**Tf-idf**

create temporary macro max2(x INT, y INT) if(x>y,x,y);

create temporary macro tfidf(tf FLOAT, df\_t INT, n\_docs INT) tf \* (log(10, CAST(n\_docs as FLOAT)/max2(1,df\_t)) + 1.0);

create table Distinct\_owner\_Id as SELECT OwnerUserId, SUM(Score) AS TotalScore FROM TABLE1 GROUP BY OwnerUserId ORDER BY TotalScore DESC LIMIT 10;

create table User\_data as Select HT.OwnerUserID,title from TABLE1 HT JOIN Distinct\_owner\_Id DO on HT.OwnerUserID = DO.OwnerUserID;

create or replace view User\_view as select ownerUserId, eachword from User\_data LATERAL VIEW explode(tokenize(title, True)) t as eachword where not is\_stopword(eachword);

create or replace view Temp\_view as select ownerUserid, eachword, freq from (select ownerUserId, tf(eachword) as word2freq from User\_view group by ownerUserId) t LATERAL VIEW explode(word2freq) t2 as eachword, freq;

create or replace view Tf\_final as select \* from (select ownerUserId, eachword, freq,rank() over (partition by ownerUserId order by freq desc) as rn from Temp\_view as t) as t where t.rn<=10 ;

select \* from Tf\_final;