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 for either Oracle or MySQL 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).

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 5.4. You should be able to use either version to complete the tutorial and assignment although the guided tutorial is written for the latest stable version 5.0.1.
* Go to <http://sourceforge.net/projects/pentaho/> and click on **Files** - **Data Integration** – **5.0.1-stable**
* Download the **pdi-ce-5.0.1-stable.zip**
* Unzip the downloaded zip file to any folder.
* Windows uses 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.

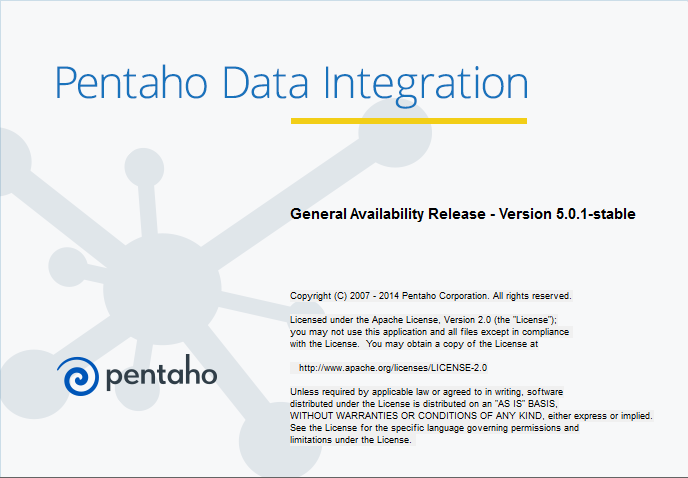


Figure 1: Pentaho Data Integration Welcome Window

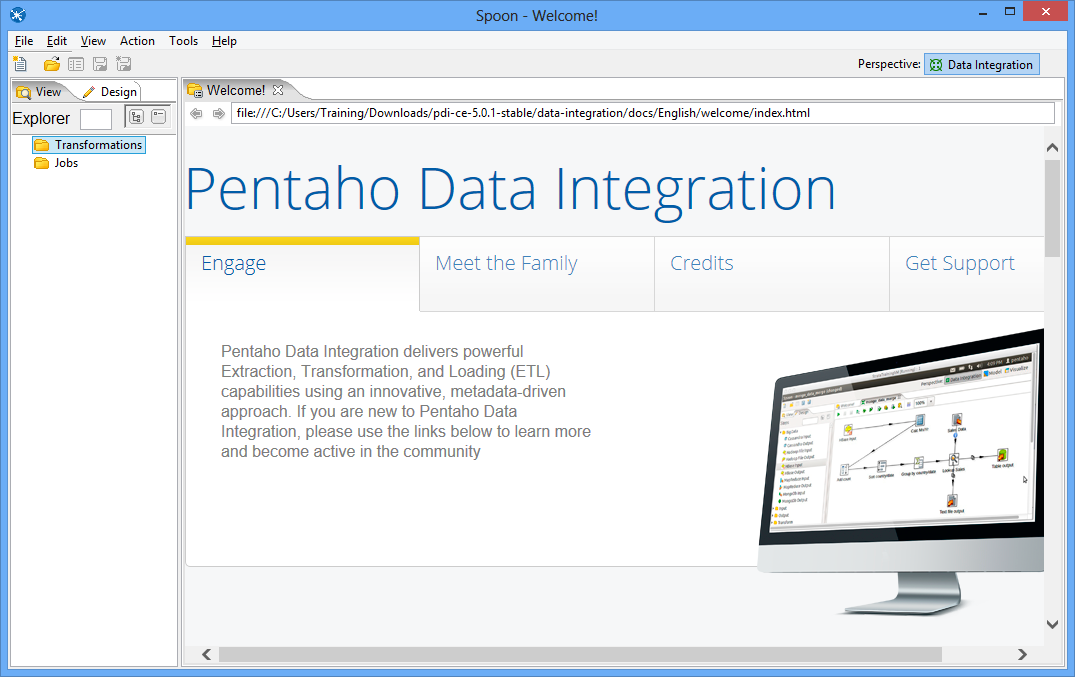


Figure 2: Spoon Opening Window

# Installing JDBC Drivers

In the guided tutorial and assignment in module 5, you will need to connect to either Oracle or MySQL 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 Oracle 12C, the JDBC driver is “ojdbc7.jar”. You can download it from the following page. You can also find JDBC files for other Oracle server versions on this page.

<http://www.oracle.com/technetwork/database/features/jdbc/index-091264.html>

* For MySQL, the JDBC driver is “mysql-connector-java-5.1.36-bin.jar”. You can download it using this page:

<http://dev.mysql.com/downloads/connector/j>

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 mysql-connector-java-5.1.36-bin.jar file.

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 mysql-connector-java-5.1.36-bin.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

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.