

# Import Data Page

## Contents:

- *General Limitations*
- *Basic Task*
  - 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**.

The screenshot shows the 'Import Data' interface. On the left, there is a sidebar with a search bar and a list of sources: Upload, S3, Trifacta, and s3user. The main area is titled 'Upload from your computer' and contains a large dashed box with the text 'Drag & drop a file here or' and a 'Choose a file' button. Below this is a table listing the uploaded files:

NAME	SIZE
+ [icon] REF_PROD.txt	13kB
+ [icon] REF_CAL.txt	56kB
+ [icon] POS-r03.txt	126kB
+ [icon] POS-r02.txt	274kB
+ [icon] POS-r01.txt	286kB

On the right, there is a panel titled '5 New Datasets' with a 'Clear All' button. It shows two dataset preview cards. The first card is for 'POS-r02.txt' and contains a table:

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

The second card is for 'POS-r03.txt' and contains a table:

#	Store_Nbr	Item_Nbr	#
200	322000	20	20
200	323000	20	20
200	325000	20	20
200	326000	20	20

At the bottom of the right panel, there is a checkbox labeled 'Add to new flow' which is checked, and two buttons: 'Cancel' and 'Continue'.

**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.
- 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:** 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 Task

### 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® 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 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.

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.

**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*.

### 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.

The screenshot shows a dialog box titled "5 New Datasets" with a "Clear All" button. It contains two dataset preview cards. The first card is for "POS-r02.txt" and includes a table with 5 rows of data. The second card is for "POS-r01.txt" and includes a table with 4 rows of data. At the bottom, there is a checkbox for "Add to new flow" and two buttons: "Cancel" and "Continue".

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

  

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

**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.