

Steps for Lab Execution

1. Once the lab is spun up, Open a terminal window.

2. Copy all the Project files to the Hive docker container.

```
# docker cp /home/labuser/Project/ hdp_hive-server:/opt/Project
```

3. Copy the JAR file to the Hive docker container.

```
# docker cp /home/labuser/Project/Data/hive-json-serde.jar  
hdp_hive-server:/opt/
```

4. Log in to the hive docker container.

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

5. Create a Project folder on HDFS.

```
# hadoop fs -mkdir /home  
# hadoop fs -mkdir /home/hadoop
```

6. Copy the Project files to HDFS location.

```
# hadoop fs -put /opt/Project/* /home/  
# hadoop fs -put /opt/Project/Data/Yelp_Data_Files/* /home/hadoop/
```

7. Open a hive shell by running the following command in the Hive docker container.

```
# hive
```

8. Refer the Project videos and Yelp_Hive_Queries.hql (Provided in the Lab environment under Project folder) file for further execution. (The file in Code.zip differs from the one provided in the Lab)

Note - Please note the difference in execution presented in videos and labs. Instead of AWS-EMR we are using docker containers. So please follow below steps for getting the environment ready for this project.

Used Docker commands:

1. docker ps - Used to list all the running containers.
2. docker exec - Used to access the running containers.
3. docker cp - Used to copy the files to & from docker containers.