Installation and Execution Problems with Pentaho Data Integration (PDI)

# Database Connection Problems

If you have an error when testing a database connection, ensure that you have the connection properties indicates as shown in the Pentaho exercise document. You should copy the text value for the Database Name item from the Pentaho exercise document, as the value is rather long. Ensure that you have no value (blank) for the Port Number and Host Name items.

# Windows Problems

If encountering problems, use the debug bat file (SpoonDebug.bat) and save errors to SpoonDebug.txt.

The java location is a common source of problems. If the java.exe file is not found, you can change the environment variable (PENTAHO\_JAVA\_HOME). If this environment variable does not exist, you should create it as a new system environment variable. You should set PENTAHO\_JAVA\_HOME to the java path. You can find the java path by clicking the View button in the Java tab of the Java Control Panel.

Another common problem involves the amount of memory to run Java, specified as a java option. In PDI 9.0, two memory options are used (-Xms and –Xmx). The default in PDI 1.8 is “–Xms1024m” and “–Xmx2048m”. If you see an error in SpoonDebug.txt about exceeding memory allocation, you can change the default values. I reduced the values by half (“–Xms512m” and “–Xmx1024m”) and PDI loaded. To change the java run time options, you need to edit spoon.bat. I suggest that you copy the file and change its extension to txt before editing in WordPad. In the bat file, search for “set PENTAHO\_DI\_JAVA\_OPTIONS”.

A third source of installation issues is the jdbc file providing the Oracle driver. PDI looks in C:\Program Files\Pentaho\Data Integration\Lib for this file. If you get an error when testing a database connection indicating that the Oracle driver cannot be found, you should make sure that Pentaho resides in the Program Files folder. I have installed Pentaho to different folders and not experienced the problem locating the Oracle driver, but some students have reported problems when Pentaho is installed to a different folder. I recommend that you use the file ojdbc8.jar file. I have experienced problems using other versions of the jdbc driver file (ojdbc7.jar and ojdbc10.jar). The Data Integration\Lib folder should only have only one jdbc driver file. After copying the jdbc file, you should close and restart PDI.

# Mac Problems

PDI 5.0.1 does not seem to work on a Mac with a newer operating system when it comes to database connections. There was some discussion on online forums of this.

PDI 7.1 seems to work with OS 10.14.1 and Java 8. The database connection will not work with newer versions of Java. If you remove a newer version of Java from your Mac make sure to remove any of the associated files from the Mac library on your computer. Pentaho was still using Java 10 files even after downgraded to 8 until removing the old files in the library.

If you get a JDBC error (Driver class ‘oracle.jdbc.driver.OracleDriver’ could not be found) when testing a database connection, try the ojdbc8.jar instead of ojdbc7.jar. You should copy the jar file to the data-integration/lib folder.

The launch time with 7.1 is substantial unless you follow this procedure to remove some plug ins:

<https://blog.twineworks.com/improving-startup-time-of-pentaho-data-integration-78d0803c559b>

The spoon.sh file executes but no activity occurs. The issue apparently had to do with the Data Integration app being "quarantined" by Mac OS X Sierra's security system. This command fixed the problem.

sudo xattr -dr com.apple.quarantine /path/to/Data\ Integration.app

PDI file may not unzip on Mac OS unless security settings are changed. This page indicates method to modify security settings.

https://truefire.zendesk.com/hc/en-us/articles/200200566-Open-Program-from-Unidentified-Developer-Mac-