Installation of Pentaho Data Integration and Database Driver for Module 5

This document provides details about installing Pentaho Data Integration and a database driver. You will need the database driver to connect to Oracle, MySQL, or PostgreSQL to complete the guided tutorial and assignment in module 5.

# Installing Pentaho Data Integration

You should install the community edition of Pentaho. The latest stable version (5.0.1) is available from the Source Forge website (<http://sourceforge.net/projects/pentaho>) although I recommend that you install the current version (9.0) in March 2020 or a later version.

You should install the community edition of Pentaho. To install Pentaho, you should follow the steps below. It is highly recommended that you use the community edition from SourceForge as the instructions in this document follow the community edition interface.

* The latest stable version is 5.0.1 although the latest version on SourceForge is 9.0.0. You should be able to use either version to complete the tutorial and assignment although the guided tutorial is written for the latest version 9.0.0. I recommend that you install the current version (9.0) in April 2020 or a later version. Use the Download button in the SourceForge website (<http://sourceforge.net/projects/pentaho>) to download the zip file for the current community edition version of Pentaho Data Integration.
* To use a different version (such as the latest stable version), go to <http://sourceforge.net/projects/pentaho/> and search in the Files link. To find a different version, look in the Client-Tools folder of the version of interest. You should select the link with PDI in the name inn the Client-Tools folder such as pdi-ce-8.2.0.0-371.zip.Go to <http://sourceforge.net/projects/pentaho/> and click on **Files** - **Data Integration** – **5.0.1-stable**
* Unzip the downloaded zip file to any folder.
* Windows users should copy the folder data-integration to the folder C:\Program Files\Pentaho. Mac and Linux users (including Oracle Virtual Box) may move the file to any folder. Note that the Oracle Virtual Box is a Linux environment.
* If you are using the Oracle Database Virtual Box Appliance, you should download and unzip the PDI zip file inside the Virtual Box. If you download the PDI zip file in Windows, you may have difficulty making a connection to an Oracle database in the module 5 assignment.

To ensure that the installation worked, you should launch Pentaho Data Integration.

* Run the file Spoon.bat by double clicking on it. You may want to create a shortcut to the spoon.bat file so starting data integration is easier. If you get a permission error or cannot execute the bat file, you should right click and select “Run as Administrator”. For Mac and Linux users, run the Spoon.sh from terminal (./spoon.sh). Note that the Oracle Virtual Box is a Linux environment.
* After you launch Pentaho Data Integration, you will see the Welcome window (Figure 1) and then the Spoon designer (Figure 2).
* Exit Spoon before installing the database driver file in the next part of the instructions.



Figure 1: Pentaho Data Integration Welcome Window

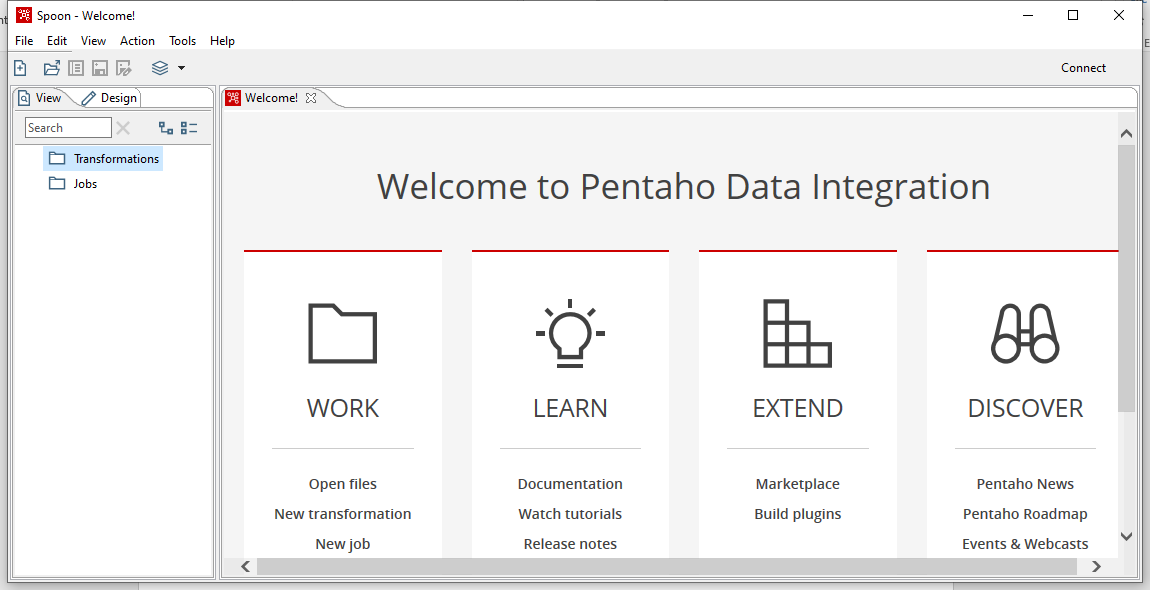


Figure 2: Spoon Opening Window

# Installing JDBC Drivers

In the guided tutorial and assignment in module 5, you will need to connect to a Oracle, MySQL, or PostgreSQL database server. Before you can connect to a database, you must install the appropriate database driver. Pentaho uses the Java Database Connectivity (JDBC) to connect to databases.

You need to install the JDBC driver for the specific version of the DBMS that you previously installed.

* For PostgreSQL, you do not need to install a database driver as the standard installation of Pentaho Data Integration includes the PostgreSQL database driver.
* The relevant JDBC driver for Oracle 19c is “ojdbc8.jar”. I am not sure that other drivers (ojdbc7.jar and ojdbc10.jar) work for Oracle 19c so I recommend not using them. Here is the current download page for the Oracle jdbc driver.

<https://www.oracle.com/database/technologies/appdev/jdbc-downloads.html>

* For MySQL, install the latest JDBC driver using this page.

<https://dev.mysql.com/downloads/connector/j/5.1.html>

The JDBC for MySQL comes in two formats: .zip and .msi. Each format has a different extraction method. Mac and Linux users (including Oracle Virtual Box) cannot use the .msi format.

For the .zip file, the unzipped folder has many files. You only need to copy the .jar file. The current jar file (May 2020) is mysql-connector-java-5.1.49.jar.

For the .msi file, you should double click to extract it to the folder C:\Program Files(x86)\MySQL\MySQL Connector J. In this folder, you will find many files and the only needed one is the .jar file. The current jar file (May 2020) is mysql-connector-java-5.1.49.jar.

For Windows users, after downloading the JDBC file, you should copy the file to the following folder; C:\Program Files\Pentaho\data-integration\lib. If you installed Pentaho to a different folder other than C:\Program Files, you should copy the .jar file to data-integration\lib folder inside the Pentaho folder.

For Mac and Linux users (including Oracle Virtual Box), you should copy the driver to the data-integration/lib directory.

After copying the JDBC file to the specified folder, you should restart Pentaho Data Integration.