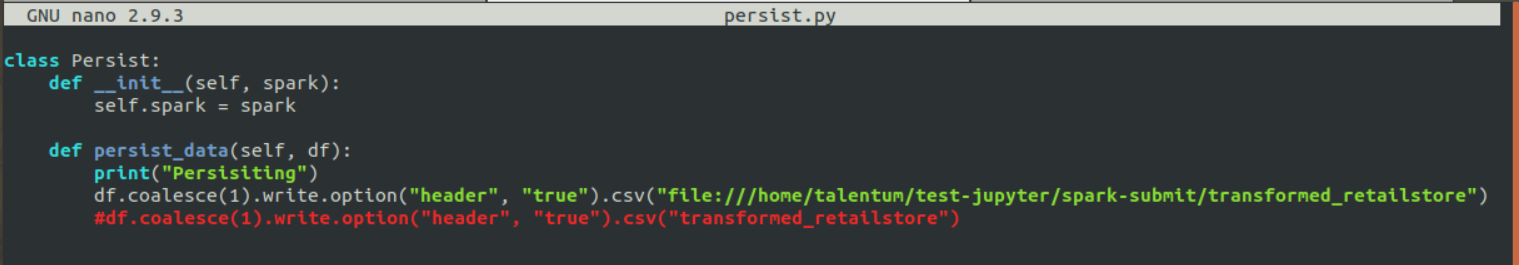
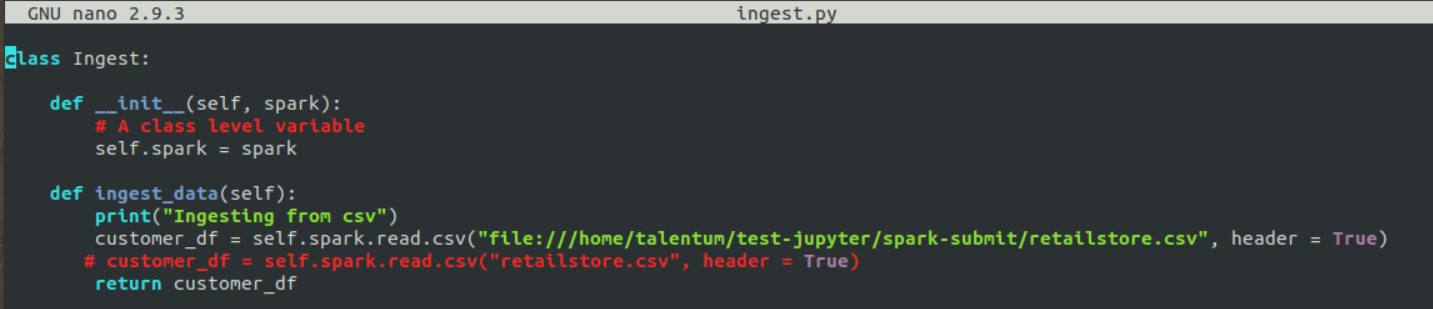
Spark Submit

Download and pull to Sandbox home directory

* Code/Dataset

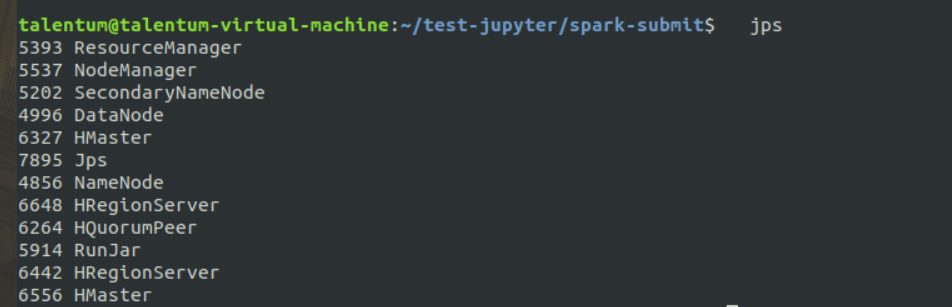
data/spark-submit.zip



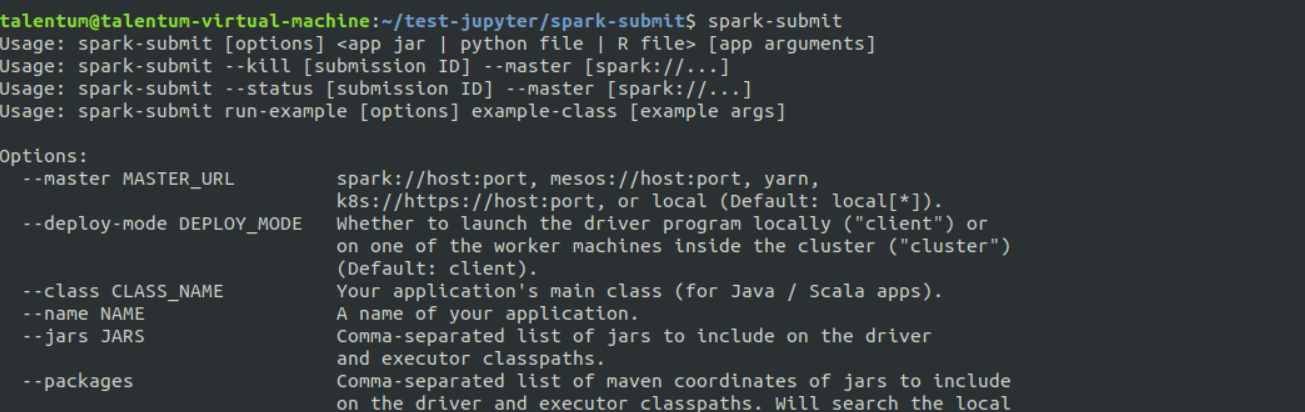


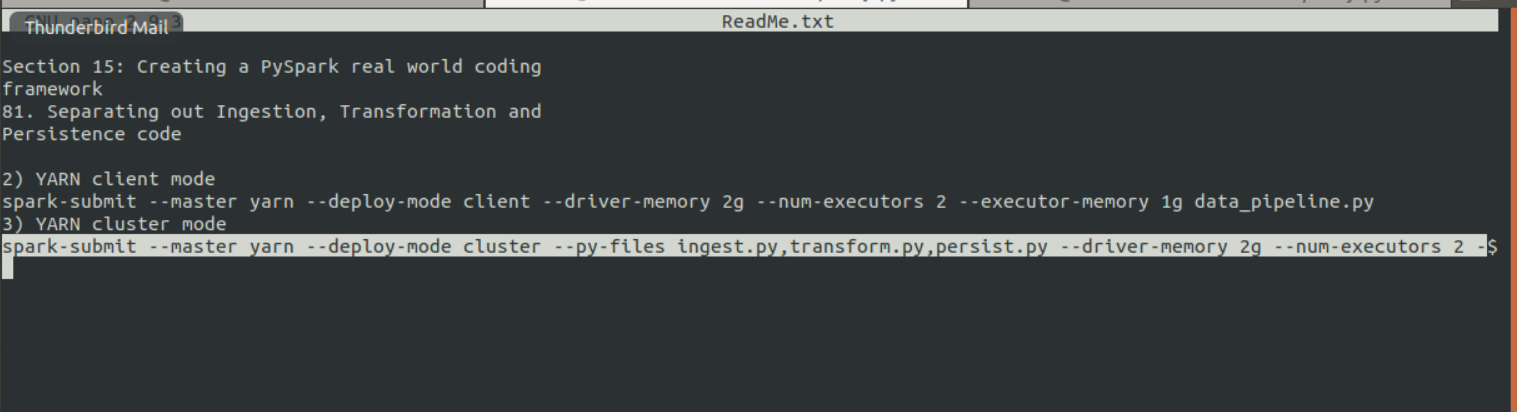
Start Hadoop

Check processes



Now see description of below commands



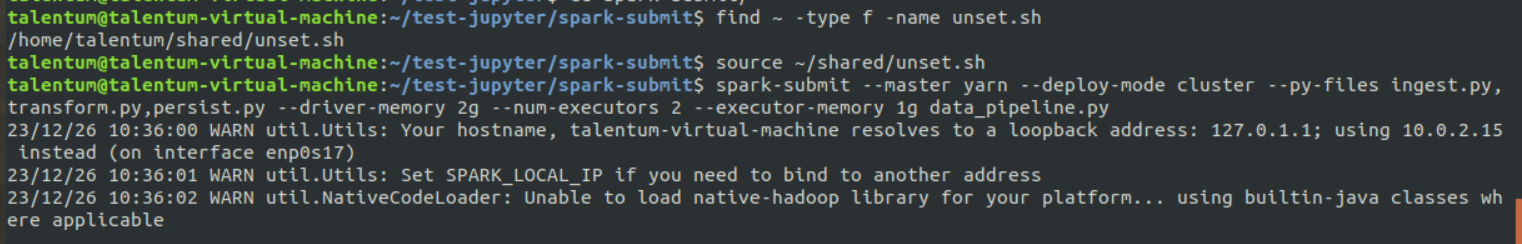


**Run this command**

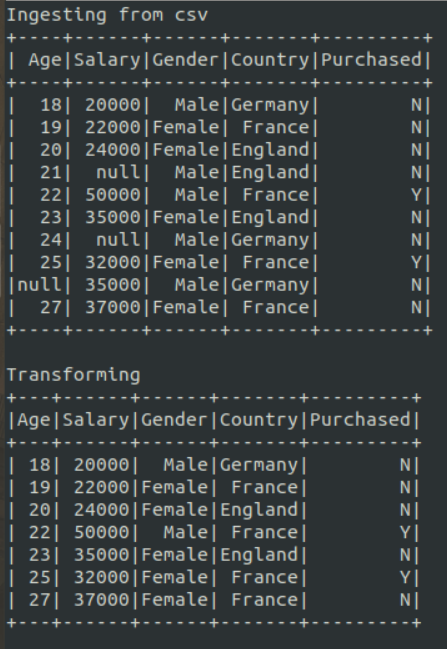
**Yarn logs -applicationID application(id)**

**Then run unset file**

**And rerun command again**

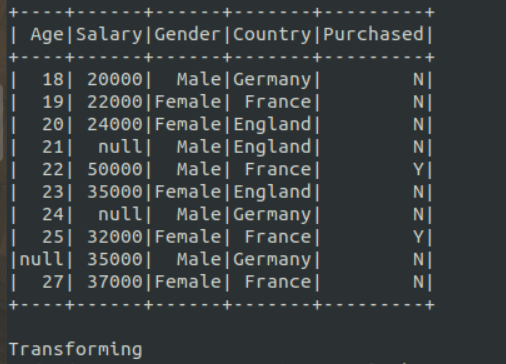


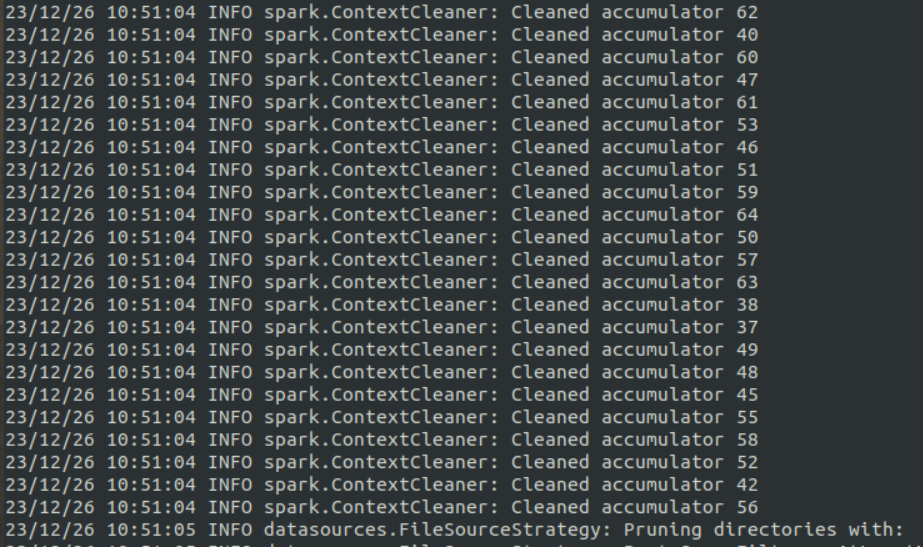
**find ~ -type f -name unset.sh**

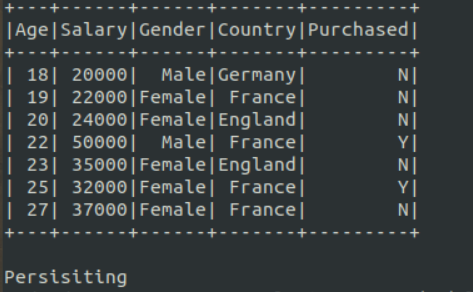


**Run client mode command**

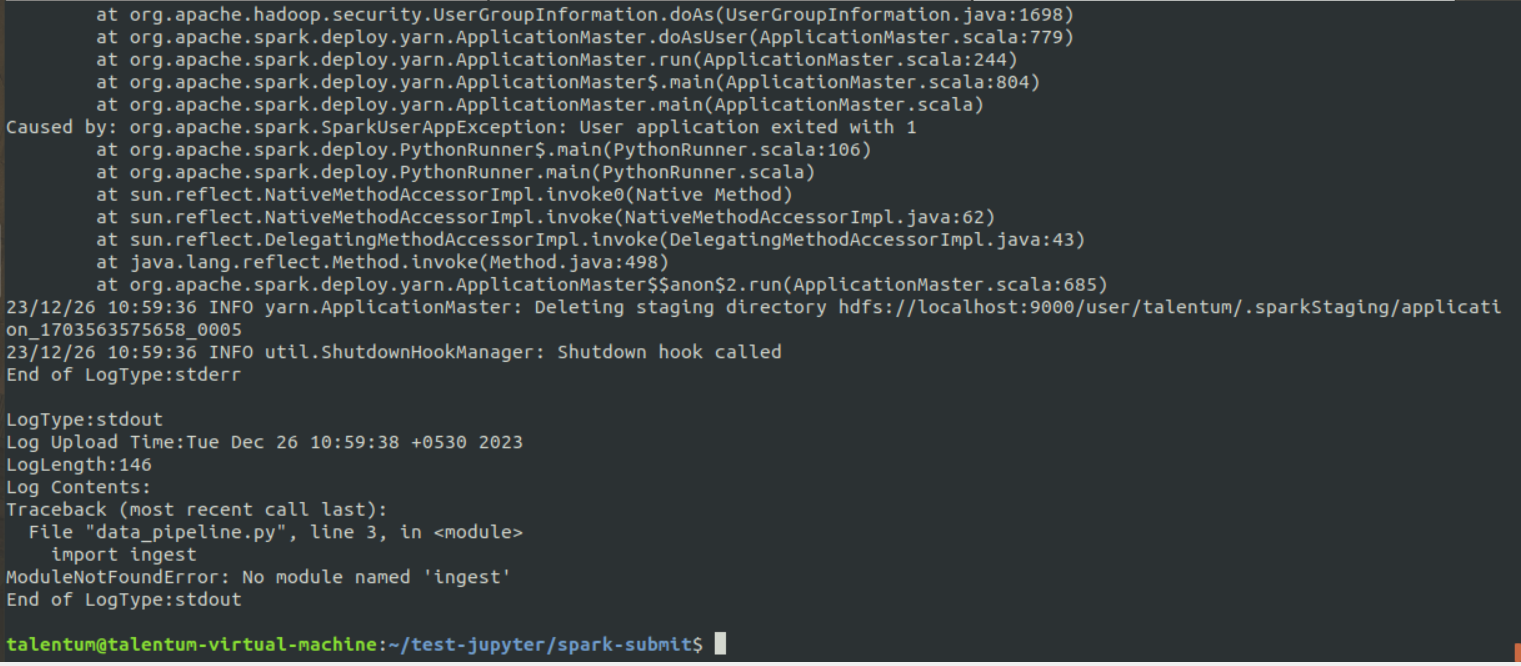
**spark-submit --master yarn --deploy-mode client --driver-memory 2g --num-executors 2 --executor-memory 1g data\_pipeline.py**



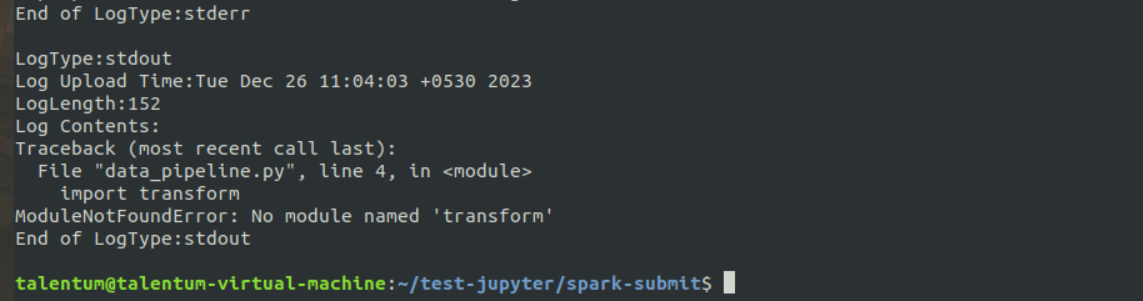




**spark-submit --master yarn --deploy-mode cluster --driver-memory 2g --num-executors 2 --executor-memory 1g data\_pipeline.py**



**spark-submit --master yarn --deploy-mode cluster --py-files ingest.py --driver-memory 2g --num-executors 2 --executor-memory 1g data\_pipeline.py**



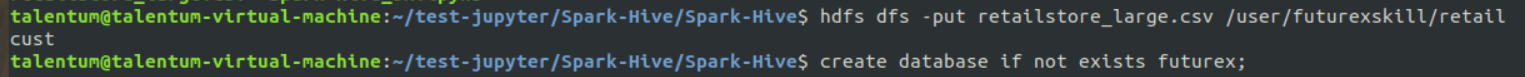
Spark Integration with Hive

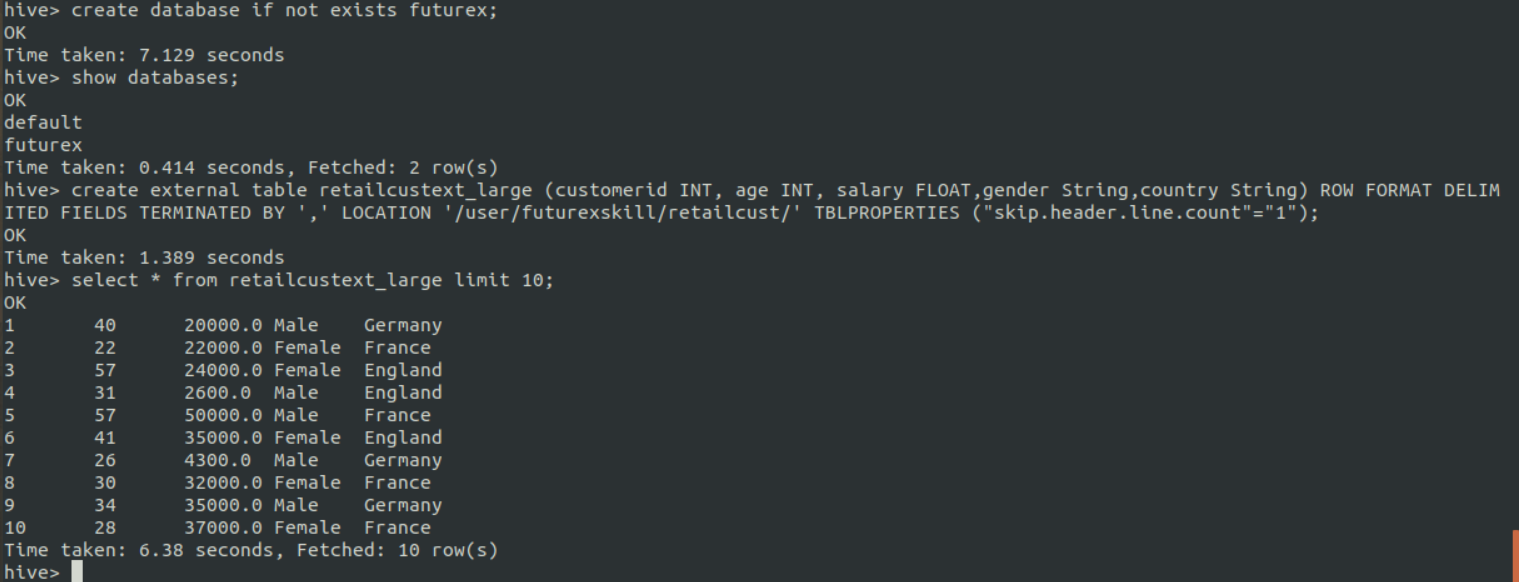
1. Download the attached file in STAGING\_AREA
2. Pull the file in LABS\_AREA/test-jupyter/spark-hive

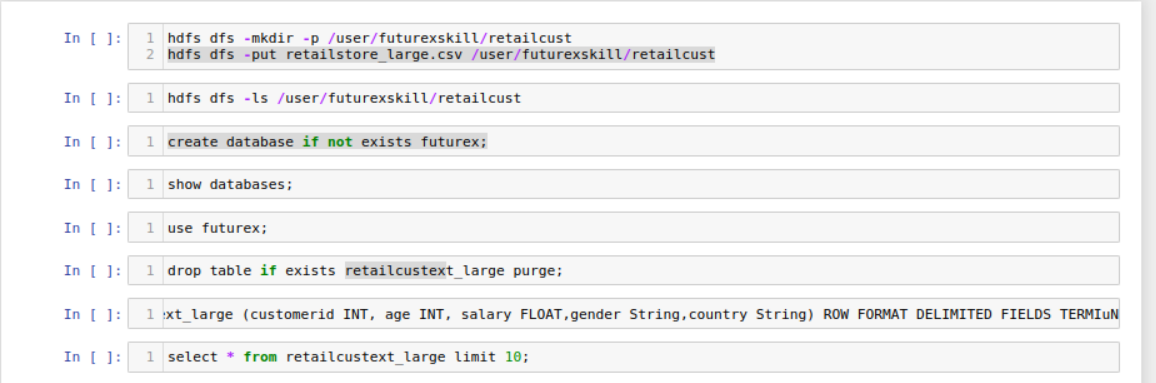
* Code/Dataset

data/Spark-Hive.zip

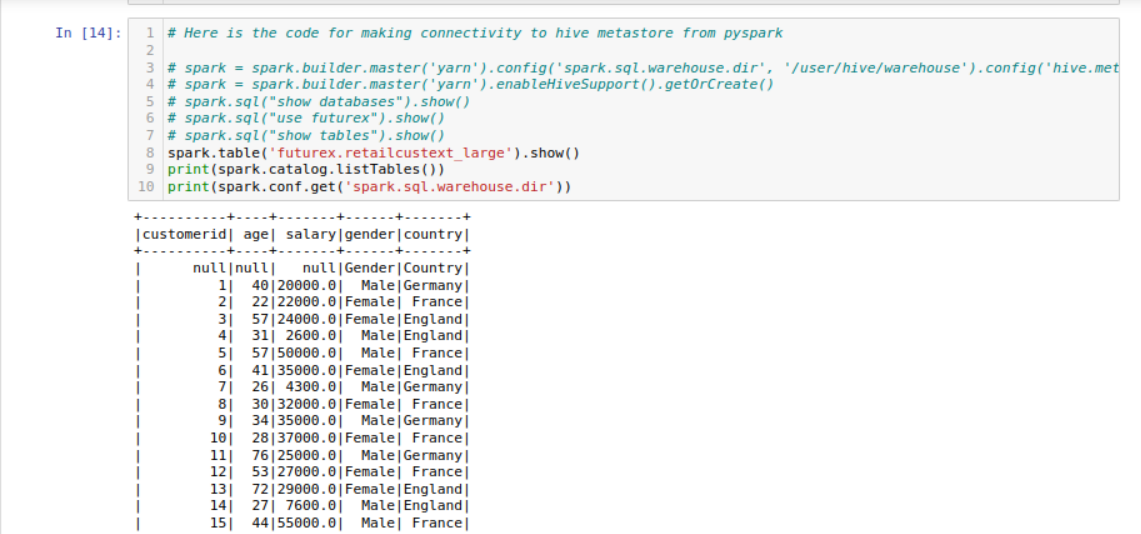


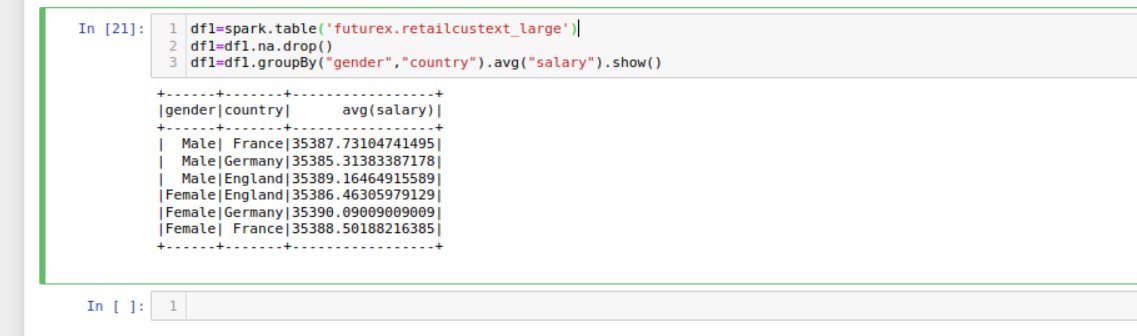


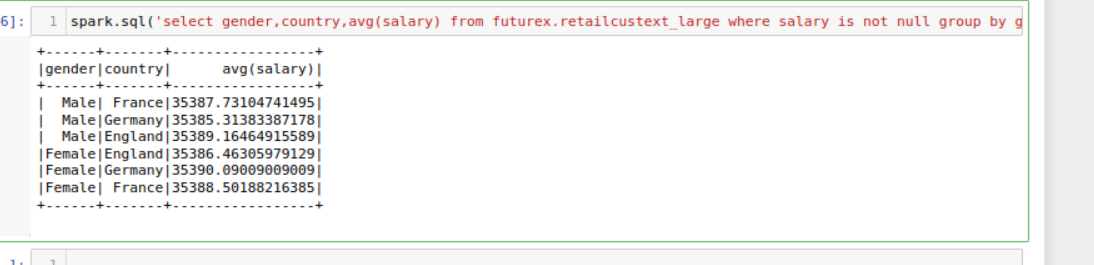






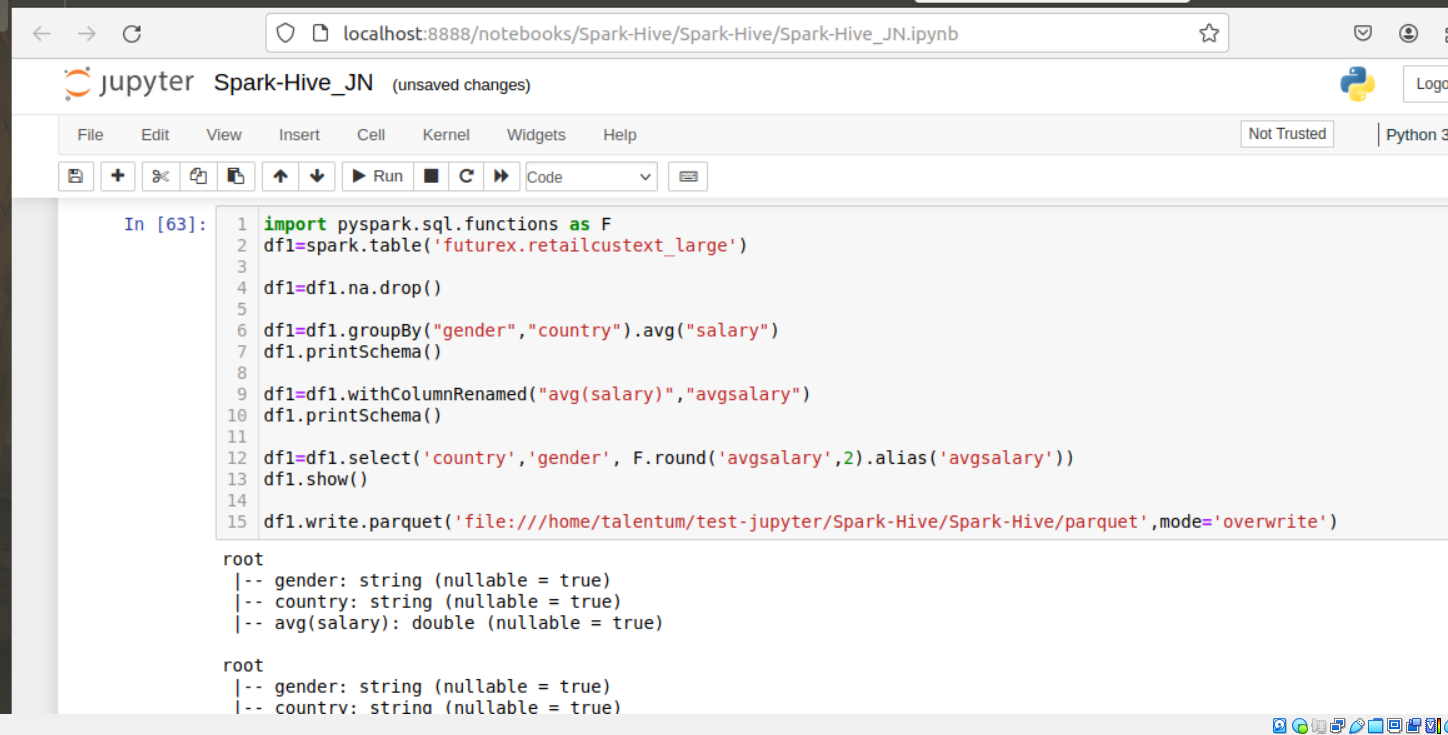


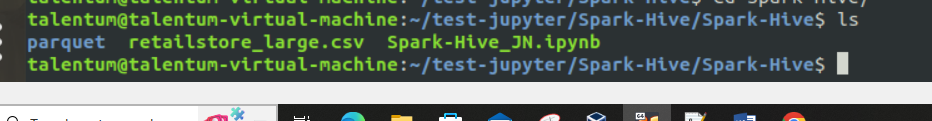


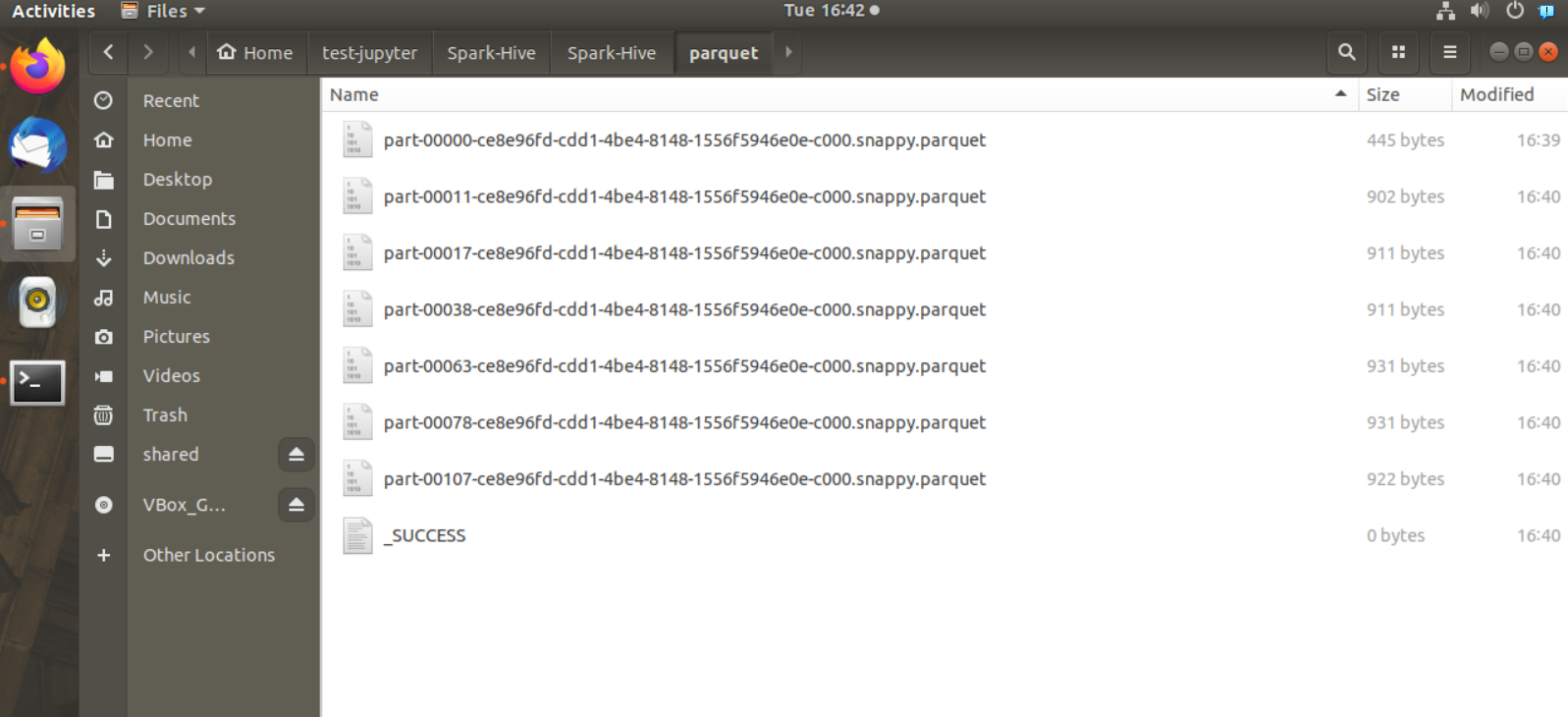




**Creating parquet file**





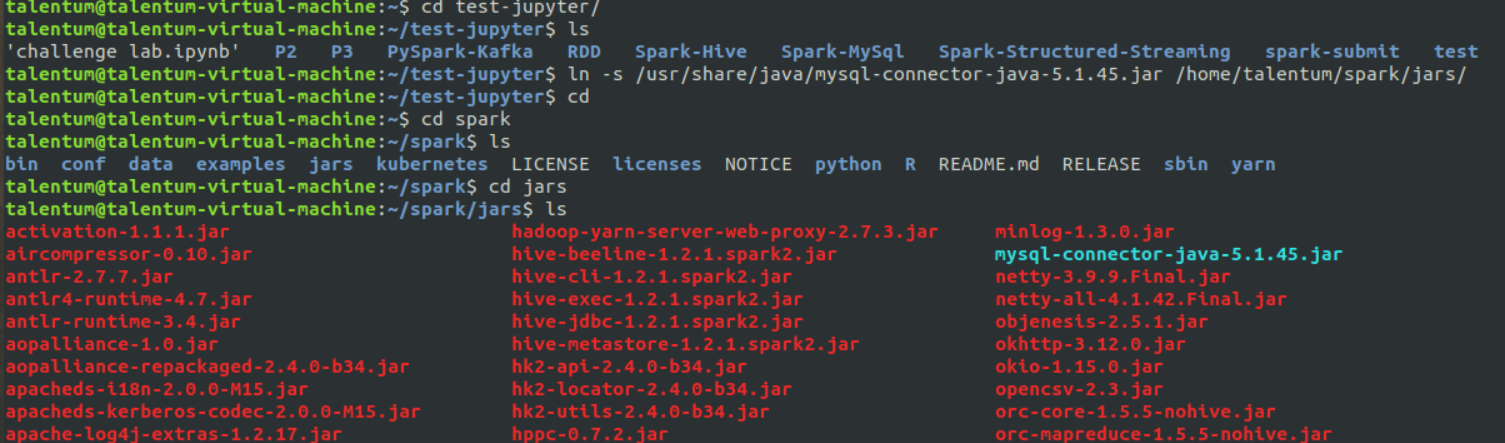


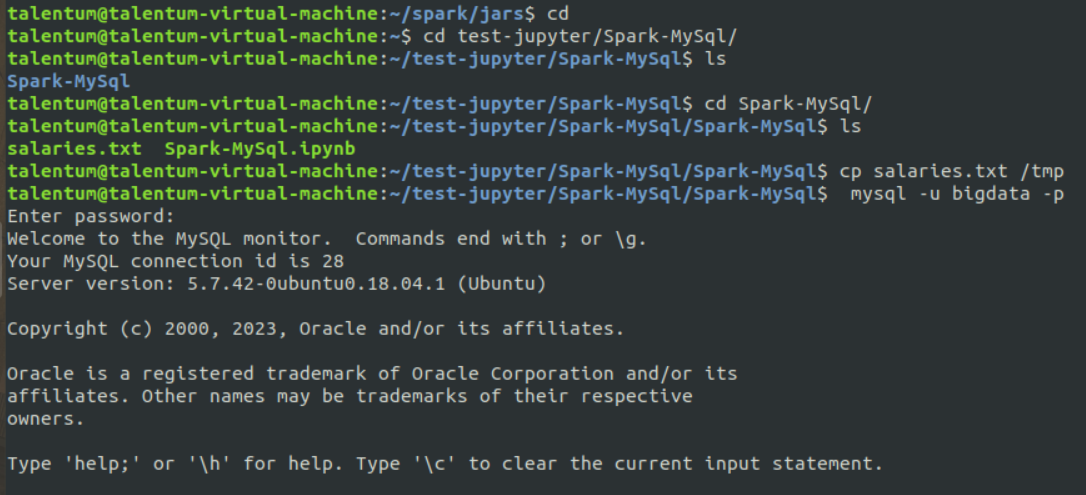
Spark JDBC

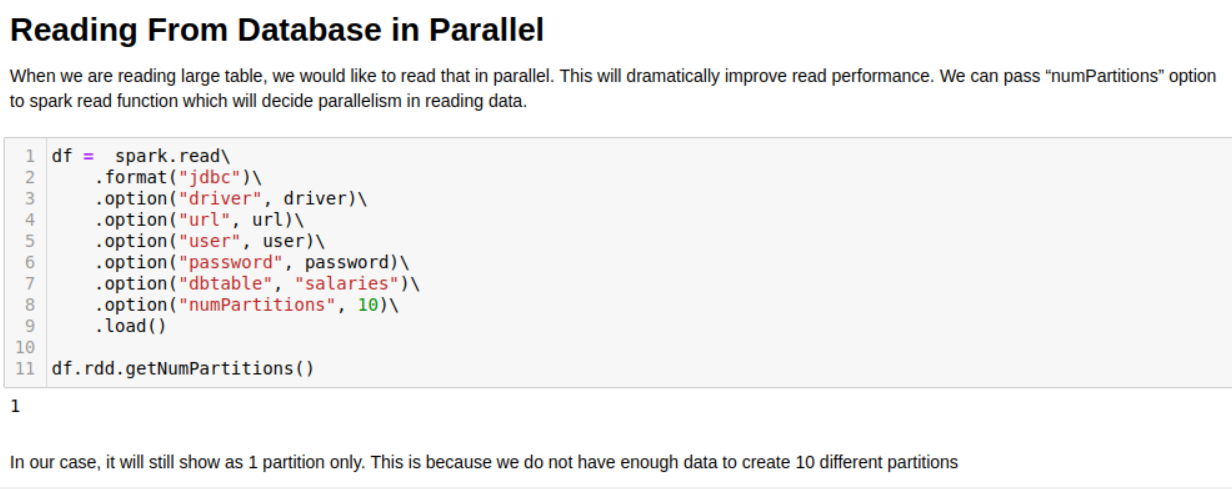
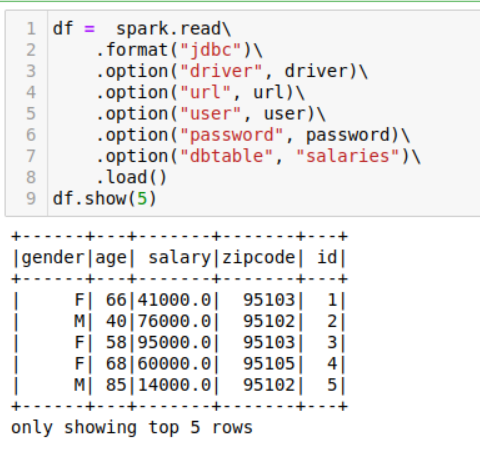
1. Download the attached file in STAGING\_AREA
2. Pull the file in LABS\_AREA/test-jupyter/spark-mysql

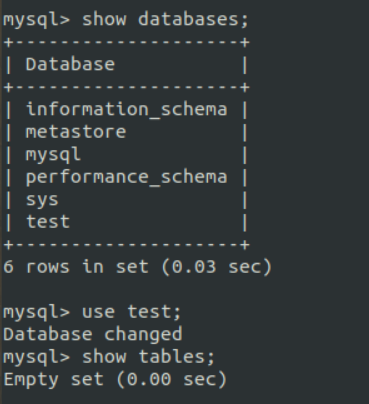
* Code/Dataset

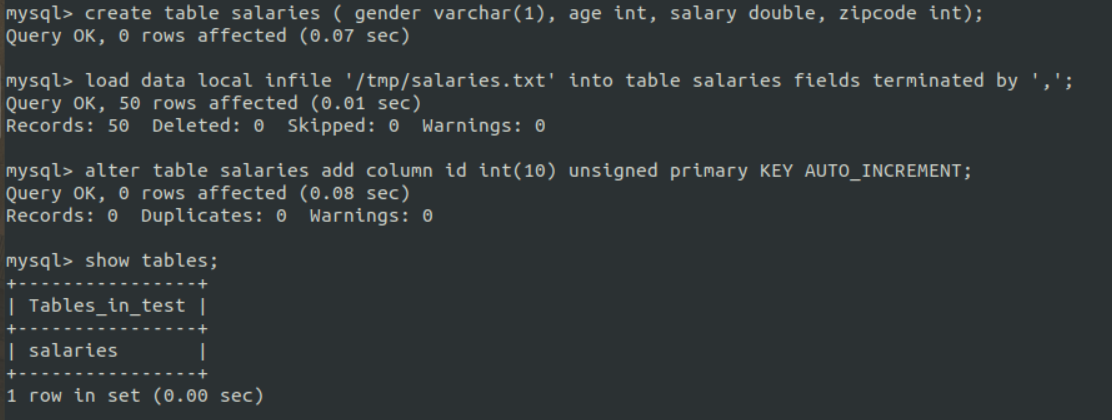
data/Spark-MySql.zip

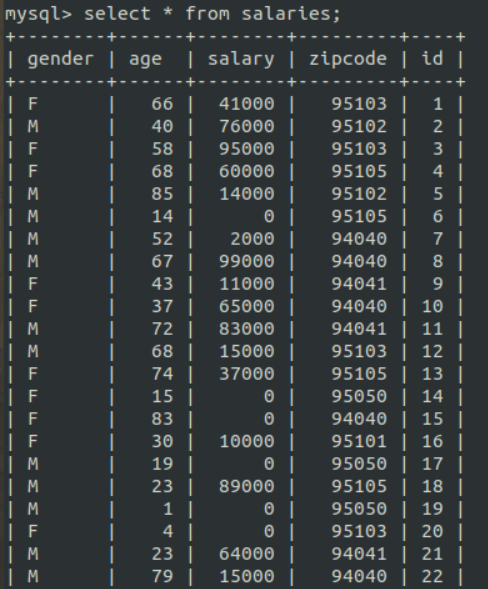












24] PySpark MlLib - Collaborative Filtering

1) Pull the extracted notes in UBUNTU\_HOME/test-jupyter/P2/M4/SM2

2) Import the notes in the Jupyter notebook.

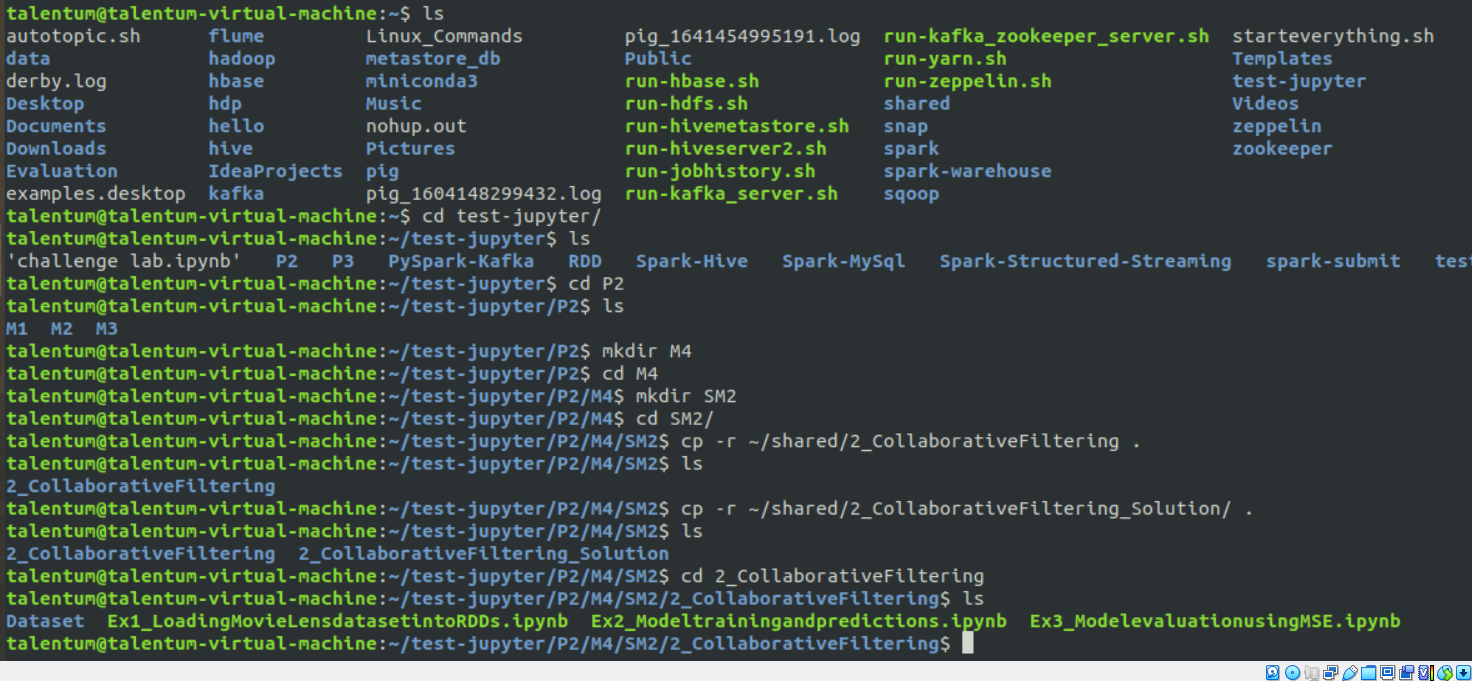
3) Follow the instructions based in the notes

* Code/Dataset

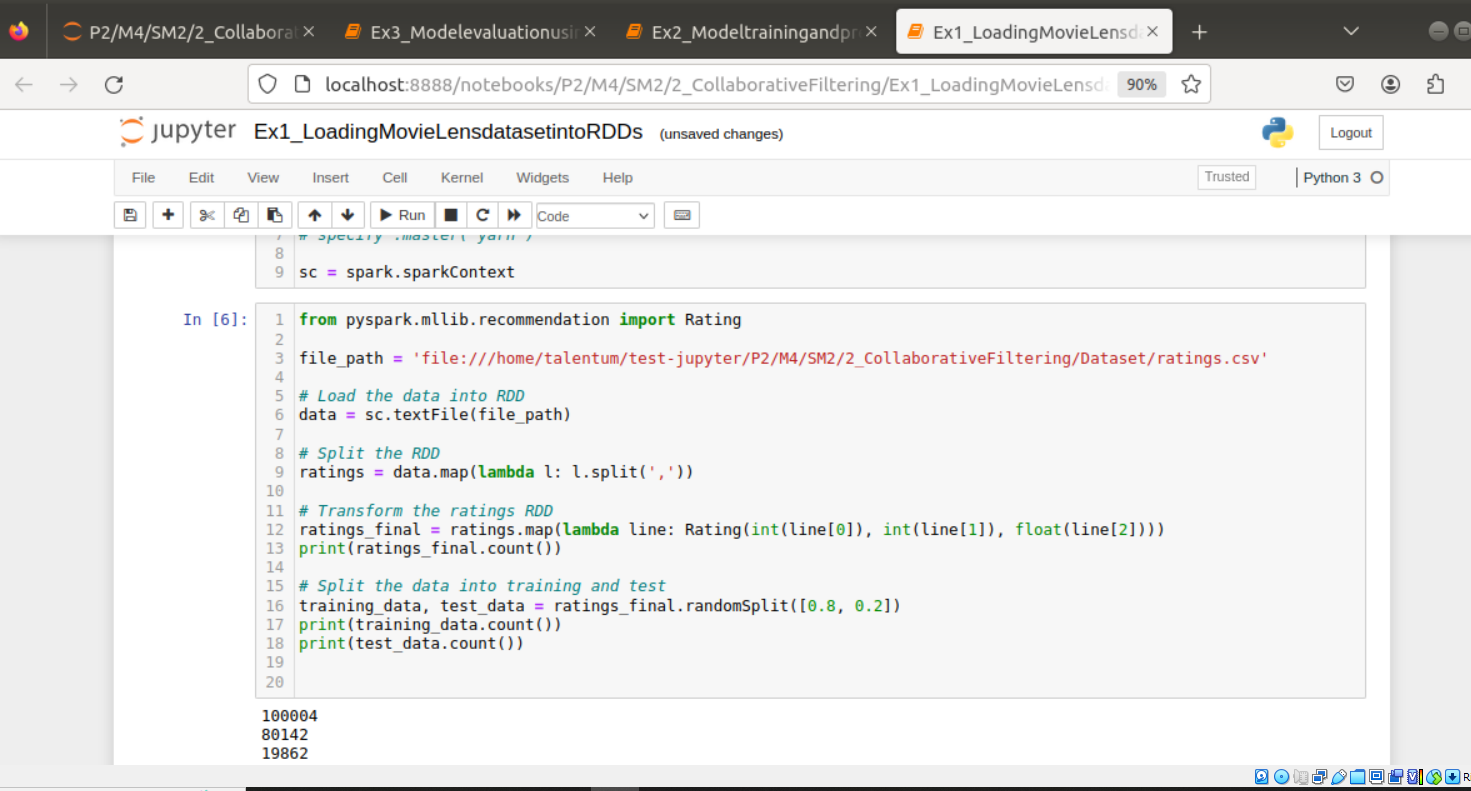
data/2\_CollaborativeFiltering.rar

data/2\_CollaborativeFiltering\_Solution.rar

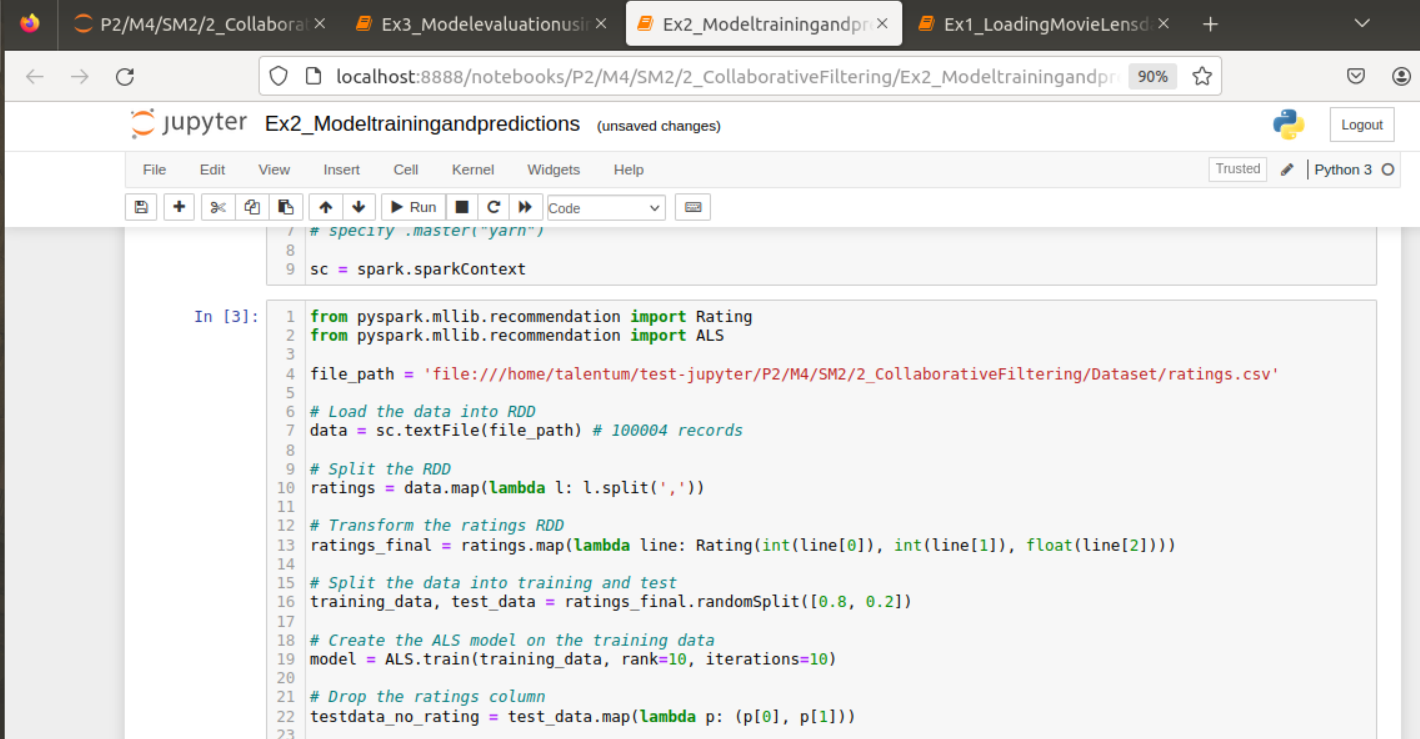
TimeLine = 25 Mins



Pwd : /home/talentum/test-jupyter/P2/M4/SM2/2\_CollaborativeFiltering



2.





3.

