

Run Job Page

Contents:

- *Running Environment*
 - *Options*
 - *Publishing Actions*
 - *Run Job*
 - *Automation*
 - *Run jobs via API*
-

In the Run Job page, you can specify transformation and profiling jobs for the currently loaded recipe. Available options include output formats and output destinations.

NOTE: When you run a job in Dataprep by Trifacta, the job is queued and executed on Dataflow. Dataprep by Trifacta observes the job in progress and reports progress as needed back into the application. Dataprep by Trifacta does not control the execution of the job.

Tip: Jobs can be scheduled for periodic execution through Flow View page. For more information, see *Add Schedule Dialog*.

Tip: Columns that have been hidden in the Transformer page still appear in the generated output. Before you run a job, you should verify that all currently hidden columns are ok to include in the output.

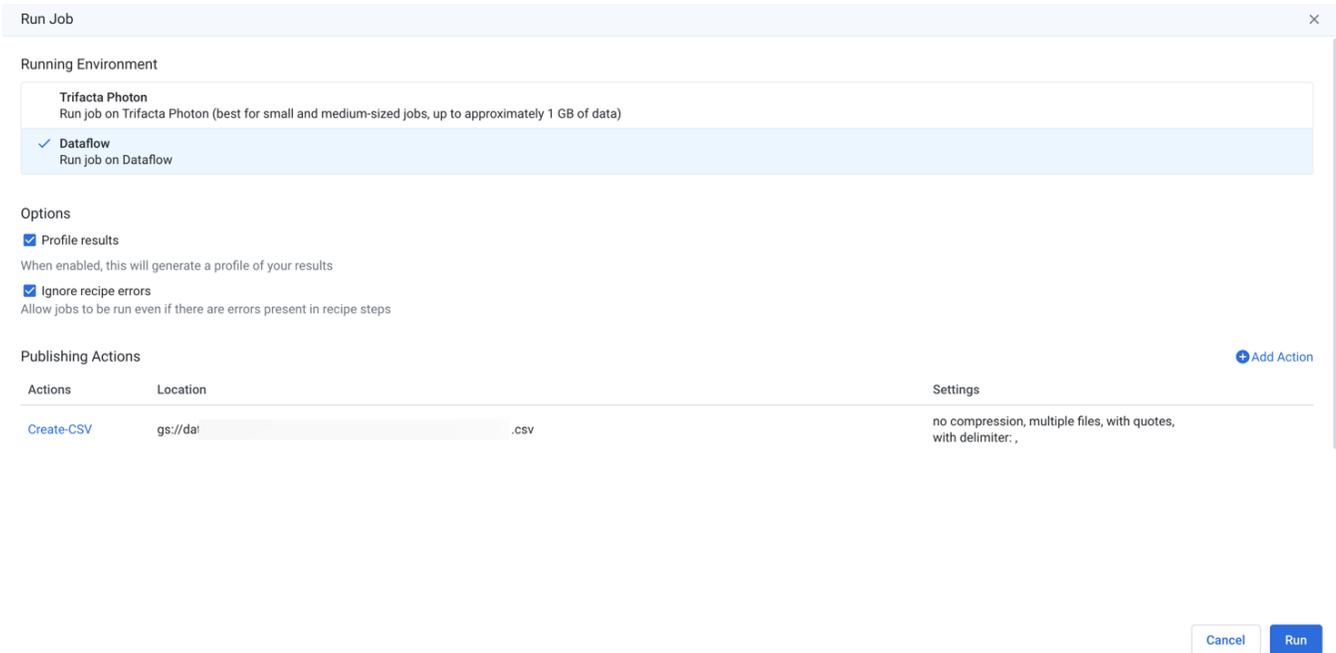


Figure: Run Job Page

Running Environment

Select the environment where you wish to execute the job. Some of the following environments may not be available to you. These options appear only if there are multiple accessible running environments.

NOTE: Running a job executes the transformations on the entire dataset and saves the transformed data to the specified location. Depending on the size of the dataset and available processing resources, this process can take a while.

Photon: Executes the job in Photon, an embedded running environment hosted on the same server as the Dataprep by Trifacta®.

 **Feature Availability:** This feature is not available in Dataprep by Trifacta Legacy only.

NOTE: Jobs that are executed on Trifacta Photon may be limited to run for a maximum of 10 minutes, after which they fail with a timeout error. If your job fails due to this limit, please switch to running the job on Dataflow.

Spark: Executes the job using the Spark running environment.

Dataflow: Executes job on Dataflow within the Google Cloud Platform. This environment is best suited for larger jobs.

Dataflow + BigQuery: For flows whose data is sourced in BigQuery or Cloud Storage, you may be able to choose to run jobs for them in BigQuery. Some limitations may apply.

- BigQuery as a running environment must be enabled in the project by a project owner. See *Configure Running Environments*.
- Individual users must enable pushdowns within their flows. See *Flow Optimization Settings Dialog*.
- For more information on limitations, see *BigQuery Running Environment*.

Options

Profile results and assess data quality rules: Optionally, you can enable this option to generate a visual profile of your job results. If your job contains recipes with data quality rules, those rules are applied to the generated results and displayed in the Job Details page.

NOTE: You must choose to profile your results if you wish to see the data quality rules applied to your results.

When the profiling job finishes, details are available through the Job Details page, including links to download results.

- Disabling profiling of your output can improve the speed of overall job execution.
- See *Job Details Page*.

NOTE: Percentages for valid, missing, or mismatched column values may not add up to 100% due to rounding. This issue applies to the Photon running environment.

Ignore recipe errors: Optionally, you can choose to ignore errors in your recipes and proceed with the job execution.

NOTE: When this option is selected, the job may be completed with warning errors. For notification purposes, these jobs with errors are treated as successful jobs, although you may be notified that the job completed with warnings.

Details are available in the Job Details page. For more information, see *Job Details Page*.

Publishing Actions

You can add, remove, or edit the outputs that are generated from this job. For more information, see *Publishing Actions*.

Run Job

To execute the job as configured, click **Run**. The job is queued for execution.

Dataflow imposes a limit on the size of the job as represented by the JSON passed in.

Tip: If this limit is exceeded, the job may fail with a `job graph too large` error. The workaround is to split the job into smaller jobs, such as splitting the recipe into multiple recipes. This is a known limitation of Dataflow.

After a job has been queued, you can track its progress toward completion. See *Job Details Page*.

Automation

Run jobs via API

You can use the available REST APIs to execute jobs for known datasets. For more information, see *API Reference*.