

Import Excel Data

In addition to CSV and other formats, Trifacta® Self-Managed Enterprise Edition can directly import Microsoft® Excel® workbooks. The worksheets of a workbook can be imported as:

- Individual datasets
- A single dataset
- A dataset with parameters

i NOTE: When importing as a parameterized dataset, all selected worksheets are imported into a single dataset.

Workbooks can be uploaded from your local system. If Trifacta Self-Managed Enterprise Edition is connected to a backend file storage system, you can also import workbooks stored in readable directories.

Limitations:

- Source row number information is not available for Excel-based datasets.
- Source Excel files with cells bracketed by single double quotes may not be properly ingested if any terminating quotes are missing.

✓ Tip: You can check the data quality bars for mismatched values or, for strings, the data histogram bars for anomalous values to see if the above issue is present. If so, deselect Detect Structure on import. Then, use a Split rows transformation applied to the affected column to break up the column as needed.

You cannot import password-protected Excel files.

- Compressed Excel files are not supported.
- Conversion of large Excel files require non-linear increases in memory requirements on the Trifacta node.
- If loading your Excel-based dataset in the Transformer page results in a blank screen, please take a new sample. The file requires conversion again with each generated sampling.
- Latest state of the Excel file may not be reflected in the Transformer page due to caching. When you run a job, the platform always collects the latest version of the data and converts it to CSV for execution.
- For datasets with parameters:
 - Generally, mixing Excel files with non-Excel files in a single dataset with parameters is not recommended.
 - Parameterized datasets utilize the first matching file for display in the UI, including initial sample generation. The type of file and schema of the dataset are determined by the first match. After creating the dataset with parameters:
 - If you modify the parameters such that an Excel file is no longer the first matching file, you may see an error similar to the following:

```
First matched file cannot be changed from Excel or vice-versa.
```

Excel is a special case of first file matching. Please modify parameters so that an Excel file is the first matching file.

- The schema of the dataset is determined by the first matching file. If you modify the parameters of your dataset after creation such that the first match file has changed, you can corrupt the dataset schema, as the new Excel file may have different sheets with different names.
 - In Flow View, you must edit the dataset settings on the imported dataset object. Select the proper sheets to use in the new first matching file.
 - If you do not fix this issue, results may be unexpected in the application.

Steps:

1. In the menu bar, click **Library**.
2. In the Library page, click **Import Data**. Select the connection to use. See *Import Data Page*.

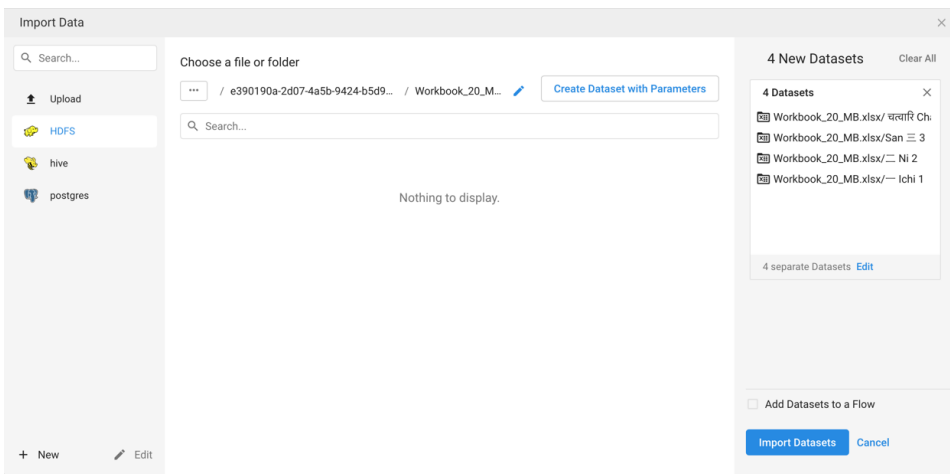


Figure: Import Excel workbook



Tip: If you experience issues uploading large XLS/XLSX files, you can convert the files to CSV files and then upload them.

3. After you select the workbook, it is uploaded and converted to CSV format and stored by the platform. Depending on the size of the workbook, this process may take a while.
4. By default, all worksheets in the workbook are imported as individual datasets. To change how the data is imported, click **Edit** in the right panel.

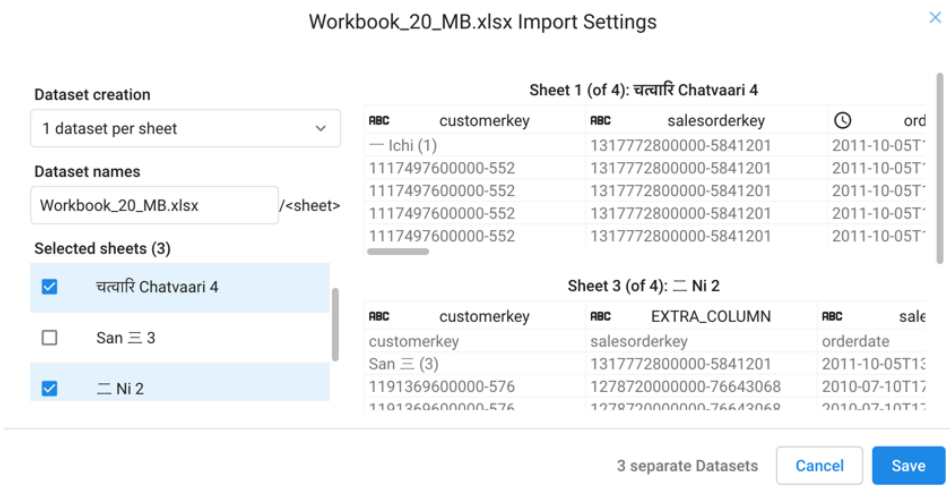


Figure: Import settings for Excel datasets

5. Dataset creation:
 - a. **1 dataset per sheet:** (Default) Each selected sheet in the workbook is imported as a separate dataset.
Specify the base name of the datasets that you are creating. If you are creating a single dataset, the name of the workbook is used.
 - b. **Selected sheets into 1 dataset:** All selected sheets in the workbook are combined and imported as a single dataset.

i **NOTE:** The schemas of each dataset must match. Columns must be listed in the same order in each dataset. The column headers are taken from the first selected dataset.

- c. **All and future sheets into 1 dataset:** If the workbook is updated periodically with new sheets that you would like to add in the future, select this option. After initial selection of sheets, all sheets that are added to the workbook in the future are automatically added as part of the imported dataset.

i **NOTE:** This option is available only if you are connected to a backend file storage system, such as HDFS, S3, or GCS.

6. Selected sheets: You can select the sheets to import. To preview the data of an individual sheet, mouse over a dataset and click **Jump**.
7. To save changes, click **Save**.
8. After your datasets have been added, you can edit the name and description information for each in the right navigation panel.
9. Optionally, you can assign the new dataset(s) to an existing flow or create a new one to contain them.