Using movie ratings data, Develop the queries in Hive for the following.

a. List all the Users who have rated the movies (Users who have rated at least one movie).

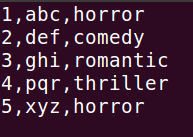
b. List all the User with the max, min, average ratings they have given against any movie.

c. List all the movies with the max, min, average ratings given by the user.

1.Create 3 text files and add contents to it as shown in the screenshot below:

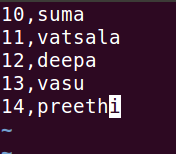


Movie(movieid, title, genre)



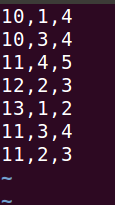


User(userid,name)

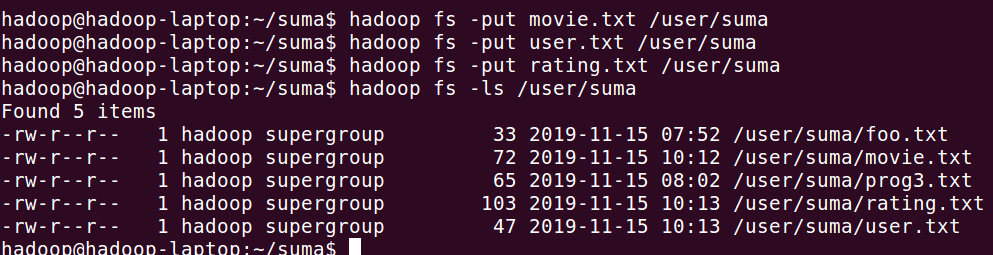




Rating(userid, movieid, rating)



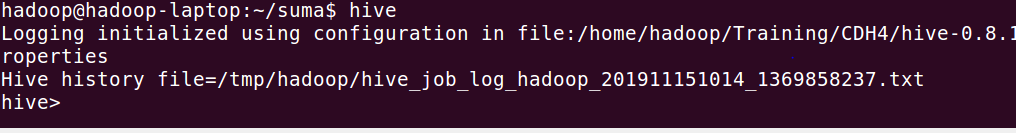
2.Copy the file from root directory to HDFS /user/suma directory using the following commands:

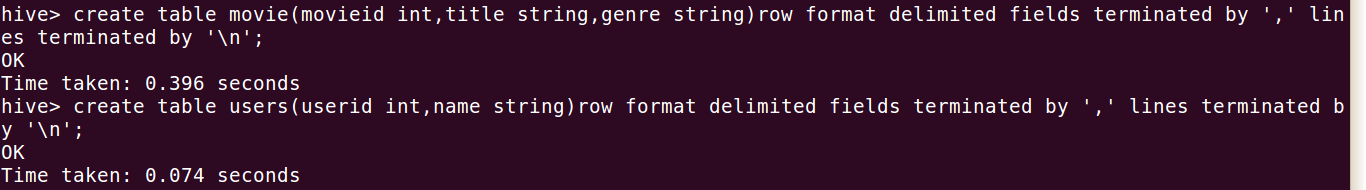


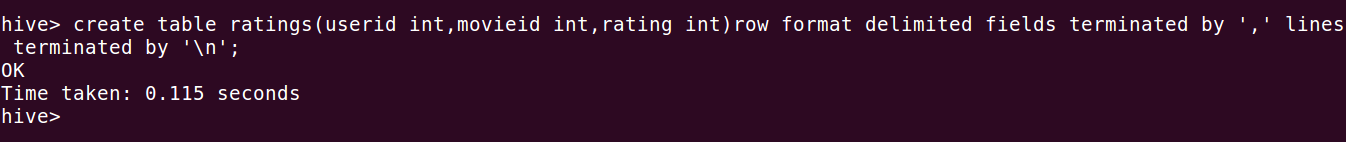
3.To launch the Hive, give the command: hive

Create a database movie\_ratings using the command: create database movie\_ratings;

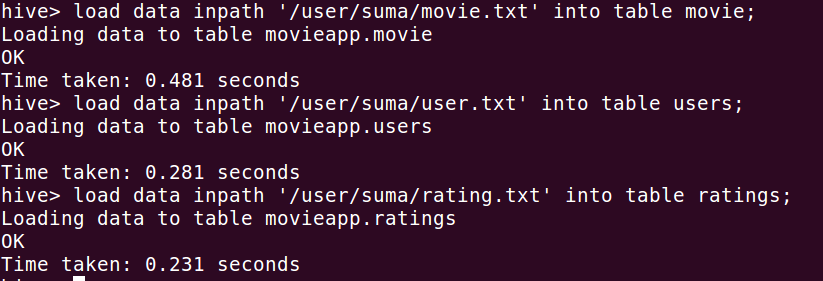
Create 3 tables movies, ratings and users as shown in the screenshot.







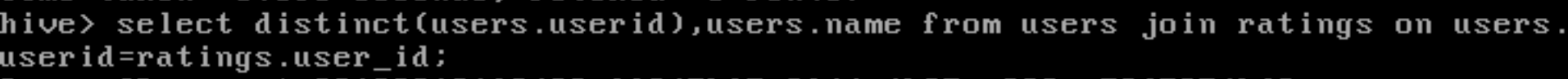
4.Load data into the tables created from the text files stored in /user/hive using the following commands: load data inpath ‘/user/hive/filename' into table tablename;

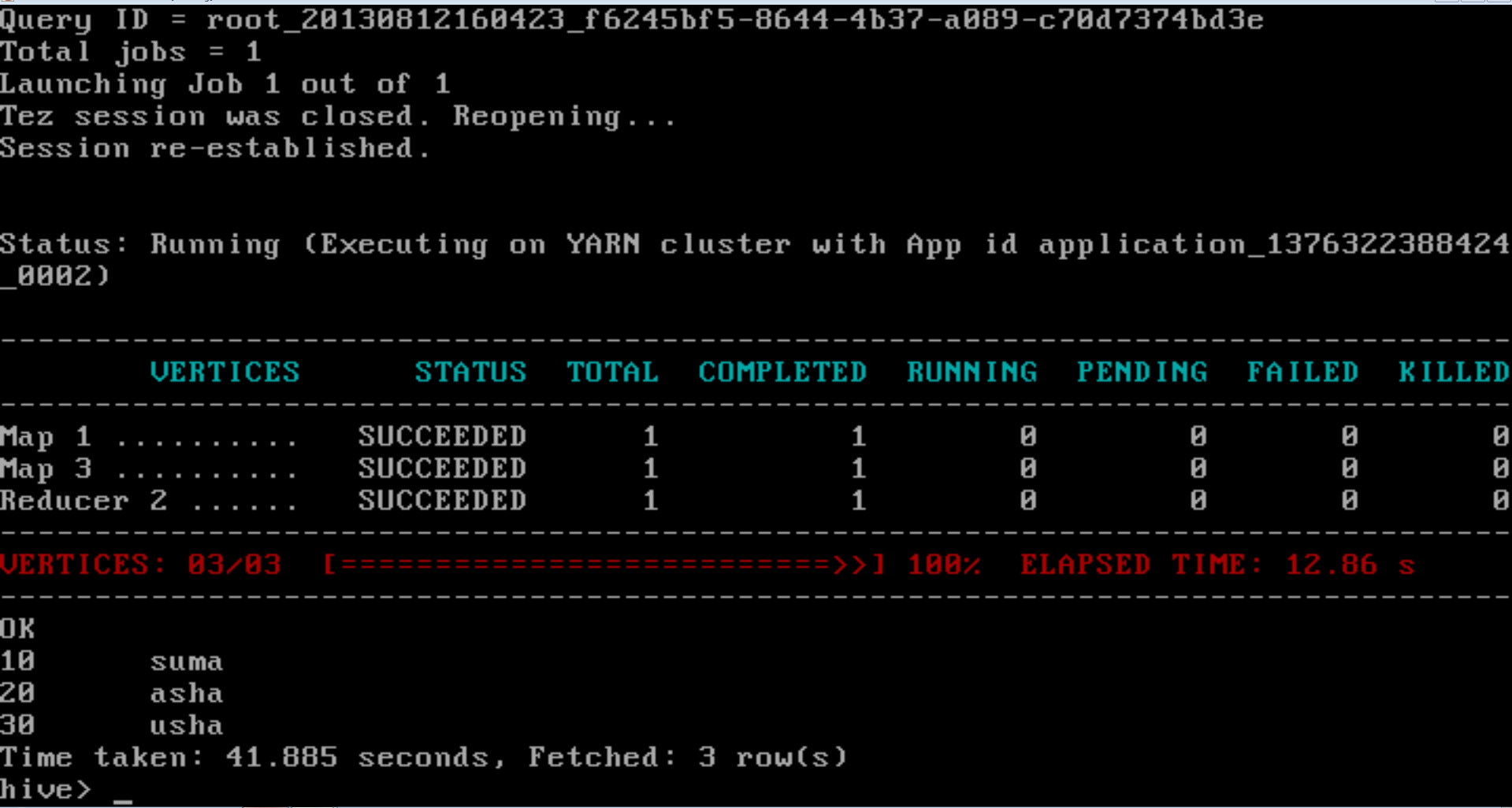


5. Till now we have created the tables and loaded data into the tables. Now execute the queries:

Query 1: List all the Users who have rated the movies (Users who have rated at least one movie)

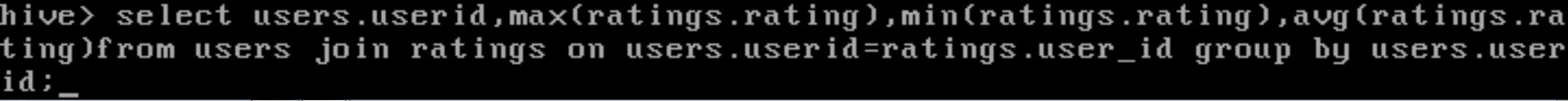
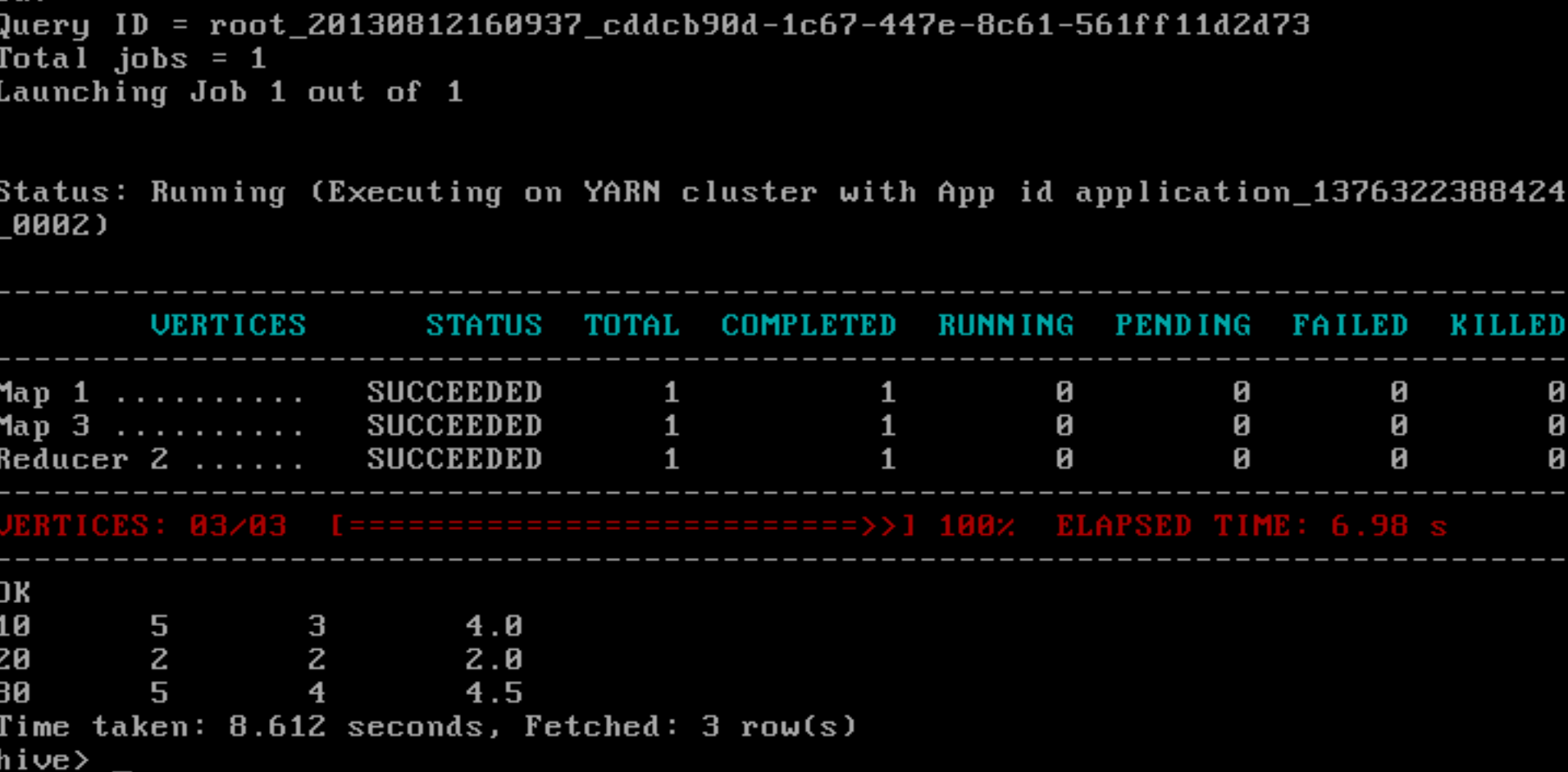
.Cmd: select distinct(users.userid int),users.name from users join ratings on users.userid=ratings.userid;





Query 2: List all the User with the max, min, average ratings they have given against any movie.

Cmd: select users.u\_id,max(ratings.rating),min(ratings.rating),avg(ratings.rating) from users join ratings on users.u\_id=ratings.u\_id group by users.u\_id;



Query 3: List all the movies with the max, min, average ratings given by the user

.

Cmd: select movies.movie\_id,max(ratings.rating),min(ratings.rating),avg(ratings.rating)from movies join ratings on movies.movie\_id=ratings.movie\_id group by movies.movie\_id;

