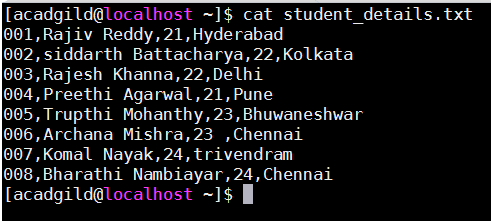
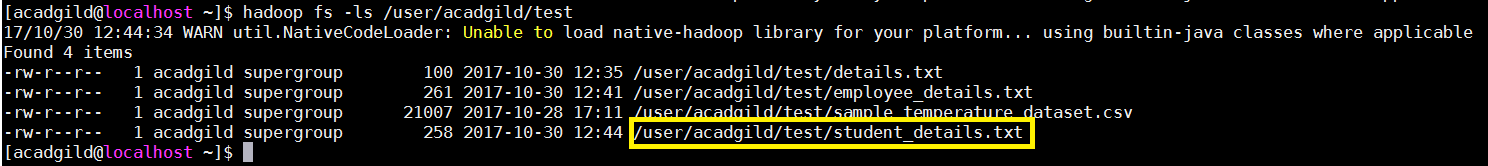
**Assignment 4.2**

Created a file student\_details.txt and uploaded the in the folder /user/acadgild/test



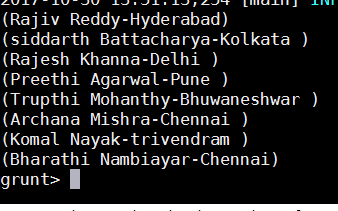


1. **Concat():**

Command :



Performed **Dump student\_name\_contact;** and got the following output.

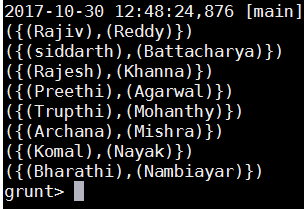


1. **TokenIze():**

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

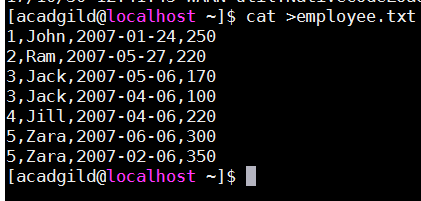


Performed **Dump student\_name\_tokenize** and got the following output

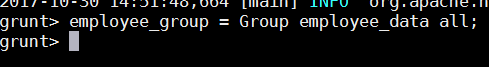


1. **Sum():**

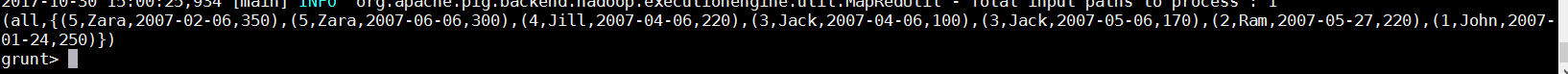
Created the employee.txt file and uploaded in /user/acadgild/test



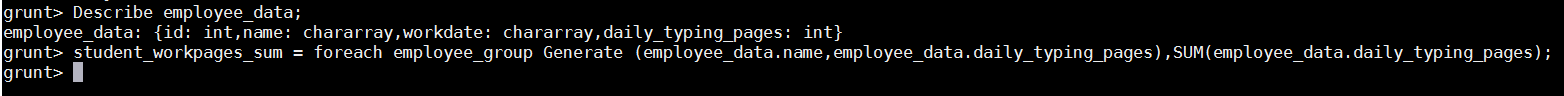
To demonstrate the sum() function, let us calculate the total number of pages typed daily of all the employee’s. First let us group the relation.



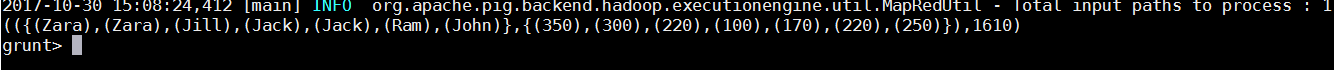
**Output:**



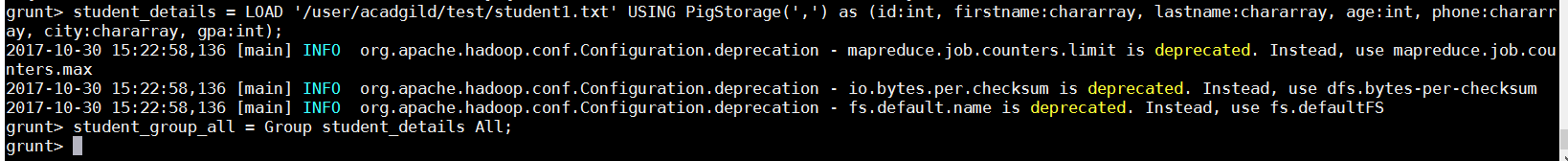
Now, we will run the sum function for calculating the number of pages typed by employee’s



**Output:**



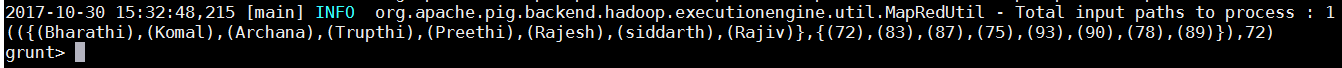
1. **Min();**



We grouped the relation and applied the MIN() function.



**Output:**



As you can see, the out is 72 which is the minimum value of all.

1. **Max();**

Taking the same dataset as in MIN() function, applied the MAX() function



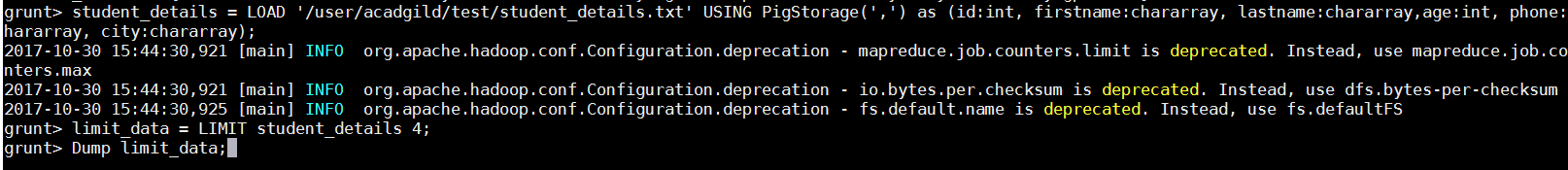
**Output:**



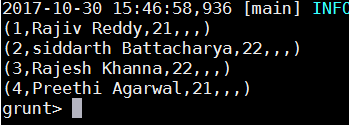
As you can see, the out is 93 which is the maximum value of all.

1. **Limit():**

Using this function we limit the number of output we want to see.



**Output:**



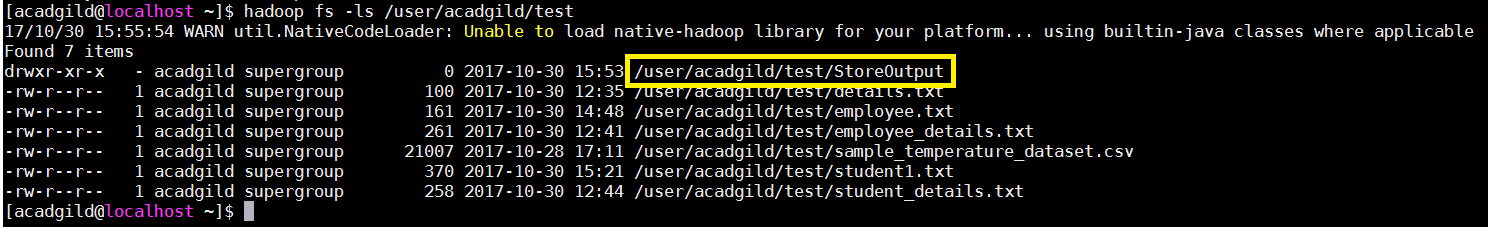
1. **Store:**

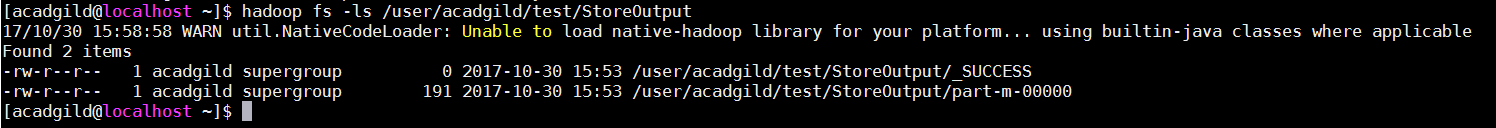
Using this command we will learn to store data in PIG.

At first we will the LOAD the file from given input and STORE the data any location we want to.

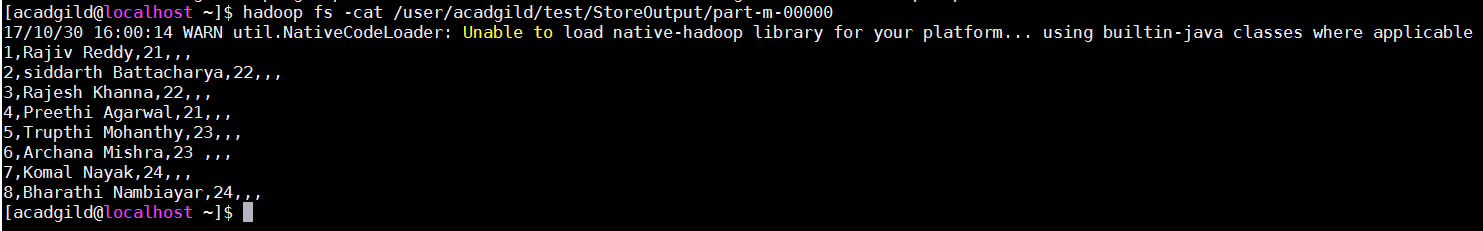


Now, we can check that the file has been stored in /user/acadgild/StoreOutput (as per the below screen shot)



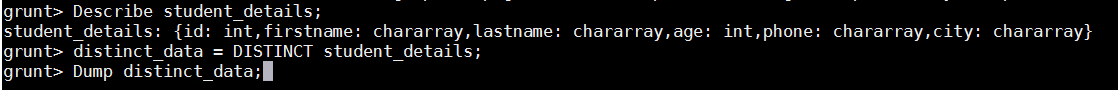


We see that we have the details in **/user/acadgild/test/Storeoutput/part-m-00000**

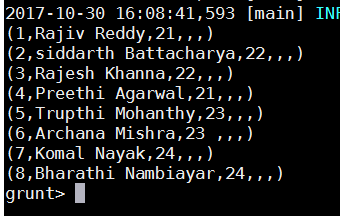


1. **Distinct();**

Using this command we can remove the duplicate tuples.

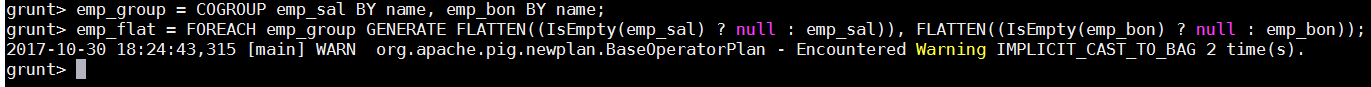


**Output:**

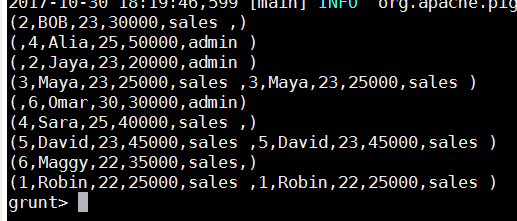


1. **Flatten():**

The FLATTEN operator looks like a UDF syntactically, but it is actually an operator that changes the structure of tuples and bags in a way that a UDF cannot. Flatten un-nests tuples as well as bags. The idea is the same, but the operation and result is different for each type of structure.

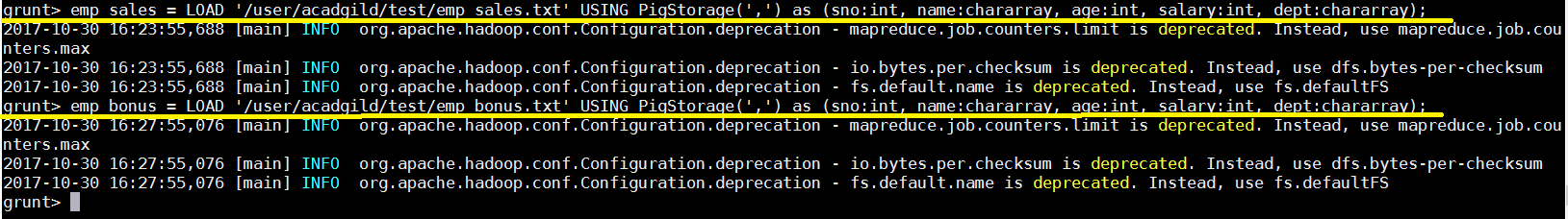


**Output:**

****

1. **IsEmpty():**

Created 2 files emp\_sales.txt and emp\_bonus.txt and LOAD the files.



Let’s list such empty bags from the **emp\_sales** relation in the group using the **IsEmpty()** function.

