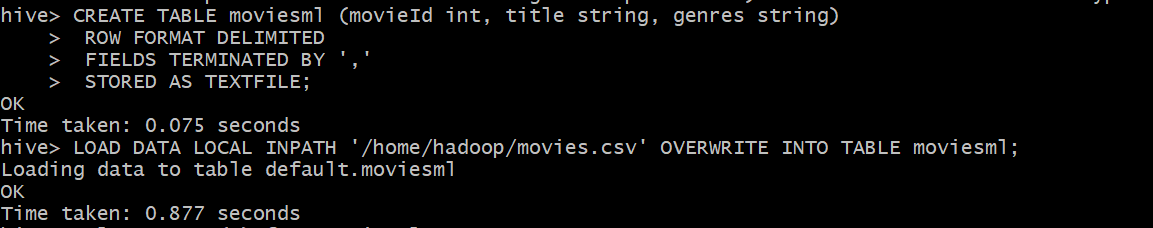
Hive Queries:

* Create Table for Movies

CREATE TABLE moviesml (movieId int, title string, genres string) ROW FORMAT DELIMITED

FIELDS TERMINATED BY ',' STORED AS TEXTFILE;

LOAD DATA LOCAL INPATH '/home/hadoop/movies.csv' OVERWRITE INTO TABLE moviesml;

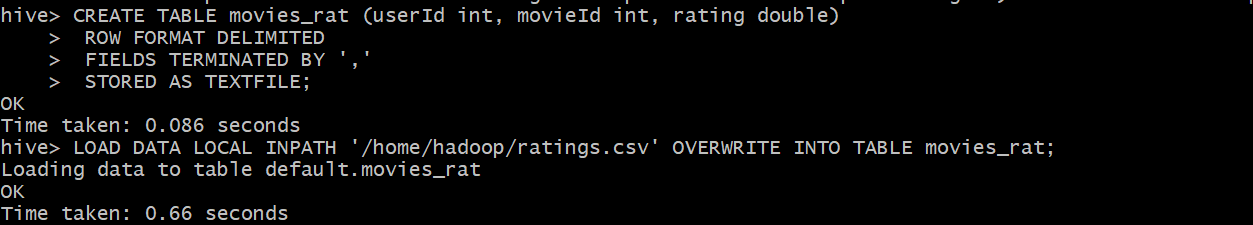


* Create Table for Ratings

CREATE TABLE movies\_rat (userId int, movieId int, rating double) ROW FORMAT DELIMITED

FIELDS TERMINATED BY ',' STORED AS TEXTFILE;

LOAD DATA LOCAL INPATH '/home/hadoop/ratings.csv' OVERWRITE INTO TABLE movies\_rat;

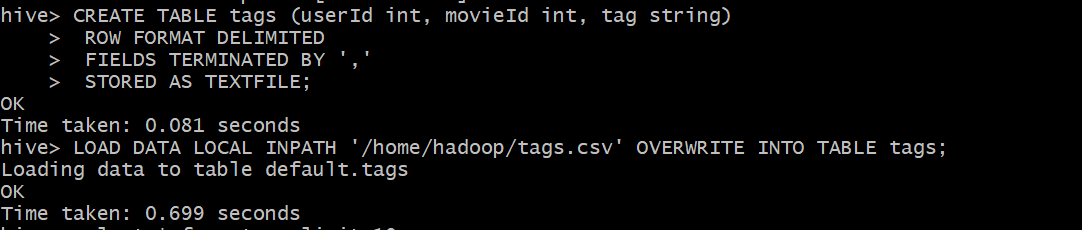


* Create table for Tags

CREATE TABLE tags (userId int, movieId int, tag string) ROW FORMAT DELIMITED

FIELDS TERMINATED BY ',' STORED AS TEXTFILE;

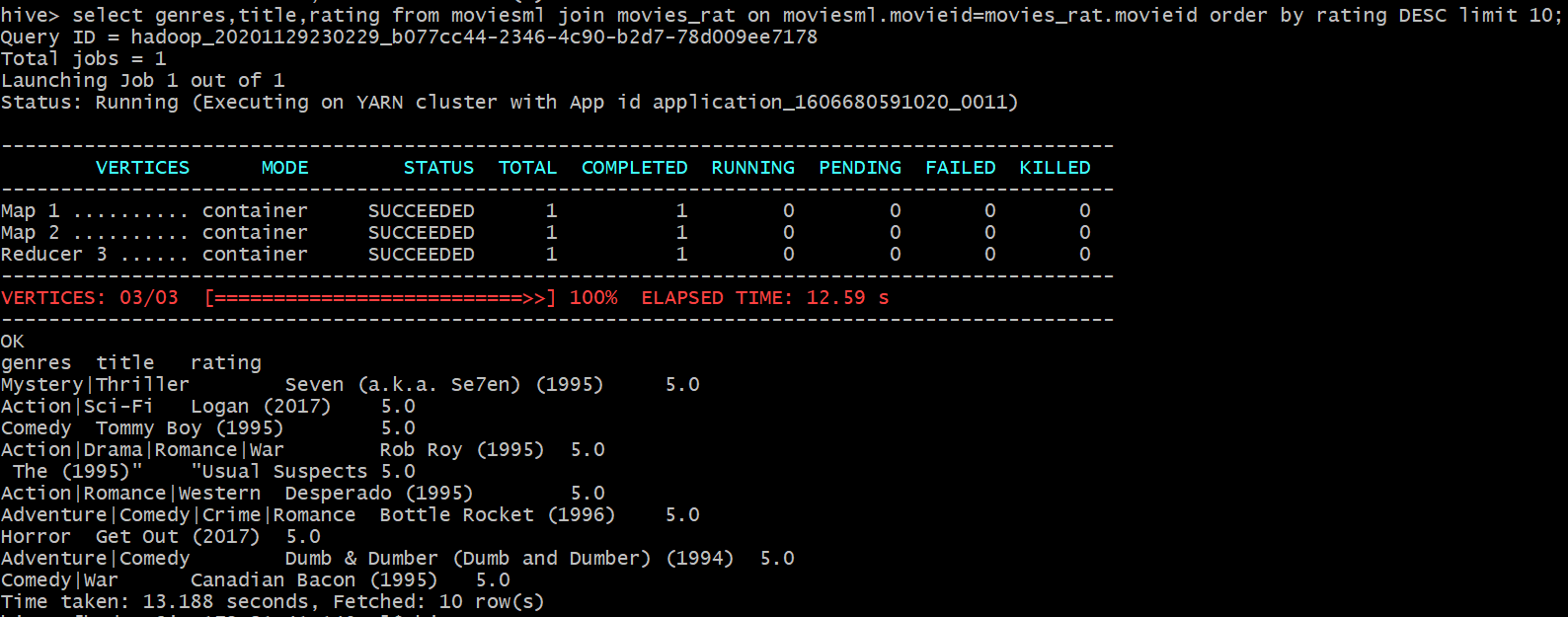
LOAD DATA LOCAL INPATH '/home/hadoop/tags.csv' OVERWRITE INTO TABLE tags;



1. To find the last 10 movies with highest rating with its genre, title, rating columns.

**Query:**

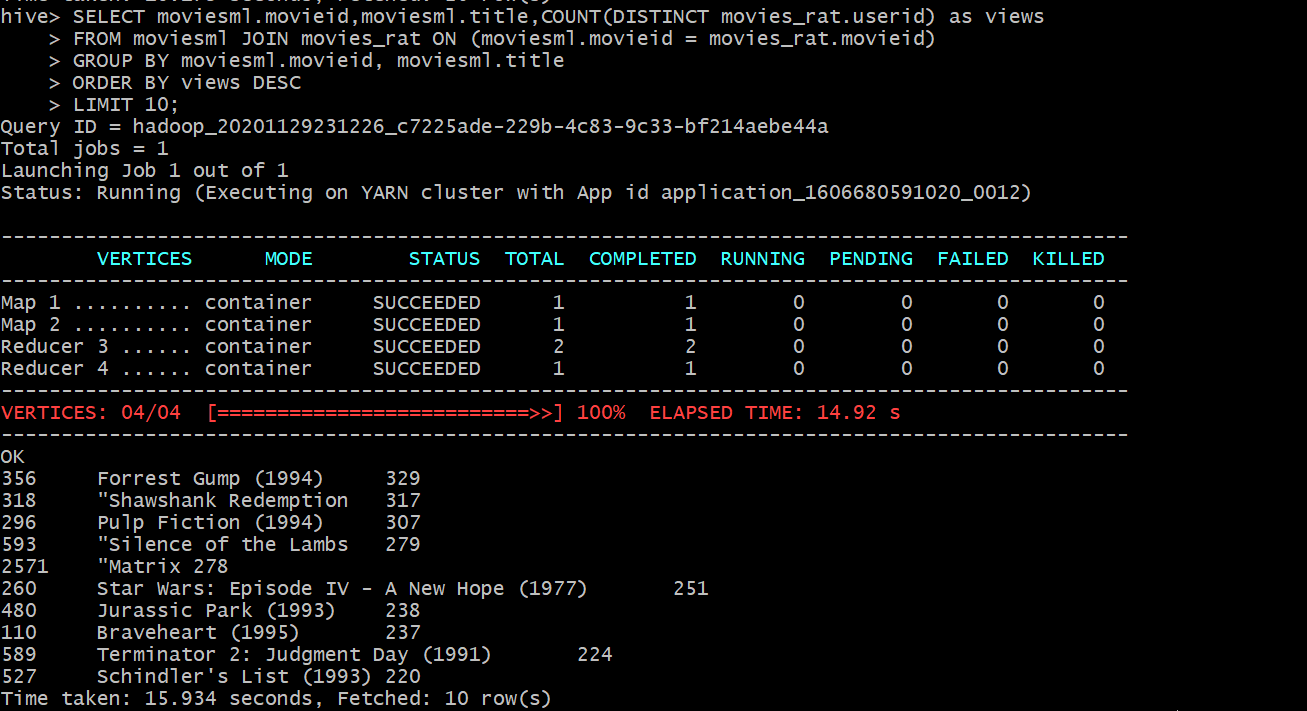
select genres,title,rating from moviesml join movies\_rat on moviesml.movieid=movies\_rat.movieid order by rating DESC limit 10;



1. To get the number of user ratings (based on highest number of ratings) to movies.

**Query:**

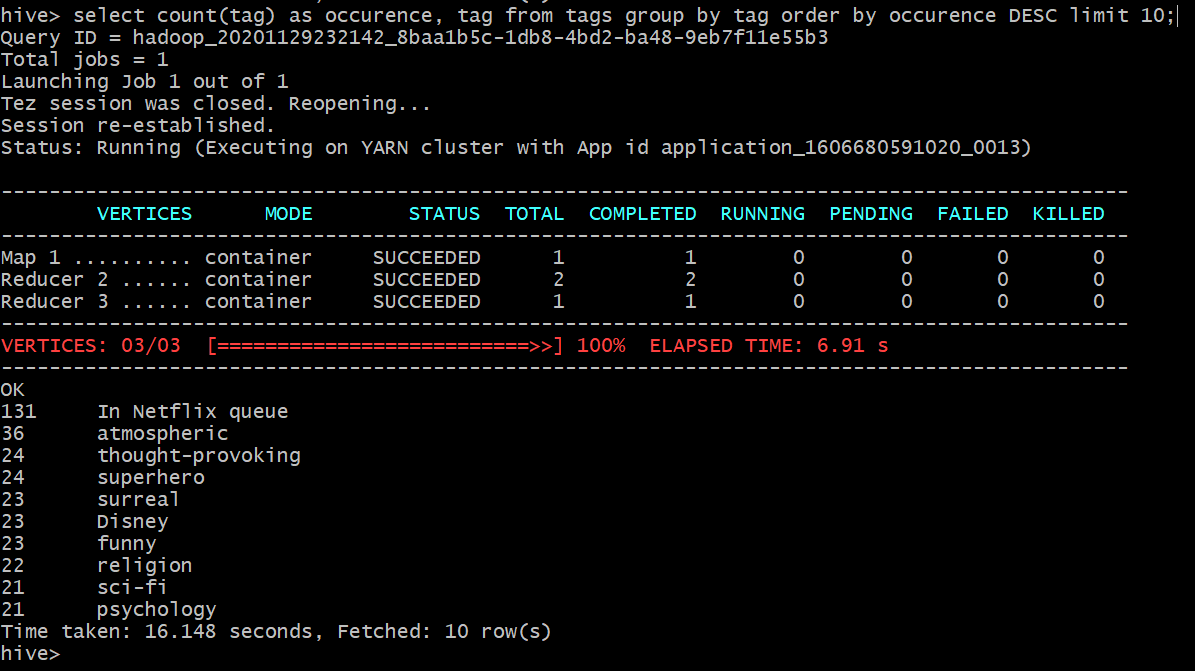
Select moviesml.movieid,movieml.title,COUNT(DISTINCT movies\_rat.userid) as views from moviesml JOIN movies\_rat ON (moviesml.moviesid = movies\_rat.movieid) GROUP BY moviesml.movieid, moviesml.title ORDER BY views DESC LIMIT 10;



1. To get the number of tags (source of rating) for movies (based on highest number of tags)

**Query:**

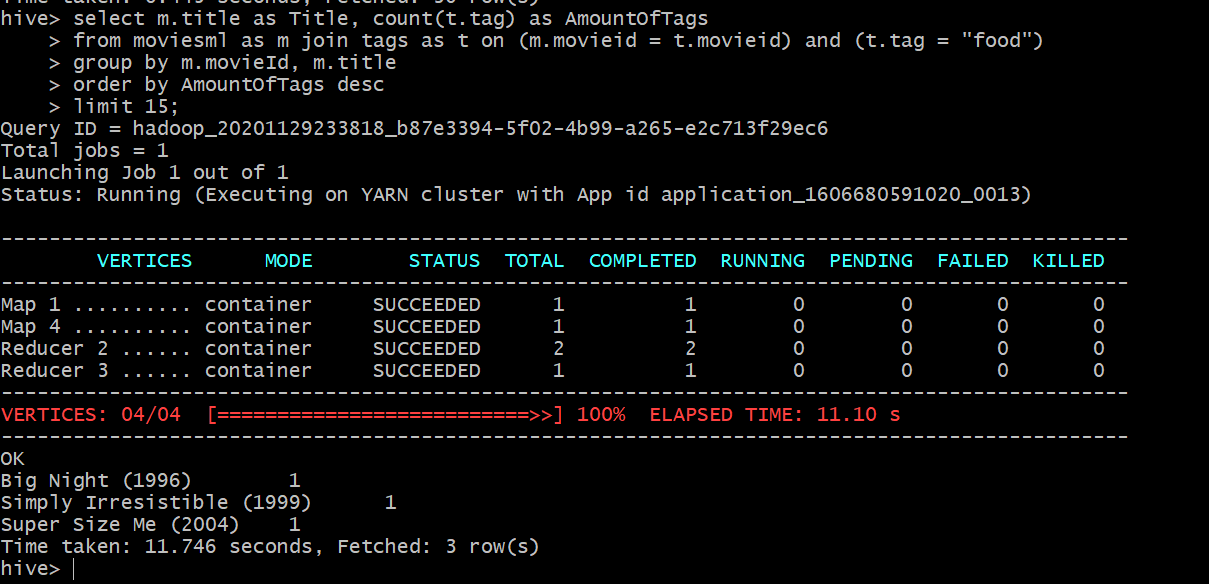
select count(tag) as occurence, tag from tags group by tag order by occurence DESC limit 10;



1. Which 15 movies (titles) have been most frequently tagged with the label "food"?

**Query:**

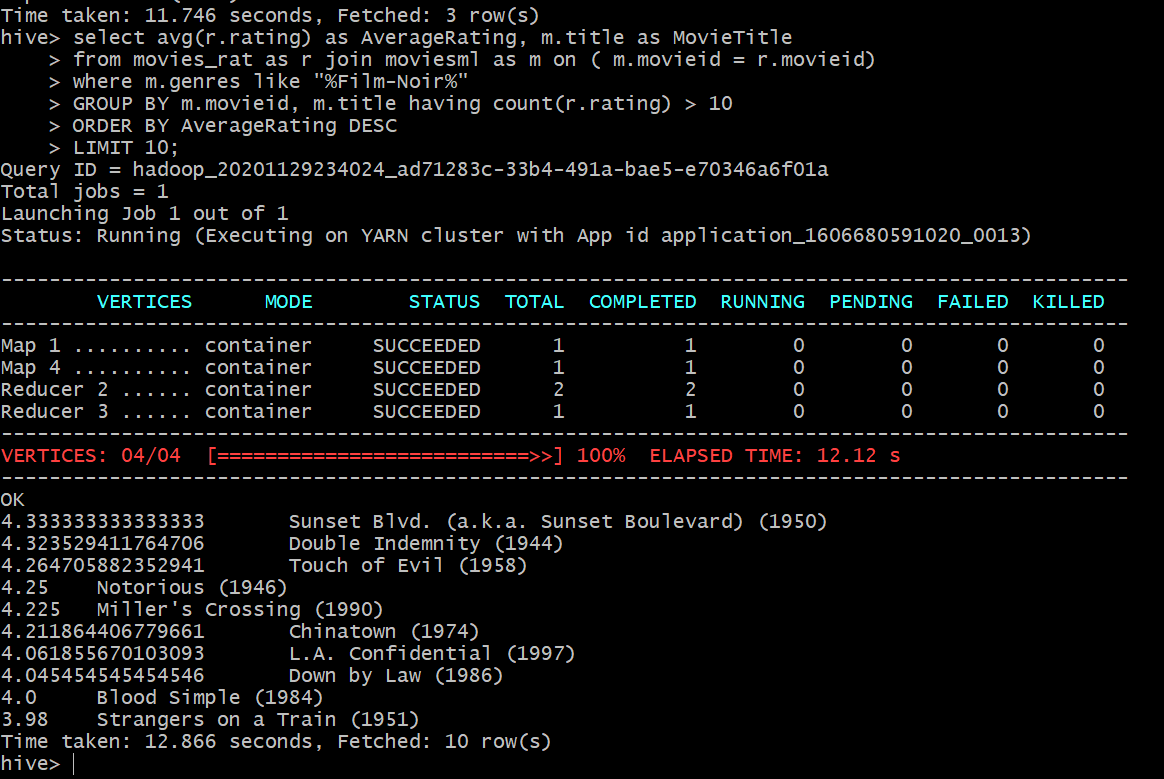
select m.title as Title, count(t.tag) as AmountOfTags from moviesml as m join tags as t on (m.movieid = t.movieid) and (t.tag = "food") group by m.movieId, m.title order by AmountOfTags desc limit 15;



1. Which are the highest-rated "Film-Noir" movies with more than 10 ratings?

**Query:**

select avg(r.rating) as AverageRating, m.title as MovieTitle from movies\_rat as r join moviesml as m on ( m.movieid = r.movieid) where m.genres like "%Film-Noir%" GROUP BY m.movieid, m.title having count(r.rating) > 10 ORDER BY AverageRating DESC LIMIT 10;



1. Which are the 10 best-rated movies (on average; list titles) with more than 1000 ratings?

**Query:**

select avg(r.rating) as AverageRating, m.title as MovieTitle from movies\_rat as r join moviesml as m on ( m.movieid = r.movieid) GROUP BY m.movieid, m.title having count(r.rating) > 100 ORDER BY AverageRating DESC LIMIT 10;

