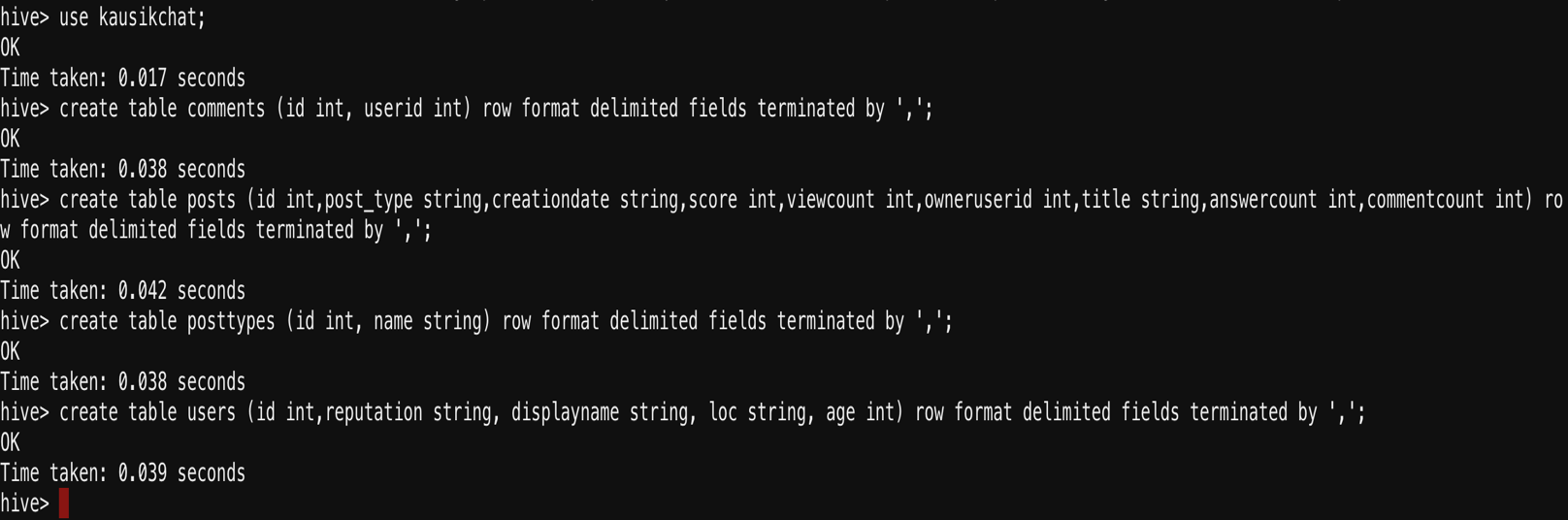
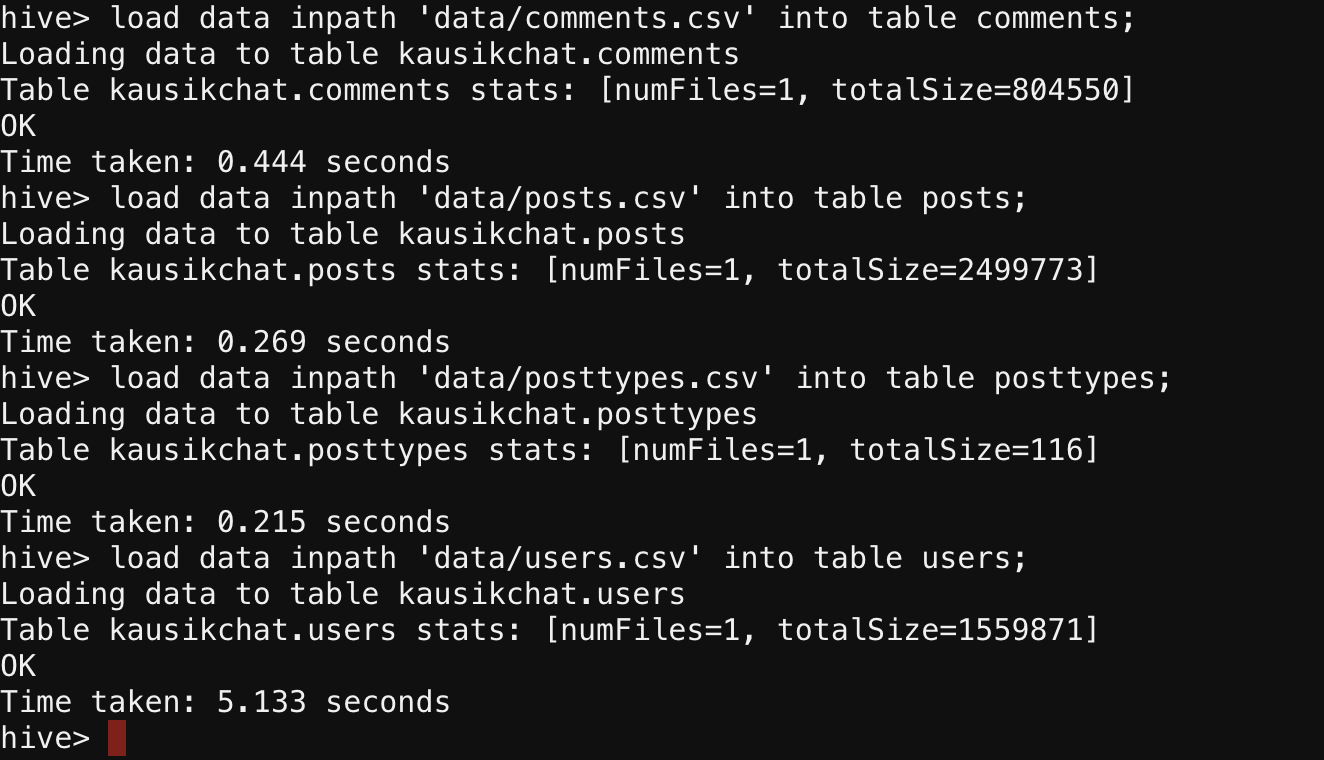
**Analysis of Stack Overflow Data using Hadoop-HIVE**

Creation of HIVE tables:

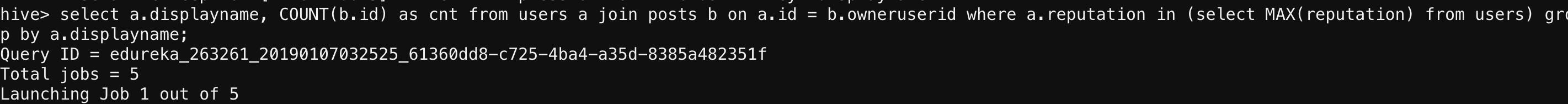


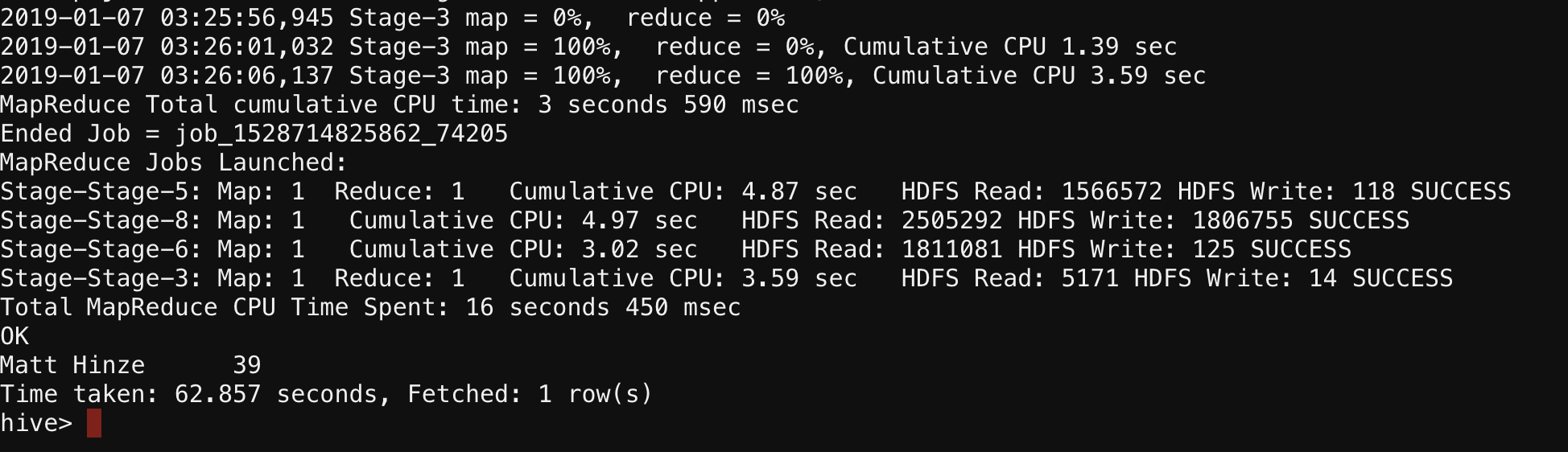
Loading data to tables:



Problem Statement

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





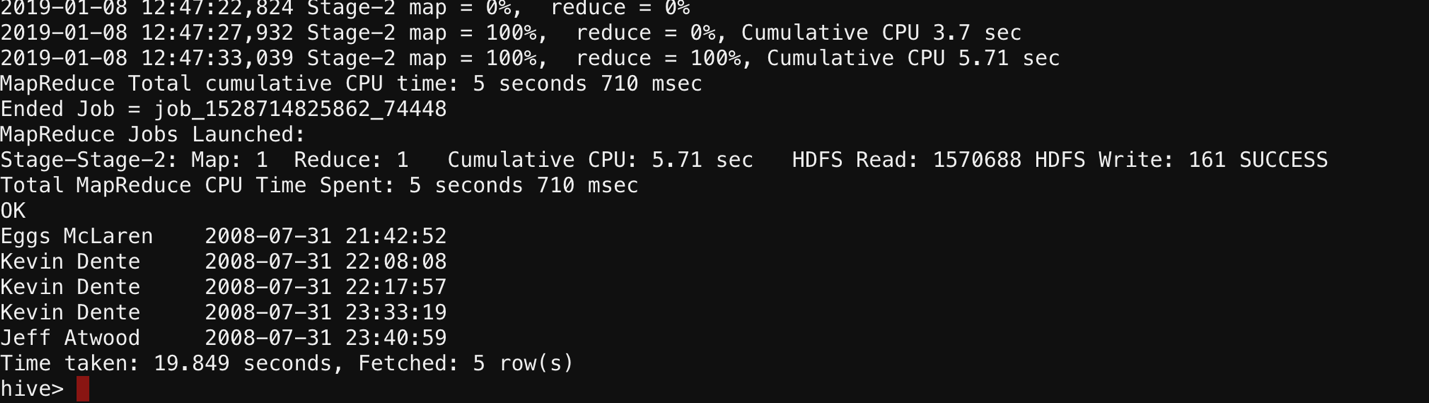
1. Find the average age of users on the Stack Overflow site.

select AVG(age) from users;



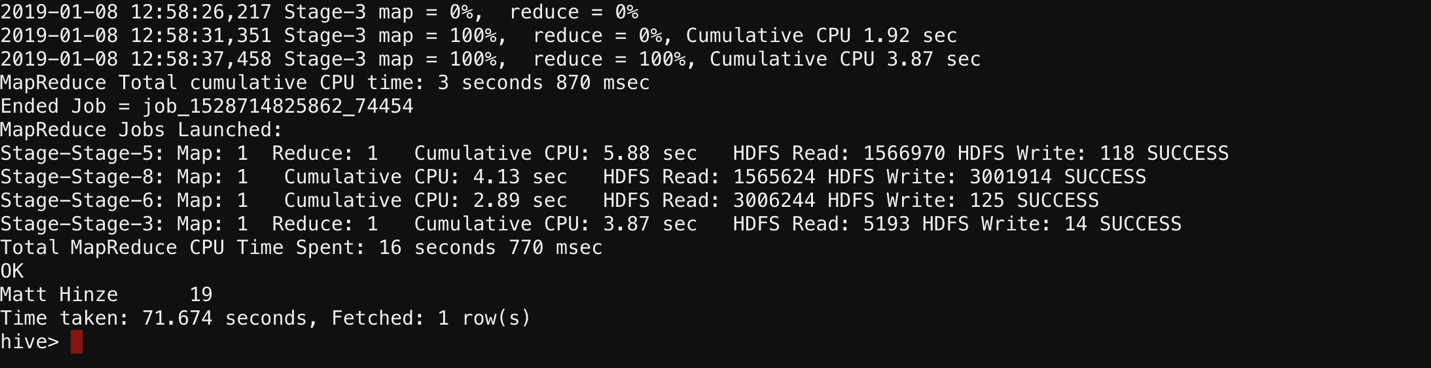
1. Find the display name of user who posted the oldest post on Stack Overflow (in terms of date).

select a.displayname, b.creationdate from users a left outer join posts b on a.id = b.owneruserid where b.creationdate is not null order by b.creationdate asc limit 5;



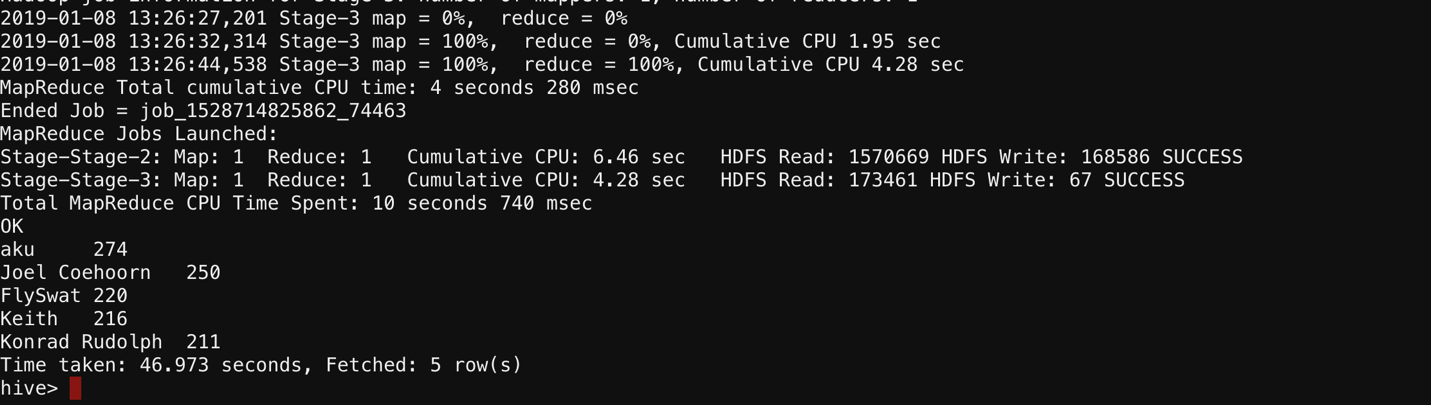
1. Find the display name and no. of comments done by the user who has got maximum reputation.

select a.displayname, SUM(b.commentcount) from users a left outer join posts b on a.id = b.owneruserid where a.displayname is not null and a.reputation in (select MAX(reputation) from users where displayname is not null) group by a.displayname;



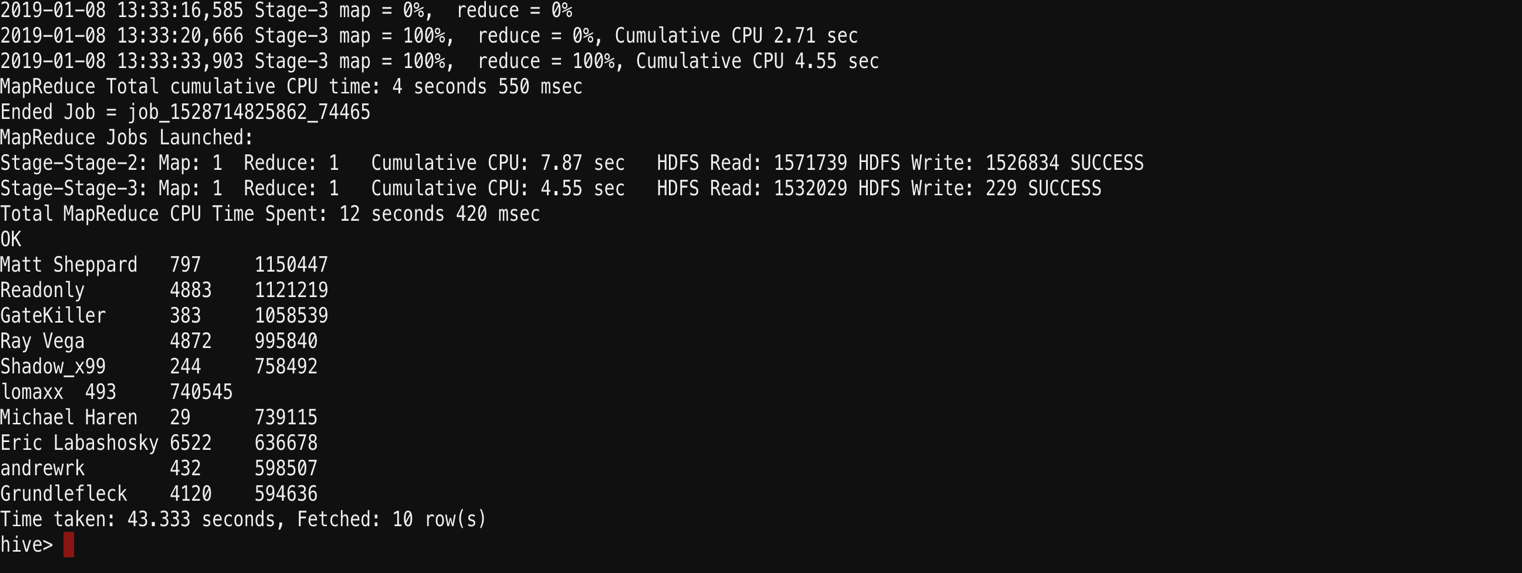
1. Find the display name of user who has created maximum no. of posts on Stack Overflow. Overflow.

select a.displayname , COUNT(b.id) cnt from users a left outer join posts b on a.id = b.owneruserid where a.displayname is not null and b.id is not null group by a.displayname order by cnt desc limit 5;



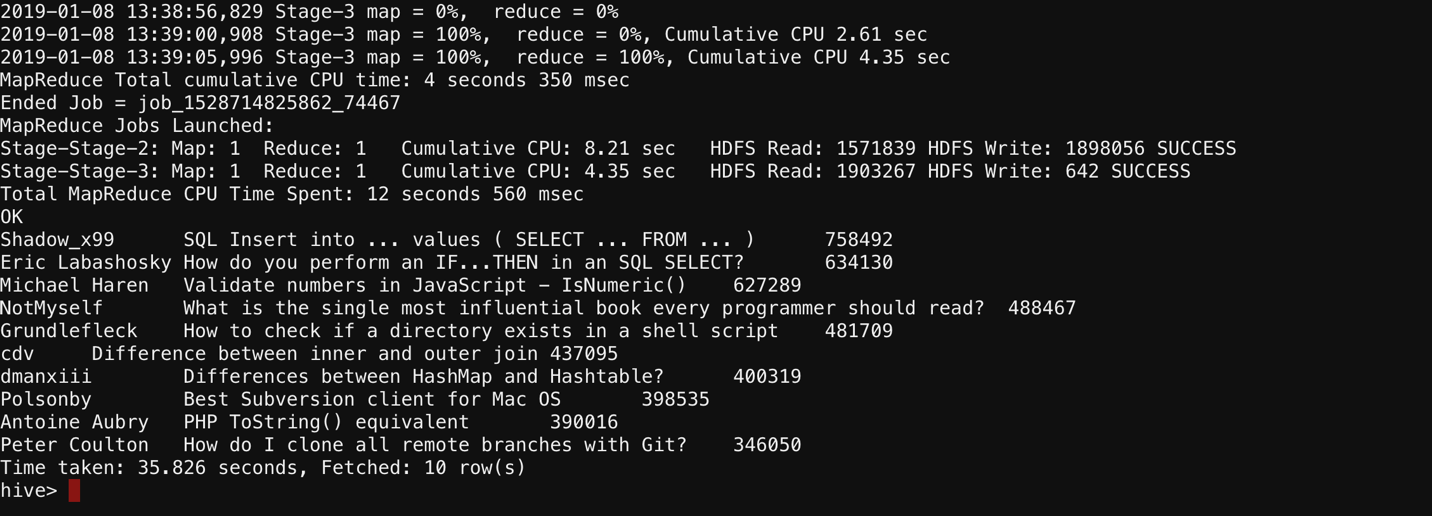
1. Find the owner name and id of user whose post has got maximum no. of view counts so far.

select a.displayname, a.id, SUM(b.viewcount) sum from users a left outer join posts b on a.id = b.owneruserid where a.id is not null group by (a.displayname,a.id) order by sum desc limit 10;



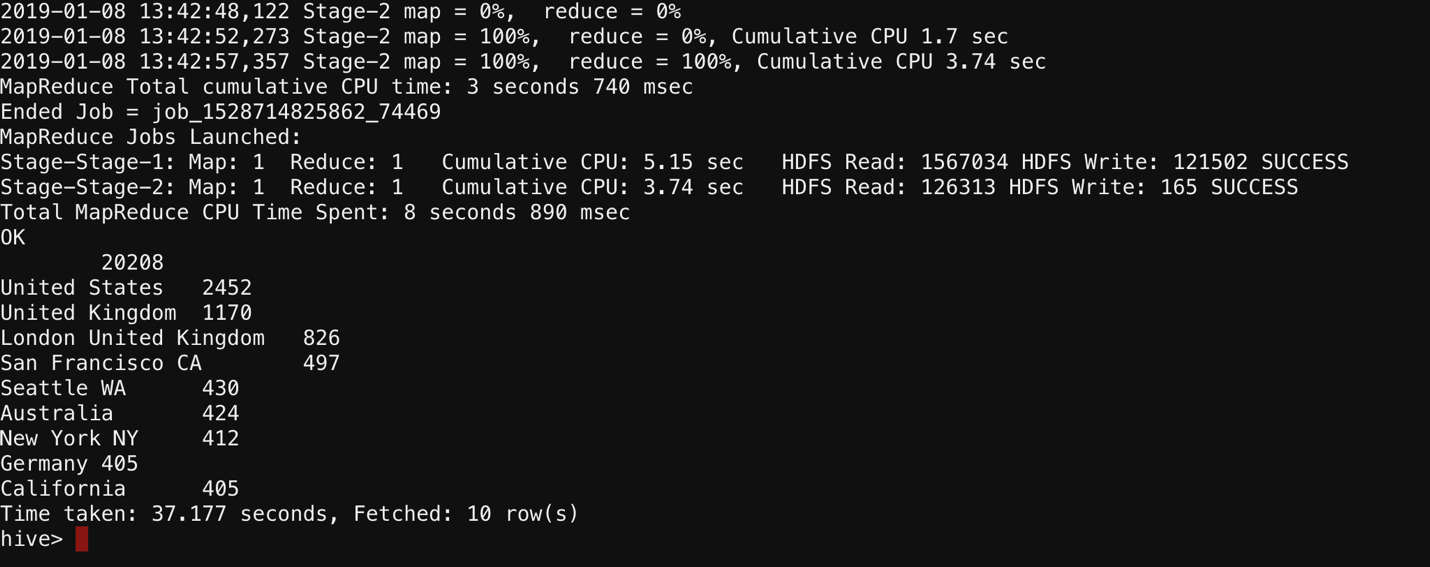
1. Find the title and owner name of the post which has maxim  
   Find the title and owner name of post who has got maximum no. of Comment count.

select a.displayname, b.title , SUM(b.viewcount) sum from users a left outer join posts b on a.id = b.owneruserid where a.id is not null group by a.displayname, b.title order by sum desc limit 10;



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

select loc , COUNT(id) cnt from users where id is not null group by loc order by cnt desc limit 10;



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

select a.loc, SUM(b.answercount) answer, SUM(b.commentcount) comment, COUNT(b.id) cnt from users a left outer join posts b where a.id is not null and a.loc = 'India' group by a.loc order by answer desc,comment desc, cnt desc;

