Start hadoop daemons

pig

REGISTER '/home/acadgild/pig-0.16.0/lib/piggybank.jar';

A = load 'ass-hdfs/studentDS.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO\_MULTILINE','UNIX','SKIP\_INPUT\_HEADER');

B = FOREACH A GENERATE (chararray)$0 AS Name, (chararray)$1 AS Col\_Undertaken, (chararray)$2 AS DOB, (chararray)$3 AS Course\_Stream, (float)$4 AS Grade, (chararray)$5 AS State, (chararray)$6 AS City;

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

C = FILTER B BY Grade < 5;

D = GROUP C all;

lessThanFive = FOREACH D GENERATE COUNT(C.Grade);

DUMP lessThanFive;



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

D = FILTER B BY State == 'alaska';

E = FOREACH D GENERATE Name;

Answer attached separate file Assign9-2-Ans-2.txt

1. How many government collages are there in Alabama?

F = FILTER B BY (Col\_Undertaken == 'goverenment' OR Col\_Undertaken == 'goverement') AND State == 'alabama';

STORE F INTO 'ass-hdfs/Student\_Answer3';

Answer attached separate file Assign9-2-Ans-3.txt

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

G = FILTER B BY State == 'oregon' AND Course\_Stream == 'BE';

STORE G INTO 'ass-hdfs/Student\_Answer4';

Dump output in file Assign9-2-Ans-4.txt