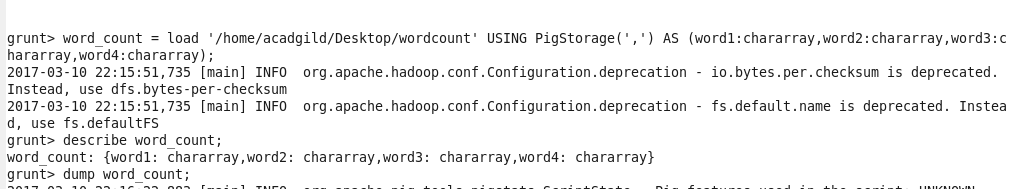
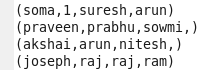
Input FILE AND its Schema



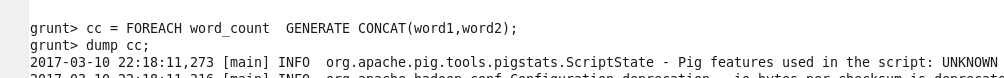


Concat:

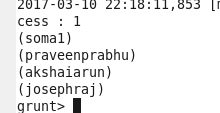
Used to concat 2 expressions

Here I concat word1 and word2 field as shown

Command



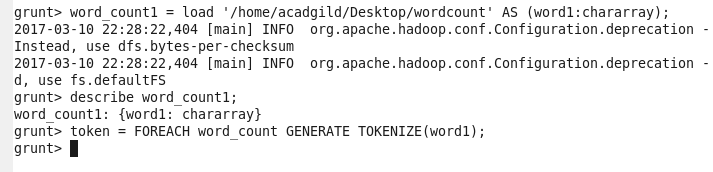
Output



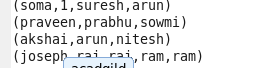
Tokenize

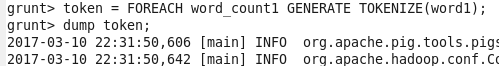
Splits a string and outputs a bag of words

It by default split a string of words (all words in a single tuple) into a bag of words (each word in a single tuple). The following characters are considered to be word separators: space, double quote("), coma(,) parenthesis(()), star(\*).

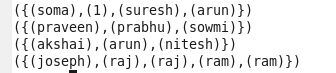


Dump word\_count1;

  
executing TOKENIZE COMMAND



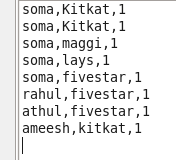
Opt:



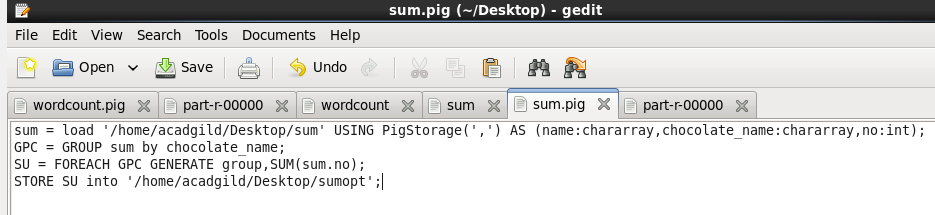
SUM :

Computes the sum of the numeric values in a single-column bag. SUM requires a preceding GROUP ALL statement for global sums and a GROUP BY statement for group sums.

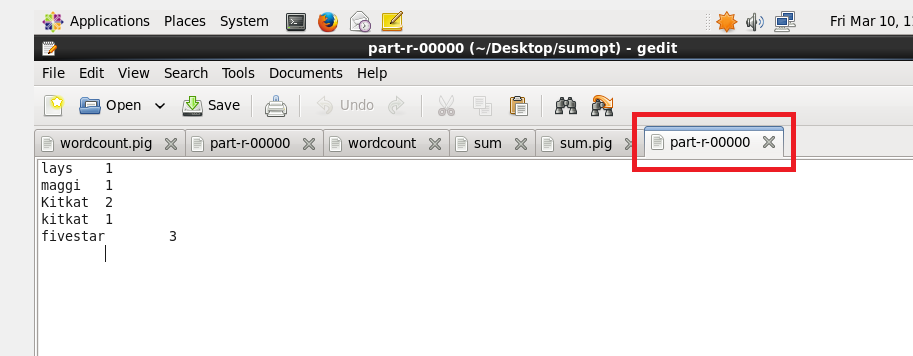
Input show the eatables bought by some friends.suppose I want to find how much of eatable is totally there:



Steps Load the file and group chocolate name and use sum function to find it as shown



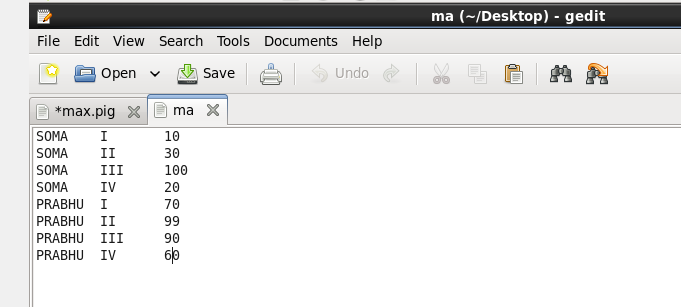
OPT:



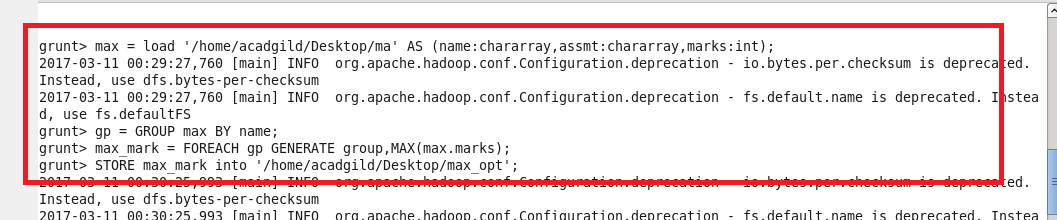
MAX

Computes the maximum of the numeric values or chararrays in a single-column bag. MAX requires a preceding GROUP ALL statement for global maximums and a GROUP BY statement for group maximum.

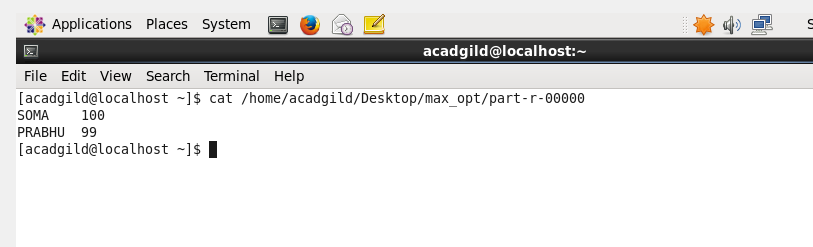
INPUT : marks of 2 students in 4 assessment



I can find the max marks by grouping the student by name and then by using Max function as shown



OPT:



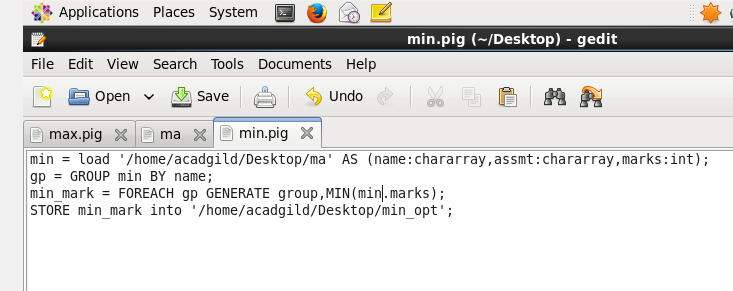
MINIMUM

It is just opposite of max and I want to find the minimum marks of each student

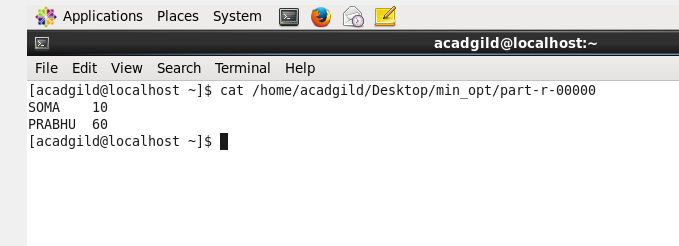
I can find the min marks by grouping the student by name and then by using MIN function as shown



Min.pig



Otpt

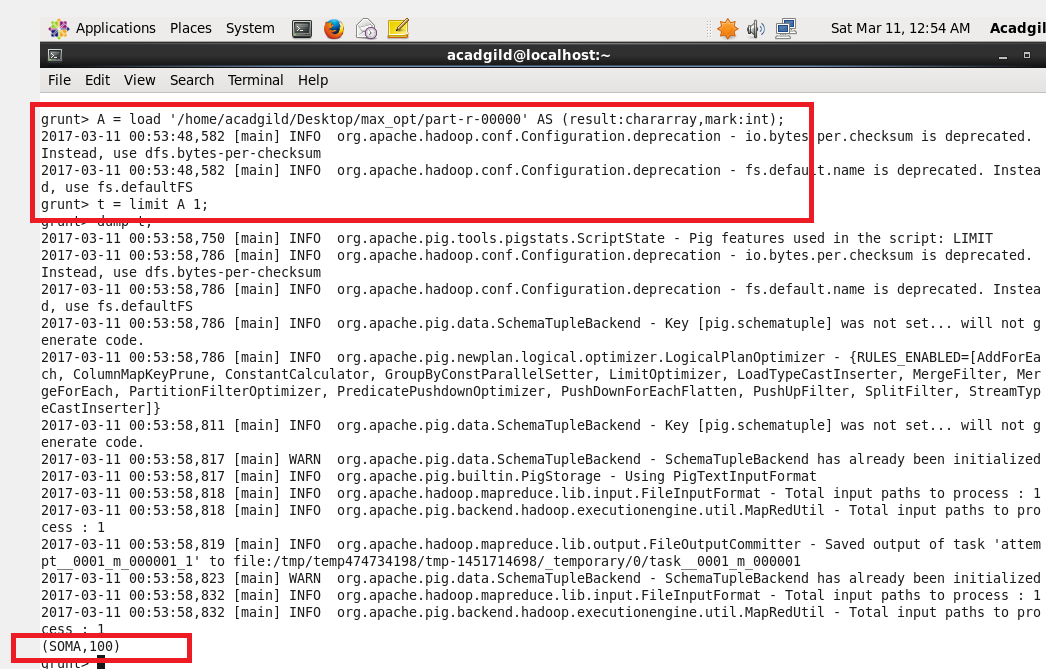


LIMIT

It is used to limit the output

For example I want the person who is the first

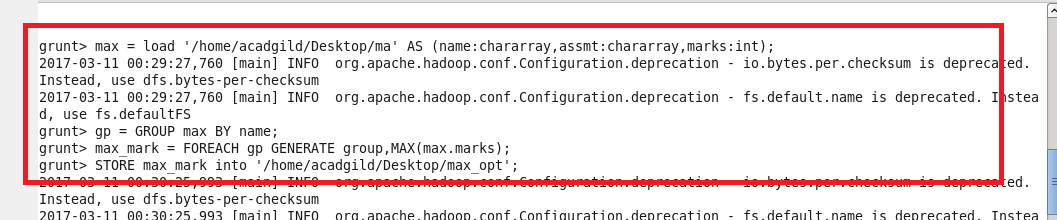
It can be done as shown in image

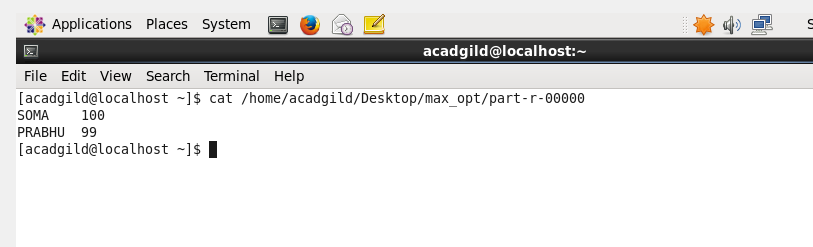


STORE

The problem with grunt shell is that once we come out of grunt shell every executed will be

Gone .So we can use Store command to store output



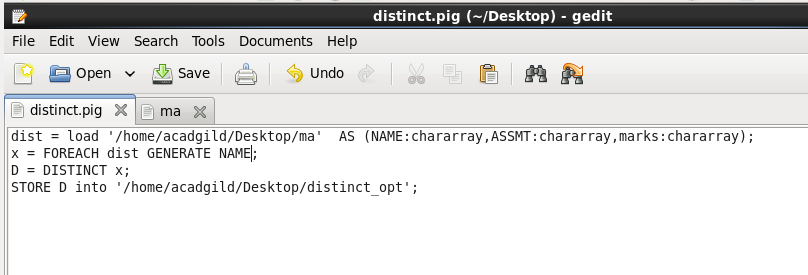


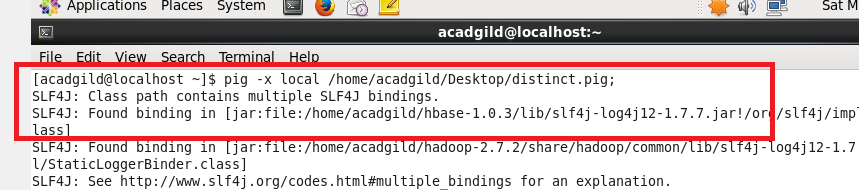
**DISTINCT**

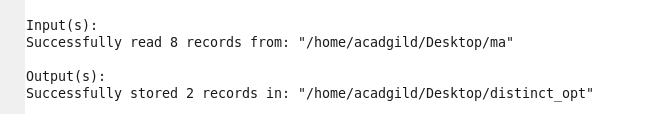
Removes duplicate tuples in a relation.

I just wanted to find the number of unique students

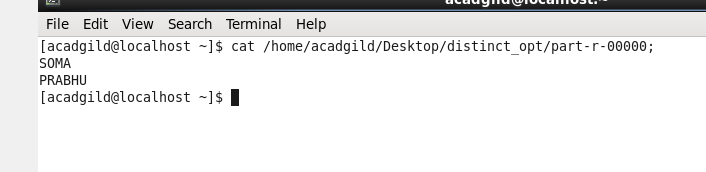
So I created the pig file





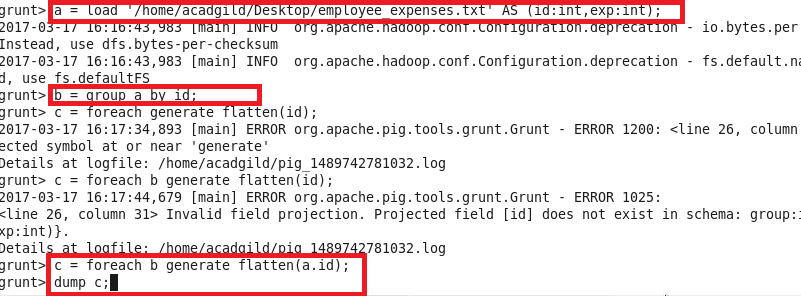


**OUTPUT**

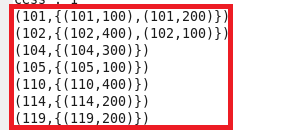
****

**FLATTEN:**

**Used to unnest a bag.HERE FIRST I AM GROUPING EXPENSES BY ID AND THEN USED FLATTEN COMMAND TO UNNEST ID FROM THE BAG AS SHOWN**



Dump b;



DUMP C;

