**Preparing Dataset.**

**Step 1:**

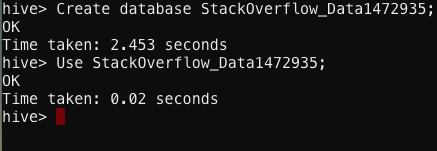
Create Database.

**Command:**

Create database StackOverflow\_Data1472935;

Use StackOverflow\_Data1472935;

**Screenshot:**

****

Step 2:

Create four tables.

**Command:**

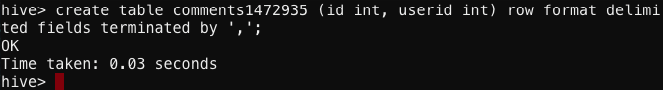
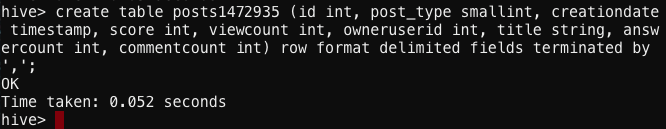
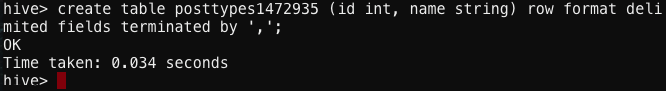
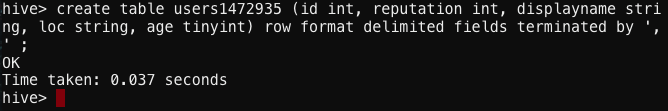
create table comments1472935 (id int, userid int) row format delimited fields terminated by ',';

create table posts1472935 (id int, post\_type smallint, creationdate timestamp, score int, viewcount int, owneruserid int, title string, answercount int, commentcount int) row format delimited fields terminated by ',';

create table posttypes1472935 (id int, name string) row format delimited fields terminated by ',';

create table users1472935 (id int, reputation int, displayname string, loc string, age tinyint) row format delimited fields terminated by ',' ;

**Screenshot:**

Step 3:

Load dataset into created tables.

**Command:**

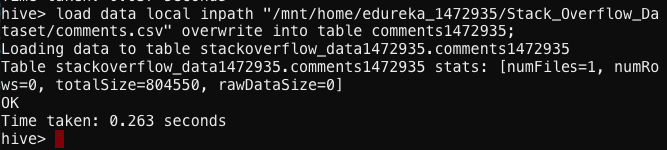
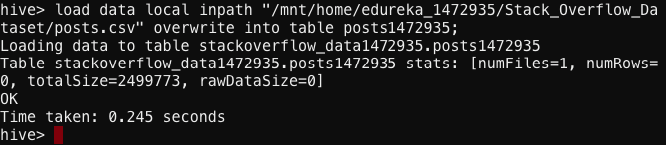
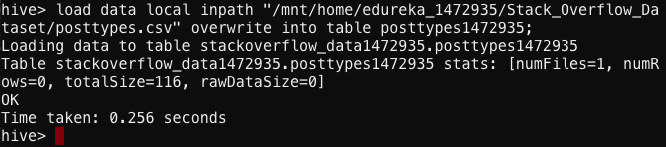
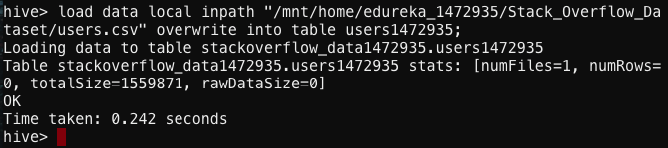
load data local inpath "/mnt/home/edureka\_1472935/Stack\_Overflow\_Dataset/comments.csv" overwrite into table comments1472935;

load data local inpath "/mnt/home/edureka\_1472935/Stack\_Overflow\_Dataset/posts.csv" overwrite into table posts1472935;

load data local inpath "/mnt/home/edureka\_1472935/Stack\_Overflow\_Dataset/posttypes.csv" overwrite into table posttypes1472935;

load data local inpath "/mnt/home/edureka\_1472935/Stack\_Overflow\_Dataset/users.csv" overwrite into table users1472935;

**Screenshots:**

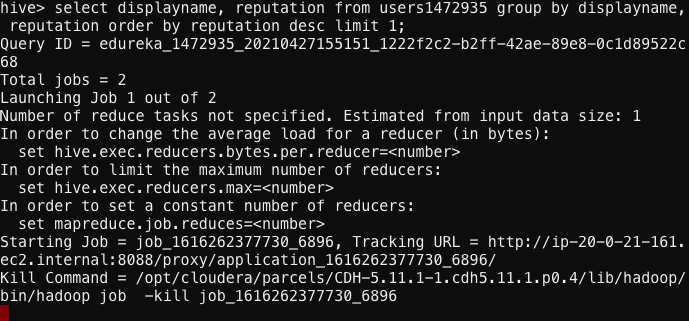
**Solutions:**

1. **Find the display name and no. of posts created by the user who has got maximum reputation.**

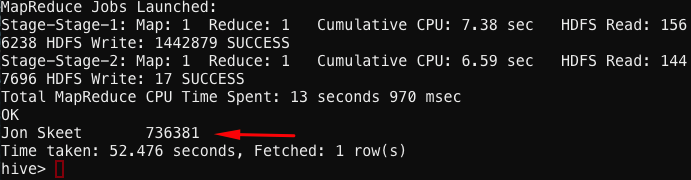
**Command:**

select displayname, reputation from users1472935 group by displayname, reputation order by reputation desc limit 1;

**Screenshots:**



**Result: Jon Skeet with 736381 posts.**

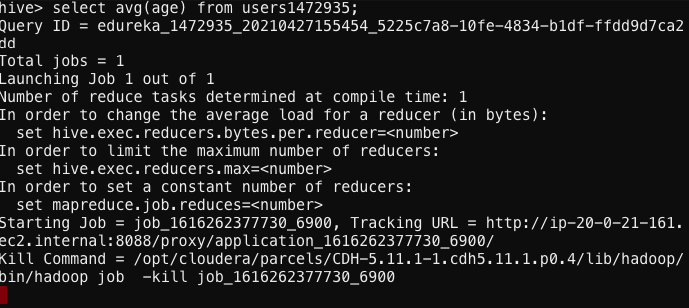


1. **Find the average age of users on the Stack Overflow site.**

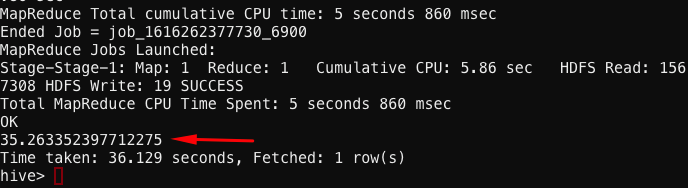
**Command:**

select avg(age) from users1472935;

**Screenshots:**

****

**Results: Average age is 35.**

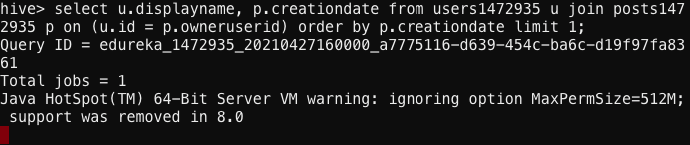
****

1. **Find the display name of user who posted the oldest post on Stack Overflow (in terms of date).**

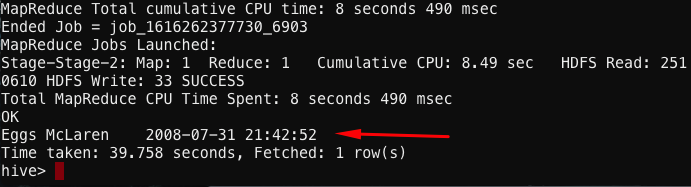
**Command:**

select u.displayname, p.creationdate from users1472935 u join posts1472935 p on (u.id = p.owneruserid) order by p.creationdate limit 1;

**Screenshots:**

****

**Result: Eggs McLaren posted the post on 2008-7-31-21:42:52.**

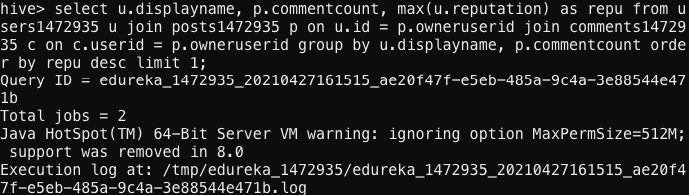


1. **Find the display name and no. of comments done by the user who has got maximum reputation.**

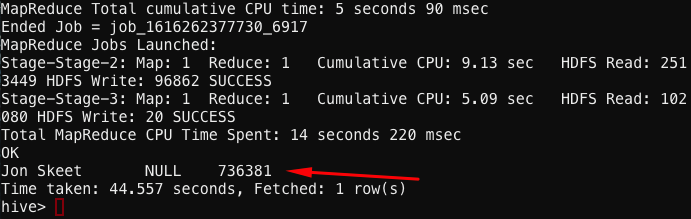
**Command:**

select u.displayname, p.commentcount, max(u.reputation) as repu from users1472935 u join posts1472935 p on u.id = p.owneruserid join comments1472935 c on c.userid = p.owneruserid group by u.displayname, p.commentcount order by repu desc limit 1;

**Screenshots:**

****

**Result:**

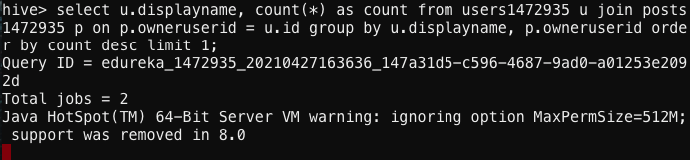


1. **1) Find the display name of user who has created maximum no. of posts on Stack Overflow.**

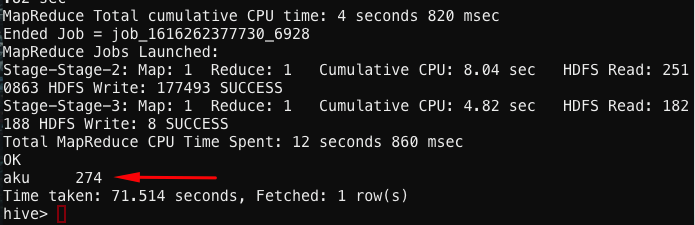
**Command:**

select u.displayname, count(\*) as count from users1472935 u join posts1472935 p on p.owneruserid = u.id group by u.displayname, p.owneruserid order by count desc limit 1;

**Screenshots:**

****

**Result:**

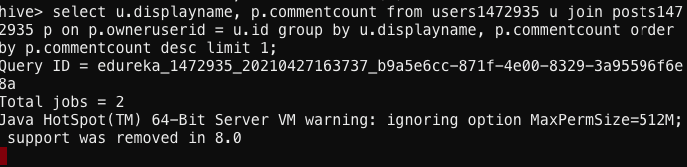


**2) Find the display name of user who has commented maximum no. of posts on Stack Overflow.**

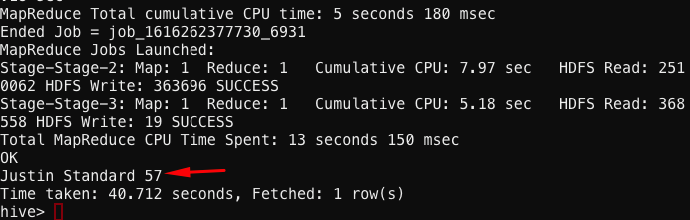
**Command:**

select u.displayname, p.commentcount from users1472935 u join posts1472935 p on p.owneruserid = u.id group by u.displayname, p.commentcount order by p.commentcount desc limit 1;

**Screenshots:**

****

**Result:**

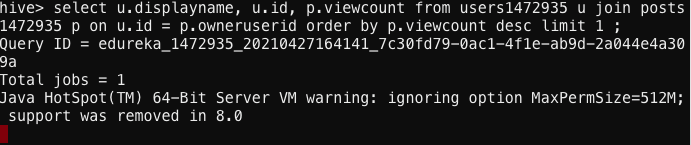


1. **Find the owner name and id of user whose post has got maximum no. of view counts so far.**

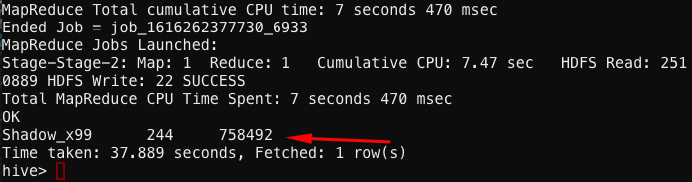
**Command:**

select u.displayname, u.id, p.viewcount from users1472935 u join posts1472935 p on u.id = p.owneruserid order by p.viewcount desc limit 1 ;

**Screenshots:**

****

**Result: Owner name: Shadow\_x99, ID: 244, viewcounts: 758492**

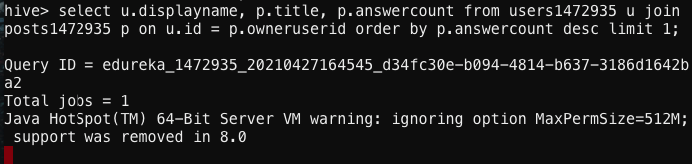


1. **1) Find the title and owner name of post who has got maximum no. of Answer count.**

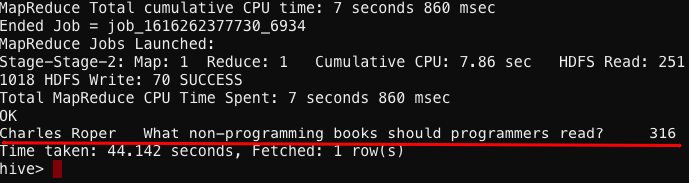
**Command:**

select u.displayname, p.title, p.answercount from users1472935 u join posts1472935 p on u.id = p.owneruserid order by p.answercount desc limit 1;

**Screenshots:**

****

**Result:**

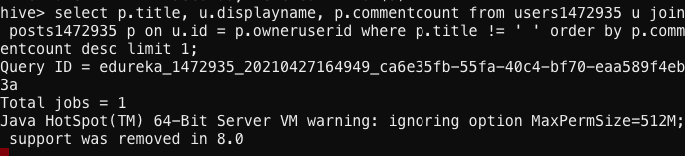
****

**2) Find the title and owner name of post who has got maximum no. of Comment count.**

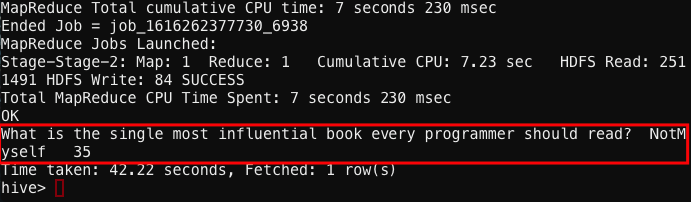
**Command:**

select p.title, u.displayname, p.commentcount from users1472935 u join posts1472935 p on u.id = p.owneruserid where p.title != '' order by p.commentcount desc limit 1;

**Screenshots:**

****

**Result:**

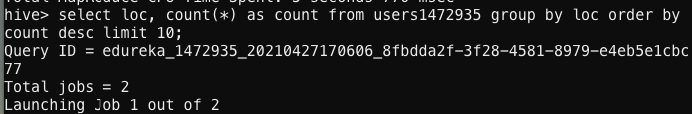


1. **Find the location which has maximum no of Stack Overflow users.**

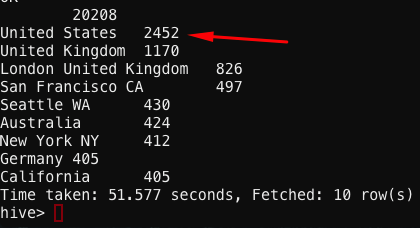
**Command:**

select loc, count(\*) as count from users1472935 group by loc order by count desc limit 10;

**Screenshots:**

****

**Result:**



1. **Find the total no. of answers, posts, comments created by Indian users.**

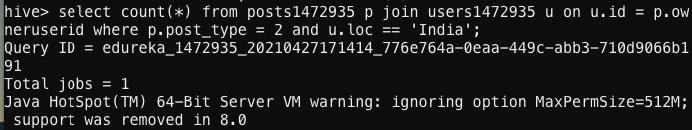
**Command:**

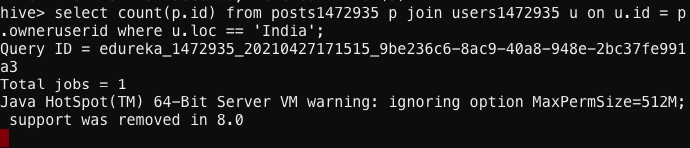
select count(\*) from posts1472935 p join users1472935 u on u.id = p.owneruserid where p.post\_type = 2 and u.loc == 'India';

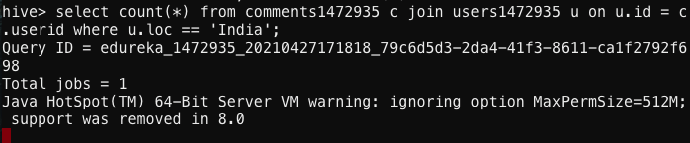
select count(p.id) from posts1472935 p join users1472935 u on u.id = p.owneruserid where u.loc == 'India';

select count(\*) from comments1472935 c join users1472935 u on u.id = c.userid where u.loc == 'India';

**Screenshots:**

****

****

****

**Results:**

