Session 9: Pig Demo-2Assignment 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

# Problem Statements

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

**Ans:**

A = load ‘studentDS’ Using PigStorage(‘,’) As ( Name:chararray,ColUndertaken:chararray,Dob:chararray,Stream:chararray,Grade:float,State:chararray,City:chararray);

B= FILTER A BY Grade<5.0;

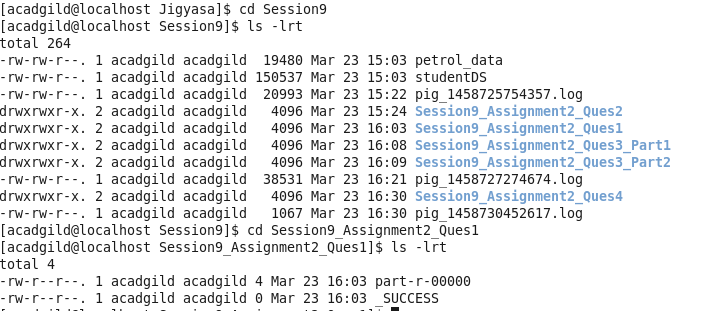
C= GROUP B ALL;

D= FOREACH C GENERATE COUNT(B);

**Output:**



STORE D into ‘Session9\_Assignment2\_Ques1’;



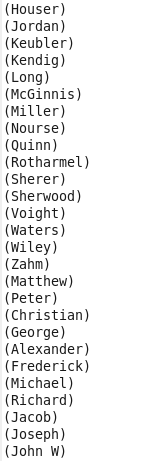
1. List the name of students who are from Alaska.

**Ans:**

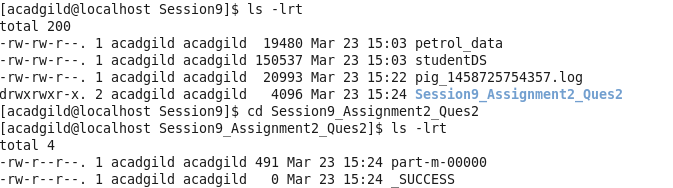
A = load ‘studentDS’ Using PigStorage(‘,’) As ( Name:chararray,ColUndertaken:chararray,Dob:chararray,Stream:chararray,Grade:float,State:chararray,City:chararray);

B= FILTER A BY State==’ alaska’;

C= FOREACH B Generate Name;



STORE C into ‘Session9\_Assignment2\_Ques2’;



1. How many government collages are there in Alabama?

**Ans:** As the 2 spellings (government and government) are given for government colleges:

A = load ‘studentDS’ Using PigStorage(‘,’) As ( Name:chararray,ColUndertaken:chararray,Dob:chararray,Stream:chararray,Grade:float,State:chararray,City:chararray);

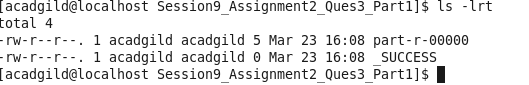
B= FILTER A BY State==’ alabama’ and ColUndertaken ==’ goverenment’;



C= GROUP B ALL;

D= FOREACH C GENERATE COUNT(B);

STORE D into ‘Session9\_Assignment2\_Ques3\_Part1’



Or

A = load ‘studentDS’ Using PigStorage(‘,’) As ( Name:chararray,Col.Undertaken:chararray,Dob:chararray,Stream:chararray,Grade:float,State:chararray,City:chararray);

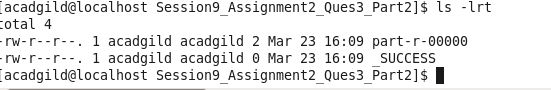
B = FILTER A BY State==’ alabama’ and ColUndertaken ==’ goverement’;



C= GROUP B ALL;

D= FOREACH C GENERATE COUNT(B);

STORE D into ‘Session9\_Assignment2\_Ques3\_Part2’;

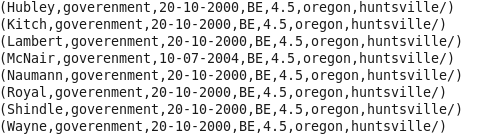


1. List the name of students who are from Oregon and persuing BE.

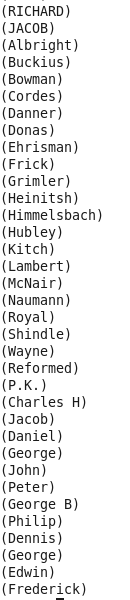
**Ans:**

A = load ‘studentDS’ Using PigStorage(‘,’) As ( Name:chararray,ColUndertaken:chararray,Dob:chararray,Stream:chararray,Grade:float,State:chararray,City:chararray);

B= FILTER A BY State==’ oregon’ and $3==’ BE’;--- We are using $3 instead of Stream because it is a key word



C= FOREACH B GENERATE Name;



STORE C into ‘Session9\_Assignment2\_Quse4’;

