**Assignment 7 Solution**

**STUDENT NAME: Guan Yue Wang**

1. **Create database command line syntax you ran to create a database called “movielens” (Hive SQL you used)**

hive> CREATE DATABASE movielens;

OK

Time taken: 1.025 seconds

1. **Create table command line syntax you ran to create a table called “rating” (Hive SQL you used)**

hive> CREATE TABLE movielens.rating

> (user\_id SMALLINT, movie\_id SMALLINT, rating SMALLINT, time BIGINT)

> ROW FORMAT DELIMITED

> FIELDS TERMINATED BY '\t'

> STORED AS TEXTFILE;

OK

Time taken: 1.611 seconds

1. **Create table command line syntax you ran to create a table called “item” (Hive SQL you used)**

hive> CREATE TABLE movielens.item

> (movie\_id SMALLINT, movie\_title STRING, release\_date STRING, video\_release\_date STRING,

> URL STRING, unknown STRING, Action BINARY, Adventure BINARY, Animation BINARY,

> Children BINARY, Comedy BINARY, Crime BINARY, Documentary BINARY, Drama BINARY, Fantasy BINARY,

> Film\_noir BINARY, Horror BINARY, Musical BINARY, Mystery BINARY, Romance BINARY, Sci\_fi BINARY)

> ROW FORMAT DELIMITED

> FIELDS TERMINATED BY '|'

> STORED AS TEXTFILE;

OK

Time taken: 0.348 seconds

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

hive> LOAD DATA LOCAL INPATH 'ml-100k/u.data'

> OVERWRITE INTO TABLE movielens.rating;

Loading data to table movielens.rating

Table movielens.rating stats: [numFiles=1, numRows=0, totalSize=1979173, rawDataSize=0]

OK

Time taken: 1.584 seconds

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

hive> LOAD DATA LOCAL INPATH 'ml-100k/u.item'

> OVERWRITE INTO TABLE movielens.item;

Loading data to table movielens.item

Table movielens.item stats: [numFiles=1, numRows=0, totalSize=236344, rawDataSize=0]

OK

Time taken: 1.119 seconds

1. **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)**

hive> SELECT i.movie\_id, i.movie\_title, i.release\_date, i.url, AVG(r.rating) avg\_rating FROM movielens.item i JOIN movielens.rating r on (i.movie\_id=r.movie\_id) GROUP BY i.movie\_id, i.movie\_title, i.release\_date, i.url;

Query ID = gwang39\_20181203220728\_c70892e8-3188-4784-8337-3f690b8e7a4e

Total jobs = 1

1. **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.**

MapReduce Jobs Launched:

Stage-Stage-2: Map: 1  Reduce: 1 Cumulative CPU: 7.42 sec   HDFS Read: 1993896 HDFS Write: 193293 SUCCESS

Total MapReduce CPU Time Spent: 7 seconds 420 msec

OK

1    Toy Story (1995)    01-Jan-1995 http://us.imdb.com/M/title-exact?Toy%20Story%20(1995)    3.8783185840707963

2    GoldenEye (1995)    01-Jan-1995 http://us.imdb.com/M/title-exact?GoldenEye%20(1995)    3.2061068702290076

3    Four Rooms (1995)    01-Jan-1995 http://us.imdb.com/M/title-exact?Four%20Rooms%20(1995)    3.033333333333333

4    Get Shorty (1995)    01-Jan-1995 http://us.imdb.com/M/title-exact?Get%20Shorty%20(1995)    3.550239234449761

5    Copycat (1995)    01-Jan-1995 http://us.imdb.com/M/title-exact?Copycat%20(1995)    3.302325581395349

6    Shanghai Triad (Yao a yao yao dao waipo qiao) (1995)    01-Jan-1995 http://us.imdb.com/Title?Yao+a+yao+yao+dao+waipo+qiao+(1995)    3.576923076923077

7    Twelve Monkeys (1995)    01-Jan-1995 http://us.imdb.com/M/title-exact?Twelve%20Monkeys%20(1995)    3.798469387755102

8    Babe (1995)    01-Jan-1995 http://us.imdb.com/M/title-exact?Babe%20(1995)    3.9954337899543377

9    Dead Man Walking (1995)    01-Jan-1995 http://us.imdb.com/M/title-exact?Dead%20Man%20Walking%20(1995)    3.8963210702341136

10    Richard III (1995)    22-Jan-1996 http://us.imdb.com/M/title-exact?Richard%20III%20(1995)    3.831460674157303

…….

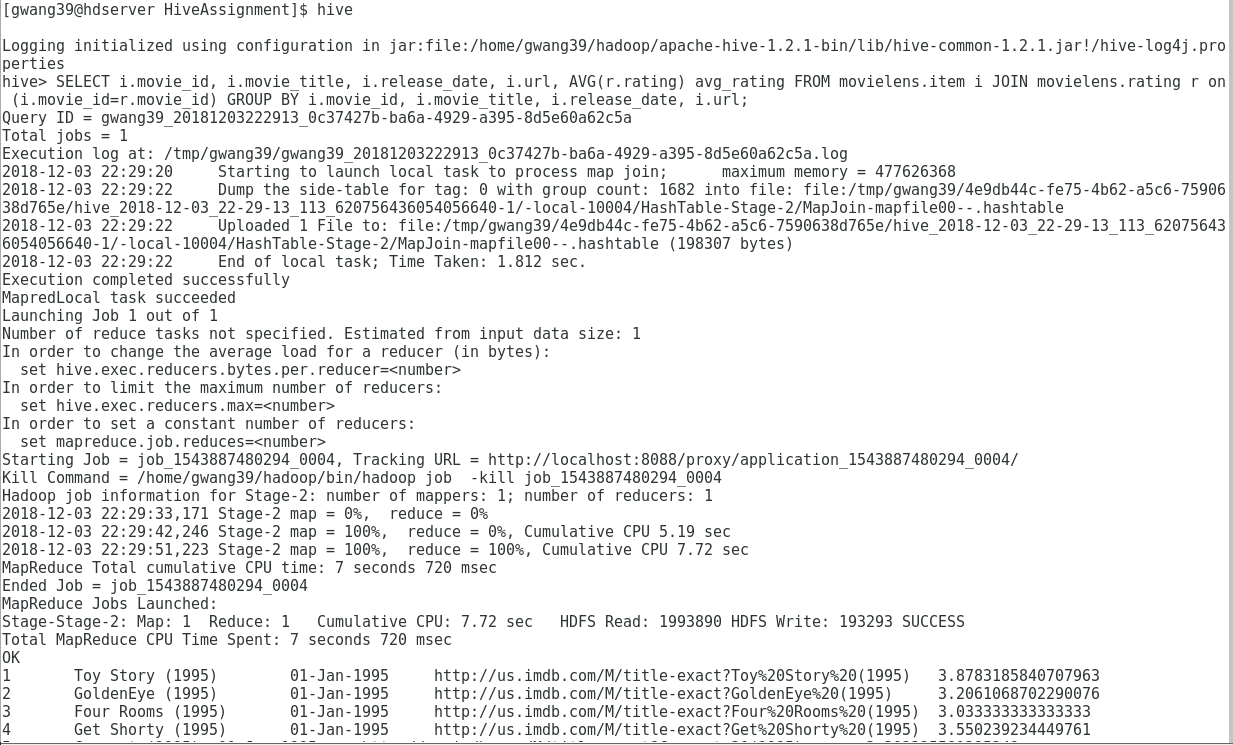
……. (RESULTS IN THE MIDDLE NOT INCLUDED DUE TO LENGTH)

…….

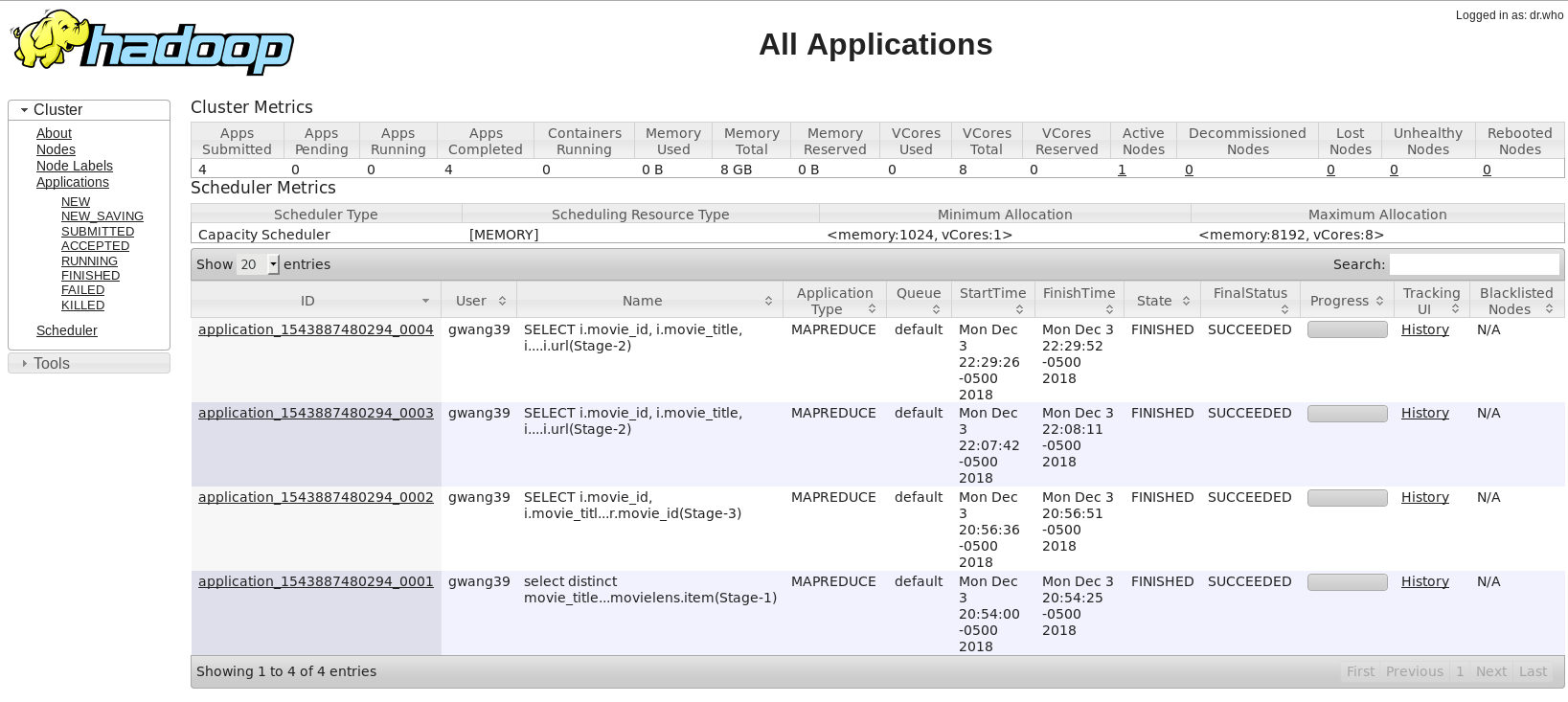
1681    You So Crazy (1994)    01-Jan-1994 http://us.imdb.com/M/title-exact?You%20So%20Crazy%20(1994)    3.0

1682    Scream of Stone (Schrei aus Stein) (1991)    08-Mar-1996 http://us.imdb.com/M/title-exact?Schrei%20aus%20Stein%20(1991)    3.0

Time taken: 45.993 seconds, Fetched: 1682 row(s)



1. **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.**



1. **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 (1994) | 01-Jan-1994 | http://us.imdb.com/M/title-exact?Houseguest%20(1994) | 2.83 |
| **495** | Around the World in 80 Days (1956) | 01-Jan-1956 | http://us.imdb.com/M/title-exact?Around%20the%20World%20in%2080%20Days%20(1956) | 3.66 |