1. Download and Install Anaconda

[Anaconda | The World's Most Popular Data Science Platform]

While installing, click on: just me(recommended)

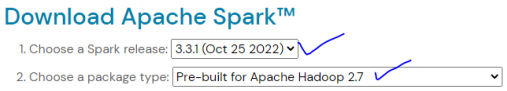
Add Anaconda3 to my PATH environment variables

2. Download and Install latest Java software

https://www.oracle.com/java/technologies/downloads/#jdk19-windows

3. Download and Extract spark-3.3.1-bin-hadoop2
   cut past the extracted file n C-drive

For Hadoop to work in cmd : Type in Google Download Apache Spark



Downloads | Apache Spark click step 3it'll direct to below link

https://dlcdn.apache.org/spark/spark-3.3.1/spark-3.3.1-bin-hadoop2.tgz

Download winutils.exe from this link

winutils/winutils.exe at master · steveloughran/winutils (github.com) **select** *(Hadoop 2.7.1)*

Copy past it in spark-3.3.1-bin-hadoop2/bin

cut past the extracted file(spark-3.3.1-bin-hadoop2) in C-drive

NOTE: Just copy past "spark-3.3.1-bin-hadoop2(with_winutils)" it contains all

4. Press windows type '**Edit the system environment**'
   Open Environmental Variable
   - **System variables** click on new
     Variable name: SPARK_HOME
     Variable values: Browse C:\spark-3.3.1-bin-hadoop2
     Variable name: HADOOP_HOME
     Variable values: Browse C:\spark-3.3.1-bin-hadoop2
   - **User Variable:**
     path: *edit: New*: %SPARK_HOME%\bin press Enter
     %JAVA_HOME%\bin press Enter
     New: *Variable name*: PYTHONPATH
     Variable values: %SPARK_HOME%\hadoop3\python\lib\py4j-0.10.9.5-src.zip

(Also: %SPARK_HOME%\hadoop3\python\lib;%SPARK_HOME%\hadoop3\python;%SPARK_HOME%\hadoop3\python\lib\py4j-0.10.9.5-src.zip , click on- Browse file: C:\spark-3.3.1-bin-hadoop2\python\lib\py4j-0.10.9.5-src.zip)

Ok, ok, ok

5. Open cmd, type java then javac
   conda activate spark
   conda install openjdk
   pip install findspark
   Pyspark
   ## Quit() to Quit**, cntl+c**: to terminate, **cls** to clear, Conda create -n spark --clone base
   Jupyter notebook

To Start Again: open cmd, **conda activate spark,** jupyter notebook

```
import findspark
findspark.init()
import pyspark
```

1.
```
from pyspark import SparkContext, SparkConf
conf= SparkConf().setAppName("app").setMaster("local")

sc= SparkContext(conf=conf)
```

sc o/p: **SparkContext**

Spark UI *(Note: you can open Spark UI and check the reports, Analysis. Etc)*

Version `v3.3.0`

Master `local`

AppName `app`

2.
```
from pyspark.sql import SparkSession

spark = SparkSession .builder \
.appName("Python Spark SQL basic example") \
.config("spark.some.config.option", "some-value") \
.getOrCreate()
```

Spark

o/p: SparkSession - in-memory

SparkContext

Spark UI

Version `v3.3.0`

Master `local`

AppName `app`