Optimize Job Processing

Contents:

This page contains a set of tips for how to improve the overall performance of job execution.

Filter data early

If you know that you are dropping some rows and columns from your dataset, add these transform steps early in your recipe. This reduction simplifies working with the content through the application and, at execution, speeds the processing of the remaining valid data. Since you may be executing your job multiple times before it is finalized, it should also speed your development process.

- To drop columns:
  - Select **Drop** from the column drop-down for individual columns. See *Column Menus*.
  - Use the `drop` transform to remove multiple discrete columns or ranges of columns. See *Drop Transform*.
- To delete rows:
  - Use the `delete` transform with a `row` parameter value to identify the rows to remove. For example, the following removes all rows that lack a value for the `id` column:

    ```
    delete row:ISMISSING(id)
    ```

    You can paste Wrangle steps into the *Transformer Page*.

  - Similarly, you can use the `keep` transform to retain the rows of interest, dropping the rows that do not match. For example, the following transform keeps all rows that lack a value in the `id` column:

    ```
    keep row:NOT(ISMISSING(colA))
    ```

Perform joins early

Join operations should be performed early in your recipe. These steps bring together your data into a single consistent dataset. By doing them early in the process, you reduce the chance of having changes to your join keys impacting the results of your join operations. See *Join Page*.

Perform unions late

Union operations should be performed later in the recipe so that you have a small chance of changes to the union operation, including dataset refreshes, affecting the recipe and the output. See *Union Page*.

Run jobs on the default running environment

When configuring a job, Trifacta analyzes the size of your dataset to determine the best of the available running environments on which to execute the job. This option is presented as the default option in the dialog. Unless you have specific reasons for doing otherwise, you should accept the default suggestion.