CIND719 - BIG DATA ANALYTICS TOOLS

Mohammed Amir

Assignment # 1 – Programming HIVE

October 23, 2016

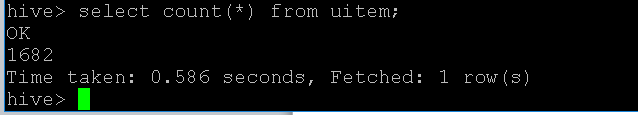
1. How many records are there in both tables?

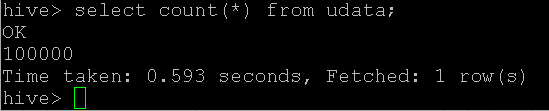
* SELECT COUNT(\*) FROM UITEM

1682 records in uitem table

* SELECT COUNT(\*) FROM UDATA

100000 records in udata table





2. Find the name of all movies released in 1990.

SELECT movie\_title

From uitem

Where year(release\_date) = 1990;



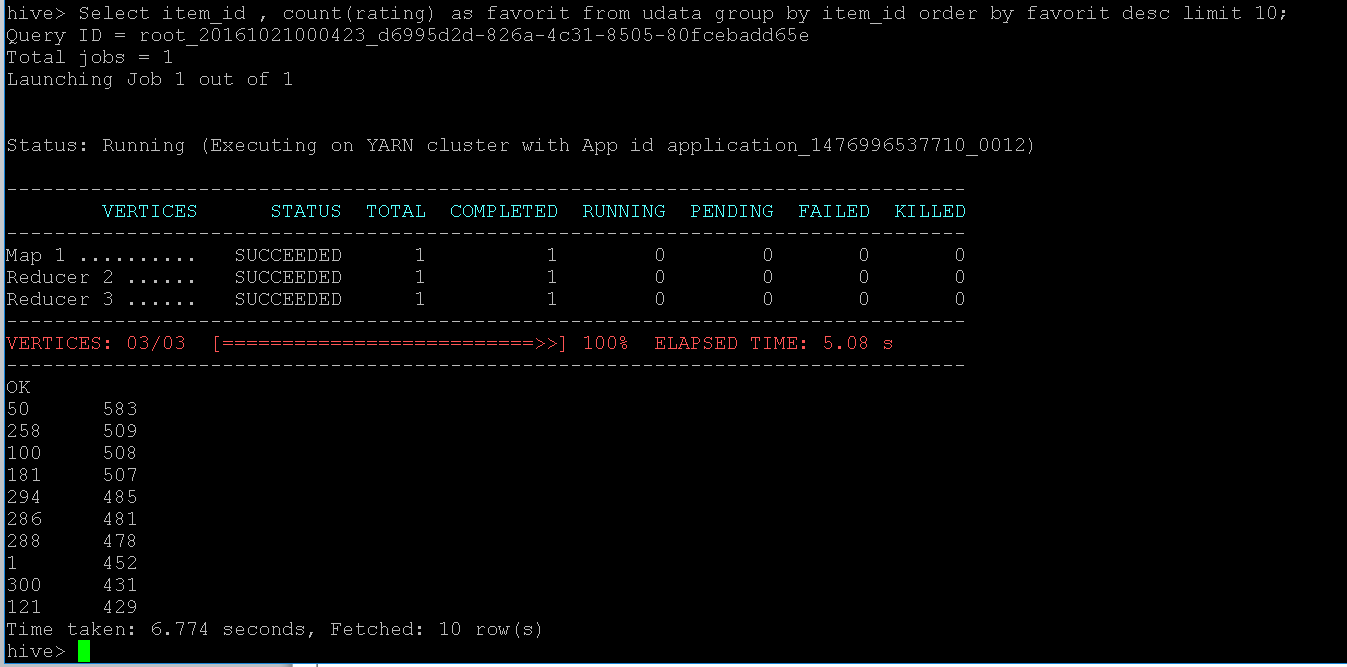
3. List the movieid of the 10 films that received the most ratings (not necessarily highest rating) in the table you created from u.data.

SELECT item\_id, count(rating) as favorit

FROM udata

GROUP BY item\_id

ORDER BY favorit desc limit 10;



4. Use a join to list the titles of the movies you found in step 3.

select count(a.rating) as favorit,

b.movie\_title from udata as a,

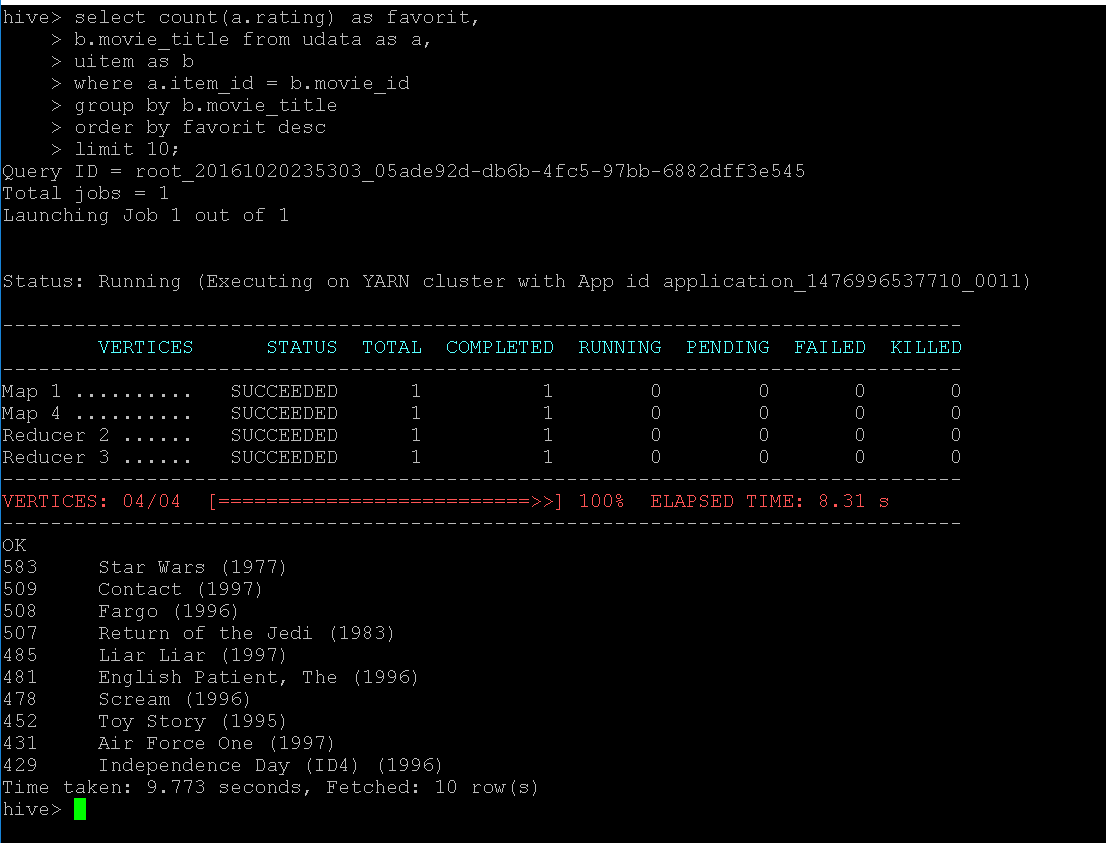
uitem as b

where a.item\_id = b.movie\_id

group by b.movie\_title

order by favorit desc

limit 10;



5. Find the highest rated sci\_fi movie. Explain how you define "highest rating".

* The highest rated sci\_fi movie is one that has the most 5 star rating. The movie that has the highest count of 5 start rating in sci\_fi category is Star Wars with 325 counts.

select b.movie\_title as Movie\_Title, count(a.rating) as Total\_5\_Star\_Rating

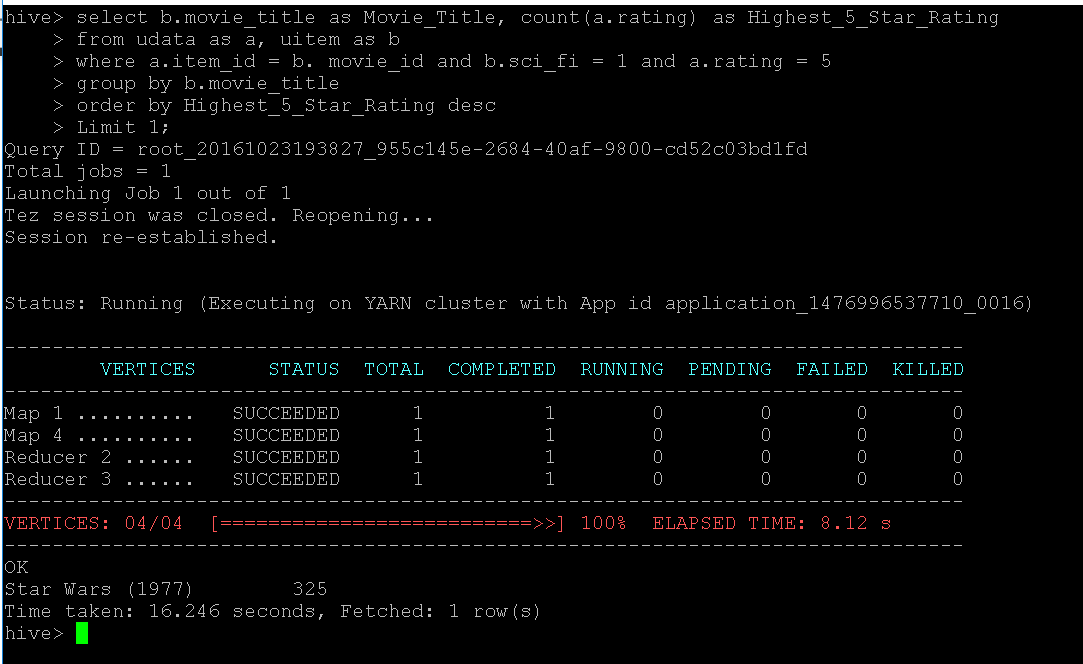
from udata as a, uitem as b

where a.item\_id = b. movie\_id and b.sci\_fi = 1 and a.rating = 5

group by b.movie\_title

order by Total\_5\_Star\_Rating desc

Limit 1;



6. BONUS: Are there any movies with no ratings? (Hint: outer join and IS NULL)

* All movies have ratings.

select b.movie\_title as Movie\_Title, a.rating

from udata as a

LEFT OUTER JOIN uitem as b

ON a.item\_id = b. movie\_id

where a.rating IS NULL;

