**IBM Project**

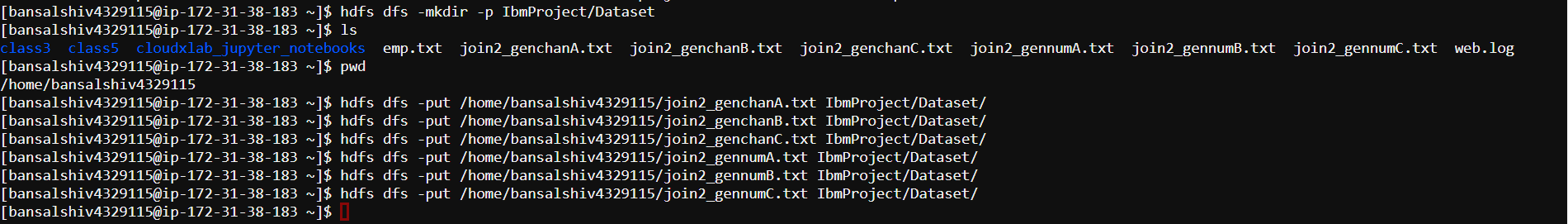
1. Create a directory in hadoop Using command

Hdfs dfs –mkdir -p IbmProject/Dataset

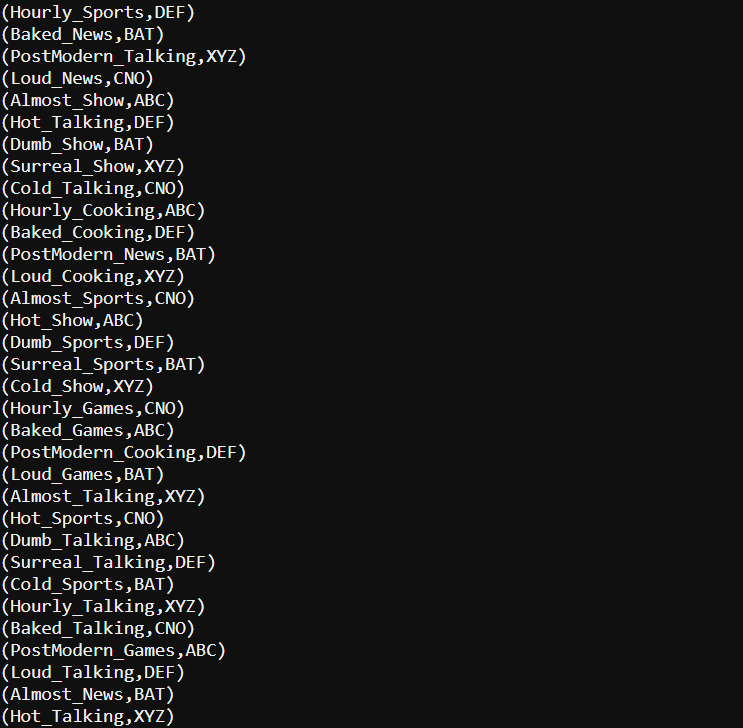
1. Put all the dataset in the directory using command as shown below

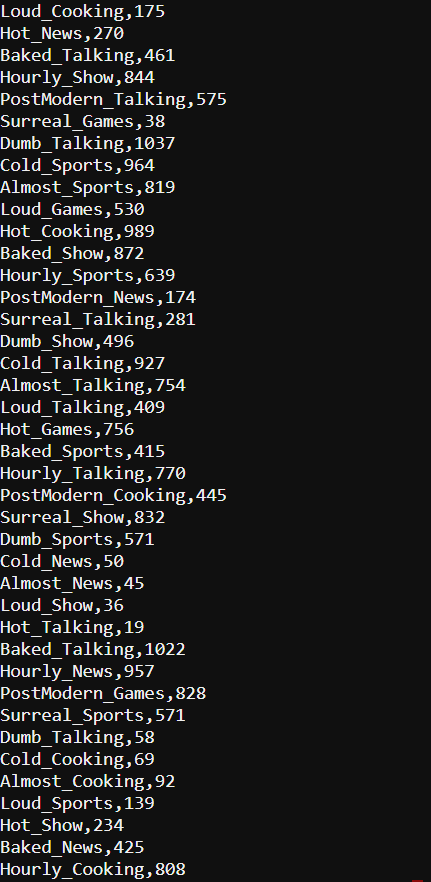
Hdfs dfs –put /home/bansalshiv4329115/join2\_genchan\*.txt IbmProject/Dataset/

Hdfs dfs –put /home/bansalshiv4329115/join2\_genchan\*.txt IbmProject/Dataset/



Here is the dataset Present in hdfs





Question 1

What is the total number of viewers for shows on ABC

Step 1 Load the data using pig

data1 = Load 'IbmProject/Dataset/join2\_genchanA.txt' Using PigStorage(',') As (title:chararray, channel:chararray);

data2 = Load 'IbmProject/Dataset/join2\_genchanB.txt' Using PigStorage(',') As (title:chararray, channel:chararray);

data3 = Load 'IbmProject/Dataset/join2\_genchanC.txt' Using PigStorage(',') As (title:chararray, channel:chararray);

data4 = Load 'IbmProject/Dataset/join2\_gennumA.txt' Using PigStorage(',') As (title:chararray, view:int);

data5 = Load 'IbmProject/Dataset/join2\_gennumB.txt' Using PigStorage(',') As (title:chararray, view:int);

data6 = Load 'IbmProject/Dataset/join2\_gennumC.txt' Using PigStorage(',') As (title:chararray, view:int);

union\_demo = UNION data1,data2 ;

unio\_demo1 = UNION union\_demo,data3;

union\_demo2 = UNION data4,data5;

union\_demo3 = UNION union\_demo2,data6;

filter\_demo = Filter union\_demo1 by channel =='ABC';

join\_demo = JOIN filter\_demo by title,union\_demo3 by title;

group\_demo =Group join\_demo by $0;

sum\_demo = FOREACH group\_demo GENERATE group, SUM(join\_demo.view);

OUTPUT



Question 2

What is the number of viewers for the BAT channel?

data1 = Load 'IbmProject/Dataset/join2\_genchanA.txt' Using PigStorage(',') As (title:chararray, channel:chararray);

data2 = Load 'IbmProject/Dataset/join2\_genchanB.txt' Using PigStorage(',') As (title:chararray, channel:chararray);

data3 = Load 'IbmProject/Dataset/join2\_genchanC.txt' Using PigStorage(',') As (title:chararray, channel:chararray);

data4 = Load 'IbmProject/Dataset/join2\_gennumA.txt' Using PigStorage(',') As (title:chararray, view:int);

data5 = Load 'IbmProject/Dataset/join2\_gennumB.txt' Using PigStorage(',') As (title:chararray, view:int);

data6 = Load 'IbmProject/Dataset/join2\_gennumC.txt' Using PigStorage(',') As (title:chararray, view:int);

union\_demo = UNION data1,data2 ;

unio\_demo1 = UNION union\_demo,data3;

union\_demo2 = UNION data4,data5;

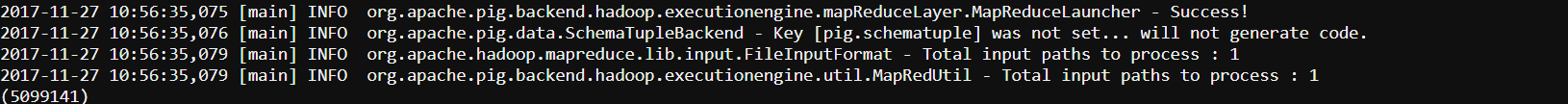
union\_demo3 = UNION union\_demo2,data6;

filter\_demo = Filter union\_demo1 by channel =='BAT';

join\_demo = JOIN filter\_demo by title,union\_demo3 by channel;

group\_demo =Group join\_demo by $1 ;

sum\_demo = FOREACH group\_demo GENERATE SUM(join\_demo.view);



Question 3 What is the most viewed show on ABC channel?

data1 = Load 'IbmProject/Dataset/join2\_genchanA.txt' Using PigStorage(',') As (title:chararray, channel:chararray);

data2 = Load 'IbmProject/Dataset/join2\_genchanB.txt' Using PigStorage(',') As (title:chararray, channel:chararray);

data3 = Load 'IbmProject/Dataset/join2\_genchanC.txt' Using PigStorage(',') As (title:chararray, channel:chararray);

data4 = Load 'IbmProject/Dataset/join2\_gennumA.txt' Using PigStorage(',') As (title:chararray, view:int);

data5 = Load 'IbmProject/Dataset/join2\_gennumB.txt' Using PigStorage(',') As (title:chararray, view:int);

data6 = Load 'IbmProject/Dataset/join2\_gennumC.txt' Using PigStorage(',') As (title:chararray, view:int);

union\_demo = UNION data1,data2 ;

unio\_demo1 = UNION union\_demo,data3;

union\_demo2 = UNION data4,data5;

union\_demo3 = UNION union\_demo2,data6;

filter\_demo = Filter union\_demo1 by channel =='ABC';

join\_demo = JOIN filter\_demo by title,union\_demo3 by title;

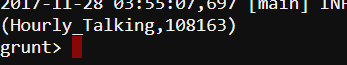
group\_demo =Group join\_demo by $ 0;

sum\_demo = FOREACH group\_demo GENERATE group, SUM(join\_demo.view) as total viewer;

order\_demo = ORDER sum\_demo by $1 DESC;

limit\_demo = order\_demo limit 1;

**OUTPUT**



Question 4

What are the aired shows on ZOO,NOX, ABC channels

data1 = Load 'IbmProject/Dataset/join2\_genchanA.txt' Using PigStorage(',') As (title:chararray, channel:chararray);

data2 = Load 'IbmProject/Dataset/join2\_genchanB.txt' Using PigStorage(',') As (title:chararray, channel:chararray);

data3 = Load 'IbmProject/Dataset/join2\_genchanC.txt' Using PigStorage(',') As (title:chararray, channel:chararray);

data4 = Load 'IbmProject/Dataset/join2\_gennumA.txt' Using PigStorage(',') As (title:chararray, view:int);

data5 = Load 'IbmProject/Dataset/join2\_gennumB.txt' Using PigStorage(',') As (title:chararray, view:int);

data6 = Load 'IbmProject/Dataset/join2\_gennumC.txt' Using PigStorage(',') As (title:chararray, view:int);

union\_demo = UNION data1,data2 ;

unio\_demo1 = UNION union\_demo,data3;

union\_demo2 = UNION data4,data5;

union\_demo3 = UNION union\_demo2,data6;

filter\_demo = Filter union\_demo1 by (channel =='ABC') OR(channel =='NOX') OR (channel =='ZOO') ;

group\_demo =Group join\_demo by channel;

output\_demo = FOREACH group\_demo GENERATE Flatten($1);

