 308
- 400 - 409
- 400 - 459
- 500 - 509

Datetime Data Type

Contents:

- *Date Range*
- *Data Validation*
- *Formatting Tokens*
- *Supported Datetime Formats*
- *Supported Time Zones*
- *Job Execution*
 - *Differences between Trifacta Photon and Spark running environments*
- *Datetime Schema via API*

Trifacta® supports a variety of Datetime formats, each of which has additional variations to it.

Date Range

Supported Date Ranges:

- **Earliest:** January 1, 1400

NOTE: Two-digit values for the year that are older than 80 years from the current year are forward-ported into the future. For example, in a job run on Dec 31, 2021, the date `01/01/41` is interpreted as 01/01/1941. However, if the job is run the next day (January 01, 2022), then the same data is interpreted as 01/01/2041. See "Two-digit year values" below.

- **Latest:** December 31, 2599

NOTE: The supported date ranges can be modified if needed. For more information, see *Configure Application Limits*.

You can use dates in the Gregorian calendar system only. Dates in the Julian calendar are not supported.

Data Validation

When values are validated against the Datetime data type, the Trifacta application does not compare them to an underlying calendar system. Instead, the application validates the values using regular expressions. This regular expression method checks for general Datetime validation and is fast to evaluate.

However, some values may follow the regular expression validation pattern but are not accurate dates. For example, every four years, February 29 is a valid date. When this date is validated against the Datetime data type, it may be detected as a valid value, while the date is changed in the application to be incremented to a close accurate date, such as March 1 in this example.

Formatting Tokens

You can use the following tokens to change the format of a column of dates:

Letter	Date or	Presentation	Examples
--------	---------	--------------	----------

	Time Component		
M	Month in year	Number	1
MM	Month in year	Number	01
MMMM	Month in year	Month	January
MMM	Month in year	Month	Jan
yy	Year	Number	16
			NOTE: Two-digit values for the year that are older than 80 years from the current year are forward-ported into the future. For example, in a job run on Dec 31, 2021, the date 01 / 01 / 41 is interpreted as 01/01/1941. However, if the job is run the next day (January 01, 2022), then the same data is interpreted as 01/01/2041. See "Two-digit year values" below.
yyyy	Year	Number	2016
D	Day in year	Number	352
d	Day in month	Number	9
dd	Day in a month	Number	09
EEE	Day in week (three-letter abbreviation)	Text	Wed
EEEE	Day in week	Text	Wednesday
h	Hour in day (1-12)	Number	2
	NO TE: Requires an AM /PM indicator (a).		
hh	Hour in am /pm (01-12)	Number	02
	NO TE: Requires an AM /PM indicator (a).		
H	Hour in day (1-12)	Number	2

HH	Hour in day (0-23)	Number	20
m	Minute in an hour	Number	9
mm	Minute in an hour	Number	09
s	Second in a minute	Number	3
ss	Second in a minute	Number	03
SSS	Millisecond	Number	218
X	Time zone	ISO 8601 time zone	-08:00
a	AM/PM indicator	String	AM

NOTE: When publishing to relational targets, Datetime values are written as date/time values in newly created tables. If you are appending to a relational table column that is in timestamp format, Datetime values can be written as timestamps.

Tip: If your DateTime column contains data in multiple formats, you must change the format of the DateTime column to one format and then add a transformation to convert that data to the other format. When all formats of your source date values are converted to a single format, the application should infer the appropriate date and time format.

Supported Separators:

- Date separators: blank space, comma, single hyphen, or forward slash
- Time separators: blank space, comma, single hyphen, colon, t or T
- Non-delimited Datetime values are supported. For example, yyyyymmdd, yyyyymmddThhmmssX.

ISO 8601 Time Zone Notes:

- Support for timezone offset from UTC indicated by +hh:mm, +hhmm, or +hh. For example, the date '2013-11-18 11:55-04:00' is recognized as a DateTime value.
- Datetime part functions (for example, Hour) truncate time zones and return local time.
- If you have a column with multiple time zones, you can convert the column to Unixtime so you can perform Date/Time operations with a standardized time zone using the UNIXTIME function. If you want to work with local times, you can truncate the time zone or use other Datetime functions.

Supported Datetime Formats

For more information on the available formats and examples of each one, see *Datetime Formats (PDF)*.

You can use the DATEFORMAT function to modify the formatting of your Datetime values.

Supported Time Zones

Time zones values (e.g. UTC-08:00) are supported.

Job Execution

Datetime data typing involves the basic type definition, plus any supported formatting options. Depending on where the job is executed, there may be variation in how the Datetime data type is interpreted.

- Some running environments may perform additional inference on the typing.

NOTE: During job execution on Spark, inputs of Datetime data type may result in row values being inferred for data type individually. For example, the String value 01/10/2020 may be inferred by date transformations as 1st Oct, 2020 or 10th Jan, 2020. Resulting outputs of Datetime values may not be deterministic in this scenario.

- Some formatting options may not be supported.

Differences between Trifacta Photon and Spark running environments

If your Datetime data does not contain time zone information, by default:

- Spark uses the time zone of the Trifacta node for Datetime values.

Tip: This use of time zone applies to any Spark-based running environment, such as EMR.

- Trifacta Photon uses the UTC time zone for Datetime values.

This difference in how the values are treated can result in differences in Datetime-based calculations, such as the DATEDIF function.

Workarounds:

You can do one of the following:

- Set the time zone for the Trifacta node to be UTC. You must also set the time zone for your Spark running environment to UTC.
- Apply the following Spark execution properties from the Run Job page:

```
"spark":
  "props": {
    ...
    "spark.driver.extraJavaOptions" : "-Duser.timezone=\"UTC\"",
    "spark.executor.extraJavaOptions" : "-Duser.timezone=\"UTC\""
  }
  ...
}
```

Datetime Schema via API

When Datetime data is returned via API calls, the schema for this information is returned as a three-element array. The additional elements to the specific are required to account for formatting options of for Datetime values.

Tip: Schema information for data types is primarily available via API calls. You may find schema information for columns in JSON versions of the visual profile and flow definitions when they are exported.

Example:

```
"end_date": [
  "Datetime",
```

```
"mm-dd-yy",  
"mm*dd*yyyy"  
]
```

Array Element	Description	Example 1	Example 2
Data type	The internal name for the data type. For Datetime columns, this schema value should always be <code>Datetime</code> .	"Datetime"	"Datetime"
Sub-format	The general format category of the data type	"mm-dd-yy"	"mm-dd-yy"
Format type	The specific formatting for the data type	"mm*dd*yyyy"	"shortMonth*d*yy"

Supported Numeric Formatting

Contents:

- *Supported Key Codes*
- *Key Codes as Separator Values*
 - *Separators for locales*
 - *Example Separators*

The following formatting can be applied to Integer and Decimal types or to String values that are being converted to numeric types.

Tip: Trifacta supports Java number formatting strings, with some exceptions.

Supported Key Codes

Code	Description	Example Format String	Example Inputs	Example Outputs
#	Insert a digit if it is present in the data.	' ### , ### '	99 999 1000 10000	99 999 1,000 10,000
0	Indicate required digits. If a digit is not available in the source, inserts zero in the data.	' 00 . ## '	20 7.1	20.00 07.1
\$	<p>You can add constants values to the expression. For example, you can insert currency markers at the beginning of your expression.</p> <div>NOTE: The following currency formats are supported: \$, "€", "£", "¥", "", "", "NT\$", "R\$", "R", "Rs", "Kr"</div> <p>Whitespace is respected, except in the following case.</p>	' \$ ## . ## '	20 2514.22 6.6666	\$ 20 \$ 2514.22 \$ 6.67
(space)	You can use space as a grouping separator. When space is used to group sets of digits, all other whitespace in the value is trimmed.	' \$ ###.## ' where space is used as grouping separator.	123456.78 £ 123456.78	\$123 456.78 \$123 456.78
%	<p>Percentage expressions can be at the back of the number formatting expression.</p> <div>NOTE: When the percentage sign is added to the format string, the value is automatically multiplied by 100. When the format string is used with the NUMVALUE function, the value is automatically divided by 100 to return the decimal value.</div>	' ## . ## % '	0.20 14.22 6.6666	20 % 1422 % 666.67 %
-	<p>Negative value indicators can be added to the front part of the number formatting string.</p> <ul style="list-style-type: none">• Negative value indicators at the end of the number are not supported.• If the source value is positive, the negative value is rendered.	' -###,###.00 '	123 -123 1234.56 -1234.56	-123.00 --123.00 -1234.56 --1234.56

	<p>NOTE: In this case, the source value is formatted to appear as its opposite.</p> <ul style="list-style-type: none"> If the source value is negative, a second dash is added to the front of the value. See examples. <p>NOTE: In this case, the value is formatted as a non-numeric value. You can add a second step to your recipe to remove the second dash from column values.</p> <p>NOTE: After the formatting has been applied, type inference may be re-applied to the column, which can change the data type of the column.</p>			
--	--	--	--	--

Key Codes as Separator Values

Some functions support the use of specifying key codes for grouping and decimal separators:

- NUMFORMAT Function** Formats a numeric set of values according to the specified number formatting. Source values can be a literal numeric value, a function returning a numeric value, or reference to a column containing an Integer or Decimal values.
- NUMVALUE Function** Converts a string formatted as a number into an Integer or Decimal value by parsing out the specified decimal and group separators. A string or a function returning formatted numbers of String type or a column containing formatted numbers of string type can be inputs.

NOTE: Separators must be specified when using the NUMVALUE function.

Separators for locales

Grouping and decimal separators can be used to format values for specific locales. Below, you can see how you can format values for locales.

Example Locale	Grouping Separator	Decimal Separator	Example Formatting	Example Output
U.S locale	Comma (,)	Period (.)	<code>NUMFORMAT(SUM(1000000,DIVIDE(1,100)), '###,###.00',' ',' ','.')'</code>	1,000,000.01
Spanish locale	Period (.)	Comma (,)	<code>NUMFORMAT(SUM(1000000,DIVIDE(1,100)), '###,###.00',' ',' ','.')'</code>	1.000.000,01
French locale	Space	Comma (,)	<code>NUMFORMAT(SUM(1000000,DIVIDE(1,100)), '###,###.00',' ',' ','.')'</code>	1 000 000,01

Example Separators

Input	Example Format String	Grouping Separator	Decimal Separator	Output

123.45	##.00	,	.	123.45
123.4	##.00	,	.	123.40
1234	#,###	,	.	1,234
1234.5	#,###.#	,	.	1,234.5
1234.56	#,###.##	,	.	1,234.56
1234	###,#	.	,	1.234
1234.56	###,#	.	,	1.234,56
1234	#,##	.	,	1.234
1234	#.###,0	.	,	1.234,0
123.45	##,#	space	,	123,45
1234	# ###	space	,	1 234
1234.5	# ###,#	space	,	1 234,5
1234.56	# ###,##	space	,	1 234,56

Type Conversions

Contents:

- *Import*
 - *Type Inference*
 - *Export*
 - *Type Conversions*
-

Import

When data is imported:

- Supported data types from the source are converted to corresponding data types supported by the application, based upon the conversions listed in this section.
- Types that are not supported but are recognized by the application are converted to String types.
- Data for types that cannot be read from the source due to technical reasons are converted to null values on import.

Type Inference

By default, the Trifacta application applies type inference for imported data. The application attempts to infer a column's appropriate data type in the application based on a review of the first lines in the sample and how those lines are interpreted by the type system.

Export

On export from the Trifacta application:

- The application maps the internal Trifacta data type to the explicit type listed in the appropriate page in this section.
- Unmapped types are converted to the equivalent of strings.

Type Conversions

Item	Description
<i>Avro Data Type Conversions</i>	This section covers data type conversions between the Trifacta® application and the Avro file format.
<i>DB2 Data Type Conversions</i>	This section covers data type conversions between the Trifacta® application and DB2 databases.
<i>Hive Data Type Conversions</i>	This section covers data type conversions between the Trifacta® application and Hive.
<i>Oracle Data Type Conversions</i>	This section covers data type conversions between the Trifacta® application and Oracle databases.
<i>MySQL Data Type Conversions</i>	This section covers data type conversions between the Trifacta® application and MySQL databases.
<i>Parquet Data Type Conversions</i>	This section covers data type conversions between the Trifacta® application and the Parquet file format.
<i>Postgres Data Type Conversions</i>	This section covers data type conversions between the Trifacta® application and PostgreSQL databases.
<i>Redshift Data Type Conversions</i>	This section covers data type conversions between the Trifacta® application and Redshift.

<i>Snowflake Data Type Conversions</i>	This section covers data type conversions between the Trifacta® application and Snowflake databases.
<i>AWS Glue Data Type Conversions</i>	This section covers data type conversions between the Trifacta® application and AWS Glue.
<i>Salesforce Data Type Conversions</i>	This section covers data type conversions between the Trifacta® application and Salesforce.
<i>SQL Server Data Type Conversions</i>	This section covers data type conversions between the Trifacta® application and SQL Server databases.
<i>SQL DW Data Type Conversions</i>	This section covers data type conversions between the Trifacta® application and SQL DW datastores.
<i>Databricks Tables Data Type Conversions</i>	This section covers data type conversions between the Trifacta® application and Databricks Tables.
<i>Tableau Hyper Data Type Conversions</i>	This section covers data type conversions between the Trifacta® application and Tableau Hyper format.
<i>Teradata Data Type Conversions</i>	This section covers data type conversions between the Trifacta® application and Teradata databases.
<i>SharePoint Data Type Conversions</i>	This section covers data type conversions between the Trifacta® application and SharePoint.

Avro Data Type Conversions

This section covers data type conversions between the Trifacta® application and the Avro file format.

NOTE: The Trifacta® data types listed in this page reflect the raw data type of the converted column. Depending on the contents of the column, the Transformer Page may re-infer a different data type, when a dataset using this type of source is loaded.

Import

Avro Data Type	Trifacta Data Type	Notes
string	String	
int	Integer	
long	Integer	
float	Decimal	
double	Decimal	
boolean	Bool	
map	Object	
array	Array	
record	Object	
enum	String	
fixed	String	

Export

On export, Trifacta data types are exported to their corresponding Avro types, with the following specific mappings:

Trifacta Data Type	Avro Data Type	Notes
Boolean	BOOLEAN	
Integer	LONG	
Decimal	DOUBLE	
String	STRING	

The fallback data type on export is STRING.

DB2 Data Type Conversions

This section covers data type conversions between the Trifacta® application and DB2 databases.

NOTE: The Trifacta® data types listed in this page reflect the raw data type of the converted column. Depending on the contents of the column, the Transformer Page may re-infer a different data type, when a dataset using this type of source is loaded.

Access/Read

Source Data Type	Supported	Trifacta Data Type
BOOLEAN	Y	Bool
VARCHAR	Y	String
INTEGER	Y	Integer

Publish/Write

Trifacta Data Type	DB2 Data Type	Notes
Bool	BOOLEAN	
	VARCHAR	
Integer	BIGINT	
	INT	
	SMALLINT	
	VARCHAR	
String	VARCHAR(1024)	
Datetime	TIMESTAMP	
	VARCHAR	
Time	VARCHAR(1024)	
Decimal	DECIMAL	
	NUMERIC	
	VARCHAR	
Map	VARCHAR(1024)	
Array	VARCHAR(1024)	

Hive Data Type Conversions

Contents:

- *Access/Read*
- *Write/Publish*
 - *Create new table*
 - *Append to existing table*
 - *Notes on Datetime columns*

This section covers data type conversions between the Trifacta® application and Hive.

NOTE: The Trifacta® data types listed in this page reflect the raw data type of the converted column. Depending on the contents of the column, the Transformer Page may re-infer a different data type, when a dataset using this type of source is loaded.

Access/Read

When a Hive data type is imported, its JDBC data type is remapped according to the following table.

Tip: Data precision may be lost during conversion. You may want to generate min and max values and compute significant digits for values in your Hive tables and then compute the same in the Trifacta application.

Source Data Type	Supported?	Trifacta Data Type	Notes
array	Y	Array	
bigint	Y	Integer	NOTE: The Trifacta platform may infer bigint columns containing very large or very small values as String data type.
binary	Y	String	
boolean	Y	Bool	
char	Y	String	
date	Y	Datetime	
decimal	Y	Decimal	
double	Y	Decimal	
float	Y	Decimal	NOTE: On import, some float columns may be interpreted as Integer data type in the Trifacta platform. To fix, you can explicitly set the column's data type to Decimal in the Transformer page.
int	Y	Integer	
map	Y	Object	
smallint	Y	Integer	
string	Y	String	

struct	Y	Object	
timestamp	Y	Datetime	
tinyint	Y	Integer	
uniontype	N		
varchar	Y	String	

Write/Publish

Create new table

NOTE: By default, the maximum length of values published to VARCHAR columns is 256 characters. As needed, this limit can be changed for multiple publication targets. For more information, see *Configure Application Limits*.

Trifacta Data Type	Hive Data Type	Notes
String	string	
Integer	bigint	<p>NOTE: The Trifacta platform may infer Integer columns containing very large or very small values as String data type. Before you publish, you should verify that your columns containing extreme values are interpreted as Integer type. You can import a target schema to assist in lining up your columns with the expected target. For more information, see <i>Overview of RapidTarget</i>.</p>
Decimal	double	
Bool	boolean	
Datetime	Timestamp /string (see Notes on Datetime columns below)	Target data type is based on the underlying data. Time zone information is retained.
Object	string	
Array	string	

Append to existing table

If you are publishing to a pre-existing table, the following data type conversions apply:

- **Columns:** Trifacta data types
- **Rows:** Target table data types

In any table cell, a Y indicates that the append operation for that data type mapping is supported.

NOTE: You cannot append to Hive map and array column types from Trifacta columns of Map and Array type, even if you imported data from this source.

	String	Integer	Datetime	Bool	Decimal	Map	Array	Out of Range error
CHAR	Y	Y	Y	Y	Y	Y	Y	

VARCHAR	Y	Y	Y	Y	Y	Y	Y	
STRING	Y	Y	Y	Y	Y	Y	Y	
INT		Y						NULL
BIGINT		Y						n/a
TINYINT								NULL
SMALLINT								NULL
DECIMAL		Y			Y			NULL
DOUBLE		Y			Y			n/a
FLOAT					Y			NULL
TIMESTAMP			Y					
BOOLEAN				Y				

Notes on Datetime columns

Run Job

Columns in new tables created for output of `Datetime` columns are written with the Hive `timestamp` data type. These columns can be appended.

- Before release 4.2.1, `Datetime` columns were written to Hive as type `String`. Jobs that were created in these releases and that write to pre-existing tables continue to behave this way.
- A single job cannot write `Datetime` values to one table as `String` type and to another table as `Timestamp` type. This type of job should be split into multiple types. The table schemas may require modification.
 - The above issue may appear as the following error when executing the job:

```
Unable to publish due to datetime data type conflict in column XXXX
```

Ad-Hoc Publishing

- When you export pre-generated results to Hive, all new tables created for `Datetime` column values continue to store `String` data type in Hive for Release 4.2.1. These columns can be appended with new `String` data.
- When you publish results from a job through the Publishing dialog to Hive, all `Datetime` column values are written as `String` type.
- If you are appending to a `Timestamp` column, the exported `Datetime` column must be in the following format: `yyyy-MM-dd HH:mm:ss.xxxx`

Oracle Data Type Conversions

This section covers data type conversions between the Trifacta® application and Oracle databases.

NOTE: The Trifacta® data types listed in this page reflect the raw data type of the converted column. Depending on the contents of the column, the Transformer Page may re-infer a different data type, when a dataset using this type of source is loaded.

NOTE: Dots (.) in the names of Oracle tables or table columns are not supported.

Access/Read

Source Data Type	Supported	Trifacta Data Type
ANYDATA	N	
ANYDATASET	N	
ANYTYPE	N	
BFILE	N	
BINARY_DOUBLE	Y	Decimal
BINARY_FLOAT	Y	Decimal
BLOB	N	
CHAR	Y	String
CLOB	Y	String
DATE	Y	Date/Time
DBURIType	Y	String
FLOAT	Y	Decimal
HTTPURIType	Y	String
INTERVAL_DAY TO SECOND	Y	String
INTERVAL_YEAR TO MONTH	Y	String
LONG RAW	Y	String (base64)
<media types>	N	
NCHAR	Y	String
NCLOB	Y	String
NUMBER	Y	Decimal / Integer
NVARCHAR2	Y	Integer
RAW	Y	String (base64)
ROWID	N	
SDO_GEOMETRY	N	
SDO_GEORASTER	N	
SDO_TOPO_GEOMETRY	N	
TIMESTAMP	Y	String

TIMESTAMP WITH LOCAL TIMEZONE	Y	String
TIMESTAMP WITH TIMEZONE	Y	String
URITYPE	Y	String
UROWID	N	
VARCHAR2	Y	String
XDBUIRType	Y	String
XMLType	N	

NOTE: Implementation of Oracle custom types is a custom engagement. For more information, please contact *Alteryx Customer Success Services*.

Write/Publish

Trifacta Data Type	Oracle Column Type	Notes
Bool	VARCHAR (256)	Oracle VARCHAR column type must include the maximum number of permitted characters.
Integer	INT	
	INTEGER	
	SMALLINT	
	DECIMAL	
	DEC	
	NUMERIC	
	NUMBER	
	FLOAT	
	VARCHAR (256)	Oracle VARCHAR column type must include the maximum number of permitted characters.
String	VARCHAR (256)	Oracle VARCHAR column type must include the maximum number of permitted characters.
Datetime	TIMESTAMP	
	VARCHAR (256)	Oracle VARCHAR column type must include the maximum number of permitted characters.
Timestamp	VARCHAR (256)	Oracle VARCHAR column type must include the maximum number of permitted characters.
Float	DECIMAL (38,9)	Oracle DECIMAL column type must include the scale (total number of permitted digits) and precision (total number of digits permitted to the right of the decimal) as parameters.
	DEC	
	NUMERIC	
	NUMBER	
	FLOAT	
	VARCHAR (256)	Oracle VARCHAR column type must include the maximum number of permitted characters.

Map	VARCHAR (256)	Oracle VARCHAR column type must include the maximum number of permitted characters.
Array	VARCHAR (256)	Oracle VARCHAR column type must include the maximum number of permitted characters.

NOTE: If you are appending to an existing table where a column's Trifacta data type is not mapped to the column data type in the target table, a validation error is thrown, as the platform writes unmapped types as String data type.

MySQL Data Type Conversions

This section covers data type conversions between the Trifacta® application and MySQL databases.

NOTE: The Trifacta® data types listed in this page reflect the raw data type of the converted column. Depending on the contents of the column, the Transformer Page may re-infer a different data type, when a dataset using this type of source is loaded.

Access/Read

Source Data Type	Supported?	Trifacta Data Type
array	N	
bigint	Y	Integer
tinyint	Y	Integer
mediumint	Y	Integer
smallint	Y	Integer
int	Y	Integer
bit [(n)]	Y	String
float	Y	Float
numeric	Y	Float
decimal	Y	Decimal
real	Y	Float
boolean	Y	Bool
character varying(n), varchar(n)	Y	String
character(n), char(n)	Y	String
date	Y	Date
double	Y	Decimal
enum	N	
set	N	
json	Y	String
text	Y	String
tinytext	Y	String
mediumtext	Y	String
longtext	Y	String
time	Y	Datetime
datetime	Y	Datetime
timestamp	Y	Datetime
year	Y	String

Write/Publish

Trifacta Data Type	PostgreSQL Column Type	Notes
Bool	BIT	
	VARCHAR	
	TINYINT(1)	
Integer	BIGINT	
	INT	
	SMALLINT	
	MEDIUMINT	
	TINYINT	
	VARCHAR	
	VARCHAR	
	VARCHAR	
String	VARCHAR	
	TINYTEXT	
	TEXT	
	MEDIUMTEXT	
	LONGTEXT	
Datetime	TIMESTAMP	
	DATETIME	
	DATE	
	VARCHAR	
Timestamp	VARCHAR	
Float	FLOAT	
	DECIMAL	
	REAL	
	NUMERIC	
	VARCHAR	
Map	VARCHAR	
Array	VARCHAR	

NOTE: If you are appending to an existing table where a column's Trifacta data type is not mapped to the column data type in the target table, a validation error is thrown, as the platform writes unmapped types as String data type.

Parquet Data Type Conversions

This section covers data type conversions between the Trifacta® application and the Parquet file format.

NOTE: The Trifacta® data types listed in this page reflect the raw data type of the converted column. Depending on the contents of the column, the Transformer Page may re-infer a different data type, when a dataset using this type of source is loaded.

Import

NOTE: Trifacta does not support ingest of Parquet files with nested values, which can occur for Map or Object data types.

Parquet Data Type	Trifacta Data Type	Notes
STRING	String	
INT	Integer	
DECIMAL	Decimal	
DATE	Datetime	
TIME	Datetime	
TIMESTAMP	Datetime	
LIST	Array	
MAP	Object	

Limitations on import:

The Parquet data format supports the use of row groups for organizing chunks of data. This row grouping is helpful for processing across distributed systems.

Trifacta places limitations on the volume of data that can be displayed in the browser. By default, these limits are set to 10 MB.

If Parquet row groups are greater than 10 MB:

- You cannot preview data from the file before import.
- When a Parquet-based dataset is loaded in the Transformer page, the screen may be blank.

Tip: You can create a new sample from inside the Transformer page. The sample is displayed normally.

Other product functions work as expected with Parquet format.

Export

On export, Trifacta data types are exported to their corresponding Parquet types, with the following specific mappings:

Trifacta Data Type	Parquet Data Type	Notes

Boolean	BOOLEAN	
Integer	INT64	
Decimal	DOUBLE	
String	BYTE_ARRAY (STRING)	

The fallback data type on export is BYTE_ARRAY (STRING).

Postgres Data Type Conversions

This section covers data type conversions between the Trifacta® application and PostgreSQL databases.

NOTE: The Trifacta® data types listed in this page reflect the raw data type of the converted column. Depending on the contents of the column, the Transformer Page may re-infer a different data type, when a dataset using this type of source is loaded.

Access/Read

Source Data Type	Supported?	Trifacta Data Type
array	N	
bigint	Y	Integer
bigserial	Y	Integer
bit [(n)]	Y	String
bit varying [(n)]	Y	String
boolean	Y	Bool
box	Y	String
bytea	Y	String
character varying(n), varchar(n)	Y	String
character(n), char(n)	Y	String
cidr	Y	String
circle	Y	String
composite	N	
date	Y	Date
daterange	N	
decimal	Y	Decimal
double precision	Y	Decimal
enum	N	
inet	Y	String
int4range	N	
int8range	N	
integer	Y	Integer
interval [fields] [(p)]	Y	String
json	Y	Object
line	N	
lseg	Y	String
macaddr	Y	String
money	Y	Decimal
numeric	Y	Decimal/Integer

numrange	N	
oid	N	
path	Y	String
point	Y	String
polygon	Y	String
real	Y	Decimal
serial	Y	Integer
smallint	Y	Integer
smallserial	Y	Integer
text	Y	String
time [(p)]	Y	Date
time [(p)] with time zone	Y	String
timestamp [(p)]	Y	Date
timestamp [(p)] with time zone	Y	Date
tsquery	Y	String
tsrange	N	
tstzrange	N	
tsvector	Y	String
txid_snapshot	Y	String
uuid	Y	String
xml	Y	String

Write/Publish

Trifacta Data Type	PostgreSQL Column Type	Notes
Bool	BOOL	
	VARCHAR	
Integer	BIGINT	
	INT	
	SMALLINT	
	VARCHAR	
String	VARCHAR	
Datetime	TIMESTAMP	
	VARCHAR	
Float	DECIMAL	
	NUMERIC	
	VARCHAR	
Map	VARCHAR	
Array	VARCHAR	

NOTE: If you are appending to an existing table where a column's Trifacta data type is not mapped to the column data type in the target table, a validation error is thrown, as the platform writes unmapped types as String data type.

Redshift Data Type Conversions

Contents:

- *Access/Read*
- *Write/Publish*
 - *Create new table*
 - *Append to existing table*

This section covers data type conversions between the Trifacta® application and Redshift.

NOTE: The Trifacta® data types listed in this page reflect the raw data type of the converted column. Depending on the contents of the column, the Transformer Page may re-infer a different data type, when a dataset using this type of source is loaded.

Access/Read

Source Data Type	Supported	Trifacta data type
string	Y	String
bigint	Y	Integer
double precision	Y	Decimal
bool	Y	Boolean
date	Y	DateTime
timestamp	Y	DateTime

Write/Publish

Create new table

NOTE: By default, the maximum length of values published to VARCHAR columns is 256 characters. As needed, this limit can be changed for multiple publication targets. For more information, see *Configure Application Limits*.

Trifacta Data Type	Redshift Data Type	Notes
String	varchar	
Integer	bigint	
Decimal	double precision	
Bool	bool	
Datetime	timestamp	When you publish results from a job through the Export Results window to Redshift, all Datetime column values are written as String type.
Object	varchar	
Array	varchar	

NOTE: Object and Array types are written in the supported format, if escaped commas and backslashes are unescaped in your recipe.

Append to existing table

If you are publishing to a pre-existing table, the following data type conversions apply:

- **Columns:** Trifacta data types
- **Rows:** Target table data types

In any table cell, a Y indicates that the append operation for that data type mapping is supported.

	String	Integer	Datetime	Bool	Float	Map	Array
TEXT	Y	Y	Y	Y	Y	Y	Y
VARCHAR	Y	Y	Y	Y	Y	Y	Y
NVARCHAR	Y	Y	Y	Y	Y	Y	Y
BPCHAR	Y	Y	Y	Y	Y	Y	Y
NCHAR							
CHAR							
CHARACTER VARYING	Y	Y	Y	Y	Y	Y	Y
SMALLINT							
INT2							
INTEGER							
INT							
INT4							
BIGINT		Y					
INT8		Y					
DECIMAL		Y			Y		
NUMERIC		Y			Y		
DOUBLE_PRECISION		Y			Y		
FLOAT		Y			Y		
FLOAT4							
FLOAT8					Y		
REAL							
BOOL				Y			
BOOLEAN				Y			
TIMESTAMP			Y				
TIMESTAMPTZ			Y				

Snowflake Data Type Conversions

This section covers data type conversions between the Trifacta® application and Snowflake databases.

NOTE: The Trifacta® data types listed in this page reflect the raw data type of the converted column. Depending on the contents of the column, the Transformer Page may re-infer a different data type, when a dataset using this type of source is loaded.

Access/Read

Source Type	Supported	Trifacta Data Type
NUMBER	Y	Integer
DECIMAL	Y	Integer
NUMERIC	Y	Integer
INT	Y	Integer
INTEGER	Y	Integer
BIGINT	Y	Integer
SMALLINT	Y	Integer
FLOAT	Y	Decimal
FLOAT4	Y	Decimal
FLOAT8	Y	Decimal
DOUBLE	Y	Decimal
DOUBLE_PRECISION	Y	Decimal
REAL	Y	Decimal
VARCHAR	Y	String
CHAR	Y	String
CHARACTER	Y	String
STRING	Y	String
TEXT	Y	String
BINARY	Y	String
VARBINARY	Y	String
BOOLEAN	Y	Bool
DATE	Y	String
DATETIME	Y	String
TIME	Y	String
TIMESTAMP	Y	String
TIMESTAMP_TZ	Y	Datetime
TIMESTAMP_LTZ	Y ¹⁾	Datetime
TIMESTAMP_NTZ	Y ²⁾	Datetime
VARIANT	N	

OBJECT	N	
ARRAY	N	

Notes:

1. Convert to Datetime using local timezone of the Snowflake connector
2. Convert to Datetime using UTC

Write/Publish

Trifacta Data Type	Supported	Snowflake Type	Notes
Array	N		
Bool	Y	BOOLEAN	
Date	Y	TIMESTAMP	See below.
Datetime	Y	TIMESTAMP	
Time	Y	TIME	
Float	Y	FLOAT	
Integer	Y	BIGINT	
Map	N		
String	Y	VARCHAR	

Notes on Date publishing:

NOTE: Trifacta Date columns can be published to existing Snowflake columns of Date, Datetime, and String type.

When Dates are published to Snowflake, the following date formats are supported. In some cases, missing data is inserted into the output value.

Trifacta date format	Snowflake date format
yy-dd-mm	yy-dd-mm
yy-mm-dd	yy-mm-dd
dd-mm-yy	dd-mm-yy
mm-dd-yy	mm-dd-yy
mm-yy	mm-yy
dd-mm	1970-dd-mm
mm-dd	1970-mm-dd
mm-yy	mm-yy-01

- On publication, all dates are written in the following format: `yyyy-mm-dd`.
- No other date formats are supported for writing to Snowflake as date values.

AWS Glue Data Type Conversions

This section covers data type conversions between the Trifacta® application and AWS Glue.

NOTE: The Trifacta® data types listed in this page reflect the raw data type of the converted column. Depending on the contents of the column, the Transformer Page may re-infer a different data type, when a dataset using this type of source is loaded.

Access/Read

Source Data Type	Supported	Trifacta Data Type	Notes
BIGINT	Y	Integer	Tables generated by the Glue crawler from double-quoted CSV files with columns that are inferred by Glue to be this data type result in empty columns when imported to the Trifacta platform. The workaround is to change the column types for these columns in Glue to STRING.
INT	Y	Integer	
STRING	Y	String	
STRUCT	Y	Object	
DOUBLE	Y	Decimal	Tables generated by the Glue crawler from double-quoted CSV files with columns that are inferred by Glue to be this data type result in empty columns when imported to the Trifacta platform. The workaround is to change the column types for these columns in Glue to STRING.
TIMESTAMP	Y	Datetime	
BOOLEAN	Y	Bool	Tables generated by the Glue crawler from double-quoted CSV files with columns that are inferred by Glue to be this data type result in empty columns when imported to the Trifacta platform. The workaround is to change the column types for these columns in Glue to STRING.

Write/Publish

Not supported.

Salesforce Data Type Conversions

This section covers data type conversions between the Trifacta® application and Salesforce.

NOTE: The Trifacta® data types listed in this page reflect the raw data type of the converted column. Depending on the contents of the column, the Transformer Page may re-infer a different data type, when a dataset using this type of source is loaded.

Access/Read

Source Data Type	Supported?	Trifacta Data Type
BOOL	Y	Bool
CHAR	Y	String
NUMBER	Y	Integer
BINARY_FLOAT	Y	Float
BINARY_DOUBLE	Y	Float

Publish/Write

Publishing to this datastore is not supported in this release.

SQL Server Data Type Conversions

This section covers data type conversions between the Trifacta® application and SQL Server databases.

NOTE: The Trifacta® data types listed in this page reflect the raw data type of the converted column. Depending on the contents of the column, the Transformer Page may re-infer a different data type, when a dataset using this type of source is loaded.

Access/Read

Source Data Type	Supported?	Trifacta Data Type
Bigint	Y	Integer
BIGINT IDENTITY	Y	Integer
Binary	Y	String (Base64)
Bit	Y	Bool
Char	Y	String
Cursor	N	
Date	Y	Date
Datetime	Y	Date
Datetime2	Y	Date
Datetimeoffset	Y	String
Decimal	Y	Integer/Decimal
DECIMAL IDENTITY	Y	Integer
Float	Y	Decimal
GEOGRAPHY	Y	String
GEOMETRY	Y	String
HIERARCHYID	Y	String
Image	Y	String
Int	Y	Integer
INT IDENTITY	Y	Integer
Money	Y	Decimal
Nchar	Y	String
Ntext	Y	String
Numeric	Y	Integer/Decimal
NUMERIC IDENTITY	Y	Integer
Nvarchar	Y	String
Nvarchar(max)	Y	String
Real	Y	Decimal
Smalldatetime	Y	Date
Smallint	Y	Integer

SMALLINT IDENTITY	Y	Integer
Smallmoney	Y	Decimal
Sql_variant	N	
Table	N	
Text	Y	String
Time	Y	String
Timestamp	Y	String
Tinyint	Y	Integer
TINYINT IDENTITY	Y	Integer
Uniqueidentifier	Y	String
Varbinary	Y	String (Base64)
Varbinary(max)	Y	String (Base64)
Varchar	Y	String
Varchar(max)	Y	String
Xml	Y	String

Write/Publish

Trifacta Data Type	SQL Server Column Type	Notes
Bool	BIT	
	VARCHAR(256)	
Integer	BIGINT	
	INT	
	SMALLINT	
	DECIMAL	
	VARCHAR(256)	
String	VARCHAR(256)	
Datetime	DATETIME	
	TIMESTAMP	
	VARCHAR(256)	
Timestamp	TIME	
	VARCHAR(256)	
Float	FLOAT	
	DECIMAL	
	NUMERIC	
	VARCHAR(256)	
Map	VARCHAR(256)	
Array	VARCHAR(256)	

NOTE: If you are appending to an existing table where a column's Trifacta data type is not mapped to the column data type in the target table, a validation error is thrown, as the platform writes unmapped types as String data type.

SQL DW Data Type Conversions

Contents:

- *Access/Read*
- *Write/Publish*
 - *Create new table*
 - *Append to existing table*

This section covers data type conversions between the Trifacta® application and SQL DW datastores.

NOTE: The Trifacta® data types listed in this page reflect the raw data type of the converted column. Depending on the contents of the column, the Transformer Page may re-infer a different data type, when a dataset using this type of source is loaded.

Access/Read

Source Data Type	Supported	Trifacta Data Type	Notes
INT	Y	Integer	
TINYINT	Y	Integer	
SMALLINT	Y	Integer	
BIGINT	Y	Integer	NOTE: The Trifacta platform may infer bigint columns containing very large or very small values as String data type. If needed, you can disable type inference for individual schematized sources. For more information, see <i>Import Data Page</i> .
FLOAT	Y	Float	
REAL	Y	Float	
BIT	Y	Bool	
SMALLMONEY	Y	String	
MONEY	Y	String	
DECIMAL	Y	Float	
NUMERIC	Y	String	
DATETIMEOFFSET	Y	String	
TIME	Y	String	
DATE	Y	String	
DATETIME	Y	String	
DATETIME2	Y	String	
SMALLDATETIME	Y	String	
CHAR	Y	String	

VARCHAR	Y	String	
NCHAR	Y	String	
NVARCHAR	Y	String	
SYSNAME	Y	String	
BINARY	Y	String	
VARBINARY	Y	String	
UNIQUEIDENTIFIER	Y	String	
TIMESTAMP	N		
GEOGRAPHY	N		
GEOMETRY	N		
HIERARCHID	N		
IMAGE	N		
TEXT	N		
NTEXT	N		
XML	N		
CURSOR	N		
ROWVERSION	N		
SQL_VARIANT	N		

Write/Publish

Create new table

NOTE: By default, the maximum length of values published to VARCHAR columns is 256 characters. As needed, this limit can be changed for multiple publication targets. For more information, see *Configure Application Limits*.

Trifacta Data Type	SQL DW Data Type	Notes
String	VARCHAR	
Integer	BIGINT	<p>NOTE: The Trifacta platform may infer Integer columns containing very large or very small values as String data type. Before you publish, you should verify that your columns containing extreme values are interpreted as Integer type. You can import a target schema to assist in lining up your columns with the expected target. For more information, see <i>Overview of RapidTarget</i>.</p>
Float	FLOAT	
Bool	BIT	

Datetime	DATETIME2	If a time-only value is published as an append to a pre-defined DATETIME2 column, then the output column is prepended with 1900-01-01.
Datetime (with time value only)	TIME	<p>If a value is published as an append to a pre-defined TIME column, then any date information is dropped from the output.</p> <div> NOTE: The platform publishes to Timestamp columns only for append operations to pre-existing tables. </div>
Datetime	VARCHAR	<p>If a Datetime value is published as an append to a pre-defined VARCHAR column, then the output column contains the string value of whatever appears in the Transformer page.</p> <p>When you publish results from a job through the Publishing dialog to SQL DW, all Datetime column values are written as String type.</p>
Map	VARCHAR	
Array	VARCHAR	
Date	VARCHAR	
All Other Data Types	VARCHAR	

Append to existing table

If you are publishing to a pre-existing table, the following data type conversions apply:

- **Columns:** Trifacta data types
- **Rows:** Target table data types

In any table cell, a **Y** indicates that the append operation for that data type mapping is supported.

	String	Integer	Datetime	Bool	Float	Map	Array
VARCHAR	Y	Y	Y	Y	Y	Y	Y
NVARCHAR	Y	Y	Y	Y	Y	Y	Y
CHAR	Y	Y	Y	Y	Y	Y	Y
NCHAR	Y	Y	Y	Y	Y	Y	Y
TINYINT							
SMALLINT							
INT							
BIGINT		Y					
DATETIME2			Y				
TIME			Y				
BIT				Y			
FLOAT		Y			Y		
REAL		Y			Y		
DECIMAL		Y			Y		

Databricks Tables Data Type Conversions

Contents:

- *Access/Read*
- *Write/Publish*
 - *Create new table*
 - *Append to existing table*
 - *Notes on Datetime columns*

This section covers data type conversions between the Trifacta® application and Databricks Tables.

NOTE: The Trifacta® data types listed in this page reflect the raw data type of the converted column. Depending on the contents of the column, the Transformer Page may re-infer a different data type, when a dataset using this type of source is loaded.

Access/Read

When a Databricks Tables data type is imported, its JDBC data type is remapped according to the following table.

Tip: Data precision may be lost during conversion. You may want to generate min and max values and compute significant digits for values in your Hive tables and then compute the same in the Trifacta application.

Source Data Type	Supported?	Trifacta Data Type	Notes
array	Y	Array	
bigint	Y	Integer	NOTE: The Trifacta platform may infer bigint columns containing very large or very small values as String data type.
binary	Y	String	
boolean	Y	Bool	
char	Y	String	
date	Y	Datetime	
decimal	Y	Decimal	
double	Y	Decimal	
float	Y	Decimal	NOTE: On import, some float columns may be interpreted as Integer data type in the Trifacta platform. To fix, you can explicitly set the column's data type to Decimal in the Transformer page.
int	Y	Integer	
map	Y	Object	
smallint	Y	Integer	
string	Y	String	

struct	Y	Object	
timestamp	Y	Datetime	
tinyint	Y	Integer	
uniontype	N		
varchar	Y	String	

Write/Publish

Create new table

NOTE: By default, the maximum length of values published to VARCHAR columns is 256 characters. As needed, this limit can be changed for multiple publication targets. For more information, see *Configure Application Limits*.

Trifacta Data Type	Databricks Tables Data Type	Notes
String	string	
Integer	bigint	<p>NOTE: The Trifacta platform may infer Integer columns containing very large or very small values as String data type. Before you publish, you should verify that your columns containing extreme values are interpreted as Integer type. You can import a target schema to assist in lining up your columns with the expected target. For more information, see <i>Overview of RapidTarget</i>.</p>
Decimal	double	
Bool	boolean	
Datetime	Timestamp /string (see Notes on Datetime columns below)	Target data type is based on the underlying data. Time zone information is retained.
Object	string	
Array	string	

Append to existing table

If you are publishing to a pre-existing table, the following data type conversions apply:

- **Columns:** Trifacta data types
- **Rows:** Target table data types

In any table cell, a Y indicates that the append operation for that data type mapping is supported.

NOTE: You cannot append to Databricks Tables map and array column types from Trifacta columns of Map and Array type, even if you imported data from this source.

	String	Integer	Datetime	Bool	Decimal	Map	Array	Out of Range error
CHAR	Y	Y	Y	Y	Y	Y	Y	
VARCHAR	Y	Y	Y	Y	Y	Y	Y	

STRING	Y	Y	Y	Y	Y	Y	Y	
INT		Y						NULL
BIGINT		Y						n/a
TINYINT								NULL
SMALLINT								NULL
DECIMAL		Y			Y			NULL
DOUBLE		Y			Y			n/a
FLOAT					Y			NULL
TIMESTAMP			Y					
BOOLEAN				Y				

Notes on Datetime columns

Run Job

Columns in new tables created for output of `Datetime` columns are written with the Databricks Tables `timestamp` data type. These columns can be appended.

A single job cannot write `Datetime` values to one table as String type and to another table as Timestamp type. This type of job should be split into multiple types. The table schemas may require modification.

- The above issue may appear as the following error when executing the job:

```
Unable to publish due to datetime data type conflict in column XXXX
```

Tableau Hyper Data Type Conversions

This section covers data type conversions between the Trifacta® application and Tableau Hyper format.

NOTE: The Trifacta® data types listed in this page reflect the raw data type of the converted column. Depending on the contents of the column, the Transformer Page may re-infer a different data type, when a dataset using this type of source is loaded.

Access/Read

Directing reading of Tableau Hyper data is not supported in the platform.

Publish/Write

Trifacta Data Type	Tableau Data Type	Notes
String	SqlType.text()	
Integer	SqlType.bigInt()	
Decimal	sqlType.doublePrecision()	
Bool	SqlType.bool()	
All other data types	SqlType.text()	

For more information on SqlTypes, see https://help.tableau.com/current/api/hyper_api/en-us/reference/java/index.html.

Teradata Data Type Conversions

This section covers data type conversions between the Trifacta® application and Teradata databases.

NOTE: The Trifacta® data types listed in this page reflect the raw data type of the converted column. Depending on the contents of the column, the Transformer Page may re-infer a different data type, when a dataset using this type of source is loaded.

Access/Read

Source Data Type	Supported	Trifacta Data Type
BYTE[(n)]	N	
VARBYTE[(n)]	N	
BYTEINT	Y	Integer
SMALLINT	Y	Integer
BIGINT	Y	Integer
INTEGER	Y	Integer
DECIMAL [(n[,m])]	Y	Decimal
NUMERIC [(n[,m])]	Y	Decimal
REAL	Y	Decimal
DOUBLE PRECISION	Y	Decimal
FLOAT	Y	Decimal
NUMBER(n[,m])	Y	Decimal
NUMBER[(*,m)]	Y	Decimal
DATE	Y	String
Time(0)	Y	Datetime
Time(0) WITH TIME ZONE	Y	Datetime
TIME (n)	Y	String
TIME (n) WITH TIME ZONE	Y	String
TIMESTAMP (0)	Y	String
Timestamp(0) WITH TIME ZONE	Y	Datetime
TIMESTAMP (n)	Y	Datetime
TIMESTAMP (n) WITH TIME ZONE	Y	String
INTERVAL DAY [(n)]	Y	String
INTERVAL DAY [(n)] TO HOUR	Y	String
INTERVAL DAY [(n)] TO MINUTE	Y	String
INTERVAL DAY [(n)] TO SECOND	Y	String
INTERVAL HOUR [(n)]	Y	String
INTERVAL HOUR [(n)] TO MINUTE	Y	String
INTERVAL HOUR [(n)] TO SECOND	Y	String

INTERVAL MINUTE [(n)]	Y	String
INTERVAL MINUTE [(n)] TO SECOND [(m)]	Y	String
INTERVAL MONTH	Y	String
INTERVAL SECOND [(n],[m]]	Y	String
INTERVAL YEAR [(n)]	Y	String
INTERVAL YEAR [(n)] TO MONTH	Y	String
CHAR[(n)]	Y	String
CHARACTER(n) CHARACTER SET GRAPHIC	N	
VARCHAR(n)	Y	String
CHAR VARYING(n)	Y	String
LONG VARCHAR	Y	String
LONG VARGRAPHIC	N	
VARGRAPHIC	N	
PERIOD(DATE)	N	
PERIOD(TIME (0))	N	
PERIOD(TIME (0)) WITH TIME ZONE	N	
PERIOD(TIME (n))	N	
PERIOD(TIME (n)) WITH TIME ZONE	N	
PERIOD(TIMESTAMP(0))	N	
PERIOD(TIMESTAMP(0)) WITH TIME ZONE	N	
PERIOD(TIMESTAMP(n))	N	
PERIOD(TIMESTAMP(n)) WITH TIME ZONE	N	
BLOB	N	
CLOB	Y	String
XML	Y	String
ARRAY	Y	Array
VARARRAY	Y	Array
TD_ANYTYPE	N	
VARIANT_TYPE	N	
Distinct	N	
Structured	N	
ST_CircularString	N	
ST_CompoundCurve	N	
ST_Curve	N	
ST_CurvePolygon	N	
ST_GeomCollection	N	
ST_Geometry	N	
ST_LineString	N	
ST_MultiCurve	N	

ST_MultiLineString	N	
ST_MultiPoint	N	
ST_MultiPolygon	N	
ST_MultiSurface	N	
ST_Point	N	
ST_Polygon	N	
ST_Surface	N	
JSON	Y	String

Write/Publish

Trifacta Data Type	Teradata Column Type	Notes
Bool	VARCHAR(256)	
Integer	INT	
	INTEGER	
	SMALLINT	
	VARCHAR(256)	
String	VARCHAR(256)	
Datetime	TIMESTAMP	
	VARCHAR(256)	
Timestamp	VARCHAR(256)	
Float	FLOAT	
	DECIMAL	
	NUMERIC	
	VARCHAR(256)	
Map	VARCHAR(256)	
Array	VARCHAR(256)	

NOTE: If you are appending to an existing table where a column's Trifacta data type is not mapped to the column data type in the target table, a validation error is thrown, as the Trifacta platform writes unmapped types as String data type.

SharePoint Data Type Conversions

This section covers data type conversions between the Trifacta® application and SharePoint.

NOTE: The Trifacta® data types listed in this page reflect the raw data type of the converted column. Depending on the contents of the column, the Transformer Page may re-infer a different data type, when a dataset using this type of source is loaded.

Access/Read

NOTE: Image, Link, and Address fields from SharePoint are imported into Trifacta as JSON strings. Incorrect formatting or modification of these strings within the product may result in errors when published back to SharePoint Lists.

Source Data Type	Supported?	Trifacta Data Type	Notes
Choice (menu)	Y	String	
Currency	Y	Float	
Date and Time	Y	Datetime	
Link and Image	Y	String (JSON)	
Lookup	Y	String	
Address	Y	String (JSON)	
Multiple lines of text	Y	String	Lines in the source are demarcated in the String value using the newline character.
Number	Y	Float	
Person or Group	Y	String (JSON)	
Single line of text	Y	String	
Task outcome	Y	String	
Yes/No	Y	Boolean	

Publish/Write

Trifacta Data Type	SharePoint List Data Types	Notes
String	String Note Text Varchar URL Choice	NOTE: Secondary types are published only if appending or truncating an existing table. See below.
Integer	Integer	
Decimal	Float Currency	

Datetime	Date and Time	
Bool	Boolean	
Map	Note String Text Varchar	NOTE: Secondary types are published only if appending or truncating an existing table. See below.
Array	Note String Text Varchar	NOTE: Secondary types are published only if appending or truncating an existing table. See below.
All other data types	String	

For more information, see http://cdn.cdata.com/help/RSF/jdbc/pg_datatypemapping.htm.

Depending on the publishing action, output data types may vary.

Appending or truncating a SharePoint List:

If you are appending or truncating an existing SharePoint List, all column data types are published to the target data types as expected, as long as the data being published can be parsed and published by SharePoint.

Writing to a new SharePoint List:

Some data types in Trifacta do not have a direct mapping to SharePoint List data types. Examples:

- Address, Link, Person, Choice, and Image data types are published to SharePoint Lists as a multi-line String.
- Currency is published as a Number.
- Datetime values in Trifacta are written back to SharePoint Lists as DateTime types, even if the source data from SharePoint is Date type.

Transformation Reference

Contents:

- *Scale to min max*
- *One hot encode*
- *Scale to mean*
- *Bin column*
- *Change column type*
- *Comment*
- *Conditional column*
- *Convert patterns*
- *Count matches*
- *Count matches between delimiters*
- *Remove duplicate rows*
- *New formula*
- *Delete columns*
- *Extract between delimiters*
- *Extract text or pattern*
- *Extract first*
- *Convert key/value to Object*
- *Extract last*
- *Extract matches to Array*
- *Extract between positions*
- *Extract mismatched*
- *Extract numbers*
- *Extract query strings*
- *Filter contains*
- *Filter custom formula*
- *Filter ends with*
- *Filter exact*
- *Filter not equals*
- *Filter from top*
- *Filter greater than*
- *Filter at interval*
- *Filter less than*
- *Filter missing*
- *Filter mismatched*
- *Filter in*
- *Filter range*
- *Filter starts with*
- *Expand Array to rows*
- *Group by*
- *Join datasets*
- *Lock type*
- *Lowercase text*
- *Pad with leading*
- *Merge columns*
- *Move columns*
- *Nest columns*
- *Pivot*
- *Prefix text*
- *Propercase text*
- *Remove symbols in text*
- *Remove whitespace in text*
- *Remove accents in text*
- *Rename columns*
- *Rename with pattern*

- *Rename with prefix*
- *Rename with row(s)*
- *Rename with suffix*
- *Rename to UPPERCASE*
- *Rename to lowercase*
- *Rename from beginning*
- *Rename from end*
- *Rename by removing special characters*
- *Replace cells*
- *Replace text or pattern*
- *Replace between delimiters*
- *Replace between positions*
- *Replace mismatched*
- *Replace missing*
- *Select*
- *Edit with formula*
- *Sort rows*
- *Split on text or pattern*
- *Split between delimiters*
- *Split with multiple delimiters*
- *Split at positions*
- *Split at interval*
- *Split between positions*
- *Split into rows*
- *Suffix text*
- *Trim whitespace*
- *Trim quotes*
- *Invoke external function*
- *Uppercase text*
- *Date format*
- *Union datasets*
- *Standardize column*
- *Create column from examples*
- *Unlock type*
- *Unnest elements*
- *Unpivot*
- *Convert values to columns*
- *Window*
- *sourcerownumber*
- *filepath*

This section contains reference information on the transformations available in Trifacta®.

Tip: Use the values in the Title column as search strings in the Search panel to begin specifying these transformations.

Name	Title	Description
scaleminmax	Scale to min max	Scale a column to a specific min max range. See <i>Prepare Data for Machine Processing</i> .
onehotencode	One hot encode	Create a column for each unique value indicating its presence or absence. See <i>Prepare Data for Machine Processing</i> .
scalestandar dize	Scale to mean	Scale a column to zero mean and unit variance. See <i>Prepare Data for Machine Processing</i> .
bincolumn	Bin column	Bin values into ranges of equal or custom size. See <i>Prepare Data for Machine Processing</i> .

changetype	Change column type	Changes the data type of a column [settype]. See <i>Change Column Data Type</i> .
comment	Comment	Adds a comment to your recipe [comment]. See <i>Add Comments to Your Recipe</i> .
conditions	Conditional column	Returns values based on conditions such as if-then-else or case statements. See <i>Apply Conditional Transformations</i> .
convertpattern	Convert patterns	Finds one or more patterns or text literals and replaces them with specified pattern values. See <i>Standardize Using Patterns</i> .
countmatches	Count matches	Counts the number of matches [countpattern]. See <i>Compute Counts</i> .
countmatchbetween	Count matches between delimiters	Counts the number of matches [countpattern]. See <i>Compute Counts</i> .
deduplicate	Remove duplicate rows	Removes duplicate rows where values in every column are the same. See <i>Deduplicate Data</i> .
derive	New formula	Creates a new column with the result of a formula.
drop	Delete columns	Delete one or more columns. See <i>Remove Data</i> .
extractbetween delimiters	Extract between delimiters	Extracts text found between two patterns. See <i>Extract Values</i> .
extractcustom	Extract text or pattern	Extracts text found between two patterns. Variant: Custom text or pattern. See <i>Extract Values</i> .
extractfirst characters	Extract first	Extracts text according to its position. Variant: Extract the first n characters. See <i>Extract Values</i> .
extractkv	Convert key/value to Object	Extracts key-value pairs into an Object [extractkv]. See <i>Extract Values</i> .
extractlast characters	Extract last	Extracts key-value pairs into an Object [extractkv]. Variant: Extract the last n characters. See <i>Extract Values</i> .
extractlist	Extract matches to Array	Extracts a list into an Array [extractlist]. See <i>Extract Values</i> .
extractrange of characters	Extract between positions	Extracts text according to its position. Variant: Extract the last n characters. See <i>Extract Values</i> .
extractmismatched	Extract mismatched	Extracts a list into an Array [extractlist]. Variant: The data type to match against. See <i>Extract Values</i> .
extractnumbers	Extract numbers	Extracts a list into an Array [extractlist]. Variant: Extract numbers from a text. See <i>Extract Values</i> .
extractquery strings	Extract query strings	Extracts a list into an Array [extractlist]. Variant: Extract fields from an URL query string. See <i>Extract Values</i> .
filtercontains	Filter contains	Filter rows that satisfy a condition. Variant: Filter rows that contain a specified value or pattern. See <i>Filter Data</i> .
filtercustom	Filter custom formula	Filter rows that satisfy a condition. Variant: Filter rows that satisfy an arbitrary formula. See <i>Filter Data</i> .
filterendswith	Filter ends with	Filter rows that satisfy a condition. Variant: Filter rows that ends with a specified value or pattern. See <i>Filter Data</i> .
filterexactly	Filter exact	Filter rows that satisfy a condition. Variant: Filter rows that match exactly a specified value. See <i>Filter Data</i> .
filternot	Filter not equals	Filters rows that do not satisfy a condition. See <i>Filter Data</i> .
filterfromtop	Filter from top	Filter rows by their position. Variant: Filter rows from the top. See <i>Filter Data</i> .
filtergreaterthan	Filter greater than	Filter rows that satisfy a condition. Variant: Filter rows with values greater than (or

an		equal to) a specified value. See <i>Filter Data</i> .
filterinterval	Filter at interval	Filter rows by their position. Variant: . Variant: The size of the interval to filter rows at. See <i>Filter Data</i> .
filterlessthan	Filter less than	Filter rows that satisfy a condition. Variant: Filter rows with values less than (or equal to) a specified value. See <i>Filter Data</i> .
filtermissing	Filter missing	Filter rows that satisfy a condition. Variant: Filter rows with missing values. See <i>Remove Data</i> .
filtermismatched	Filter mismatched	Filter rows that satisfy a condition. Variant: Filter rows with mismatched values. See <i>Filter Data</i> .
filteroneof	Filter in	Filter rows that satisfy a condition. Variant: Filter rows that match any of the specified values. See <i>Filter Data</i> .
filterrange	Filter range	Filter rows by their position. Variant: Filter rows within a range. See <i>Filter Data</i> .
filterstartswith	Filter starts with	Filter rows that satisfy a condition. Variant: Filter rows that starts with a specified value or pattern. See <i>Filter Data</i> .
flatten	Expand Array to rows	Converts each element in an Array into a new row. See <i>Working with Arrays</i> .
groupby	Group by	Group data and perform aggregated calculations on it. See <i>Create Aggregations</i> .
join	Join datasets	Adds additional columns from other data sources [join]. See <i>Join Window</i> .
locktype	Lock type	Lock column to current type.
lowercase	Lowercase text	Format text in columns. Variant: Convert text in column to lowercase. See <i>Modify String Values</i> .
leftpad	Pad with leading	Format text in columns. Variant: Add the necessary number of characters to each value to make them of the same length. See <i>Modify String Values</i> .
merge	Merge columns	Concatenates the values from two or more columns into a new column [merge]. See <i>Add Two Columns</i> .
move	Move columns	Moves one or more columns before or after another column [move]. See <i>Move Columns</i> .
nest	Nest columns	Converts columns into an Object or Array [nest]. See <i>Working with Arrays</i> .
pivot	Pivot	Creates a new column for each unique value in a column [pivot]. See <i>Pivot Data</i> .
prefix	Prefix text	Format text in columns. Variant: Specify a prefix to be added at the beginning of each selected column name. See <i>Modify String Values</i> .
propercase	Propercase text	Format text in columns. Variant: Convert text in column to ProperCase. See <i>Modify String Values</i> .
removesymbols	Remove symbols in text	Format text in columns. Variant: Remove all non-alphanumeric characters from the text. See <i>Remove Data</i> .
removewhitespace	Remove whitespace in text	Format text in columns. Variant: Remove all whitespace found in the text. See <i>Remove Data</i> .
removeaccents	Remove accents in text	Remove accent marks from text. See <i>Modify String Values</i> .
rename	Rename columns	Renames one or more columns [rename]. See <i>Rename Columns</i> .
renamepattern	Rename with pattern	Renames one or more columns [rename]. See <i>Rename Columns</i> .
renameprefix	Rename with prefix	Renames one or more columns [rename]. See <i>Rename Columns</i> .
renameheader	Rename with row(s)	Renames one or more columns [rename]. See <i>Rename Columns</i> .
renamesuffix	Rename with suffix	Renames one or more columns [rename]. See <i>Rename Columns</i> .

renameupper	Rename to UPPERCASE	Renames one or more columns [rename]. See <i>Rename Columns</i> .
renamelower	Rename to lowercase	Renames one or more columns [rename]. See <i>Rename Columns</i> .
renamekeepleft	Rename from beginning	Renames one or more columns [rename]. See <i>Rename Columns</i> .
renamekeepright	Rename from end	Renames one or more columns [rename]. See <i>Rename Columns</i> .
renamesanitize	Rename by removing special characters	Renames one or more columns [rename]. See <i>Rename Columns</i> .
replacecell	Replace cells	Renames one or more columns [rename]. See <i>Rename Columns</i> .
replacepattern	Replace text or pattern	Replace text matching a pattern. See <i>Replace Cell Values</i> .
replacebetweenpatterns	Replace between delimiters	Replace text between delimiters. Variant: Replace text between delimiters. See <i>Replace Cell Values</i> .
replacebetweenpositions	Replace between positions	Replace text between delimiters. Variant: Replaces text based on position. See <i>Replace Cell Values</i> .
replacemismatched	Replace mismatched	Replace mismatched values. See <i>Replace Cell Values</i> .
replacemissing	Replace missing	Replace missing values. See <i>Replace Cell Values</i> .
select	Select	Create a new table of columns <i>Selectd</i> from your current dataset. See <i>Select</i> .
set	Edit with formula	Sets the values of one or more columns to the result of a formula [set].
sort	Sort rows	Sorts the rows based on the values in one or more columns.
splitondelimiter	Split on text or pattern	Split by delimiter. Variant: Text or pattern. See <i>Split Column</i> .
splitbetween delimiters	Split between delimiters	Split by delimiter. Variant: Between two delimiters. See <i>Split Column</i> .
splitmultiple delimiters	Split with multiple delimiters	Split by delimiter. Variant: By multiple delimiters. See <i>Split Column</i> .
splitpositions	Split at positions	Split by character position. Variant: By positions. See <i>Split Column</i> .
splitinterval	Split at interval	Split by character position. Variant: At regular interval. See <i>Split Column</i> .
splitbetween positions	Split between positions	Split by character position. Variant: Between two positions. See <i>Split Column</i> .
splitrows	Split into rows	Splits raw data into rows [splitrows]. See <i>Split Column</i> .
suffix	Suffix text	Format text in columns. Variant: Specify a suffix to be added to the end of each selected column name. See <i>Modify String Values</i> .
trimwhitespace	Trim whitespace	Format text in columns. Variant: Remove all whitespaces found at the beginning and end of the text. See <i>Modify String Values</i> .
trimquotes	Trim quotes	Format text in columns. Variant: Remove quotes found at the beginning and end of the text. See <i>Modify String Values</i> .
udf	Invoke external function	Creates a new column with the result of an external function. <div>NOTE: This transformation requires additional configuration.</div>

uppercase	Uppercase text	Format text in columns. Variant: Convert text in column to UPPERCASE. See <i>Modify String Values</i> .
dateformat	Date format	Change format for Datetime columns. See <i>Format Dates</i> .
union	Union datasets	Adds additional rows from other data source [union]. See <i>Union Page</i> .
standardize	Standardize column	Single-column standardization for standardizing column values. See <i>Standardize Page</i> .
columnbyexample	Create column from examples	Create a new column by providing example values. See <i>Create Column by Example</i> .
unlocktype	Unlock type	Unlock column type.
unnest	Unnest elements	Extracts elements from an Object or Array into columns. See <i>Working with Arrays</i> .
unpivot	Unpivot	Turns columns into rows. Produces a key column with unnested values. See <i>Pivot Data</i> .
valuestocols	Convert values to columns	Creates a new column for each unique value in a column [valuestocols]. See <i>Pivot Data</i> .
window	Window	Performs row-based calculations across multiple ordered rows [window]. See <i>Window Functions</i> .
sourcerownumber	sourcerownumber	Generate a new column containing the row number for each row from the source, if available. See <i>Source Metadata References</i> .
filepath	filepath	Generate a new column containing the path to the source file, if available. See <i>Source Metadata References</i> .

Plan Metadata References

Contents:

- *General Syntax*
 - *\$plan References*
 - *\$http References*
 - *Response references*
 - *\$slack References*
 - *\$delete References*
 - *\$flow References*
 - *Output references*
 - *Exploring Metadata*
 - *Metadata functions*
 - *Metadata structure*
 - *Parameter References*
 - *Flow parameters*
 - *Environment parameters*
 - *Additional References*
-

In the body and header of HTTP tasks in your plans, you can reference the following elements of metadata from the plan run for additional contextual information.

General Syntax

All plan metadata references follow the following basic syntax:

```
{{ $plan.path.to.reference }}
```

- All references can be entered with \$ in the Trifacta application. These references are turned into {{ \$ in the code definition. The double-curly braces forms the environment for metadata replacement.

Tip: In the Trifacta application, you can start by typing \$.

- Nodes in the tree are separated with a . period.

Reference values that contain whitespace must be listed in the following manner:

```
{{ $plan.path['path with white space in it'].rest.of.path }}
```

Notes:

- In the Trifacta application, you can use double-quotes when specifying a whitespace value. However, these double-quotes get escaped in the actual request. It is safer and more consistent to use single quotes.

Whitespace values typically appear when referencing the display name values for underlying objects, like recipes executed as part of a flow task.

\$plan References

These references apply to the plan definition or current plan run.

Text to enter:

```
$plan.
```

Reference	Description
name	Name of the plan that is run.
duration	Length of time that the plan ran or has run so far <div>Tip: To return a more readable form of this duration value, use the following reference: <pre>{{ \$plan.duration humanizeDuration }}</pre></div>
startTime	Timestamp for when the plan run began
runId	Internal identifier for this run of the plan
user	Internal identifier of the user who launched this run.
taskCount	Count of tasks in the plan run.

\$http References

These references apply to HTTP tasks in the plan run.

Enter the following, after which you can see the two-letter codes for the HTTP tasks that have already executed in the current plan run:

```
$http_ax.
```

Reference	Description
name	Name of the HTTP task
status	Current status of the task execution
duration	Length of time that the task ran or has run so far
startTime	Timestamp for when the task began. A null value if the task has not begun.
endTime	Timestamp for when the task ended. A null value if it has not ended yet.
statusCode	Status code (if any) returned from the receiving endpoint
response	Response information. See below.

Response references

These references apply to the response returned as part of the task execution.

Enter the following, after which you can see the two-letter codes for the HTTP tasks that have already executed in the current plan run:

```
$http_ax.response.
```

Reference	Description
body	Body of the response
json	JSON-formatted version of the response
headers	Headers returned with the response

\$slack References

You can reference metadata from Slack tasks in the current plan run using the following reference types:

```
$slack_ax.
```

Supported metadata is identical to the metadata for HTTP tasks. See the previous section for details.

\$delete References

You can reference metadata from Delete tasks in the current plan run using the following reference types:

```
$delete_92.
```

Supported metadata is identical to the metadata for HTTP tasks. See the previous section for details.

Reference	Description
name	Name of the HTTP task
status	Current status of the task execution
path	Full path to the file(s) or folder that was deleted.
duration	Length of time that the task ran or has run so far
startTime	Timestamp for when the task began. A null value if the task has not begun.
endTime	Timestamp for when the task ended. A null value if it has not ended yet.
deletedFilesCount	Number of files that were successfully deleted.

\$flow References

These references apply to flow tasks in the plan run.

Enter the following, after which you can see the two-letter codes for the HTTP tasks that have already executed in the current plan run:

```
$flow_ax.
```

Reference	Description
name	Name of the flow task
status	Current status of the task execution

duration	Length of time that the task ran or has run so far
startTime	Timestamp for when the task began. A null value if the task has not begun.
endTime	Timestamp for when the task ended. A null value if it has not ended yet.
jobIds	Internal identifiers for the jobs that were run as part of this flow task
flowName	Name of the flow underlying this flow task
output	Metadata from the flow task's output. See below.
params	Parameters created in the flow can be referenced in the task.

Output references

These references apply to the outputs that are generated in the flow tasks of the plan run.

Enter the following for flow task 7p with output My Output Name:

```
$flow_7p['My Output Name'].
```

Reference	Description
name	Name of the flow
status	Current status of the flow
duration	Length of time that the flow execution ran or has run so far
startTime	Timestamp for when the flow execution began. A null value if the run has not begun.
endTime	Timestamp for when the flow execution ended. A null value if it has not ended yet.
lastUpdate	Timestamp for when the flow was last modified
jobIds	Internal identifier(s) for the job that was run or is running for the flow. Can contain multiple identifiers.
user	Internal identifier for the user who executed the job
jobType	The type of job that was executed. Values: <ul style="list-style-type: none"> • manual - executed through the Trifacta application • scheduled - executed according to a defined schedule • api - executed via API
fileSize	If the output generates a file or files, this value captures the size in KB of the output.
environment	Running environment where the job was executed
columnCount	Count of columns generated in the output
rowCount	Count of rows generated in the output
dataTypeCount	Count of Trifacta data types detected in the output
validValuesCount	Count of valid values in the output
mismatchedValuesCount	Count of mismatched values in the output
emptyValuesCount	Count of missing or empty values in the output
columns	Column information from the selected output for the flow. See below.
sources	Source filename and table information from the imported datasets. See below.
publishing actions	Data on the publishing actions defined for the output. See below.

Output column references

Tip: The metrics calculated for output columns may also appear in profiles of your output data.

These references are available for output columns in the following syntax:

```
$flow_7p['My Output Name'].output.
```

Reference	Description
name	Column name
type	Data type of column
validValuesCount	Count of valid values in the column
mismatchedValuesCount	Count of mismatched values in the column
emptyValuesCount	Count of empty values in the column
topValues	List of top values in the column
minimumValue	Lowest value in the column
lowerQuartileValue	25th percentile value in the column
medianValue	50th percentile value in the column
upperQuartileValue	75th percentile value in the column
maximumValue	Maximum value in the column

Sources references

These references apply to the datasource files or tables that were used to generate the output.

Reference	Description
name	Name of the datasource file or table

Publishing actions references

These references apply to the publishing actions defined for the output.

Reference	Description
name	Name of the publishing action
action	Definition of the publishing action
location	Location where the publishing action is published
type	Type of publishing action

Exploring Metadata

You can use the following functions and techniques to further explore the metadata returned from your plan execution.

Metadata functions

The following functions can be applied to select metadata references to further filter the raw values.

humanizeDuration

In raw form, the `duration` metadata references return values that look like the following:

```
PT7.523S
```

You can apply the `HumanizeDuration` function to render the above into a more readable format:

```
{{ $plan.duration | humanizeDuration }}
```

The output of the above applied to the first value is the following:

```
7.523 seconds
```

uuid

You can generate a universally unique identifier, which can be delivered as part of a messaging payload:

```
{{ uuid() }}
```

Metadata structure

Some metadata references return complex or nested objects, which may return data that looks like the following:

```
[object Object],[object Object],[object Object],[object Object],[object Object],[object Object],[object Object]
```

In these cases, the nested data inside the object is not exposed by the basic reference. To explore further, you can use either of the following solutions:

Text: Create a for loop through the returned objects:

```
{% for value in $flow_6f.Job.columns.MyColumnName.topValues %}{{value.key}}, {{value.count}} {% endfor %}
```

JSON: add the `| dump` command to the end of your reference. You can modify the following example reference and try to insert in the Body textbox:

```
{{ $flow_6f.Job.columns.MyColumnName.topValues | dump | replace("'", '\"') }}
```

Tip: When the data is returned, you can use loop structures to retrieve specific values for display.

For more information, see <https://mozilla.github.io/nunjucks/templating.html#dump>.

Parameter References

Flow parameters

To reference flow parameters as part of a flow task, use the following pattern:

```
$plan.params.<myParamName>
```

Environment parameters

You can reference environment parameters by name in your plan task metadata. For example:

```
env.MyEnvironmentParameter1
```

Additional References

Plan metadata reference information leverages the Nunjucks templating language, which provides additional capabilities such as loops, conditions, filters, and helper functions.

NOTE: These additional capabilities are available through the language, but their implementation in the Trifacta application has not been certified. For Nunjucks capabilities not listed on this page, you should experiment with them in a development environment first.

For more information, see <https://mozilla.github.io/nunjucks/templating.html>.

cron Schedule Syntax Reference

Contents:

- Overview of cron
 - Cron syntax
 - Special characters
- Examples
 - Hourly
 - Daily
 - Weekly
 - Weekdays
 - Monthly
 - Yearly
 - Other examples
 - Unsupported cron expressions
 - Invalid cron expressions

This section describes the syntax for defining scheduled executions using cron in Trifacta®. Typically, this method is used for repeated schedules.

Flow schedules:

- Flow owners can define scheduled executions of flows from within the Flow View page.
- Collaborators can review and cannot edit schedules.

NOTE: Time zone settings defined in the Trifacta application page where you are specifying your cron schedule are used with the schedule. To use UTC time zone, select UTC in the drop-down.

Overview of cron

Trifacta allows you to make use of cron, a widely used syntax, for specifying times that recur at regular intervals. You can use cron to specify schedules on a per-minute or annual basis and arbitrary intervals in between.

Cron syntax

A cron scheduled is defined as a space-separated string of values. The following cron example defines a schedule to be triggered at 11:30:00pm on February 1:

minute	hour	day of month	month	day of week
30	23	1	2	*

When all values are matched, the cron job is triggered.

NOTE: Specification of seconds is not supported.

Wildcards:

In the above cron expression, the wildcard * can be used to match any accepted value, which means that the cron value type is not a factor in determining this schedule. Since the wildcard is applied to the day of week value, the schedule can be triggered on any day of the week.

NOTE: You must use the * character in either the day-of-week or day-of-month fields. Specifying both fields in the same cron expression is not supported.

Legend:

Except for the final field (year), all fields are required in the cron expression. Special characters are described below the table.

Value	Type	Description	Supported Special Characters
30	minute	0-59	, - * /
23	hour	0-23	, - * /
1	day of month	1-31	, - * / L W
2	month	1-12	, - * /
*	day of week	0 - 6 or Sun - Sat 0, Sun, SUN = Sunday 1, Mon, MON = Monday ... 6, Sat, SAT = Saturday	, - * / L #
*	year	(Optional) You can specify year settings if needed. Default is *.	, - * /

Special characters

You can use the following special characters in your cron expressions.

Character	Description
*	("all values") - Wildcard to match all possible values in the field. For example, if you wanted your trigger to fire every minute of the 10pm hour, the minute character in the expression is *. An example is below.
-	Specify a range of values. For example, you could use 1 – 5 in the day-of-week field to match the work days of the week (Monday through Friday). An example is below.
,	Specify a discrete set of values. For example, an entry of 1 , 10 , 20 , 30 for the day of month field is triggered on the 1st, 10th, 20th, and 30th (if possible) of the month.
/	Specify increments of the field in the units of the field. For example, 5 / 20 in the minutes field matches on the 5th, 25th, and 45th minute of each hour.
L	<p>Last value accepted in the range is accepted in the following fields:</p> <ul style="list-style-type: none"> Day-of-month: Specifies the last day of the month for the currently selected month value. <ul style="list-style-type: none"> In January, this value matches with 31. In February, this value matches with 28 for non-leap years. In April, this value matches with 30. Day-of-week: <ul style="list-style-type: none"> By itself, it specifies the last day of the week, which matches with 6 (Saturday). When used with another value, it specifies the last matching value for the month. For example, 3L is the last Wednesday of the month.
W	<p>Specifies the nearest matching weekday. For example, an entry of 22W in the day-of-month field matches on the nearest weekday to the 22nd of the month. If the 22nd is a Saturday, then the cron job matches on the 24th (the following Monday).</p> <div> <p>Tip: LW can be used in the day-of-month field to match on the last weekday of the month.</p> </div>

#	Specifies the nth day of the month. Examples for the day-of-week field: <ul style="list-style-type: none"> 3#4 - fourth Tuesday of the month 5#2 - second Thursday of the month
---	---

Examples

Below are some example cron schedules.

Hourly

Runs at minute 15 of every hour:

```
15 * * * *
```

Daily

Runs every day at 10pm:

```
0 22 * * *
```

Runs every minute of the 10pm hour every day:

```
* 22 * * *
```

Weekly

Runs every Tuesday at 3am:

```
0 3 * * 2
```

Weekdays

Runs each weekday at 8pm:

```
0 20 * * 1-5
```

Note that the above schedule runs at 10pm on Monday night and each night of the week at that time.

To refresh the flow for each weekday morning, you might choose to start the schedules on Sunday, in which the day-of-week value starts with 0 and ends with 4.

Monthly

Runs the first day of each month at 2:30am:

```
30 2 1 * *
```

Runs at 3:30pm on the nearest weekday (W) to the 25th of the month:

```
30 15 25W * *
```

- If the 25th is a Saturday, the above triggers on Friday the 24th.
- If the 25th is a Sunday, the above triggers on Monday the 26th.

Yearly

Runs at midnight of January 1 each year:

```
0 0 1 1 * *
```

Other examples

Expression	Meaning
0 12 * * *	Fire at 12pm (noon) every day
15 10 * * *	Fire at 10:15am every day
15 10 * * *	Fire at 10:15am every day
15 10 * * * *	Fire at 10:15am every day
15 10 * * * 2017	Fire at 10:15am every day during the year 2017
* 14 * * *	Fire every minute starting at 2pm and ending at 2:59pm, every day
0/5 14 * * *	Fire every 5 minutes starting at 2pm and ending at 2:55pm, every day
0/5 14,18 * * *	Fire every 5 minutes starting at 2pm and ending at 2:55pm, AND fire every 5 minutes starting at 6pm and ending at 6:55pm, every day
0-5 14 * * *	Fire every minute starting at 2pm and ending at 2:05pm, every day
10,44 14 * 3 WED	Fire at 2:10pm and at 2:44pm every Wednesday in the month of March.
15 10 * * MON-FRI	Fire at 10:15am every Monday, Tuesday, Wednesday, Thursday and Friday
15 10 15 * *	Fire at 10:15am on the 15th day of every month

15 10 L * *	Fire at 10:15am on the last day of every month
15 10 L-2 * *	Fire at 10:15am on the 2nd-to-last last day of every month
15 10 * * 5L	Fire at 10:15am on the last Friday of every month
15 10 * * 5L 2017-2019	Fire at 10:15am on every last friday of every month during the years 2017, 2018 and 2019
15 10 * * 5#3	Fire at 10:15am on the third Friday of every month
0 12 1/5 * *	Fire at 12pm (noon) every 5 days every month, starting on the first day of the month.
11 11 11 11 *	Fire every November 11th at 11:11am.

Unsupported cron expressions

NOTE: The Trifacta application does not support mixing / and – special characters in the same expressions.

Instead of expressing ranges in your cron syntax, you can reference all possible options.

Invalid expression	Valid expression
0 23 * 1-11/2 * *	0 23 * 2,4,6,8,10 * *

Invalid cron expressions

Expression	Meaning	Reason
15 10 * * * 2001	Fire at 10:15am every day during the year 2001	This cron expression is invalid because it will not generate any events in the future.
* * *	-	The cron expression should contain 6 or 7 fields.

Supported Time Zone Values

American Time Zones

The following American time zones are mapped to the time zone values supported in Trifacta®:

Tip: Trifacta does make adjustments for Daylight Savings Time.

US Time Zone	Time Zone Value	Standard Time (UTC)	Daylight Savings Time (UTC)	Included States and Territories	Notes
Atlantic	America / Puerto Rico	UTC-04:00	n/a	Puerto Rico, US Virgin Islands	Daylight Savings Time is not observed.
Eastern	US / Eastern	UTC-05:00	UTC-04:00	Entire: Connecticut, Delaware, Georgia, Maine, Maryland, Massachusetts, New Hampshire, New Jersey, New York, North Carolina, Ohio, Pennsylvania, Rhode Island, South Carolina, Vermont, Virginia, West Virginia Partial: Florida, Indiana, Kentucky, Michigan, Tennessee	
Central	US / Central	UTC-06:00	UTC-05:00	Entire: Alabama, Arkansas, Illinois, Iowa, Louisiana, Minnesota, Mississippi, Missouri, Oklahoma, Wisconsin Partial: Florida, Indiana, Kansas, Kentucky, Michigan, Nebraska, North Dakota, South Dakota, Tennessee, Texas	
Mountain	US / Mountain	UTC-07:00	UTC-06:00	Entire: Arizona, Colorado, Montana, New Mexico, Utah, Wyoming Partial: Idaho, Kansas, Nebraska, Nevada, North Dakota, Oregon, South Dakota, Texas	Most of Arizona does not observe Daylight Savings Time. Use US / Arizona in this area.
Pacific	US / Pacific	UTC-08:00	UTC-07:00	Entire: California, Washington Partial: Idaho, Nevada, Oregon	
Alaska	US / Alaska	UTC-09:00	UTC-08:00	Partial: Alaska	Daylight Savings Time is not observed in the Aleutian Islands. Use US / Aleutian in this area.
Hawaii	US / Hawaii	UTC-10:00	UTC-09:00	Entire: Hawaii Partial: Alaska	Daylight Savings Time is not observed in Hawaii.

Global Time Zone Values

For the functions that support use of specified time zones, you can apply the following string values as parameters to specify the time zone:

Time Zone Value	Standard time (UTC)	Daylight Saving time (UTC)	Partial country coverage

Africa/Abidjan	UTC+00:00	UTC+00:00	
Africa/Accra	UTC+00:00	UTC+00:00	
Africa/Addis_Ababa	UTC+03:00	UTC+03:00	
Africa/Algiers	UTC+01:00	UTC+01:00	
Africa/Asmara	UTC+03:00	UTC+03:00	
Africa/Bamako	UTC+00:00	UTC+00:00	
Africa/Bangui	UTC+01:00	UTC+01:00	
Africa/Banjul	UTC+00:00	UTC+00:00	
Africa/Bissau	UTC+00:00	UTC+00:00	
Africa/Blantyre	UTC+02:00	UTC+02:00	
Africa/Brazzaville	UTC+01:00	UTC+01:00	
Africa/Bujumbura	UTC+02:00	UTC+02:00	
Africa/Cairo	UTC+02:00	UTC+02:00	
Africa/Casablanca	UTC+01:00	UTC+01:00	
Africa/Ceuta	UTC+01:00	UTC+01:00	Ceuta, Melilla
Africa/Conakry	UTC+00:00	UTC+00:00	
Africa/Dakar	UTC+00:00	UTC+00:00	
Africa/Dar_es_Salaam	UTC+03:00	UTC+03:00	
Africa/Djibouti	UTC+03:00	UTC+03:00	
Africa/Douala	UTC+01:00	UTC+01:00	
Africa/El_Aaiun	UTC+00:00	UTC+01:00	
Africa/Freetown	UTC+00:00	UTC+00:00	
Africa/Gaborone	UTC+02:00	UTC+02:00	
Africa/Harare	UTC+02:00	UTC+02:00	
Africa/Johannesburg	UTC+02:00	UTC+02:00	
Africa/Juba	UTC+03:00	UTC+03:00	
Africa/Kampala	UTC+03:00	UTC+03:00	
Africa/Khartoum	UTC+02:00	UTC+02:00	
Africa/Kigali	UTC+02:00	UTC+02:00	
Africa/Kinshasa	UTC+01:00	UTC+01:00	Dem. Rep. of Congo (west)
Africa/Lagos	UTC+01:00	UTC+01:00	
Africa/Libreville	UTC+01:00	UTC+01:00	
Africa/Lome	UTC+00:00	UTC+00:00	
Africa/Luanda	UTC+01:00	UTC+01:00	
Africa/Lubumbashi	UTC+02:00	UTC+02:00	Dem. Rep. of Congo (east)
Africa/Lusaka	UTC+02:00	UTC+02:00	
Africa/Malabo	UTC+01:00	UTC+01:00	
Africa/Maputo	UTC+02:00	UTC+02:00	
Africa/Maseru	UTC+02:00	UTC+02:00	
Africa/Mbabane	UTC+02:00	UTC+02:00	

Africa/Mogadishu	UTC+03:00	UTC+03:00	
Africa/Monrovia	UTC+00:00	UTC+00:00	
Africa/Nairobi	UTC+03:00	UTC+03:00	
Africa/Ndjamena	UTC+01:00	UTC+01:00	
Africa/Niamey	UTC+01:00	UTC+01:00	
Africa/Nouakchott	UTC+00:00	UTC+00:00	
Africa/Ouagadougou	UTC+00:00	UTC+00:00	
Africa/Porto-Novo	UTC+01:00	UTC+01:00	
Africa/Sao_Tome	UTC+00:00	UTC+00:00	
Africa/Timbuktu	UTC+00:00	UTC+00:00	
Africa/Tripoli	UTC+02:00	UTC+02:00	
Africa/Tunis	UTC+01:00	UTC+01:00	
Africa/Windhoek	UTC+02:00	UTC+02:00	
America/Adak	UTC10:00	UTC09:00	Aleutian Islands
America/Anchorage	UTC09:00	UTC08:00	Alaska (most areas)
America/Anguilla	UTC04:00	UTC04:00	
America/Antigua	UTC04:00	UTC04:00	
America/Araguaina	UTC03:00	UTC03:00	Tocantins
America/Argentina /Buenos_Aires	UTC03:00	UTC03:00	Buenos Aires (BA, CF)
America/Argentina /Catamarca	UTC03:00	UTC03:00	Catamarca (CT); Chubut (CH)
America/Argentina /ComodRivadavia	UTC03:00	UTC03:00	
America/Argentina/Cordoba	UTC03:00	UTC03:00	Argentina (most areas: CB, CC, CN, ER, FM, MN, SE, SF)
America/Argentina/Jujuy	UTC03:00	UTC03:00	Jujuy (JY)
America/Argentina/La_Rioja	UTC03:00	UTC03:00	La Rioja (LR)
America/Argentina/Mendoza	UTC03:00	UTC03:00	Mendoza (MZ)
America/Argentina /Rio_Gallegos	UTC03:00	UTC03:00	Santa Cruz (SC)
America/Argentina/Salta	UTC03:00	UTC03:00	Salta (SA, LP, NQ, RN)
America/Argentina /San_Juan	UTC03:00	UTC03:00	San Juan (SJ)
America/Argentina/San_Luis	UTC03:00	UTC03:00	San Luis (SL)
America/Argentina/Tucuman	UTC03:00	UTC03:00	Tucuman (TM)
America/Argentina/Ushuaia	UTC03:00	UTC03:00	Tierra del Fuego (TF)
America/Aruba	UTC04:00	UTC04:00	
America/Asuncion	UTC04:00	UTC03:00	
America/Atikokan	UTC05:00	UTC05:00	EST - ON (Atikokan); NU (Coral H)
America/Atka	UTC10:00	UTC09:00	
America/Bahia	UTC03:00	UTC03:00	Bahia
America/Bahia_Banderas	UTC06:00	UTC05:00	Central Time - Bahia de Banderas
America/Barbados	UTC04:00	UTC04:00	

America/Belem	UTC03:00	UTC03:00	Para (east); Amapa
America/Belize	UTC06:00	UTC06:00	
America/Blanc-Sablon	UTC04:00	UTC04:00	AST - QC (Lower North Shore)
America/Boa_Vista	UTC04:00	UTC04:00	Roraima
America/Bogota	UTC05:00	UTC05:00	
America/Boise	UTC07:00	UTC06:00	Mountain - ID (south); OR (east)
America/Buenos_Aires	UTC03:00	UTC03:00	
America/Cambridge_Bay	UTC07:00	UTC06:00	Mountain - NU (west)
America/Campo_Grande	UTC04:00	UTC03:00	Mato Grosso do Sul
America/Cancun	UTC05:00	UTC05:00	Eastern Standard Time - Quintana Roo
America/Caracas	UTC04:00	UTC04:00	
America/Catamarca	UTC03:00	UTC03:00	
America/Cayenne	UTC03:00	UTC03:00	
America/Cayman	UTC05:00	UTC05:00	
America/Chicago	UTC06:00	UTC05:00	Central (most areas)
America/Chihuahua	UTC07:00	UTC06:00	Mountain Time - Chihuahua (most areas)
America/Coral_Harbour	UTC05:00	UTC05:00	
America/Cordoba	UTC03:00	UTC03:00	
America/Costa_Rica	UTC06:00	UTC06:00	
America/Creston	UTC07:00	UTC07:00	MST - BC (Creston)
America/Cuiaba	UTC04:00	UTC03:00	Mato Grosso
America/Curacao	UTC04:00	UTC04:00	
America/Danmarkshavn	UTC+00:00	UTC+00:00	National Park (east coast)
America/Dawson	UTC08:00	UTC07:00	Pacific - Yukon (north)
America/Dawson_Creek	UTC07:00	UTC07:00	MST - BC (Dawson Cr, Ft St John)
America/Denver	UTC07:00	UTC06:00	Mountain (most areas)
America/Detroit	UTC05:00	UTC04:00	Eastern - MI (most areas)
America/Dominica	UTC04:00	UTC04:00	
America/Edmonton	UTC07:00	UTC06:00	Mountain - AB; BC (E); SK (W)
America/Eirunepe	UTC05:00	UTC05:00	Amazonas (west)
America/El_Salvador	UTC06:00	UTC06:00	
America/Ensenada	UTC08:00	UTC07:00	
America/Fort_Nelson	UTC07:00	UTC07:00	MST - BC (Ft Nelson)
America/Fort_Wayne	UTC05:00	UTC04:00	
America/Fortaleza	UTC03:00	UTC03:00	Brazil (northeast: MA, PI, CE, RN, PB)
America/Glace_Bay	UTC04:00	UTC03:00	Atlantic - NS (Cape Breton)
America/Godthab	UTC03:00	UTC02:00	Greenland (most areas)
America/Goose_Bay	UTC04:00	UTC03:00	Atlantic - Labrador (most areas)
America/Grand_Turk	UTC05:00	UTC04:00	
America/Grenada	UTC04:00	UTC04:00	

America/Guadeloupe	UTC04:00	UTC04:00	
America/Guatemala	UTC06:00	UTC06:00	
America/Guayaquil	UTC05:00	UTC05:00	Ecuador (mainland)
America/Guyana	UTC04:00	UTC04:00	
America/Halifax	UTC04:00	UTC03:00	Atlantic - NS (most areas); PE
America/Havana	UTC05:00	UTC04:00	
America/Hermosillo	UTC07:00	UTC07:00	Mountain Standard Time - Sonora
America/Indiana/Indianapolis	UTC05:00	UTC04:00	Eastern - IN (most areas)
America/Indiana/Knox	UTC06:00	UTC05:00	Central - IN (Starke)
America/Indiana/Marengo	UTC05:00	UTC04:00	Eastern - IN (Crawford)
America/Indiana/Petersburg	UTC05:00	UTC04:00	Eastern - IN (Pike)
America/Indiana/Tell_City	UTC06:00	UTC05:00	Central - IN (Perry)
America/Indiana/Vevay	UTC05:00	UTC04:00	Eastern - IN (Switzerland)
America/Indiana/Vincennes	UTC05:00	UTC04:00	Eastern - IN (Da, Du, K, Mn)
America/Indiana/Winamac	UTC05:00	UTC04:00	Eastern - IN (Pulaski)
America/Indianapolis	UTC05:00	UTC04:00	
America/Inuvik	UTC07:00	UTC06:00	Mountain - NT (west)
America/Iqaluit	UTC05:00	UTC04:00	Eastern - NU (most east areas)
America/Jamaica	UTC05:00	UTC05:00	
America/Jujuy	UTC03:00	UTC03:00	
America/Juneau	UTC09:00	UTC08:00	Alaska - Juneau area
America/Kentucky/Louisville	UTC05:00	UTC04:00	Eastern - KY (Louisville area)
America/Kentucky/Monticello	UTC05:00	UTC04:00	Eastern - KY (Wayne)
America/Knox_IN	UTC06:00	UTC05:00	
America/Kralendijk	UTC04:00	UTC04:00	
America/La_Paz	UTC04:00	UTC04:00	
America/Lima	UTC05:00	UTC05:00	
America/Los_Angeles	UTC08:00	UTC07:00	Pacific
America/Louisville	UTC05:00	UTC04:00	
America/Lower_Princes	UTC04:00	UTC04:00	
America/Maceio	UTC03:00	UTC03:00	Alagoas, Sergipe
America/Managua	UTC06:00	UTC06:00	
America/Manaus	UTC04:00	UTC04:00	Amazonas (east)
America/Marigot	UTC04:00	UTC04:00	
America/Martinique	UTC04:00	UTC04:00	
America/Matamoros	UTC06:00	UTC05:00	Central Time US - Coahuila, Nuevo Leon, Tamaulipas (US border)
America/Mazatlan	UTC07:00	UTC06:00	Mountain Time - Baja California Sur, Nayarit, Sinaloa
America/Mendoza	UTC03:00	UTC03:00	
America/Menominee	UTC06:00	UTC05:00	Central - MI (Wisconsin border)
America/Merida	UTC06:00	UTC05:00	Central Time - Campeche, Yucatan

America/Metlakatla	UTC09:00	UTC08:00	Alaska - Annette Island
America/Mexico_City	UTC06:00	UTC05:00	Central Time
America/Miquelon	UTC03:00	UTC02:00	
America/Moncton	UTC04:00	UTC03:00	Atlantic - New Brunswick
America/Monterrey	UTC06:00	UTC05:00	Central Time - Durango; Coahuila, Nuevo Leon, Tamaulipas (most areas)
America/Montevideo	UTC03:00	UTC03:00	
America/Montreal	UTC05:00	UTC04:00	
America/Montserrat	UTC04:00	UTC04:00	
America/Nassau	UTC05:00	UTC04:00	
America/New_York	UTC05:00	UTC04:00	Eastern (most areas)
America/Nipigon	UTC05:00	UTC04:00	Eastern - ON, QC (no DST 196773)
America/Nome	UTC09:00	UTC08:00	Alaska (west)
America/Noronha	UTC02:00	UTC02:00	Atlantic islands
America/North_Dakota /Beulah	UTC06:00	UTC05:00	Central - ND (Mercer)
America/North_Dakota /Center	UTC06:00	UTC05:00	Central - ND (Oliver)
America/North_Dakota /New_Salem	UTC06:00	UTC05:00	Central - ND (Morton rural)
America/Ojinaga	UTC07:00	UTC06:00	Mountain Time US - Chihuahua (US border)
America/Panama	UTC05:00	UTC05:00	
America/Pangnirtung	UTC05:00	UTC04:00	Eastern - NU (Pangnirtung)
America/Paramaribo	UTC03:00	UTC03:00	
America/Phoenix	UTC07:00	UTC07:00	MST - Arizona (except Navajo)
America/Port_of_Spain	UTC04:00	UTC04:00	
America/Port-au-Prince	UTC05:00	UTC04:00	
America/Porto_Acre	UTC05:00	UTC05:00	
America/Porto_Velho	UTC04:00	UTC04:00	Rondonia
America/Puerto_Rico	UTC04:00	UTC04:00	
America/Punta_Arenas	UTC03:00	UTC03:00	Region of Magallanes
America/Rainy_River	UTC06:00	UTC05:00	Central - ON (Rainy R, Ft Frances)
America/Rankin_Inlet	UTC06:00	UTC05:00	Central - NU (central)
America/Recife	UTC03:00	UTC03:00	Pernambuco
America/Regina	UTC06:00	UTC06:00	CST - SK (most areas)
America/Resolute	UTC06:00	UTC05:00	Central - NU (Resolute)
America/Rio_Branco	UTC05:00	UTC05:00	Acre
America/Rosario	UTC03:00	UTC03:00	
America/Santa_Isabel	UTC08:00	UTC07:00	
America/Santarem	UTC03:00	UTC03:00	Para (west)
America/Santiago	UTC04:00	UTC03:00	Chile (most areas)
America/Santo_Domingo	UTC04:00	UTC04:00	

America/Sao_Paulo	UTC03:00	UTC03:00	Brazil (southeast: GO, DF, MG, ES, RJ, SP, PR, SC, RS)
America/Scoresbysund	UTC01:00	UTC+00:00	Scoresbysund/Ittoqqortoormiit
America/Shiprock	UTC07:00	UTC06:00	
America/Sitka	UTC09:00	UTC08:00	Alaska - Sitka area
America/St_Barthelemy	UTC04:00	UTC04:00	
America/St_Johns	UTC03:30	UTC02:30	Newfoundland; Labrador (southeast)
America/St_Kitts	UTC04:00	UTC04:00	
America/St_Lucia	UTC04:00	UTC04:00	
America/St_Thomas	UTC04:00	UTC04:00	
America/St_Vincent	UTC04:00	UTC04:00	
America/Swift_Current	UTC06:00	UTC06:00	CST - SK (midwest)
America/Tegucigalpa	UTC06:00	UTC06:00	
America/Thule	UTC04:00	UTC03:00	Thule/Pituffik
America/Thunder_Bay	UTC05:00	UTC04:00	Eastern - ON (Thunder Bay)
America/Tijuana	UTC08:00	UTC07:00	Pacific Time US - Baja California
America/Toronto	UTC05:00	UTC04:00	Eastern - ON, QC (most areas)
America/Tortola	UTC04:00	UTC04:00	
America/Vancouver	UTC08:00	UTC07:00	Pacific - BC (most areas)
America/Virgin	UTC04:00	UTC04:00	
America/Whitehorse	UTC08:00	UTC07:00	Pacific - Yukon (south)
America/Winnipeg	UTC06:00	UTC05:00	Central - ON (west); Manitoba
America/Yakutat	UTC09:00	UTC08:00	Alaska - Yakutat
America/Yellowknife	UTC07:00	UTC06:00	Mountain - NT (central)
Antarctica/Casey	UTC+11:00	UTC+11:00	Casey
Antarctica/Davis	UTC+07:00	UTC+07:00	Davis
Antarctica/DumontDURville	UTC+10:00	UTC+10:00	Dumont-d'Urville
Antarctica/Macquarie	UTC+11:00	UTC+11:00	Macquarie Island
Antarctica/Mawson	UTC+05:00	UTC+05:00	Mawson
Antarctica/McMurdo	UTC+12:00	UTC+13:00	New Zealand time - McMurdo, South Pole
Antarctica/Palmer	UTC03:00	UTC03:00	Palmer
Antarctica/Rothera	UTC03:00	UTC03:00	Rothera
Antarctica/South_Pole	UTC+12:00	UTC+13:00	
Antarctica/Syowa	UTC+03:00	UTC+03:00	Syowa
Antarctica/Troll	UTC+00:00	UTC+02:00	Troll
Antarctica/Vostok	UTC+06:00	UTC+06:00	Vostok
Arctic/Longyearbyen	UTC+01:00	UTC+02:00	
Asia/Aden	UTC+03:00	UTC+03:00	
Asia/Almaty	UTC+06:00	UTC+06:00	Kazakhstan (most areas)
Asia/Amman	UTC+02:00	UTC+03:00	
Asia/Anadyr	UTC+12:00	UTC+12:00	MSK+09 - Bering Sea

Asia/Aqtau	UTC+05:00	UTC+05:00	Mangghystau/Mankistau
Asia/Aqtobe	UTC+05:00	UTC+05:00	Aqtobe/Aktobe
Asia/Ashgabat	UTC+05:00	UTC+05:00	
Asia/Ashkhabad	UTC+05:00	UTC+05:00	
Asia/Atyrau	UTC+05:00	UTC+05:00	Atyrau/Atirau/Gur'yev
Asia/Baghdad	UTC+03:00	UTC+03:00	
Asia/Bahrain	UTC+03:00	UTC+03:00	
Asia/Baku	UTC+04:00	UTC+04:00	
Asia/Bangkok	UTC+07:00	UTC+07:00	
Asia/Barnaul	UTC+07:00	UTC+07:00	MSK+04 - Altai
Asia/Beirut	UTC+02:00	UTC+03:00	
Asia/Bishkek	UTC+06:00	UTC+06:00	
Asia/Brunei	UTC+08:00	UTC+08:00	
Asia/Calcutta	UTC+05:30	UTC+05:30	
Asia/Chita	UTC+09:00	UTC+09:00	MSK+06 - Zabaykalsky
Asia/Choibalsan	UTC+08:00	UTC+08:00	Dornod, Sukhbaatar
Asia/Chongqing	UTC+08:00	UTC+08:00	
Asia/Chungking	UTC+08:00	UTC+08:00	
Asia/Colombo	UTC+05:30	UTC+05:30	
Asia/Dacca	UTC+06:00	UTC+06:00	
Asia/Damascus	UTC+02:00	UTC+03:00	
Asia/Dhaka	UTC+06:00	UTC+06:00	
Asia/Dili	UTC+09:00	UTC+09:00	
Asia/Dubai	UTC+04:00	UTC+04:00	
Asia/Dushanbe	UTC+05:00	UTC+05:00	
Asia/Famagusta	UTC+02:00	UTC+02:00	Northern Cyprus
Asia/Gaza	UTC+02:00	UTC+03:00	Gaza Strip
Asia/Harbin	UTC+08:00	UTC+08:00	
Asia/Hebron	UTC+02:00	UTC+03:00	West Bank
Asia/Ho_Chi_Minh	UTC+07:00	UTC+07:00	
Asia/Hong_Kong	UTC+08:00	UTC+08:00	
Asia/Hovd	UTC+07:00	UTC+07:00	Bayan-Olgii, Govi-Altai, Hovd, Uvs, Zavkhan
Asia/Irkutsk	UTC+08:00	UTC+08:00	MSK+05 - Irkutsk, Buryatia
Asia/Istanbul	UTC+03:00	UTC+03:00	
Asia/Jakarta	UTC+07:00	UTC+07:00	Java, Sumatra
Asia/Jayapura	UTC+09:00	UTC+09:00	New Guinea (West Papua / Irian Jaya); Maluku/Moluccas
Asia/Jerusalem	UTC+02:00	UTC+03:00	
Asia/Kabul	UTC+04:30	UTC+04:30	
Asia/Kamchatka	UTC+12:00	UTC+12:00	MSK+09 - Kamchatka
Asia/Karachi	UTC+05:00	UTC+05:00	

Asia/Kashgar	UTC+06:00	UTC+06:00	
Asia/Kathmandu	UTC+05:45	UTC+05:45	
Asia/Katmandu	UTC+05:45	UTC+05:45	
Asia/Khandyga	UTC+09:00	UTC+09:00	MSK+06 - Tomponsky, Ust-Maysky
Asia/Kolkata	UTC+05:30	UTC+05:30	
Asia/Krasnoyarsk	UTC+07:00	UTC+07:00	MSK+04 - Krasnoyarsk area
Asia/Kuala_Lumpur	UTC+08:00	UTC+08:00	Malaysia (peninsula)
Asia/Kuching	UTC+08:00	UTC+08:00	Sabah, Sarawak
Asia/Kuwait	UTC+03:00	UTC+03:00	
Asia/Macao	UTC+08:00	UTC+08:00	
Asia/Macau	UTC+08:00	UTC+08:00	
Asia/Magadan	UTC+11:00	UTC+11:00	MSK+08 - Magadan
Asia/Makassar	UTC+08:00	UTC+08:00	Borneo (east, south); Sulawesi/Celebes, Bali, Nusa Tenggara; Timor (west)
Asia/Manila	UTC+08:00	UTC+08:00	
Asia/Muscat	UTC+04:00	UTC+04:00	
Asia/Novokuznetsk	UTC+07:00	UTC+07:00	MSK+04 - Kemerovo
Asia/Novosibirsk	UTC+07:00	UTC+07:00	MSK+04 - Novosibirsk
Asia/Omsk	UTC+06:00	UTC+06:00	MSK+03 - Omsk
Asia/Oral	UTC+05:00	UTC+05:00	West Kazakhstan
Asia/Phnom_Penh	UTC+07:00	UTC+07:00	
Asia/Pontianak	UTC+07:00	UTC+07:00	Borneo (west, central)
Asia/Pyongyang	UTC+09:00	UTC+09:00	
Asia/Qatar	UTC+03:00	UTC+03:00	
Asia/Qyzylorda	UTC+05:00	UTC+05:00	Qyzylorda/Kyzylorda/Kzyl-Orda
Asia/Rangoon	UTC+06:30	UTC+06:30	
Asia/Riyadh	UTC+03:00	UTC+03:00	
Asia/Saigon	UTC+07:00	UTC+07:00	
Asia/Sakhalin	UTC+11:00	UTC+11:00	MSK+08 - Sakhalin Island
Asia/Samarkand	UTC+05:00	UTC+05:00	Uzbekistan (west)
Asia/Seoul	UTC+09:00	UTC+09:00	
Asia/Shanghai	UTC+08:00	UTC+08:00	Beijing Time
Asia/Singapore	UTC+08:00	UTC+08:00	
Asia/Srednekolymsk	UTC+11:00	UTC+11:00	MSK+08 - Sakha (E); North Kuril Is
Asia/Taipei	UTC+08:00	UTC+08:00	
Asia/Tashkent	UTC+05:00	UTC+05:00	Uzbekistan (east)
Asia/Tbilisi	UTC+04:00	UTC+04:00	
Asia/Tehran	UTC+03:30	UTC+04:30	
Asia/Tel_Aviv	UTC+02:00	UTC+03:00	
Asia/Thimbu	UTC+06:00	UTC+06:00	
Asia/Thimphu	UTC+06:00	UTC+06:00	

Asia/Tokyo	UTC+09:00	UTC+09:00	
Asia/Tomsk	UTC+07:00	UTC+07:00	MSK+04 - Tomsk
Asia/Ujung_Pandang	UTC+08:00	UTC+08:00	
Asia/Ulaanbaatar	UTC+08:00	UTC+08:00	Mongolia (most areas)
Asia/Ulan_Bator	UTC+08:00	UTC+08:00	
Asia/Urumqi	UTC+06:00	UTC+06:00	Xinjiang Time
Asia/Ust-Nera	UTC+10:00	UTC+10:00	MSK+07 - Oymyakonsky
Asia/Vientiane	UTC+07:00	UTC+07:00	
Asia/Vladivostok	UTC+10:00	UTC+10:00	MSK+07 - Amur River
Asia/Yakutsk	UTC+09:00	UTC+09:00	MSK+06 - Lena River
Asia/Yangon	UTC+06:30	UTC+06:30	
Asia/Yekaterinburg	UTC+05:00	UTC+05:00	MSK+02 - Urals
Asia/Yerevan	UTC+04:00	UTC+04:00	
Atlantic/Azores	UTC01:00	UTC+00:00	Azores
Atlantic/Bermuda	UTC04:00	UTC03:00	
Atlantic/Canary	UTC+00:00	UTC+01:00	Canary Islands
Atlantic/Cape_Verde	UTC01:00	UTC01:00	
Atlantic/Faeroe	UTC+00:00	UTC+01:00	
Atlantic/Faroe	UTC+00:00	UTC+01:00	
Atlantic/Jan_Mayen	UTC+01:00	UTC+02:00	
Atlantic/Madeira	UTC+00:00	UTC+01:00	Madeira Islands
Atlantic/Reykjavik	UTC+00:00	UTC+00:00	
Atlantic/South_Georgia	UTC02:00	UTC02:00	
Atlantic/St_Helena	UTC+00:00	UTC+00:00	
Atlantic/Stanley	UTC03:00	UTC03:00	
Australia/ACT	UTC+10:00	UTC+11:00	
Australia/Adelaide	UTC+09:30	UTC+10:30	South Australia
Australia/Brisbane	UTC+10:00	UTC+10:00	Queensland (most areas)
Australia/Broken_Hill	UTC+09:30	UTC+10:30	New South Wales (Yancowinna)
Australia/Canberra	UTC+10:00	UTC+11:00	
Australia/Currie	UTC+10:00	UTC+11:00	Tasmania (King Island)
Australia/Darwin	UTC+09:30	UTC+09:30	Northern Territory
Australia/Eucla	UTC+08:45	UTC+08:45	Western Australia (Eucla)
Australia/Hobart	UTC+10:00	UTC+11:00	Tasmania (most areas)
Australia/LHI	UTC+10:30	UTC+11:00	
Australia/Lindeman	UTC+10:00	UTC+10:00	Queensland (Whitsunday Islands)
Australia/Lord_Howe	UTC+10:30	UTC+11:00	Lord Howe Island
Australia/Melbourne	UTC+10:00	UTC+11:00	Victoria
Australia/North	UTC+09:30	UTC+09:30	
Australia/NSW	UTC+10:00	UTC+11:00	

Australia/Perth	UTC+08:00	UTC+08:00	Western Australia (most areas)
Australia/Queensland	UTC+10:00	UTC+10:00	
Australia/South	UTC+09:30	UTC+10:30	
Australia/Sydney	UTC+10:00	UTC+11:00	New South Wales (most areas)
Australia/Tasmania	UTC+10:00	UTC+11:00	
Australia/Victoria	UTC+10:00	UTC+11:00	
Australia/West	UTC+08:00	UTC+08:00	
Australia/Yancowinna	UTC+09:30	UTC+10:30	
Brazil/Acre	UTC05:00	UTC05:00	
Brazil/DeNoronha	UTC02:00	UTC02:00	
Brazil/East	UTC03:00	UTC02:00	
Brazil/West	UTC04:00	UTC04:00	
Canada/Atlantic	UTC04:00	UTC03:00	
Canada/Central	UTC06:00	UTC05:00	
Canada/Eastern	UTC05:00	UTC04:00	
Canada/Mountain	UTC07:00	UTC06:00	
Canada/Newfoundland	UTC03:30	UTC02:30	
Canada/Pacific	UTC08:00	UTC07:00	
Canada/Saskatchewan	UTC06:00	UTC06:00	
Canada/Yukon	UTC08:00	UTC07:00	
CET	UTC+01:00	UTC+02:00	
Chile/Continental	UTC04:00	UTC03:00	
Chile/EasterIsland	UTC06:00	UTC05:00	
CST6CDT	UTC06:00	UTC05:00	
Cuba	UTC05:00	UTC04:00	
EET	UTC+02:00	UTC+03:00	
Egypt	UTC+02:00	UTC+02:00	
Eire	UTC+00:00	UTC+01:00	
EST	UTC05:00	UTC05:00	
EST5EDT	UTC05:00	UTC04:00	
Etc/GMT	UTC+00:00	UTC+00:00	
Etc/GMT+0	UTC+00:00	UTC+00:00	
Etc/GMT+1	UTC01:00	UTC01:00	
Etc/GMT+10	UTC10:00	UTC10:00	
Etc/GMT+11	UTC11:00	UTC11:00	
Etc/GMT+12	UTC12:00	UTC12:00	
Etc/GMT+2	UTC02:00	UTC02:00	
Etc/GMT+3	UTC03:00	UTC03:00	
Etc/GMT+4	UTC04:00	UTC04:00	
Etc/GMT+5	UTC05:00	UTC05:00	

Etc/GMT+6	UTC06:00	UTC06:00	
Etc/GMT+7	UTC07:00	UTC07:00	
Etc/GMT+8	UTC08:00	UTC08:00	
Etc/GMT+9	UTC09:00	UTC09:00	
Etc/GMT0	UTC+00:00	UTC+00:00	
Etc/GMT-0	UTC+00:00	UTC+00:00	
Etc/GMT-1	UTC+01:00	UTC+01:00	
Etc/GMT-10	UTC+10:00	UTC+10:00	
Etc/GMT-11	UTC+11:00	UTC+11:00	
Etc/GMT-12	UTC+12:00	UTC+12:00	
Etc/GMT-13	UTC+13:00	UTC+13:00	
Etc/GMT-14	UTC+14:00	UTC+14:00	
Etc/GMT-2	UTC+02:00	UTC+02:00	
Etc/GMT-3	UTC+03:00	UTC+03:00	
Etc/GMT-4	UTC+04:00	UTC+04:00	
Etc/GMT-5	UTC+05:00	UTC+05:00	
Etc/GMT-6	UTC+06:00	UTC+06:00	
Etc/GMT-7	UTC+07:00	UTC+07:00	
Etc/GMT-8	UTC+08:00	UTC+08:00	
Etc/GMT-9	UTC+09:00	UTC+09:00	
Etc/Greenwich	UTC+00:00	UTC+00:00	
Etc/UCT	UTC+00:00	UTC+00:00	
Etc/Universal	UTC+00:00	UTC+00:00	
Etc/UTC	UTC+00:00	UTC+00:00	
Etc/Zulu	UTC+00:00	UTC+00:00	
Europe/Amsterdam	UTC+01:00	UTC+02:00	
Europe/Andorra	UTC+01:00	UTC+02:00	
Europe/Astrakhan	UTC+04:00	UTC+04:00	MSK+01 - Astrakhan
Europe/Athens	UTC+02:00	UTC+03:00	
Europe/Belfast	UTC+00:00	UTC+01:00	
Europe/Belgrade	UTC+01:00	UTC+02:00	
Europe/Berlin	UTC+01:00	UTC+02:00	Germany (except for Büsingen am Hochrhein)
Europe/Bratislava	UTC+01:00	UTC+02:00	
Europe/Brussels	UTC+01:00	UTC+02:00	
Europe/Bucharest	UTC+02:00	UTC+03:00	
Europe/Budapest	UTC+01:00	UTC+02:00	
Europe/Busingen	UTC+01:00	UTC+02:00	Büdingen am Hochrhein
Europe/Chisinau	UTC+02:00	UTC+03:00	
Europe/Copenhagen	UTC+01:00	UTC+02:00	
Europe/Dublin	UTC+00:00	UTC+01:00	

Europe/Gibraltar	UTC+01:00	UTC+02:00	
Europe/Guernsey	UTC+00:00	UTC+01:00	
Europe/Helsinki	UTC+02:00	UTC+03:00	
Europe/Isle_of_Man	UTC+00:00	UTC+01:00	
Europe/Istanbul	UTC+03:00	UTC+03:00	
Europe/Jersey	UTC+00:00	UTC+01:00	
Europe/Kaliningrad	UTC+02:00	UTC+02:00	MSK01 - Kaliningrad
Europe/Kiev	UTC+02:00	UTC+03:00	Ukraine (most areas)
Europe/Kirov	UTC+03:00	UTC+03:00	MSK+00 - Kirov
Europe/Lisbon	UTC+00:00	UTC+01:00	Portugal (mainland)
Europe/Ljubljana	UTC+01:00	UTC+02:00	
Europe/London	UTC+00:00	UTC+01:00	
Europe/Luxembourg	UTC+01:00	UTC+02:00	
Europe/Madrid	UTC+01:00	UTC+02:00	Spain (mainland)
Europe/Malta	UTC+01:00	UTC+02:00	
Europe/Mariehamn	UTC+02:00	UTC+03:00	
Europe/Minsk	UTC+03:00	UTC+03:00	
Europe/Monaco	UTC+01:00	UTC+02:00	
Europe/Moscow	UTC+03:00	UTC+03:00	MSK+00 - Moscow area
Asia/Nicosia	UTC+02:00	UTC+03:00	Cyprus (most areas)
Europe/Oslo	UTC+01:00	UTC+02:00	
Europe/Paris	UTC+01:00	UTC+02:00	
Europe/Podgorica	UTC+01:00	UTC+02:00	
Europe/Prague	UTC+01:00	UTC+02:00	
Europe/Riga	UTC+02:00	UTC+03:00	
Europe/Rome	UTC+01:00	UTC+02:00	
Europe/Samara	UTC+04:00	UTC+04:00	MSK+01 - Samara, Udmurtia
Europe/San_Marino	UTC+01:00	UTC+02:00	
Europe/Sarajevo	UTC+01:00	UTC+02:00	
Europe/Saratov	UTC+04:00	UTC+04:00	MSK+01 - Saratov
Europe/Simferopol	UTC+03:00	UTC+03:00	Crimea
Europe/Skopje	UTC+01:00	UTC+02:00	
Europe/Sofia	UTC+02:00	UTC+03:00	
Europe/Stockholm	UTC+01:00	UTC+02:00	
Europe/Tallinn	UTC+02:00	UTC+03:00	
Europe/Tirane	UTC+01:00	UTC+02:00	
Europe/Tiraspol	UTC+02:00	UTC+03:00	
Europe/Ulyanovsk	UTC+04:00	UTC+04:00	MSK+01 - Ulyanovsk
Europe/Uzhgorod	UTC+02:00	UTC+03:00	Ruthenia
Europe/Vaduz	UTC+01:00	UTC+02:00	

Europe/Vatican	UTC+01:00	UTC+02:00	
Europe/Vienna	UTC+01:00	UTC+02:00	
Europe/Vilnius	UTC+02:00	UTC+03:00	
Europe/Volgograd	UTC+04:00	UTC+04:00	MSK+01 - Volgograd
Europe/Warsaw	UTC+01:00	UTC+02:00	
Europe/Zagreb	UTC+01:00	UTC+02:00	
Europe/Zaporozhye	UTC+02:00	UTC+03:00	Zaporozh'ye/Zaporizhia; Lugansk/Luhansk (east)
Europe/Zurich	UTC+01:00	UTC+02:00	
GB	UTC+00:00	UTC+01:00	
GB-Eire	UTC+00:00	UTC+01:00	
GMT	UTC+00:00	UTC+00:00	
GMT+0	UTC+00:00	UTC+00:00	
GMT0	UTC+00:00	UTC+00:00	
GMT-0	UTC+00:00	UTC+00:00	
Greenwich	UTC+00:00	UTC+00:00	
Hongkong	UTC+08:00	UTC+08:00	
HST	UTC10:00	UTC10:00	
Iceland	UTC+00:00	UTC+00:00	
Indian/Antananarivo	UTC+03:00	UTC+03:00	
Indian/Chagos	UTC+06:00	UTC+06:00	
Indian/Christmas	UTC+07:00	UTC+07:00	
Indian/Cocos	UTC+06:30	UTC+06:30	
Indian/Comoro	UTC+03:00	UTC+03:00	
Indian/Kerguelen	UTC+05:00	UTC+05:00	
Indian/Mahe	UTC+04:00	UTC+04:00	
Indian/Maldives	UTC+05:00	UTC+05:00	
Indian/Mauritius	UTC+04:00	UTC+04:00	
Indian/Mayotte	UTC+03:00	UTC+03:00	
Indian/Reunion	UTC+04:00	UTC+04:00	
Iran	UTC+03:30	UTC+04:30	
Israel	UTC+02:00	UTC+03:00	
Jamaica	UTC05:00	UTC05:00	
Japan	UTC+09:00	UTC+09:00	
Kwajalein	UTC+12:00	UTC+12:00	
Libya	UTC+02:00	UTC+02:00	
MET	UTC+01:00	UTC+02:00	
Mexico/BajaNorte	UTC08:00	UTC07:00	
Mexico/BajaSur	UTC07:00	UTC06:00	
Mexico/General	UTC06:00	UTC05:00	
MST	UTC07:00	UTC07:00	

MST7MDT	UTC07:00	UTC06:00	
Navajo	UTC07:00	UTC06:00	
NZ	UTC+12:00	UTC+13:00	
NZ-CHAT	UTC+12:45	UTC+13:45	
Pacific/Apia	UTC+13:00	UTC+14:00	
Pacific/Auckland	UTC+12:00	UTC+13:00	New Zealand (most areas)
Pacific/Bougainville	UTC+11:00	UTC+11:00	Bougainville
Pacific/Chatham	UTC+12:45	UTC+13:45	Chatham Islands
Pacific/Chuuk	UTC+10:00	UTC+10:00	Chuuk/Truk, Yap
Pacific/Easter	UTC06:00	UTC05:00	Easter Island
Pacific/Efate	UTC+11:00	UTC+11:00	
Pacific/Enderbury	UTC+13:00	UTC+13:00	Phoenix Islands
Pacific/Fakaofu	UTC+13:00	UTC+13:00	
Pacific/Fiji	UTC+12:00	UTC+13:00	
Pacific/Funafuti	UTC+12:00	UTC+12:00	
Pacific/Galapagos	UTC06:00	UTC06:00	Galapagos Islands
Pacific/Gambier	UTC09:00	UTC09:00	Gambier Islands
Pacific/Guadalcanal	UTC+11:00	UTC+11:00	
Pacific/Guam	UTC+10:00	UTC+10:00	
Pacific/Honolulu	UTC10:00	UTC10:00	Hawaii
Pacific/Johnston	UTC10:00	UTC10:00	
Pacific/Kiritimati	UTC+14:00	UTC+14:00	Line Islands
Pacific/Kosrae	UTC+11:00	UTC+11:00	Kosrae
Pacific/Kwajalein	UTC+12:00	UTC+12:00	Kwajalein
Pacific/Majuro	UTC+12:00	UTC+12:00	Marshall Islands (most areas)
Pacific/Marquesas	UTC09:30	UTC09:30	Marquesas Islands
Pacific/Midway	UTC11:00	UTC11:00	Midway Islands
Pacific/Nauru	UTC+12:00	UTC+12:00	
Pacific/Niue	UTC11:00	UTC11:00	
Pacific/Norfolk	UTC+11:00	UTC+11:00	
Pacific/Noumea	UTC+11:00	UTC+11:00	
Pacific/Pago_Pago	UTC11:00	UTC11:00	
Pacific/Palau	UTC+09:00	UTC+09:00	
Pacific/Pitcairn	UTC08:00	UTC08:00	
Pacific/Pohnpei	UTC+11:00	UTC+11:00	Pohnpei/Ponape
Pacific/Ponape	UTC+11:00	UTC+11:00	
Pacific/Port_Moresby	UTC+10:00	UTC+10:00	Papua New Guinea (most areas)
Pacific/Rarotonga	UTC10:00	UTC10:00	
Pacific/Saipan	UTC+10:00	UTC+10:00	
Pacific/Samoa	UTC11:00	UTC11:00	

Pacific/Tahiti	UTC+10:00	UTC+10:00	Society Islands
Pacific/Tarawa	UTC+12:00	UTC+12:00	Gilbert Islands
Pacific/Tongatapu	UTC+13:00	UTC+14:00	
Pacific/Truk	UTC+10:00	UTC+10:00	
Pacific/Wake	UTC+12:00	UTC+12:00	Wake Island
Pacific/Wallis	UTC+12:00	UTC+12:00	
Pacific/Yap	UTC+10:00	UTC+10:00	
Poland	UTC+01:00	UTC+02:00	
Portugal	UTC+00:00	UTC+01:00	
PRC	UTC+08:00	UTC+08:00	
PST8PDT	UTC08:00	UTC07:00	
ROC	UTC+08:00	UTC+08:00	
ROK	UTC+09:00	UTC+09:00	
Singapore	UTC+08:00	UTC+08:00	
Turkey	UTC+03:00	UTC+03:00	
UCT	UTC+00:00	UTC+00:00	
Universal	UTC+00:00	UTC+00:00	
US/Alaska	UTC09:00	UTC08:00	
US/Aleutian	UTC10:00	UTC09:00	
US/Arizona	UTC07:00	UTC07:00	
US/Central	UTC06:00	UTC05:00	
US/Eastern	UTC05:00	UTC04:00	
US/East-Indiana	UTC05:00	UTC04:00	
US/Hawaii	UTC10:00	UTC10:00	
US/Indiana-Starke	UTC06:00	UTC05:00	
US/Michigan	UTC05:00	UTC04:00	
US/Mountain	UTC07:00	UTC06:00	
US/Pacific	UTC08:00	UTC07:00	
US/Pacific-New	UTC08:00	UTC07:00	
US/Samoa	UTC11:00	UTC11:00	
UTC	UTC+00:00	UTC+00:00	
WET	UTC+00:00	UTC+01:00	
W-SU	UTC+03:00	UTC+03:00	
Zulu	UTC+00:00	UTC+00:00	

Locale Settings

Trifacta® supports a tier-based scheme for applying locale settings.

NOTE: Locale settings apply only to the inference and validation of data in the Trifacta application. Underlying data is not affected by changing locale.

NOTE: After saving changes to your locale, refresh your page. Subsequent executions of the data inference service use the new locale settings.

NOTE: When locale is changed, data type validation is affected only on subsequent executions of the data type inference service. If you are using structured datasets, such as schematized JDBC sources, data types may be attached to the datasets that you have already imported. These data types are not affected.

Locale settings can be configured for:

- **Project or workspace level:** global local for all users in the environment
- **User level:** In User Preferences, individual users can set their personal locale.

Supported Locales

The following locales are available:

- United States
- Australia
- Austria
- Belgium
- Bulgaria
- Canada
- Croatia
- Cyprus
- Czech Republic
- Denmark
- Estonia
- Finland
- France
- Germany
- Hungary
- Greece
- Ireland
- Italy
- Japan
- Latvia
- Lithuania
- Luxembourg
- Malta
- New Zealand
- Netherlands
- Poland
- Portugal
- Romania
- Singapore
- Slovakia

- Slovenia
- Spain
- Sweden
- United Kingdom

Product Areas Affected by Locale

Datetime

- Data type inference more accurately recognizes Datetime values based on locale settings.
- Suggestion cards containing Datetime transformations and example previews utilize locale settings.
- Number of weeks in the year may vary. For example, the WEEKNUM function calculated on the last few dates of the calendar year may return different values depending on the locale where the calculation is performed:

NOTE: The output calculations for number of weeks may vary between browser, Trifacta Photon, and the running environment where the job may be executed, based upon locale.

- **EN-US:** Maximum of 52 weeks
- **ISO 8601:** Maximum of 53 weeks
- You can add steps similar to the following, which adds a 53rd week as a possible output for WEEKNUM:

```
derive type: single value: dateformat(date(year(myDate), 1, 1), &apos;yyyy\MM\dd&apos;) as: &apos;NewYearsDayforMyYear&apos;
```

```
derive type: single value: if(datedif(NewYearsDayforMyYear, myDate, day) > (52 * 7), 53, weeknum(myDate)) as: &apos;weekNumforMyDate&apos;
```

Character Encoding

Contents:

- *Character Encoding on Input*
 - *Character Encoding within the Application*
 - *Character Encoding on Output*
-

This section describes how Trifacta® manages character encoding on import, within the application, and on export.

Overview of Character Encoding

Character encoding refers to the mechanism by which numeric digital data is used to represent characters, including alphanumeric characters and punctuation, in languages around the world. To ensure that different machines can represent the same thing on-screen, each machine can reference one or more of the supported file encoding types, which are standards for representation of characters. For example, a machine in the United Kingdom will represent the letter "A" sent from a machine in the United States if they are using the same encoding file encoding types.

In many languages around the world, the representation of all characters requires hundreds and even thousands of characters. As a result, encodings for these regions may require a larger number of bits to represent all aspects of the language.

The platform supports a global file encoding type. By default, this encoding type is UTF-8. For more information, see *Configure Global File Encoding Type*.

Character Encoding on Input

By default, Trifacta supports UTF-8 on input. As needed, individual users can change the file encoding of input files. For example, a file that is ingested with a double-byte encoding can be identified as such for the product in the file settings during import, so that the data can be properly parsed during input.

Character Encoding within the Application

Within the Trifacta application, you can use the following functions to modify character encodings:

Item	Description
<i>BASE64ENCODE Function</i>	Converts an input value to base64 encoding with optional padding with an equals sign (=). Input can be of any type. Output type is String.
<i>BASE64DECODE Function</i>	Converts an input base64 value to text. Output type is String.
<i>UNICODE Function</i>	Generates the Unicode index value for the first character of the input string.

Character Encoding on Output

All files are published with UTF-8 encoding.

Supported File Encoding Types

Supported Global File Encoding Types for Input

- UTF-8 (default)
- IBM00858
- IBM437
- IBM775
- IBM850
- IBM852
- IBM855
- IBM857
- IBM862
- IBM866
- ISO-8859-1
- ISO-8859-2
- ISO-8859-3
- ISO-8859-4
- ISO-8859-5
- ISO-8859-6
- ISO-8859-7
- ISO-8859-8
- ISO-8859-9
- ISO-8859-13
- ISO-8859-15
- KOI8-R
- KOI8-U
- US-ASCII
- UTF-16
- UTF-16BE
- UTF-16LE
- UTF-32
- windows-1250
- windows-1251
- windows-1252
- windows-1253
- windows-1254
- windows-1255
- windows-1256
- windows-1257
- x-IBM737
- x-IBM874
- x-UTF-16LE-BOM

Supported Global File Encoding Types for Output

Output files are written in UTF-8 encoding.

Sort Order

Contents:

- *General Sort Order*
 - *Sort Order for Strings*
 - *Sort Order for Integers and Decimals*
-

This section describes how values are sorted within Trifacta®. Sorting can be applied through the following mechanisms:

- Clicking a column header in a workspace table, such as the Flows, Library, or Jobs pages.
- Applying a Sort transform.
- Applying the ARRAYSORT function to an array.

NOTE: Following listings represent sorting in ascending order. Descending order sorting is in the reverse of the listings below.

General Sort Order

For any column, sorting is performed in the following order:

1. Sorting based on data type
2. Mismatched values
3. Null/empty values

Sort Order for Strings

Since all values are valid for String data type, this sort order represents the most common representation for sort order.

1. Sorting based on data type:
 - a. Numeric values (low value to high value)
 - b. Whitespace
 - c. Special characters
 - d. Alphabetical
 - i. Case-insensitive
 - ii. Accented characters (ä) are below unaccented character (a)
2. Mismatched values
3. Null/empty values

Sort Order for Integers and Decimals

For Integers and Decimals, sorting happens in the following order:

1. Sorting based on data type:
 - a. Numeric values (low value to high value)
2. Mismatched values

3. Null/empty values

Join Types

Contents:

- *Inner Join*
 - *Left Join*
 - *Right Join*
 - *Full Outer Join*
 - *Cross Join*
 - *Self Join*
 - *Joins Together*
-

The following types of joins are supported. For example, the following tables contains information about employees and departments.

Employee table:

Name	DepartmentID	Role
Dave Smith	001	Product Marketing Manager
Julie Jones	002	Software Engineer
Scott Tanner	001	Director of Demand Gen
Ted Connors	002	Software Engineer
Margaret Lane	001	VP of Marketing
Mary Martin	004	Receptionist

Department table:

Name	DepartmentID
Marketing	001
Engineering	002
Accounting	003

In the above example, `DepartmentID` is the key to use in both tables for any joins.

Inner Join

An **inner join** requires that key values exist in both tables for the records to appear in the results table. Records appear in the merge only if there are matches in both tables for the key values.

- If you want to include rows containing non-matching values, you must use some form of an outer join. See below.

For the preceding example tables, an inner join on the `DepartmentID` table produces the following result table:

Employee.Name	Employee.DepartmentID	Employee.Role	Department.Name	Department.DepartmentID
Dave Smith	001	Product Marketing Manager	Marketing	001
Julie Jones	002	Software Engineer	Engineering	002
Scott Tanner	001	Director of Demand Gen	Marketing	001

Ted Connors	002	Software Engineer	Engineering	002
Margaret Lane	001	VP of Marketing	Marketing	001

Notes:

- All fields are included in the merged result set. Fields from the first dataset are listed first.
- The row for Mary Martin is excluded, since there is no reference in the Department table for her department identifier. The row for Accounting is excluded, since there is no reference in the Employee table for the department identifier.
 - To include these rows, you either need to augment the data or perform a form of an outer join.
- A null value in one table does not match a null value in another table. So, rows with null values in a join key are never included in an inner join. These values should be fixed.

Tip: An inner join can be used to eliminate rows with null values in their key fields.

Left Join

A left join (or left outer join) does not require that there be matching records for each value in the key value of the source (left) table. Each row in the left table appears in the results, regardless of whether there are matches in the right table.

For the preceding example tables, a left join on the `DepartmentID` table produces the following result table:

Employee.Name	Employee.DepartmentID	Employee.Role	Department.Name	Department.DepartmentID
Dave Smith	001	Product Marketing Manager	Marketing	001
Julie Jones	002	Software Engineer	Engineering	002
Scott Tanner	001	Director of Demand Gen	Marketing	001
Ted Connors	002	Software Engineer	Engineering	002
Margaret Lane	001	VP of Marketing	Marketing	001
Mary Martin	004	Receptionist	NULL	NULL

Notes:

- In this left join, the Mary Martin row has been added to the result, since her record in the Employee table does contain an entry for the `DepartmentID`. However, since there are no corresponding values in the Department table, the corresponding fields in the result table are `NULL` values.

Right Join

A right join (or right outer join) is the reverse of a left join. A right join does not require that there be matching records for each value in the key value of the secondary (right) table. Each row in the right table appears in the results, regardless of whether there are matches in the left table.

For the preceding example tables, a right join on the `DepartmentID` table produces the following result table:

Employee.Name	Employee.DepartmentID	Employee.Role	Department.Name	Department.DepartmentID
Dave Smith	001	Product Marketing Manager	Marketing	001
Julie Jones	002	Software Engineer	Engineering	002
Scott Tanner	001	Director of Demand Gen	Marketing	001
Ted Connors	002	Software Engineer	Engineering	002

Margaret Lane	001	VP of Marketing	Marketing	001
NULL	NULL	NULL	Accounting	003

Notes:

- In this right join, the Accounting entry is added. However, since there is no entry in the Employee table for the DepartmentID value, those fields are NULL values in the result set.

Full Outer Join

A **full outer join** combines the effects of a left join and a right join. If there is a match between the key values, a row is written in the result.

- If there is no match for a key value that appears in either table, a single record is written to the result, with NULL values inserted for the fields from the other table.

Employee.Name	Employee.DepartmentID	Employee.Role	Department.Name	Department.DepartmentID
Dave Smith	001	Product Marketing Manager	Marketing	001
Julie Jones	002	Software Engineer	Engineering	002
Scott Tanner	001	Director of Demand Gen	Marketing	001
Ted Connors	002	Software Engineer	Engineering	002
Margaret Lane	001	VP of Marketing	Marketing	001
Mary Martin	004	Receptionist	NULL	NULL
NULL	NULL	NULL	Accounting	003

Notes:

- Any duplicated rows between joining from left-to-right and from right-to-left are removed from the results.

Cross Join

A **cross join** combines each row of the first data set with each row of the second dataset, where every combination is represented in the output. As a result, the number of total rows in the join are:

```
Rows(DatasetA) * Rows(DatasetB)
```

NOTE: Depending on the size of your datasets, a cross join can greatly expand the size of the output, which may increase costs in some environments.

Self Join

A **self join** is a join operation between a dataset and a copy of itself. For example, you can use a self-join to invert the structure of hierarchical data, such as brand-product or manager-employee.

Trifacta supports joins between a recipe and any upstream recipe or dataset.

- You cannot join a recipe to itself.
- You can join it to its source imported dataset. When a self-join is performed with a recipe connected to its source dataset, only one line connects the imported dataset with the recipe in Flow View. This is as designed.
- You can join a recipe to any recipe upstream of it. Examples:

- You can create an empty recipe after the recipe from which you wish to self-join. In this new empty recipe, you add the join step back to the original recipe.
- You can insert an empty recipe between an imported dataset and the recipe where the self-join is performed. When you perform the self join in the first recipe, you join to the empty recipe you just created in between.
- In both examples, you can see multiple lines in Flow View to indicate the self-join.

Joins Together

The following diagram summarizes the relationships between the types of supported joins. In each venn diagram, the area of intersection is the set of records that contain shared key values.

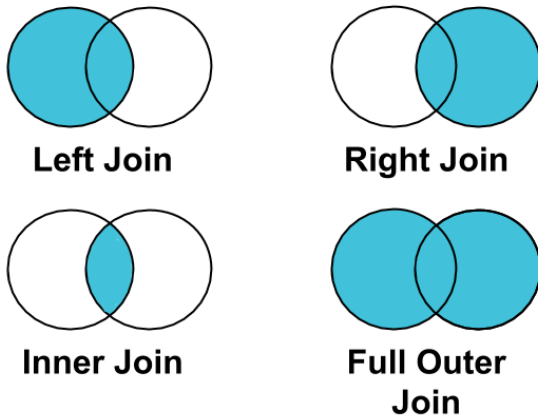


Figure: Join Types

Join Metrics

When you create a join, Trifacta® attempts to match up columns as the keys in your join. For each set of join keys, you can review the following metrics related to the join.

Match percentage:

When you hover over the percentage of matches between key values, you can see the details that make up the calculation:

Metric	Description
All Rows	Total count of rows in the dataset
Matches	Total count of values in the join key of the selected column with matching values in the join key of the other dataset.
Non-Matches	Total count of values in the join key of the selected column with values that do not have a match in the join key of the other dataset.

The percentage is calculated by summing the count of matches for both datasets and dividing that by the total count of rows across both datasets:

```
(Matches_Current_Dataset + Matches_Joined-in-Dataset) /  
(All_Rows_Current_Dataset + All_Rows_Joined-in-Dataset)
```

Rows in output:

When you hover over the Rows in Output metric, you can see the following values:

Metric	Description
All Rows	Total count of rows in each dataset.
Included	Count of rows from each dataset that are included in the output.
Excluded	Count of rows from each dataset that are excluded from the output.

Supported SQL Syntax

Contents:

- *Basic Syntax*
 - *Supported syntax by datastore*
 - *General Examples*
 - *Column aliasing*
 - *Collect whole table*
 - *Filter columns*
 - *Filter rows*
 - *Multi-line statement for imported datasets*
-

This section provides general information on how the Trifacta uses SQL to interact with your databases, including syntax requirements and examples.

Basic Syntax

Your SQL statements must be valid for the syntax expected by the target relational system. In particular, object delimiters may vary between systems.

NOTE: The proper syntax depends on your database system. Please consult the documentation for your product for details.

Tip: Although some relational systems do not require object delimiters around column names, it is recommended that you add them to all applicable objects.

Tip: Avoid using column type identifiers (e.g. `int`) and other SQL keywords as object names. Some systems may generate invalid SQL errors.

NOTE: In the following sections, Oracle syntax is used in the examples. Please modify the examples for your target system.

Supported syntax by datastore

Individual datastores may have differences in the supported syntax. For more information, please see the documentation for your datastore.

General Examples

Here are some basic SQL examples to get started.

Column aliasing

If your select statement results in multiple columns with same name, the query fails to validate or fails on execution, such as selecting all columns in a JOIN. In these cases, columns must be properly aliased.

NOTE: This error will be caught either during validating or during dataset import.

For example, in the following JOIN, the EMPLOYEE and DEPARTMENT tables have column names department_id and department_id.

```
SELECT * FROM EMPLOYEE INNER JOIN DEPARTMENT ON (department_id = department_id);
```

The above query generates an error. Columns must be properly aliased, as in the following:

```
SELECT e.id, e.department_id, e.first_name, e.last_name, d.department_name FROM EMPLOYEE AS E INNER JOIN DEPARTMENT d ON (e.department_id = d.department_id);
```

Collect whole table

```
SELECT * FROM "DB1"."table2";
```

Filter columns

```
SELECT lastName,firstName FROM "DB1"."table2";
```

Filter rows

```
SELECT lastName,firstName FROM "DB1"."table2" WHERE invoiceAmt > 10000;
```

Multi-line statement for imported datasets

The following example uses a multi-line SQL sequence to import a dataset:

NOTE: Multi-line SQL support is considered an advanced use case. This feature must be enabled.

The following example inserts values in the TABLE_INVENTORY table and then queries the table. It utilizes Oracle syntax:

```
INSERT INTO "SALES"."TABLE_INVENTORY" ("ID", "AVAILABILITY") VALUES (1, 10);  
SELECT * FROM "SALES"."TABLE_INVENTORY";
```


Sample Types

Contents:

- *Initial Data Samples*
 - *First Rows Samples*
 - *Random Samples*
 - *Filter-Based Samples*
 - *Anomaly-Based Samples*
 - *Stratified Samples*
 - *Cluster-Based Samples*
-

This section provides an overview of the types of samples that Trifacta® can generate.

Sample filters:

Several sampling types support the application of filters to the source data. In this case, a **filter** can be defined to limit the scope of rows that are used to generate the sample. For example, suppose you apply a filter like the following:

```
orderId == '100'
```

The rows of data available for generating the sample are reduced to include only the rows where the value of the `orderId` column is 100.

Tip: Sample filters are very useful for allowing you to generate samples that are much more specific to the steps that are trying to build at the present time in your recipe.

Scan method:

Depending on the type of sample, you may be able to select the method by which the data is scanned:

- **Quick Scan:** Representative sample is scanned and executed in-memory on the Trifacta node. Although the scope of the scanned data is smaller, these samples are much faster to generate.
 - If a Quick Scan sample fails, the Trifacta application may attempt to perform the scan on an available clustered running environment.
- **Full Scan:** Data is sampled from the full set of available data. The sampling job is executed on an available clustered running environment. These sampling jobs can take longer to execute. Depending on your environment, additional costs may be incurred.

Initial Data Samples

These samples are collected automatically when you first load a new dataset into the Transformer page. These sample contain the first 10 MB of data from the first file or table in the dataset.

Tip: In the Transformer page, these samples are labeled as **Initial Data**.

First Rows Samples

NOTE: The First rows sampling technique requires the Trifacta Photon running environment.

This sample is taken from the first set of rows in the imported dataset based on the current cursor location in the recipe. The first N rows in the dataset are collected based on the recipe steps up to the configured sample size.

- This sample may span multiple datasets and files, depending on how the recipe is constructed.
- The first rows sample is different from the initial sample, which is gathered without reference to any recipe steps.

These samples are fast to generate. These samples may load faster in the application than samples of other types.

Tip: If you have chained together multiple recipes, all steps in all linked recipes must be run to provide visual updates. If you are experiencing performance problems related to this kind of updating, you can select a recipe in the middle of the chain of recipes and switch it off the initial sample to a different sample. When invoked, the recipes from the preceding datasets do not need to be executed, which can improve performance.

Random Samples

Random selection of a subset of rows in the dataset. These samples are comparatively fast to generate. You can apply quick scan or full scan to determine the scope of the sample.

Filter-Based Samples

Find specific values in one or more columns. From the rows that have matching set of values, a random sample is generated.

You must define your filter in the Filter textbox.

Anomaly-Based Samples

Find mismatched or missing data or both in one or more columns.

You specify one or more columns and whether the anomaly is:

1. mismatched
2. missing
3. either of the above

Optionally, you can define an additional filter on any column.

Stratified Samples

Find all unique values within a column and create a sample that contains the unique values, up to the sample size limit. The distribution of the column values in the sample reflects the distribution of the column values in the dataset. Sampled values are sorted by frequency, relative to the specified column.

Optionally, you can apply a filter to this one.

Tip: Collecting samples containing all unique values can be useful if you are performing mapping transformations, such as values to columns. If your mapping contains too many unique values among your key-value pairs, you can try to delete all columns except the one containing key-value pairs in a step, collect the sample, add the mapping step, and then delete the step where all other columns are removed.

Cluster-Based Samples

Cluster sampling collects contiguous rows in the dataset that correspond to a random selection from the unique values in a column. All rows corresponding to the selected unique values appear in the sample, up to the maximum sample size. This sampling is useful for time-series analysis and advanced aggregations.

Optionally, you can apply an advanced filter to the column.

Support Bundle Contents

Contents:

- *Job Logs*
 - *cdf script*
 - *job.log*
 - *photon flags file*
 - *photon cli info file*
 - *photon cli log info file*
 - *Support Bundle*
 - *conf files folder*
 - *configuration service folder*
 - *service logs folder*
 - *Process files*
 - *ulimit.txt*
 - *version.txt*
 - *Binary files*
-

When you download job log files, the following contents may be included in the exported support bundle.

Example ZIP contents:

- 75 - job log folder
- 76 - job log folder
- support-bundle - support bundle folder

Tip: If the support bundle contents fails to generate, please review `log-bundle-creation-errors.txt` for details, which is located inside the `support-bundle` folder.

Job Logs

When you execute a job, it is broken down into individual for each phase of the process.

NOTE: The job log files downloaded from the Trifacta application may contain unnecessary messages from other executed jobs. In some cases, it may not be possible to filter out these messages.

There may be separate folders for each of the following processes:

- **Ingest:** Data is ready into the platform from the sources.
- **Convert:** Some imported datasets must be converted from their source format to a format that is natively readable by the product. Typically, these jobs convert binary files into CSVs for use.

NOTE: The files generated during Convert jobs are retained only for the duration of job execution, after which they are purged.

- **Transform:** All recipes in the job are executed at scale against datasources.
- **Profile:** Results of the transformation are profiled, if profiling has been enabled for the job.
- **Publish:** Results are written to the specified output locations and formats.

cdf script

This file contains the script that is passed to the running environment to transform your dataset.

NOTE: This script is in a compiled language that is passed to the running environment for job execution. It is not in Wrangle and is not intended for user consumption.

Example filenames(s):

```
cdf-script-7932315341561207019.py
```

job.log

Log file for the specific job.

Example filenames(s):

```
job.log
```

photon flags file

This file identifies the settings used by the Photon running environment during job execution.

Example filenames(s):

```
photon-3840706514533636937.flags
```

photon cli info file

Location and format of the log files for the instance of Photon where the job was executed.

Example filenames(s):

```
photon-cli.bin.INFO
```

photon cli log info file

Location and format of the log files for the instance of Photon where the job was executed.

Example filenames(s):

```
photon-cli.bin.ip-10-0-0-9.us-west-2.compute.internal.trifacta.log.INFO.20191101-234021.16850
```

Support Bundle

When support bundling is enabled, the following folders and files are included in the download. Support bundling is enabled by default.

NOTE: Log files that are configured for JSON output format cannot be included in the support bundle.

NOTE: You can disable or configure the contents of the support bundle. For more information, see *Configure Support Bundling*.

conf files folder

Current version and archived versions of `trifacta-conf.json`, the core platform configuration file. If the platform is connected to a Hadoop cluster, additional cluster configuration files are retrieved from the local Trifacta node and included in the support bundle.

Example filenames(s):

```
archives/trifacta-conf.json_2019-11-01T073409Z
archives/trifacta-conf.json_20191024-000_2019-11-01T073419.404Z
archives/trifacta-conf.json_20191024-000_2019-11-01T073435.390Z
```

configuration service folder

Configuration files extracted from the configuration service, which governs configuration at the system, workspace, and user level.

Example filenames(s):

```
actual-settings.json
default-settings.json
system-overrides.json
user-overrides.json
workspace-overrides.json
workspace-tier-overrides.json
```

service logs folder

Log files for services of Trifacta.

Example filenames(s):

```
artifact-storage-service.log
batch-job-runner.log
configuration-service.log
orchestration-service.log
```

Tip: If you are experiencing issues with the execution of plans, please review `orchestration-service.log` for messages.

Tip: Your downloaded bundle may also include access logs for the services.

For more information on these log files, see *System Services and Logs*.

Process files

The following files include information on the processes that are created and used by the Trifacta platform:

Example filenames(s):

- `process-info.json` - list of processes that are run under the Trifacta user

ulimit.txt

Output of executing the `ulimit` operating system command. `ulimit` returns the values for key operating system parameters.

Example filenames(s):

```
ulimit.txt
```

version.txt

Build number of the software from which this bundle was downloaded.

Example filenames(s):

```
version.txt
```

Binary files

The following files are stored in binary format and not intended for customer consumption.

Example filename(s):

```
iv.data  
key.data.enc  
support-bundle.zip.enc
```



Copyright © 2022 - Trifacta, Inc.
All rights reserved.