**Create a sample dataset and implement the below Pig commands on the same dataset.**

Sample data**:** File Pig\_opeartions.txt

001,Rajiv,Reddy,21,9848022337,Hyderabad,89000

001,Rajiv,Reddy,21,9848022337,Hyderabad,89000

002,Siddarth,Battacharya,22,9848022338,Kolkata,78000

003,Rajesh,Khanna,22,9848022339,Delhi,90000

003,Rajesh,Khanna,22,9848022339,Delhi,90000

004,Preethi,Agarwal,21,9848022330,Pune,93000

005,Trupthi,Mohanthy,23,9848022336,Bhuwaneshwar,75000

006,Archana,Mishra,23,9848022335,Chennai,87000

007,Komal,Nayak,24,9848022334,Trivendram,83000

008,Bharathi,Nambiayar,24,9848022333,Chennai,72000

009,Ram,Reddy,21,9848022337,Hyderabad,89000

010,Seema,Marwa,22,9848022338,Kolkata,76000

File Manager.txt

001,Philips moore

002,Angila Bosco

003,Richard Herry

006,Murli Pcachury

007,Rendy Hillson

008,Patric Ponty

010,Illiaz smith

1) Concat

empdtls = LOAD 'pig\_operations.txt' USING PigStorage(',') as (id:int, firstname:chararray, lastname:chararray, age:int, phone:chararray, city:chararray, monthlyincome:int);

emp\_name\_concat = foreach empdtls Generate CONCAT (firstname,lastname);

dump emp\_name\_concat;

(RajivReddy)

(SiddarthBattacharya)

(RajeshKhanna)

(PreethiAgarwal)

(TrupthiMohanthy)

(ArchanaMishra)

(KomalNayak)

(BharathiNambiayar)

(RamReddy)

(SeemaMarwa)

2) Tokenize

empdtls = LOAD 'pig\_operations.txt' AS (line:chararray);

X\_tokenize = FOREACH empdtls GENERATE TOKENIZE(line);

DUMP X\_tokenize;

({(001),(Rajiv),(Reddy),(21),(9848022337),(Hyderabad),(89000)})

({(002),(Siddarth),(Battacharya),(22),(9848022338),(Kolkata),(78000)})

({(003),(Rajesh),(Khanna),(22),(9848022339),(Delhi),(90000)})

({(004),(Preethi),(Agarwal),(21),(9848022330),(Pune),(93000)})

({(005),(Trupthi),(Mohanthy),(23),(9848022336),(Bhuwaneshwar),(75000)})

({(006),(Archana),(Mishra),(23),(9848022335),(Chennai),(87000)})

({(007),(Komal),(Nayak),(24),(9848022334),(Trivendram),(83000)})

({(008),(Bharathi),(Nambiayar),(24),(9848022333),(Chennai),(72000)})

({(009),(Ram),(Reddy),(21),(9848022337),(Hyderabad),(89000)})

({(010),(Seema),(Marwa),(22),(9848022338),(Kolkata),(76000)})

3) Sum

empdtls = LOAD 'pig\_operations.txt' USING PigStorage(',') as (id:int, firstname:chararray, lastname:chararray, age:int,phone:chararray, city:chararray, monthlyincome:int);

groupByCity = GROUP empdtls BY city;

sumByCity = FOREACH groupByCity GENERATE group,SUM(empdtls.monthlyincome);

DUMP sumByCity;

(Pune,93000)

(Delhi,90000)

(Chennai,159000)

(Kolkata,154000)

(Hyderabad,178000)

(Trivendram,83000)

(Bhuwaneshwar,75000)

---sum all salary

empdtls = LOAD 'pig\_operations.txt' USING PigStorage(',') as (id:int, firstname:chararray, lastname:chararray, age:int,phone:chararray, city:chararray, monthlyincome:int);

emp\_constant\_salary = foreach empdtls Generate 1 as Constant,monthlyincome;

groupByConstant = GROUP emp\_constant\_salary BY Constant;

sumAll = FOREACH groupByConstant GENERATE SUM(emp\_constant\_salary.monthlyincome);

dump sumAll;

(832000)

4) Min

empdtls = LOAD 'pig\_operations.txt' USING PigStorage(',') as (id:int, firstname:chararray, lastname:chararray, age:int,phone:chararray, city:chararray, monthlyincome:int);

groupByCity = GROUP empdtls BY city;

minsalByCity = FOREACH groupByCity GENERATE group,MIN(empdtls.monthlyincome);

DUMP minsalByCity;

(Pune,93000)

(Delhi,90000)

(Chennai,72000)

(Kolkata,76000)

(Hyderabad,89000)

(Trivendram,83000)

(Bhuwaneshwar,75000)

--minsalary in all employee

empdtls = LOAD 'pig\_operations.txt' USING PigStorage(',') as (id:int, firstname:chararray, lastname:chararray, age:int,phone:chararray, city:chararray, monthlyincome:int);

groupByCity = GROUP empdtls BY city;

minsalByCity = FOREACH groupByCity GENERATE group,MIN(empdtls.monthlyincome);

minsal = FOREACH groupByCity GENERATE 1 as aa, MIN(empdtls.monthlyincome) as minmonthlyincomebycity;

groupByAA = GROUP minsal BY aa;

minAll = FOREACH groupByAA GENERATE MIN(minsal.minmonthlyincomebycity);

dump minAll;

(72000)

-----------without group get minSalary

empdtls = LOAD 'pig\_operations.txt' USING PigStorage(',') as (id:int, firstname:chararray, lastname:chararray, age:int,phone:chararray, city:chararray, monthlyincome:int);

emp\_constant\_salary = foreach empdtls Generate 1 as Constant,monthlyincome;

groupByConstant = GROUP emp\_constant\_salary BY Constant;

minAll = FOREACH groupByConstant GENERATE MIN(emp\_constant\_salary .monthlyincome);

dump minAll;

(72000)

5) Max

empdtls = LOAD 'pig\_operations.txt' USING PigStorage(',') as (id:int, firstname:chararray, lastname:chararray, age:int,phone:chararray, city:chararray, monthlyincome:int);

groupByCity = GROUP empdtls BY city;

maxsalByCity = FOREACH groupByCity GENERATE group,MAX(empdtls.monthlyincome);

DUMP maxsalByCity;

(Pune,93000)

(Delhi,90000)

(Chennai,87000)

(Kolkata,78000)

(Hyderabad,89000)

(Trivendram,83000)

(Bhuwaneshwar,75000)

------MAX of all the salary

empdtls = LOAD 'pig\_operations.txt' USING PigStorage(',') as (id:int, firstname:chararray, lastname:chararray, age:int,phone:chararray, city:chararray, monthlyincome:int);

groupByCity = GROUP empdtls BY city;

maxsalByCity = FOREACH groupByCity GENERATE group,MAX(empdtls.monthlyincome);

maxsal = FOREACH groupByCity GENERATE 1 as aa, MIN(empdtls.monthlyincome) as maxmonthlyincomebycity;

groupByAA = GROUP maxsal BY aa;

maxAll = FOREACH groupByAA GENERATE MAX(maxsal.maxmonthlyincomebycity);

dump maxAll;

(93000)

-----------without group get maxSalary

empdtls = LOAD 'pig\_operations.txt' USING PigStorage(',') as (id:int, firstname:chararray, lastname:chararray, age:int,phone:chararray, city:chararray, monthlyincome:int);

emp\_constant\_salary = foreach empdtls Generate 1 as Constant,monthlyincome;

groupByConstant = GROUP emp\_constant\_salary BY Constant;

maxAll = FOREACH groupByConstant GENERATE MAX(emp\_constant\_salary.monthlyincome);

dump maxAll;

(93000)

6) Limit

empdtls = LOAD 'pig\_operations.txt' USING PigStorage(',') as (id:int, firstname:chararray, lastname:chararray, age:int,phone:chararray, city:chararray, monthlyincome:int);

emp = foreach empdtls Generate id,CONCAT(firstname,lastname) as name,age,monthlyincome;

emp\_limit= limit emp 3;

dump emp\_limit;

(1,RajivReddy,21,89000)

(2,SiddarthBattacharya,22,78000)

(3,RajeshKhanna,22,90000)

7) Store

empdtls = LOAD 'pig\_operations.txt' USING PigStorage(',') as (id:int, firstname:chararray, lastname:chararray, age:int,phone:chararray, city:chararray, monthlyincome:int);

emp = foreach empdtls Generate id,CONCAT(firstname,lastname) as name,age,monthlyincome;

dump emp;

store emp into 'PigStore' ;

store emp into 'PigStore' Using PigStorage(':');

8) Distinct

empdtls = LOAD 'pig\_operations.txt' USING PigStorage(',') as (id:int, firstname:chararray, lastname:chararray, age:int,phone:chararray, city:chararray, monthlyincome:int);

empdistinct= DISTINCT empdtls;

dump empdistinct;

(1,Rajiv,Reddy,21,9848022337,Hyderabad,89000)

(2,Siddarth,Battacharya,22,9848022338,Kolkata,78000)

(3,Rajesh,Khanna,22,9848022339,Delhi,90000)

(4,Preethi,Agarwal,21,9848022330,Pune,93000)

(5,Trupthi,Mohanthy,23,9848022336,Bhuwaneshwar,75000)

(6,Archana,Mishra,23,9848022335,Chennai,87000)

(7,Komal,Nayak,24,9848022334,Trivendram,83000)

(8,Bharathi,Nambiayar,24,9848022333,Chennai,72000)

(9,Ram,Reddy,21,9848022337,Hyderabad,89000)

(10,Seema,Marwa,22,9848022338,Kolkata,76000)

9) Flatten

empdtls = LOAD 'pig\_operations.txt' USING PigStorage(',') as (id:int, firstname:chararray, lastname:chararray, age:int,phone:chararray, city:chararray, monthlyincome:int);

groupByCity = GROUP empdtls BY city;

c\_flatten = FOREACH groupByCity generate FLATTEN(group,age),COUNT(empdtls);

(Pune,21,1)

(Delhi,22,2)

(Chennai,23,1)

(Chennai,24,1)

(Kolkata,22,2)

(Hyderabad,21,3)

(Trivendram,24,1)

(Bhuwaneshwar,23,1)

10) IsEmpty

empdtls = LOAD 'pig\_operations.txt' USING PigStorage(',') as (id:int, firstname:chararray, lastname:chararray, age:int,phone:chararray, city:chararray, monthlyincome:int);

mgrdtls = LOAD 'Manager.txt' USING PigStorage(',') as (id:int, managername:chararray);

grpdt = cogroup empdtls by (id),mgrdtls by (id);

filterisempty = filter grpdt by IsEmpty(mgrdtls);

filterresult = foreach filterisempty generate flatten(empdtls);

dump filterresult;

(4,Preethi,Agarwal,21,9848022330,Pune,93000)

(5,Trupthi,Mohanthy,23,9848022336,Bhuwaneshwar,75000)

(9,Ram,Reddy,21,9848022337,Hyderabad,89000)