

# Run Job Page

## Contents:

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

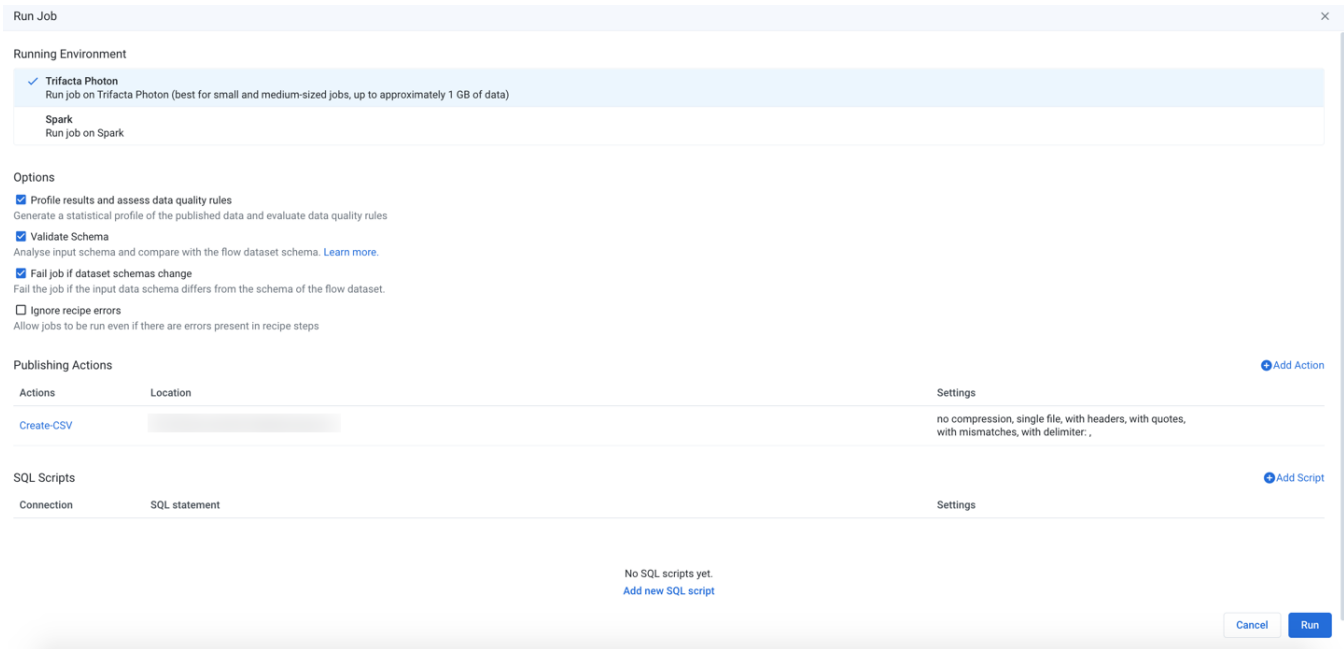
In the Run Job page, you can specify transformation and profiling jobs for the currently loaded recipe. Available options include output formats and output destinations.

You can also configure the environment where the job is to be executed.

**NOTE:** If the job is executed in an environment other than Trifacta Photon, the job is queued for execution in the environment. Jobs executed on a remote cluster may incur additional overhead to spin up execution nodes, which is typically within 10-15 seconds. During job execution, Designer Cloud observes the job in progress and reports progress as needed back into the application. Designer Cloud 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.

**Tip:** The application attempts to identify the best running environment for you. You should choose the default option, which factors in the available environments and the size of your dataset to identify the most efficient processing environment.

**Photon:** Executes the job in Photon, an embedded running environment hosted on the same server as the Designer Cloud.

**Spark:** Executes the job using the Spark running environment.

### Advanced Execution Options:

- If Spark job overrides have been enabled in your environment, you can apply overrides to the specified job. See *Spark Execution Properties Settings*.

**Spark + Snowflake:** Executes job on Snowflake.

- Snowflake 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*.

**NOTE:** Supported file formats are CSV, TXT, and JSON only.

For more information on limitations, see *Snowflake 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.

**Validate Schema:** When enabled, the schemas of the datasources for this job are checked for any changes since the last time that the datasets were loaded. Differences are reported in the Job Details page as a Schema validation stage.

**Tip:** A **schema** defines the column names, data types, and ordering in a dataset.

**Fail job if dataset schemas change:** When Validate Schema is enabled, you can set this flag to automatically fail the job if there are differences between the stored schemas for your datasets and the schemas that are detected when the job is launched.

**NOTE:** If you attempt to refresh the schema of a parameterized dataset based on a set of files, only the schema for the first file is checked for changes. If changes are detected, the other files must contain those changes as well. This can lead to changes being assumed or undetected in later files and potential data corruption in the flow.

**Tip:** This setting prevents data corruption for downstream consumers of your executed jobs.

**Tip:** The default for validate schema is set at the workspace level. In the Run Job page, these settings are overrides for individual jobs.

For more information, see *Overview of Schema Management*.

**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. 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*.