

alteryx | TRIFACTA

Product Overview

Version: 9.7

Doc Build Date: 12/31/2023

Disclaimers

Except as otherwise provided in an express written agreement, Alteryx Inc. (“ Alteryx”) makes no representations or warranties with respect to the software and documentation contained herein and specifically disclaims any implied warranties of merchantability or fitness for any particular purpose. Furthermore, Alteryx reserves the right to revise the software and documentation from time to time without the obligation of Alteryx to notify any person of such revisions or changes.

Copyright and Trademark Notices

© 2013 - 2022 Alteryx Inc. All Rights Reserved.

Alteryx Inc.

17200 Laguna Canyon Road

Irvine, CA 92618

Phone: +1 888 836 4274

Alteryx and Trifacta are registered trademarks of Alteryx Inc.

Credits

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

See also [Third-Party License Information](#).

Product Overview

Contents:

- *What is Designer Cloud Powered by Trifacta Enterprise Edition?*
 - *Why use Designer Cloud Powered by Trifacta Enterprise Edition?*
 - *Machine Learning*
 - *How does it work?*
 - *Connectivity*
 - *Sampling*
 - *Distributed Processing Environments*
 - *Deployment Scenarios*
 - *What do you build in Designer Cloud Powered by Trifacta Enterprise Edition?*
 - *What else can you do in Designer Cloud Powered by Trifacta Enterprise Edition?*
 - *Visual Profiling*
 - *Transform to Target*
 - *Security and Authentication*
 - *Operationalization*
 - *Automation*
 - *Getting Started*
-

This section provides a short overview of Designer Cloud Powered by Trifacta® Enterprise Edition, its key features, and how they interact with each other.

What is Designer Cloud Powered by Trifacta Enterprise Edition?

Designer Cloud Powered by Trifacta® Enterprise Edition enables you to explore, combine, and transform diverse datasets for downstream analysis.

Within an enterprise, data required for key decisions typically resides in various silos. It comes in different formats, featuring different types. It is often inconsistent. It may require refactoring in some form for different audiences. All of this work must be done before you can begin extracting information valuable to the organization.

Data preparation (or data wrangling) has been a constant challenge for decades, and that challenge has only amplified as data volumes have exploded.

Why use Designer Cloud Powered by Trifacta Enterprise Edition?

Estimates vary, but something like 60% of an analyst's time is consumed with preparing data for use, leaving two days per week to actually analyze it. That's expensive and inefficient.



Did you know: This new category of software, called **data wrangling** or **data prep**, was invented by the founders of Alteryx, who created the first self-service data preparation tool. This joint Stanford/UC Berkeley release was called *Data Wrangler*.

Some organizations have pushed these cleansing efforts onto IT, which may take weeks to come up with a custom, scripted solution that requires inevitable back-and-forth between coder and analyst. As formats, feeds, and requirements change, these rigid solutions require frequent updating, which cannot be done by the people who really know the data. Instead of producing insights, analysts are filing requests and waiting for weeks for solutions.

The Trifacta solution delivers the tools to wrangle data to the people who understand the meaning of the data. With Designer Cloud Powered by Trifacta Enterprise Edition, analysts have the means to apply their expertise to the preparation of the data, in a way that is faster and more productive.

Designer Cloud Powered by Trifacta Enterprise Edition helps to do the following:

1. Cut time to prepare actionable data
2. Avoid IT bottlenecks and reliance on data scientists in data prep
3. Deliver tools to prepare data to the people who understand the data
4. Eliminate manual prep work
5. Surface data quality issues in a way that's easy to fix them

Featuring a leading-edge interface, powerful machine intelligence, and advanced distributed processing, Designer Cloud Powered by Trifacta Enterprise Edition renders the time-consuming, complex, and error-prone process of preparing datasets of any volume into a point-and-click exercise. What took six weeks in the IT lab can be done in less than two hours at the analyst's desk. Predictive Transformation

Humans are pretty good at identifying singular problems; software is better at solving them at scale. The platform leverages this concept through **predictive transformation**.

You see an issue in your sampled data. Whether it is part of a value, multiple values in a column, or the entire column itself, you select it. Immediately, the platform surfaces a set of suggestions for you. How would you like to transform this data?

The screenshot displays the Designer Cloud interface. At the top, there's a header with '2013 POS' and 'POS-r01'. Below this is a toolbar with various icons for data manipulation. The main area shows a data table with columns: Store_Nbr, WM_Week, Whse_Nbr, Whse_Name, and POS_Sal. The Whse_Name column is highlighted, and a 'Suggestions' panel on the right offers transformations like 'Extract values matching', 'Replace', 'Count values matching', and 'Split on values matching'. The 'Extract values matching' suggestion is selected, showing a preview of the results. The 'Replace' suggestion is also visible, showing a preview of the results. The 'Count values matching' suggestion is also visible, showing a preview of the results. The 'Split on values matching' suggestion is also visible, showing a preview of the results.

Figure: Select data elements to receive context-specific suggestions on transformations to apply to the element or to patterns that describe it. Preview the results before you add the change.

Through an innovative interface and leading-edge machine-learning techniques, Designer Cloud Powered by Trifacta Enterprise Edition surfaces potential actions on the data in an intuitive manner for rapid triage. Select something in the data grid, and the platform provides a set of context-specific recommendations of actions to take on the selected data and similar matching items. Click the recommendation, and your data is transformed. Modify the particulars of the transformation to get it right. No coding required.

Machine Learning

As you make selections, the platform's predictions become smarter and better. What you select today with this dataset informs the platform recommendations for transforming tomorrow's dataset.

Additionally, customers may opt-in to send anonymized usage data to Alteryx, so that the transformations being crafted across thousands of users can influence the machine-learning algorithms deployed in subsequent releases.

How does it work?

The scale and complexity of these transformations can quickly overwhelm even the most powerful of machines. Designer Cloud Powered by Trifacta Enterprise Edition utilizes a number of techniques to deliver high performance at scale.

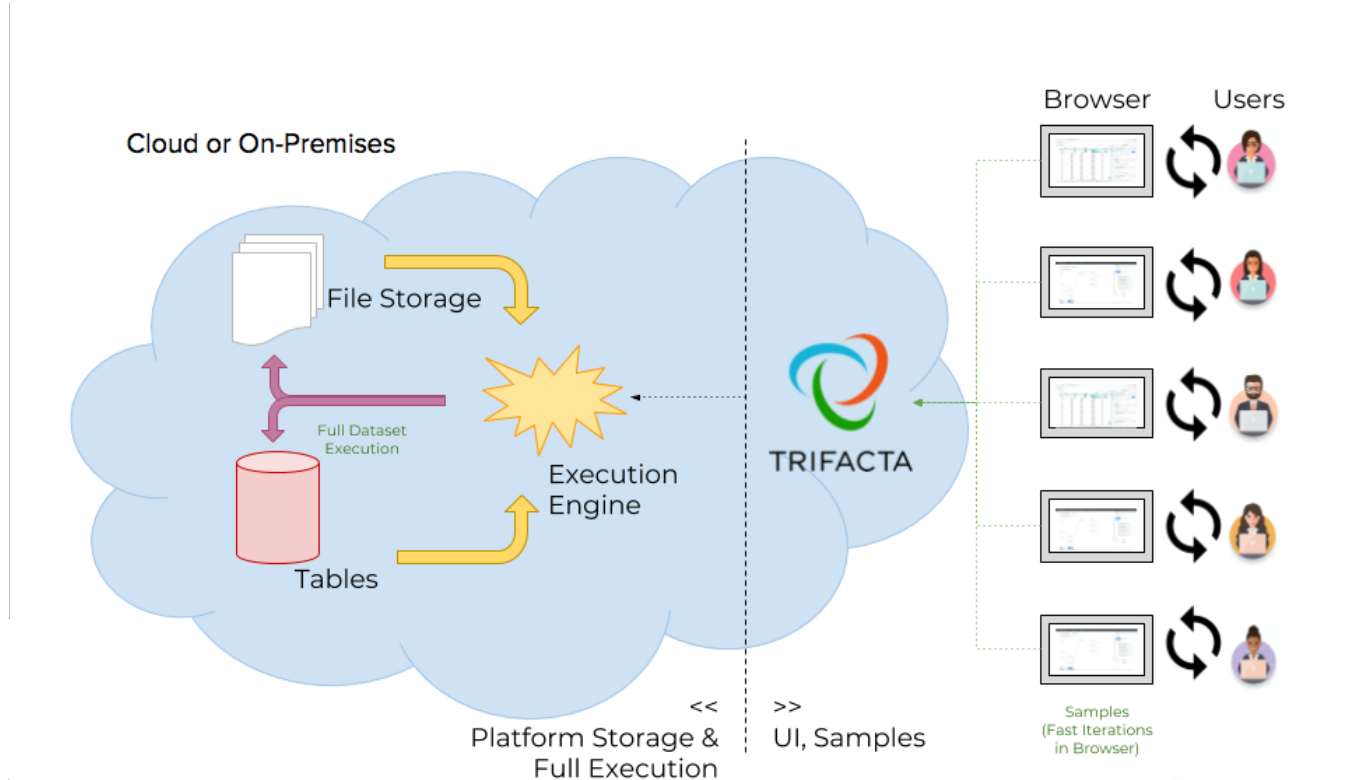


Figure: Platform interactions and data movements

Basic Steps:

Through a standards-compliant web browser, users interact with the platform through the user interface to complete the following tasks:

1. Create connections to sources of data. Some connections, such as desktop upload, are automatically created for you.
2. Through those connections, create references to specific datasets (files and tables).
3. Load samples of data into the user interface.
4. Build sequences of steps to transform the sampled data through innovative user interface controls.
5. Execute a job to perform that sequence of steps across the entire dataset, yielding an output dataset delivered to the preferred destination in the proper format.

NOTE: For large datasets, the transformation work is distributed across the nodes available to the integrated cluster.

6. As needed, export the results from the platform.

The above steps create a single sequence of steps from a single dataset. Datasets and sequences (recipes) can be combined or chained together to address much more complicated data wrangling requirements.

Connectivity



Feature Availability: This feature may not be available in all product editions. For more information on available features, see *Compare Editions*.

The platform supports creation of connections to the following:

- JDBC (relational) connections to a wide variety of database sources
- Read/write connectivity to distributed file-based storage
- Upload/download

Sampling

For larger datasets, loading all rows can quickly overwhelm the desktop system through which they are being viewed. Even if the local environment can handle the data volume, performing transformations becomes a cumbersome experience. For systems that do not support data sampling, the local desktop effectively becomes the limiting factor, variable as it is, on the size of the dataset.

Designer Cloud Powered by Trifacta Enterprise Edition overcomes limitations of the local desktop by sampling from datasets. When a dataset is loaded for use for the first time, a sample of the first set of rows is immediately taken. This sample size does not exceed 10 MB.

At any time, you can collect a new sample using one of several statistically useful sampling techniques. Random, filtered, anomaly-based, and stratified sampling are only some of the techniques available for use.

Samples are maintained and can be selected for use again at any time.

Distributed Processing Environments

Designer Cloud Powered by Trifacta Enterprise Edition includes the Trifacta Photon environment, an embedded running environment that is suitable for sampling and smaller-scale jobs.

For large-scale jobs, Designer Cloud Powered by Trifacta Enterprise Edition can integrate with a variety of leading-edge running environments. When it is time to execute a job, the platform distributes the workload to the nodes of the cluster's running environment, where the transformations are executed on the segment of source data stored on the node. These transformations occur on in-memory versions of the source data, with the results returned to Designer Cloud Powered by Trifacta Enterprise Edition for final assembly and export.

Tip: Processing of jobs within the Trifacta Photon running environment or across a distributed processing environment happens asynchronously. So, you kick off your job, resume working, and collect the results when the job is finished.

The platform supports integration with:

- Cloudera
- Azure Databricks
- AWS Databricks
- Amazon EMR

Deployment Scenarios

Designer Cloud Powered by Trifacta Enterprise Edition can be deployed within your enterprise infrastructure or onto various cloud-based infrastructures.

Deploy through Microsoft Azure if your sources are in ADLS Gen1, ADLS Gen2, or WASB. Install the product on AWS to wrangle S3 data. Deploy the product where it can take best advantage of the local data sources.

You can move results and flows between instances, as needed.

Tip: Don't forget to check out the free version of Designer Cloud Powered by Trifacta. Features may be released here before they are released in other products. Please visit <https://cloud.trifacta.com>.

What do you build in Designer Cloud Powered by Trifacta Enterprise Edition?

In Designer Cloud Powered by Trifacta Enterprise Edition, the primary object that you create is the recipe. A **recipe** is a sequence of transformation steps that you create to transform your source dataset. When you select suggestions, choose options from the handy toolbar, or select values from a data histogram, you begin building new steps in your recipe. After selecting, you can modify them through the Transform Builder, a context panel where your configured transformation can be modified and the changes previewed before saving them.

When you finish your recipe, you run a job to generate results. A **job** executes your set of recipe steps on the source data, without modifying the source, for delivery to a specified **output**, which defines location, format, compression and other settings.



Did you know: Designer Cloud Powered by Trifacta Enterprise Edition does not modify source data at all. All data is brought into the system through a reference to the source.

Datasets, recipes, and outputs can be grouped together into objects called flows. A **flow** is a unit of organization in the platform. Depending on your product, flows can be shared between users, scheduled for automated execution, and exported and imported into the platform. In this manner, you can build and test your recipes, chain together sets of datasets and recipes in a flow, share your work with others, and operationalize your production datasets for automated execution.

What else can you do in Designer Cloud Powered by Trifacta Enterprise Edition?

In addition to the above, the following key features simplify the data prep process and bring enterprise-grade tools for managing your production wrangling efforts.

Visual Profiling

For individual columns in your dataset, data histograms and data quality information immediately identify potential issues with the column. Select from these color-coded bars, and specific suggestions for transformations are surfaced for you. When you make a selection, you can optionally choose to display only the rows or columns affected by the change.

The screenshot shows a data transformation tool interface. On the left, a column named 'POS_Cost' is displayed with a red bar at the top indicating mismatched values. Below the bar, a list of values is shown, including several '.00' values, one '-.55', and one 'undefined'. A red bar is also visible at the bottom of the column, indicating the range of affected rows. On the right, a 'Suggestions' panel is open, showing four categories of suggestions: 'Delete rows', 'Keep rows', 'Create a new column', and 'Set'. Each category has a description of the action and buttons for 'Edit' and 'Add'.

Suggestions Panel:

- Delete rows**
with mismatched values in POS_Cost
Buttons: Edit, Add
- Keep rows**
with mismatched values in POS_Cost
- Create a new column**
flag mismatched values in POS_Cost
- Set**
mismatched values to `NULL()`
mismatched values to 0

At the bottom of the interface, there is a checkbox labeled 'Show only affected' which is checked, and a label 'Rows'.

Figure: Click the red bar to select all mismatched values in the column. Show only the affected rows. Review suggestions for how to fix these specific values.

Tip: Dig into the column details to explore distributions of values based on the column's data type.

As part of your transformation job, you can optionally generate a visual profile of your dataset, which allows you to quickly identify areas for additional iteration.

Transform to Target

Since data wrangling targets are often other systems with well-defined input requirements, the structure of the target data is typically known in advance. To assist in your data wrangling efforts, you can import a representation of the target structure into your flow, assign it to a recipe, and then use specific tools to rapidly transform your dataset to this target. Match up fields in your source to the target using name, data type, and position in the schema. Then, pattern-match source data to expectations in the target field to ensure that you are delivering the appropriate data for downstream requirements.

Security and Authentication

The platform supports enterprise-grade authentication method through LDAP/AD and supports cluster-based security through secure authentication and Kerberos. Platform interactions can be configured to work over SSL, and database integration can be protected through encrypted key technologies.

Operationalization

After you have finalized development of your flow, you can operationalize its execution. Using a simple interface, you can define when the flow is executed on a periodic basis.

Tip: Datasets can be parameterized. For example, you can store a set of files in a single directory and reference all of them through a single dataset with some parameterized value. When referencing flows are executed, the transformation steps are applied to all source files.

Automation

Scheduling

You can schedule the execution of your flows through the application. See *Overview of Scheduling*.

Plans



Feature Availability: This feature may not be available in all product editions. For more information on available features, see *Compare Editions*.

Through the Designer Cloud application, you can create plans, which are sequence of flow execution tasks. For more information, see *Plans Page*.

REST APIs

The platform supports automation through externally available REST APIs.



Copyright © 2022 - Trifacta, Inc.
All rights reserved.