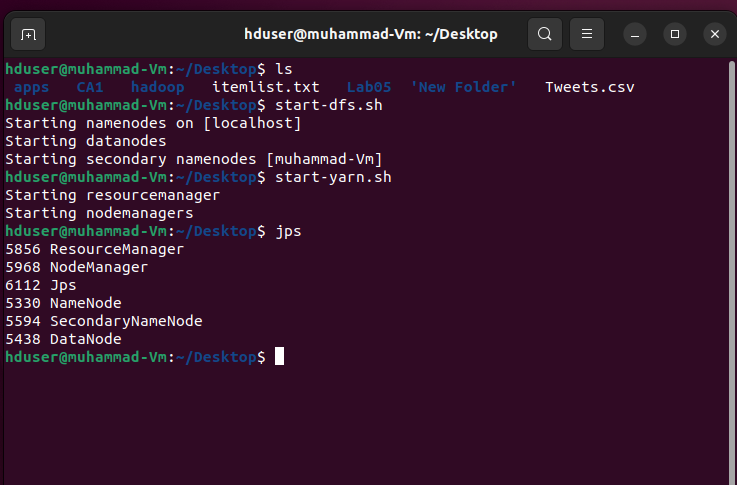
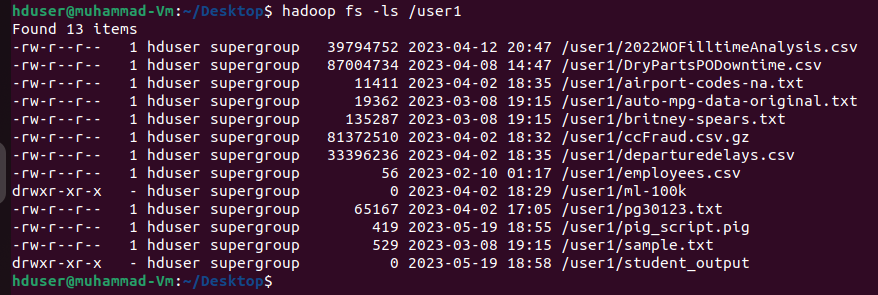
Cdcd Appendix

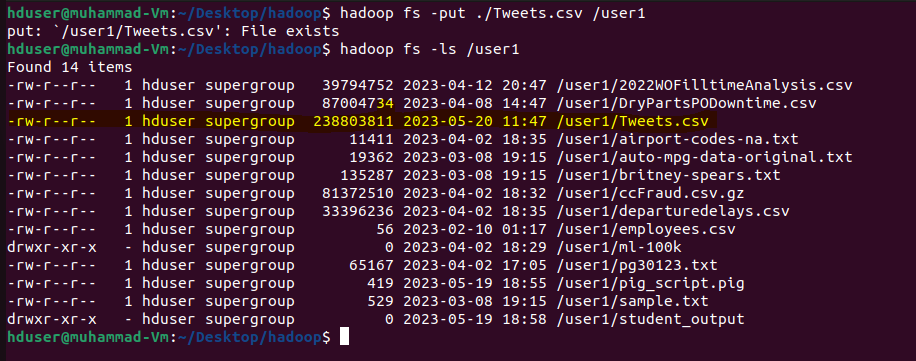
1. Tweets.csv present on the Ubuntu Desktop
2. Start Hadoop

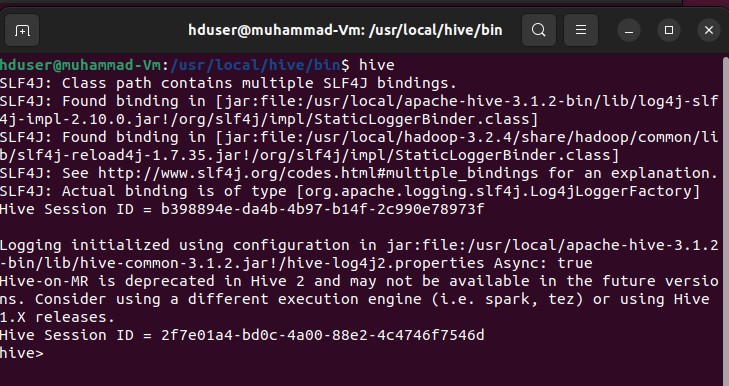


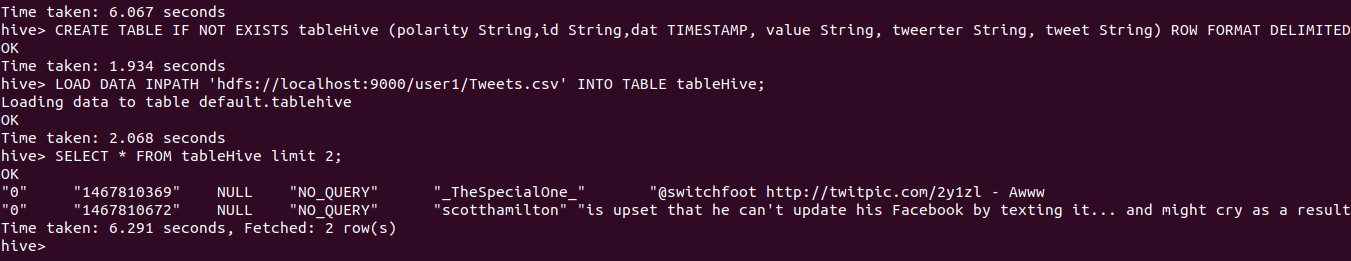
1. Checking what files are on the user1 hadoop system currently



1. Move the Tweets.csv file from the Hadoop folder on the desktop to the HDFS system





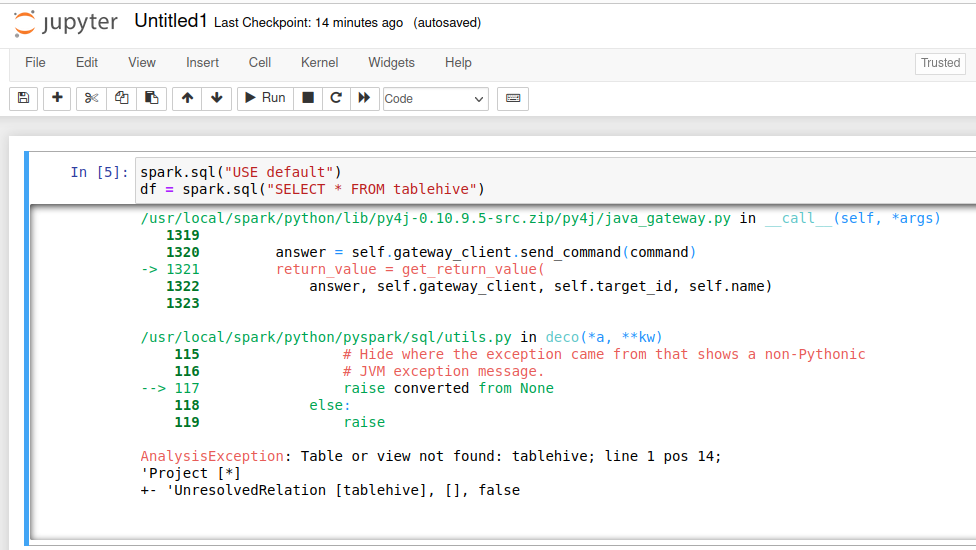


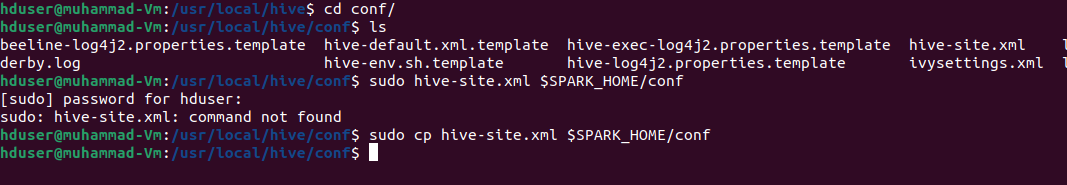
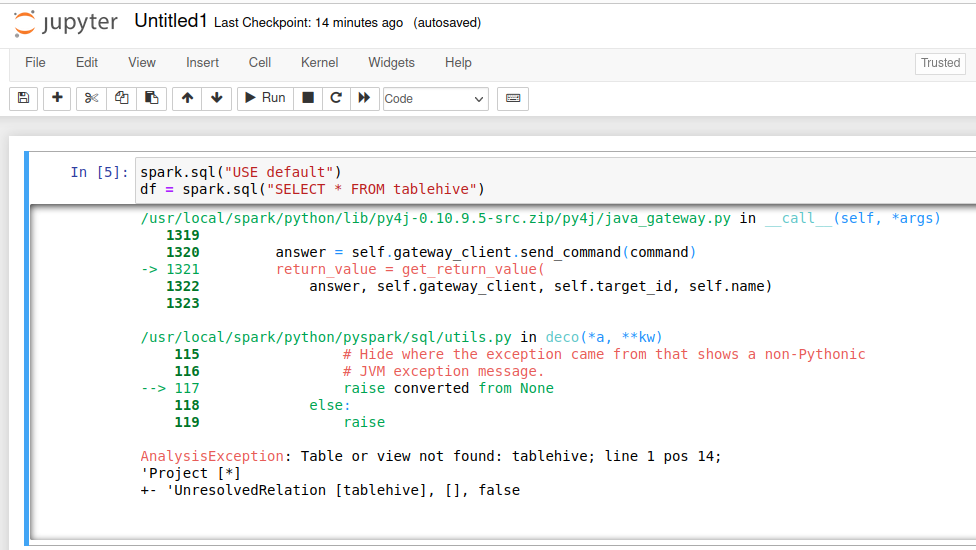
Copying the hive-site.xml from the hive/conf folder to the spark configuration folder:

Ref :https://www.projectpro.io/recipes/read-table-of-data-from-hive-database-pyspark

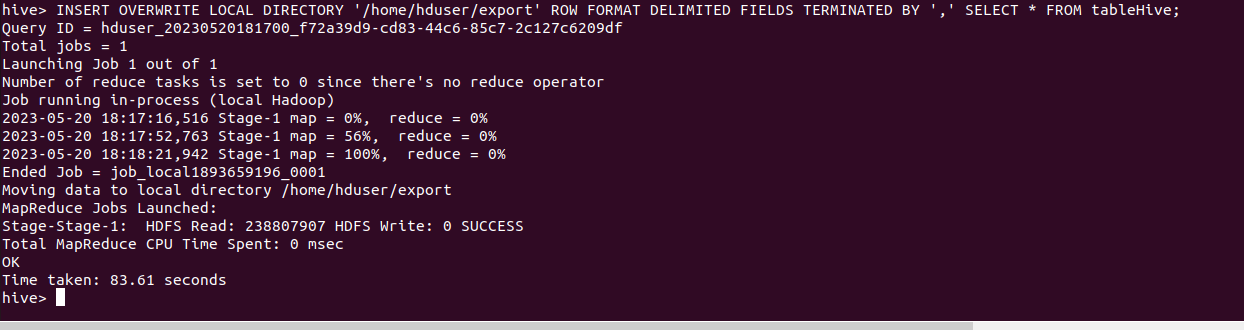
<https://phoenixnap.com/kb/hive-create-external-table>

Trying performing processing using pyspark by calling the table from hive :





Export it to Ubuntu local drive directory



Created a new csv file in pyspark containing the filered Tweets and saved it into Hadoop

