Assignment 9.2

Associated Data Files

<https://github.com/prateekATacadgild/mAPrEDuce/blob/master/studentDS>

Data Set Description: Data Dictionary

1. Name : chararray

2. Col.Undertaken : chararray

3. Dob : chararray

4. Stream : chararray

5. Grade : float

6. State : chararray

7. City : chararray

**students = load '/home/ec2-user/pigdata/studentDS.txt' using PigStorage(',') as(Name:chararray,Col.Undertaken:chararray,Dob:chararray,Stream:chararray,Grade:FLOAT,State:chararray,City:chararray);**

Problem Statements

1. **Find the number of students who scored less than 5(50%).**

students\_filtered = FILTER students By Grade<5;

X = GROUP students\_filtered ALL;

CNT = foreach X generate COUNT(students\_filtered);

dump CNT;

(188)

2**. List the name of students who are from Alaska.**

students\_alaska = FILTER students By State matches '(?i).\*Alaska.\*';

1. **How many government collages are there in Alabama?**

Students\_from\_ govtUndertaken\_coll\_Alabama = FILTER students1 By state matches '(?i).\*Alabama.\*' AND **Col.Undertaken** =='goverenment';

Not enough data set to find out the number of government colleges. No college name mentioned.

1. **List the name of students who are from Oregon and pursuing BE**

X = FILTER students By state matches '(?i).\*Oregon.\*' AND Stream =='BE';