

Profiling Options

Contents:

- [Overview](#)
- [Configure](#)
 - [Run Profiling as a Second Job in Spark](#)
- [Disable Profiling Option](#)

A **visual profile** can be optionally generated as part of execution of a job. While profiling can extend the time it takes to complete the job, a visual profile can provide key statistical information on the columns of your data and the overall dataset itself. This section provide information on each of the available profilers and how to configure them.

NOTE: To enable use of a profiler, the running environment on which it is hosted must also be enabled. See [Running Environment Options](#).

The Trifacta® platform supports the following options for profiling your data.

Overview

Type of Profiler	Requires Hadoop cluster?	Supported Running Environment	Description
Scala Spark Profiler	Yes	all	Generates a profile on the Trifacta Photon running environment using the Spark Job Service. Does not require any Spark installation on the cluster. <div style="border: 1px solid green; padding: 5px; text-align: center;">Tip: This is the best-performing profiler in most use cases.</div> See Configure for Spark .
Trifacta Photon Profiler	No	Trifacta Photon	Default profiler when Trifacta Photon running environment is enabled, and the Scala Spark Profiler has not been enabled.

Configure

Making changes to your profiling type:

1. Apply the configuration changes listed below.
2. Save your changes and restart the platform.
3. Restart your browser and login again.

Run Profiling as a Second Job in Spark

By default, profiling jobs execute on the running environment where the transformation job was executed. You can configure the Trifacta Photon running environment to execute profiling jobs as a separate job in the Spark running environment. This separation allows users to begin working with the transformed data while waiting for the profiling job to complete.

Steps:

1. You can apply this change through the *Admin Settings Page* (recommended) or `trifacta-conf.json`. For more information, see *Platform Configuration Methods*.
2. Locate the following property and set its value to `true`:

```
"photon.runProfileWithSpark": false,
```

3. Save your changes and restart the platform.

When profiling is selected for jobs executing on the Trifacta Photon running environment, their jobs are executed on the Spark running environment as a separate job.

Disable Profiling Option

Profiling is invoked at job execution time by the user. See *Run Job Page*.

To disable user choice through the UI, set the following parameter:

```
"profiler.userOption": false,
```

When the above setting is disabled:

- Any available checkbox no longer works. User cannot choose whether to profile or not.
- Profiles are always executed.

NOTE: This setting does not affect profiling through the APIs. Profiling can always be enabled or disabled for jobs that are executed via API.

To disable user choice through the UI, set the following parameter:

```
"profiler.userOption": false,
```

NOTE: Profiling can still be enabled for jobs that are executed via API.