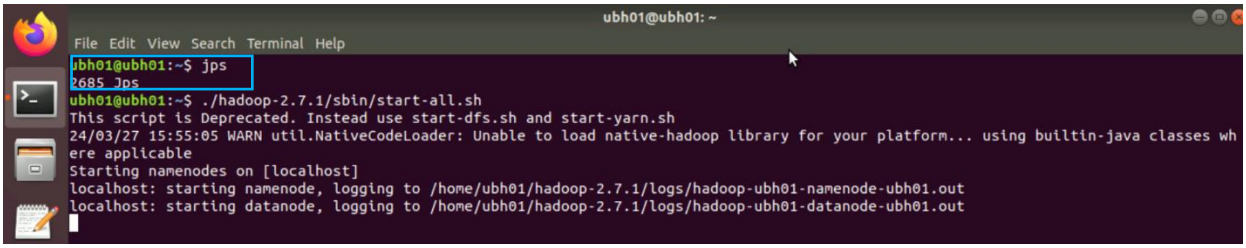
ASSIGNMENT-1

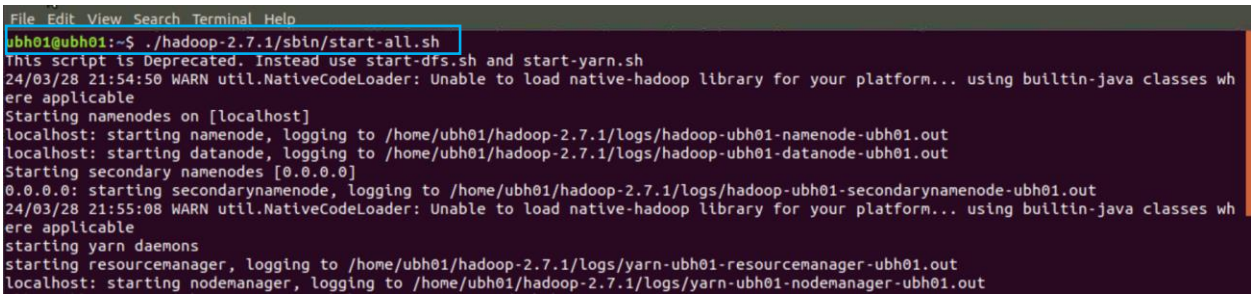
(HIVE)

Employee ID : 2320074

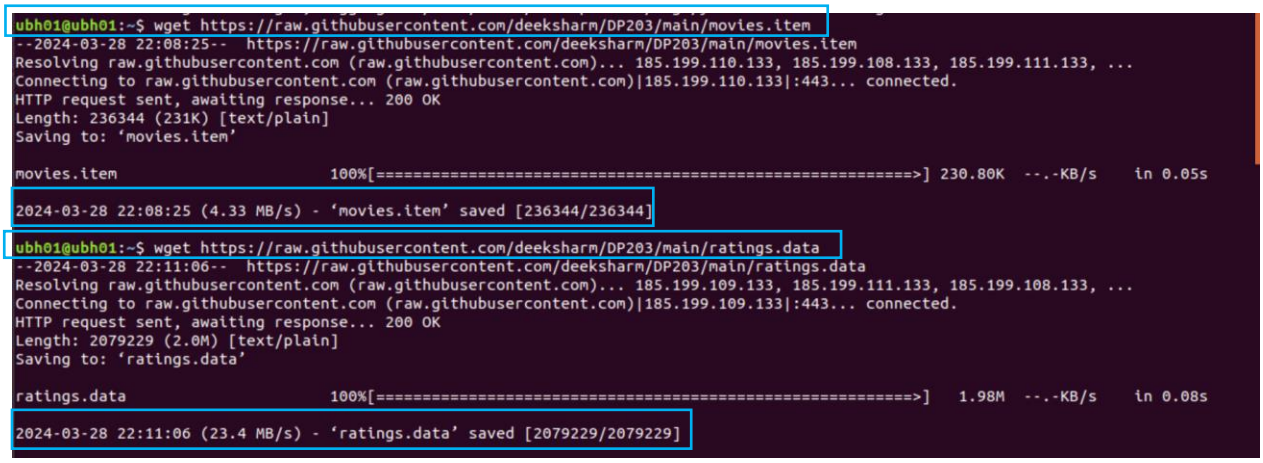
CSDAIA24AZ003



Step 1: Start the session.



Step 2: Start the HDFS in the virtual machine.



Step 3: Now get the files(movies.item, ratings.data) into local system from external source by using wget command.

```
ubh01@ubh01:~$ hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/ubh01/apache-hive-2.3.2-bin/lib/log4j-slf4j-impl-2.6.2.jar!/org/slf4j/impl/StaticLoggerBinder
.class]
SLF4J: Found binding in [jar:file:/home/ubh01/hadoop-2.7.1/share/hadoop/common/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLog
gerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]

Logging initialized using configuration in jar:file:/home/ubh01/apache-hive-2.3.2-bin/lib/hive-common-2.3.2.jar!/hive-log4j2.properti
es Async: true
Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e.
 spark, tez) or using Hive 1.X releases.
```

Step 4: Now start the hive in the virtual machine.

```
hive> create database dharini;
Loading class `com.mysql.jdbc.Driver'. This is deprecated. The new driver class is `com.mysql.cj.jdbc.Driver'. The driver is automati
cally registered via the SPI and manual loading of the driver class is generally unnecessary.
OK
Time taken: 5.644 seconds
hive> show databases;
OK
default
dharini
sumitdb
Time taken: 0.241 seconds, Fetched: 3 row(s)
hive> use dharini;
OK
Time taken: 0.076 seconds
hive> create table movieeee(movie_id int,movie_title varchar(255),release_date date,video_release_date date,IMdB_URL varchar(255),unk
own int,Action int,Adventure int,Animation int,Childrens int,Comedy int,Crime int,Documentary int,Drama int,Fantasy int,Film_Noir int
,Horror int,Musical int,Mystery int,Romance int,Sci_Fi int,Thriller int,War int,Western int)row format delimited fields terminated by
'|' lines terminated by '\n' stored as textfile;
OK
Time taken: 1.336 seconds
hive> create table ratingsss(user_id int,m_id int,item int,ts int)row format delimited fields terminated by '\t' lines terminated by
'\n' stored as textfile;
OK
Time taken: 0.24 seconds
hive>
```

Step 5: Create a database in the hive.

Step 6: check the database that has been created using show databases command.

Step 7: Now, to create tables in the database we have created, use “use <database_name>” command.

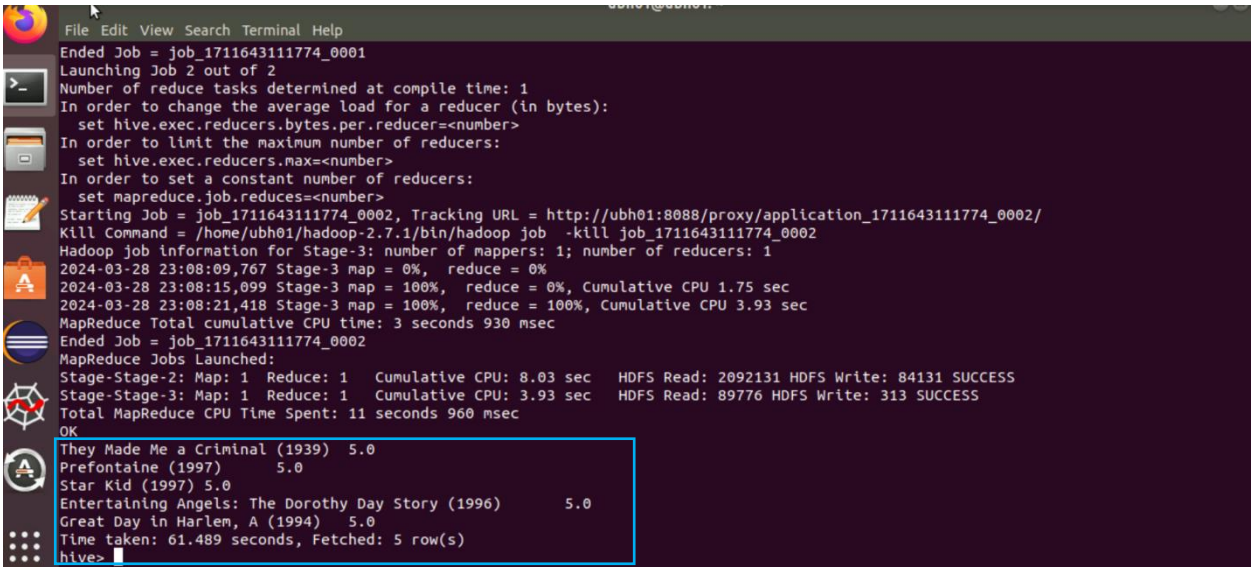
Step 8: Now create the two tables named movieeee and ratingsss in the database we created in hive.

```
hive> load data local inpath 'movies.item' into table movieeee;
Loading data to table dharini.movieeee
OK
Time taken: 1.934 seconds
hive> load data local inpath 'ratings.data' into table ratingsss;
Loading data to table dharini.ratingsss
OK
Time taken: 0.784 seconds
hive>
```

Step 9: Now load the data from files we saved in the local system to the tables we created in the hive.

```
hive> select m.movie_title,avg(r.item) as average_rating from movieeee as m join ratingsss as r on m.movie_id=r.m_id group by m.movie
_id,m.movie_title order by average_rating desc limit 5;
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution eng
ine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = ubh01_20240328230721_000cdaa9-a2f4-4f93-8723-9dbd6e42ff2f
Total jobs = 2
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/ubh01/apache-hive-2.3.2-bin/lib/log4j-slf4j-impl-2.6.2.jar!/org/slf4j/impl/StaticLoggerBinder
.class]
SLF4J: Found binding in [jar:file:/home/ubh01/hadoop-2.7.1/share/hadoop/common/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLog
gerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
2024-03-28 23:07:31     Starting to launch local task to process map join;      maximum memory = 477626368
2024-03-28 23:07:33     Dump the side-table for tag: 0 with group count: 1682 into file: file:/tmp/ubh01/7abc9e9c-b1ce-4d84-a30f-cf49
1df5130b/hive_2024-03-28_23-07-21_105_7425958013031461092-1/-local-10006/HashTable-Stage-2/MapJoin-mapfile00--.hashtable
2024-03-28 23:07:33     Uploaded 1 File to: file:/tmp/ubh01/7abc9e9c-b1ce-4d84-a30f-cf491df5130b/hive_2024-03-28_23-07-21_105_7425958
```

Step 10: Now write a query to fetch the 5 records having top ratings for movie.



Step 11: Check the records at the output.