

# Profile Your Source Data

You might want to execute a profile of the data that you imported from the source. As soon as you create a recipe from a source, you can execute a job to profile the dataset.

By profiling the data as soon as you load it into the Transformer page, you can assess the following:

- Identify problems in the source and potentially correct them in the source system.
- Create a baseline to evaluate the data wrangling work you do in Cloud Dataprep by TRIFACTA® INC..
- Identify mismatched or missing values.

**Tip:** You can also use this technique to generate an output of your source data, which is useful if you do not have read access to the source outside of Cloud Dataprep by TRIFACTA INC..

## Steps:

1. Create an imported dataset from your source. Add it to a flow. See *Import Data Page*.
  - a. Depending on how your data is structured, you may choose to disable Detect Structure. For more information, see *Initial Parsing Steps*.
2. In Flow view, create a recipe for your imported dataset. See *Flow View Page*.
3. In Flow view, edit the newly created recipe. It is opened in the Transformer page. See *Transformer Page*.
4. If needed, add a header step to your dataset.
5. Click **Run**.
6. In the Run Job page, select the following options:
  - a. If you have the option of selecting a running environment, select the default one. This option may not be available in your product.
  - b. CSV format (you need at least one format to generate your dataset's profile).
  - c. Select to profile results.
7. Click **Run**.
8. When the results are generated, click **View Results**.
9. A profile of your dataset is displayed.

In the generated profile, you can identify:

- Missing or mismatched values in each column
- Statistical break-out by quartile
- Beginning dataset size and baseline job execution speed

**Tip:** You might want to write down the overall statistics for the dataset, which may be useful when validating the changes you have applied through recipe.

You might also download the dataset for recordkeeping. See *Job Details Page*.

## Preserve Source Visual Profile

If you wish to preserve the capability of running a profile or gathering results from your source, you can do the following:

1. In Flow View, select the recipe that was used to create the source profile.
2. Rename this recipe to something like, *SourceData*.
3. Create an output off of this recipe. Run the job if you have not yet created the visual profile.
4. Select the recipe again. Now, click **Add New Recipe**.
5. Edit this new recipe and build out your transformation steps.

6. Whenever you need to regenerate the profile for the source, select the `SourceData` recipe and select the output from it. Then, run a job for it.

**Tip:** This technique is useful if you are replacing the source dataset with refreshed data on a periodic basis.

See *Flow View Page*.