**Problem Statement**

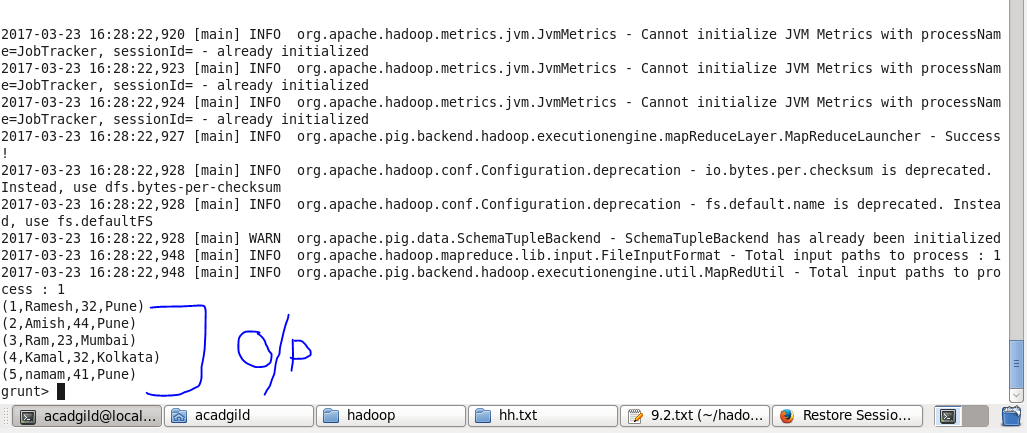
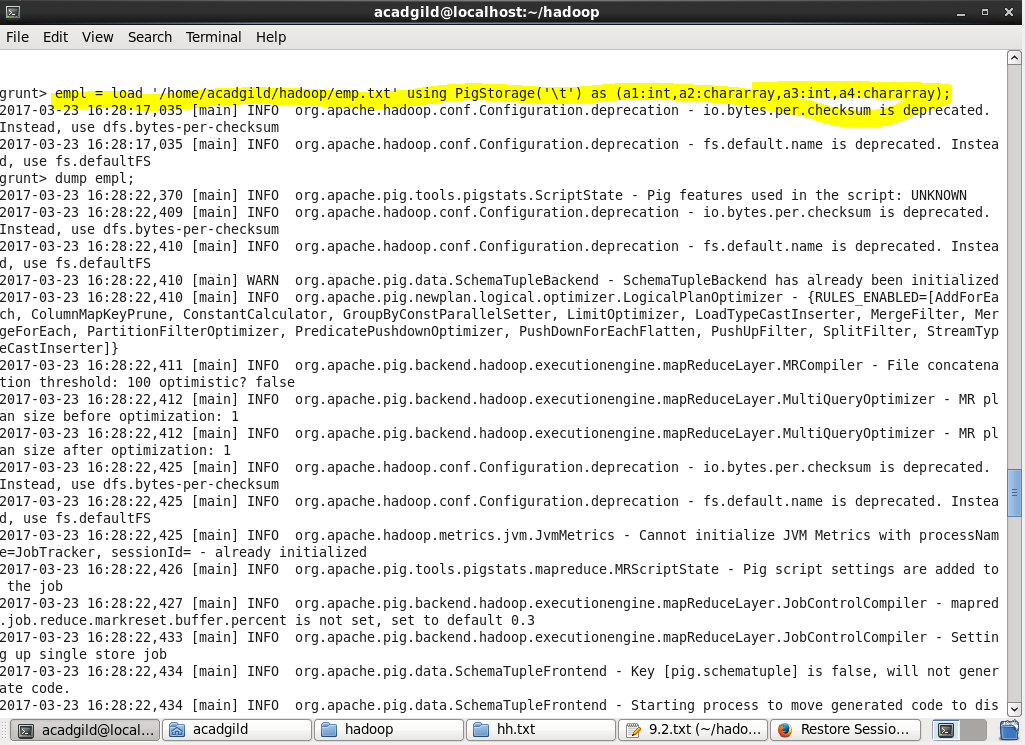
1. LOAD – LOAD command is used to load data from file system.

Syntax- A=LOAD ‘myfile.txt’ USING PigStorage(‘ ‘) AS (f1:int,f2:int);

myfile.txt - is the file to be loaded.

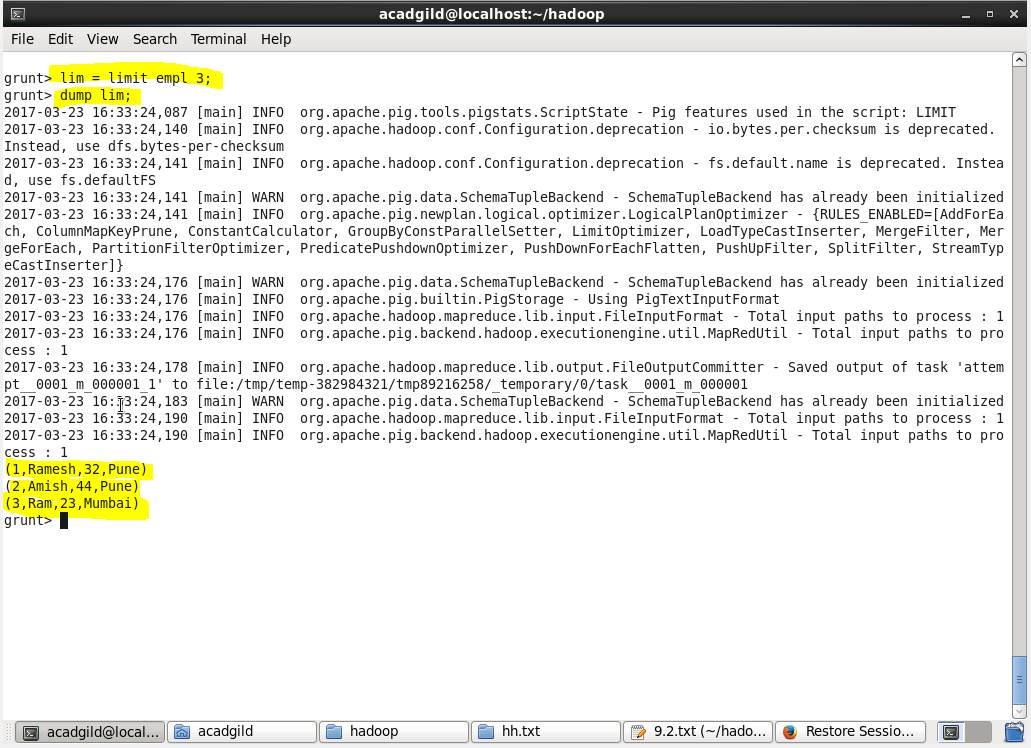
PigStorage - is specifying separator.

(f1:int,f2:int) – specify the schema and data types for fields.



1. LIMIT- Used to limit number of outputs to the desired number.

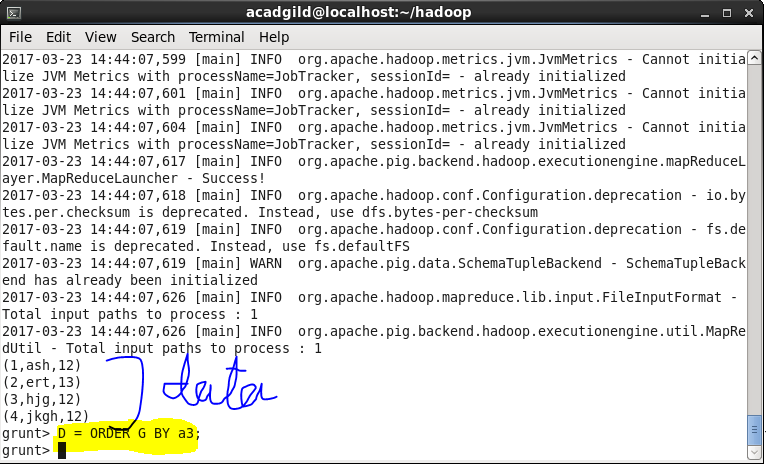
Syntax- alias = LIMIT alias n;

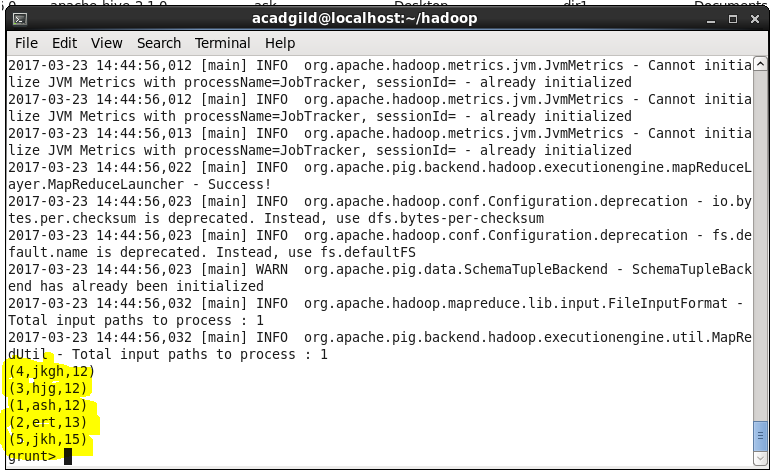


1. ORDER BY – Sorts a relation based on one or more fields.

Syntax- A= LOAD ‘data’ AS (a1:int,a2:int,a3:int)

X= ORDER A BY a3 DESC;

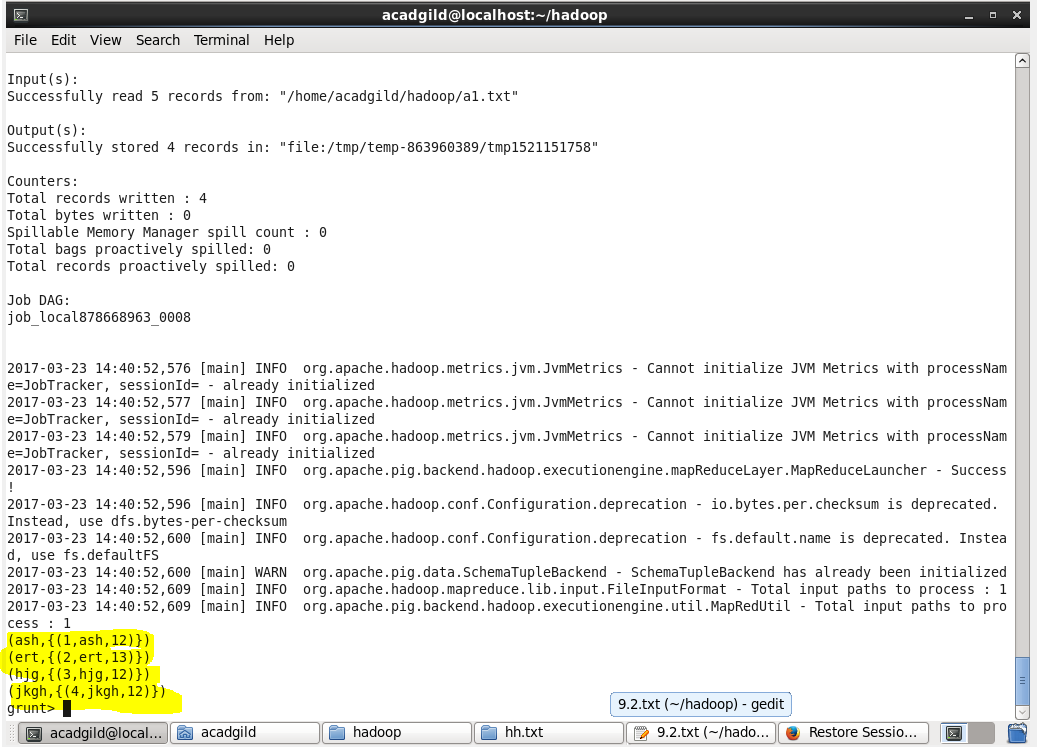
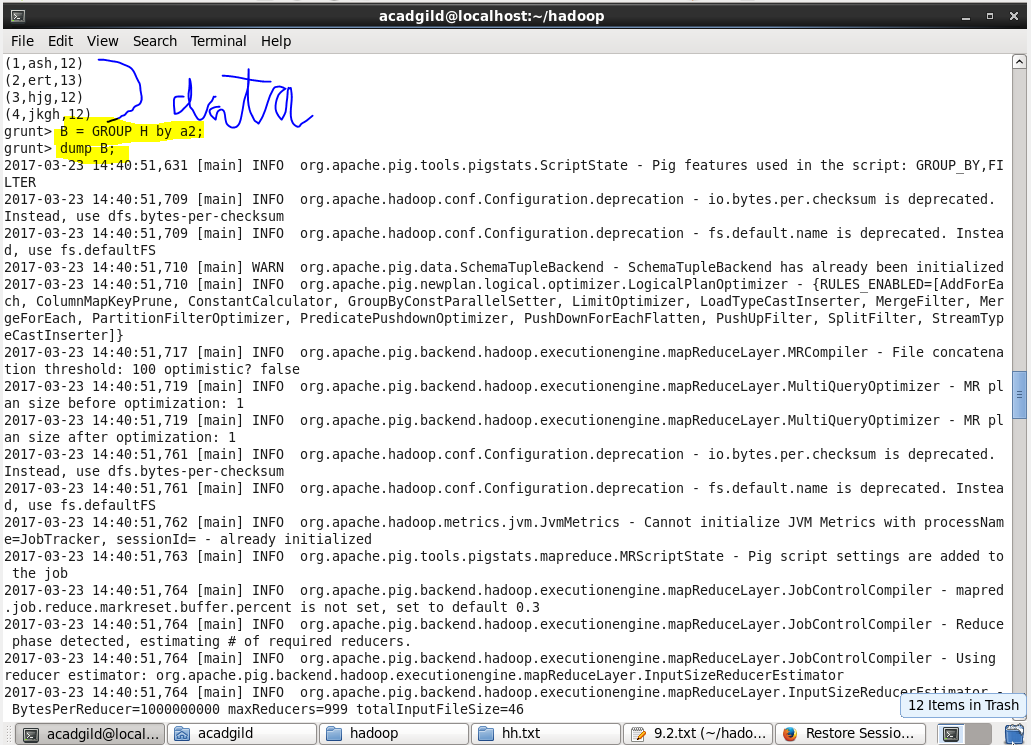




1. GROUP BY- Group the data in one or more relations. The GROUP operator groups together tuples that have the same group key (key field)

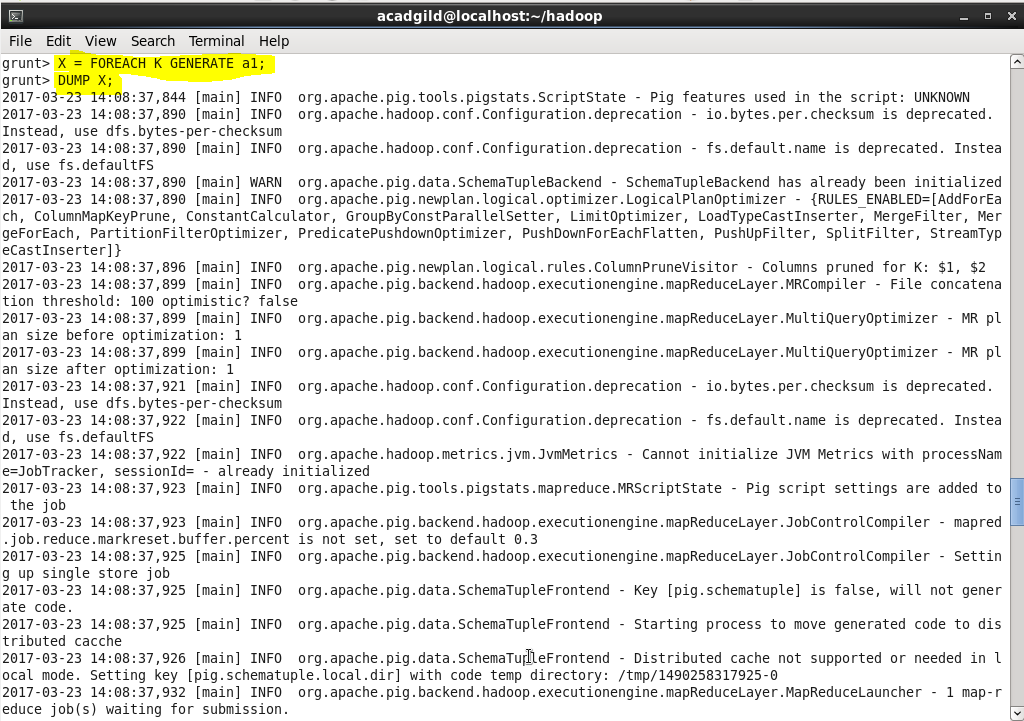
Syntax- A= LOAD ‘student’ AS (name:chararray,age:int,gpa:float);

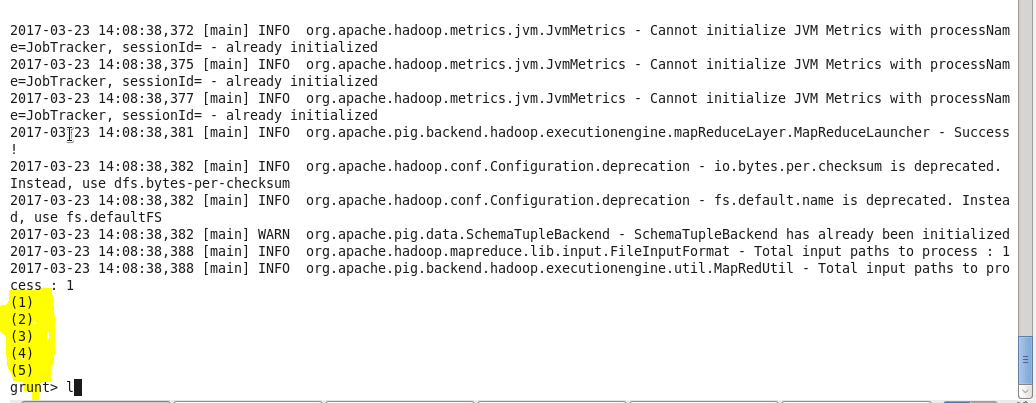
B=GROUP A BY age;



1. FOREACH – Generates transformation of data for each row as specified.

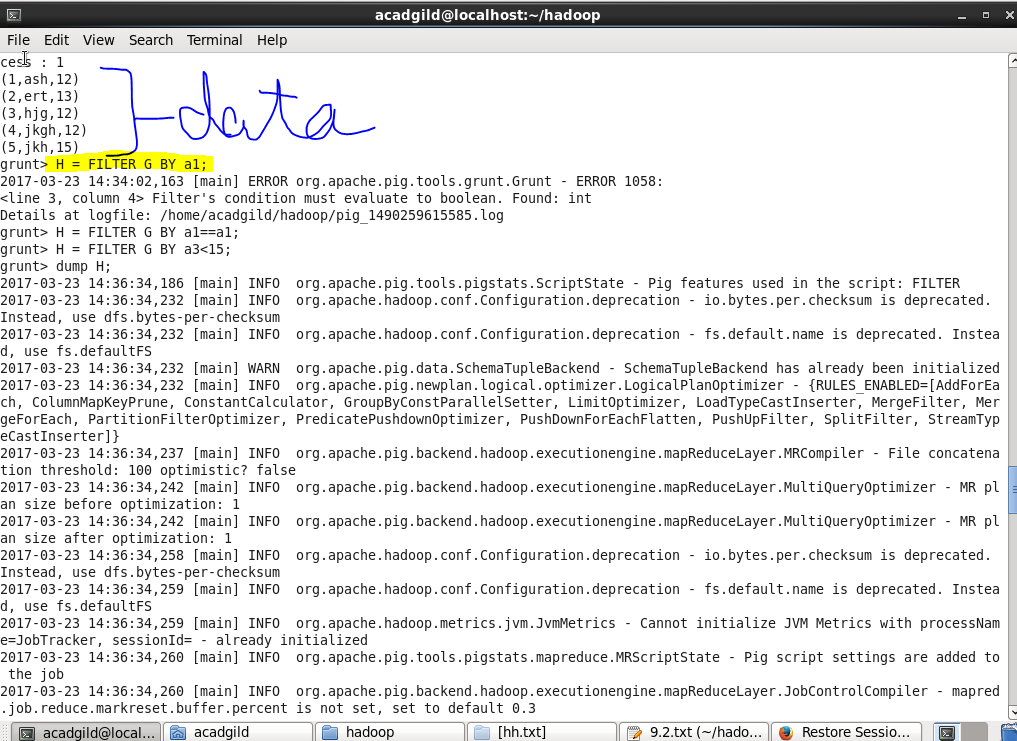
Syntax- X=FOREACH A GENERATE a1;

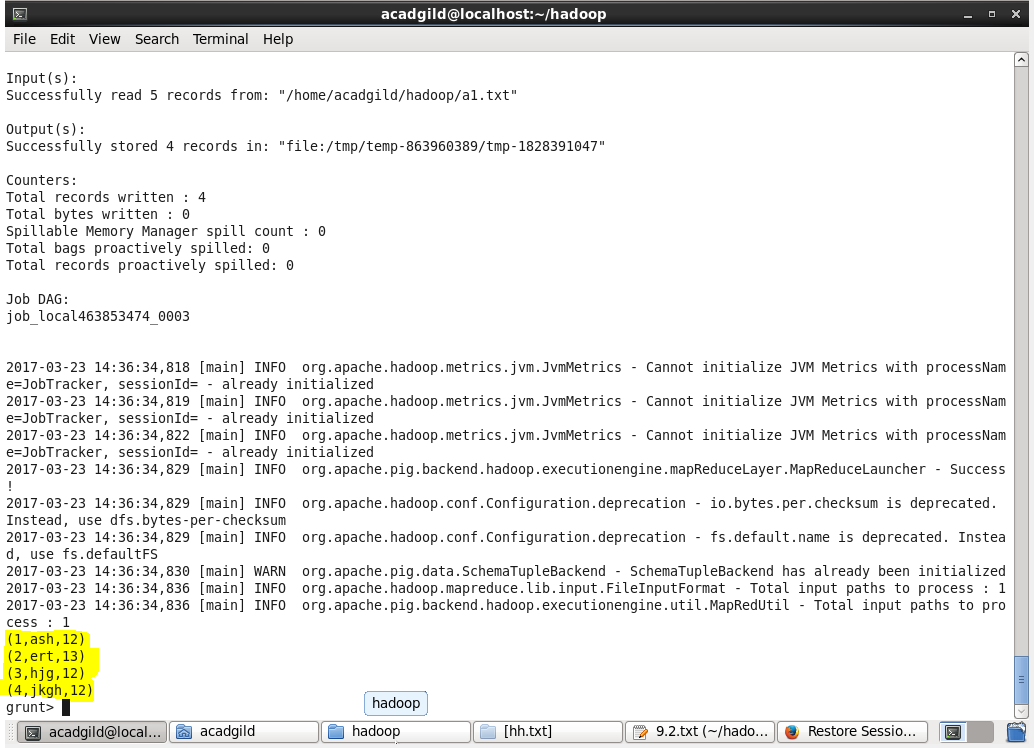




1. FILTER – Select tuples from a relation based on some condition.

Syntax- X=FILTER A by f3==f3;





1. STORE- Store or saves file to the file system.

Syntax- STORE A INTO ‘myoutput’ USING PigStorage (‘\*’);

