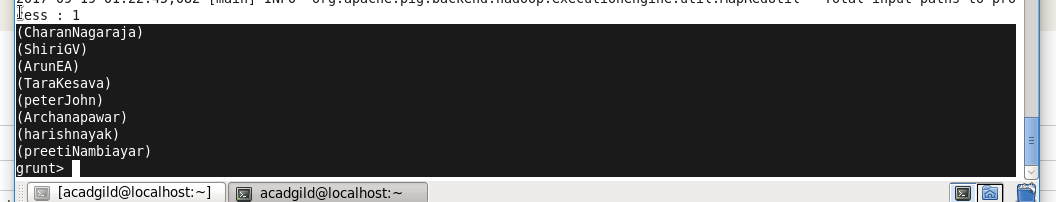
**Session 4: Schedulers in YARN & Introductionto Pig**

**Assignment 2 Question**

**Concat** : The CONCAT() function of Pig Latin is used to concatenate two or more expressions of the same type.

**Command** : concatDemo = FOREACH studentDetails GENERATE CONCAT(firstName,lastName) as FullName;

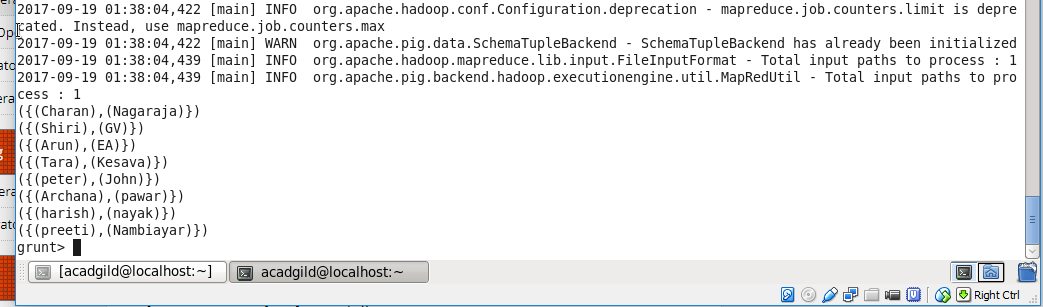
studentDetails = LOAD '/home/acadgild/employee.txt' USING PigStorage(',') as (id : int , firstName : chararray , lastName : chararray , contactNo :bytearray , location : chararray);



**Tokenize** : The TOKENIZE() function of Pig Latin is used to split a string (which contains a group of words) in a single tuple and returns a bag which contains the output of the split operation.

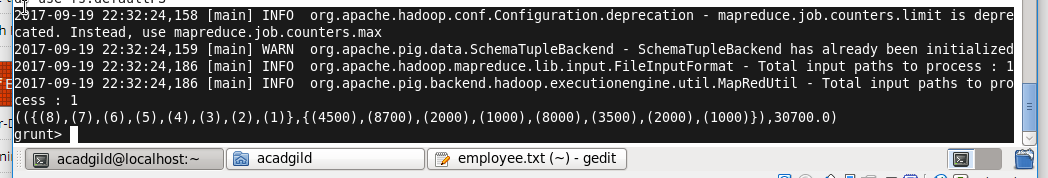
employee = LOAD '/home/acadgild/employee.txt' USING PigStorage(',') as (id : int , firstname : chararray , lastname: chararray, age :bytearray , location : chararray, salary:bytearray);

**Command** : emp\_salary = FOREACH employee GENERATE id , TOKENIZE(name);



**Sum**: SUM is used to get the total of the numeric values of a column in a single-column bag. While computing the total, the SUM() function ignores the NULL values.

**Command** : emp\_salary = FOREACH employee GENERATE (employee.salary), SUM (employee.salary);



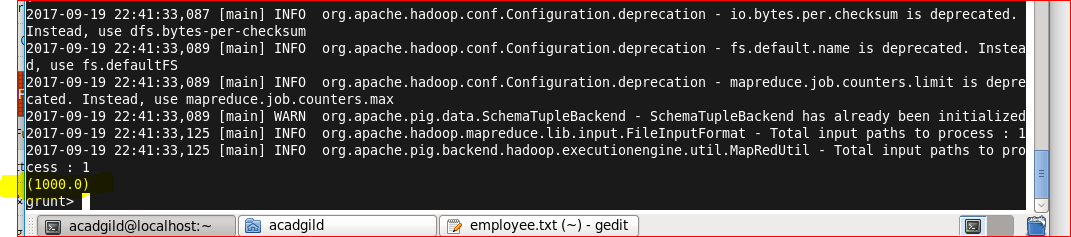
**Min:** The MIN() function of Pig Latin is used to get the minimum (lowest) value (numeric or chararray) for a certain column in a single-column bag. While calculating the minimum value, the MIN() function ignores the NULL values.

**commands**

employee = LOAD '/home/acadgild/employee.txt' USING PigStorage(',') as (id : int , firstname : chararray , lastname: chararray, age :bytearray , location : chararray, salary:bytearray);

emp\_salary = GROUP employee all;

min\_salary = FOREACH emp\_salary GENERATE MIN(employee.salary);

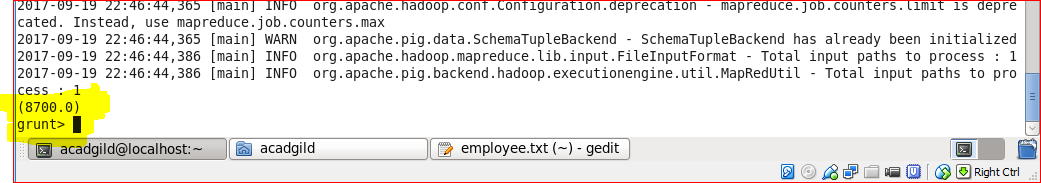


**Max** **:** The MAX() function of Pig Latin is used to get the maximum (highest) value (numeric or chararray) for a certain column in a single-column bag. While calculating the minimum value, the MIN() function ignores the NULL values.

employee = LOAD '/home/acadgild/employee.txt' USING PigStorage(',') as (id : int , firstname : chararray , lastname: chararray, age :bytearray , location : chararray, salary:bytearray);

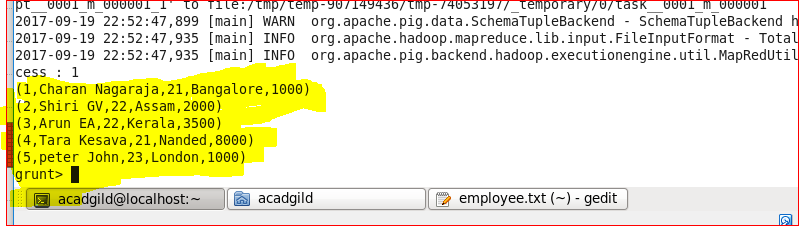
emp\_salary = GROUP employee all;

min\_salary = FOREACH emp\_salary GENERATE MAX(employee.salary);

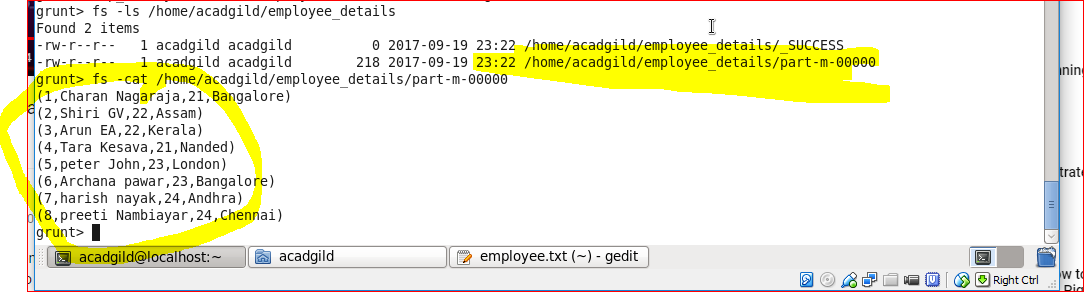


**Limit :** The LIMIT operator is used to get a limited number of tuples from a relation.

**Command** : emp\_salary = LIMIT employee 5;

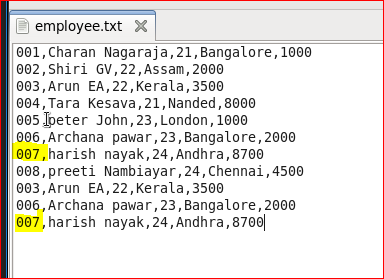


**Store**: It is used to store the loaded data in the file system.

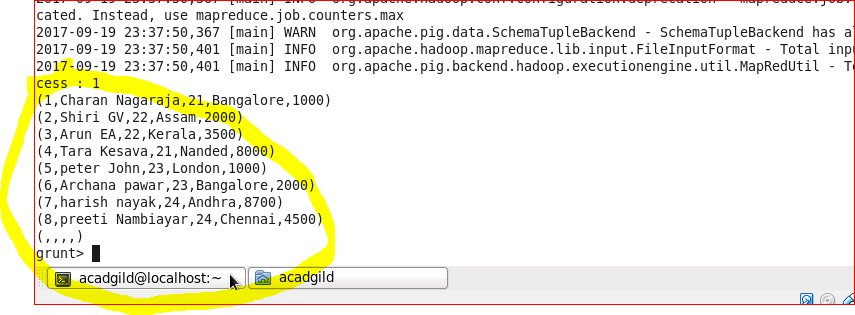


**Distinct** : The DISTINCT operator is used to remove redundant (duplicate) tuples from a relation.

Below Dataset:-



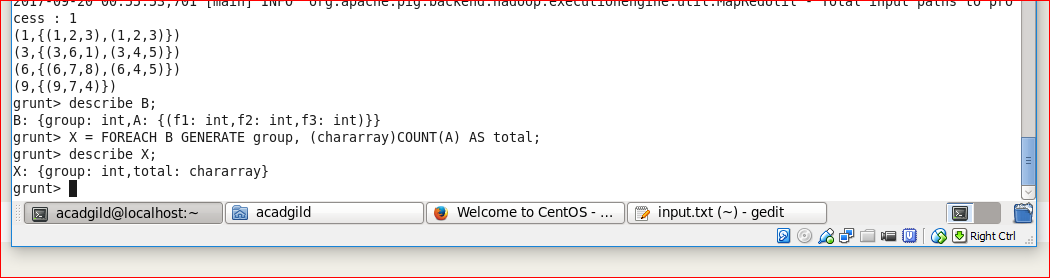
Output:-



**Flatten** : FLATTEN is used to un-nest tuple or bag.

**command**

X = FOREACH B GENERATE group, (chararray)COUNT(A) AS total;



**IsEmpty** : The IsEmpty() function of Pig Latin is used to check if a bag or map is empty.

The emp\_sales relation holds the tuples that are not there in the relation emp\_bonus.

