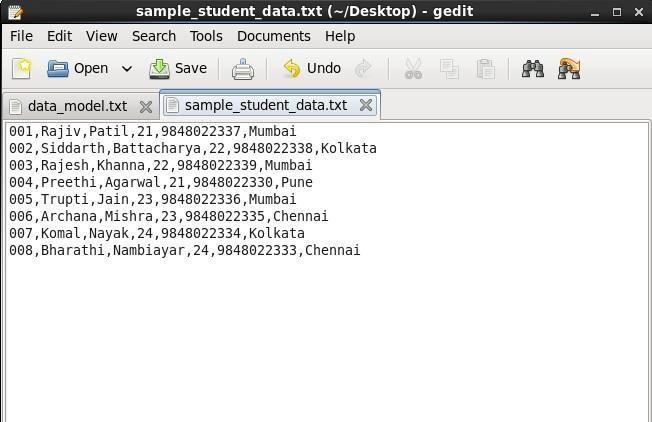
 pig -x local

//Loading data\_model.txt

Grunt>DataModels = LOAD ‘/home/cloudera/data\_model.txt’ using PigStorage(‘;’) AS (name: chararray,address:tuple(city:chararray,pincode:chararray),result:bag{info:tuple(sub:chararray, marks:int)},m:map[int]);

Grunt>dump DataModels;

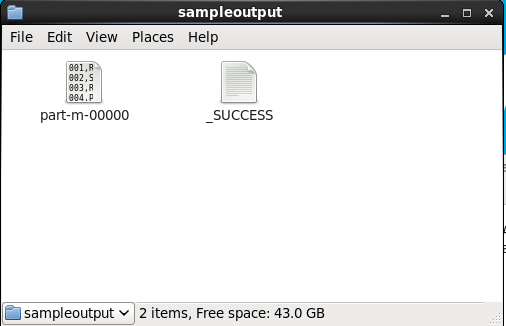
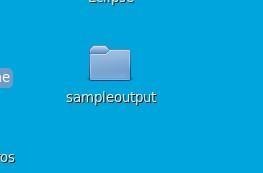


>student\_info = LOAD '/home/cloudera/student\_data.txt' using PigStorage(',') AS (id:chararray,fname:chararray,lname:chararray,age:int,phone:chararray,city:chararray);

>dump student\_info;

//Store Data:

>store student\_info into 'sampleoutput' using PigStorage('|');



Pig Operations:

# Operators

# describe

grunt>describe student\_info;

Filter Data:

**filter**

grunt>filterstudent = filter student\_info BY age > 22;

grunt>dump filterstudent;

**explain**

grunt>EXPLAIN filterstudent;

**foreach**

grunt>foreachstudent = foreach filterstudent generate id, fname, age, city;

grunt>dump foreachstudent;

**GROUP:**

Grunt>groupstudent = GROUP student\_info by city;

Grunt>dump groupstudent;

Grunt>describe groupstudent;

Grunt>groupstudent2 = group student\_info by (city, age);

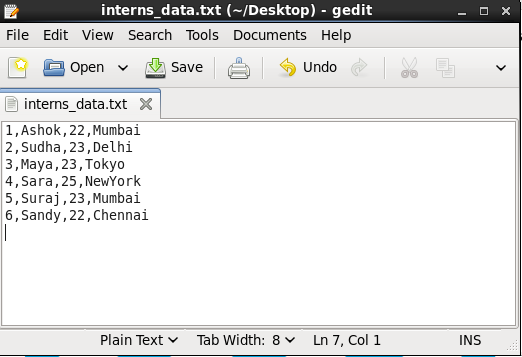
Grunt>dump groupstudent2;

Grunt>groupstudent3 = group student\_info all;

Grunt>dump groupstudent3;

**ILLUSTRATE:**

Grunt>Illustrate foreachstudent;



Grunt>I = LOAD '/home/cloudera/interns\_data.txt' using PigStorage(',') AS (id:chararray,fname:chararray,age:int,city:chararray);

dump I;

**COGROUP:**

Grunt>CG = COGROUP student\_info by age, I by age; grunt> dump CG;

**Join**

**//selfjoin**

grunt>student\_infoB = LOAD '/home/cloudera/Desktop/sample\_student\_data.txt' using PigStorage(',') AS (id:chararray,fname:chararray,lname:chararray,age:int,phone:chararray,city:chararray);

grunt>SelfJoin = Join student\_info by age, student\_infoB by age;

grunt>dump SelfJoin;

**//Inner Join:**

grunt>InnerJoin = JOIN student\_info By city, I by city;

grunt>dump InnerJoin;

# //Left Join:

grunt>LeftJoin = JOIN student\_info By city LEFT, I by city;

grunt> dump LeftJoin;

# //Right Join:

grunt>RightJoin = JOIN student\_info By city RIGHT OUTER, I by city;

grunt> dump RightJoin;

# //Full Join:

grunt>FullJoin = JOIN student\_info By city FULL OUTER, I by city;

grunt>dump FullJoin;

SPLITING:

grunt>split student\_info into younger\_students if age<23, older\_students if age>=23;

grunt>dump student\_info;

FILTERING:

grunt>filter\_city = FILTER student\_info by city == 'Mumbai'; grunt>dump filter\_city;

grunt>all\_city = foreach student\_info generate city;

grunt>dump all\_city;

**distinct**

grunt>distinct\_cities = Distinct all\_city;

grunt>dump distinct\_cities;

**Writing and executing pig latin script**

Q.Write a Pig Latin Script to get the count of each word of a text file, save it in WordCount.pig and execute script to get the required output.(Use any text file as input file. It should have a few statements).

Create files demo1.txt and wordcount.pig in /home/cloudera

**demo1.txt**

MCA NMITD Dadar west MCA

**wordcount.pig**

input2 = LOAD '/home/cloudera/demo1.txt' AS (line:Chararray) ;

Words = FOREACH input2 GENERATE FLATTEN(TOKENIZE(line,' ')) AS word;

Grouped = GROUP Words BY word;

wordcount = FOREACH Grouped GENERATE COUNT(Words), group;

dump wordcount;

grunt> exec /home/cloudera/wordcount.pig