

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**.

Import Data

Search... (/)

Upload

HDFS

S3

hive

Upload from your computer

Drag & drop a file here or

Choose a file

NAME	SIZE
+ CUSTOMER DATABASE - 2.txt	676kB
+ CUSTOMER DATABASE.csv	676kB
ATM.log	8MB/11MB
+ ATM_TRANSCODES.csv	51B

+ New Edit

3 New Datasets Clear All

CUSTOMER DATABASE.csv

Add a Description

ACCOUNT	END_DATE
4208107376653317	2014-01-03
4208106585213952	2014-10-27
4208109505943761	2012-02-18
4208113686690378	2015-06-30
4208104023796892	2015-08-04

Edit settings

CUSTOMER DATABASE - 2.

Add a Description

ACCOUNT	END_DATE
4208107376653317	2014-01-03
4208106585213952	2014-10-27

Add Datasets to a Flow

ATM Demo

Import & Add to Flow Cancel

Figure: Import Data page

General Limitations

NOTE: For file-based sources, Trifacta® Wrangler 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.

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 (_).

Basic Workflow

1. Connect to sources

NOTE: Compressed files are recognized and can be imported based on their file extensions.

Upload: Trifacta® Wrangler 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 100 MB 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.

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

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*.

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.

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

You can modify settings used during import for individual files. In the card for an individual dataset, click **Edit Settings**.

NOTE: In some cases, there may be discrepancies between row counts in the previewed data versus the data grid after the dataset has been imported, due to rounding in row counts performed in the preview.

- **Per-file encoding:** By default, Trifacta Wrangler attempts to interpret the encoding used in 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.
- **Detect structure:** By default, Trifacta Wrangler 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*.

4. Import selections

Single dataset

If you have selected a single dataset for import:

1. To immediately wrangle it, click **Import & Wrangle**. The dataset is imported. A recipe is created for it, added to a flow, and loaded in the Transformer page for wrangling. See *Transformer Page*.
2. To import the dataset, click **Import**. The imported dataset is created. You can add it to a flow and create a recipe for it later. See *Library 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.

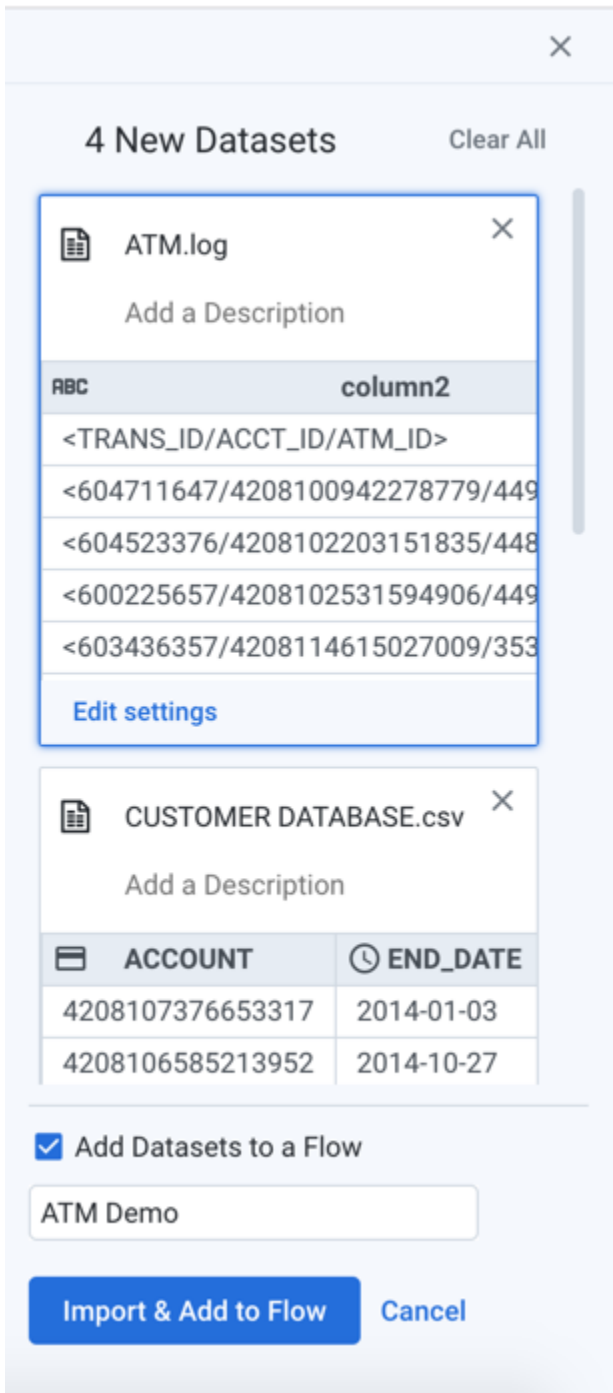


Figure: Import Multiple Datasets

If you have selected multiple datasets for import:

- To import the selected datasets, click **Import Datasets**. The imported datasets are created. You can begin working with these imported datasets now or at a later time. If you are not wrangling the datasets immediately, the datasets you just imported are listed at the top of the Library page. See *Library Page*.
- To import the selected datasets and add them to a flow:
 1. Click the Add Dataset to a Flow checkbox.
 2. Click the textbox to see the available flows, or start typing a new name.
 3. Click **Import & Add to Flow**.

4. Create or select the flow to which to add them:

The screenshot shows a dialog box titled "Add 6 datasets to...". At the top, there are navigation arrows and a search bar containing the letter "T". Below the search bar, there are three tabs: "All Flows" (selected), "Owned by me", and "Shared with me". A table lists several flows, each with a "Test-Notification" name, a unique owner ID, and a timestamp. The third row is highlighted in blue and has a checkmark in the left margin. At the bottom of the dialog, there are three buttons: "Create new flow", "Cancel", and "Add".

Name	Owner	Contains	Last Updated
Test-Notification	Test User21147982	2 Datasets, 2 Recipes	Today at 1:48 PM
Test-Notification	Test User26026347	2 Datasets, 2 Recipes	Today at 1:43 PM
Test-Notification	Test User57416298	2 Datasets, 2 Recipes	Today at 1:38 PM
Test-Notification	Test User93058100	2 Datasets, 2 Recipes	Today at 1:35 PM
Test-Notification	Test User31301016	2 Datasets, 2 Recipes	Today at 1:33 PM
Test-Notification	Test User76694503	2 Datasets, 2 Recipes	Today at 1:19 PM
Test-Notification	Test User63450053	2 Datasets, 2 Recipes	Today at 1:17 PM

Figure: Add imported datasets to a flow

- a. **Filter by Type:** Click one of the tabs in the dialog to display only the applicable flows.
 - b. **Search:** Start typing letters to filter the list of flows.
 - c. **Create new flow:** Enter a name and description for the new flow to which to add the datasets.
 - d. To add the datasets to the selected flow, click **Add**.
5. The datasets are imported, and the associated recipes are created. These datasets and recipes are added to the selected flow.
6. For any dataset that has been added to a flow, you can review and perform actions on it. See *Flow View Page*.
- To remove a dataset from import, click the X in the dataset card.