**Session 9: Pig**

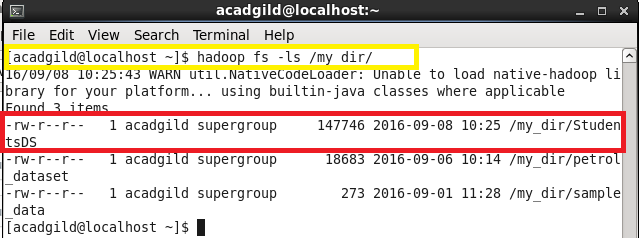
**Assignment 2**

**Steps**

1. Put the StudentsDS into hadoop file system using **command put as given below**.

hadoop fs –put /home/acadgild/SampleDatasets/StudentsDS /my\_dir/

2. Type **ls command** to see list as given below.



3. Start the grunt shell using **pig –x mapreduce command**

4. Load the data

A = LOAD '/data\_directory\_sample.txt' using PigStorage(',') as (name:chararray,col:chararray,dob:chararray,Stream:chararray,Grade:float,state:chararray,city:chararray);

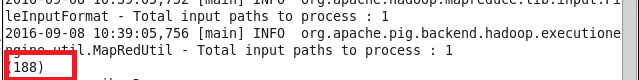
**Q 1. Find the number of students who scored less than 5(50%).**

B = FILTER A by Grade < 5;

C = GROUP B ALL:

STD\_COUNT = FOREACH C GENERATE COUNT(B);

dump STD\_COUNT;



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

D = FILTER A by state =='alaska';

STD\_NAME = FOREACH D GENERATE name;

dump STD\_NAME;



**Q 3. How many government collages are there in Alabama?**

E = FILTER A by state =='alabama' AND col == 'goverenment';

F = GROUP E ALL;

STATE\_CNT = FOREACH F GENERATE group, COUNT(E);

dump STATE\_CNT;



**Q 4. List the name of students who are from Oregon and persuing BE.**

G = FILTER A by state =='oregon' AND $3== 'BE';

ORE\_NAME = FOREACH G GENERATE name;

dump ORE\_NAME;

