

ASSESSMSNT

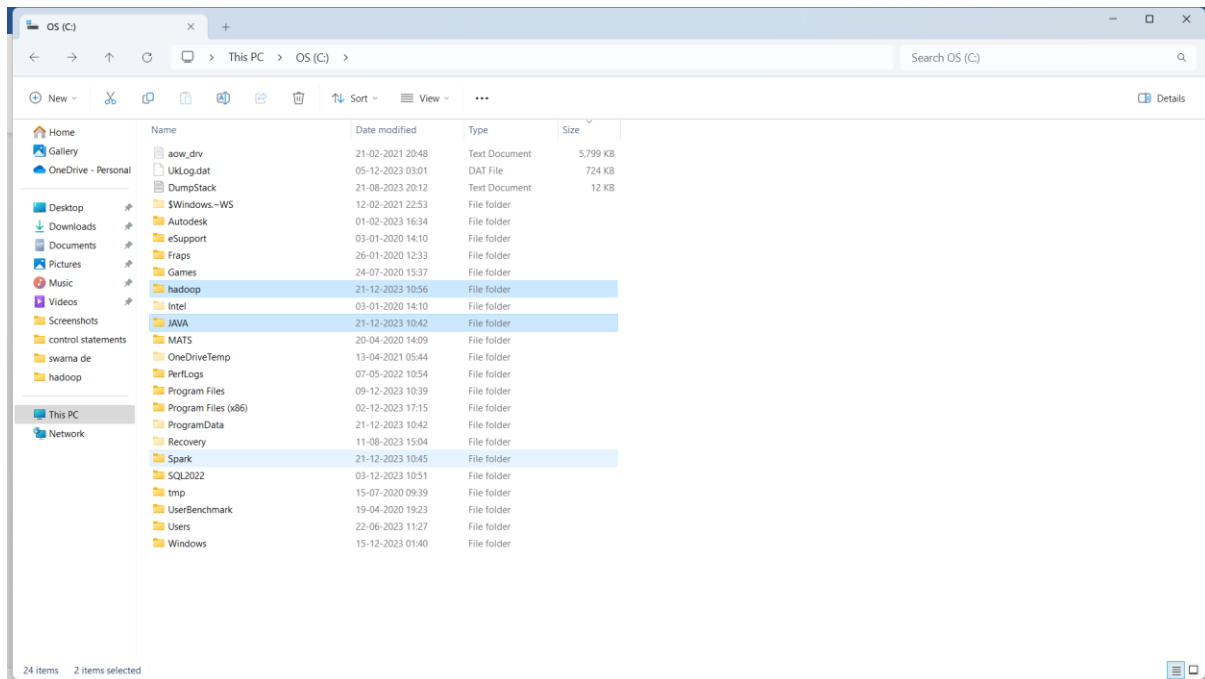
NAME-SWARNA LATHA GUVVALA

DATE-21/12/2023

STEP-1: DOWNLOAD THE BELOW LINKS FOR INSTALLATION PROCESS

- 1.Download JDK: <https://www.oracle.com/in/java/techno...>
2. Download Python: <https://www.python.org/downloads/>
3. Download Spark: <https://spark.apache.org/downloads.html>
4. Winutils repo link: <https://github.com/steveloughran/winutils>

STEP-2: EXTRACT DOWNLOADED FILES IN NEW FLOADER IN PC



STEP-3: EDIT ENVIRONMENT VARIABLES

Environment Variables:

HADOOP_HOME- C:\hadoop\bin

JAVA_HOME- C:\JAVA\JDK

SPARK_HOME- C:\Spark\spark-3.5.0-bin-hadoop3\spark-3.5.0-bin-hadoop3

PYTHON_HOME-C:\Users\Sri Vishnu
Reddy\AppData\Local\Programs\Python\Python39\python.exe

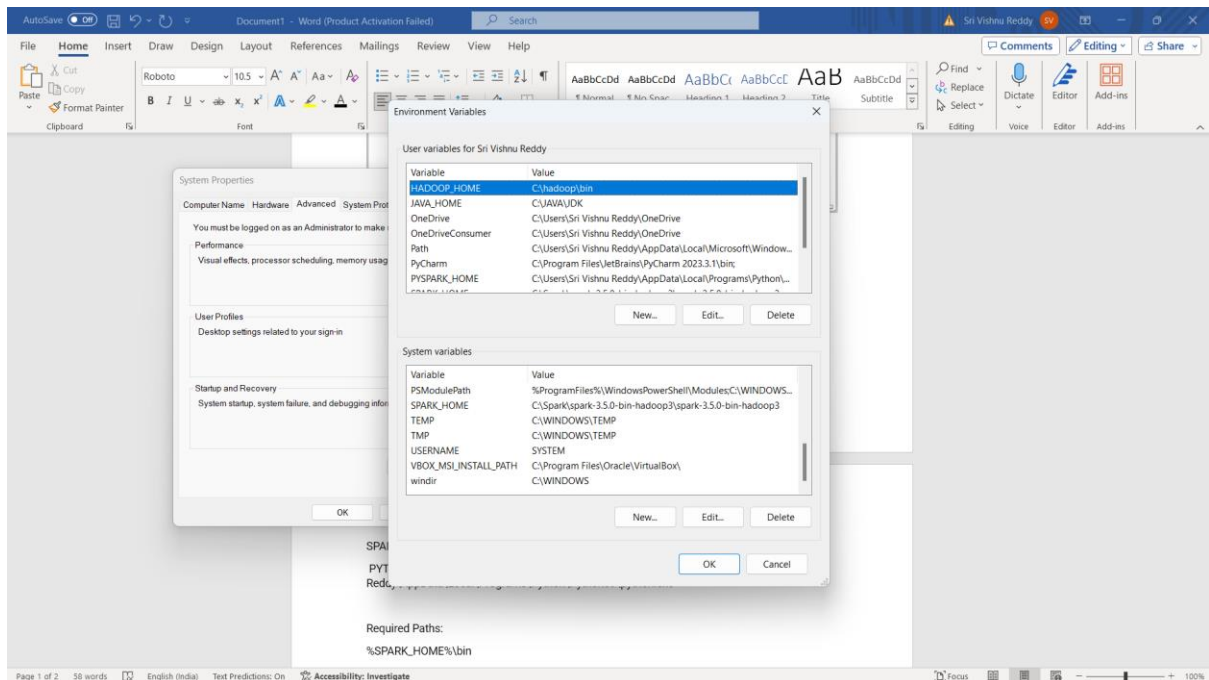
Required Paths:

%SPARK_HOME%\bin

%HADOOP_HOME%\bin

%JAVA_HOME%\bin

%PYHTON_HOME%\bin



```
Select Command Prompt - spark-shell
Microsoft Windows [Version 10.0.22621.2861]
(c) Microsoft Corporation. All rights reserved.

C:\Users\Sri Vishnu Reddy>java --version
java 19.0.2 2023-01-17
Java(TM) SE Runtime Environment (build 19.0.2+7-44)
Java HotSpot(TM) 64-Bit Server VM (build 19.0.2+7-44, mixed mode, sharing)

C:\Users\Sri Vishnu Reddy>python --version
Python 3.11.6

C:\Users\Sri Vishnu Reddy>spark-shell
23/12/22 09:15:25 WARN Shell: Did not find winutils.exe: java.io.FileNotFoundException: Hadoop bin directory does not exist: C:\hadoop\bin -see https://wiki.apache.org/hadoop/WindowsProblems
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
23/12/22 09:15:38 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Spark context Web UI available at http://LAPTOP-F3I6TQIE:4040
Spark context available as 'sc' (master = local[*], app id = local-1703216748911).
Spark session available as 'spark'.
Welcome to

  ____      _
 / ___|  __| | | |
 \___ \  | | | | | |
  ___) | | | | | | |
 |_____|_|_|_|_|_|_|

version 3.5.0

Using Scala version 2.12.18 (Java HotSpot(TM) 64-Bit Server VM, Java 19.0.2)
Type in expressions to have them evaluated.
Type :help for more information.

scala> █
```

```
Command Prompt - pyspark
Microsoft Windows [Version 10.0.22621.2861]
(c) Microsoft Corporation. All rights reserved.

C:\Users\Sri Vishnu Reddy>pyspark
Python 3.11.6 (tags/v3.11.6:8b6ee5b, Oct 2 2023, 14:57:12) [MSC v.1935 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
23/12/22 09:17:27 WARN Shell: Did not find winutils.exe: java.io.FileNotFoundException: Hadoop bin directory does not exist: C:\hadoop\bin -see https://wiki.apache.org/hadoop/WindowsProblems
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
23/12/22 09:17:28 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Welcome to

  ____      _
 / ___|  __| | | |
 \___ \  | | | | | |
  ___) | | | | | | |
 |_____|_|_|_|_|_|_|

version 3.5.0

Using Python version 3.11.6 (tags/v3.11.6:8b6ee5b, Oct 2 2023 14:57:12)
Spark context Web UI available at http://LAPTOP-F3I6TQIE:4040
Spark context available as 'sc' (master = local[*], app id = local-1703216851173).
SparkSession available as 'spark'.
>>>
```

Spark Jobs ^(?)

User: Sri Vishnu Reddy
Total Uptime: 1.2 min
Scheduling Mode: FIFO

- ▼ Event Timeline
☐ Enable zooming

