

UI Integration - Create Dataset

Contents:

- [Pre-requisites](#)
- [Authentication](#)
- [Sources of Data](#)
- [Step-by-Step Guide](#)

Using the following URL endpoint, you can create a dataset from another application through the Trifacta application.

NOTE: This integration is not supported in the Wrangler Enterprise desktop application.

Pre-requisites

- If you are calling from a source application, you must be logged into that application first. See [Authentication](#) below.
- You must authenticate with the Trifacta platform before you are redirected to the target destination. See [API - UI Integrations](#).
- This URL integration is supported on HDFS and S3 datastores.
- It is assumed that there are no conflicting datasets with the names that are used to create the dataset in this set of steps. No name validation is performed as part of this action.

Authentication

NOTE: Before using any UI integration, you must first login to the application. If you are not logged in, you are redirected to the login page, where you can input your credentials before reaching your target URL.

In addition to authentication with the Trifacta platform, the authenticated user must also have the appropriate permissions to access the assets on the datastore. This includes:

- Permissions to access the folder or directory
- Appropriate impersonated user configured for the account, if secure impersonation is enabled.
- If this dataset is going to be executed later via command line interface, you must create the dataset with the same user that will execute the job.

For more information:

Topic	Section
HDFS: permissions and security	See Configure Hadoop Authentication .
HDFS: usage	See Using HDFS . See HDFS Browser .
S3: permissions and security	See Enable S3 Access .
S3: usage	See Using S3 . See S3 Browser .

Sources of Data

You can use this integration to create datasets from single files or a single directory. Below are some example URLs for sources from Hadoop HDFS or S3:

Datastore	Source Type	Example URL	Results
HDFS	Directory	hdfs:///user/warehouse/campaign_data/	User can choose the file through the UI to use for the dataset.
	File	hdfs:///user/warehouse/campaign_data/d000001_01.csv	User can complete the steps through the UI to create the dataset.
S3	Directory	s3:///3fad-demo/data/biosci/source/	User can choose the file through the UI to use for the dataset.
	File	s3:///3fad-demo/data/biosci/source/1-DRUG15Q1.txt	User can complete the steps through the UI to create the dataset.

NOTE: The above results assume that the user has the appropriate permissions to access the file or directory. If the user lacks permissions, an HTTP 404 error is displayed.

Step-by-Step Guide

Steps:

1. Acquire the target URL for the datastore through the Trifacta® application or through the datastore itself. Examples URLs:
 - a. HDFS (file):

```
hdfs:///user/warehouse/campaign_data/d000001_01.csv
```

- b. S3 (directory):

```
s3:///3fad-demo/data/biosci/source/
```

2. Navigate the browser to the appropriate URL in the Trifacta platform. The following example applies to the HDFS file example from above. It must be preceded by the base URL for the platform. For more information, see *API - UI Integrations*.

```
<base_url>/import/data?uri=hdfs:///user/warehouse/campaign_data/d000001_01.csv
```

3. For file-based URLs, the file is selected automatically.
4. For directory-based URLs, the user can select which ones to include through the browser. Click the Add Datasets to a Flow. Add the dataset to an existing flow or create a new one for it.
5. After the datasets have been imported, open the flow in which your import is located. For the datasets that you wish to execute, you should do the following in the Flow View page:
 - a. Click the icon for the dataset.
 - b. From the URL, retrieve the identifiers for the flow and the dataset. These values are needed for later execution through the command line interface.
 - c. Example:

Dataset URL	flowId	datasetId
http://example.com:3005/flows/31#dataset=186	31	186

The flow is consistent across all datasets that you imported through the above steps.

6. You can open the datasets and wrangle them as needed.
7. Complete any required actions from within your source application.

You can run jobs on the dataset through the following interfaces:

- **UI:** See *Run Job Page*.
- **API:** See *API JobGroups Create v3*.
- **CLI:** See *CLI for Jobs*.