Designing and Setting Up a Data Warehouse – Lab Setup

Assignment 2 focuses on designing a data warehouse and setting it up in practice. Therefore, this assignment is a “hands-on” exercise, in which you will create a “classic” data warehouse for an organization, develop ETL processes to load data from the OLTP environment, set up a ROLAP cube, and run a series of queries on it that revolve around important business questions.

This document describes the Lab setup.

Table of Contents

1. Install Java Dependencies, if required

To run Pentaho Data Integration (PDI) and Pentaho Schema Workbench (PSW), Java Runtime Environment (JRE) and Java Development Kit (JDK) are required.

For example, in Windows 10 to check if you already have these installed, go to this path in your file explorer: **"C:\Program Files\Java" or "C:\Program Files (x86)\Java".**

If this folder exists and inside you see folders **"jdk-17" and "jre1.8.0_391"** (or similar) then you have the required files. If this folder doesn't exist or you don't see one or both folders, then you need to download JRE and/or JDK.

1.1. Install Java Runtime Environment (JRE)

Without the bare minimum requirement of a JRE, we cannot run any Java program. We recommend installing 64-bit version of Java Runtime Environment.

- https://www.java.com/en/download/manual.jsp
  For example, for Windows 10 the filename would be: jre-8u391-windows-x64.exe

We will be installing JDK at the very next step, but since Java 11 the official installer provided by Oracle does not any longer include JRE bundled with JDK.

1.2. Install Java Development Kit (JDK)

To run Pentaho Data Integration, Java Development Kit is also required.

- https://www.oracle.com/java/technologies/downloads/
- For Example, for Windows 10 the filename would be: jdk-17_windows-x64_bin.exe

Installation of Java Development Kit is a simple process. It is safe to follow default configuration.

1.3. Setting JAVA_HOME system variable

Windows

- Check if JAVA_HOME is already set,
  - Open Command Prompt and execute : SET JAVA_HOME
  - If output is JAVA_HOME=path, then your JAVA_HOME is set, make sure the path is correct
  - If output is "Environment Variable JAVA_HOME not defined", then execute following steps
- In the Windows Explorer right-click on the **This PC** icon, then select **Properties**
- Open the **Advanced** Tab and click the **Environment Variables** button
- Under System Variable, click **New**
- Enter the variable name as **JAVA_HOME**
- Enter the variable value as the install path for the Java (e.g.: C:\Program Files\Java\jre-1.8)
- You can also change the value later on by using Edit button.
- Click OK
- Open a New Command Prompt
- Execute: "SET JAVA_HOME"
- Output should be like JAVA_HOME=path

Linux

Install for Single User:

- Check if JAVA_HOME is already set:
  - Open Console and execute: echo $JAVA_HOME
  - If output is a path, then your JAVA_HOME is set, make sure the path is correct
  - If output is empty, then execute following steps
- Execute: vi ~/.bashrc OR vi ~/.bash_profile
- add line : export JAVA_HOME=/usr/java/jre1.8.0_391
- save the file
- source ~/.bashrc OR source ~/.bash_profile
- Execute : echo $JAVA_HOME
- Output should print the path

Install for all users:

- Login as root or execute commands with sudo
- Execute: bash
- Execute: vi /etc/bashrc OR vi /etc/profile
- do the same steps as done for single user

2. Install Pentaho Data Integration (PDI)

We install version 9.4.0.0-343, filename: pdi-ce-9.4.0.0-343.zip.

Please open

- https://www.hitachivantara.com/en-us/products/pentaho-platform/data-integration-analytics/pentaho-community-edition.html

Click the "Download Now", acknowledge the open source licenses, proceed with "Proceed to Download", and choose to download the "pdi-ce-9.4.0.0-343.zip" file.

Or use direct URL for the pdi-ce-9.4.0.0-343.zip file.

Unzip "data-integration" folder to a chosen location (e.g., C:\Pentaho\)

3. Install Pentaho Schema Workbench (PSW)

We install version 9.4.0.0-343, filename: psw-ce-9.4.0.0-343.zip

Please open

- https://www.hitachivantara.com/en-us/products/pentaho-platform/data-integration-analytics/pentaho-community-edition.html

Click the "Download Now", acknowledge the open source licenses, proceed with "Proceed to Download", and choose to download the "psw-ce-9.4.0.0-343.zip " file.

Or use direct URL for the psw-ce-9.4.0.0-343.zip file.

Unzip " schema-workbench" folder to a chosen location (e.g., C:\Pentaho\)

4. Install MySQL

MySQL Installer, Version: 8.0.35 (or later)

- https://dev.mysql.com/downloads/mysql/
- https://dev.mysql.com/downloads/windows/installer/8.0.html

We will be using MySQL Workbench to connect to MySQL server and run the queries. The MySQL Installer application can install, upgrade, and manage most MySQL products, including MySQL Workbench. Managing all of your MySQL products, including Workbench, with MySQL Installer is the recommended approach. It handles all requirements and prerequisites, configurations, and upgrades.

When executing MySQL Installer, the installation of MySQL Workbench is selected by default.

Note: When installing MySQL server on a Windows (!) machine make sure to tick "Show advanced and logging options". And then on the "Advanced options" tab choose "Preserve given case" (see screenshots). It is important to get this step right at the time of installation (!).

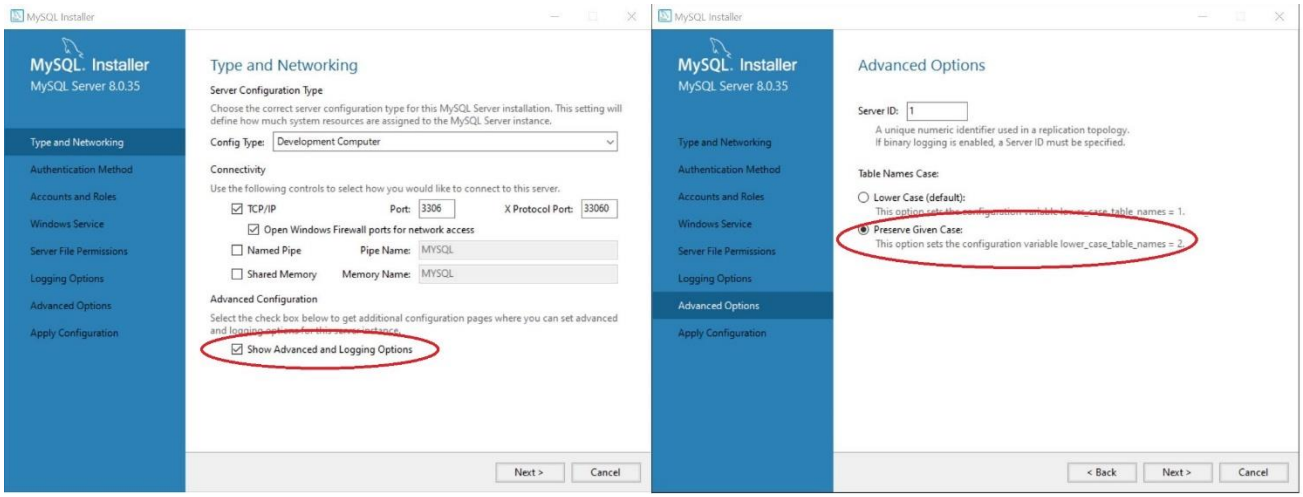


Figure 1. Preserve_given_case setting on a Windows machine

5. Install MySQL JDBC driver (MySQL Connector/J)

MySQL provides connectivity for client applications, developed in the Java programming language, using "MySQL Connector/J", a driver that implements the Java Database Connectivity API.

Download MySQL Connector/J, version: 5.1.49 (platform independent, ZIP archive):

- https://downloads.mysql.com/archives/c-j/

or direct URL for the ZIP archive:

- https://downloads.mysql.com/archives/get/p/3/file/mysql-connector-java-5.1.49.zip

Please note that although any later version of this driver will most certainly work with PDI, it might NOT work with PSW.

The "installation" is straightforward - we just copy *.jar file from the downloaded zip file to the "lib" folder of the corresponding tool (PDI or PSW).

6. Configuration

6.1. Create MySQL user and set up necessary permissions

Launch MySQL Workbench and connect to the default instance with your root credentials using the password you've set up during the installation.

Open a new query window and paste the following code.

This code snippet creates a user "**bi2023**" with the password "**bi2023W!**"

Please use only this user's credentials to implement all database connections in PDI and PSW.

```
-- <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
-- Step 1. Create MySQL user

-- delete user if already exists
DROP USER IF EXISTS bi2023@localhost;

-- create a user to implement all the tasks of the assignment
CREATE USER bi2023@localhost IDENTIFIED BY 'bi2023W!';

-- we must flush the privileges each time we add or edit a user
FLUSH PRIVILEGES;
GRANT ALL PRIVILEGES ON *.* TO bi2023@localhost;
FLUSH PRIVILEGES;
 -- >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
```

6.2. Create databases and grant privileges

Launch MySQL Workbench and connect to the default instance as the newly created "bi2023" user using the password you already know from the previous step.

Open a new query window and paste the following code. Please make sure you replace <**groupNo**> **with Your Group Number (with the leading zero)**.

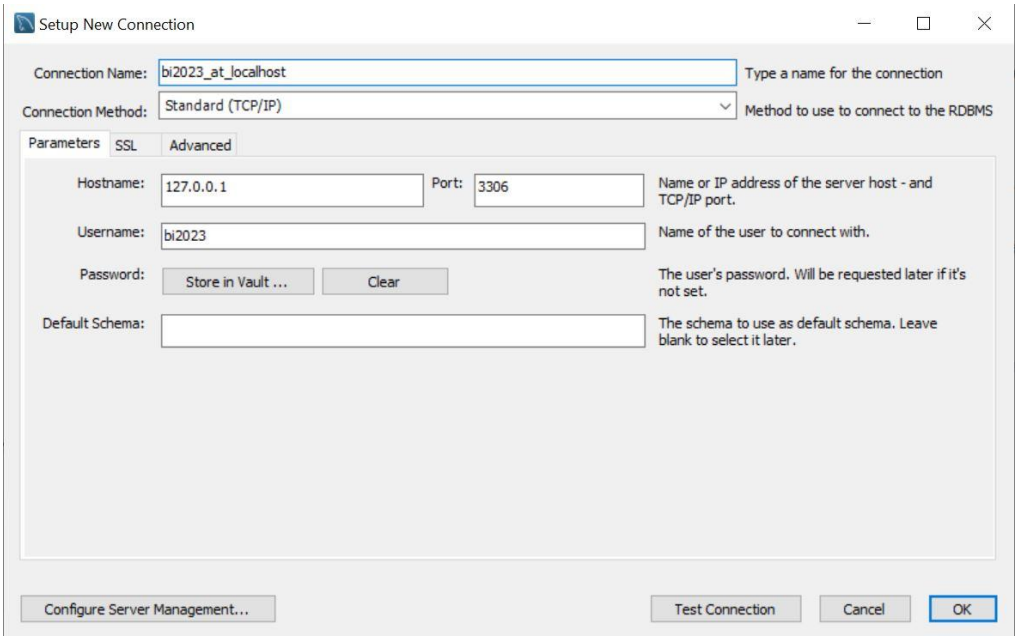
```
-- <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
-- Step 2. Drop if exists and re-create databases
-- please make sure you replace <groupNo> with YOUR GROUP NUMBER

DROP DATABASE IF EXISTS BI_Bikes_<groupNo>;
DROP DATABASE IF EXISTS BI_BikesDW_<groupNo>;
CREATE DATABASE IF NOT EXISTS BI_Bikes_<groupNo>;
CREATE DATABASE IF NOT EXISTS BI_BikesDW_<groupNo>;
-- >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
```

7. Creating a new MySQL Connection

Launch MySQL Workbench to open the home screen. Existing connections are shown in the **MySQL Connections** section.

Click the **[+]** icon near the **MySQL Connections** label to open the **Setup New Connection** wizard. Define the **Connection Name** value, for example, **bi2023_at_localhost**. You can also save your password in Vault by clicking on **Store in Vault** button, so that you don't have to enter password on each connection.



8. Source data preparation

Download the project folder, which already contains csv data dump of the transactional database, from TUWEL (BI_Projects.zip).