**PYSPARK QUESTIONS**

1. Create a spark job programmatically from command line, creating dataframe on devices.json file and filter out models "iFruit" and "Titanic" and save the result on HDFS in csv format. Submit the job from the command line.
2. Check from the Spark UI number of executors, time taken for the job to complete.
3. Create a join program joining EMP, DEPT, and SALGRADE files using your own logic and fields. After executing the program find out what type of join is used by the spark.
4. Persist the above dataframe., run the same query again the last dataframe, now check from execution plan how spark is executing the same.
5. Now unpersist the above dataframe, convert the dataframe into table and cache the table.
6. Create a dataframe on ratings.csv file from hdfs making sure that Spark creates with single partition.