// Steps for Code execution in Lab environment //

1. Open Terminal

1.1. Copy the JAR file to Spark Jars using the following command.

 # docker cp /home/labuser/Project/Installation_and_Execution/Docker_Hive_Spark_Dependencies/Docker_Hive_Spark_Dependencies/postgresql-42.3.1.jar hdp_spark-master:/spark/jars/

1.2. Copy hive-site.xml file from Hive container to Spark container

source path - hdp_hive-server:/opt/hive/conf/hive-site.xml
target path - hdp_spark-master:/spark/conf/

 # docker cp hdp_hive-server:/opt/hive/conf/hive-site.xml /home/labuser/Downloads/
 # docker cp /home/labuser/Downloads/hive-site.xml hdp_spark-master:/spark/conf/

2. Copy .sql files to ra_mysql.

 # docker cp /home/labuser/Project/Code ra_mysql:/opt/

3. Login to MySQL (ra_mysql) docker container.

 # docker exec -i -t ra_mysql bash

3. Open a mysql shell.

 # mysql -u root -p
Enter the password- **example**

4. Execute the partial dataset creation file step by step while referring to the project videos.

5. Execute the views creation file step by step while referring to the project videos.

6. Open another terminal and login to SQOOP (ra_sqoop) docker container.

 # docker exec -i -t ra_sqoop bash

7. Execute SQOOP import commands provided in the **03_Sqoop-import.txt** file present at /home/labuser/Project/Installation_and_Execution/ location in the local file system of lab environment.

8. Exit the SQOOP container.

 # exit

9. Login to Hive Docker container (hdp_hive-server).

 # docker exec -i -t hdp_hive-server bash

10. Open a hive shell and execute the **04_Hive_tables_creation_cust_sales_stores.hql** step by step while referring to project videos..

 # hive

11. Login to Spark docker container (hdp_spark-master) and open a spark shell.

 # docker exec -i -t hdp_spark-master bash
 # ./spark/bin/spark-shell

12. Execute the **05_customer_demographic.scala** file in spark-shell which is stored in the Project/Code/ location on desktop.

13. Copy the parquet files from spark container(/spark/customer_demographics_xml_mined/) to local file system using docker cp command. Make sure you copy the files to an empty directory in the local file system.

 # docker cp /spark/customer_demographics_xml_mined/ /home/labuser/Downloads/

14. Copy the parquet files from the local files system to the Hive container. (/opt location)

 # docker cp /home/labuser/Downloads/ hdp_hive-server:/opt

15. Login to Hive container and open a hive shell.

 # docker exec -i -t hdp_hive-server bash
 # hive

16. Execute the **customer_demographic_creation.hql** step by step & please refer the project videos for in-depth understanding while execution.