Running Environment Options

Contents:

The Trifacta® platform can be configured to integrate with a variety of environments for processing transformation jobs. When you run a job through the application, you have the option of selecting the running environment on which you wish to run the job.

> **Tip:** In general, you should accept the default environment that is presented for job execution. The application attempts to match the scope of your job to the most appropriate running environment.

This section applies to execution of transform jobs. For more information on options for profiling jobs, see *Profiling Options*.

Available Running Environments

For more information, see *Overview of Job Execution*.

Configure

Configure Running Environments

To apply this configuration change, login as an administrator to the Trifacta node. Then, edit `trifacta-conf.json`. Some of these settings may not be available through the *Admin Settings Page*. For more information, see *Platform Configuration Methods*.

The following parameters define the available running environments:

```
"webapp.runWithSparkSubmit": true,
"webapp.runinEMR": false,
"webapp,runInDatabricks' : true,
"webapp.runInDataflow": false,
```

For more information on configuring the running environment for EMR, see *Configure for EMR*.

Below, you can see the configuration settings required to enable each running environment.

- The Spark running environment requires a Hadoop cluster as the backend job execution environment.
    - In the Run Job page, select **Spark**.
- The Trifacta Photon running environment executes on the Trifacta node and provide processing to the front-end client and at execution time.
    - In the Run Job page, select **Photon**.
    - For more information on disabling the Trifacta Photon running environment, see *Configure Photon Running Environment*.

Type	Running Environment	Configuration Parameters	Notes

| Hadoop Backend | Spark | `webapp.`
`runWithSparkSubmit`
`= true` | The Spark running environment is the default configuration. |
| Client Front-end and non-Hadoop Backend | Trifacta Photon | In the Workspace Settings page, set `Photon Execution` to `Enabled`. | Trifacta Photon is the default running environment for the front-end of the application. It is enabled by default. For more information, see *Workspace Settings Page*. |

> **NOTE:** Do not modify the `runInDataflow` setting.

Configure Default Running Environment

When you specify a job, the default running environment is pre-configured for you, based on the following parameter:

> **NOTE:** If your environment has no running environment such as Spark for running large-scale jobs, this parameter is not used. All jobs are run on the Trifacta node.

```
"webapp.client.maxExecutionBytes.photon": 1000000000,
```

The default environment presented to you is based on the size of the primary datasource. For the above setting of 1 GB:

Running Environment	Default Condition
Trifacta Photon	Size of primary datasource is less than or equal to the above value in bytes.
Spark	Size of primary datasource is greater than the above value in bytes.

> **NOTE:** This setting defines only the environment that is recommended to you as a predefined selection. If a second running environment is available, you can choose to select it, although it is not recommended to choose an environment other than the default. See *Run Job Page*.

> **Setting this value too high forces more jobs onto the Trifacta Photon running environment, which may cause slow performance and can potentially overwhelm the server.**

> **Tip:** To force the default setting to always be a Hadoop or bulk running environment, set this value to `0`. All users are recommended to use the bulk option instead of the Trifacta Photon running environment. However, smaller jobs may take longer than expected to execute.