**HIVE QUERIES**

**--Creating table to put the data in--**

CREATE TABLE TABLE1(Id INT ,Score INT,Body VARCHAR(10000),OwnerUserId INT, Title STRING,Tags STRING,FavoriteCount INT)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ',';

**--Loading the data into Table--**

LOAD DATA INPATH 'hdfs://ct-project1-m/final\_data.txt' INTO TABLE TABLE1;

Q1 The top 10 posts by score

SELECT Id, Title, Score from TABLE1 order by Score DESC LIMIT 10;

Q2 The top 10 users by post score

select owneruserid, sum(score) as OverallScore from TABLE1 where owneruserid IS NOT NULL group by OwnerUserId order by OverallScore desc limit 10;

Q3 The number of distinct users, who used the word “cloud” in one of their posts

select count (distinct owneruserid) from TABLE1 where (lower(body) like '%cloud%' or lower(title) like '%cloud%' or lower(tags) like '%cloud%');