

Add Headers to Your Columns

Contents:

- *If your data has a header row in row 1*
- *If your data has a header row after row 1*
- *If your data does not have a header row*

A column is referenced by the name of the column, which can be inferred from the first row of data in your dataset.

When a dataset is loaded, the application inserts a few transform steps automatically. If the application can identify that the first row of data is likely to contain the column headers for the dataset, this row is promoted to be used as the first version of the names of each column.

In some cases, however, this auto-generation of column headers may not work as expected, or you may have chosen at import time to not detect the structure of the dataset.

This section describes how you can generate column headers from within the application.

If your data has a header row in row 1

If the initial transforms do not promote your first row of data to be the column headers, you can use the following transform to promote the first row of data to be the column headers:

Transformation Name	Rename columns with a row
Parameter: Option	Use row as header
Parameter: Row	1

In some cases, the first row of data might not contain the headers or might not contain all of them.

For example, you may have some columns that contain nested data, and the column headers may not be immediately accessible.

 **Tip:** After you unnest data in one or more columns, the first row might contain column headers. You can apply the `header` transform to promote these new values to be the names of the columns. The other column headers should not be overwritten.

If your data has a header row after row 1

In some cases, data may be imported such that header information is stored in a row other than the first one in your dataset.

 **NOTE:** If source row number information is not available due to changes in the dataset, this transform may not be available.

Steps:

1. Hover your mouse over the black dot to the left of the row that contains your header information. The popup displays something similar to the following:

Row 12
Source Row 12

2. Add a transform step using the source row number that you found:

Transformation Name	Rename columns with a row
Parameter: Option	Use row as header
Parameter: Row	12

3. Add it to your recipe.

If your data does not have a header row

If for some reason your source data does not include header information, you can insert header information using the following method.

i **NOTE:** In general, it is easiest to manually rename columns through the application. See *Rename Columns*.

However, if your data contains a large number of columns, manually renaming each column may be time-consuming, and each column rename adds a step to your recipe. For wide datasets, this solution may be easier to execute and to maintain.

Steps:

1. Open the dataset in the Transformer page.
2. Open an application such as Microsoft Excel, which can write out CSV files.
3. For each column in the Transformer page, add a string name for the column in the other application.
 - a. If you are using Excel, insert this column name in the top row of the spreadsheet, with each new column added in the cell to the right of the previous one.
 - b. If you are using another application, make sure that you are inserting commas between each value and putting your column names between double-quotes.

i **NOTE:** You may need to create a dummy second row, which forces the application to treat the imported dataset as multiple columns. Otherwise, it may treat the incoming CSV as a single columnar value.

4. Among the column headers, locate a column by which you are comfortable sorting the dataset.
 - a. For example, if your dataset includes transaction information, you may want to sort the data by the primary key `TransactionId` column.
 - b. Rename this column to prepend the name with `aaa`. For our `TransactionId` column, the new column name would be the following:

```
aaaTransactionId
```

- c. This modification enables the sorting of the values in the dataset, starting with this row. All rows in the dataset are sorted according to this column.
5. Save the column header file in CSV format.
 6. Login to Trifacta Self-Managed Enterprise Edition. Create a new dataset from this CSV file. See *Import Data Page*.
 7. Load the dataset without headers.
 8. In the Transformer page, click the Tools menu. Select **Union**.

9. In the Union tool, append this dataset to the dataset that contains its headers. Include all columns. See *Union Page*.
10. Execute the union to create the appended dataset.
11. In the append dataset, the column header row is the last one. Locate the column that contains the column header by which you want to sort.
12. Check the data type in this column. Change it to String type, if it is not currently. Don't worry about mismatched values; you will switch it back at the end of these steps.
13. Next to the current column name for this value, click the drop-down caret. Select **Sort ascending...**
14. The appended dataset is sorted according to the values in this column. The `aaaTransactionId` value should be at the top of the dataset.
15. Now, use this row to create your headers:

Transformation Name	Rename columns with a row
Parameter: Option	Use row as header
Parameter: Row	1

16. Rename the column header to remove the `aaa` identifier.
17. Change the data type back to its original value, if necessary.