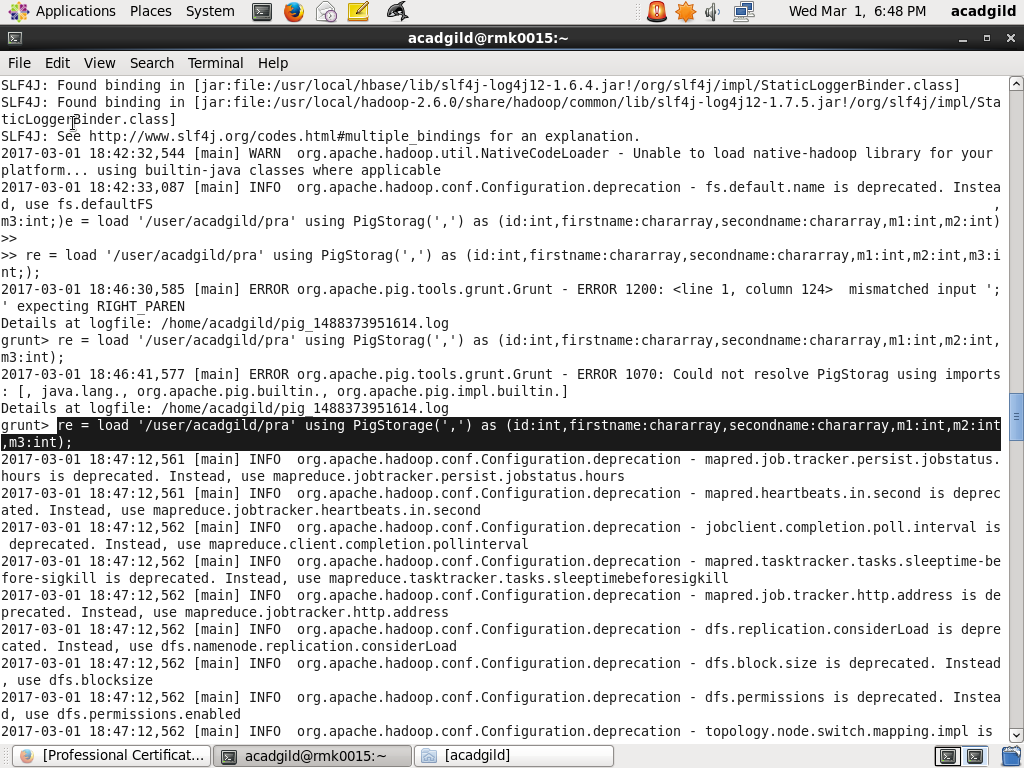
ASSIGNMENT7.1

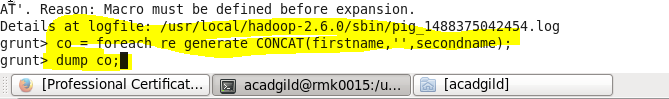
re = load ‘/user/acdgild/pig’ using PigStorage(‘,’) as (id:int ,firstname:chararray, secondname: chararray ,m1:int,m2:int,m3:int);



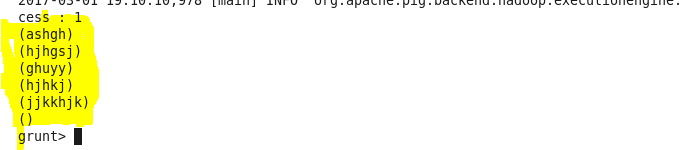
1)CONCAT:

The concat function in pig latin is used to concat two or more expressions.

ex: co = foreach re CONCAT(firstname,'\_',lastname);



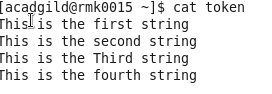
dump co;

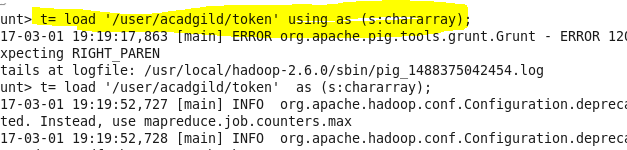


2)TOKENIZE:

The tokenize function is used to split a string in a single tuple and returns a bag which contains the output of the split operation.

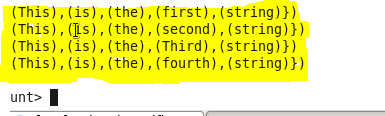
ex:





X = foreach A generate TOKENIZE(f1);

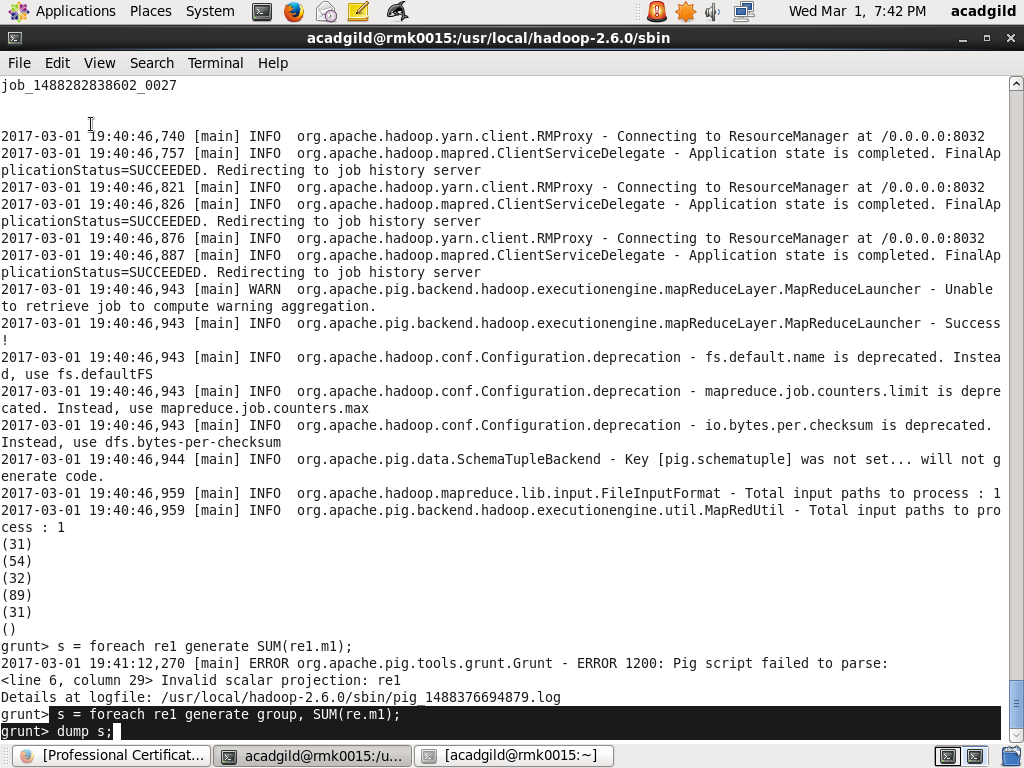
C:\Users\rmk\Pictures\3,7.1.PNG

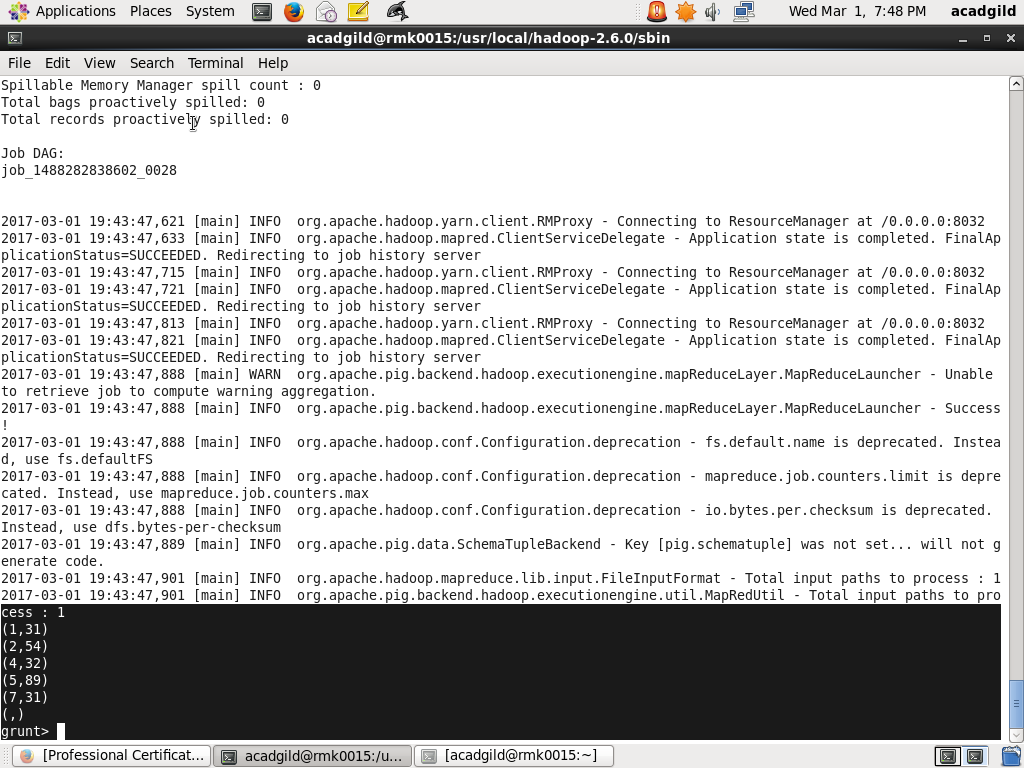


3)SUM:

The sum command in the pig is used to find the total of values of a column.sum function ignores the null values.

ex: s = foreach re1 generate SUM(re.m1);



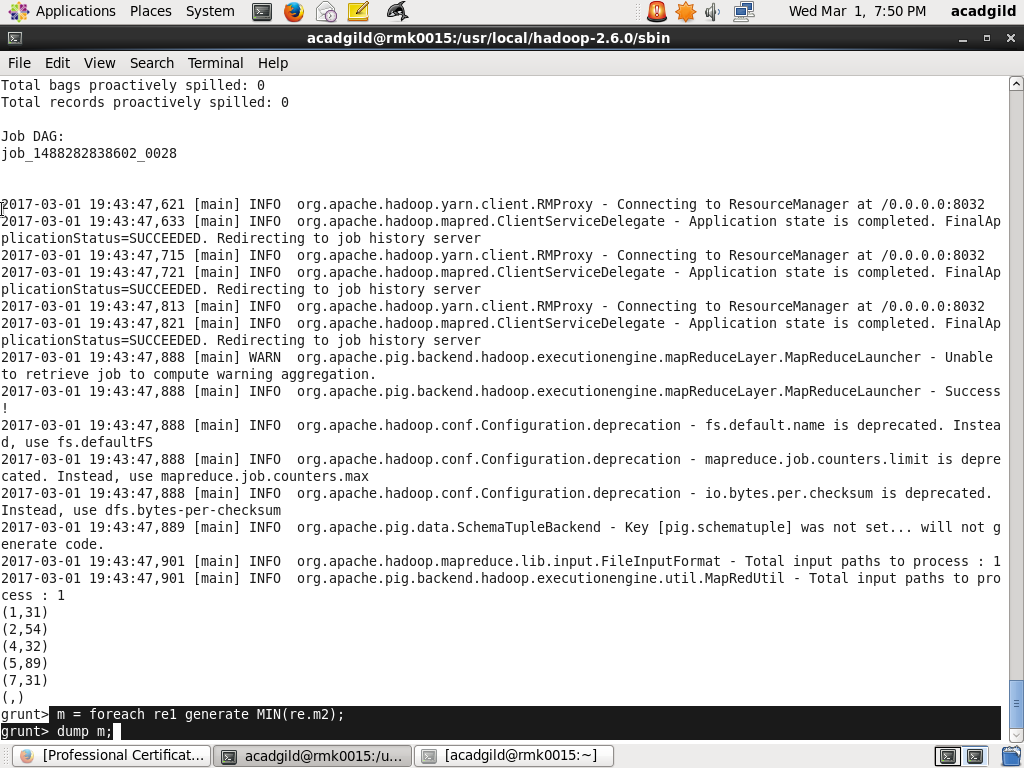


4)MIN:

The min function is used to find the minimum of values in a column.

ex:

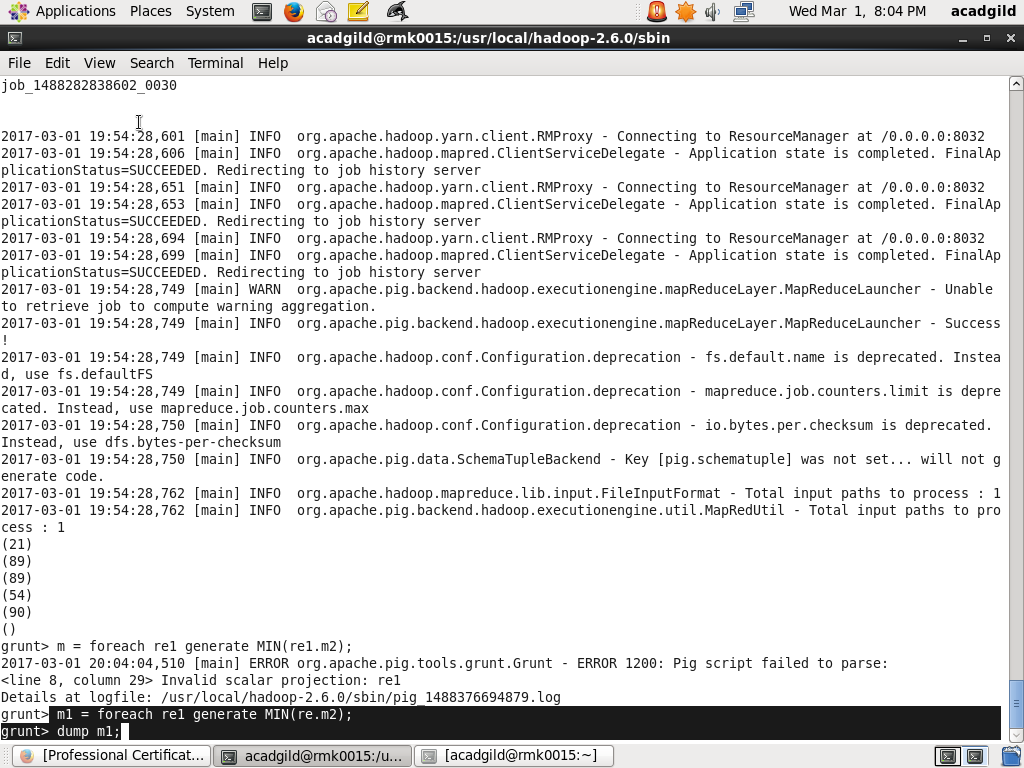
m = foreach re1 generate MIN(re.m2);

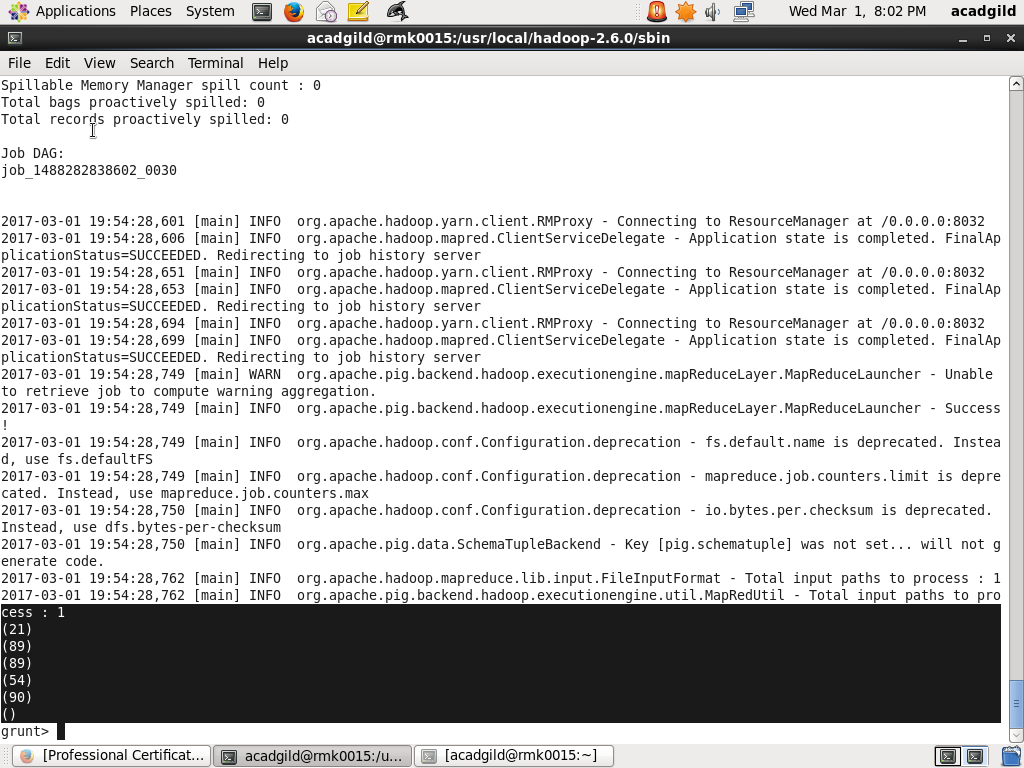


5)MAX:

The max function in pig is used to find the maximum of values in a column.

ex: m 1= foreach re1 generate MIN(re.m2);

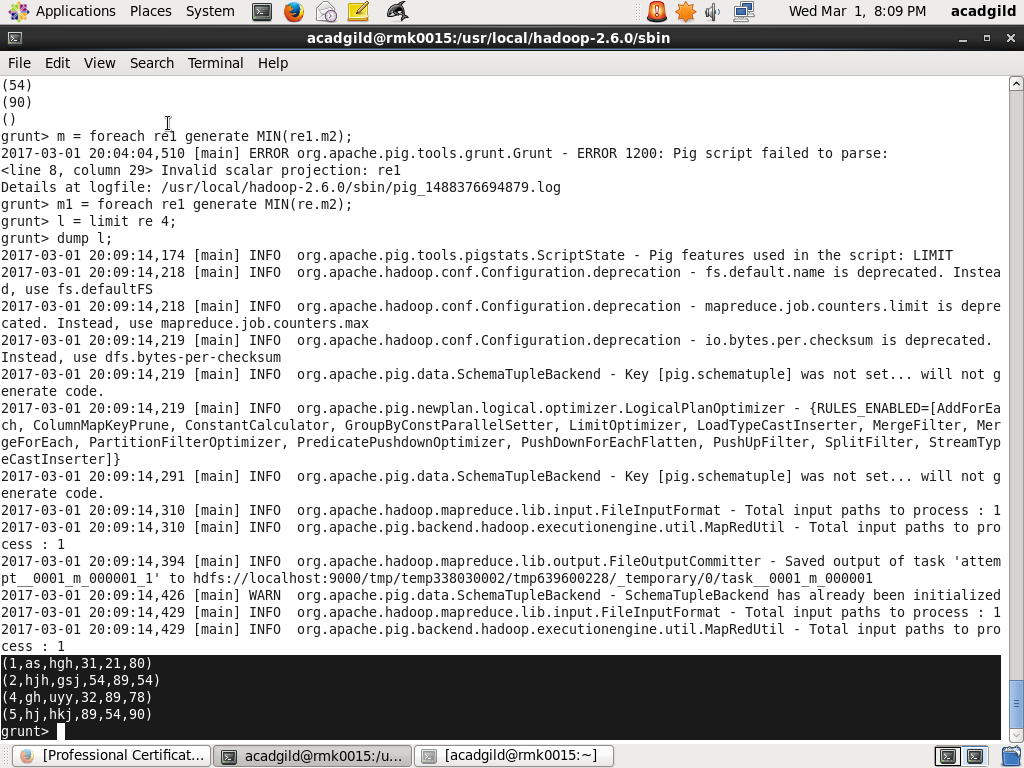




6)LIMIT:

It is used to get the limited number of values from a relation.

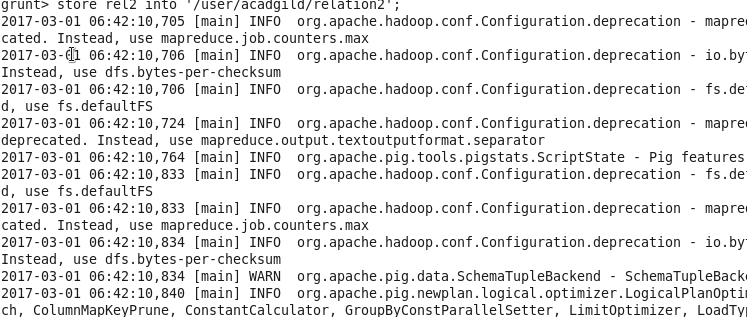
ex:l = LIMIT re 4;



7)STORE:

The store command is used to save the relation in the file system.

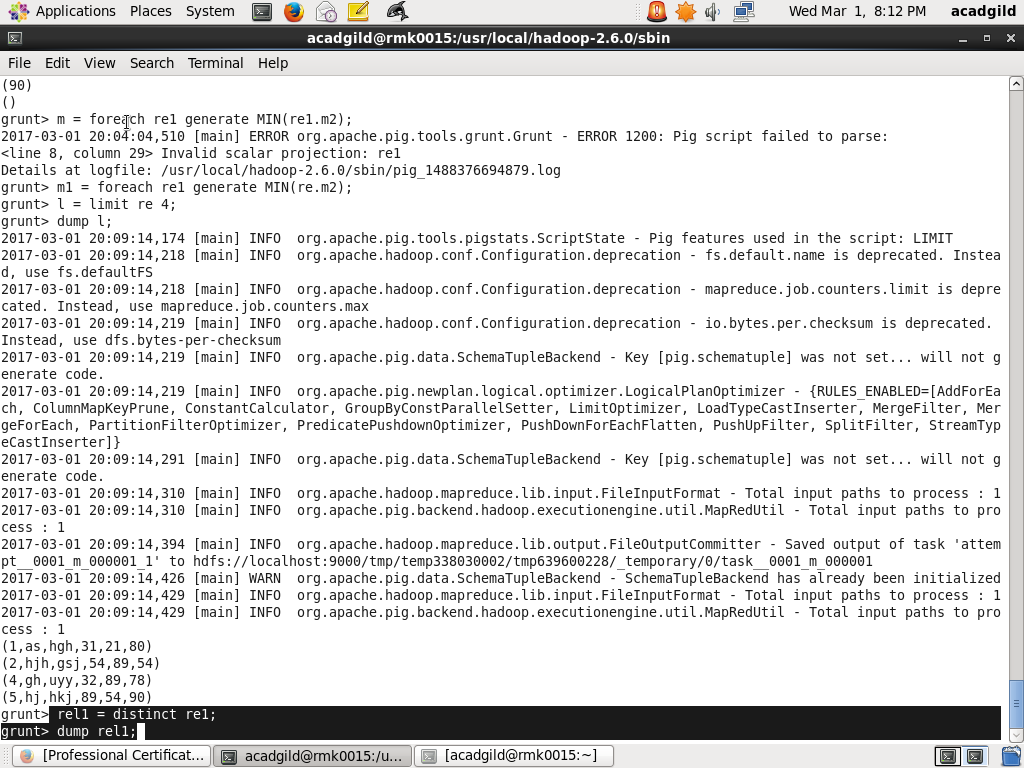
ex: store rel into ‘/user/acadgild/relation2’;

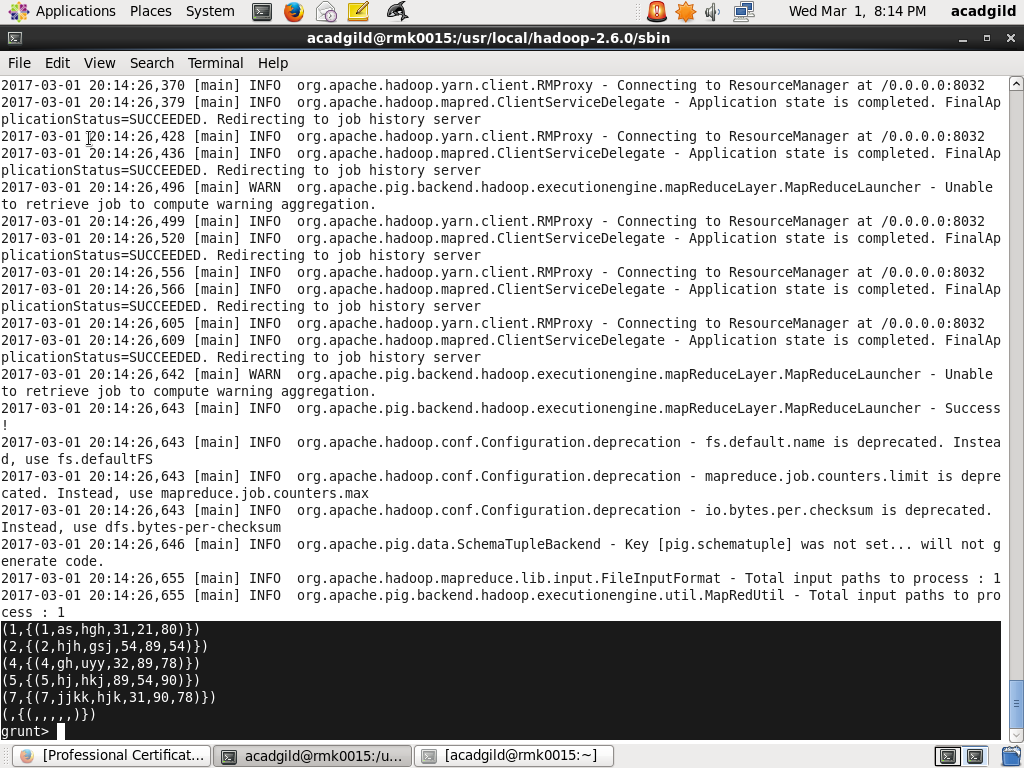


8)DISTINCT:

The Distinct commnad is used to remove the duplicate tuples from a relation.

ex:rel1=DISTINCT re1;





9)FLATTEN:

This command is used when we have a data in a bag or a tuple and we want to remove the level of nesting.

ex:r1= foreach r2 generate FLATTEN(name) as name;

10)ISEMPTY:

It is used to check whether the bag or map is empty.

ex:Y = filter X by IsEmpty(X.name);

