

Spark Execution Properties Settings

When you specify a job, you may pass to the Spark running environment a set of Spark property values to apply to the execution of the job. These property values override the global Spark settings for your deployment.

NOTE: A workspace administrator must enable Spark job overrides and configure the set of available parameters. For more information, see *Enable Spark Job Overrides*.

Spark overrides are applied to individual output objects.

- You can specify overrides for ad-hoc jobs through the Run Job page.
- You can specify overrides when you configure a scheduled job execution.

User-specific Spark overrides: If you have enabled user-specific overrides for Spark jobs, those settings take precedence over the settings that are applied through this feature. For more information, see *Configure User-Specific Props for Cluster Jobs*.

In the appropriate page, click the Advanced Execution Settings caret.

Run Job

Running Environment

- Photon (Trifacta Server)
Run job on Photon (Trifacta Server) engine
- Spark
Run job on Spark engine

Advanced environment options ^

Spark Execution Properties ?

spark.driver.memory

spark.executor.memory

spark.executor.cores

transformer.dataframe.checkpoint.threshold

Figure: Spark Execution Properties

Default Spark overrides:

The first four properties are available for all Spark job overrides:

Before you modify these parameters, you should review with your cluster administrator what are appropriate settings for each parameter. In some cases, you can set these values to cause failures on the cluster. No validation is performed for inputted values.

Spark parameter	Description
spark.driver.memory	Amount of RAM in GB on each Spark node that is made available for the Spark drivers. By raising this number: <ul style="list-style-type: none">• The drivers for your job are allocated more memory on each Spark node.• There is less memory available for other uses on the node.
spark.executor.memory	Amount of RAM in GB on each Spark node that is made available for the Spark executors. By raising this number: <ul style="list-style-type: none">• The Spark executors for your job are allocated more memory.• There is less memory available for other uses on the node.
spark.executor.cores	Number of cores on each Spark executor that is made available to Spark. By raising this number: <ul style="list-style-type: none">• The maximum number of cores available for your job is raised on each Spark executor.• There are fewer cores for other uses on the node.
transformer.dataframe.checkpoint.threshold	When checkpointing is enabled, the Spark DAG is checkpointed when the approximate number of expressions in this parameter has been added to the DAG. Checkpointing assists in managing the volume of work that is processed through Spark at one time; by checkpointing after a set of steps, the Trifacta platform can reduce the chances of execution errors for your jobs. By raising this number: <ul style="list-style-type: none">• You increase the upper limit of steps between checkpoints.• You may reduce processing time.• It may result in a higher number of job failures.

For more details on setting these parameters, see *Tune Cluster Performance*.

Other Spark overrides:

Your workspace administrator may have enabled other Spark properties to be overridden. These parameters appear at the bottom of the list.

Please check with your administrator for appropriate settings for these properties.