{JIRA Case: <http://jira.pentaho.com/browse/DOC-3146>}

{MindTouch Articles: <https://help.pentaho.com/Documentation/7.0/0H0/Set_Up_Pentaho_to_Connect_to_a_Hortonworks_Cluster#Version-Specific_Notes> and <https://help.pentaho.com/Documentation/7.1/0H0/Set_Up_Pentaho_to_Connect_to_a_Hortonworks_Cluster#Version-Specific_Notes>}

# Version-Specific Notes

The following notes are special topics for HDP.

## HDP 2.5 Notes

The following note address issues with HDP 2.5.

### Sqoop Support

If you receive an error message stating *Generating splits for a textual index column allowed only in case of "-Dorg.apache.sqoop.splitter.allow\_text\_splitter=true" property passed as a parameter* while trying to use the **split-by** option to the **[Sqoop Import](http://wiki.pentaho.com/display/EAI/Sqoop+Import)** job entry with the HDP 2.5 shim, perform the following steps to set the org.apache.sqoop.splitter.allow\_text\_splitter property to true:

1. Open your KJB file that contains a **Sqoop Import** entry in the PDI client (also known as Spoon).
2. Double-click the **Sqoop Import** entry to access the **Sqoop Import** property dialog box.
3. Click the **Advanced Options** link in the lower left corner of the dialog box.
4. In the **Custom** tab, add the Dorg.apache.sqoop.splitter.allow\_text\_splitter argument and set value to true.
5. Click OK and save your KJB file.

You should now be able to use the split-by option to the **Sqoop Import** entry.

## HDP 2.4 Notes

The following note address issues with HDP 2.4.

### Simba Spark SQL Driver Support

**If you are using Pentaho 7.0 or later,** the HDP 2.4 shim supports the Simba Spark SQL driver. You will need to download, install, and configure the driver to use Simba Spark SQL with the HDP 2.4 shim.

1. Download the [Simba Spark SQL driver](http://www.simba.com/drivers/spark-jdbc-odbc/).
2. Extract the ZIP file, and then copy the following 3 files into the lib/ directory of the HDP shim:

* SparkJDBC41.jar
* TCLIServiceClient.jar
* QI.jar

1. In the **Database Connection** window, select **SparkSQL** option. The default port for the Spark thrift server is 10015.
2. For secure connections, set the following additional parameters on the JDBC URL through the **Options** tab:

* KrbServiceName
* KrbHostFQDN
* KrbRealm

1. For unsecure connections, if your Spark SQL configuration specifies hive.server2.authentication=NONE, then make sure to include an appropriate **User Name** in the **Database Connection** window.  Otherwise, the connection is assumed to be NOSASL authentication, which will cause a connection failure after timeout.
2. Stop and restart the component.

## HDP 2.3 Notes

The following note addresses issues with HDP 2.3.

### Pentaho can connect to HDP 2.3 cluster using the HDP 2.2 or HDP 2.3 shims

You can use either the HDP 2.2 or HDP 2.3 shims to connect to a HDP 2.3 clusters:

* If you use the HDP 2.2 shim to connect to an HDP 2.3 cluster, only HDP 2.2 functionality is supported.
* If you want to support HDP 2.3 functionality, use the HDP 2.3 shim to connect to the HDP 2.3 cluster instead.

Shims can be downloaded from the Pentaho Support Portal.