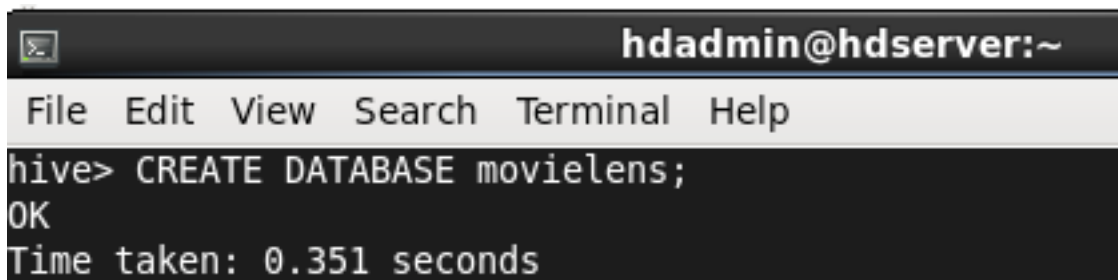# Assignment 7 Solution

STUDENT NAME: Chin-Ting Ko

1) Create database command line syntax you ran to create a database called "movielens" (Hive SQL you used)

USE DATABASE movielens;

2) Create table command line syntax you ran to create a table called "rating" (Hive SQL you used)

```
hdadmin@hdserver:~
File  Edit  View  Search  Terminal  Help
hive> CREATE TABLE rating (userid STRING, itemid STRING, rating STRING, timestampdata STRING)
    > ROW FORMAT DELIMITED
    > FIELDS TERMINATED BY '\t'
    > STORED as TEXTFILE;
OK
Time taken: 0.983 seconds
hive> show tables;
OK
rating
Time taken: 0.166 seconds, Fetched: 1 row(s)
```

```
hdadmin@hdserver:~
File  Edit  View  Search  Terminal  Help
hive> describe rating;
OK
userid                  string
itemid                  string
rating                  string
timestampdata           string
Time taken: 0.355 seconds, Fetched: 4 row(s)
hive>
```

3)Create table command line syntax you ran to create a table called "item" (Hive SQL you used)

```
                                  hdadmin@hdserver:~
File  Edit  View  Search  Terminal  Help
hive> CREATE TABLE item (movieid STRING, movietitle STRING, releasedate STRING, releasedata STRING, im
db STRING, unknown STRING, action STRING, adventure STRING, animation STRING, children STRING, comedy
STRING, crime STRING, documentary STRING, drama STRING, fantasy STRING, filmnoir STRING, horror STRING
, musical STRING, mystery STRING, romance STRING, scifi STRING, thriller STRING, war STRING, western S
TRING)STORED as TEXTFILE;
OK  >
Time taken: 0.355 seconds
hive> show tables;
OK
item
rating
Time taken: 0.141 seconds, Fetched: 2 row(s)
hive> describe item;
OK
movieid                 string
movietitle              string
releasedate             string
releasedata             string
imdb                    string
unknown                 string
action                  string
adventure               string
animation               string
children                string
comedy                  string
crime                   string
documentary             string
drama                   string
fantasy                 string
filmnoir                string
horror                  string
musical                 string
mystery                 string
romance                 string
scifi                   string
thriller                string
war                     string
western                 string
Time taken: 0.116 seconds, Fetched: 24 row(s)
hive> 
```

4)Syntax you used to load data into rating table. (Hive or whatever syntax you used).

```
hdadmin@hdserver:~

File  Edit  View  Search  Terminal  Help
hive> LOAD DATA LOCAL INPATH 'u.data'
    > OVERWRITE INTO TABLE rating;
Loading data to table default.rating
Table default.rating stats: [numFiles=1, numRows=0, totalSize=1979173, rawDataSize=0]
OK
Time taken: 2.516 seconds
hive>
```

5) Syntax you used to load data into item table. (Hive or whatever syntax you used).

```
hdadmin@hdserver:~
File  Edit  View  Search  Terminal  Help
hive> LOAD DATA LOCAL INPATH 'u.item'
    > OVERWRITE INTO TABLE item;
Loading data to table default.item
Table default.item stats: [numFiles=1, numRows=0, totalSize=236344, rawDataSize=0]
OK
Time taken: 2.537 seconds
hive>
```

6) **Select statement with join and group by to retrieve the records, or whatever statement(s) you used to extract data as described in the assignment (Hive SQL you used)**


```
hdadmin@hdserver:~
File  Edit  View  Search  Terminal  Help
hive> select item.movieid, movietitle, releasedate, imdb, avg(rating.rating) avg_rating
    > from item join rating on (rating.itemid = item.movieid)
    > group by item.movieid, movietitle, releasedate, imdb
    > order by avg_rating desc
    > limit 20;
Query ID = hdadmin_20161105072421_3cda4df5-5375-4765-8ffb-600ac932bb7f
Total jobs = 2
Java HotSpot(TM) Server VM warning: You have loaded library /usr/local/hadoop/hadoop-2.7.2/lib/native/libhadoop.so.1.0.0
which might have disabled stack guard. The VM will try to fix the stack guard now.
It's highly recommended that you fix the library with 'execstack -c <libfile>', or link it with '-z noexecstack'.
16/11/05 07:24:25 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-jav
a classes where applicable
Execution log at: /tmp/hdadmin/hdadmin_20161105072421_3cda4df5-5375-4765-8ffb-600ac932bb7f.log
2016-11-05 07:24:26     Starting to launch local task to process map join;     maximum memory = 477364224
2016-11-05 07:24:27     Dump the side-table for tag: 0 with group count: 1682 into file: file:/tmp/hdadmin/dcb1accb-b39d-
4fb4-b43a-72e6492b8604/hive_2016-11-05_07-24-21_533_12141448258595564247-1/-local-10005/HashTable-Stage-2/MapJoin-mapfile0
0--.hashtable
2016-11-05 07:24:27     Uploaded 1 File to: file:/tmp/hdadmin/dcb1accb-b39d-4fb4-b43a-72e6492b8604/hive_2016-11-05_07-24-
21_533_12141448258595564247-1/-local-10005/HashTable-Stage-2/MapJoin-mapfile00--.hashtable (199612 bytes)
2016-11-05 07:24:27     End of local task; Time Taken: 1.798 sec.
Execution completed successfully
```

7) Run the program on Hive command line and place the screenshot of successful run/ execution right here. You can limit your run to 10 results (if sending output to screen), or you can send the output to a file. At this point show the execution of Hive query to list of all movies rated with the following information per line -- movie id, movie title, release date, IMDB URL, average rating for the movie.

```
                                    hdadmin@hdserver:~
 File  Edit  View  Search  Terminal  Help
Ended Job = job_1478355130597_0004
MapReduce Jobs Launched:
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 8.34 sec   HDFS Read: 1993176 HDFS Write: 212294 SUCCESS
Stage-Stage-3: Map: 1  Reduce: 1   Cumulative CPU: 4.02 sec   HDFS Read: 218000 HDFS Write: 2378 SUCCESS
Total MapReduce CPU Time Spent: 12 seconds 360 msec
OK
1467    Saint of Fort Washington, The (1993)    01-Jan-1993    http://us.imdb.com/M/title-exact?Saint%20of%20Fort%20Wash
ington,%20The%20(1993)  5.0
1599    Someone Else's America (1995)    10-May-1996    http://us.imdb.com/M/title-exact?Someone%20Else's%20America%20(19
95)     5.0
1201    Marlene Dietrich: Shadow and Light (1996)    02-Apr-1996    http://us.imdb.com/M/title-exact?Marlene%20Dietri
ch:%20Shadow%20and%20Light%20(1996)%20(TV)    5.0
1500    Santa with Muscles (1996)    08-Nov-1996    http://us.imdb.com/M/title-exact?Santa%20with%20Muscles%20(1996)5
.0
1293    Star Kid (1997) 16-Jan-1998    http://us.imdb.com/M/title-exact?imdb-title-120478    5.0
1189    Prefontaine (1997)    24-Jan-1997    http://us.imdb.com/M/title-exact?Prefontaine%20(1997)    5.0
814    Great Day in Harlem, A (1994)    01-Jan-1994    http://us.imdb.com/M/title-exact?Great%20Day%20in%20Harlem,%20A%2
0(1994) 5.0
1653    Entertaining Angels: The Dorothy Day Story (1996)    27-Sep-1996    http://us.imdb.com/M/title-exact?Entertai
ning%20Angels:%20The%20Dorothy%20Day%20Story%20(1996)    5.0
1122    They Made Me a Criminal (1939) 01-Jan-1939    http://us.imdb.com/M/title-exact?They%20Made%20Me%20a%20Criminal%
20(1939)       5.0
1536    Aiqing wansui (1994)    22-Jul-1996    http://us.imdb.com/M/title-exact?Aiqing%20Wansui%20(1994)      5.0
1449    Pather Panchali (1955) 22-Mar-1996    http://us.imdb.com/M/title-exact?Pather%20Panchali%20(1955)    4.625
1594    Everest (1998) 10-Mar-1998    http://us.imdb.com/Title?Everest+(1998) 4.5
119     Maya Lin: A Strong Clear Vision (1994) 01-Jan-1994    http://us.imdb.com/M/title-exact?Maya%20Lin:%20A%20Strong
%20Clear%20Vision%20(1994)    4.5
1398    Anna (1996)    13-Nov-1996    http://us.imdb.com/M/title-exact?Anna%20(1996)  4.5
1642    Some Mother's Son (1996)    27-Dec-1996    http://us.imdb.com/M/title-exact?Some%20Mother's%20Son%20(1996) 4
.5
408     Close Shave, A (1995)    28-Apr-1996    http://us.imdb.com/M/title-exact?Close%20Shave,%20A%20(1995)    4.4910714
28571429
318     Schindler's List (1993) 01-Jan-1993    http://us.imdb.com/M/title-exact?Schindler's%20List%20(1993)    4.4664429
53020135
169     Wrong Trousers, The (1993)    01-Jan-1993    http://us.imdb.com/M/title-exact?Wrong%20Trousers,%20The%20(1993)
4.466101694915254
483     Casablanca (1942)    01-Jan-1942    http://us.imdb.com/M/title-exact?Casablanca%20(1942)    4.45679012345679
114     Wallace & Gromit: The Best of Aardman Animation (1996) 05-Apr-1996    http://us.imdb.com/Title?Wallace+%26+Grom
it%3A+The+Best+of+Aardman+Animation+(1996)    4.447761194029851
Time taken: 62.513 seconds, Fetched: 20 row(s)
hive> █
```

8)A screen shot of the JobHistoryServer when you ran the job. Click on the Job, get into the details, and then take a screen shot.

MapReduce Job job_1478355130597_0004 - Mozilla Firefox

Application application…   ×   MapReduce Job job_14…   ×   ✛

← ⓘ | localhost:19888/jobhistory/job/job_1478355130597_0004 | C̃ | 🔍 Search   ☆ 📋 ▽ ⬇

Logged

MapReduce Job job_1478355130597_0004

▸ Application

▾ Job
  Overview
  Counters
  Configuration
  Map tasks
  Reduce tasks

▸ Tools

Job

Job Name:	select item.movieid, movietitle, releas…20(Stage-3)
User Name:	hdadmin
Queue:	default
State:	SUCCEEDED
Uberized:	false
Submitted:	Sat Nov 05 07:24:56 PDT 2016
Started:	Sat Nov 05 07:25:08 PDT 2016
Finished:	Sat Nov 05 07:25:22 PDT 2016
Elapsed:	13sec
Diagnostics:	
Average Map Time	4sec
Average Shuffle Time	3sec
Average Merge Time	0sec
Average Reduce Time	1sec

ApplicationMaster

Attempt Number	Start Time	Node
1	Sat Nov 05 07:25:02 PDT 2016	localhost:8042

Task Type	Total	Complete
Map	1	1
Reduce	1	1

Attempt Type	Failed	Killed	Successful
Maps	0	0	1
Reduces	0	0	1

9) For the movie id 376 and 495, what is the movie title, release date, IMDB URL, average rating that you got as result using Hive. Please fill the data grid below:

Movie id	Title	Release date	IMDB URL	Average rating
376	Houseguest	01-Jan-1994	http:// us.imdb.co m/ M/title- exact? Houseguest %20(1994)	2.833
495	Around the world in 80 days	01- Jan-1956	http:// us.imdb.co m/ M/title- exact? Around %20the %20World %20in %2080%20 Da ys%20(195 6)	3.661