**Link** http://grouplens.org/datasets/movielens/

u.user u.item u.data

1> Create an External table pointing to some location /home/hduser/moviedata

to hold the above 3 details

Udata table:

hive (moverating)> create external table udata(user\_id int,item\_id int,rating int,time string)row format delimited fields terminated by '\t' location '/user/hduser/moviedata/rating\_data';

Itemdata table;

hive (moverating)> create external table itemdata(movieid int,movietitle string,releasedate string,videoreleasedate string,url string,unknow int,action int,adventure int,animation int,children int,comedy int,crime int,documentary int,drama int,fantasy int,film\_noir int,horror int,musical int,mystery int,romance int,sci\_fi int,thriller int,war int,wester int)row format delimited fields terminated by '|' location '/user/hduser/moviedata/itemdata';

Userdata table

Hive(movierating)>create external table userdata(userid int,age int,gender string,occupation string,zip int)row format delimited fields terminated by '|' location '/user/hduser/moviedata/userdata';

2>Find out the movie with highest no of 5 ratings(u data)

hive (moverating)> create table ratingcount as select item\_id,count(rating) as count from udata where rating=5 group by item\_id;

create table ratingcount as select item\_id,count(rating) as count from udata where rating=5 group by item\_id

item\_id count

50 325

3>Find all the female above 30 Years who has seen a movie of type HORROR.

hive (moverating)> select u.userid as users from userdata u join udata ud on u.userid=ud.user\_id join itemdata id on ud.item\_id = id.movieid where u.gender ='F' and u.age >30 and id.horror=1 group by u.userid sort by users;

4>Find all male above 25 years who has seen movie of type Action and Mystry.

hive (moverating)> select u.userid as users from userdata u join udata ud on u.userid=ud.user\_id join itemdata id on ud.item\_id = id.movieid where u.gender ='M' and u.age >25 and id.action=1 and id.mystery=1 group by u.userid sort by users;

5>Find out top 5 movies with highest number of votes with Average votes

select item\_id,count(user\_id) as no\_of\_users,avg(rating) as rate from udata group by item\_id sort by rate desc limit 5;

item\_id no\_of\_users rate

814 1 5.0

1653 1 5.0

1293 3 5.0

1189 3 5.0

1122 1 5.0

6>Find out the category which is being watched most in a particular Area

create table movie\_category as

select movieid,

case

when action = 1 then 'action'

when adventure = 1 then 'adventure'

when animation = 1 then 'animation'

when children = 1 then 'children'

when comedy = 1 then 'comedy'

when crime = 1 then 'crime'

when documentary = 1 then 'documentary'

when drama = 1 then 'drama'

when fantasy = 1 then 'fantasy'

when film\_noir = 1 then 'film\_noir'

when horror = 1 then 'horror'

when musical = 1 then 'musical'

when mystery = 1 then 'mystery'

when romance = 1 then 'romance'

when sci\_fi = 1 then 'sci\_fi'

when thriller = 1 then 'thriller'

when war = 1 then 'war'

when wester = 1 then 'western'

end

as category from itemdata;

hive (moverating)> create table movie\_category1 as select zip,collect\_set(userid) as users,collect\_list(category) as categories from userdata ud join udata u on u.user\_id =ud.userid join movie\_category m on m.movieid = u.item\_id group by ud.zip

hive (moverating)> create table movie\_category2 as select zip,cgry,count(zip) as catgory\_count from movie\_category1 lateral view explode(categories) cat as cgry group by zip,cgry;

hive (moverating)> create table movie\_category3 as select zip,max(catgory\_count) as most\_watch\_catgory from movie\_category2 group by zip;

hive (moverating)> select m.zip,collect\_set(m.cgry) from movie\_category2 m join movie\_category3 m1 on m.zip=m1.zip and m.catgory\_count = m1.most\_watch\_catgory group by m.zip;