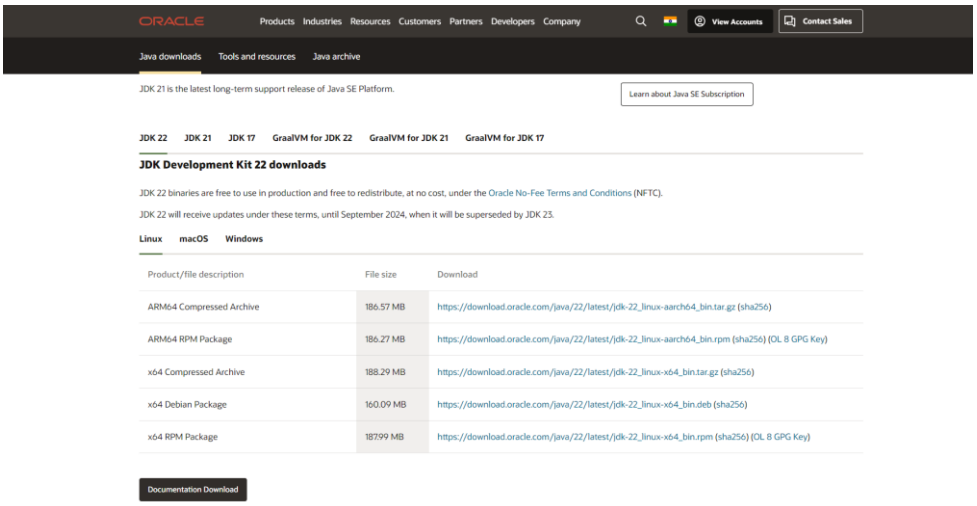
Spark Installation Guide (Windows)

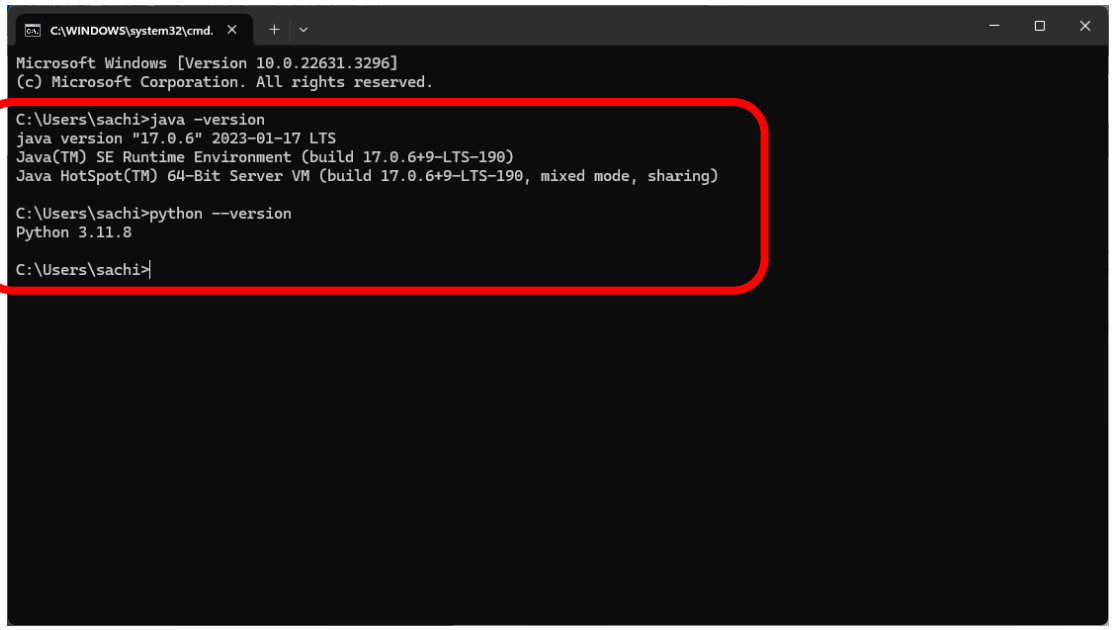
1. Installation of Java and Python necessary:

- Java Download:
- https://www.oracle.com/java/technologies/java-se-glance.html



2. Check Python and Java Versions using Windows CMD:

- java -version
- python –version

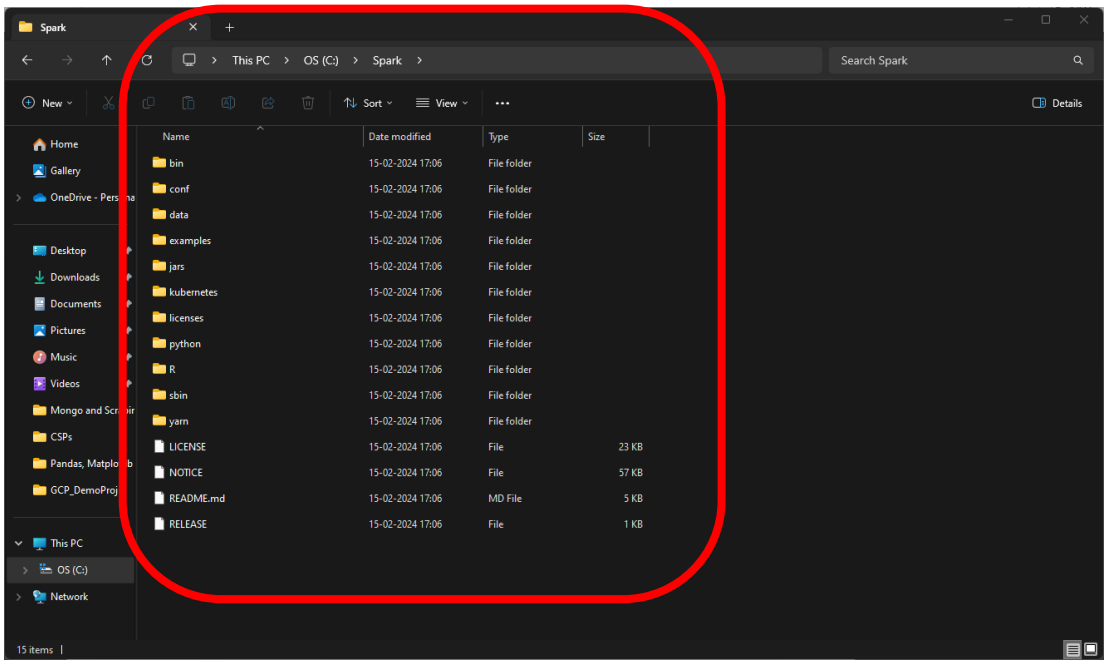


3. Download Spark Files and copy it in C:/SPARK folder:

- https://spark.apache.org/downloads.html

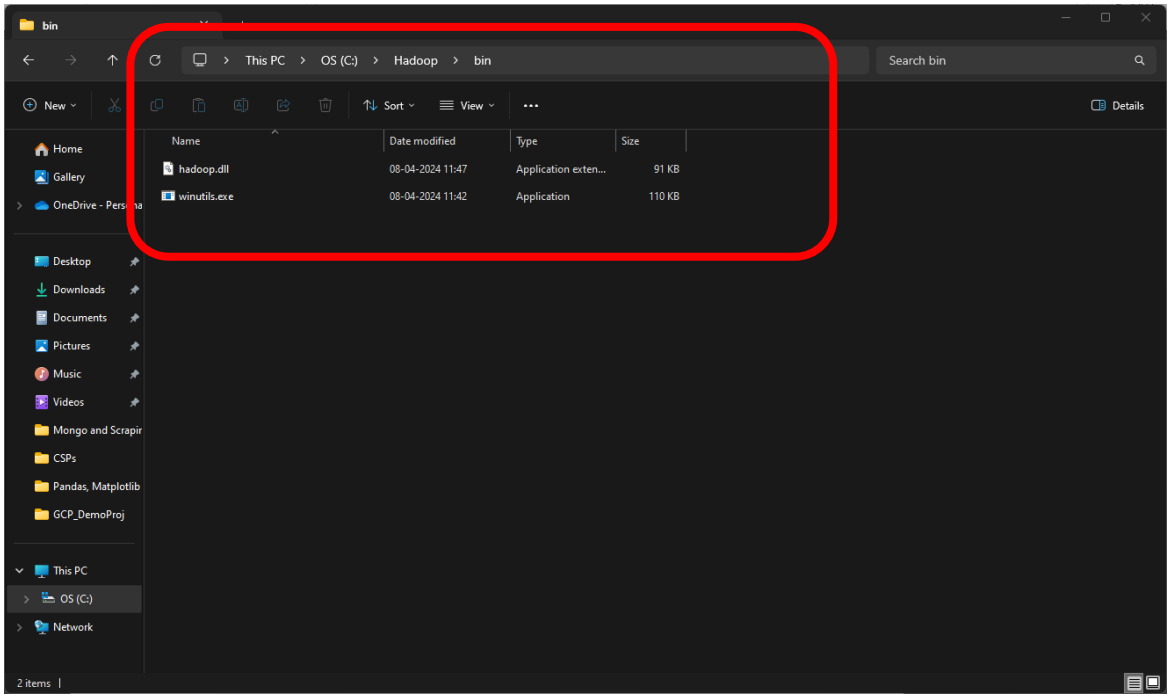
Alternatively, we can use direct download link:

- https://dlcdn.apache.org/spark/spark-3.5.1/spark-3.5.1-bin-hadoop3.tgz



4. Download WinUtils and copy it in C:/HADOOP/BIN folder:

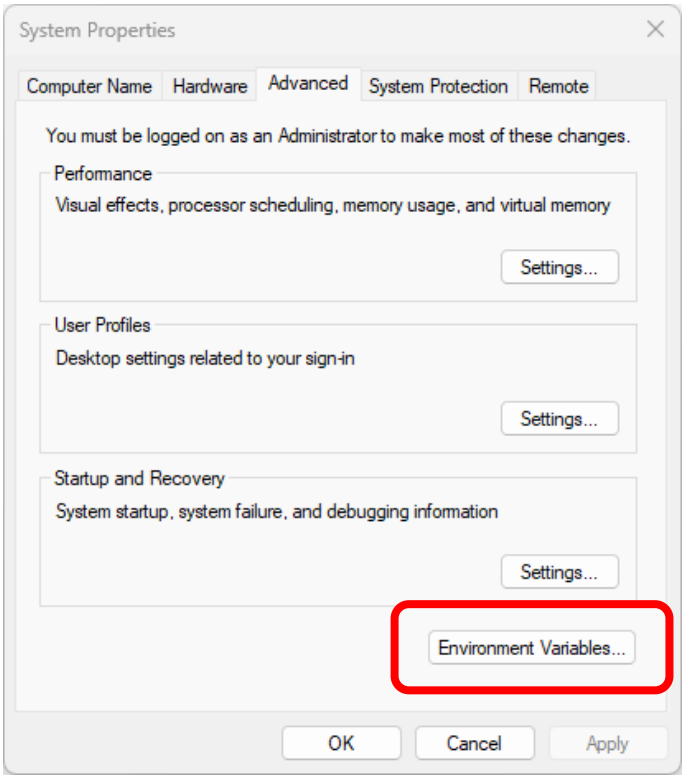
- https://github.com/steveloughran/winutils/tree/master/hadoop-3.0.0/bin
- https://github.com/steveloughran/winutils/blob/master/hadoop-3.0.0/bin/winutils.exe
- https://github.com/steveloughran/winutils/blob/master/hadoop-3.0.0/bin/hadoop.dll

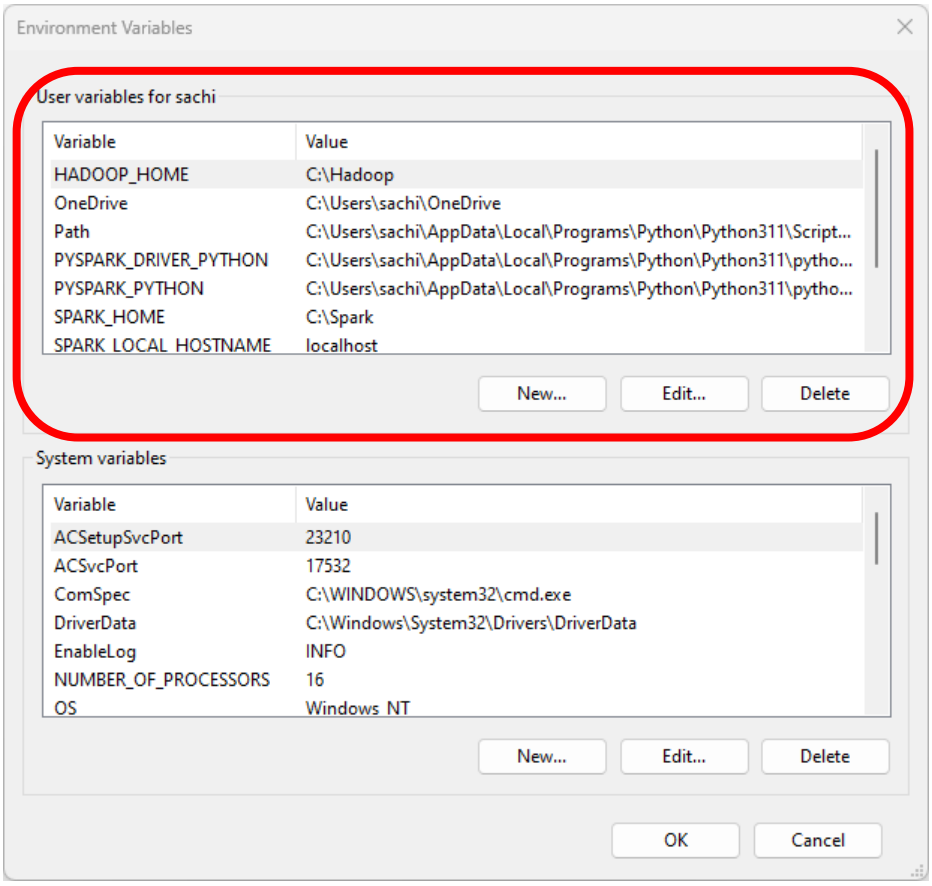


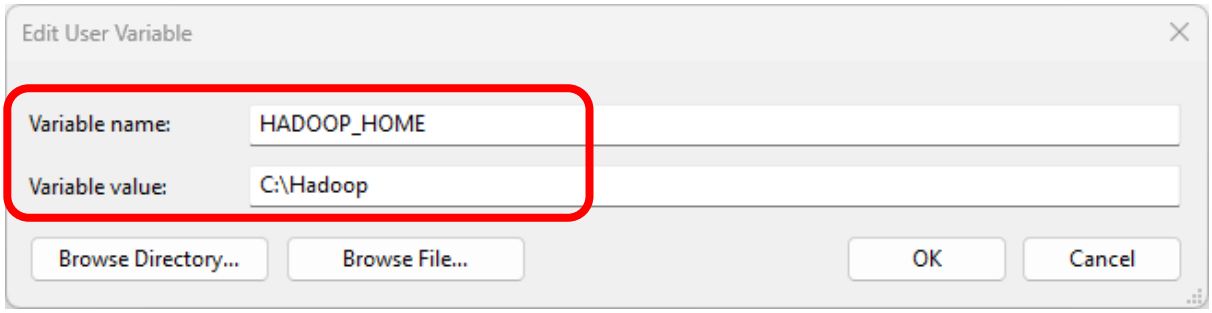
5. Set the ENVIRONMENT PATH:

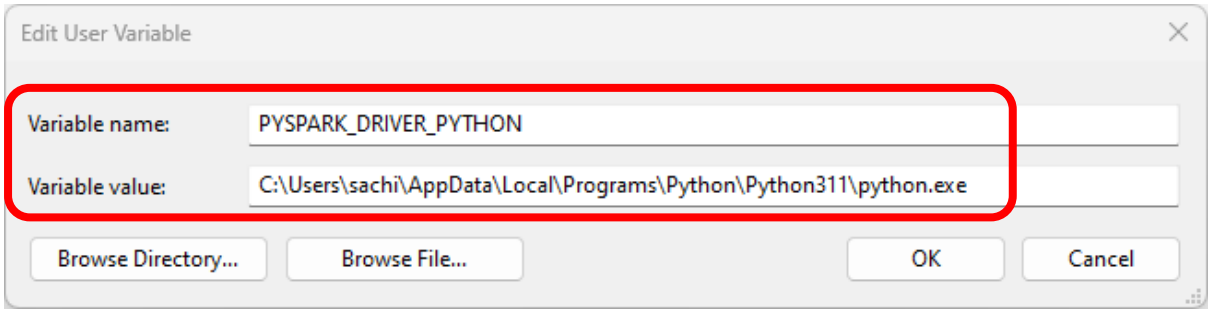
- SPARK_HOME = /**PATH**/
- HADOOP_HOME = /**PATH**/

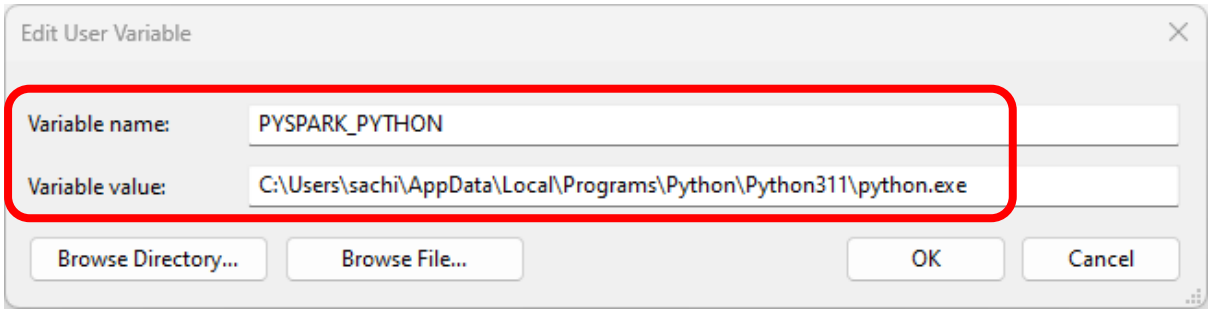
- SPARK_LOCAL_HOSTNAME = localhost
- PYSPARK_PYTHON = /**PYTHON EXECUTABLE**/
- PYSPARK_DRIVER_PYTHON = /**PYTHON EXECUTABLE**/

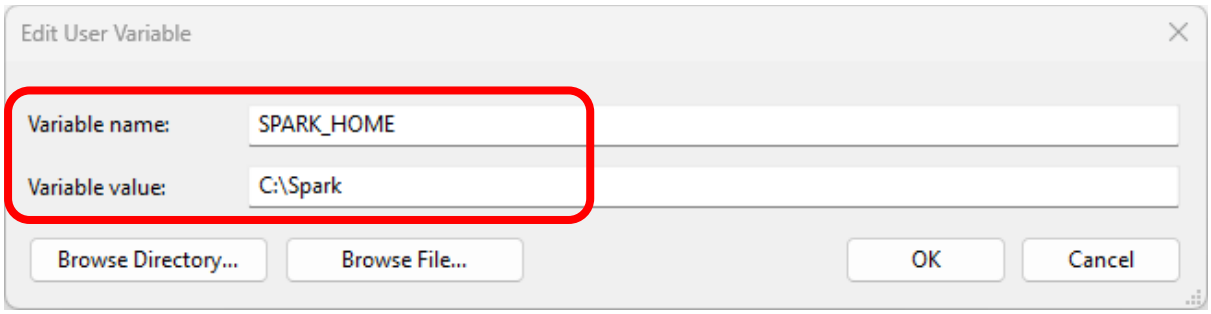


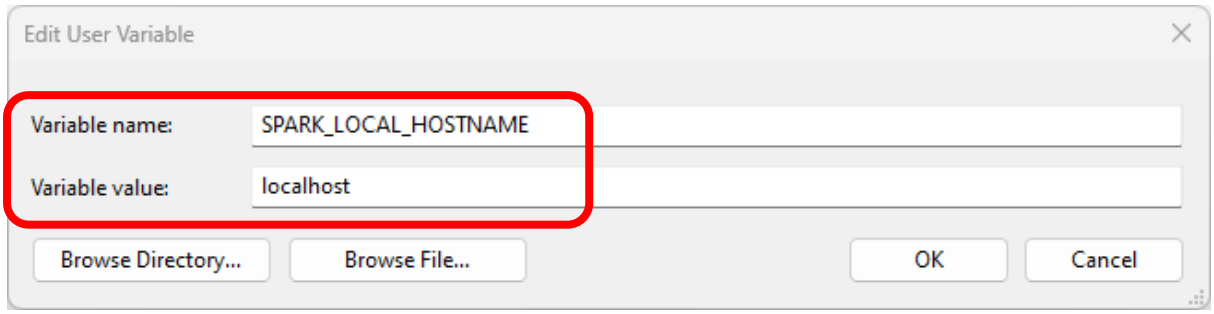


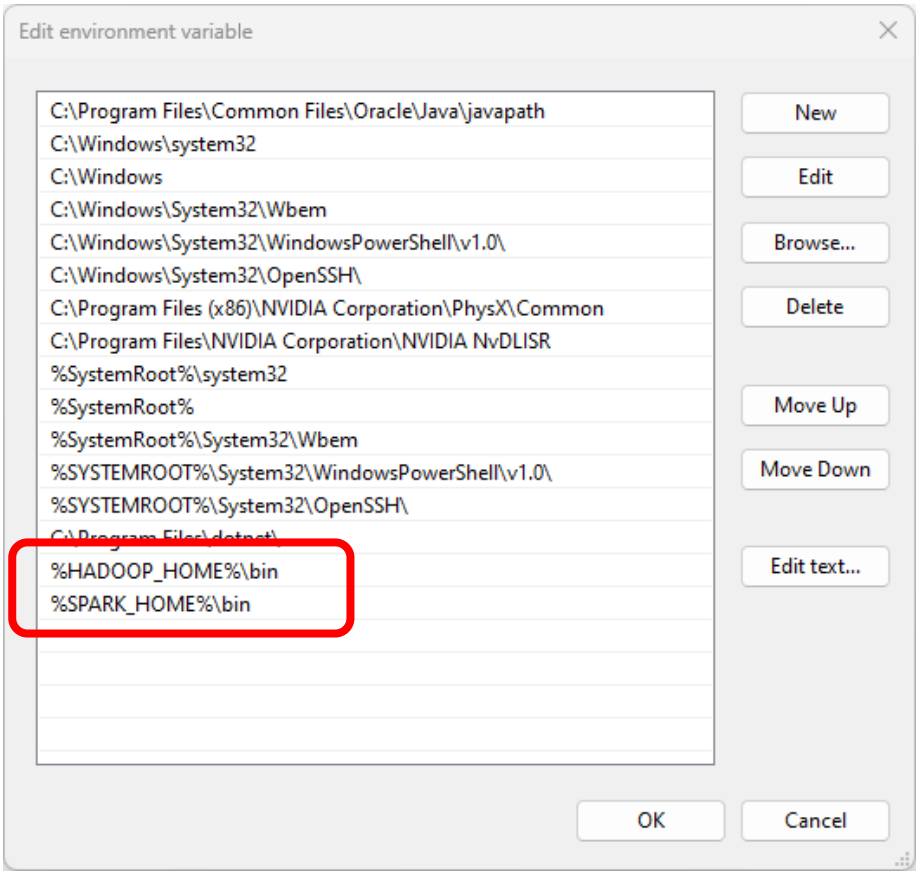

Edit User Variable
Variable name:
HADOOP_HOME
Variable value:
C:\Hadoop
Browse Directory...
Browse File...
OK
Cancel


Edit User Variable
Variable name:
PYSPARK_DRIVER_PYTHON
Variable value:
C:\Users\sachi\AppData\Local\Programs\Python\Python311\python.exe
Browse Directory...
Browse File...
OK
Cancel


Edit User Variable
Variable name:
PYSPARK_PYTHON
Variable value:
C:\Users\sachi\AppData\Local\Programs\Python\Python311\python.exe
Browse Directory...
Browse File...
OK
Cancel


Edit User Variable
Variable name:
SPARK_HOME
Variable value:
C:\Spark
Browse Directory...
Browse File...
OK
Cancel


Edit User Variable
Variable name:
SPARK_LOCAL_HOSTNAME
Variable value:
localhost
Browse Directory...
Browse File...
OK
Cancel



6. Run Command pyspark in cmd terminal to check if Spark is loaded