



ASSIGNMENT-1

(HIVE)

Employee ID : 2320074

CSDAIA24AZ003

```
ubh01@ubh01: ~  
File Edit View Search Terminal Help  
ubh01@ubh01:~$ jps  
2685 jps  
ubh01@ubh01:~$ ./hadoop-2.7.1/sbin/start-all.sh  
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh  
24/03/27 15:55:05 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
```

Step 1: Start the session.

```
File Edit View Search Terminal Help  
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 21:54:50 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 21:55:08 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
```

Step 2: Start the HDFS in the virtual machine.

```
ubh01@ubh01:~$ wget https://raw.githubusercontent.com/deeksharm/DP203/main/movies.item  
--2024-03-28 22:08:25-- https://raw.githubusercontent.com/deeksharm/DP203/main/movies.item  
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: 236344 (231K) [text/plain]  
Saving to: 'movies.item'  
  
movies.item 100%[=====] 230.80K --.-KB/s in 0.05s  
2024-03-28 22:08:25 (4.33 MB/s) - 'movies.item' saved [236344/236344]  
  
ubh01@ubh01:~$ wget https://raw.githubusercontent.com/deeksharm/DP203/main/ratings.data  
--2024-03-28 22:11:06-- https://raw.githubusercontent.com/deeksharm/DP203/main/ratings.data  
Resolving raw.githubusercontent.com (raw.githubusercontent.com)... 185.199.109.133, 185.199.111.133, 185.199.108.133, ...  
Connecting to raw.githubusercontent.com (raw.githubusercontent.com)|185.199.109.133|:443... connected.  
HTTP request sent, awaiting response... 200 OK  
Length: 2079229 (2.0M) [text/plain]  
Saving to: 'ratings.data'  
  
ratings.data 100%[=====] 1.98M --.-KB/s in 0.08s  
2024-03-28 22:11:06 (23.4 MB/s) - 'ratings.data' saved [2079229/2079229]
```

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/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.

```

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 automatically 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),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,Western int)row format delimited fields terminated by '|' lines terminated by '\n' stored as textfile;
OK
Time taken: 1.336 seconds
hive> create table ratingssss(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 ratingssss in the database we created in hive.

```

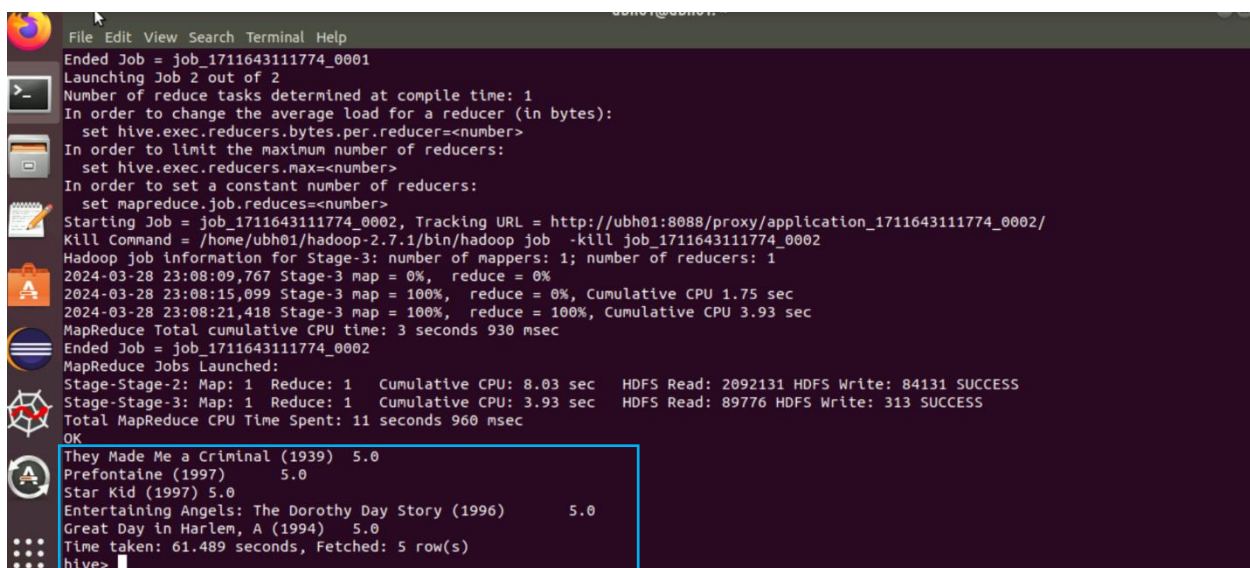
hive> load data local inpath 'movies.item' into table movieeee;
Loading data to table dharini.movieeee
OK
Time taken: 1.934 seconds
hive> load data local inpath 'ratings.data' into table ratingssss;
Loading data to table dharini.ratingssss
OK
Time taken: 0.784 seconds
hive>

```

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 ratings 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 engine (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/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 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-bice-4d84-a30f-cf491df5130b/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-bice-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.



```
File Edit View Search Terminal Help
Ended Job = job_1711643111774_0001
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_1711643111774_0002, Tracking URL = http://ubh01:8088/proxy/application_1711643111774_0002/
Kill Command = /home/ubh01/hadoop-2.7.1/bin/hadoop job -kill job_1711643111774_0002
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 1
2024-03-28 23:08:09,767 Stage-3 map = 0%, reduce = 0%
2024-03-28 23:08:15,099 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 1.75 sec
2024-03-28 23:08:21,418 Stage-3 map = 100%, reduce = 100%, Cumulative CPU 3.93 sec
MapReduce Total cumulative CPU time: 3 seconds 930 msec
Ended Job = job_1711643111774_0002
MapReduce Jobs Launched:
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 8.03 sec HDFS Read: 2092131 HDFS Write: 84131 SUCCESS
Stage-Stage-3: Map: 1 Reduce: 1 Cumulative CPU: 3.93 sec HDFS Read: 89776 HDFS Write: 313 SUCCESS
Total MapReduce CPU Time Spent: 11 seconds 960 msec
OK
They Made Me a Criminal (1939) 5.0
Prefontaine (1997) 5.0
Star Kid (1997) 5.0
Entertaining Angels: The Dorothy Day Story (1996) 5.0
Great Day in Harlem, A (1994) 5.0
Time taken: 61.489 seconds, Fetched: 5 row(s)
hive>
```

Step 11: Check the records at the output.