Run Jobs and Automate

Contents:

- *Job Execution Process*
  - *Output objects*
  - *Job types*
- *Run Jobs to Generate Results*
- *Schedule Jobs*
- *Parameterize Your Trifacta Objects*
- *Orchestrate Job Sequences*
  - *Schedule Plans*

When you are ready to apply your recipe across your entire dataset, you run a job. When your recipe is finalized, you can schedule a job for regular execution, so that downstream stakeholders are assured of having fresh data.

Job Execution Process

A **job** is a complex set of tasks to ingest your data from its datasources and deliver your data and recipe to the selected running environment for execution.

A **running environment** is an execution engine designed for transforming large datasets based on a set of scripted steps. A running environment can be:

- Trifacta® Photon is an in-memory running environment, local to the Trifacta node. Trifacta Photon enables faster execution of small- to medium-sized jobs.
- **Remote running environments** provide cloud- or cluster-based running environments for execution of large jobs of any scale. Trifacta supports various remote running environments, depending on your deployment.

Output objects

A job is executed through an **output object**, which is required for every job.

> **Tip:** If an output object does not exist for the job you are trying to run, the Trifacta application creates one for you.

An output object definition includes the following:

- **Running environment:** For best results, select the default.
- **Visual profiling:** Select the visual profiling option to generate a visual profile of the job results. Visual profiling is handled as a separate job executed after the transformation job is complete.
- **Publishing actions**: Define one or more publishing actions to specify:
  - Output datastore, path/database, and output file or table name
  - Output format
  - Update action: create new, append, or replace.
  - Parameterization: create output parameters as needed.
- Other output settings

> **Tip:** A "job" encompasses multiple sub-jobs, which manage the processes of ingestion, conversion, transfer, transformation, profiling, and generating of results as needed to complete the job.

Job types

Jobs can be of the following types:

- **Manual:** Need results? Click **Run** to launch a job right now.
- **Scheduled:** If you need results to be scheduled at a specific time, you can set up a scheduled execution.

> **NOTE:** Both types of jobs require output objects. For any recipe, you can create different output destinations for manual or scheduled jobs.

> **Tip:** Jobs can also be triggered using REST APIs, if you prefer to handle job scheduling outside of the Trifacta application.

Run Jobs to Generate Results

> **NOTE:** Running a job consumes resources. Depending on your environment, resource consumption may cost money. Your project owner or workspace administrator may be able to provide guidance on resources and their costs.

To run a job right now, you can do either of the following:

1. In the Transformer page, click **Run**.
2. In Flow View, click the output to generate. In the right panel, click **Run**.

> **Tip:** By default, a manual job generates a CSV with visual profiling to the default output location using the optimal running environment for the job size. In the Run Jobs page, you can define or update your output object and its publishing actions, as needed.

For more information, see *Generate Results*.

Schedule Jobs

Through Flow View, you can create outputs for your scheduled destinations and define the schedule for when those outputs are generated.

For more information, see *Schedule Jobs*.

Parameterize Your Trifacta Objects

In the Trifacta application, a **parameter** is a storage object that can be defined to capture a variable, a pattern or wildcard, or a set of timestamp values. You can apply parameters to:

- imported datasets
- flows
- output objects
- your project or workspace

For example, if you are having a set of files stored with parallel names in a single directory, you can create a dataset with parameters to capture all of these files into a single dataset. So, instead of having to union all of the files together (and re-union them if new files are added), you can create a single imported dataset object to capture all of them, and if new files added to the directory follow the same pattern, the **dataset with parameters** gets automatically updated.

For more information, see *Parameters*.

Orchestrate Job Sequences

> ⚠ **Feature Availability:** This feature may not be available in all product editions.

You can use plans to orchestrate sequences of job executions. A **plan** is a sequence of tasks executed in the Trif acta application. In addition to flow tasks, which execute specific outputs, you can create HTTP tasks to message external systems or, if needed, to execute REST API endpoints within the Trifacta platform.

For more information, see *Plans*.

Schedule Plans

Plans can be scheduled, too.

See *Plans*.