Create Custom Data Types

Contents:

- *Create dictionary file for custom data type*
- *Example - Sizes*
- *Create the data type*
- *Use a custom data type*

A custom data type requires that you create and upload a CSV dictionary file. This dictionary file includes all accepted values for the custom data type.

> **NOTE:** The method described in this section validates against a fixed set of values. If you would like to validate your custom type against a pattern, you can specify the pattern using RegEx. See *Create Custom Data Types Using RegEx*.

A **dictionary** represents one or more columns of reference data, which can be used for data validation. For defining custom types, if a value is included in the dictionary, it is a valid member of the custom type. For example, you may wish to create a custom type called `storeId`, which contains a valid identifiers for stores in your enterprise.

> **After a custom type has been added, it cannot be removed or disabled.**

Create dictionary file for custom data type

For your custom data type, you must create a dictionary of values in your local environment. This file is then uploaded to Trifacta® Wrangler Enterprise.

File characteristics:

- CSV file format
- File can be multi-column, but data validation only uses one of the columns.
- File can contain up to 250,000 values. If your data type contains more values than this limit, you might see values in your dataset identified as members of the data type, when they are not.
- Remove any header row from your file.
- Values are case-insensitive during matching.
- Special restrictions on the newline (\n) character are described below.

Notes and Limitations:

While you can use newline as a delimiter, dictionaries do not support using the newline (\n) character within a cell value. if your dictionary includes this character in cell values, it is dropped from use in the generated dictionary. In the following CSV example data, the first row is acceptable, while the second is not:

```
"Arizona"\n"Alaska"\n
"Arizona\n"\n"Alaska"
```

Example - Sizes

For example, your data contains size information from Extra Small (XS) to Extra Extra Large (XXL). You can create a one-column dictionary file with values for these sizes on separate lines. This dictionary file could be use to validate the custom type `Sizes`. Your data might look like the following. Note that the column has no header.

Extra Small
Small
Medium
Large
Extra Large
Extra Extra Large
XS
S
M
L
XL
XXL
Extra-Small
Extra-Large
Extra-Extra-Large

You can download this source file: *Dict-Sizes.csv*.

Create the data type

Please complete the following steps to create the custom type.

> **NOTE:** After a custom type has been created, a platform restart is required. Please contact your Trifacta administrator.

Steps:

1. Click the data type drop-down in a column where you want to apply the custom type.
2. Click **More**. Scroll down and click **Custom Type**.
3. The Custom Type dialog is displayed. Click the Create New Custom Type tab.
4. Click **Upload Dictionary**. Select the CSV file you created. Click **Open**.

> **NOTE:** After you upload a dictionary file, it cannot be removed. If necessary, upload a new version with a different filename.

5. The file is uploaded:

Figure: Custom Data Type dialog

6. Click the caret next to the filename to review the contents of the dictionary.
7. Select one of the column headers in the left side of the dialog. On the right side, you can review values in the selected column.

> **NOTE:** You must expand the custom dictionary to see values before you can save the custom type. This is a known issue.

8. Select the column you want to use for validating the custom type.
9. Enter a name for the data type.

> **NOTE:** This name appears in the data type drop-down for each column. Also, it can be referenced explicitly in transforms that utilize a named data type as a parameter.

10. Click **Save**.
11. Restart the platform. See *Start and Stop the Platform*.

For more information, see *Custom Type Dialog*.

Use a custom data type

Steps:

1. Select **Custom Type** from the column drop-down.
2. In the Custom Type dialog, click the Use Existing Custom Type tab.

> **NOTE:** If you cannot see a recently created custom data type, you may need to logout and login again.

3. Select the name of the custom type. Click **Save**.
4. When the data type is saved, the values in the column are validated against this custom type.
5. Make sure you review the missing and mismatched values for the column.

> **Tip:** You can also reference the data type by name in your transforms.

For more information, see *Custom Type Dialog*.