**Hive Queries:**

* ***Most viewed questions in each category***

select id,tags.TagName,questionpost.ViewCount from questionpost,tags,(select tags.TagName,max(ViewCount) as tagcount from questionpost,tags where questionpost.Tags like concat('%',tags.TagName,'%') group by tags.TagName)as testview where testview.TagName=tags.TagName and testview.tagcount=questionpost.ViewCount and questionpost.Tags like concat('%',tags.TagName,'%') and tags.TagName like "%java%" order by tags.TagName;

* ***Most Scored Question in each category***

select id,tags.TagName,questionpost.score from questionpost,tags,(select tags.TagName,max(questionpost.score) as maxscore from questionpost,tags where questionpost.Tags like concat('%',tags.TagName,'%') group by tags.TagName) as tagscore where tags.TagName=tagscore.TagName and questionpost.score=tagscore.maxscore and questionpost.Tags like concat('%',tags.TagName,'%') and tags.TagName like "%java%" order by tags.TagName;

* ***How many upvotes for each tag***

select taguser.TagName,(sum(upvotes))as upvotes from users,

(select tags.TagName,questionpost.owneruserid from questionpost,tags where questionpost.Tags like concat('%',tags.TagName,'%') order by TagName) taguser where taguser.owneruserid=users.id and taguser.TagName like "%java%" group by taguser.TagName;

* ***Users with highest accept rate of their answers***

SELECT u.Id,COUNT(\*) AS NumAnswers, SUM(CASE WHEN q.AcceptedAnswerId = a.Id THEN 1 ELSE 0 END) AS NumAccepted,(SUM(CASE WHEN q.AcceptedAnswerId = a.Id THEN 1 ELSE 0 END)\*100.0/COUNT(\*)) AS AcceptedPercent FROM answerpost a,users u,questionpost q WHERE u.Id=a.OwnerUserId AND a.ParentId= q.Id GROUP BY u.Id HAVING COUNT(\*)>= 5 ORDER BY AcceptedPercent DESC, NumAnswers DESC;

* ***Users whose post has a maximum comment count***

select Users.Id,DisplayName,QuestionPost.id,CommentCount from Users,QuestionPost,(select max(commentcount) as maxcommentcount from QuestionPost) as MaxQuestionPost where QuestionPost.owneruserid=Users.Id and commentcount = MaxQuestionPost.maxcommentcount;

* ***Users having top 5 gold badges***

select Users.Id,collect\_set(Users.DisplayName) as displayname,sum(UsersHave.count) as goldbadgecount from Users,UsersHave where Users.Id=UsersHave.Id and class=1 group by Users.Id order by goldbadgecount desc limit 5;

* ***Most contraversial questionposts***

select id,up,down,score from questionpost,(select postid,sum( case when VoteTypeId=2 then 1 else 0 end) as up,sum( case when VoteTypeId=3 then 1 else 0 end) as down from Votes where VoteTypeId in (2,3) group by PostId order by down desc) as votestats where votestats.postid=questionpost.id and (votestats.down) > (2\*votestats.up) and (questionpost.score) <= (select avg(score) from questionpost) order by up desc;

* ***TOP 20 most viewed questions in each category***

In this query, for every tag name view count is calculated and based on the viewcount,top 20 most viewed questions are displayed.Output consists of title of the question,with which tag it belongs to and the viewcount of that question.

select id,TagName,ViewCount from questionpost,tags where (TagName,ViewCount) in (select tags.TagName,max(ViewCount) from questionpost,tags where questionpost.Tags like concat(%,tags.TagName,%) group by tags.TagName) and questionpost.Tags like concat(%,tags.TagName,%) order by TagName ;

* ***Top 20 Valuable users in the database***

Under the consideratin that valuable users as anyone who has upvotes-downvotes count greater than the average difference of upvotes and downvotes and who has total no of badges greater than average no of badges in the data.For every user on checking the above mentioned two conditions and if the user satisfies them, he is cconsidered as a valuable user.Output consists of Users DisplayName ,difference of votes as votevount and totalbadges a user has as BadgeCount for all valuable user ids

select users.DisplayName,(UpVotes-DownVotes) as VoteCount,badgecount.BadgeCount from users,(select id,sum(count) as BadgeCount from usershave group by id) as badgecount where users.id=badgecount.id and (users.UpVotes-users.DownVotes)>=(select avg(users.UpVotes-users.DownVotes) from users)and badgecount.BadgeCount >=(select avg(count) as BadgeCount from usershave);

* ***Top 20 Most contraversial questionposts***

The most contraversial questionposts are considered as those posts which have a score less than avg score of all posts and whose downvotes are greater than double the upvotes

Output consists of question title with number of upvotes, downvotes and score.

select id,up,down,score from questionpost,(select postid,sum( case when VoteTypeId=2 then 1 else 0 end) as up,sum( case when VoteTypeId=3 then 1 else 0 end) as down from Votes where VoteTypeId in (2,3) group by PostId order by down desc) as votestats

where votestats.postid=questionpost.id and (votestats.down) > (2\*votestats.up) and (questionpost.score) <= (select avg(score) from questionpost) order by up desc;

* ***Top 30 Users with highest accept rate of their answers***

In this query,users with highest accept rate of their answers are calculated.Accepted percentage is calculated from number of answers a particular user has given to the posts and among them number of answers that are accepted.

The output consists of user display name and their accpted percentage.

select u.Id,count(\*) as NumAnswers, sum(case when q.AcceptedAnswerId = a.Id then 1 else 0 end) AS NumAccepted,(sum(case when q.AcceptedAnswerId = a.Id then 1 else 0 end)\*100.0/count(\*)) AS AcceptedPercent from answerpost a,users u,questionpost q where u.Id=a.OwnerUserId and a.ParentId= q.Id and (q.OwnerUserId <> u.Id or q.OwnerUserId) is NULL group by u.Id having count(\*)>= 5 order by AcceptedPercent desc, NumAnswers desc;

* ***Top 10 countries whose citizen questions are most viewed questions***

In this query, countries whose citizens have most viewed questions are displayed.For every location in users table we see all the users in that location and then for each user in the location , check the question that users posted by connecting with QuestionPost using the UserId and Owneruserid .Then for every user,sum of viewcount of all posts is calculated and repeat the process for all the users in that location

The output consists of Country Name with count of users.

select count(\*) as totalcount,lower(trim((substring\_index(Location,,,-1)))) as country from users where Location is not null and users.id in (select users.id from questionpost,tags,users where (TagName,ViewCount) in (select tags.TagName,max(ViewCount) from questionpost,tags where questionpost.Tags like concat(%,tags.TagName,%) group by tags.TagName) and questionpost.Tags like concat(%,tags.TagName,%) and users.id=questionpost.owneruserid order by TagName) group by country order by totalcount desc limit 5;