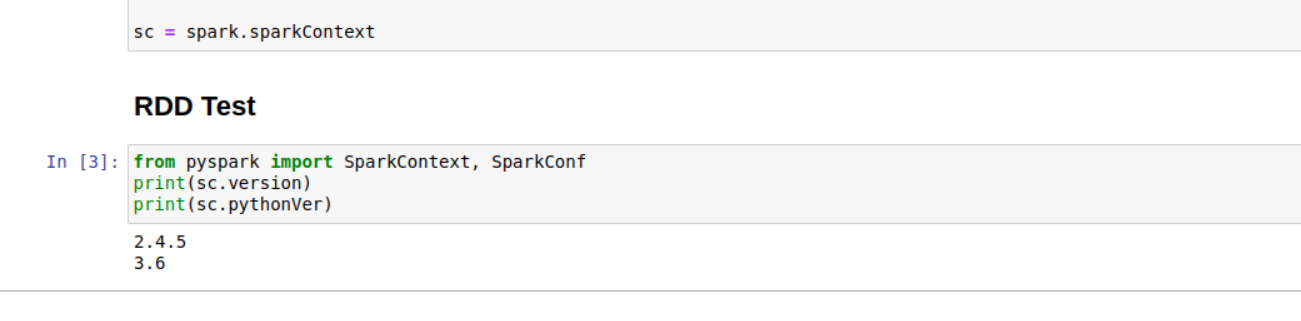
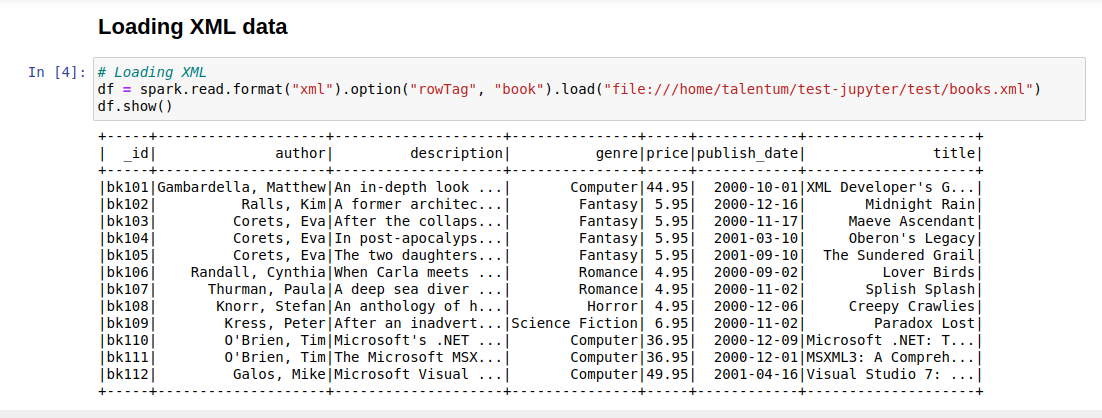


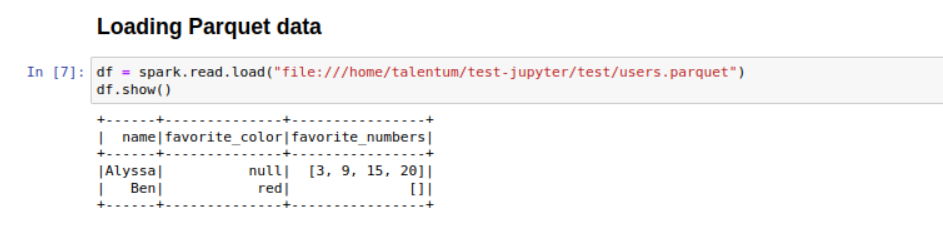
Run it











1] BigData Fundamentals With Pyspark (M1 -> Introduction to Big Data analysis with Spark)

(M1->SM1)

1) Download the attached note in STAGING\_AREA

2) Create a folder UBUNTU\_HOME/test-jupyter/P2/M1/SM1 in Ubuntu machine

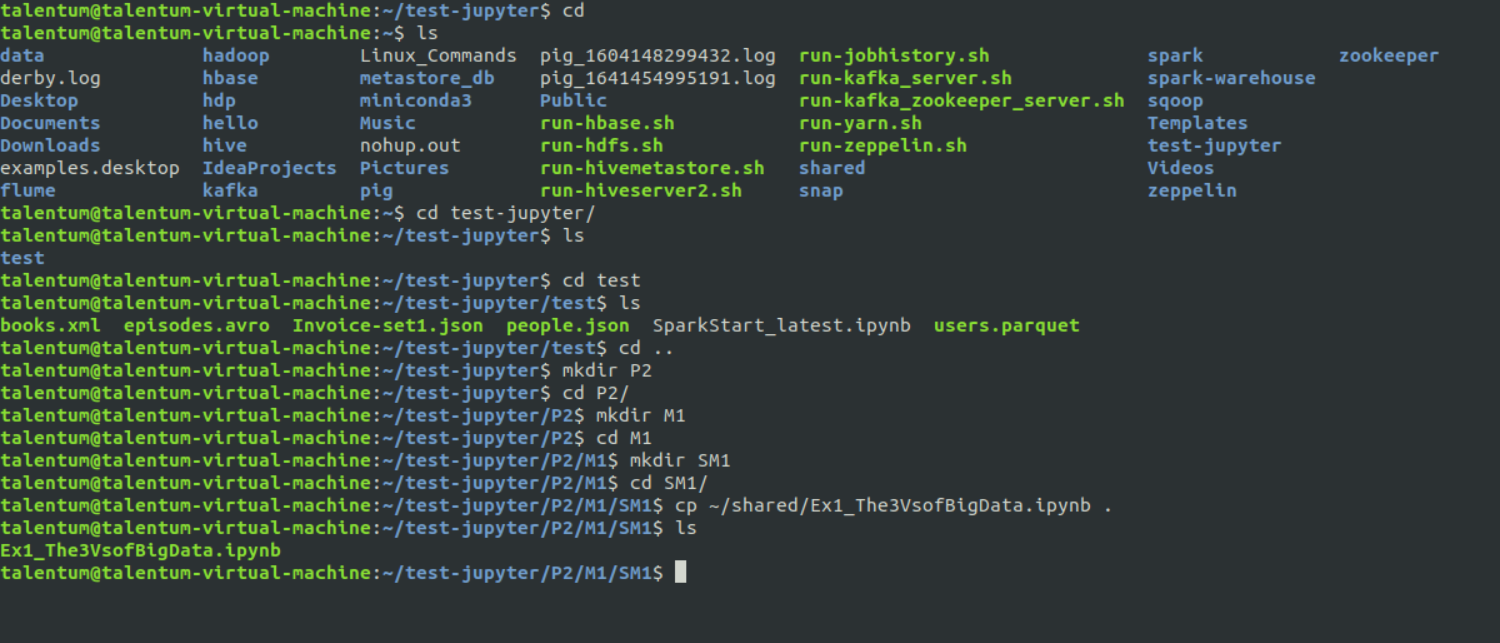
3) Pull the note in above location from your STAGING\_AREA

4) Open it in Jupyter Notebook running on Ubuntu Machine

* Code/Dataset

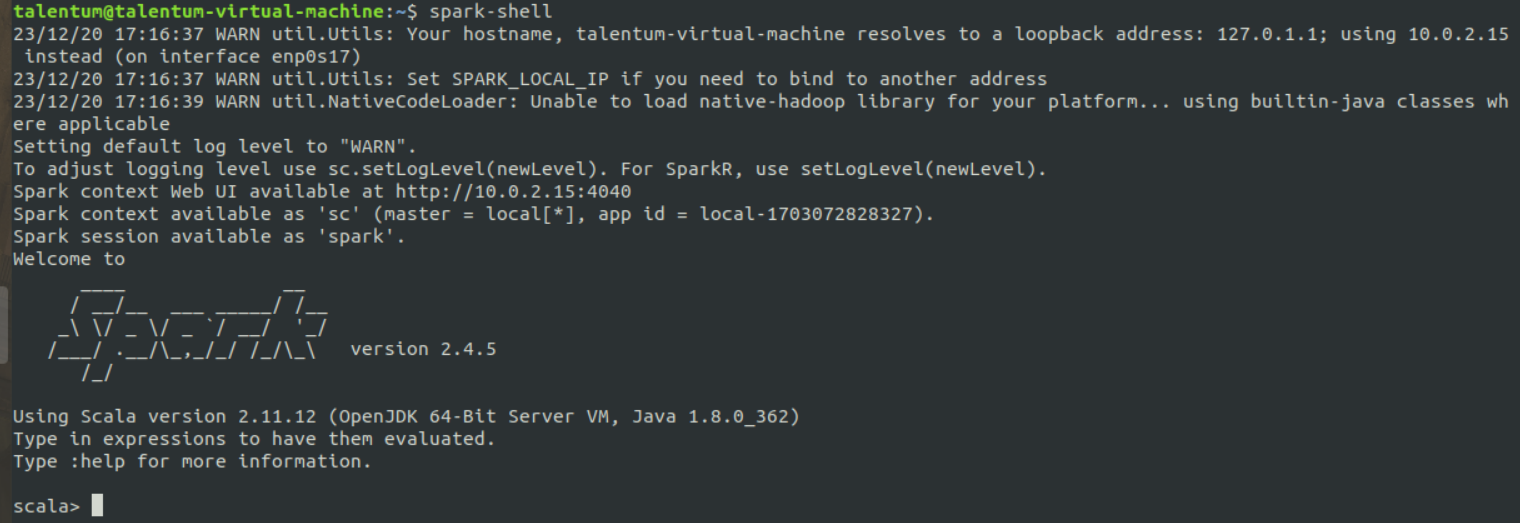
data/1\_Whatisbigdata.zip

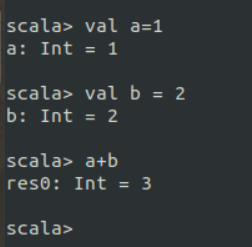
**TimeLine -** 5 Mins

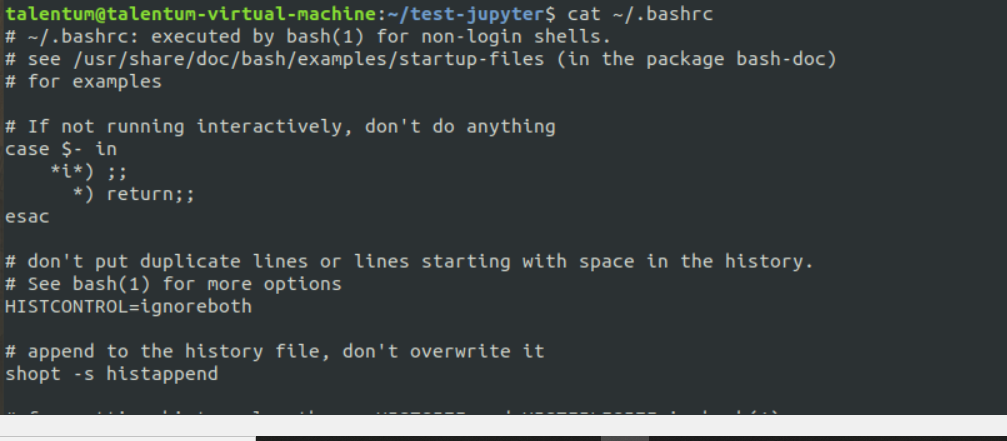


To open spark scala shell

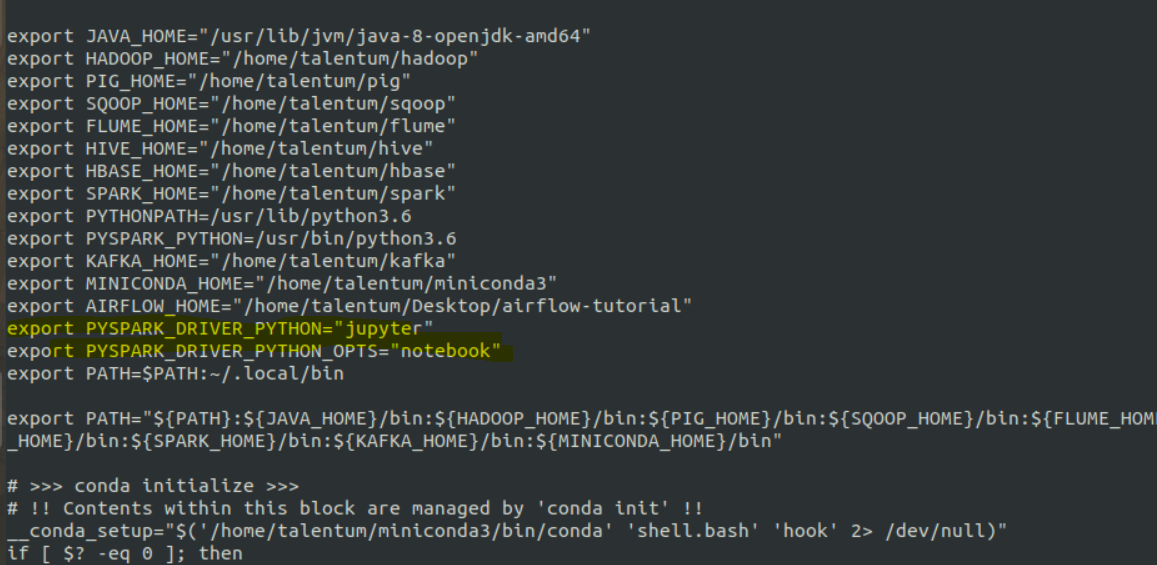
Spark-shell -> command

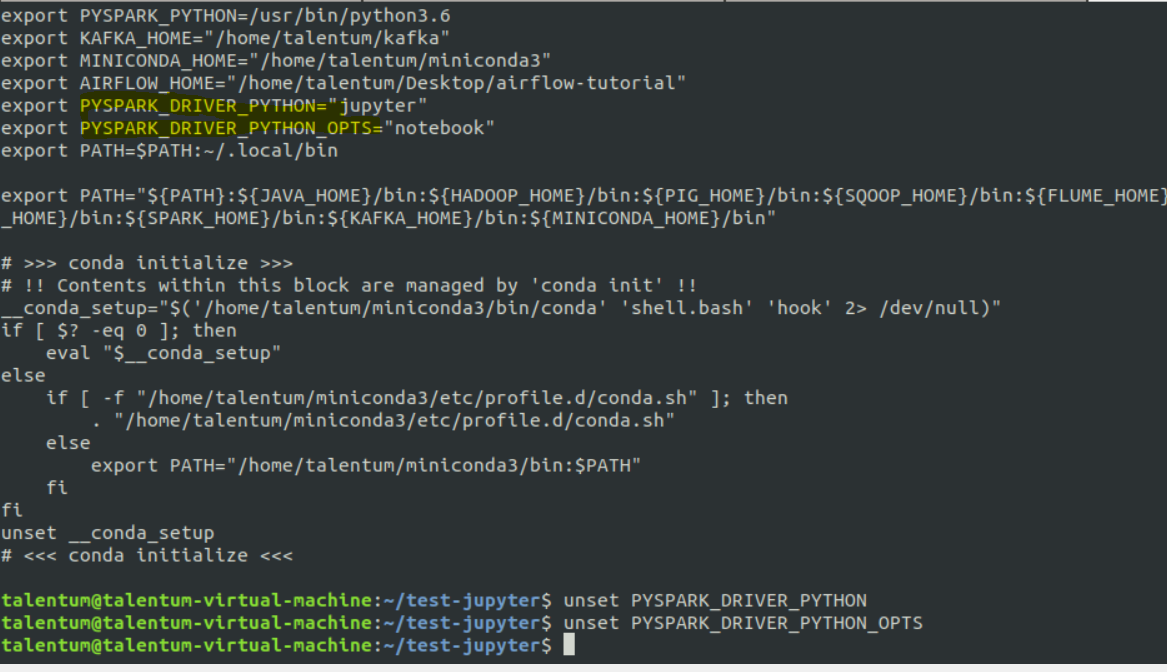


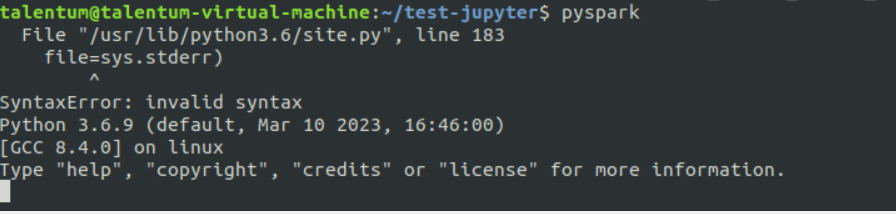




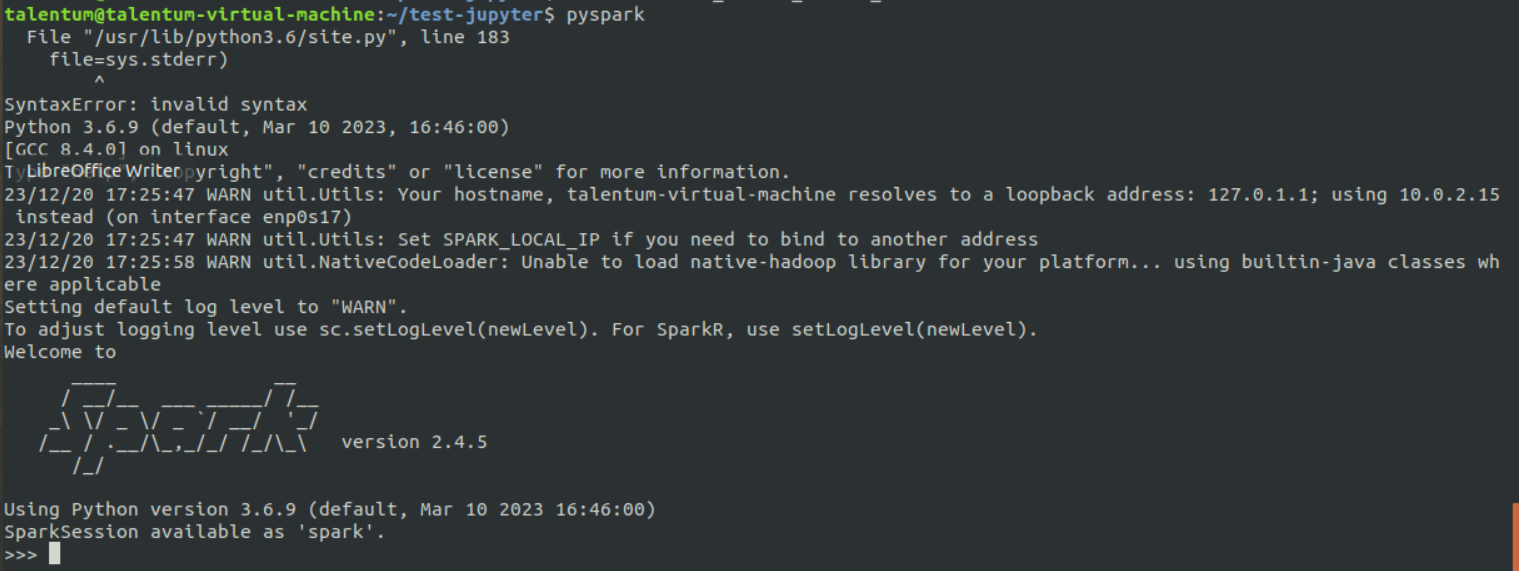
Unset





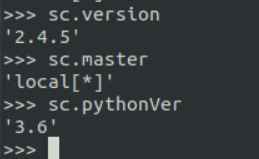


Now it will launch spark python shell





Spark Context SC



2] BigData Fundamentals With Pyspark (M1 Introduction to Big Data analysis with Spark)

PySpark: Spark with Python (M1->SM2)

1) Download RAR on STAGING\_AREA

2) Extract the RAR

3) Put the Jupyter notes in UBUNTU\_HOME/test-jupyter/P2/M1/SM2

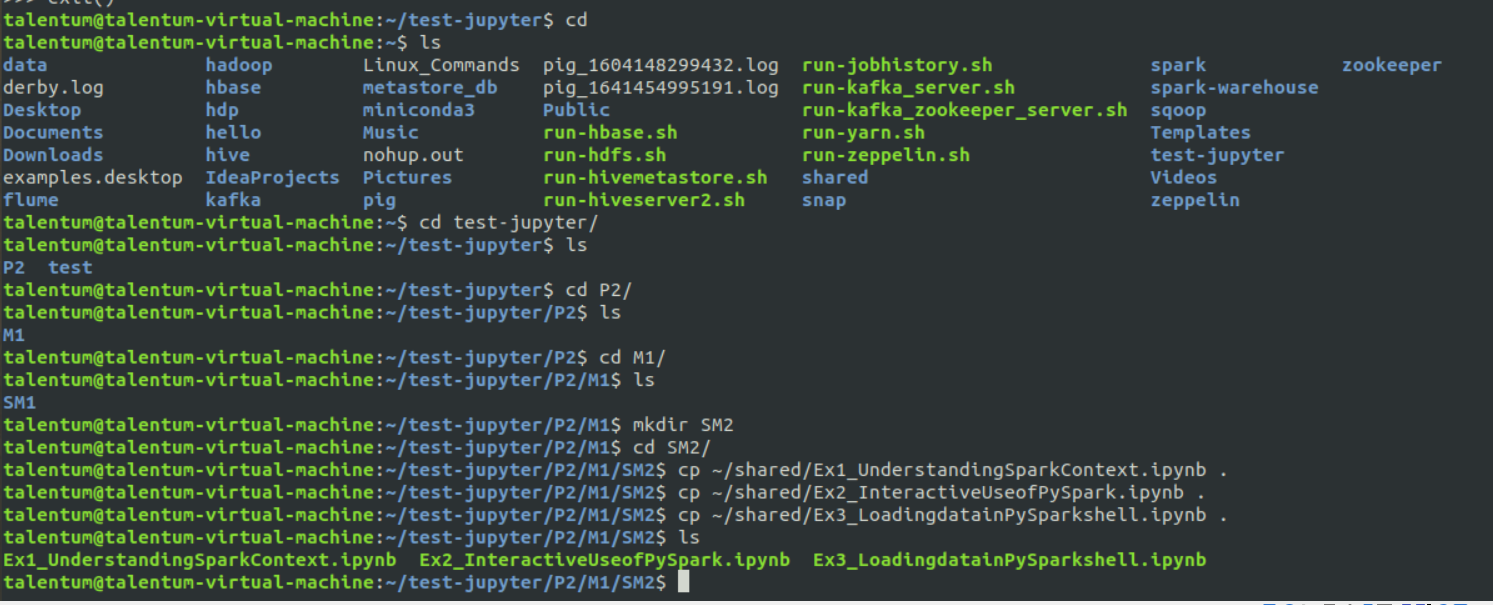
4) Import the notes in Jupyter Notebooks

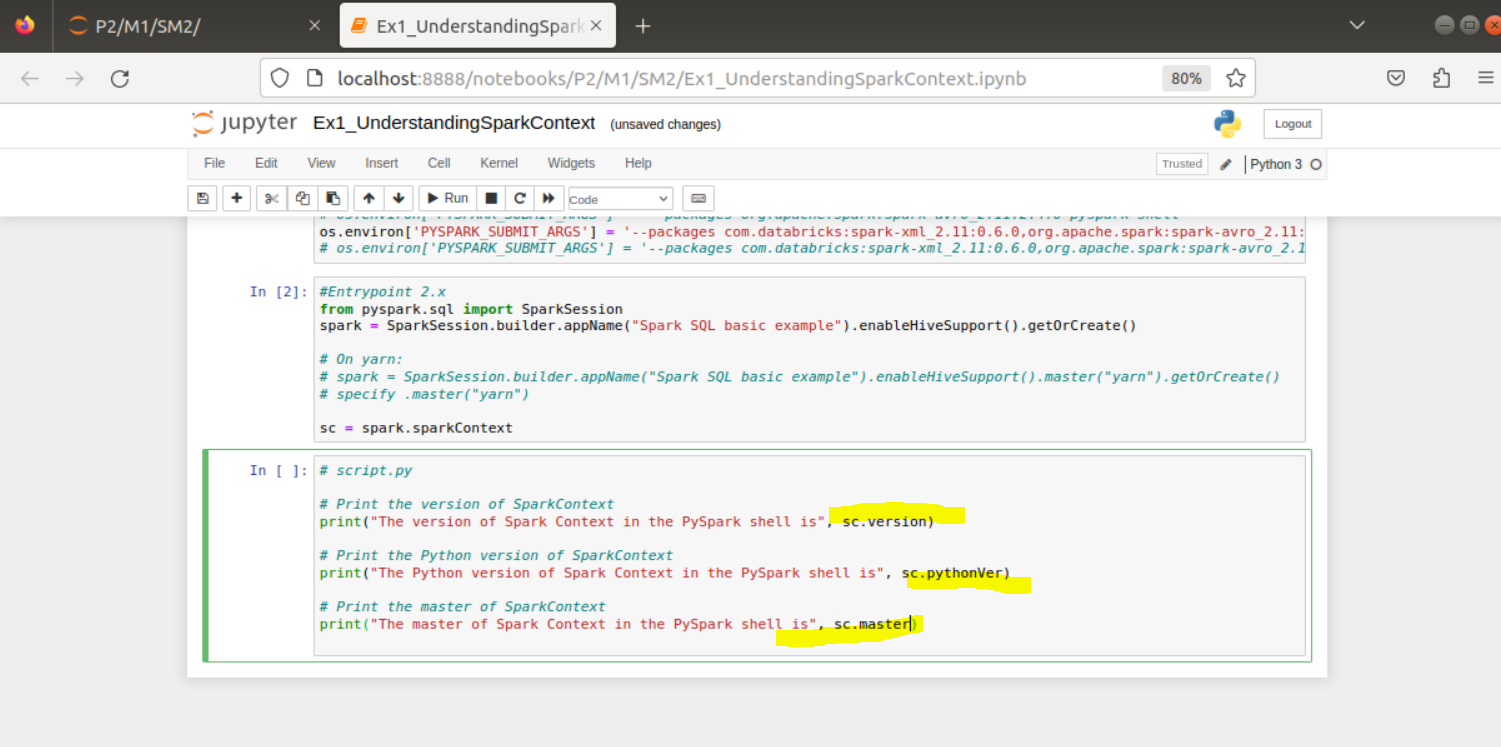
5) Follow the instructions in the notes and give the solutions

* Code/Dataset -

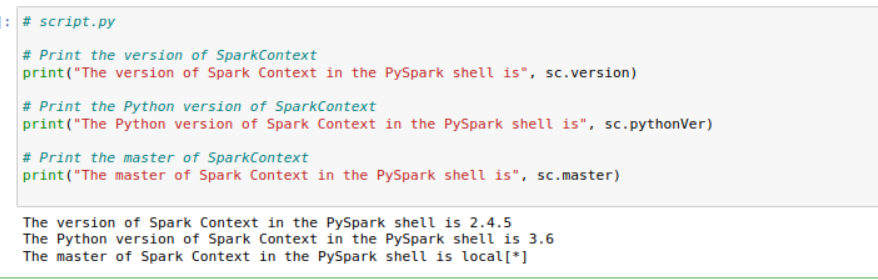
data/2\_Pyspark-SparkWithPython.rar

Time = 20 Mins.



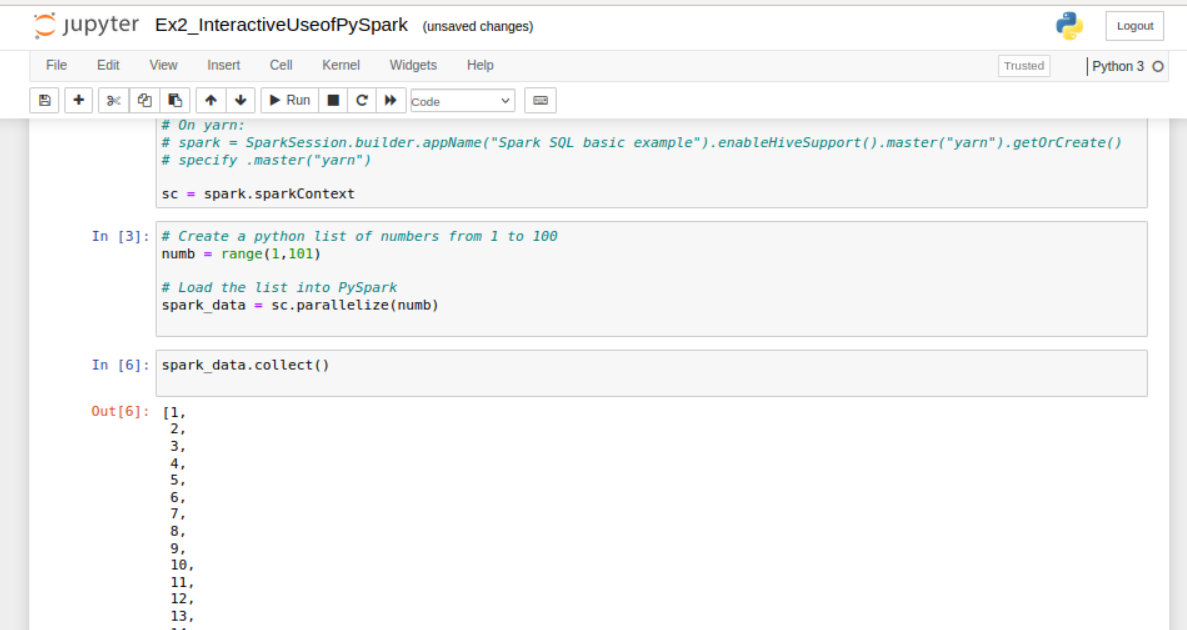


Run



Stop

Start 2nd



Stop

Start third



Trying same commands in pyspark

First unset PYSPARK\_DRIVER\_PYTHON

Unset PYSPARK\_DRIVER\_PYTHON\_OPTS

