**Terminal Commands for 1st Powershell:**

* py-env/scripts/activate
* python main.py

**Terminal Commands for 2nd Powershell:**

* py-env/scripts/activate
* docker compose up -d –build

**Important Paths to be set or check inside Virtual Environment:**

**(py-env) PS D:\PRACTICE WORK\SAMPLE PROJECTS\KS-Project> echo $env:HADOOP\_HOME**

C:\hadoop\hadoop-3.3.6

**(py-env) PS D:\PRACTICE WORK\SAMPLE PROJECTS\KS-Project> echo $env:SPARK\_HOME**

C:\spark\spark-3.5.2-bin-hadoop3

**(py-env) PS D:\PRACTICE WORK\SAMPLE PROJECTS\KS-Project> echo $env:JAVA\_HOME**

C:\Program Files\Java\jdk-11

**(py-env) PS D:\PRACTICE WORK\SAMPLE PROJECTS\KS-Project> echo $env:HADOOP\_WINUTILS**

C:\hadoop\hadoop-3.3.6\bin\winutils.exe

**(py-env) PS D:\PRACTICE WORK\SAMPLE PROJECTS\KS-Project> echo $env:HADOOP\_NATIVE\_LIB**

C:\hadoop\hadoop-3.3.6\lib\native

**(py-env) PS D:\PRACTICE WORK\SAMPLE PROJECTS\KS-Project> echo $env:PYTHONPATH**

D:\PRACTICE WORK\SAMPLE PROJECTS\KS-Project\py-env\Scripts\python.exe

**(py-env) PS D:\PRACTICE WORK\SAMPLE PROJECTS\KS-Project> echo $env:PYSPARK\_PYTHON**

D:\PRACTICE WORK\SAMPLE PROJECTS\KS-Project\py-env\Scripts\python.exe

**Container Kafka-broker-1:**

kafka-topics --list --bootstrap-server kafka-broker:9092

kafka-console-consumer --topic properties --bootstrap-server kafka-broker:9092 --from-beginning

**Container Cassandra:**

cqlsh

SELECT \* FROM property\_streams.properties;

**Containers spark-master-1, spark-worker-1, spark-worker-2-1, spark-worker-3-1 :**

cd jobs

pip install -r jobs/requirements.txt

OR

pip install -r jobs/requirements.txt

**(py-env) PS D:\PRACTICE WORK\SAMPLE PROJECTS\KS-Project>** docker exec -it ks-project-spark-master-1 spark-submit --master spark://172.18.0.4:7077 --packages com.datastax.spark:spark-cassandra-connector\_2.12:3.5.1,org.apache.spark:spark-sql-kafka-0-10\_2.12:3.5.1 jobs/spark-consumer.py

Above command is used to run Spark, Kafka, Cassandra Connector

**(spark://172.18.0.4:7077) These spark version in above command have to be check every time on localhost:9090 whenever you installed requirements.txt file inside jobs folder in containers spark-master-1, spark-worker-1, spark-worker-2-1, spark-worker-3-1. And according to the spark localhost change the version of spark in above command and consumer.py file where clusters or spark is connecting with Cassandra.**

**I’ll also recommend you that if you are facing Hadoop issues then please must watch this video**

[**https://www.youtube.com/watch?v=2Gmk7OTU2vc**](https://www.youtube.com/watch?v=2Gmk7OTU2vc)