

Spark Local Installation

Installing java 8

<https://www.oracle.com/in/java/technologies/javase/javase8-archive-downloads.html>

To verify the installation: `java -version`

Installing Scala

<https://www.scala-lang.org/download/2.13.11.html>

To verify the installation: `scala -version`

Installing Apache Spark

<https://spark.apache.org/downloads.html>

To verify the downloaded file use the command

```
certutil -hashfile C:\users\xs334-amakum\Downloads\spark-3.4.0-bin-hadoop3-scala2.13.tgz SHA512
```

Match the output given by the above code with the value given under checksum in the downloads page

After downloading use the following commands in PowerShell:

Step 1: `cd \`

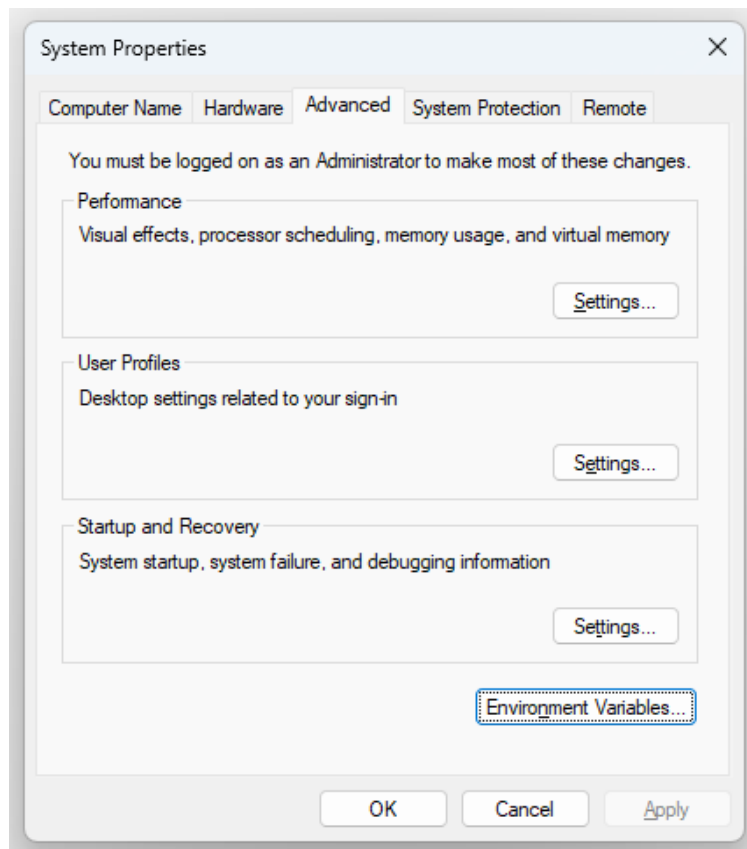
Step 2: `mkdir spark`

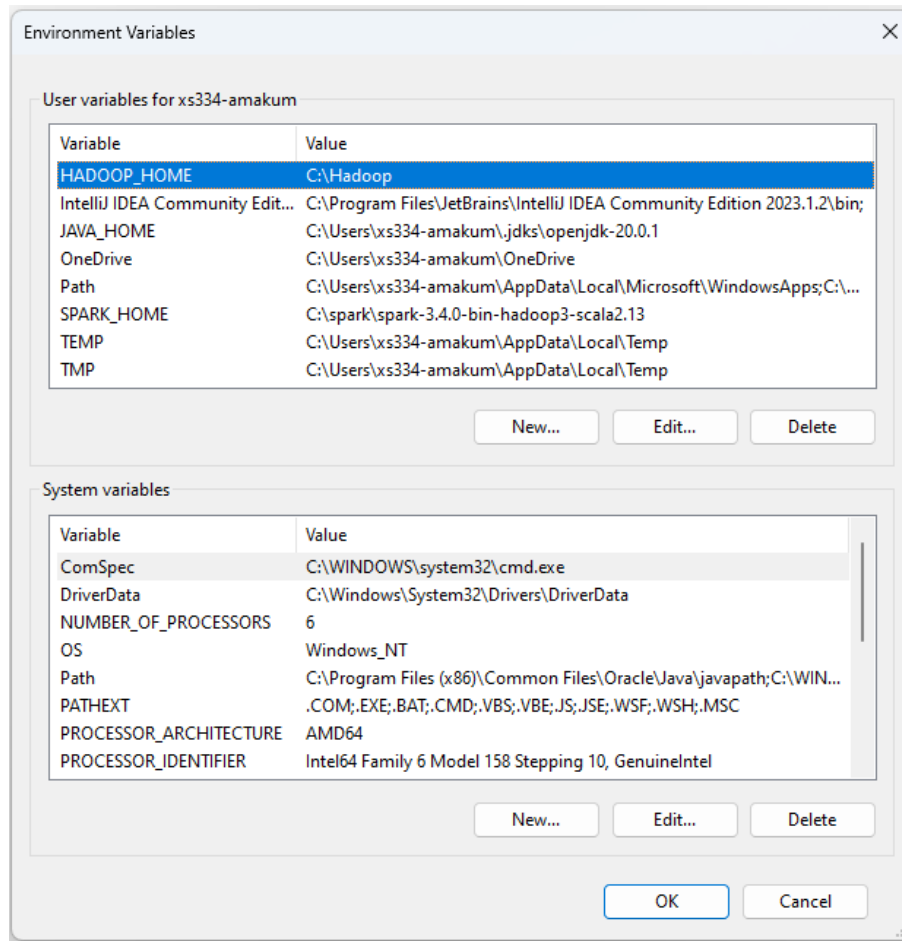
Step 3: Now extract the downloaded spark (.tgz file) in the spark directory

Step 4: Move to the directory containing the extracted file

Step 5: Download winutils.exe from : <https://github.com/cdarlint/winutils>

Step 6: Set the environment variables like **HADOOP_HOME**, **JAVA_HOME**, **SPARK_HOME**, etc.





Step 7: Set the path variables

Step 8: Finally, to run spark, go to the directory containing the extracted spark file and then enter **spark-shell**

```
PS C:\spark\spark-3.4.0-bin-hadoop3-scala2.13> spark-shell
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
Welcome to

    /---/
   /_/_/_/_/_/_/_/_/_/_/_/_/_/_/_/_/_\_
  _\ \/_/_\_/_/_\_/_/_/_/_/_/_/_/_/_/_
 /___/_/.__/_/_/_/_/_/_/_/_/_/_/_/__\ version 3.4.0
  /_/

Using Scala version 2.13.8 (OpenJDK 64-Bit Server VM, Java 20.0.1)
Type in expressions to have them evaluated.
```

To stop the spark-shell enter **:quit**