Big Data Analytics | Lab 7

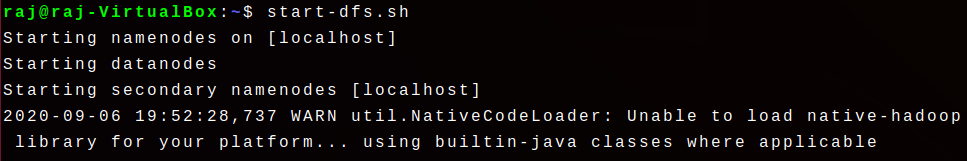
Explore Hive: - Data Warehouse System:



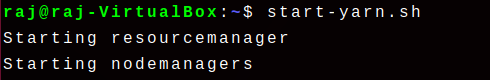
Raj Panchal  
17CEUBG104

**LAB SETUP:**

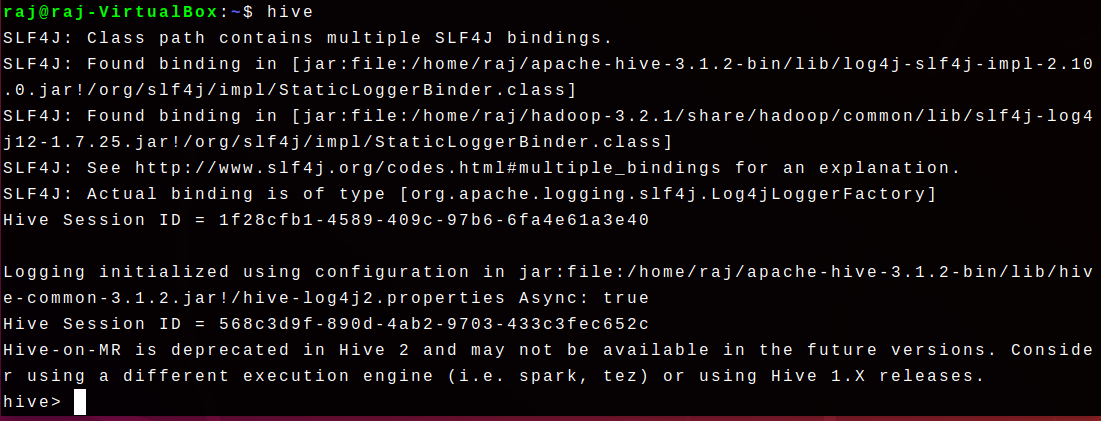
**$ start-dfs.sh** - To start name node, data node, secondary name node.



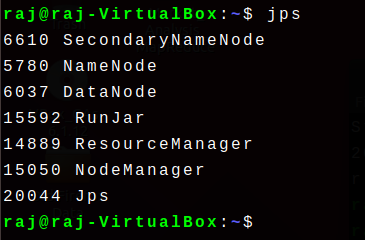
**$ start-yarn.sh** - To start resource manager and node managers.



**$ hive** - To start hive.



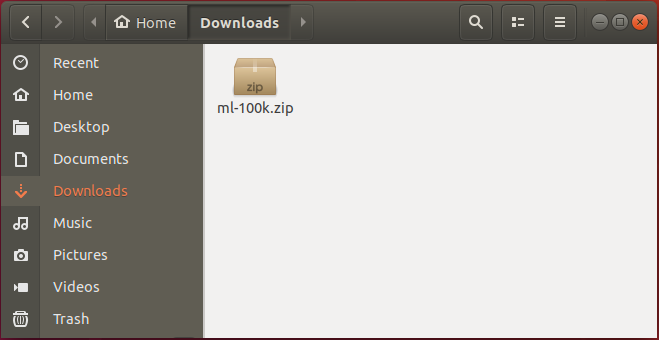
**$ jps** - To check java process status for running Hadoop components.



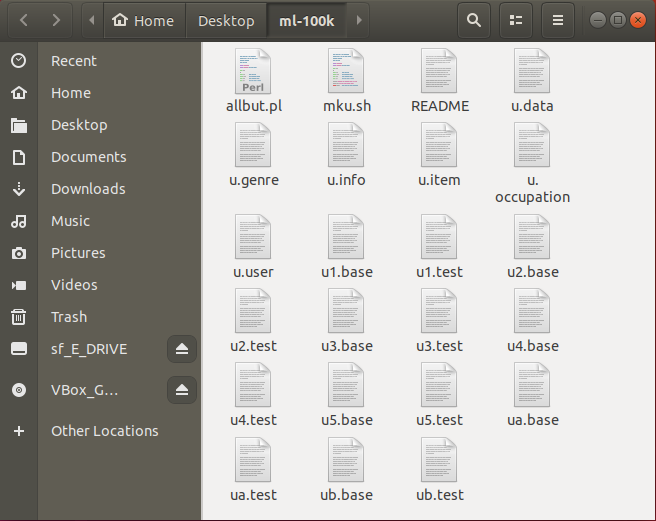
**\*\*RunJar named process is representing the Hive.**

**Task 1: Retrieve the data set. (** [**http://files.grouplens.org/datasets/movielens/ml-100k.zip**](http://files.grouplens.org/datasets/movielens/ml-100k.zip) **)**

Step 1: Download the dataset from the given link in task.



Step 2: Extract the dataset and confirm presence of the data.



**Task 2: Run below HQL.**

hive (default)> CREATE TABLE u\_data (

userid INT,

movieid INT,

rating INT,

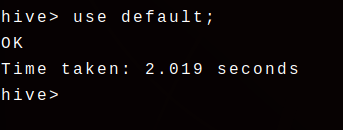
unixtime STRING)

ROW FORMAT DELIMITED

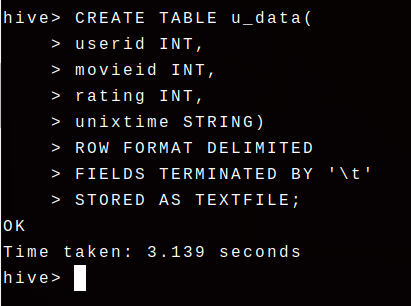
FIELDS TERMINATED BY '\t'

STORED AS TEXTFILE;

Step 1: Set default database as current database.

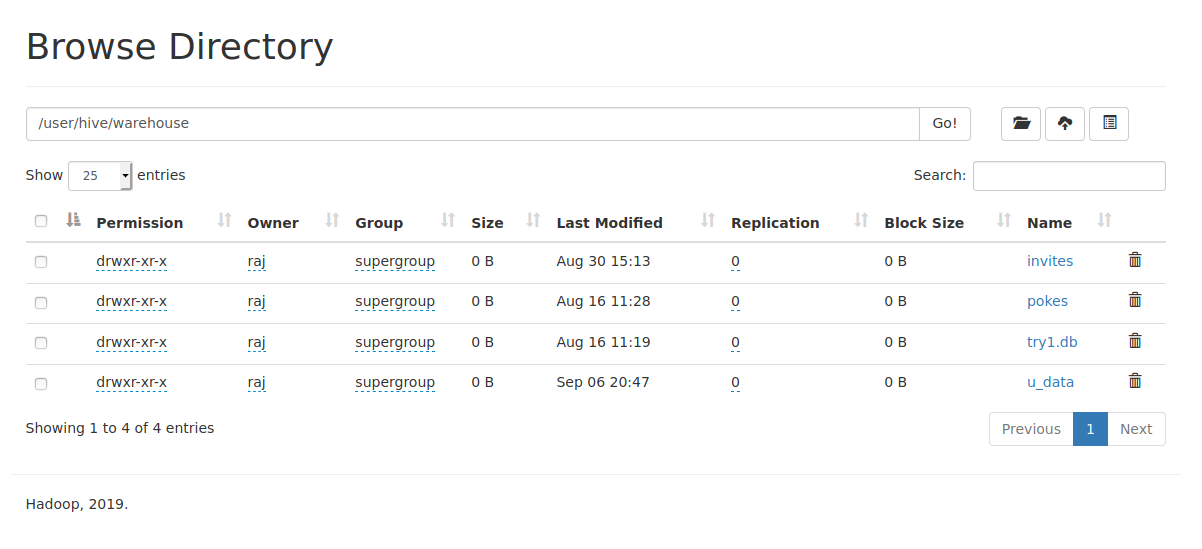
****

Step 2: Write above query and run it.

****

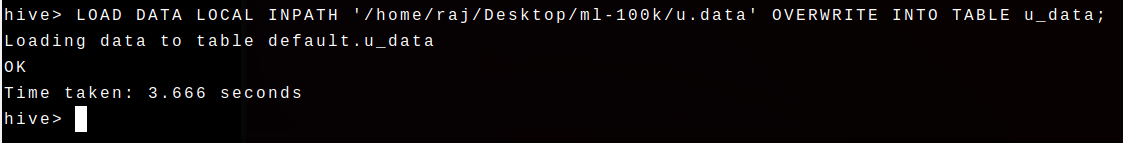
This query creates a new table named u\_data with column userid, movieid, rating and unixtime.

Locate this table using the HDFS GUI interface.

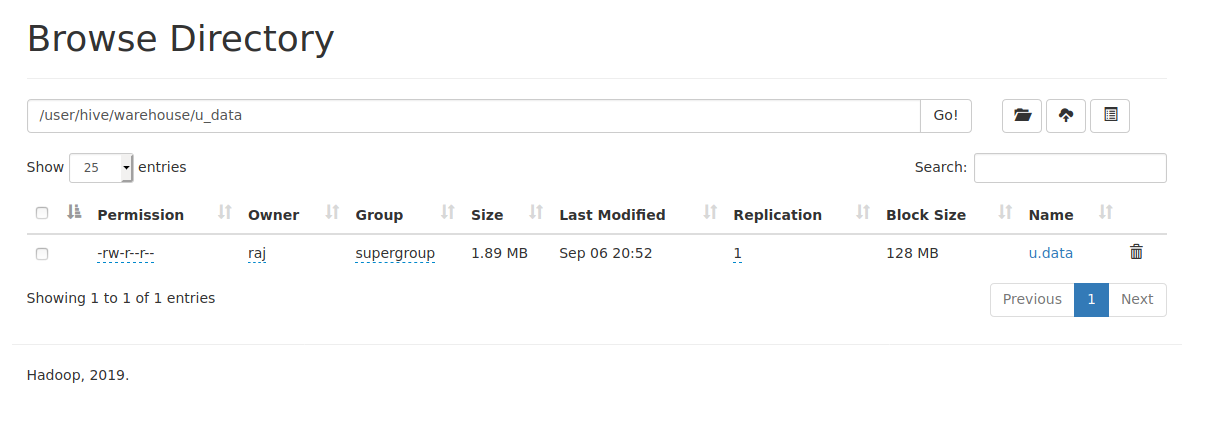
****

**Task 3: Load the data from local file u.data to table u\_data.**

Step 1: Write below command to load the data from local file u.data to table u\_data.

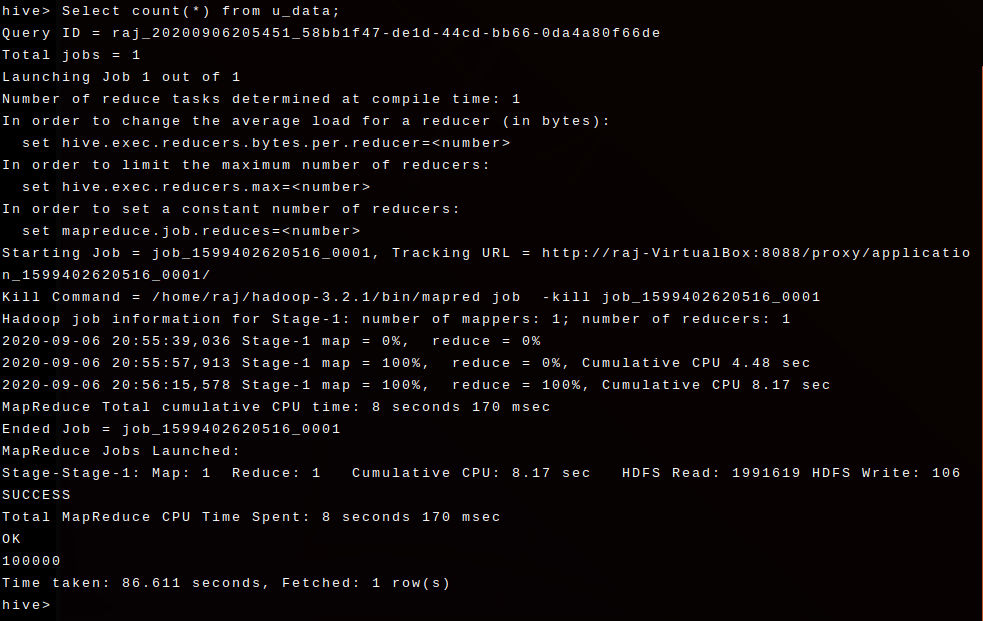
****

Step 2: Confirm presence of file u.data in warehouse using the HDFS GUI**.**

****

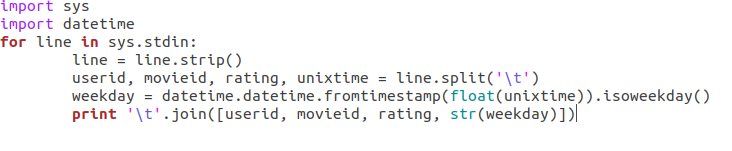
**Task 4: Find total numbers of records in u\_data table**

Step: Run the given query to retrieve numbers of records in u\_data table.

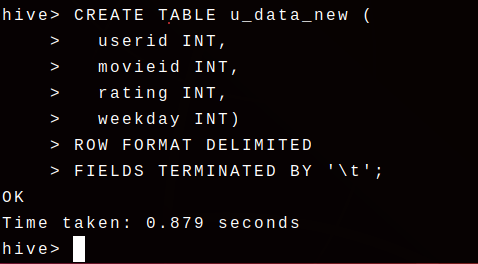
****

**Task 5: Find weekday and corresponding count group by weekday**

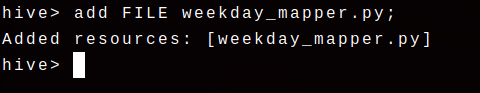
Step 1: Create weekday\_mapper.py.

****

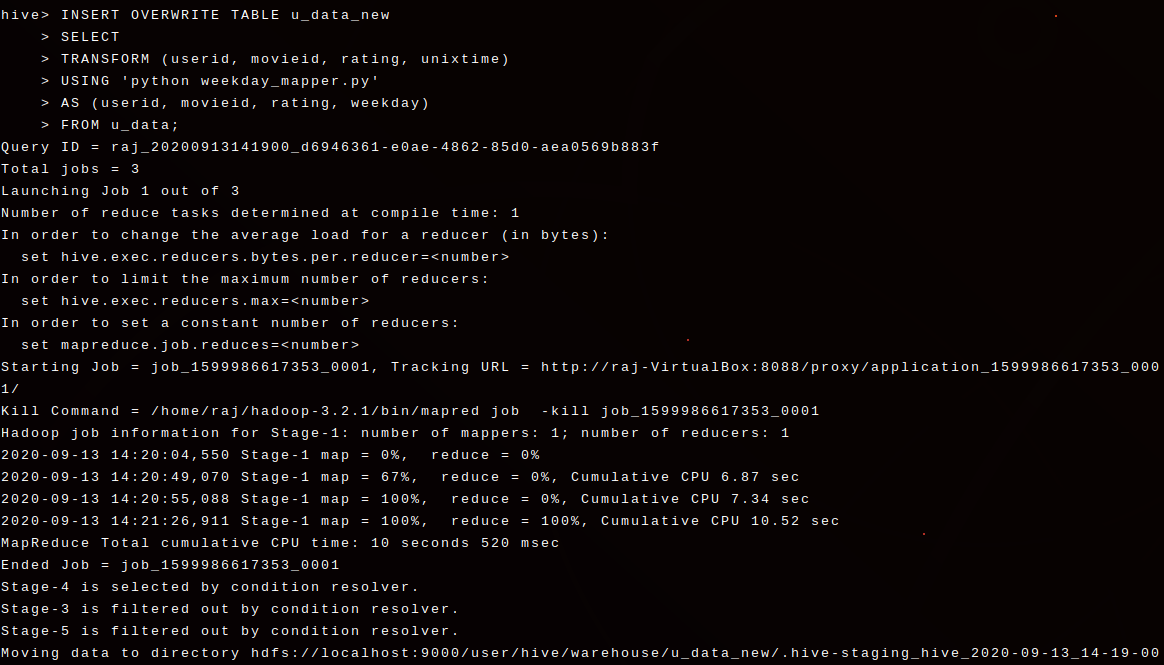
Step 2: Create table using below query.



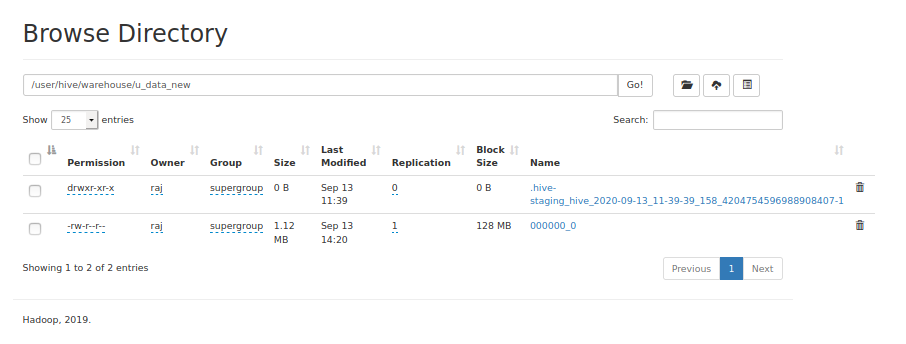
Step 3: Add weekday\_mapper.py to hive using below query.



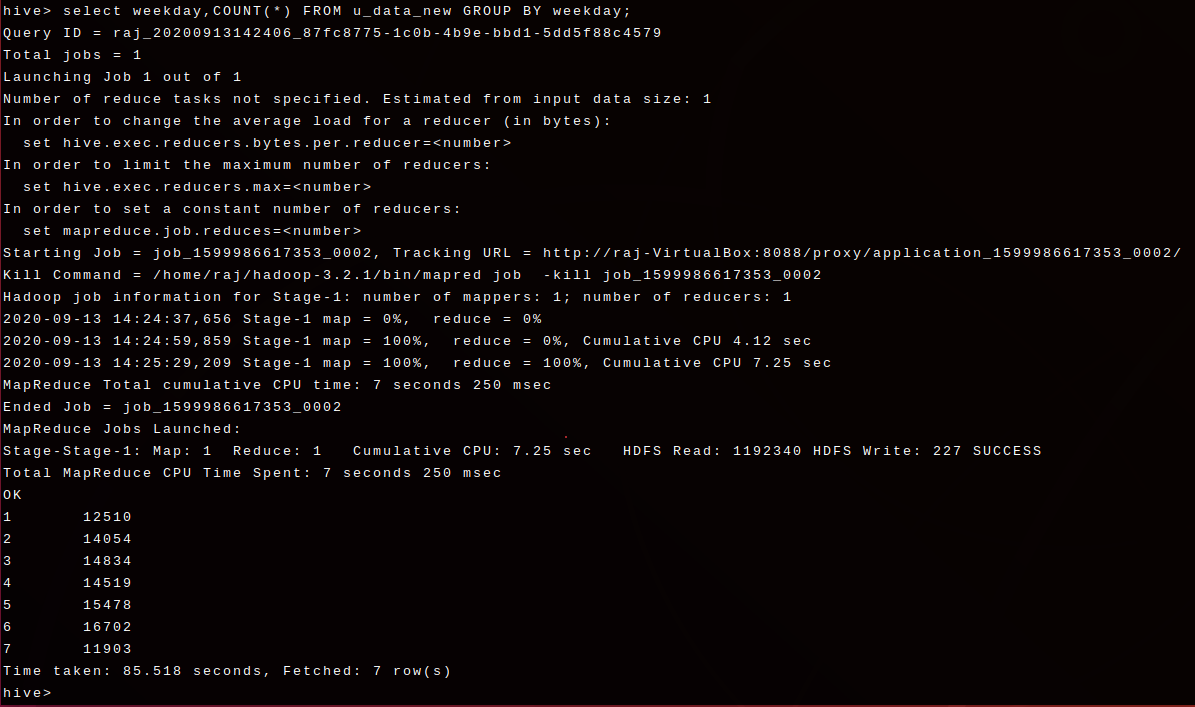
Step 4: Insert data into u\_data\_new table using weekday\_mapper.py.



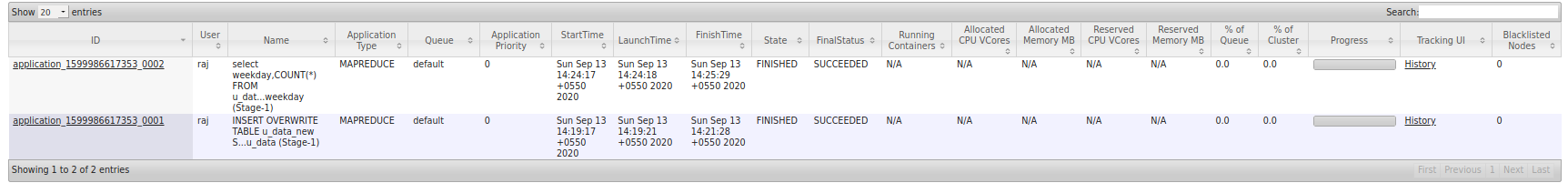
Step 5: Confirm presence of data in u\_data\_new table.



Step 6: Run below query.



You can see the status of the job on localhost:8088.



Document Link : <https://docs.google.com/document/d/1vVgoesyEr7608R9COPGppldhqWVtnYUw0EeNfItaKLk/edit?usp=sharing>