Stack overflow Data Analysis

create table comments(id int, userid int) row format delimited fields terminated by "," tblproperties("skip.header.line.count"="1") ;

load data local inpath 'comments.csv' into table comments;

create table posttypes(id int, name string) row format delimited fields terminated by "," tblproperties("skip.header.line.count"="1");

load data local inpath 'posttypes.csv' into table posttypes;

create table user(id int, reputation int, displyname string, loc string, age string) row format delimited fields terminated by "," tblproperties("skip.header.line.count"="1");

load data local inpath 'users.csv' into table user;

create table post(id int, posttype int, creationdate timestamp, score int,viewcount int,owneruserid int,title string,answercount int, commentcount int) row format delimited fields terminated by "," tblproperties("skip.header.line.count"="1");

load data local inpath 'posts.csv' into table post;

1. select u.displyname,u.reputation, score from post p join user u on p.owneruserid=u.id order by reputation desc limit 1;

1. select AVG(AGE) FROM user;

1. select u.displyname,p.creationdate from user u join post p on u.id=p.owneruserid order by creationdate limit 1;

1. select u.displyname,u.reputation, commentcount from post p join user u on p.owneruserid=u.id order by reputation desc limit 1;

1. select u.displyname,count(p.id) as cnt from post p join user u on p.owneruserid=u.id group by u.displyname order by cnt desc limit 1;

select u.displyname,sum(p.commentcount) as cnt from post p join user u on p.owneruserid=u.id group by u.displyname order by cnt desc limit 1;

1. select u.displyname,u.id,sum(p.viewcount) as cnt from post p join user u on p.owneruserid=u.id group by u.displyname,u.id order by cnt desc limit 1;

1. select p.title,u.displyname,p.answercount from post p join user u on p.owneruserid=u.id order by p.answercount desc limit 1;

1. select p.title,u.displyname,p.commentcount from post p join user u on p.owneruserid=u.id order by p.commentcount desc limit 1;

1. select loc, count(id)as cnt from user group by loc having loc!="" order by cnt desc limit 1;

create table indianuser(loc string, answer int,post int,comment int) row format delimited fields terminated by ",";

insert into indianuser select u.loc, sum(p.answercount),count(p.id),sum(p.commentcount) from post p join user u on p.owneruserid=u.id where u.loc like "%India" group by u.loc;

select sum(answer),sum(post),sum(comment) from indianuser;