

# Spark installation

**Install Anaconda from official site :**

- Go to: <https://www.anaconda.com/products/distribution>
- Install according to your OS and window.

**Install Java in Oracle website:**

- Java 11, Java 17** can download we can work on this both version.
- Install on your window and run it.

**Install Apache spark:**

- Download Spark 3.5.x pre-built with Hadoop.
- Go to: <https://spark.apache.org/downloads.html>
- It will download as zip file and extract the zip file to c drive.
- Rename the folder name as **spark** .

**Install winutils :**

- Download winutils from **github 3.3.6 version**.
- Go to : <https://github.com/cdarlint/winutils>
- Create a folder **Hadoop** in c drive inside the folder create another folder of **bin** and paste the winutils inside the bin folder .
- Paste the winutils inside **spark ->bin->paste it** .

**Install MySql workbench :**

- Download from: [dev.mysql.com](https://dev.mysql.com)
- Run the workbench with root

-Refer the steps in youtube :

[https://youtu.be/hiS\\_mWZmml0?si=aD2jFIHvsHhwPI5F](https://youtu.be/hiS_mWZmml0?si=aD2jFIHvsHhwPI5F)

Because of missing some steps the we cannot work in MySql.

Install MySql connector :

- Use MySQL Installer from dev.mysql.com

-Downloaded as zip file and extract the file.

-Inside it **jar file** will be there ,copy the jar file and and paste it in **jar folder** inside the **spark folder**.

Path setting :

-Go to environmental variable and in system variable click path->edit->new.

-paste path of **spark ,java ,Mysql workbench and hadoop** with bin in it.

-paste these all in path

%JAVA\_HOME%\bin

%SPARK\_HOME%\bin

%SPARK\_HOME%\sbin

%HADOOP\_HOME%\bin

-In system variable create new variables with path

JAVA\_HOME with path without bin

SPARK\_HOME with path without bin

HADOOP\_HOME with path without bin

Checking :

-Open your cmd and command this to verify whether your path are set and installed properly.

-For Java:

Cmd : **java -version**

Output:

Coorect : java version "17.0.2" 2022-01-18 LTS

Java(TM) SE Runtime Environment (build 17.0.2+8-LTS-86)

Java HotSpot(TM) 64-Bit Server VM (build 17.0.2+8-LTS-86, mixed mode)

Wrong : 'java' is not recognized as an internal or external command, operable program or batch file.

-For Sql:

Cmd : **mysql --version**

Output:

Coorect mysql Ver 8.0.35 for Win64 on x86\_64 (MySQL Community Server - GPL)

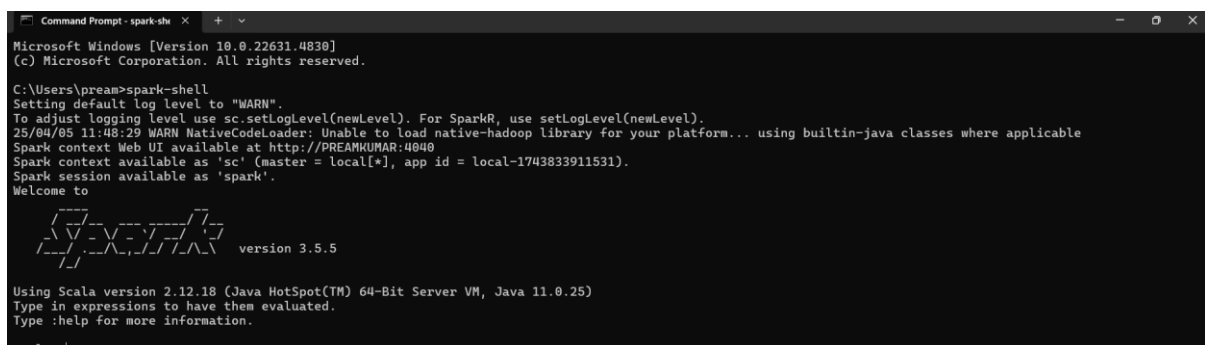
Wrong : 'mysql' is not recognized as an internal or external command, operable program or batch file.

-spark:

Cmd : **spark-shell**

Output:

Coorect :



```
Microsoft Windows [Version 10.0.22631.4830]
(c) Microsoft Corporation. All rights reserved.

C:\Users\pream>spark-shell
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
25/04/05 11:48:29 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Spark context Web UI available at http://PREAMKUMAR:4040
Spark context available as 'sc' (master = local[*], app id = local-1743833911531).
Spark session available as 'spark'.
Welcome to

  ____              __
 / ___/___  ___  ___/  /___  ___
/ /  / __ \/ __ \/ __/ __/ __/
/___/_/ /_/_/ /___/_/ /___/_/

version 3.5.5

Using Scala version 2.12.18 (Java HotSpot(TM) 64-Bit Server VM, Java 11.0.25)
Type in expressions to have them evaluated.
Type :help for more information.

scala>
```

Wrong : 'spark-shell' is not recognized as an internal or external command, operable program or batch file.

