Creating Tables and Loading Data:

u.data:

- create table **udata** (userid bigint, itemid bigint, rating smallint, time bigint) row format delimited fields terminated by '\t';
- load data inpath '/user/root/u.data' overwrite into table udata;

u.user:

- create table **uuser** (userid bigint, age int, gender string, occupation string, zipcode string) row format delimited fields terminated by '|';
- load data inpath '/user/root/u.user' overwrite into table uuser;

Answer 1

```
SELECT a.rating, count(distinct(b.userid))
FROM uuser b
JOIN udata a on a.userid = b.userid
WHERE a.rating = 5 and b.gender='F'
GROUP BY rating;
```

root@sandbox:~

```
hive> SELECT a.rating, count(distinct(b.userid))
    > FROM uuser b
    > JOIN udata a on a.userid = b.userid
    > WHERE a.rating = 5 and b.gender='F'
    > GROUP BY rating;
Query ID = root_20201005195139_bdffcc14-20d4-4353-bd6a-a8df6c606570
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1601833207261
_0007)

--------------------------------------------------------------------------------
        VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------------
Map 1 ..........   SUCCEEDED      1          1        0        0       0       0
Map 2 ..........   SUCCEEDED      1          1        0        0       0       0
Reducer 3 ......   SUCCEEDED      1          1        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 03/03  [==========================>>] 100%  ELAPSED TIME: 6.24 s
--------------------------------------------------------------------------------
OK
5       271
Time taken: 7.33 seconds, Fetched: 1 row(s)
hive>
```

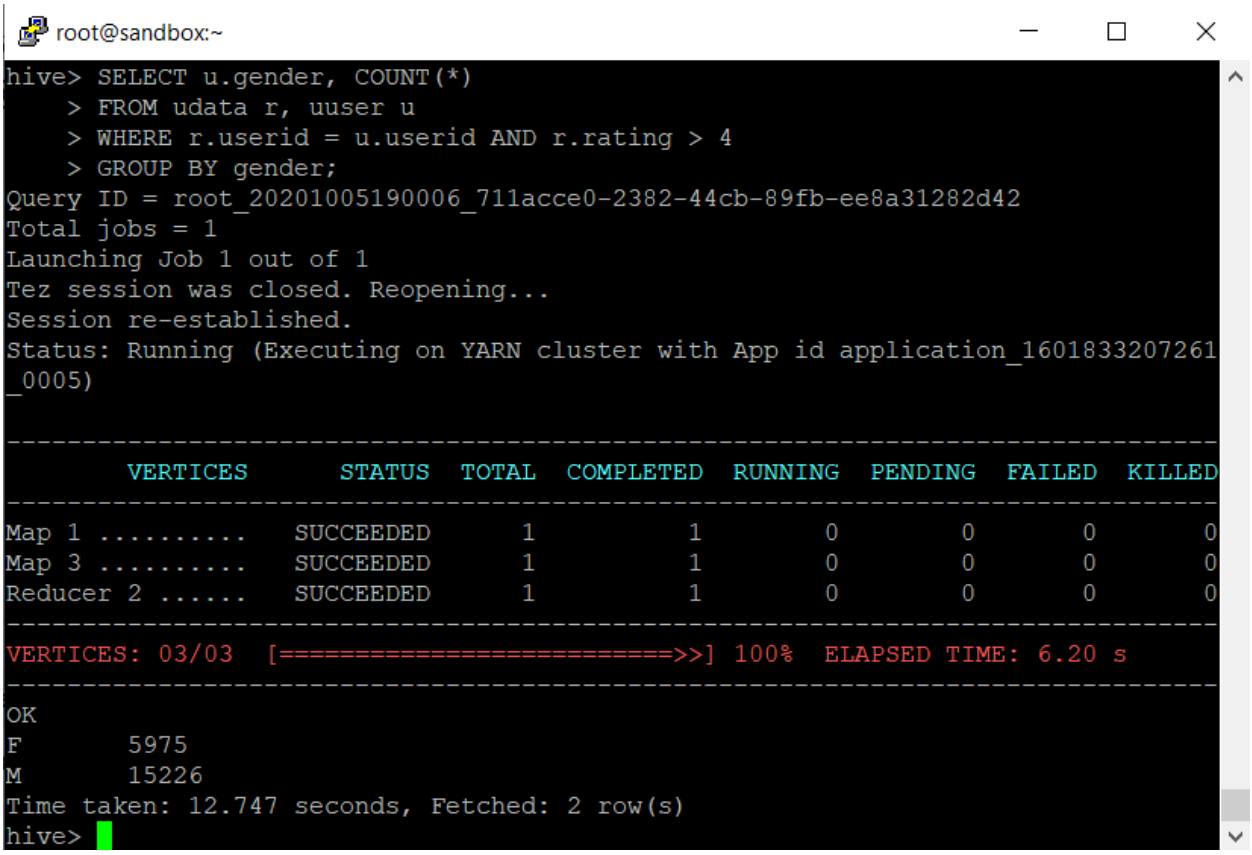
Answer 2

SELECT u.gender, COUNT(*)

FROM udata r, uuser u

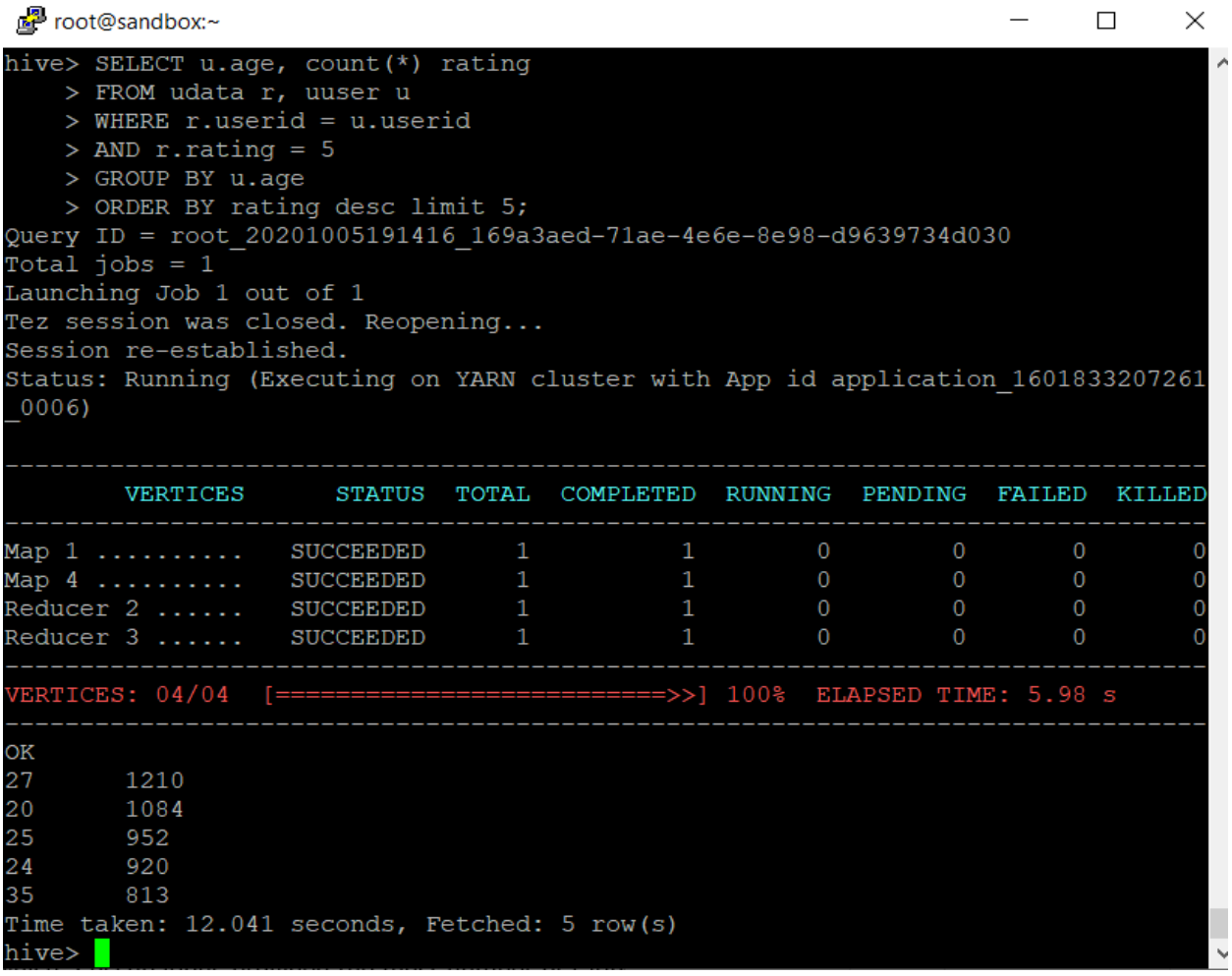
WHERE r.userid = u.userid AND r.rating > 4

GROUP BY gender;



Answer 3

```
SELECT u.age, count(*) rating
FROM udata r, uuser u
WHERE r.userid = u.userid
AND r.rating = 5
GROUP BY u.age
ORDER BY rating desc limit 5;
```



Answer 4

SELECT a.occupation,a.gender, count(b.rating)

AS total_occupation

FROM udata b

JOIN uuser a on a.userid = b.userid where a.gender= 'F'

GROUP BY a.occupation, a.gender

ORDER BY total_occupation desc limit 5;