

Workflow Basics

Contents:

- *Overview*
 - *Prerequisites*
 - *Basic Workflow*
-

Learn the basics of how to import, wrangle, execute jobs, profile, and export your data from Trifacta® Self-Managed Enterprise Edition.

Overview

Trifacta® Self-Managed Enterprise Edition enables analysts, data specialists, and other domain experts to quickly cleanse and transform datasets of varying sizes for use throughout the enterprise. Using an innovative set of web-based tools, you can import complex datasets and wrangle them for use in virtually any target system. Key capabilities include:

- **Import** from flat file, databases, or distributed storage systems
- Locate and remove or modify **missing or mismatched** data
- **Unnest complex** data structures
- Identify statistical **outliers** in your data for review and management
- Perform **lookups** from one dataset into another reference dataset
- Aggregate columnar data using a variety of **aggregation functions**
- **Normalize** column values for more consistent usage and statistical modeling
- Merge datasets with **joins**
- Append one dataset to another through **union** operations

Most of these operations can be executed with a few mouse clicks. This section provides a basic overview of common workflows through Trifacta Self-Managed Enterprise Edition.

Prerequisites

Before you begin, please verify the following:

- **Trifacta account:** You have a Trifacta account and can login.
- If you are importing data from a location other than your local file system, you must have the appropriate role in your user account. See *Import Basics*.
- **Example data:** You should use a sample set of data during this workflow.

Basic Workflow

1. **Import data:** Integrate data from a variety of sources of data.

Tip: When you login for the first time, you can immediately upload a dataset to begin transforming it.

See *Import Basics*.

2. **Profile your data:** Before, during, and after you transform your data, you can use the visual profiling tools to quickly analyze and make decisions about your data. See *Profiling Basics*.

3. **Build transform recipes:** Use the various views in the Transformer Page to build your transform recipes and preview the results on sampled data. See *Transform Basics*.
4. **Sample your data:** In Trifacta Self-Managed Enterprise Edition, you create your recipes while working with a sample of your overall dataset. As needed, you can take new samples, which can provide new perspectives and enhance performance in complex flows. See *Sampling Basics*.
5. **Run job:** Launch a job to run your recipe on the full dataset. Review results and iterate as needed. See *Running Job Basics*.
6. **Export results:** Export the generated results data for use outside of Trifacta Self-Managed Enterprise Edition. See *Export Basics*.

Object overview: You should review the overview of the objects that are created and maintained in Trifacta Self-Managed Enterprise Edition. See *Object Overview*.