

Configure Running Environments

Contents:

- *Trifacta Photon*
 - *Dataflow*
 - *BigQuery*
-

This section provides overview information on how to configure the running environments accessible from your deployment of the Trifacta application.

A **running environment** is the set of services that are used to execute a job.

- A job can include tasks to do the following:
 - Ingest data
 - Transform data
 - Profile data
 - Sample data
 - Generate results
- A running environment can be hosted on the Trifacta node or across a cluster that is connected to the product.

Trifacta Photon

Hosted on the Trifacta node, Trifacta Photon is an in-memory running environment designed for high performance on small- to medium-sized jobs.

Feature Availability: This feature is available in the following editions:

- Dataprep by Trifacta Enterprise Edition
- Dataprep by Trifacta Professional Edition
- Dataprep by Trifacta Starter Edition
- Dataprep by Trifacta Premium
- Dataprep by Trifacta Standard

NOTE: Trifacta Photon lives in the Trifacta VPC. These jobs are not executed in a customer VPC. Data is streamed to the Trifacta VPC for transformation and is not stored within the VPC.

NOTE: You cannot cancel jobs that have been launched on Trifacta Photon.

Configuration:

Trifacta Photon may require enablement in your project or workspace:

- For more information, see *Dataprep Project Settings Page*.

Dataflow

Dataflow is a fully managed, serverless data processing service that is hosted in the Google Cloud Platform. Managed by Google, this service is enabled by default when you enable Dataprep by Trifacta in any of your Google Cloud Platform projects.

Configuration:

- Access to the Dataflow service is governed through permissions in the IAM roles for users. Access is enabled by default. For more information, see *Required Dataprep User Permissions*.
- Jobs are run on Dataflow using service accounts. The default Compute Engine service account deployed to your project has sufficient permissions to run Dataflow jobs. For more information, see *Google Service Account Management*.

BigQuery

Feature Availability: This feature is available in the following editions:

- Dataprep by Trifacta Enterprise Edition
- Dataprep by Trifacta Professional Edition
- Dataprep by Trifacta Premium
- Dataprep by Trifacta Standard

BigQuery is a cloud-based data warehouse platform that is fully integrated into the Google Cloud Platform. BigQuery supports a standard SQL dialect for querying datasets and tables and enables the writing of results back from the product. For more information, see <https://cloud.google.com/products/bigquery/>.

For datasets and outputs that are hosted in BigQuery, you can configure the Trifacta application to perform the transformation steps of your job in BigQuery. In this manner, no data needs to be transferred to and from the data warehouse, and performance should be significantly better.

Tip: For jobs that are executed in BigQuery, you can optionally enable the execution of the visual profile in BigQuery, too. This option is enabled for individual flows. For more information, see *Flow Optimization Settings Dialog*.

Limitations:

NOTE: BigQuery is not a running environment that you explicitly select or specify as part of a job. If all of the requirements are met, then the job is executed in BigQuery when you select Dataflow. For more information on limitations, see *Overview of Job Execution*.

Configuration:

- A project owner must enable the following features in the project:
 - BigQuery Execution
 - Logical and physical optimization of jobs
 - For more information, see *BigQuery Running Environment*.
- For individual flows, all general and BigQuery optimizations must be enabled. For more information, see *Flow Optimization Settings Dialog*.

