

Configure Spark Running Environment

Contents:

- *Known Limitations and Issues*
 - *Enable Spark Execution Environment*
 - *Use Spark Execution Environment*
 - *Change Limits*
-

This section provides information on how to enable and configure the Spark running environment, which leverages Spark's faster in-memory processing to deliver better execution performance.

i NOTE: Spark is the default running environment for Hadoop job execution in Release 4.0 and later. Unless you are upgrading from a pre-Release 4.0 environment, no additional configuration is required.

Known Limitations and Issues

i NOTE: When a recipe containing a user-defined function is applied to text data, any non-printing (control) characters cause records to be truncated by the running environment during Hadoop job execution. In these cases, please execute the job on the Trifacta® Server.

- You cannot publish through Cloudera Navigator for Spark jobs.

Enable Spark Execution Environment

The Spark execution environment is enabled by default.

i NOTE: If you have not done so already, please enable and configure the Spark Job Service. See *Configure for Spark*.

i NOTE: If you have upgraded from a pre-Release 4.0 system, your running environment may default to the one defined in your previous release. For more information on enabling, see *Running Environment Options*.

Use Spark Execution Environment

When Spark execution is enabled, it is available like any other execution environment in the application. When executing a job, select the **Run on Hadoop** option from the drop-down in the Run Job page. See *Run Job Page*.

Change Limits

For more information on changing limits and other tuning parameters, see *Configure for Spark*.