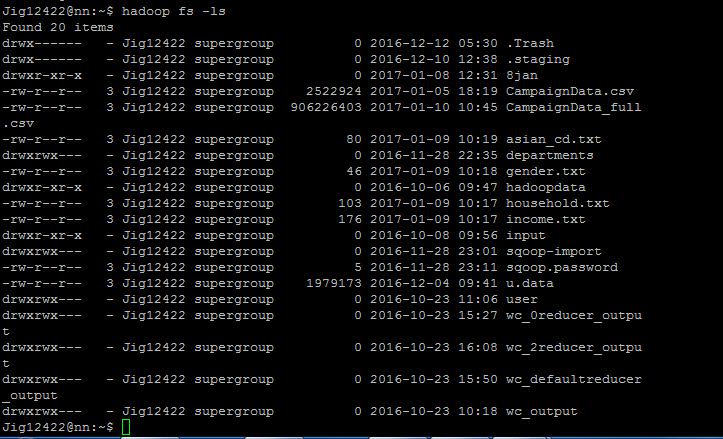
**Capstone Project**

**Email Marketing Campaign**

The file **CampaignData\_full.csv** has been copied to hdfs, using the hadoop fs –copyFromLocal command.



**Pig Latin Scripts**

grunt> register /home/data/piggybank.jar ;

grunt> define CSVLoader org.apache.pig.piggybank.storage.CSVLoader();

grunt> TEmailCampaignData\_full = LOAD 'CampaignData\_full.csv' USING CSVLoader AS (CLICK\_FLG:chararray, OPEN\_FLG:chararray, ADDR\_VER\_CD:chararray, AQI:chararray, ASIAN\_CD:chararray, AUTO\_IN\_MARKET:chararray, BIRD\_QTY:chararray, BUYER\_DM\_BOOKS:chararray, BUYER\_DM\_COLLECT\_SPC\_FOOD:chararray, BUYER\_DM\_CRAFTS\_HOBBI:chararray, BUYER\_DM\_FEMALE\_ORIEN:chararray, BUYER\_DM\_GARDEN\_FARM:chararray, BUYER\_DM\_GENERAL:chararray, BUYER\_DM\_GIFT\_GADGET:chararray, BUYER\_DM\_MALE\_ORIEN:chararray, BUYER\_DM\_UPSCALE:chararray, BUYER\_MAG\_CULINARY\_INTERS:chararray, BUYER\_MAG\_FAMILY\_GENERAL:chararray, BUYER\_MAG\_FEMALE\_ORIENTED:chararray, BUYER\_MAG\_GARDEN\_FARMING:chararray, BUYER\_MAG\_HEALTH\_FITNESS:chararray, BUYER\_MAG\_MALE\_SPORT\_ORIENTED:chararray, BUYER\_MAG\_RELIGIOUS:chararray, CATS\_QTY:chararray, CEN\_2000\_MATCH\_LEVEL:chararray, CLUB\_MEMBER\_CD:chararray, COUNTRY\_OF\_ORIGIN:chararray, DECEASED\_INDICATOR:chararray, DM\_RESPONDER\_HH:chararray, DM\_RESPONDER\_INDIV:chararray, DMR\_CONTRIB\_CAT\_GENERAL:chararray, DMR\_CONTRIB\_CAT\_HEALTH\_INST:chararray, DMR\_CONTRIB\_CAT\_POLITICAL:chararray, DMR\_CONTRIB\_CAT\_RELIGIOUS:chararray, DMR\_DO\_IT\_YOURSELFERS:chararray, DMR\_MISCELLANEOUS:chararray, DMR\_NEWS\_FINANCIAL:chararray, DMR\_ODD\_ENDS:chararray, DMR\_PHOTOGRAPHY:chararray, DMR\_SWEEPSTAKES:chararray, DOG\_QTY:chararray, DWELLING\_TYPE:chararray, DWELLING\_UNIT\_SIZE:chararray, EST\_LOAN\_VALUE\_RATIO:chararray, ETECH\_GROUP:chararray, ETHNIC\_GROUP\_CODE:chararray, ETHNIC\_INSIGHT\_MTCH\_FLG:chararray, ETHNICITY\_DETAIL:chararray, EXPERIAN\_INCOME\_CD:chararray, EXPERIAN\_INCOME\_CD\_V4:chararray, GNDR\_OF\_CHLDRN\_0\_3:chararray, GNDR\_OF\_CHLDRN\_10\_12:chararray, GNDR\_OF\_CHLDRN\_13\_18:chararray, GNDR\_OF\_CHLDRN\_4\_6:chararray, GNDR\_OF\_CHLDRN\_7\_9:chararray, HH\_INCOME:chararray, HHLD\_DM\_PURC\_CD:chararray, HOME\_BUSINESS\_IND:chararray, I1\_BUSINESS\_OWNER\_FLG:chararray, I1\_EXACT\_AGE:chararray, I1\_GNDR\_CODE:chararray, I1\_INDIV\_HHLD\_STATUS\_CODE:chararray, INDIV\_EDUCATION:chararray, INDIV\_EDUCATION\_CONF\_LVL:chararray, INDIV\_MARITAL\_STATUS:chararray, INDIV\_MARITAL\_STATUS\_CONF\_LVL:chararray, INS\_MATCH\_TYPE:chararray, LANGUAGE:chararray, LENGTH\_OF\_RESIDENCE:chararray, MEDIAN\_HOUSING\_VALUE:chararray, MEDIAN\_LEN\_OF\_RESIDENCE:chararray, MM\_INCOME\_CD:chararray, MOSAIC\_HH:chararray, MULTI\_BUYER\_INDIV:chararray, NEW\_CAR\_MODEL:chararray, NUM\_OF\_ADULTS\_IN\_HHLD:chararray, NUMBER\_OF\_CHLDRN\_18\_OR\_LESS:chararray, OCCUP\_DETAIL:chararray, OCCUP\_MIX\_PCT:chararray, PCT\_CHLDRN:chararray, PCT\_DEROG\_TRADES:chararray, PCT\_HOUSEHOLDS\_BLACK:chararray, PCT\_OWNER\_OCCUPIED:chararray, PCT\_RENTER\_OCCUPIED:chararray, PCT\_TRADES\_NOT\_DEROG:chararray, PCT\_WHITE:chararray, PHONE\_TYPE\_CD:chararray, PRES\_OF\_CHLDRN\_0\_3:chararray, PRES\_OF\_CHLDRN\_10\_12:chararray, PRES\_OF\_CHLDRN\_13\_18:chararray, PRES\_OF\_CHLDRN\_4\_6:chararray, PRES\_OF\_CHLDRN\_7\_9:chararray, PRESENCE\_OF\_CHLDRN:chararray, PRIM\_FEM\_EDUC\_CD:chararray, PRIM\_FEM\_OCC\_CD:chararray, PRIM\_MALE\_EDUC\_CD:chararray, PRIM\_MALE\_OCC\_CD:chararray, RECIPIENT\_RELIABILITY\_CD:chararray, RELIGION:chararray, SCS\_MATCH\_TYPE:chararray, TRW\_INCOME\_CD:chararray, TRW\_INCOME\_CD\_V4:chararray, USED\_CAR\_CD:chararray, Y\_OWNS\_HOME:chararray, Y\_PROBABLE\_HOMEOWNER:chararray, Y\_PROBABLE\_RENTER:chararray, Y\_RENTER:chararray, YRS\_SCHOOLING\_CD:chararray, Z\_CREDIT\_CARD:chararray, age\_hh\_mem1:chararray, age\_hh\_mem2:chararray, age\_hh\_mem3:chararray, age\_hh\_mem4:chararray, age\_hh\_mem5:chararray, age\_hh\_mem6:chararray, age\_hh\_mem7:chararray, age\_hh\_mem8:chararray, gender\_hh\_mem1:chararray, gender\_hh\_mem2:chararray, gender\_hh\_mem3:chararray, gender\_hh\_mem4:chararray, gender\_hh\_mem5:chararray, gender\_hh\_mem6:chararray, gender\_hh\_mem7:chararray, gender\_hh\_mem8:chararray, statcd\_hh\_mem1:chararray, statcd\_hh\_mem2:chararray, statcd\_hh\_mem3:chararray, statcd\_hh\_mem4:chararray, statcd\_hh\_mem5:chararray, statcd\_hh\_mem6:chararray, statcd\_hh\_mem7:chararray, statcd\_hh\_mem8:chararray, trait\_cd1:chararray, trait\_cd2:chararray, trait\_cd3:chararray, trait\_cd4:chararray, trait\_cd5:chararray, trait\_cd6:chararray, trait\_cd7:chararray, trait\_cd8:chararray, trait\_cd9:chararray, trait\_cd10:chararray, trait\_cd11:chararray, trait\_cd12:chararray, trait\_cd13:chararray, trait\_cd14:chararray, trait\_cd15:chararray, trait\_cd16:chararray, trait\_cd17:chararray, trait\_cd18:chararray, trait\_cd19:chararray, trait\_cd20:chararray, trait\_cd21:chararray, trait\_cd22:chararray, trait\_cd23:chararray, trait\_cd24:chararray, trait\_cd25:chararray, trait\_cd26:chararray, trait\_cd27:chararray, trait\_cd28:chararray, trait\_cd29:chararray, trait\_cd30:chararray, trait\_cd31:chararray, trait\_cd32:chararray, trait\_cd33:chararray, trait\_cd34:chararray, trait\_cd35:chararray, trait\_cd36:chararray, trait\_cd37:chararray, trait\_cd38:chararray, trait\_cd39:chararray, trait\_cd40:chararray, trait\_cd41:chararray, trait\_cd42:chararray, trait\_cd43:chararray, trait\_cd44:chararray, trait\_cd45:chararray, trait\_cd46:chararray, trait\_cd47:chararray, trait\_cd48:chararray, trait\_cd49:chararray, trait\_cd50:chararray, trait\_cd51:chararray, trait\_cd52:chararray, trait\_cd53:chararray, trait\_cd54:chararray, trait\_cd55:chararray, trait\_cd56:chararray, trait\_cd57:chararray, trait\_cd58:chararray, trait\_cd59:chararray, trait\_cd60:chararray, trait\_cd61:chararray, trait\_cd62:chararray, trait\_cd63:chararray, trait\_cd64:chararray, trait\_typ\_cd1:chararray, trait\_typ\_cd2:chararray, trait\_typ\_cd3:chararray, trait\_typ\_cd4:chararray, trait\_typ\_cd5:chararray, trait\_typ\_cd6:chararray, trait\_typ\_cd7:chararray, trait\_typ\_cd8:chararray, trait\_typ\_cd9:chararray, trait\_typ\_cd10:chararray, trait\_typ\_cd11:chararray, trait\_typ\_cd12:chararray, trait\_typ\_cd13:chararray, trait\_typ\_cd14:chararray, trait\_typ\_cd15:chararray, trait\_typ\_cd16:chararray, trait\_typ\_cd17:chararray, trait\_typ\_cd18:chararray, trait\_typ\_cd19:chararray, trait\_typ\_cd20:chararray, trait\_typ\_cd21:chararray, trait\_typ\_cd22:chararray, trait\_typ\_cd23:chararray, trait\_typ\_cd24:chararray, trait\_typ\_cd25:chararray, trait\_typ\_cd26:chararray, trait\_typ\_cd27:chararray, trait\_typ\_cd28:chararray, trait\_typ\_cd29:chararray, trait\_typ\_cd30:chararray, trait\_typ\_cd31:chararray, trait\_typ\_cd32:chararray, trait\_typ\_cd33:chararray, trait\_typ\_cd34:chararray, trait\_typ\_cd35:chararray, trait\_typ\_cd36:chararray, trait\_typ\_cd37:chararray, trait\_typ\_cd38:chararray, trait\_typ\_cd39:chararray, trait\_typ\_cd40:chararray, trait\_typ\_cd41:chararray, trait\_typ\_cd42:chararray, trait\_typ\_cd43:chararray, trait\_typ\_cd44:chararray, trait\_typ\_cd45:chararray, trait\_typ\_cd46:chararray, trait\_typ\_cd47:chararray, trait\_typ\_cd48:chararray, trait\_typ\_cd49:chararray, trait\_typ\_cd50:chararray, trait\_typ\_cd51:chararray, trait\_typ\_cd52:chararray, trait\_typ\_cd53:chararray, trait\_typ\_cd54:chararray, trait\_typ\_cd55:chararray, trait\_typ\_cd56:chararray, trait\_typ\_cd57:chararray, trait\_typ\_cd58:chararray, trait\_typ\_cd59:chararray, trait\_typ\_cd60:chararray, trait\_typ\_cd61:chararray, trait\_typ\_cd62:chararray, trait\_typ\_cd63:chararray, trait\_typ\_cd64:chararray, CITY:chararray, POSTAL\_CD:chararray, STATE\_PROVINCE:chararray, new\_id:chararray, list\_auth\_cd:chararray, list\_auth\_flg:chararray, new\_list\_id:chararray, list\_new\_brand\_cd:chararray, list\_new\_ed\_cd:chararray, new\_mailing\_id:chararray, Mailing\_Category:chararray, Mailed\_Date:chararray);

grunt> EmailCampaignData1 = FOREACH TEmailCampaignData\_full generate CLICK\_FLG as CLICK\_FLG,OPEN\_FLG as OPEN\_FLG,ADDR\_VER\_CD as ADDR\_VER\_CD,AQI as AQI, (ASIAN\_CD == '' ? 'null' : (ASIAN\_CD is null ? 'null': ASIAN\_CD)) as ASIAN\_CD, AUTO\_IN\_MARKET as AUTO\_IN\_MARKET,BIRD\_QTY as BIRD\_QTY ,DM\_RESPONDER\_HH as DM\_RESPONDER\_HH, GNDR\_OF\_CHLDRN\_0\_3 as GNDR\_OF\_CHLDRN\_0\_3 ,GNDR\_OF\_CHLDRN\_10\_12 as GNDR\_OF\_CHLDRN\_10\_12,GNDR\_OF\_CHLDRN\_13\_18 as GNDR\_OF\_CHLDRN\_13\_18,GNDR\_OF\_CHLDRN\_4\_6 as GNDR\_OF\_CHLDRN\_4\_6,GNDR\_OF\_CHLDRN\_7\_9 as GNDR\_OF\_CHLDRN\_7\_9 ,HOME\_BUSINESS\_IND as HOME\_BUSINESS\_IND,I1\_BUSINESS\_OWNER\_FLG as I1\_BUSINESS\_OWNER\_FLG ,I1\_EXACT\_AGE as I1\_EXACT\_AGE,I1\_GNDR\_CODE as I1\_GNDR\_CODE, (I1\_INDIV\_HHLD\_STATUS\_CODE == '' ? 'null' : I1\_INDIV\_HHLD\_STATUS\_CODE) as I1\_INDIV\_HHLD\_STATUS\_CODE,INDIV\_MARITAL\_STATUS as INDIV\_MARITAL\_STATUS,LANGUAGE as LANGUAGE,MM\_INCOME\_CD as MM\_INCOME\_CD,NUM\_OF\_ADULTS\_IN\_HHLD as NUM\_OF\_ADULTS\_IN\_HHLD,NUMBER\_OF\_CHLDRN\_18\_OR\_LESS as NUMBER\_OF\_CHLDRN\_18\_OR\_LESS,PRES\_OF\_CHLDRN\_0\_3 as PRES\_OF\_CHLDRN\_0\_3,PRES\_OF\_CHLDRN\_10\_12 as PRES\_OF\_CHLDRN\_10\_12, PRES\_OF\_CHLDRN\_13\_18 as PRES\_OF\_CHLDRN\_13\_18,PRES\_OF\_CHLDRN\_4\_6 as PRES\_OF\_CHLDRN\_4\_6,PRES\_OF\_CHLDRN\_7\_9 as PRES\_OF\_CHLDRN\_7\_9,PRESENCE\_OF\_CHLDRN as PRESENCE\_OF\_CHLDRN,RELIGION as RELIGION,age\_hh\_mem1 as age\_hh\_mem1,age\_hh\_mem2 as age\_hh\_mem2,age\_hh\_mem3 as age\_hh\_mem3, age\_hh\_mem4 as age\_hh\_mem4,age\_hh\_mem5 as age\_hh\_mem5,age\_hh\_mem6 as age\_hh\_mem6,age\_hh\_mem7 as age\_hh\_mem7,age\_hh\_mem8 as age\_hh\_mem8,gender\_hh\_mem1 as gender\_hh\_mem1, gender\_hh\_mem2 as gender\_hh\_mem2, gender\_hh\_mem3 as gender\_hh\_mem3,gender\_hh\_mem4 as gender\_hh\_mem4,gender\_hh\_mem5 as gender\_hh\_mem5,gender\_hh\_mem6 as gender\_hh\_mem6,gender\_hh\_mem7 as gender\_hh\_mem7,gender\_hh\_mem8 as gender\_hh\_mem8,statcd\_hh\_mem1 as statcd\_hh\_mem1,statcd\_hh\_mem2 as statcd\_hh\_mem2,statcd\_hh\_mem3 as statcd\_hh\_mem3,statcd\_hh\_mem4 as statcd\_hh\_mem4,statcd\_hh\_mem5 as statcd\_hh\_mem5,statcd\_hh\_mem6 as statcd\_hh\_mem6,statcd\_hh\_mem7 as statcd\_hh\_mem7,statcd\_hh\_mem8 as statcd\_hh\_mem8,CITY as CITY,new\_id as new\_id,new\_list\_id as new\_list\_id,new\_mailing\_id as new\_mailing\_id,Mailing\_Category as Mailing\_Category, (TRW\_INCOME\_CD\_V4 == '' ? 'null' : TRW\_INCOME\_CD\_V4) as TRW\_INCOME\_CD\_V4, Mailed\_Date as Mailed\_Date, SUBSTRING(Mailed\_Date,5,13) As Date,CONCAT('20',SUBSTRING(Mailed\_Date,11,13)) As Year,(SUBSTRING(Mailed\_Date,5,7)=='01' ? 'January' : ( SUBSTRING(Mailed\_Date,5,7)=='02' ? 'February' : (SUBSTRING(Mailed\_Date,5,7)=='03' ? 'March' :(SUBSTRING(Mailed\_Date,5,7)=='04' ? 'April' :(SUBSTRING(Mailed\_Date,5,7)=='05' ? 'May' :(SUBSTRING(Mailed\_Date,5,7)=='06' ? 'June' :(SUBSTRING(Mailed\_Date,5,7)=='07' ? 'July' :(SUBSTRING(Mailed\_Date,5,7)=='08' ? 'August' :(SUBSTRING(Mailed\_Date,5,7)=='09' ? 'September' :(SUBSTRING(Mailed\_Date,5,7)=='10' ? 'October' :(SUBSTRING(Mailed\_Date,5,7)=='11' ? 'November' :'December'))))))))))) As Month,SUBSTRING(Mailed\_Date,0,3) as Day,SUBSTRING(Mailed\_Date,14,19) as TimeStamp,SUBSTRING(Mailed\_Date,20,22) as AMPM;

grunt> IncomeHouseHold = LOAD 'income.txt' USING PigStorage('|') as (group:chararray,income:chararray);

grunt> AsianEthnicity = LOAD 'asian\_cd.txt' USING PigStorage('|') as (ethnicitycode:chararray,ethnicity:chararray);

grunt> HouseHoldStatus = LOAD 'household.txt ' USING PigStorage('|') as (statuscode:chararray,status:chararray);

grunt> Gender = LOAD 'gender.txt ' USING PigStorage('|') as (gendercode:chararray,genderdesc:chararray);

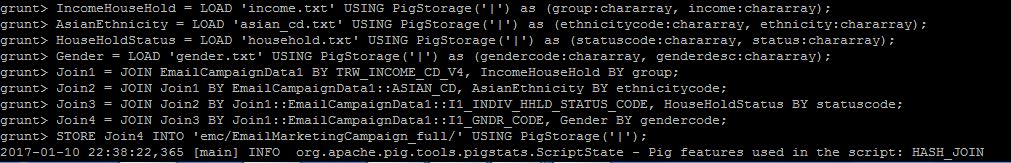
grunt> Join1 = JOIN EmailCampaignData1 BY TRW\_INCOME\_CD\_V4 , IncomeHouseHold BY group;

grunt> Join2 = JOIN Join1 BY EmailCampaignData1::ASIAN\_CD , AsianEthnicity BY ethnicitycode;

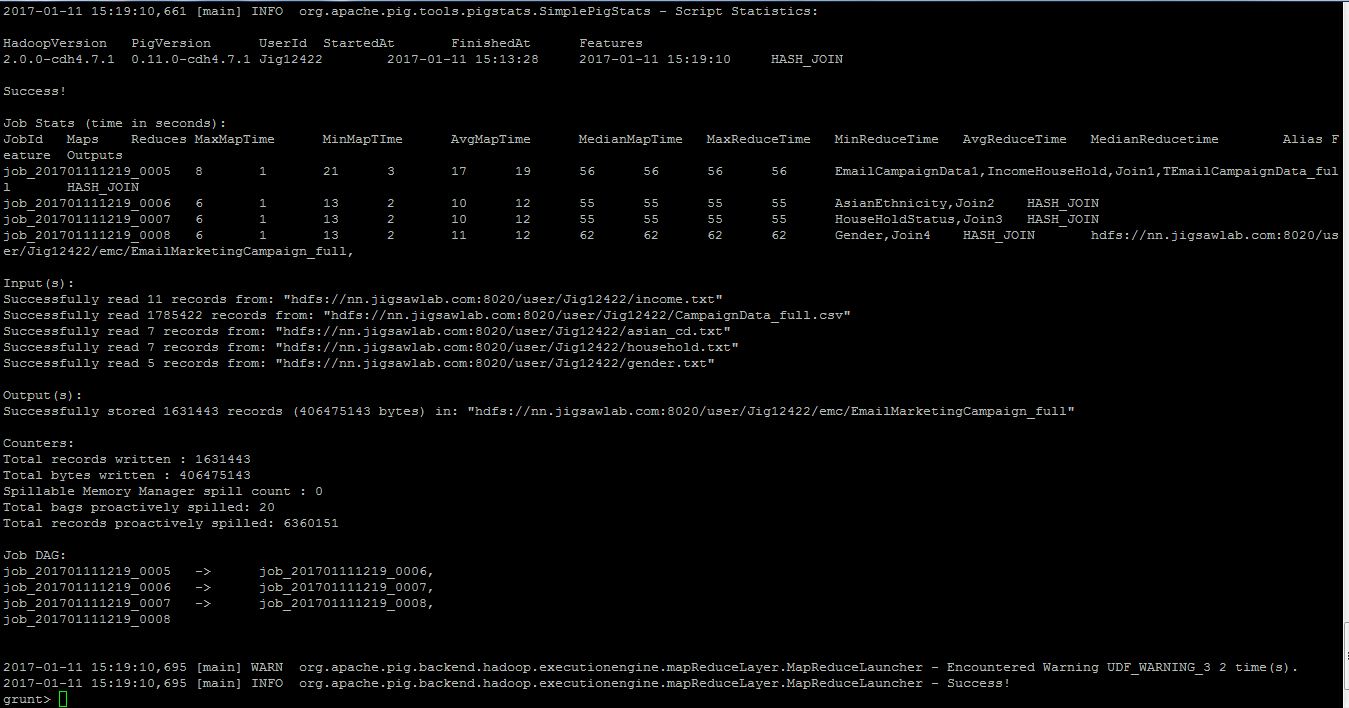
grunt> Join3 = JOIN Join2 BY Join1::EmailCampaignData1::I1\_INDIV\_HHLD\_STATUS\_CODE , HouseHoldStatus BY statuscode;

grunt> Join4 = JOIN Join3 BY Join1::EmailCampaignData1::I1\_GNDR\_CODE, Gender BY gendercode;

grunt> STORE Join4 INTO 'emc/EmailMarketingCampaign\_full/' USING PigStorage('|');

****

**The below screenshot shows that the Map Reduce job in Pig Latin has run successfully. The output of the job is stored at /user/Jig12422/emc/EmailMarketingCampaign\_full.**

****

**CONTENTS OF ‘asian\_cd.txt’ FILE:**

00|Unknown

05|Chinese

24|Japanese

25|Korean

47|Vietnamese

48|Asian

null|No match

**CONTENTS OF ‘gender.txt’ FILE:**

B|Both

F|Female

M|Male

U|Unknown

null|no match

**CONTENTS OF ‘household.txt’ FILE:**

D|Deceased

H|Head

P|Aged parent living home

U|Unknown

W|Spouse

Y|Young adult (Age 19-25)

null|No match

**CONTENTS OF ‘income.txt’ FILE:**

J|<$15,000

K|$15,000-$24,999

L|$25,000-$34,999

M|$35,000-$49,999

N|$50,000-$74,999

O|$75,000-$99,999

P|$100,000-$119,999

Q|$120,000-$149,999

R|$150,000+

U|Unknown

null|No match

**Hive DDL & DML**

hive> CREATE DATABASE Capstone;

hive> Use Capstone;

hive> CREATE EXTERNAL TABLE emchive(CLICK\_FLG string,OPEN\_FLG string,ADDR\_VER\_CD string,AQI string,ASIAN\_CD string,AUTO\_IN\_MARKET string,BIRD\_QTY string,DM\_RESPONDER\_HH string,GNDR\_OF\_CHLDRN\_0\_3 string,GNDR\_OF\_CHLDRN\_10\_12 string,GNDR\_OF\_CHLDRN\_13\_18 string,GNDR\_OF\_CHLDRN\_4\_6 string,GNDR\_OF\_CHLDRN\_7\_9 string,HOME\_BUSINESS\_IND string,I1\_BUSINESS\_OWNER\_FLG string,I1\_EXACT\_AGE string,I1\_GNDR\_CODE string,I1\_INDIV\_HHLD\_STATUS\_CODE string,INDIV\_MARITAL\_STATUS string,LANGUAGE string,MM\_INCOME\_CD string,NUM\_OF\_ADULTS\_IN\_HHLD string,NUMBER\_OF\_CHLDRN\_18\_OR\_LESS string,PRES\_OF\_CHLDRN\_0\_3 string,PRES\_OF\_CHLDRN\_10\_12 string,PRES\_OF\_CHLDRN\_13\_18 string,PRES\_OF\_CHLDRN\_4\_6 string,PRES\_OF\_CHLDRN\_7\_9 string,PRESENCE\_OF\_CHLDRN string,RELIGION string,age\_hh\_mem1 string,age\_hh\_mem2 string,age\_hh\_mem3 string,age\_hh\_mem4 string,age\_hh\_mem5 string,age\_hh\_mem6 string,age\_hh\_mem7 string,age\_hh\_mem8 string,gender\_hh\_mem1 string,gender\_hh\_mem2 string,gender\_hh\_mem3 string,gender\_hh\_mem4 string,gender\_hh\_mem5 string,gender\_hh\_mem6 string,gender\_hh\_mem7 string,gender\_hh\_mem8 string,statcd\_hh\_mem1 string,statcd\_hh\_mem2 string,statcd\_hh\_mem3 string,statcd\_hh\_mem4 string,statcd\_hh\_mem5 string,statcd\_hh\_mem6 string,statcd\_hh\_mem7 string,statcd\_hh\_mem8 string,CITY string,new\_id string,new\_list\_id string,new\_mailing\_id string,Mailing\_Category string,TRW\_INCOME\_CD\_V4 string,Mailed\_Date string,datepart string, Mail\_Year string,Mail\_Month string,Mail\_Day string,Mail\_TimeStamp string, AMPM string, income\_group string, income string,ethnicitycode string, ethnicity string, hhstatuscode string, hhstatus string, gendercode string, gender string) ROW FORMAT DELIMITED

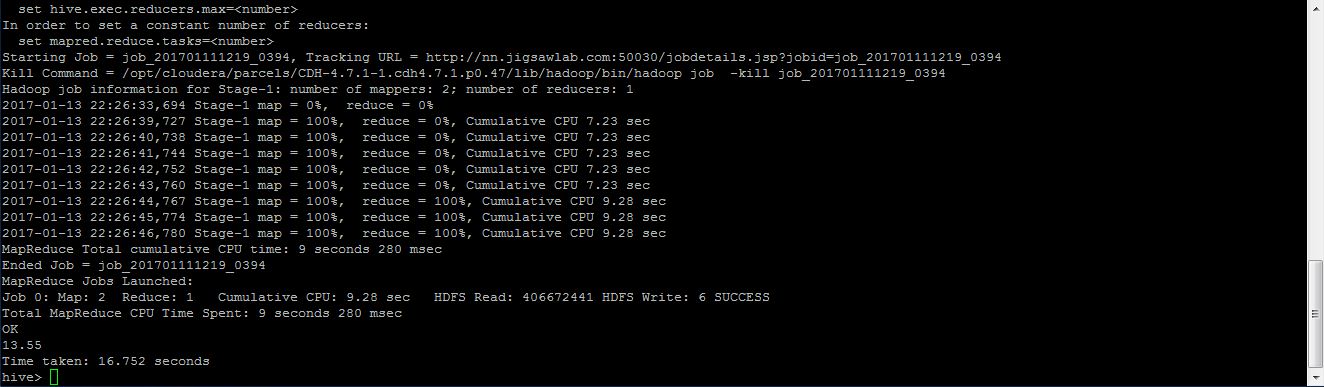
FIELDS TERMINATED BY '|'

LOCATION '/user/Jig12422/emc/EmailMarketingCampaign\_full';

**Codes for solving question set 1**

**1A. Overall CTOR**

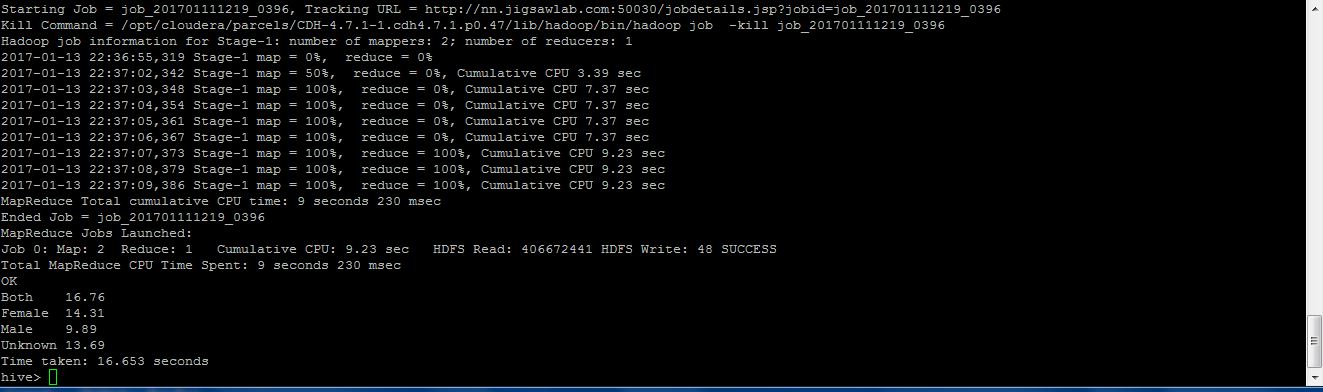
hive> SELECT ROUND(SUM(CASE WHEN CLICK\_FLG = 'Y' THEN 1 ELSE 0 END)/count(\*) \* 100,2) CTOR FROM emchive WHERE OPEN\_FLG = 'Y';



hive> CREATE TABLE CTOR as SELECT ROUND(SUM(CASE WHEN CLICK\_FLG = 'Y' THEN 1 ELSE 0 END)/count(\*) \* 100,2) CTOR FROM emchive WHERE OPEN\_FLG = 'Y';

**1B. CTOR by Gender**

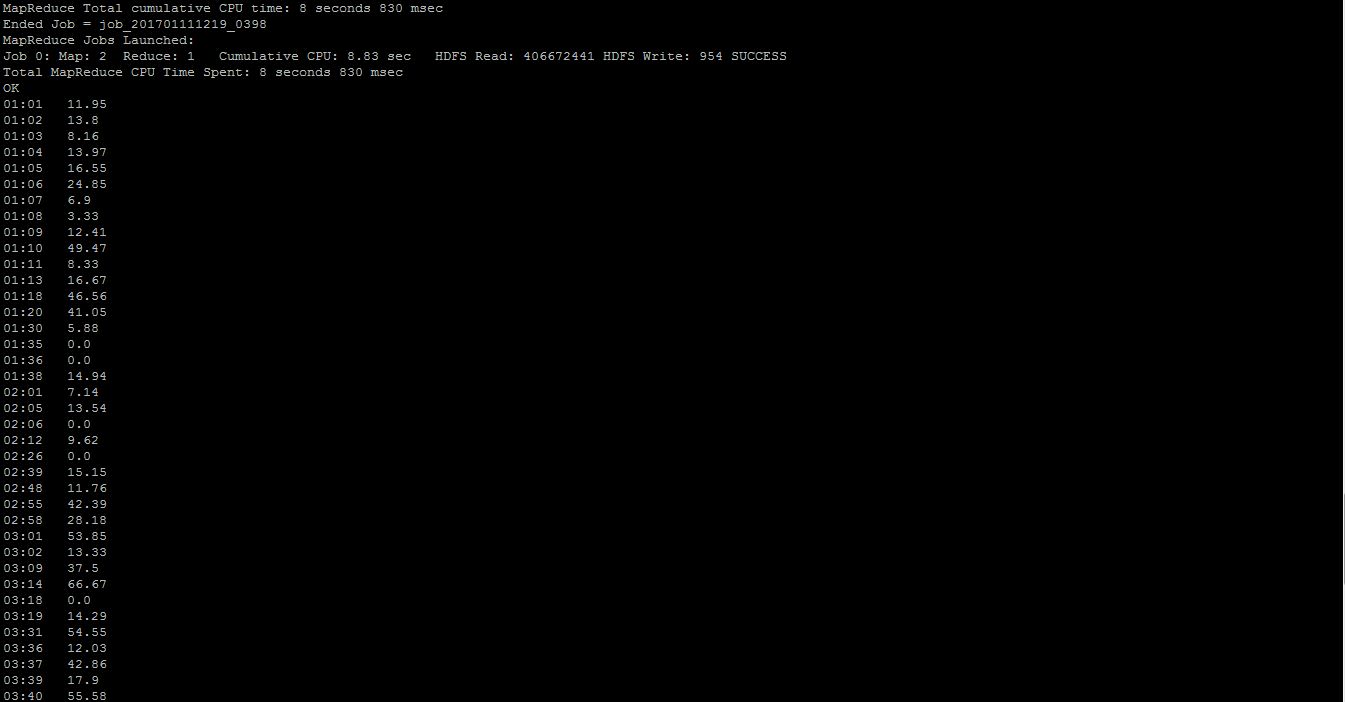
hive> SELECT gender, ROUND(SUM(CASE WHEN CLICK\_FLG = 'Y' THEN 1 ELSE 0 END)/count(\*) \* 100,2) CTOR FROM emchive WHERE OPEN\_FLG ='Y' GROUP BY gender;



hive> CREATE TABLE CTOR\_Gender as SELECT gender, ROUND(SUM(CASE WHEN CLICK\_FLG = 'Y' THEN 1 ELSE 0 END)/count(\*) \* 100,2) CTOR FROM emchive WHERE OPEN\_FLG ='Y' GROUP BY gender;

**1C. CTOR by Time of the day**

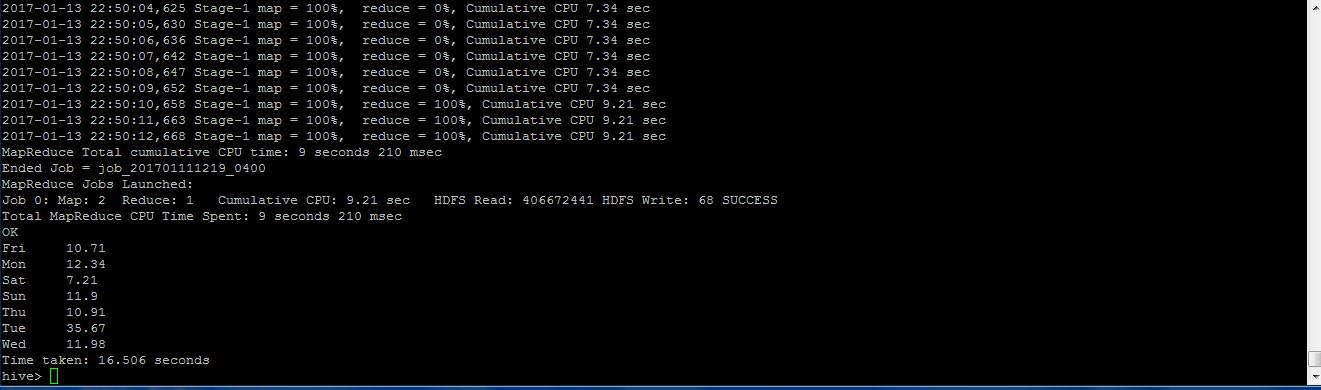
hive> SELECT Mail\_TimeStamp, ROUND(SUM(CASE WHEN CLICK\_FLG = 'Y' THEN 1 ELSE 0 END)/count(\*) \* 100,2) CTOR FROM emchive WHERE OPEN\_FLG='Y' GROUP BY Mail\_TimeStamp;



hive> CREATE TABLE CTOR\_TimeOfDay as SELECT Mail\_TimeStamp, ROUND(SUM(CASE WHEN CLICK\_FLG = 'Y' THEN 1 ELSE 0 END)/count(\*) \* 100,2) CTOR FROM emchive WHERE OPEN\_FLG='Y' GROUP BY Mail\_TimeStamp;

**1D. CTOR by Day of the week.**

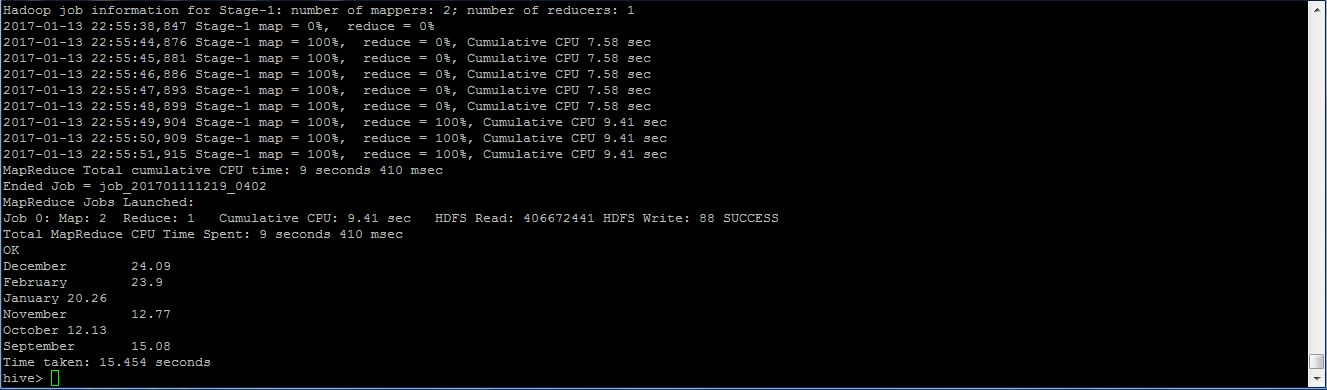
hive> SELECT Mail\_Day, ROUND(SUM(CASE WHEN CLICK\_FLG = 'Y' THEN 1 ELSE 0 END)/count(\*) \* 100,2) CTOR FROM emchive WHERE OPEN\_FLG='Y' GROUP BY Mail\_Day;



hive> CREATE TABLE CTOR\_DayOfTheWeek as SELECT Mail\_Day, ROUND(SUM(CASE WHEN CLICK\_FLG = 'Y' THEN 1 ELSE 0 END)/count(\*) \* 100,2) CTOR FROM emchive WHERE OPEN\_FLG='Y' GROUP BY Mail\_Day;

**1E. CTOR by Month**

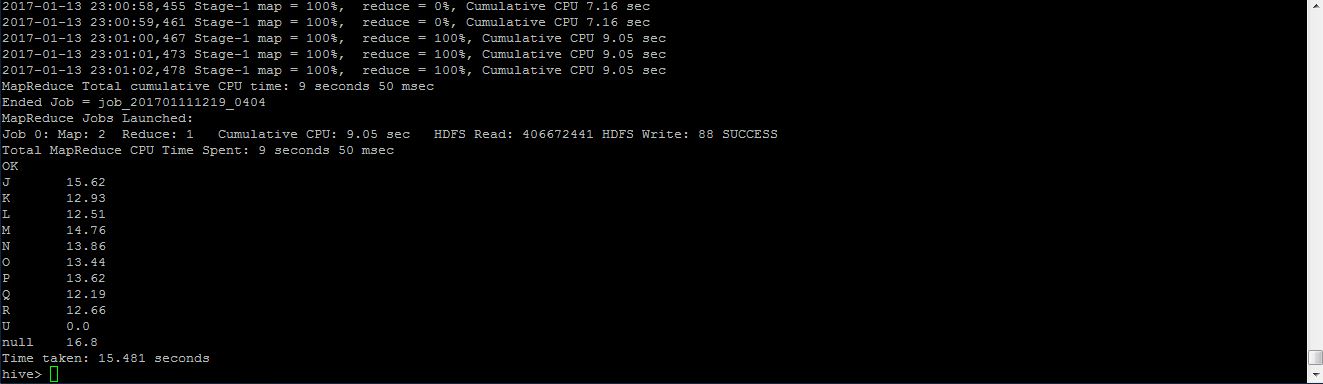
hive> SELECT Mail\_Month, ROUND(SUM(CASE WHEN CLICK\_FLG = 'Y' THEN 1 ELSE 0 END)/count(\*) \* 100,2) CTOR FROM emchive WHERE OPEN\_FLG ='Y' GROUP BY Mail\_Month;



hive> CREATE TABLE CTOR\_Month as SELECT Mail\_Month, ROUND(SUM(CASE WHEN CLICK\_FLG = 'Y' THEN 1 ELSE 0 END)/count(\*) \* 100,2) CTOR FROM emchive WHERE OPEN\_FLG ='Y' GROUP BY Mail\_Month;

**1F. CTOR by Lead's Income Group**

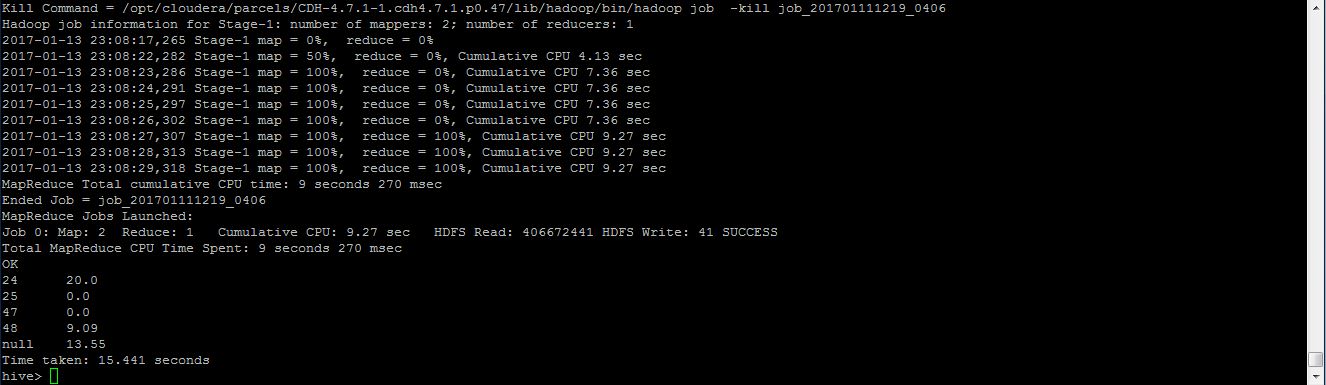
hive> SELECT TRW\_INCOME\_CD\_V4, ROUND(SUM(CASE WHEN CLICK\_FLG = 'Y' THEN 1 ELSE 0 END)/count(\*) \* 100,2) CTOR FROM emchive WHERE OPEN\_FLG='Y' GROUP BY TRW\_INCOME\_CD\_V4;



hive> CREATE TABLE CTOR\_LeadsIncomeGroup as SELECT TRW\_INCOME\_CD\_V4, ROUND(SUM(CASE WHEN CLICK\_FLG = 'Y' THEN 1 ELSE 0 END)/count(\*) \* 100,2) CTOR FROM emchive WHERE OPEN\_FLG='Y' GROUP BY TRW\_INCOME\_CD\_V4;

**1G. CTOR by Lead's Ethnicity**

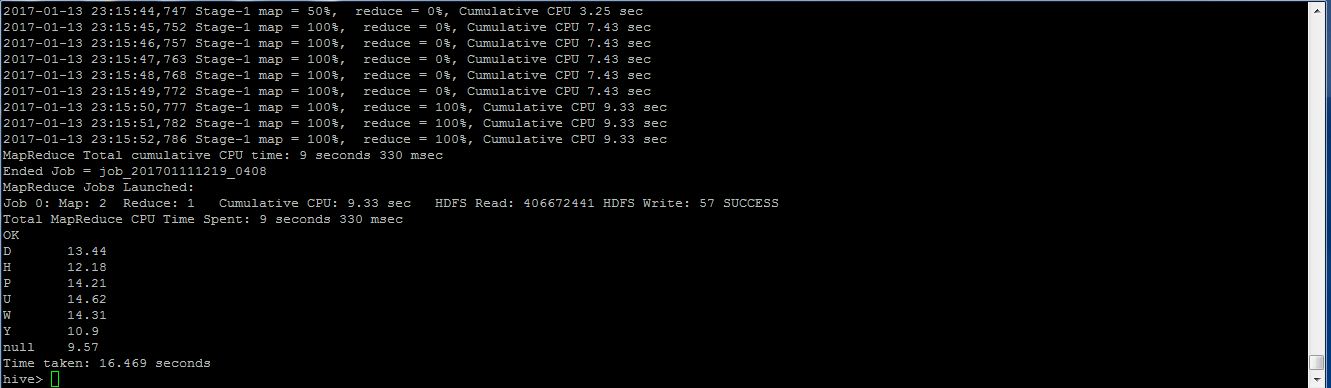
hive> SELECT ASIAN\_CD, ROUND(SUM(CASE WHEN CLICK\_FLG = 'Y' THEN 1 ELSE 0 END)/count(\*) \* 100,2) CTOR FROM emchive WHERE OPEN\_FLG='Y' GROUP BY ASIAN\_CD;



hive> CREATE TABLE CTOR\_LeadsEthnicity as SELECT ASIAN\_CD, ROUND(SUM(CASE WHEN CLICK\_FLG = 'Y' THEN 1 ELSE 0 END)/count(\*) \* 100,2) CTOR FROM emchive WHERE OPEN\_FLG='Y' GROUP BY ASIAN\_CD;

**1H. CTOR by Lead's Household Status**

hive> SELECT I1\_INDIV\_HHLD\_STATUS\_CODE, ROUND(SUM(CASE WHEN CLICK\_FLG = 'Y' THEN 1 ELSE 0 END)/count(\*) \* 100,2) CTOR FROM emchive WHERE OPEN\_FLG='Y' GROUP BY I1\_INDIV\_HHLD\_STATUS\_CODE;



hive> CREATE TABLE CTOR\_LeadsHouseHold as SELECT I1\_INDIV\_HHLD\_STATUS\_CODE, ROUND(SUM(CASE WHEN CLICK\_FLG = 'Y' THEN 1 ELSE 0 END)/count(\*) \* 100,2) CTOR FROM emchive WHERE OPEN\_FLG='Y' GROUP BY I1\_INDIV\_HHLD\_STATUS\_CODE;

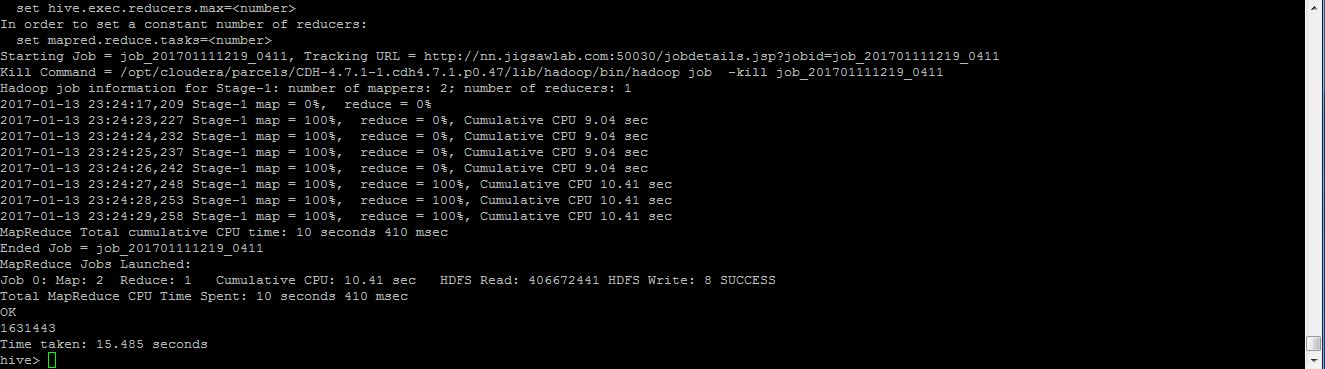
**Note: First the Hive queries are run and their outputs have been taken as screenshots. Then, tables have been created with the same queries to store the output in hdfs for processing in Tableau.**

**Codes for solving question set 2**

**2. Household Members Information**

**A. Find count of leads with information about members of their household. If a lead has information about 3 members, and another has information about 2 members and another has none, then the answer to this question is 2. (use statcd\_hh\_mem - statcd\_hh\_mem8 column)**

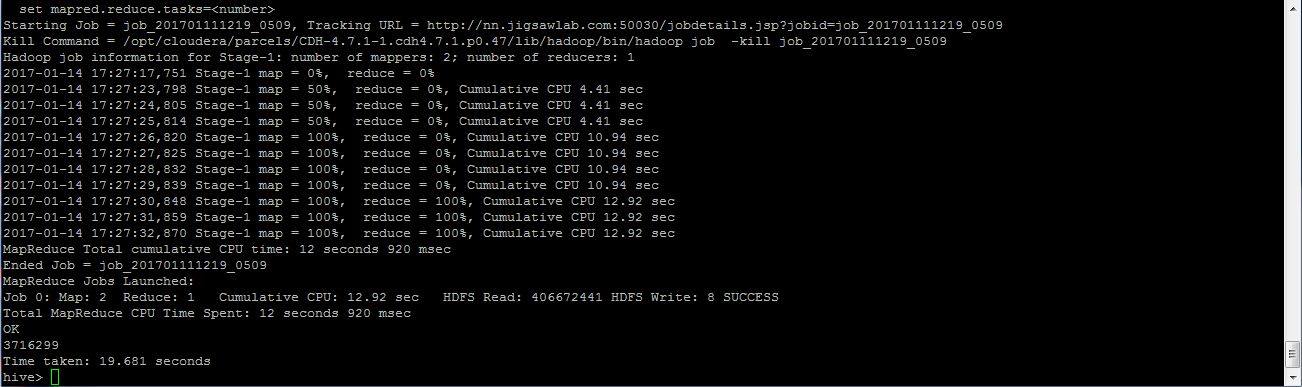
hive> SELECT SUM(CASE WHEN statcd\_hh\_mem1 != "" OR statcd\_hh\_mem2 != "" OR statcd\_hh\_mem3 != "" OR statcd\_hh\_mem4 != "" OR statcd\_hh\_mem5 != "" OR statcd\_hh\_mem6 != "" OR statcd\_hh\_mem7 != "" OR statcd\_hh\_mem8 != "" THEN 1 ELSE 0 END) AS total FROM emchive;



hive> CREATE TABLE 2A as SELECT SUM(CASE WHEN statcd\_hh\_mem1 != "" OR statcd\_hh\_mem2 != "" OR statcd\_hh\_mem3 != "" OR statcd\_hh\_mem4 != "" OR statcd\_hh\_mem5 != "" OR statcd\_hh\_mem6 != "" OR statcd\_hh\_mem7 != "" OR statcd\_hh\_mem8 != "" THEN 1 ELSE 0 END) AS total FROM emchive;

**2B. Find count of total number of household members information is available. For example, if a lead has 3 household members, and another has about 2 members, and the other has none, then the total count of household members is 3+2+0 = 5.**

hive> SELECT SUM(CASE WHEN statcd\_hh\_mem1 != "" THEN 1 ELSE 0 END) + SUM(CASE WHEN statcd\_hh\_mem2 != "" THEN 1 ELSE 0 END) + SUM(CASE WHEN statcd\_hh\_mem3 != "" THEN 1 ELSE 0 END) + SUM(CASE WHEN statcd\_hh\_mem4 != "" THEN 1 ELSE 0 END) + SUM(CASE WHEN statcd\_hh\_mem5 != "" THEN 1 ELSE 0 END) + SUM(CASE WHEN statcd\_hh\_mem6 != "" THEN 1 ELSE 0 END) + SUM(CASE WHEN statcd\_hh\_mem7 != "" THEN 1 ELSE 0 END) + SUM(CASE WHEN statcd\_hh\_mem8 != "" THEN 1 ELSE 0 END) AS total FROM emchive;



hive> CREATE TABLE 2B as SELECT SUM(CASE WHEN statcd\_hh\_mem1 != "" THEN 1 ELSE 0 END) + SUM(CASE WHEN statcd\_hh\_mem2 != "" THEN 1 ELSE 0 END) + SUM(CASE WHEN statcd\_hh\_mem3 != "" THEN 1 ELSE 0 END) + SUM(CASE WHEN statcd\_hh\_mem4 != "" THEN 1 ELSE 0 END) + SUM(CASE WHEN statcd\_hh\_mem5 != "" THEN 1 ELSE 0 END) + SUM(CASE WHEN statcd\_hh\_mem6 != "" THEN 1 ELSE 0 END) + SUM(CASE WHEN statcd\_hh\_mem7 != "" THEN 1 ELSE 0 END) + SUM(CASE WHEN statcd\_hh\_mem8 != "" THEN 1 ELSE 0 END) AS total FROM emchive;

**2C. Find count of household members by type (Head of Household, Spouse etc.).**

hive> select (case when member=''” then 'no match'

 when member='W' then 'Spouse'

 when member='U' then 'Unknown'

 when member='Y' then 'Young Adult'

 when member='H' then 'Head'

 when member='D' then 'deceased'

 when member='P' then 'Aged Parent Living Home' end), count(\*) From (select statcd\_hh\_mem1 as member from emchive

union all

select statcd\_hh\_mem2 as member from emchive

union all

select statcd\_hh\_mem3 as member from emchive

union all

select statcd\_hh\_mem4 as member from emchive

union all

select statcd\_hh\_mem5 as member from emchive

union all

select statcd\_hh\_mem6 as member from emchive

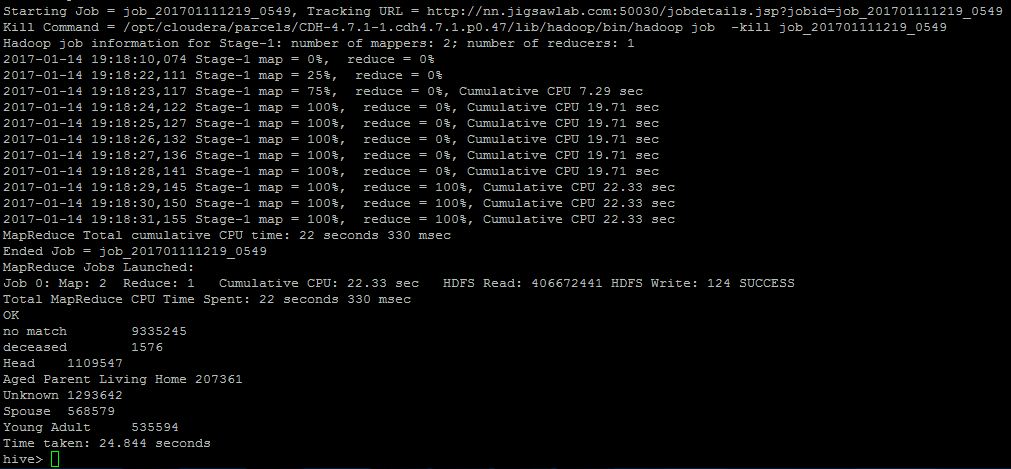
union all

select statcd\_hh\_mem7 as member from emchive

union all

select statcd\_hh\_mem8 as member from emchive) myTab

group by member;



hive> CREATE TABLE 2C as select (case when member=''” then 'no match'

 when member='W' then 'Spouse'

 when member='U' then 'Unknown'

 when member='Y' then 'Young Adult'

 when member='H' then 'Head'

 when member='D' then 'deceased'

 when member='P' then 'Aged Parent Living Home' end), count(\*) From (select statcd\_hh\_mem1 as member from emchive

union all

select statcd\_hh\_mem2 as member from emchive

union all

select statcd\_hh\_mem3 as member from emchive

union all

select statcd\_hh\_mem4 as member from emchive

union all

select statcd\_hh\_mem5 as member from emchive

union all

select statcd\_hh\_mem6 as member from emchive

union all

select statcd\_hh\_mem7 as member from emchive

union all

select statcd\_hh\_mem8 as member from emchive) myTab

group by member;

**2D. %age of household members type. For example, if there are 5 Head of Household, 10 Spouse and 85 in the other categories, then the %age of Spouses is 10.**

hive> CREATE EXTERNAL TABLE householdtype(hhtype  string, hhtypedesc string) ROW FORMAT DELIMITED

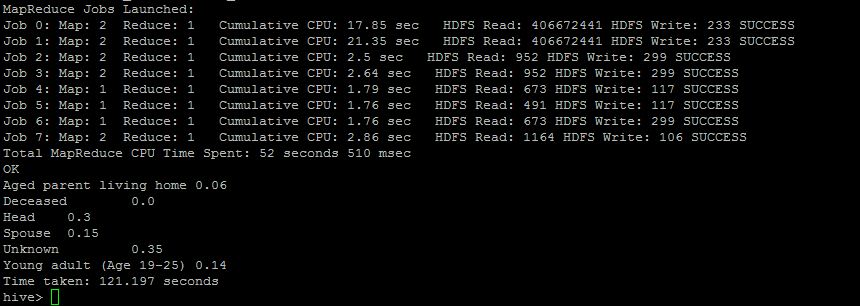
FIELDS TERMINATED BY '|';

hive> LOAD DATA INPATH    ‘household’  into table householdtype;

hive> CREATE VIEW vw\_hhcount AS SELECT hh.hhtypedesc, typecount.hhcount FROM

(SELECT t.hhtype, count(1) hhcount FROM emchive LATERAL VIEW explode(array(statcd\_hh\_mem1, statcd\_hh\_mem2, statcd\_hh\_mem3, statcd\_hh\_mem4, statcd\_hh\_mem5, statcd\_hh\_mem6, statcd\_hh\_mem7, statcd\_hh\_mem8)) t AS hhtype WHERE t.hhtype != ''” GROUP BY t.hhtype) typecount JOIN householdtype hh ON typecount.hhtype = hh.hhtype ORDER BY hh.hhtypedesc;

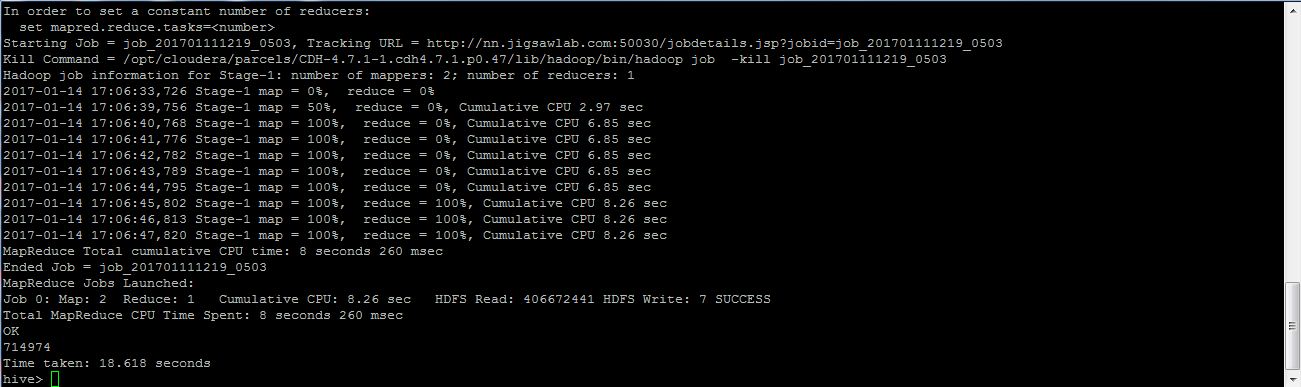
hive> SELECT t1.hhtypedesc, ROUND(t1.hhcount/t2.total, 2) percentoftotal FROM vw\_hhcount t1 JOIN (SELECT SUM(hhcount) total FROM vw\_hhcount) t2;



hive> CREATE TABLE 2D as SELECT t1.hhtypedesc, ROUND(t1.hhcount/t2.total, 2) percentoftotal FROM vw\_hhcount t1 JOIN (SELECT SUM(hhcount) total FROM vw\_hhcount) t2;

**2E. How many known households have children? (use PRESENCE\_OF\_CHLDRN column)**

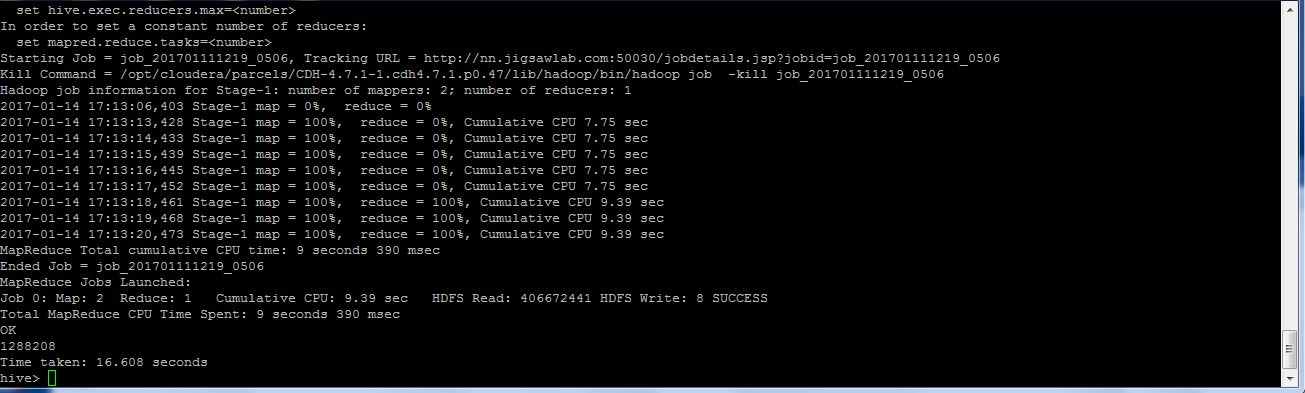
hive> SELECT count(\*) hhwithchildren FROM emchive WHERE PRESENCE\_OF\_CHLDRN = 'Y';



hive> CREATE TABLE 2E as SELECT count(\*) hhwithchildren FROM emchive WHERE PRESENCE\_OF\_CHLDRN = 'Y';

**2F. Overall, how many children are there? (use NUMBER\_OF\_CHLDRN\_18\_OR\_LESS column and PRESENCE\_OF\_CHLDRN )**

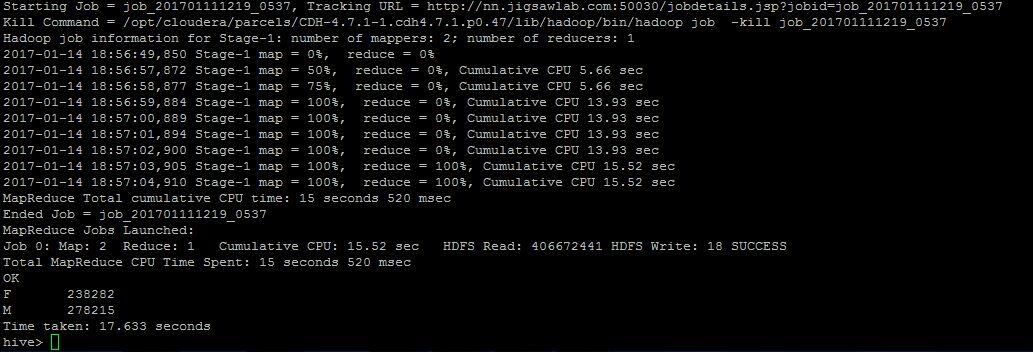
hive> SELECT SUM(CAST(NUMBER\_OF\_CHLDRN\_18\_OR\_LESS AS TINYINT)) FROM emchive WHERE PRESENCE\_OF\_CHLDRN = 'Y';



hive> CREATE TABLE 2F as SELECT SUM(CAST(NUMBER\_OF\_CHLDRN\_18\_OR\_LESS AS TINYINT)) FROM emchive WHERE PRESENCE\_OF\_CHLDRN = 'Y';

**2G. How many of the children are male and how many are female? (use GNDR\_OF\_CHLDRN\_0\_3 - GNDR\_OF\_CHLDRN\_13\_18 column)**

hive> SELECT t2.gender, count(1) gendercount FROM emchive t1 LATERAL VIEW explode(array(GNDR\_OF\_CHLDRN\_0\_3, GNDR\_OF\_CHLDRN\_4\_6, GNDR\_OF\_CHLDRN\_7\_9, GNDR\_OF\_CHLDRN\_10\_12, GNDR\_OF\_CHLDRN\_13\_18)) t2 AS gender WHERE t2.gender='M' OR t2.gender='F' GROUP BY t2.gender;

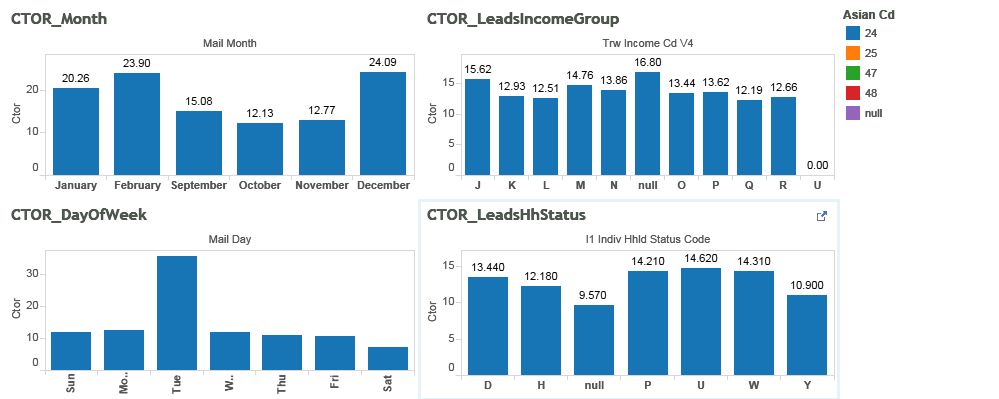
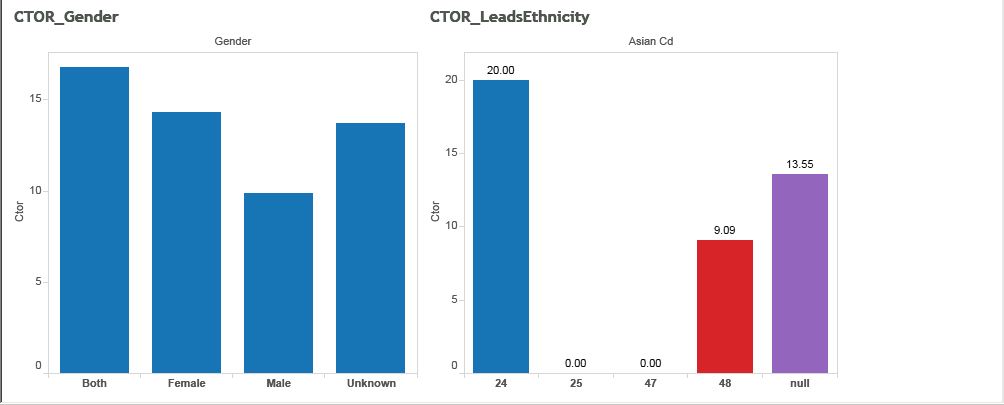


hive> CREATE TABLE 2G AS SELECT t2.gender, count(1) gendercount FROM emchive t1 LATERAL VIEW explode(array(GNDR\_OF\_CHLDRN\_0\_3, GNDR\_OF\_CHLDRN\_4\_6, GNDR\_OF\_CHLDRN\_7\_9, GNDR\_OF\_CHLDRN\_10\_12, GNDR\_OF\_CHLDRN\_13\_18)) t2 AS gender WHERE t2.gender='M' OR t2.gender='F' GROUP BY t2.gender;

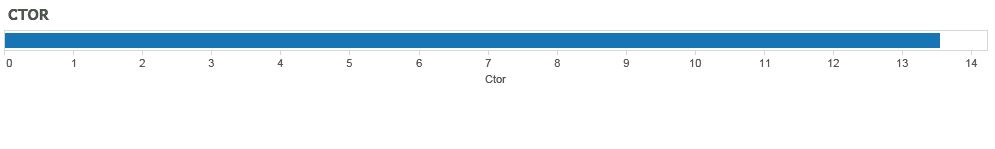
**Tableau**

**Tableau Report for Question1 - CTOR**

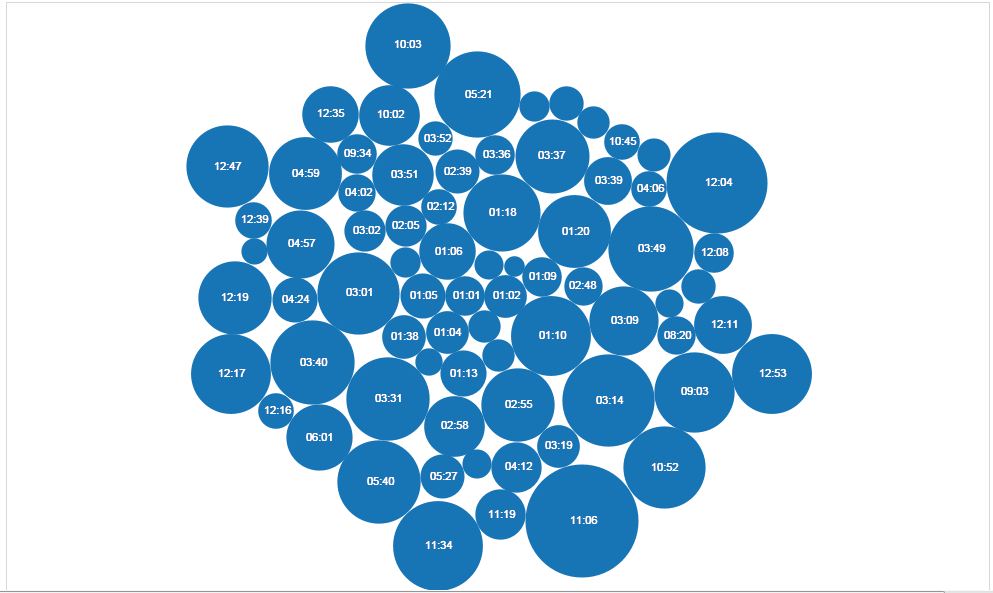
**Dashboard 1**

**Dashboard 2**



**CTOR\_TimeOfDay**

****

**Tableau Report for Question 2 – Household Member’s Information**