API ImportedDatasets Create v3

⚠ As of Release 6.4, the v3 APIs are End of Life (EOL). Before using this release or a later one, you should migrate to using the latest available API version. For more information, see *API Version Support Matrix*. Latest version of this endpoint: *API ImportedDatasets Create v4*.

Contents:

Create an imported dataset from an available resource. Created dataset is owned by the authenticated user.

> **NOTE:** When an imported dataset is created via API, it is always imported as an unstructured dataset. Any recipe that references this dataset should contain initial parsing steps required to structure the data.

> **NOTE:** Do not create an imported dataset from a file that is being used by another imported dataset. If you delete the newly created imported dataset, the file is removed, and the other dataset is corrupted. Use a new file or make a copy of the first file first.

Version: v3

Required Permissions

> **NOTE:** Each request to the Trifacta® platform must include authentication credentials. See *API Authentication*.

Request and Response

Request Type: `POST`

Endpoint:

```
/v3/importedDatasets
```

Response Status Code - Success: `201 - Created`

Examples by Type

Below, you can review the basic request body for creating imported datasets for various types of sources:

- File (HDFS or S3 source)
- Hive
- Relational
- Relation with Custom SQL Query

File (HDFS and S3 sources)

Request Body - HDFS file:

Below, the `bucket` value is set to `null`. This parameter applies only to S3 sources.

> **NOTE:** The `path` value should not include the HDFS protocol, host, or port information. You only need to provide the path on HDFS.

```
{
  "path": "/trifacta/uploads/1/4aee9852-cf92-47a8-8c6a-9ff2adeb3b4a/POS-r02.txt",
  "type": "hdfs",
  "bucket": null,
  "name": "POS-r02b.txt",
  "description": "POS-r02 - copy"
}
```

Request Body - S3 file:

For S3 sources, a bucket must be specified. Below, the `bucket` value is set to `myBucket`.

> **NOTE:** The `path` value should not include the S3 protocol, host, or port information. You only need to provide the path on S3.

```
{
  "path": "/trifacta/uploads/1/4aee9852-cf92-47a8-8c6a-9ff2adeb3b4a/POS-r02.txt",
  "type": "s3",
  "bucket": "myBucket",
  "name": "POS-r02b.txt",
  "description": "POS-r02 - copy"
}
```

Response Body - file:

Following example is for an HDFS file. For an S3 file, `type=s3`.

```
{
  "id": 8,
  "size": "281032",
  "path": "/trifacta/uploads/1/4aee9852-cf92-47a8-8c6a-9ff2adeb3b4a/POS-r02.txt",
  "isSharedWithAll": false,
  "type": "hdfs",
  "bucket": null,
  "isSchematized": false,
  "createdBy": 1,
  "updatedBy": 1,
  "updatedAt": "2017-02-08T18:38:56.640Z",
  "createdAt": "2017-02-08T18:38:56.560Z",
  "connectionId": null,
  "parsingScriptId": 14,
  "cpProject": null
}
```

Hive

Request Body - Hive:

Notes:

- Note that the `type=jdbc`.
- The `columns` key is optional. If not provided, all columns in the source table are included.

```
{
  "visible": true,
  "numFlows": 0,
  "size": -1,
  "type": "jdbc",
  "jdbcType": "TABLE",
  "jdbcPath": [
    "DB1"
  ],
  "jdbcTable": "MyHiveTable",
  "columns": [
    "column1",
    "column2"
  ],
  "connectionId": 16,
  "name": "My Hive Table"
}
```

Response Example - Hive:

```
{
  "jdbcTable": "MyHiveTable",
  "jdbcPath": [
    "DB1"
  ],
  "columns": [
    "column1",
    "column2"
  ],
  "filter": null,
  "raw": null,
  "isSharedWithAll": false,
  "id": 192,
  "size": "-1",
  "type": "jdbc",
  "connectionId": 16,
  "createdBy": 1,
  "updatedBy": 1,
  "updatedAt": "2017-02-17T18:09:17.745Z",
  "createdAt": "2017-02-17T18:09:16.407Z",
  "path": null,
  "bucket": null,
  "parsingScriptId": 366,
  "cpProject": null,
  "isSchematized": true
}
```

Relational

Request Body - Relational:

Notes:

- If you know the `size` value for the table, please provide. It is helpful for performance reasons and validation but is not required.
- The `columns` key is optional. If not provided, all columns in the source table are included.

```
{
  "visible": true,
  "numFlows": 0,
  "size": 65536,
  "type": "jdbc",
  "jdbcType": "TABLE",
  "jdbcPath": [
    "OracleDB_1"
  ],
  "jdbcTable": "MyOracleTable",
  "columns": [
    "I",
    "J",
    "K"
  ],
  "connectionId": 7,
  "name": "My Oracle Table"
}
```

Response Example - Relational:

```
{
  "jdbcTable": "MyOracleTable",
  "jdbcPath": [
    "OracleDB_1"
  ],
  "columns": [
    "I",
    "J",
    "K"
  ],
  "filter": null,
  "raw": null,
  "isSharedWithAll": false,
  "id": 195,
  "size": "65536",
  "type": "jdbc",
  "connectionId": 7,
  "createdBy": 1,
  "updatedBy": 1,
  "updatedAt": "2017-02-17T18:10:48.662Z",
  "createdAt": "2017-02-17T18:10:47.441Z",
  "path": null,
  "bucket": null,
  "parsingScriptId": 372,
  "cpProject": null,
  "isSchematized": true
}
```

Relational with Custom SQL Query

You can submit custom SQL queries to relational or hive connections. These custom SQLs can be used to pre-filter the data inside the database, improving performance of the query and the overall dataset.

- For more information, see *Enable Custom SQL Query*.

Request Body:

Notes:

- See previous notes on queries to relational sources.
- As part of the request body, you must submit the custom SQL query as the value for the `raw` property.

The following example is valid for Oracle databases. Note the escaping of the double-quote marks.

> **NOTE:** Syntax for the custom SQL query varies between relational systems. For more information on syntax examples, see *Create Dataset with SQL*.

```
{
  "visible": true,
  "numFlows": 0,
  "type": "jdbc",
  "jdbcType": "TABLE",
  "connectionId": 7,
  "raw": "SELECT INST#,BUCKET#,INST_LOB# FROM \"AUDSYS\".\"CLI_SWP$7395268a$1$1\"",
  "size": -1,
  "name": "SQL Dataset 1"
}
```

Response Body:

In the response, note that the source of the data is defined by the `connectionId` value and the SQL defined in the `raw` value.


```
{
  "jdbcTable": null,
  "jdbcPath": null,
  "columns": null,
  "filter": null,
  "raw": [
    "SELECT INST#,BUCKET#,INST_LOB# FROM \"AUDSYS\".\"CLI_SWP$7395268a$1$1\""
  ],
  "isSharedWithAll": false,
  "id": 196,
  "size": "-1",
  "type": "jdbc",
  "connectionId": 7,
  "createdBy": 1,
  "updatedBy": 1,
  "updatedAt": "2017-02-17T19:09:10.117Z",
  "createdAt": "2017-02-17T19:07:12.757Z",
  "path": null,
  "bucket": null,
  "parsingScriptId": 378,
  "cpProject": null,
  "isSchematized": true
}
```


Reference

For more information on the properties of an imported dataset, see *API ImportedDatasets Get v3*.