



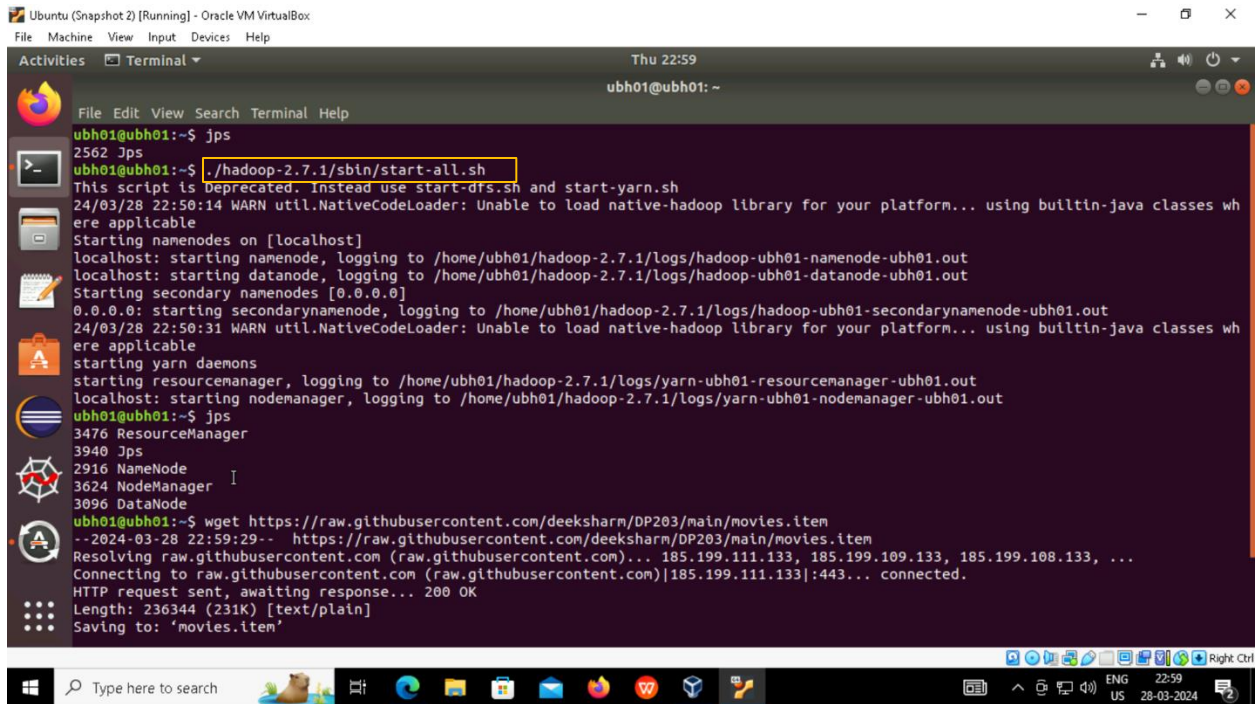
ASSIGNMENT-2 HIVE

Name: Bandi Yuktha reddy

EmpId: 2319933

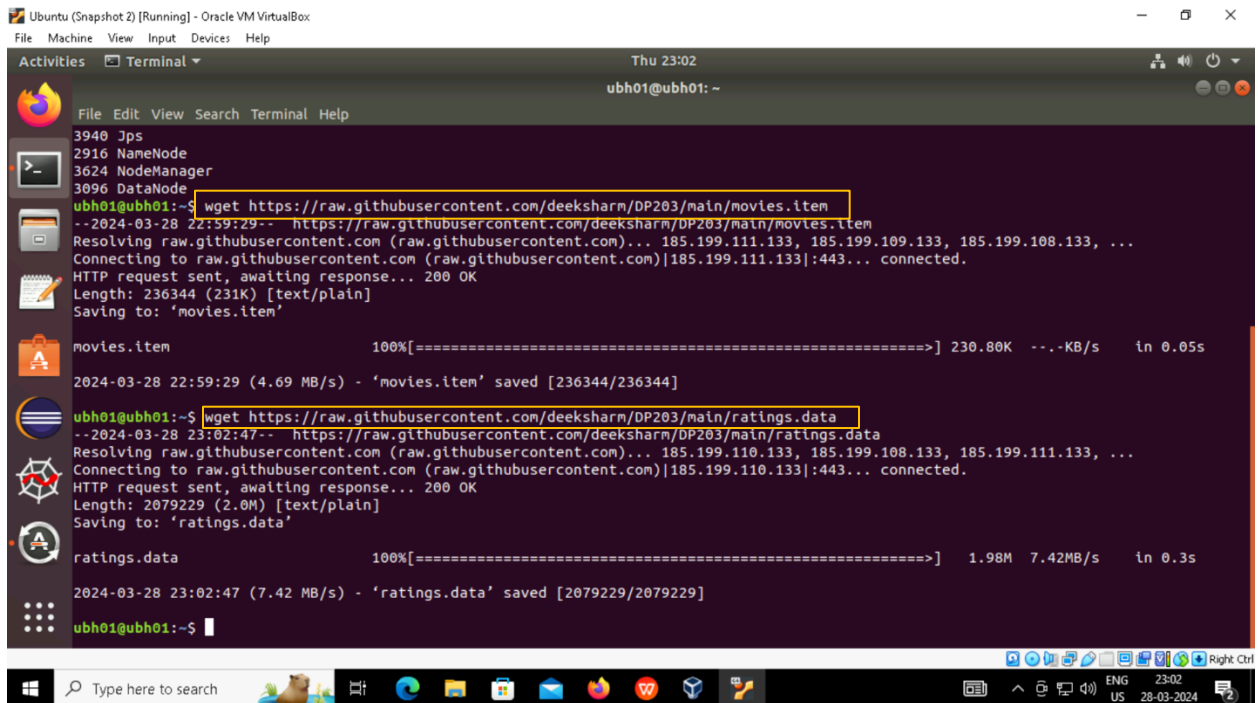
CohertCode: CSDAIA24AZ002

Step 1: Start Hadoop.



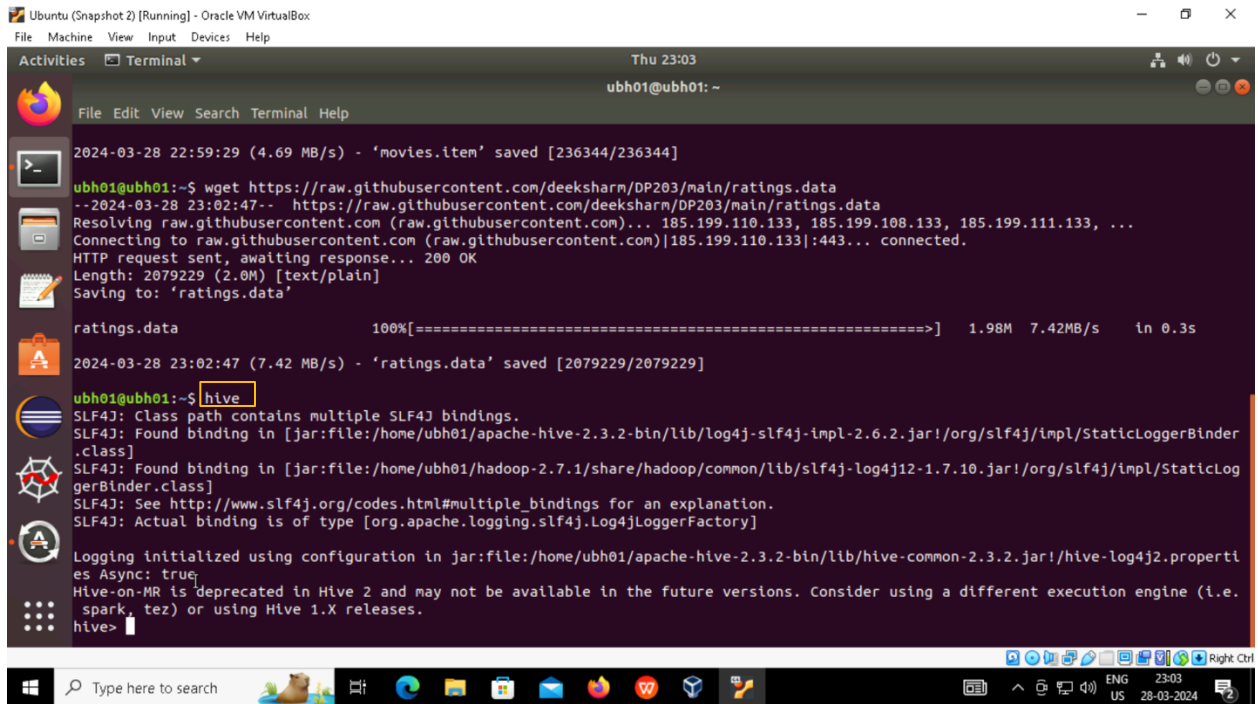
```
ubh01@ubh01:~$ jps
2562 Jps
ubh01@ubh01:~$ ./hadoop-2.7.1/sbin/start-all.sh
This script is deprecated. Instead use start-dfs.sh and start-yarn.sh
24/03/28 22:50:14 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes wh
ere applicable
Starting namenodes on [localhost]
localhost: starting namenode, logging to /home/ubh01/hadoop-2.7.1/logs/hadoop-ubh01-namenode-ubh01.out
localhost: starting datanode, logging to /home/ubh01/hadoop-2.7.1/logs/hadoop-ubh01-datanode-ubh01.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /home/ubh01/hadoop-2.7.1/logs/hadoop-ubh01-secondarynamenode-ubh01.out
24/03/28 22:50:31 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes wh
ere applicable
starting yarn daemons
starting resourcemanager, logging to /home/ubh01/hadoop-2.7.1/logs/yarn-ubh01-resourcemanager-ubh01.out
localhost: starting nodemanager, logging to /home/ubh01/hadoop-2.7.1/logs/yarn-ubh01-nodemanager-ubh01.out
ubh01@ubh01:~$ jps
3476 ResourceManager
3940 Jps
2916 NameNode
3624 NodeManager
3096 DataNode
ubh01@ubh01:~$ wget https://raw.githubusercontent.com/deeksharm/DP203/main/movies.item
--2024-03-28 22:59:29-- https://raw.githubusercontent.com/deeksharm/DP203/main/movies.item
Resolving raw.githubusercontent.com (raw.githubusercontent.com)... 185.199.111.133, 185.199.109.133, 185.199.108.133, ...
Connecting to raw.githubusercontent.com (raw.githubusercontent.com)|185.199.111.133|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 236344 (231K) [text/plain]
Saving to: 'movies.item'
```

Step2: Get the files (movies.item and ratings.data) into Local System.



```
ubh01@ubh01:~$ jps
3940 Jps
2916 NameNode
3624 NodeManager
3096 DataNode
ubh01@ubh01:~$ wget https://raw.githubusercontent.com/deeksharm/DP203/main/movies.item
--2024-03-28 22:59:29-- https://raw.githubusercontent.com/deeksharm/DP203/main/movies.item
Resolving raw.githubusercontent.com (raw.githubusercontent.com)... 185.199.111.133, 185.199.109.133, 185.199.108.133, ...
Connecting to raw.githubusercontent.com (raw.githubusercontent.com)|185.199.111.133|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 236344 (231K) [text/plain]
Saving to: 'movies.item'
movies.item 100%[=====] 230.80K --.-KB/s in 0.05s
2024-03-28 22:59:29 (4.69 MB/s) - 'movies.item' saved [236344/236344]
ubh01@ubh01:~$ wget https://raw.githubusercontent.com/deeksharm/DP203/main/ratings.data
--2024-03-28 23:02:47-- https://raw.githubusercontent.com/deeksharm/DP203/main/ratings.data
Resolving raw.githubusercontent.com (raw.githubusercontent.com)... 185.199.110.133, 185.199.108.133, 185.199.111.133, ...
Connecting to raw.githubusercontent.com (raw.githubusercontent.com)|185.199.110.133|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 2079229 (2.0M) [text/plain]
Saving to: 'ratings.data'
ratings.data 100%[=====] 1.98M 7.42MB/s in 0.3s
2024-03-28 23:02:47 (7.42 MB/s) - 'ratings.data' saved [2079229/2079229]
ubh01@ubh01:~$
```

Step3: Activate HIVE.



```
2024-03-28 22:59:29 (4.69 MB/s) - 'movies.item' saved [236344/236344]
ubh01@ubh01:~$ wget https://raw.githubusercontent.com/deeksharm/DP203/main/ratings.data
--2024-03-28 23:02:47-- https://raw.githubusercontent.com/deeksharm/DP203/main/ratings.data
Resolving raw.githubusercontent.com (raw.githubusercontent.com)... 185.199.110.133, 185.199.108.133, ...
Connecting to raw.githubusercontent.com (raw.githubusercontent.com)|185.199.110.133|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 2079229 (2.0M) [text/plain]
Saving to: 'ratings.data'

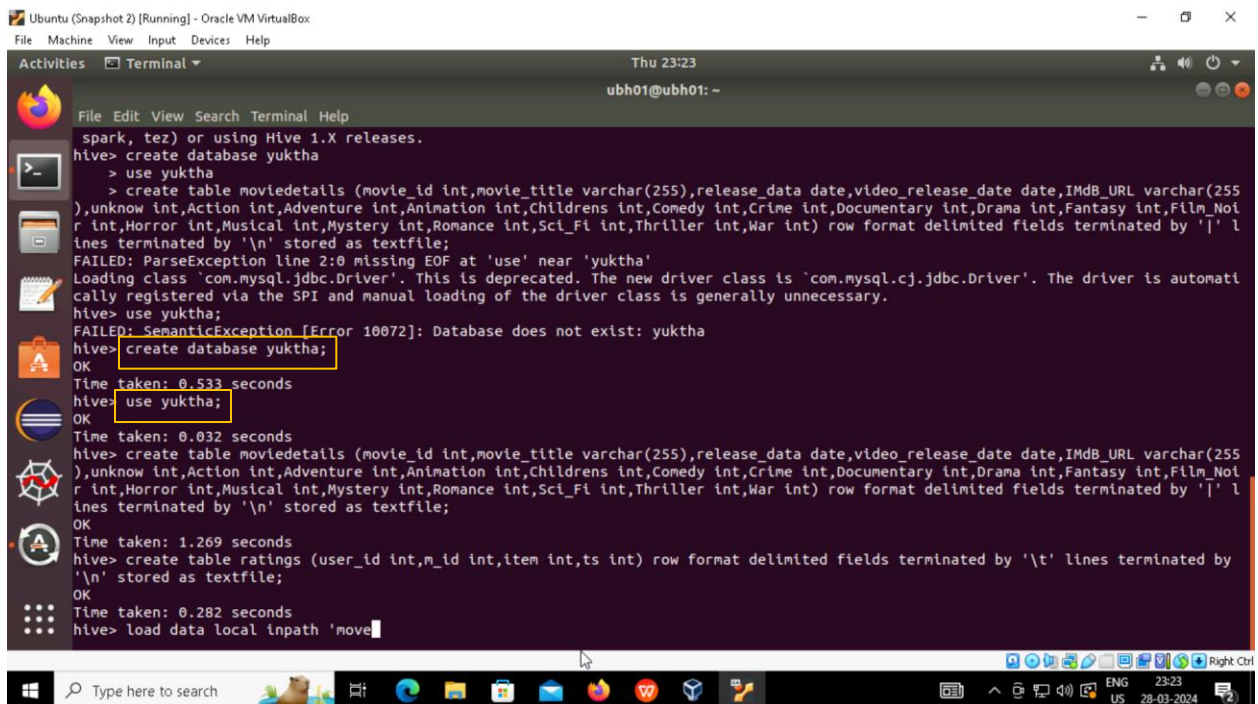
ratings.data                                100%[=====] 1.98M  7.42MB/s  in 0.3s

2024-03-28 23:02:47 (7.42 MB/s) - 'ratings.data' saved [2079229/2079229]

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/StaticLoggerBinder.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]

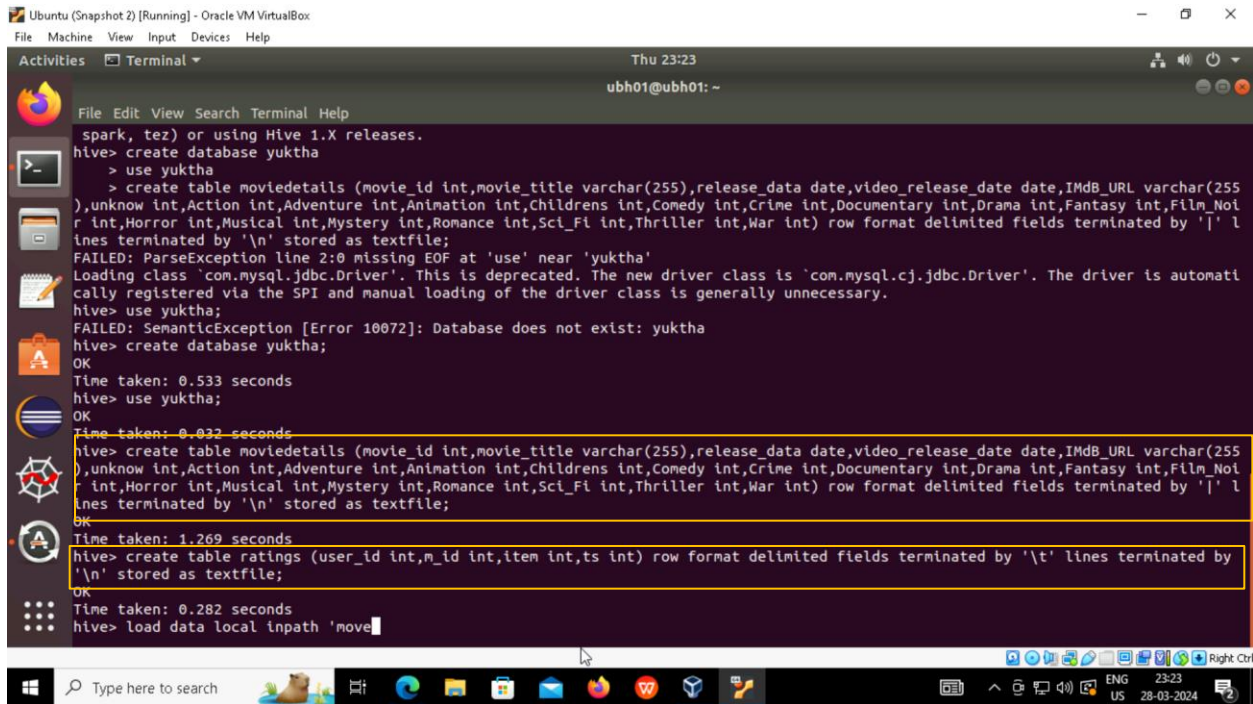
Logging initialized using configuration in jar:file:/home/ubh01/apache-hive-2.3.2-bin/lib/hive-common-2.3.2.jar!/hive-log4j2.properties Async: true
Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using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

Step4: Create database(Yuktha).



```
hive> create database yuktha
OK
Time taken: 0.032 seconds
hive> use yuktha;
OK
Time taken: 0.032 seconds
hive> create table moviedetails (movie_id int,movie_title varchar(255),release_data date,video_release_date date,IMdB_URL varchar(255),unknown 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) row format delimited fields terminated by '|' lines terminated by '\n' stored as textfile;
FAILED: ParseException line 2:0 missing EOF at 'use' near 'yuktha'
Loading class 'com.mysql.jdbc.Driver'. This is deprecated. The new driver class is 'com.mysql.cj.jdbc.Driver'. The driver is automatically registered via the SPI and manual loading of the driver class is generally unnecessary.
hive> use yuktha;
FAILED: SemanticException [Error 10072]: Database does not exist: yuktha
hive> create database yuktha;
OK
Time taken: 0.533 seconds
hive> use yuktha;
OK
Time taken: 0.032 seconds
hive> create table moviedetails (movie_id int,movie_title varchar(255),release_data date,video_release_date date,IMdB_URL varchar(255),unknown 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) row format delimited fields terminated by '|' lines terminated by '\n' stored as textfile;
OK
Time taken: 1.269 seconds
hive> create table ratings (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: 0.282 seconds
hive> load data local inpath 'move
```

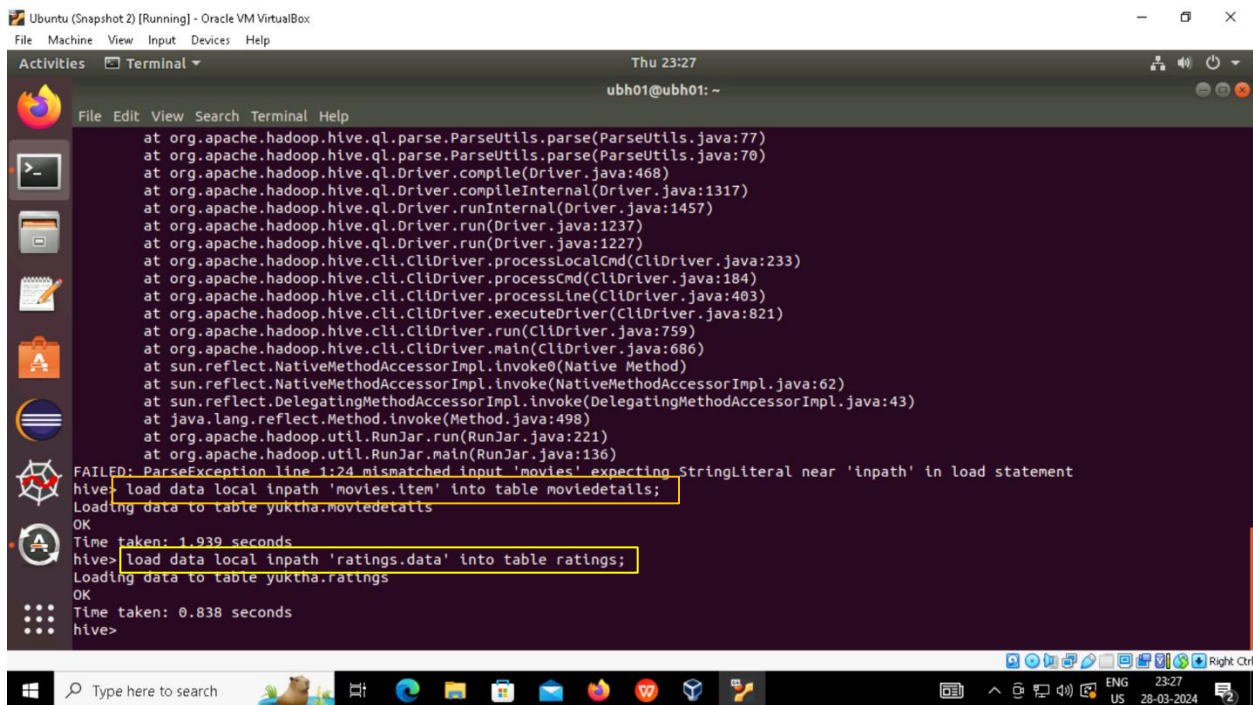
Step5: Create two new tables inside HIVE.



The screenshot shows a terminal window titled "Ubuntu (Snapshot 2) [Running] - Oracle VM VirtualBox". The terminal prompt is "ubh01@ubh01: ~". The user enters the following commands:

```
spark, tez) or using Hive 1.X releases.
hive> create database yuktha
> use yuktha
> create table moviedetails (movie_id int,movie_title varchar(255),release_data date,video_release_date date,IMDb_URL varchar(255),unknown 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) row format delimited fields terminated by '|' lines terminated by '\n' stored as textfile;
FAILED: ParseException line 2:0 missing EOF at 'use' near 'yuktha'
Loading class 'com.mysql.jdbc.Driver'. This is deprecated. The new driver class is 'com.mysql.cj.jdbc.Driver'. The driver is automatically registered via the SPI and manual loading of the driver class is generally unnecessary.
hive> use yuktha;
FAILED: SemanticException [Error 10072]: Database does not exist: yuktha
hive> create database yuktha;
OK
Time taken: 0.533 seconds
hive> use yuktha;
OK
Time taken: 0.032 seconds
hive> create table moviedetails (movie_id int,movie_title varchar(255),release_data date,video_release_date date,IMDb_URL varchar(255),unknown 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) row format delimited fields terminated by '|' lines terminated by '\n' stored as textfile;
OK
Time taken: 1.269 seconds
hive> create table ratings (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: 0.282 seconds
hive> load data local inpath 'move
```

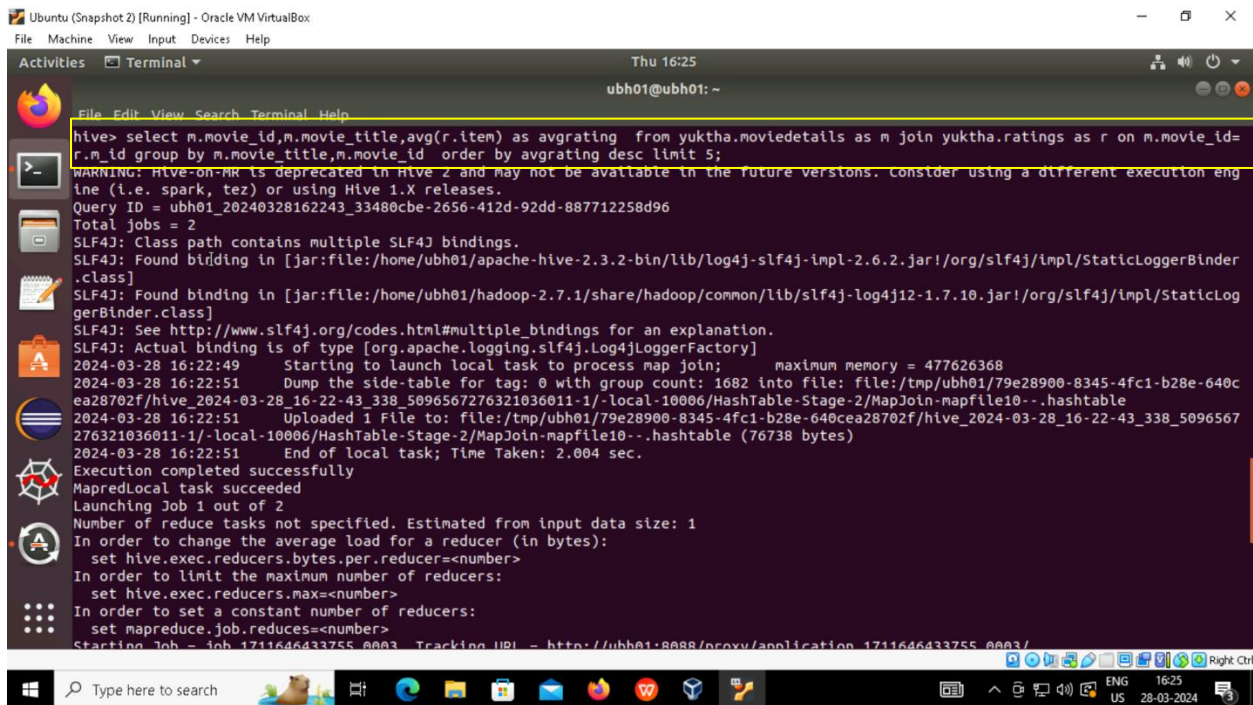
Step 6: Load the data from local files movies and rating into this tables.



The screenshot shows a terminal window titled "Ubuntu (Snapshot 2) [Running] - Oracle VM VirtualBox". The terminal prompt is "ubh01@ubh01: ~". The user enters the following commands:

```
at org.apache.hadoop.hive.ql.parse.ParseUtils.parse(ParseUtils.java:77)
at org.apache.hadoop.hive.ql.parse.ParseUtils.parse(ParseUtils.java:70)
at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:468)
at org.apache.hadoop.hive.ql.Driver.compileInternal(Driver.java:1317)
at org.apache.hadoop.hive.ql.Driver.runInternal(Driver.java:1457)
at org.apache.hadoop.hive.ql.Driver.run(Driver.java:1237)
at org.apache.hadoop.hive.ql.Driver.run(Driver.java:1227)
at org.apache.hadoop.hive.cli.CliDriver.processLocalCmd(CliDriver.java:233)
at org.apache.hadoop.hive.cli.CliDriver.processCmd(CliDriver.java:184)
at org.apache.hadoop.hive.cli.CliDriver.processLine(CliDriver.java:403)
at org.apache.hadoop.hive.cli.CliDriver.executeDriver(CliDriver.java:821)
at org.apache.hadoop.hive.cli.CliDriver.run(CliDriver.java:759)
at org.apache.hadoop.hive.cli.CliDriver.main(CliDriver.java:686)
at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
at java.lang.reflect.Method.invoke(Method.java:498)
at org.apache.hadoop.util.RunJar.run(RunJar.java:221)
at org.apache.hadoop.util.RunJar.main(RunJar.java:136)
FAILED: ParseException line 1:24 mismatched input 'movies' expecting StringLiteral near 'inpath' in load statement
hive> load data local inpath 'movies.item' into table moviedetails;
Loading data to table yuktha.moviedetails
OK
Time taken: 1.939 seconds
hive> load data local inpath 'ratings.data' into table ratings;
Loading data to table yuktha.ratings
OK
Time taken: 0.838 seconds
hive>
```

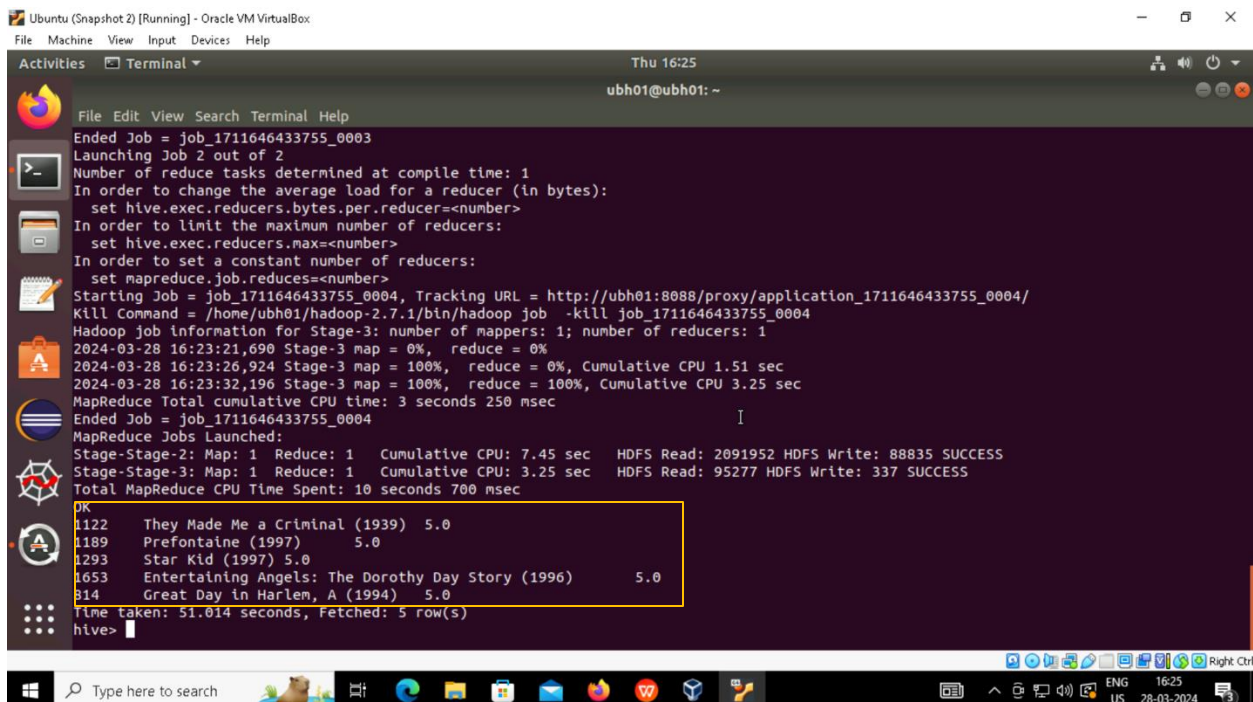

Step 7: Now write a query to retrieve the top 5 movies based on ratings.



The terminal window shows the execution of a Hive query. The query is: `hive> select m.movie_id,m.movie_title,avg(r.item) as avgrating from yuktha.moviedetails as m join yuktha.ratings as r on m.movie_id=r.m_id group by m.movie_title,m.movie_id order by avgrating desc limit 5;` The output shows the query ID, total jobs, and the execution of a MapReduce task. The task is completed successfully, and the output is displayed in a table format.

```
hive> select m.movie_id,m.movie_title,avg(r.item) as avgrating from yuktha.moviedetails as m join yuktha.ratings as r on m.movie_id=r.m_id group by m.movie_title,m.movie_id order by avgrating 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 engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = ubh01_20240328162243_33480cbe-2656-412d-92dd-887712258d96
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/StaticLoggerBinder.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 16:22:49 Starting to launch local task to process map join; maximum memory = 477626368
2024-03-28 16:22:51 Dump the side-table for tag: 0 with group count: 1682 into file: file:/tmp/ubh01/79e28900-8345-4fc1-b28e-640cea28702f/hive_2024-03-28_16-22-43_338_5096567276321036011-1-local-10006/HashTable-Stage-2/MapJoin-mapfile10--.hashtable
2024-03-28 16:22:51 Uploaded 1 File to: file:/tmp/ubh01/79e28900-8345-4fc1-b28e-640cea28702f/hive_2024-03-28_16-22-43_338_5096567276321036011-1-local-10006/HashTable-Stage-2/MapJoin-mapfile10--.hashtable (76738 bytes)
2024-03-28 16:22:51 End of local task; Time Taken: 2.004 sec.
Execution completed successfully
MapredLocal task succeeded
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1711646433755_0003 Tracking URL = http://ubh01:8088/proxy/application_1711646433755_0003/
```

Step8: View the output.



The terminal window shows the execution of a Hive query. The query is: `hive> select m.movie_id,m.movie_title,avg(r.item) as avgrating from yuktha.moviedetails as m join yuktha.ratings as r on m.movie_id=r.m_id group by m.movie_title,m.movie_id order by avgrating desc limit 5;` The output shows the query ID, total jobs, and the execution of a MapReduce task. The task is completed successfully, and the output is displayed in a table format.

```
Ended Job = job_1711646433755_0003
Launching Job 2 out of 2
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1711646433755_0004, Tracking URL = http://ubh01:8088/proxy/application_1711646433755_0004/
Kill Command = /home/ubh01/hadoop-2.7.1/bin/hadoop job -kill job_1711646433755_0004
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 1
2024-03-28 16:23:21,690 Stage-3 map = 0%, reduce = 0%
2024-03-28 16:23:26,924 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 1.51 sec
2024-03-28 16:23:32,196 Stage-3 map = 100%, reduce = 100%, Cumulative CPU 3.25 sec
MapReduce Total cumulative CPU time: 3 seconds 250 msec
Ended Job = job_1711646433755_0004
MapReduce Jobs Launched:
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 7.45 sec HDFS Read: 2091952 HDFS Write: 88835 SUCCESS
Stage-Stage-3: Map: 1 Reduce: 1 Cumulative CPU: 3.25 sec HDFS Read: 95277 HDFS Write: 337 SUCCESS
Total MapReduce CPU Time Spent: 10 seconds 700 msec
OK
1122 They Made Me a Criminal (1939) 5.0
1189 Prefontaine (1997) 5.0
1293 Star Kid (1997) 5.0
1653 Entertaining Angels: The Dorothy Day Story (1996) 5.0
814 Great Day in Harlem, A (1994) 5.0
Time taken: 51.014 seconds, Fetched: 5 row(s)
hive>
```