**Assignment 1: Pig Commands Report**

**Assignment**: Find out successful students in a class.

1. **Problem statement**: Write a Pig script or Pig Statements to analyze the given datasets and print the student names who have successfully cleared the exam. You can only use PIG commands to achieve the desired result.

**Pig Commands used in the Assignment:**

**1.LOAD:** It is used to load data from the filesystem or HDFS storage to a pig relation.

E.g. loadin2 = LOAD ‘\root\students.txt’ USING PigStorage (‘,’) AS (name: chararray, rollno: int);

loading2 = LOAD ‘\root\result.txt’ USING PigStorage (‘,’) AS (name: chararray, rollno: int);

**2.DUMP:** It is used in Pig Latin statements & display the results on the screen.

E.g. DUMP loading1;

DUMP loading2;

**3.FI** **LTER:** It selects tuples from a relation based on a condition.

E.g. A = FILTER B BY results==’pass’;

**4.JOIN:** It is used to perform an inner join ,equijoin or two or more relations based on .

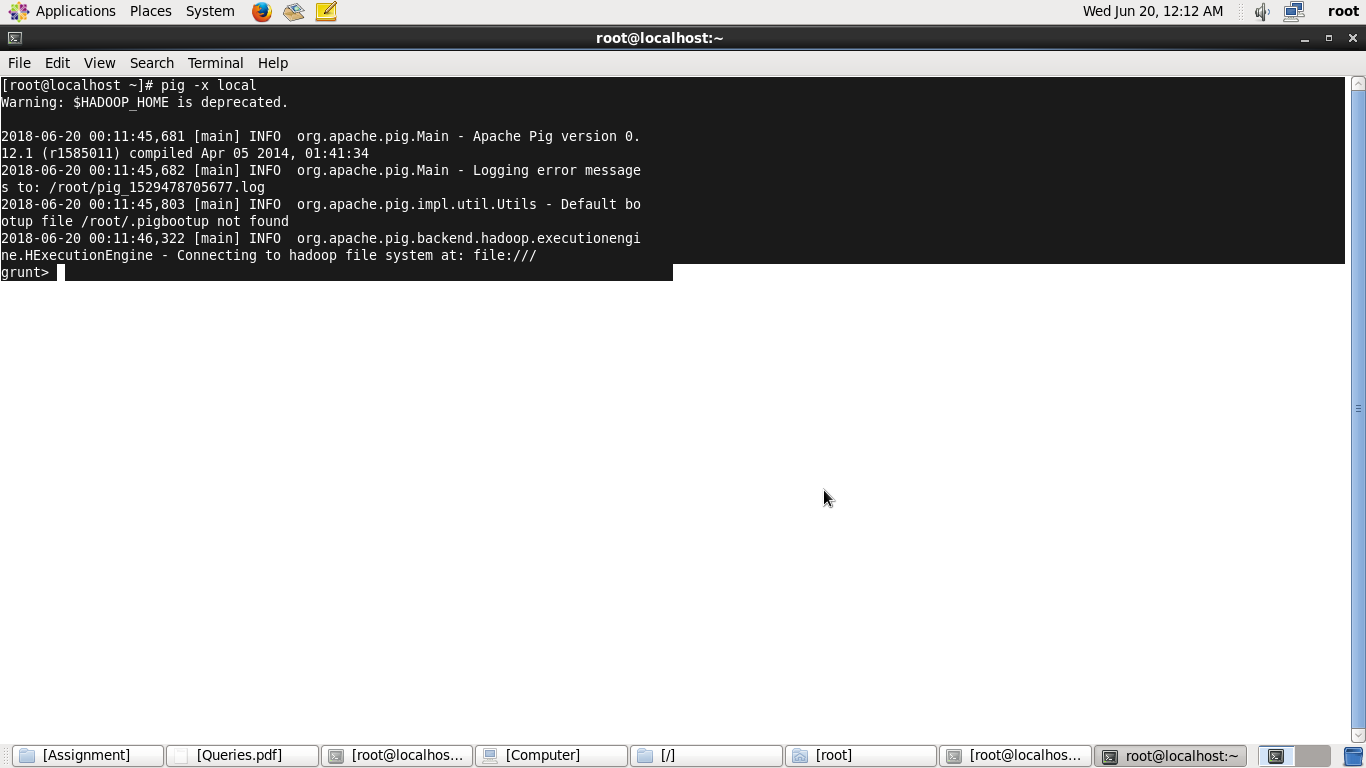
A = JOIN A BY rollno, loading1 BY rollno;

**5.FOREACH GENERATE:** This Operators generate data transformation based on columns of data. It removes or add fields to the relation

E.g. d = FOREACH A GENERATE name;

**Screenshots:**

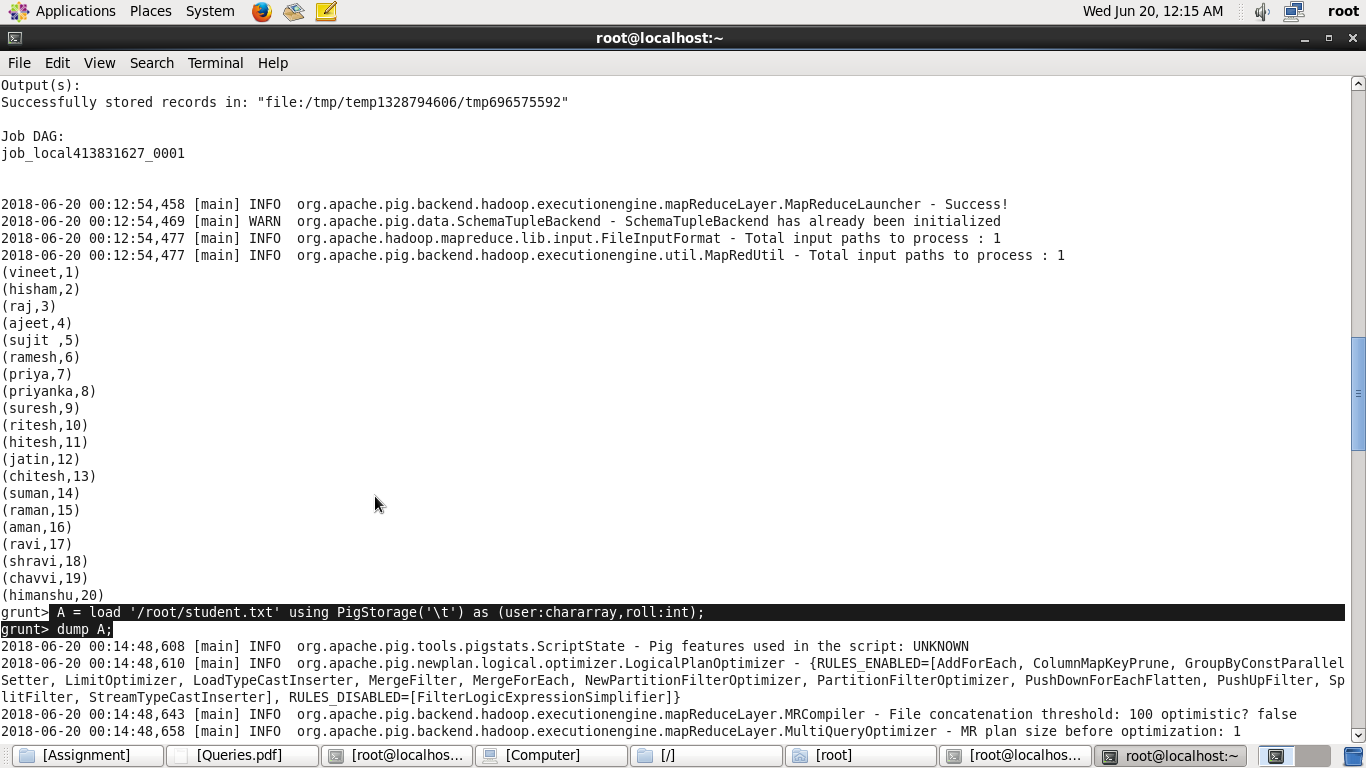
1)First we open the terminal and we enter the command 'pig -x local' which takes us to the local execution mode.



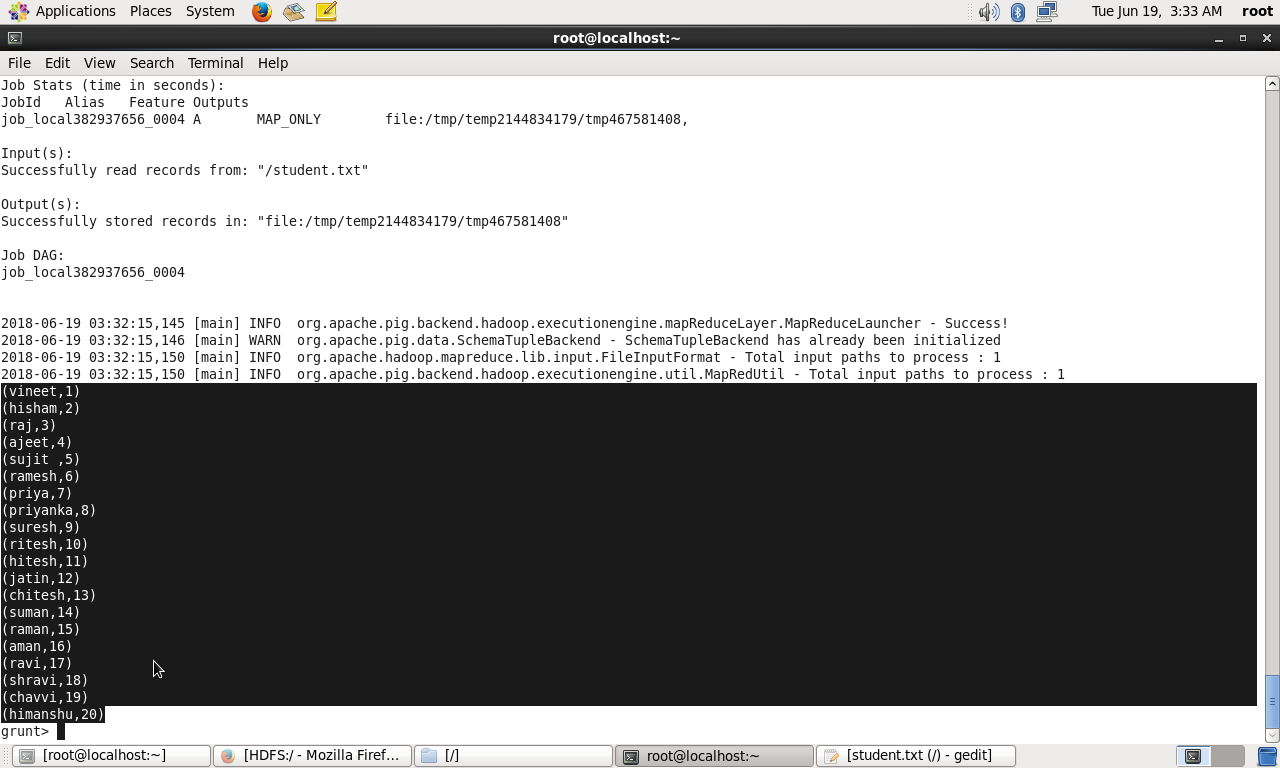
**fig: command to enter the pig local execution mode**

2) After we enter I local mode, we load the dataset 'student.txt' in pig relation using LOAD command and dump it using DUMP command:

**grunt> A = load '/student.txt' using PigStorage('\t') as(user:chararray,roll:int);**

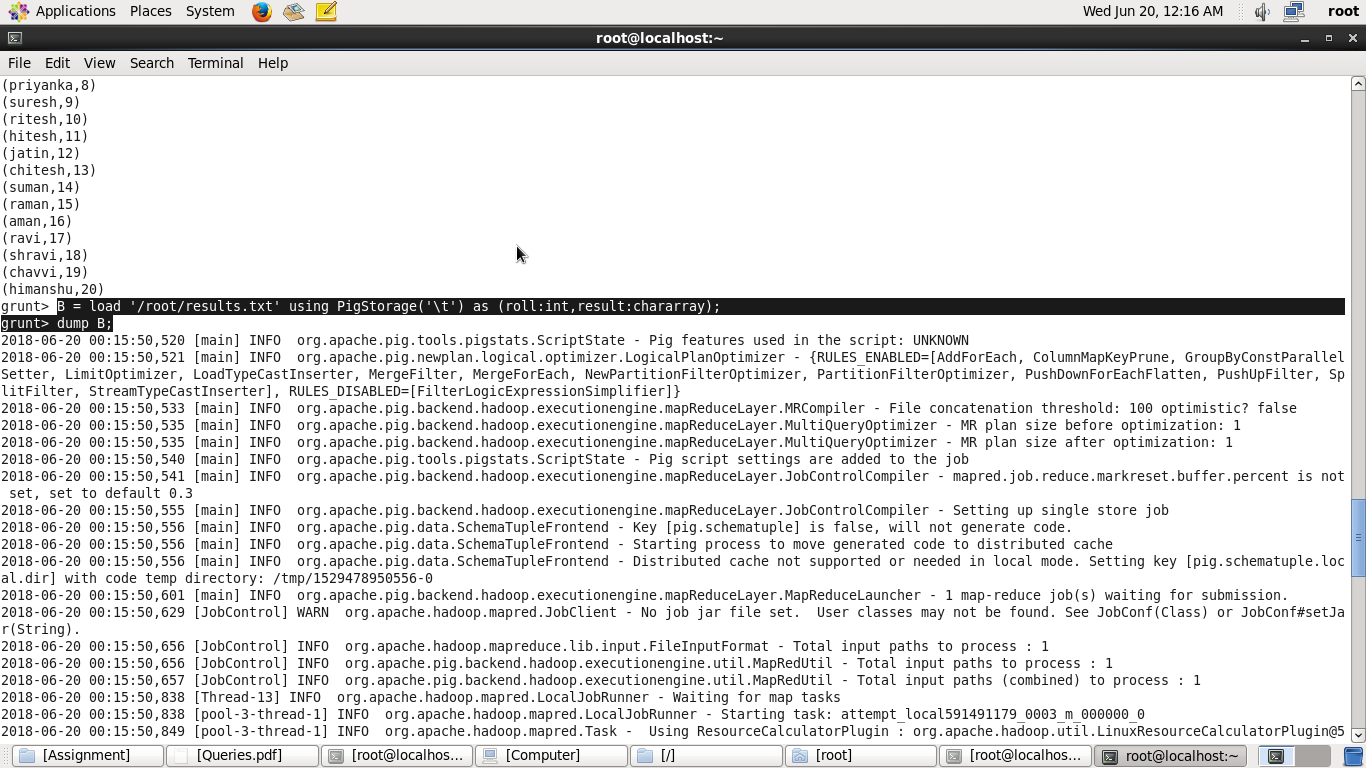
** fig1: load command**

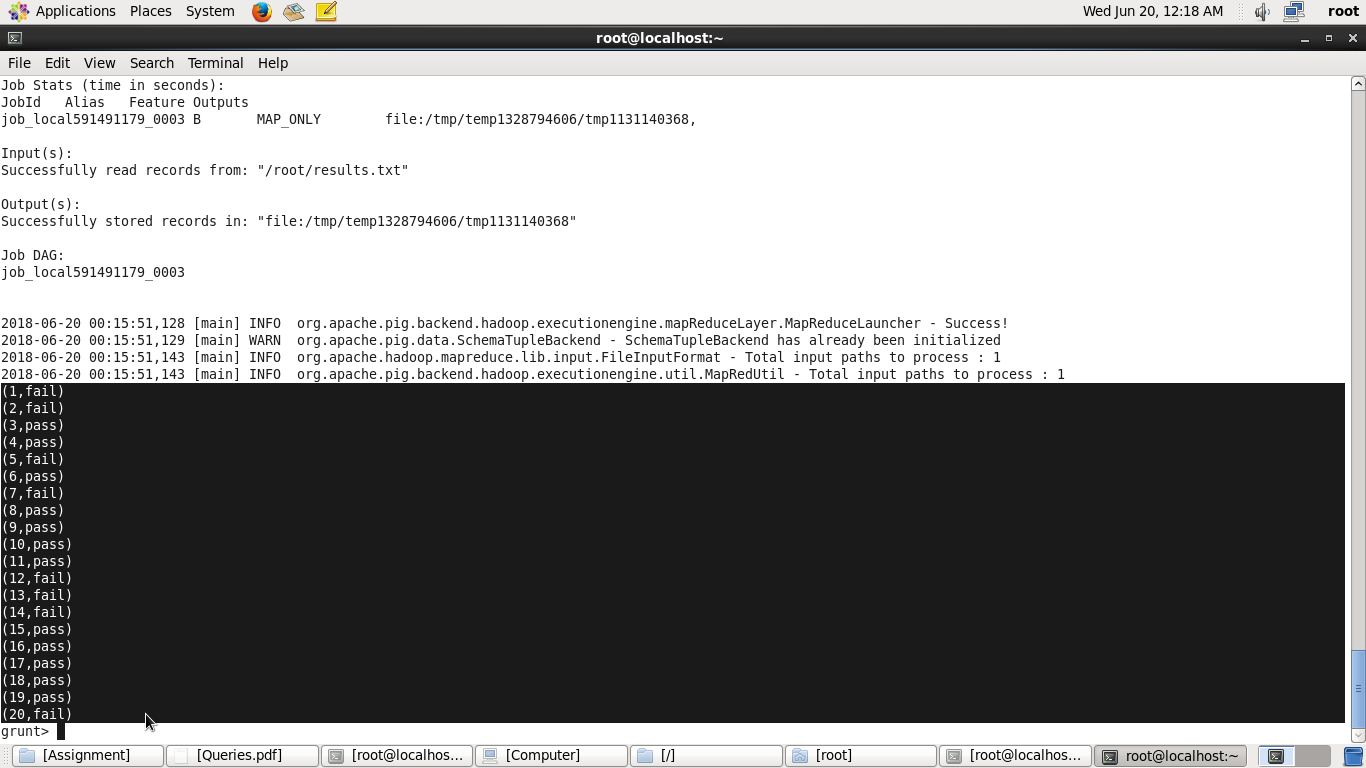
**grunt> dump A;**

** fig2: dump command**

3) Now we load the dataset 'results. txt' in pig relation using LOAD command and dump it using DUMP command.

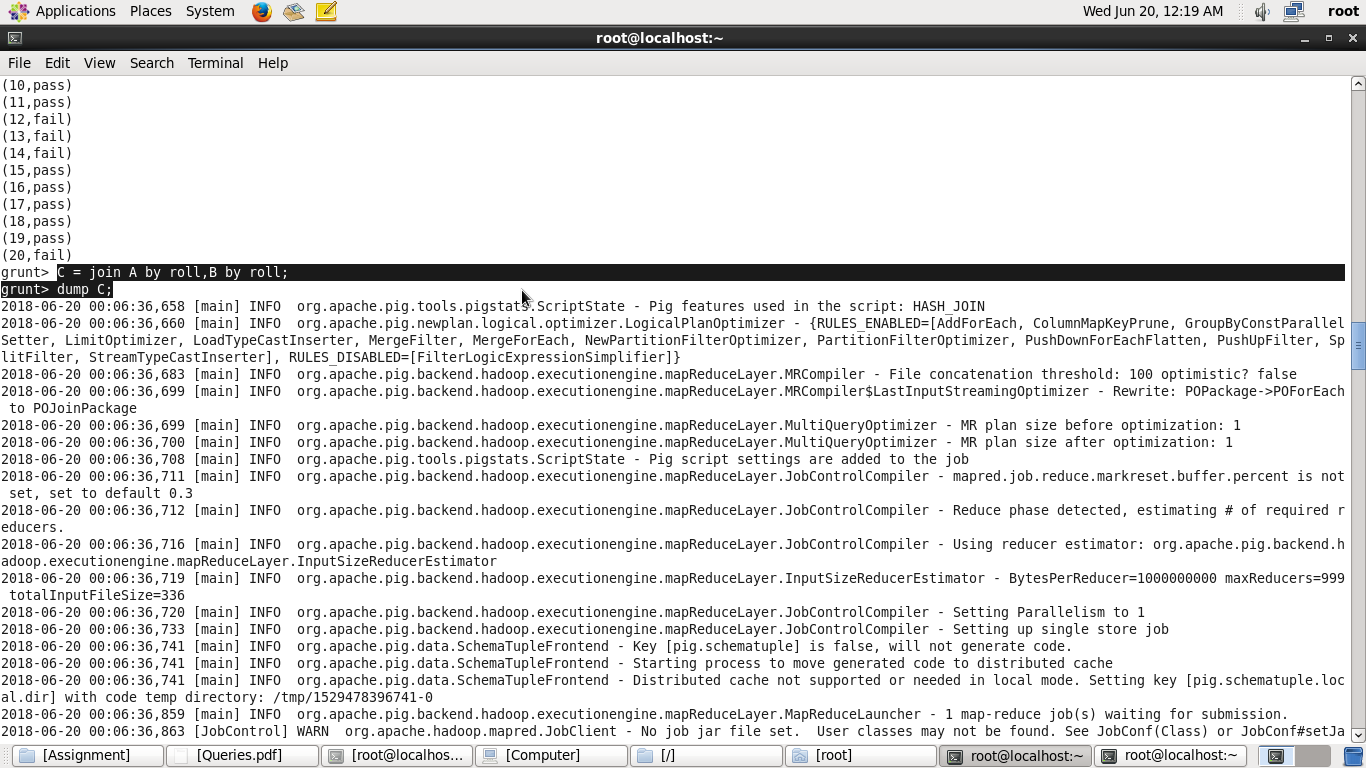
**grunt> B = load '/results.txt' using PigStorage(',') as(roll:int,result:chararray);**

** fig3: load command**

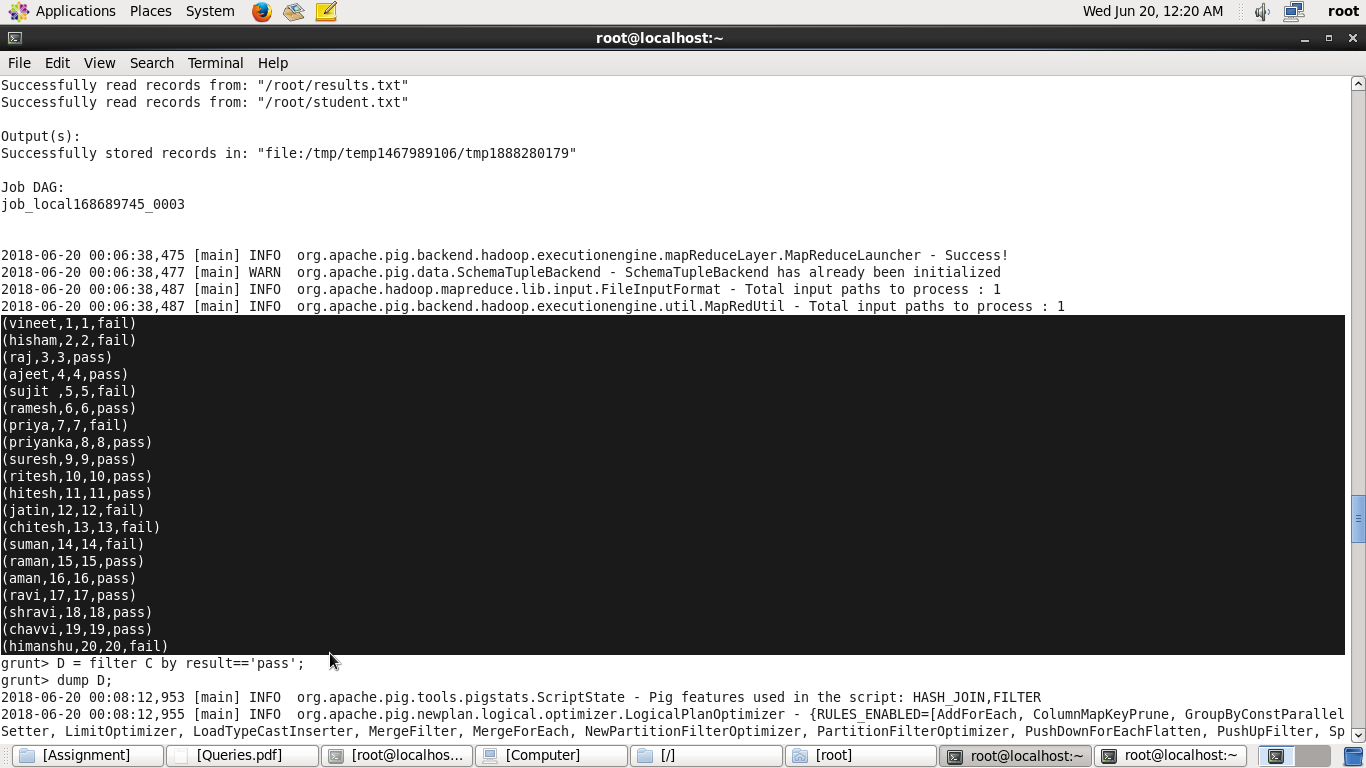
**grunt > dump B;  fig 4: dump command**

4)Now we join the two relations using the JOIN command

**grunt> C = join A by roll, c by roll;**

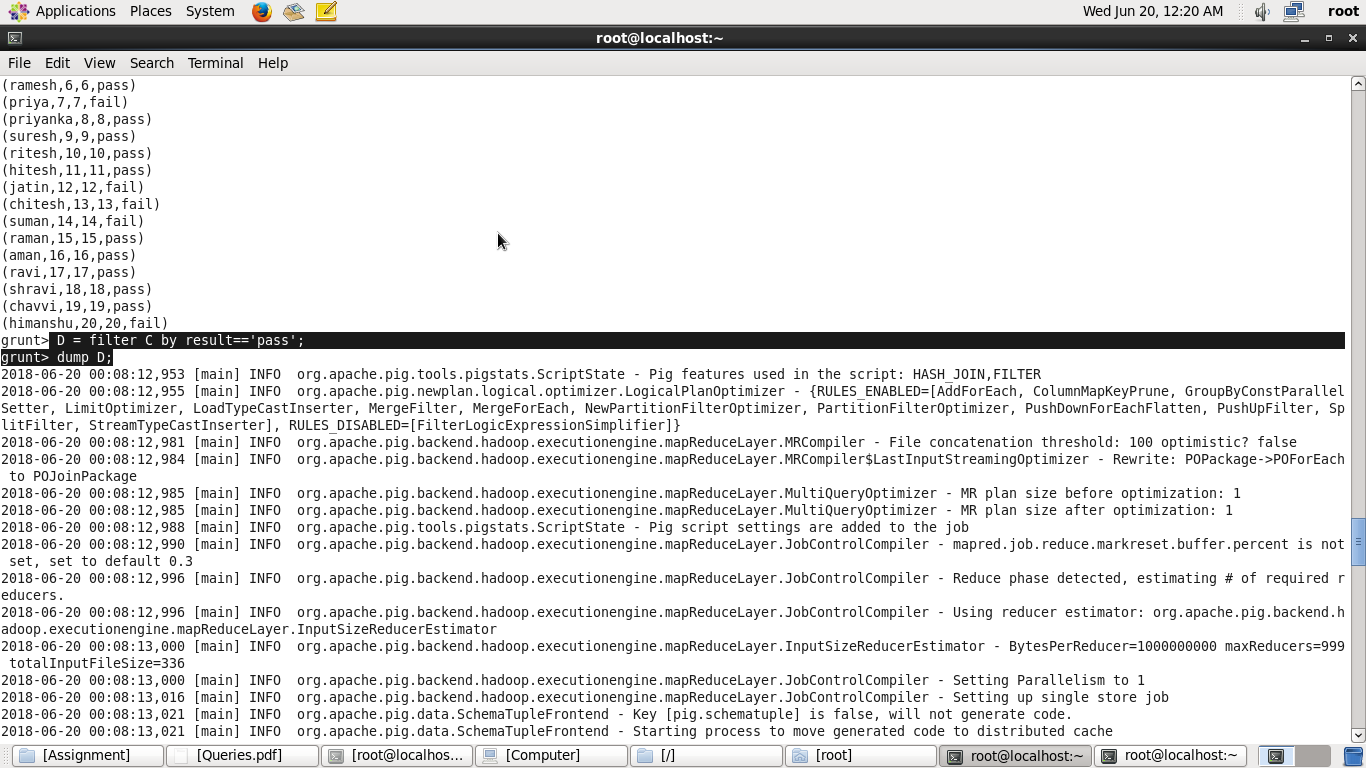
** fig5: join command**

**grunt > dump C;**

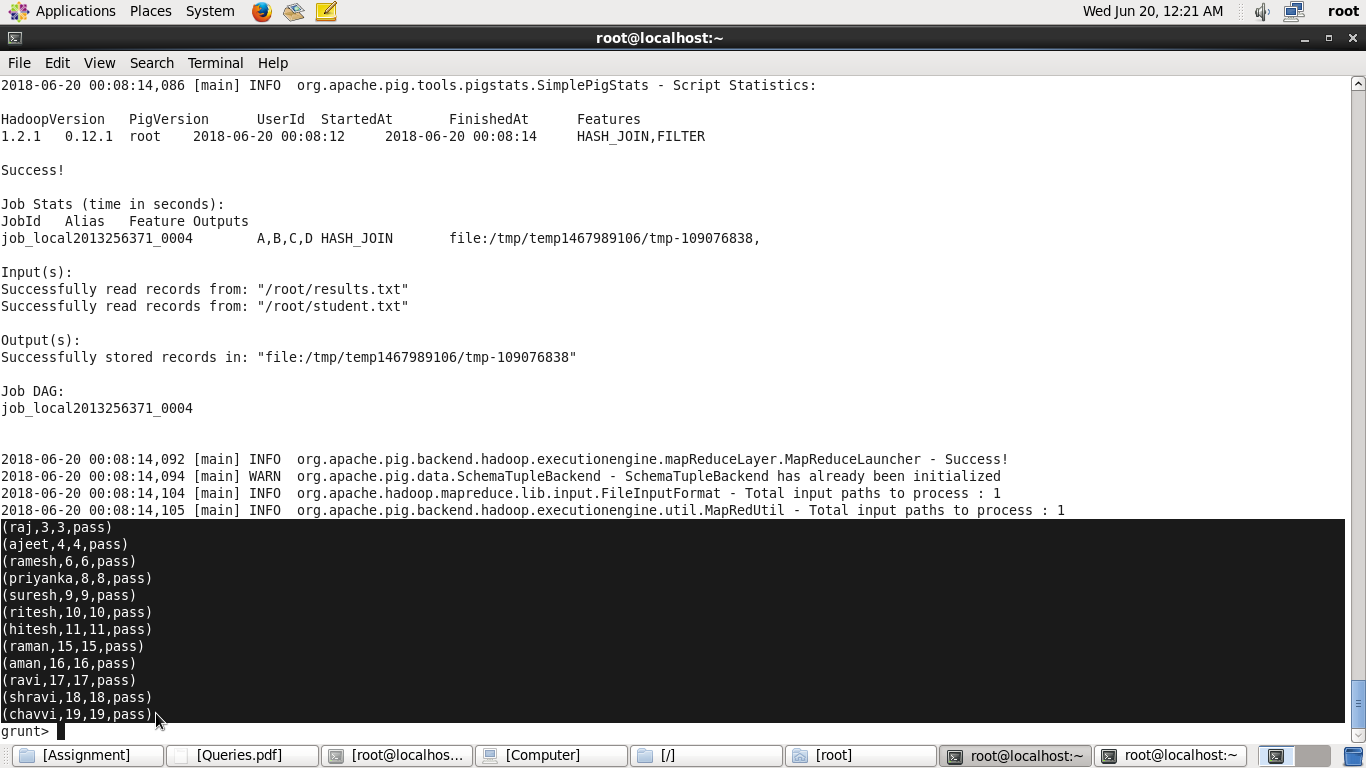
** fig6: dump command**

5)After joining the two relations we print only the names of the students who have passed using the FILTER command.

**grunt > D = filter f by result =='pass';**

** fig7: filter command**

**grunt > dump D;**

** fig8: dump command**