**Assignment 4\_2**

**Create a sample dataset and implement the below Pig commands on the same dataset.**

**1) Concat**

**2) Tokenize**

**3) Sum**

**4) Min**

**5) Max**

**6) Limit**

**7) Store**

**8) Distinct**

**9) Flatten**

**10) IsEmpty**

**Input File :**

**StudentId , StudentFname , StudentLname,Age,PhoneNum,City,Marks**

001,Rajiv,Reddy,21,9848022337,Hyderabad,89

002,siddarth,Battacharya,22,9848022338,Kolkata,78

003,Rajesh,Khanna,22,9848022339,Delhi,90

004,Preethi,Agarwal,21,9848022330,Pune,93

005,Trupthi,Mohanthy,23,9848022336,Bhuwaneshwar,75

006,Archana,Mishra,23,9848022335,Chennai,81

007,Komal,Nayak,24,9848022334,trivendram,83

008,Bharathi,Nambiayar,24,9848022333,Chennai,72

009,Rajiv,Reddy2,21,9848022337,Hyderabad,82

010,siddarth,Battacharya2,22,9848022338,Kolkata,78

011,Rajesh,Khanna2,22,9848022339,Delhi,90

012,Preethi,Agarwal2,21,9848022330,Pune,94

013,Trupthi,Mohanthy2,23,9848022336,Bhuwaneshwar,75

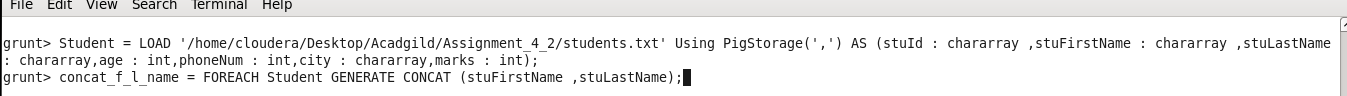
014,Archana,Mishra2,23,9848022335,Chennai,81

015,Komal,Nayak2,24,9848022334,trivendram,81

016,Bharathi,Nambiayar2,24,9848022333,Chennai,70

**1)Concat :** Use the CONCAT function to concatenate two expressions. The result values of the two expressions must have identical types.If either subexpression is null, the resulting expression is null.

**Concating First Name and Last Name :**

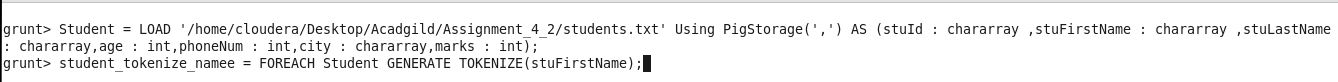


DUMP concat\_f\_l **;**

**OUTPUT**



**2) 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.

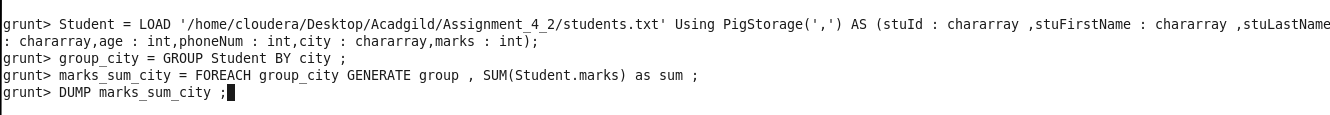


DUMP student\_tokenize\_namee ;



**3)SUM :** You can use the **SUM()** function of Pig Latin 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.

**Sum of Marks Per city**

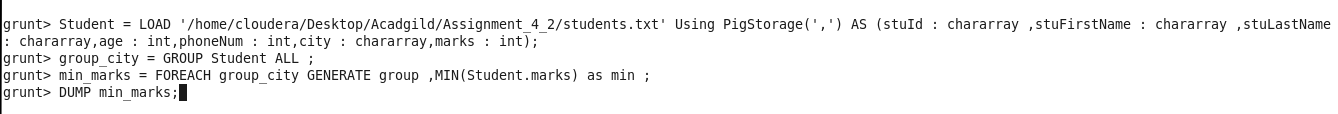


**Output**



**4) 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.

**Minimum Marks in the Students List**

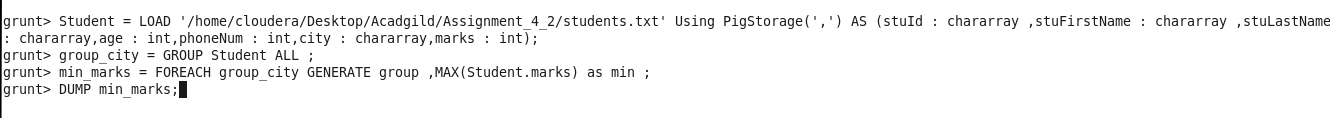


**Output**



**5) MAX :** The Pig Latin **MAX()** function is used to calculate the highest value for a column (numeric values or chararrays) in a single-column bag. While calculating the maximum value, the **Max()** function ignores the NULL values.

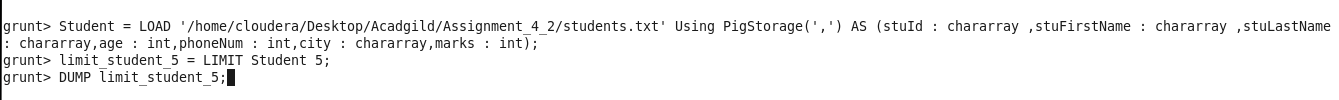
**Maximum Marks in the Students List**



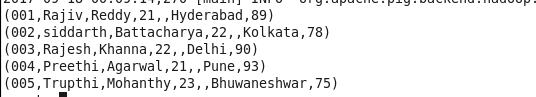
**Output**



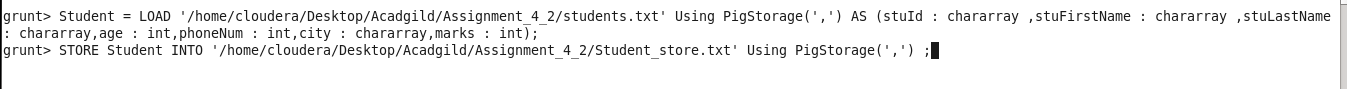
**6)LIMIT :** The **LIMIT** operator is used to get a limited number of tuples from a relation.



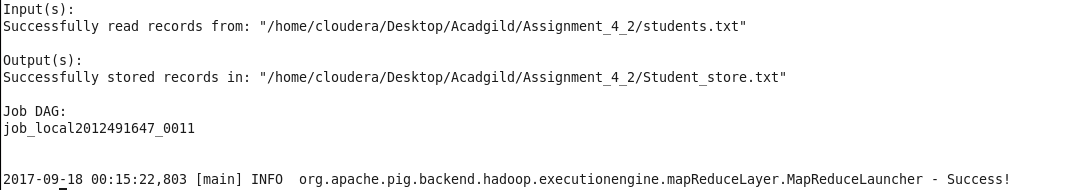
**Output**

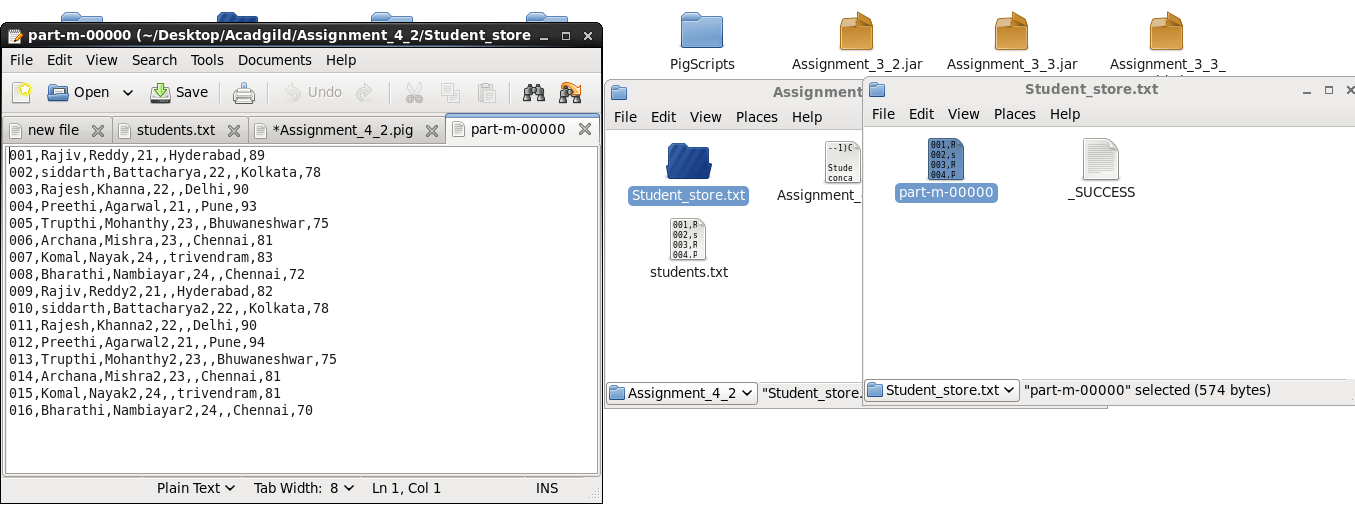


**7) STORE :** You can store the loaded data in the file system using the **store** operator.

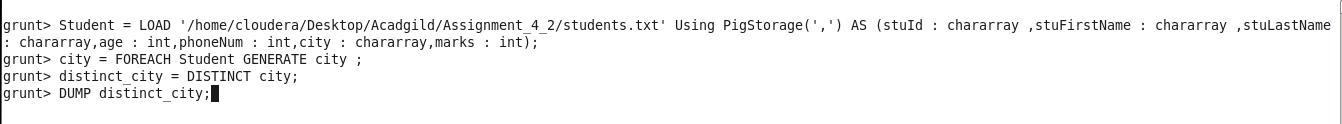


**Output**





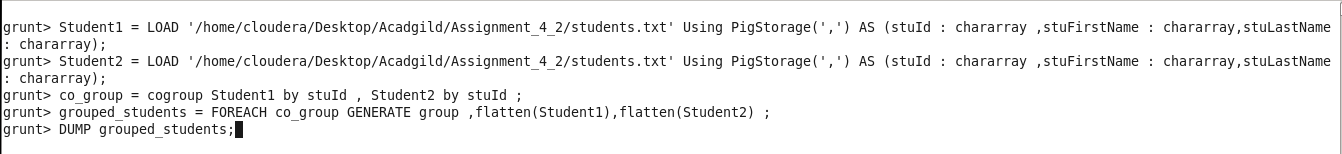
8)**Distinct :**The **DISTINCT** operator is used to remove redundant (duplicate) tuples from a relation.



**OUTPUT**

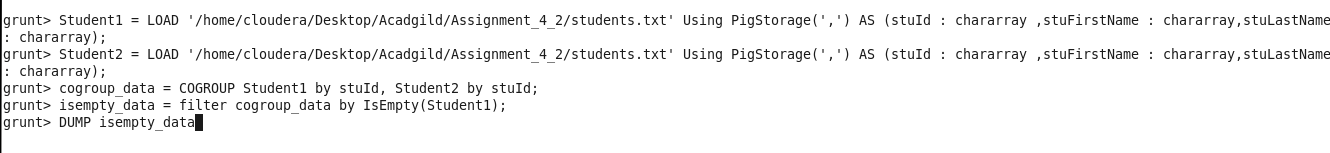


**9) Flatten :** The FLATTEN operator looks like a UDF syntactically, but it is actually an operator that changes the structure of tuples and bags in a way that a UDF cannot. Flatten un-nests tuples as well as bags. The idea is the same, but the operation and result is different for each type of structure.For tuples, flatten substitutes the fields of a tuple in place of the tuple. For example, consider a relation that has a tuple of the form (a, (b, c)). The expression GENERATE $0, flatten($1), will cause that tuple to become (a, b, c).



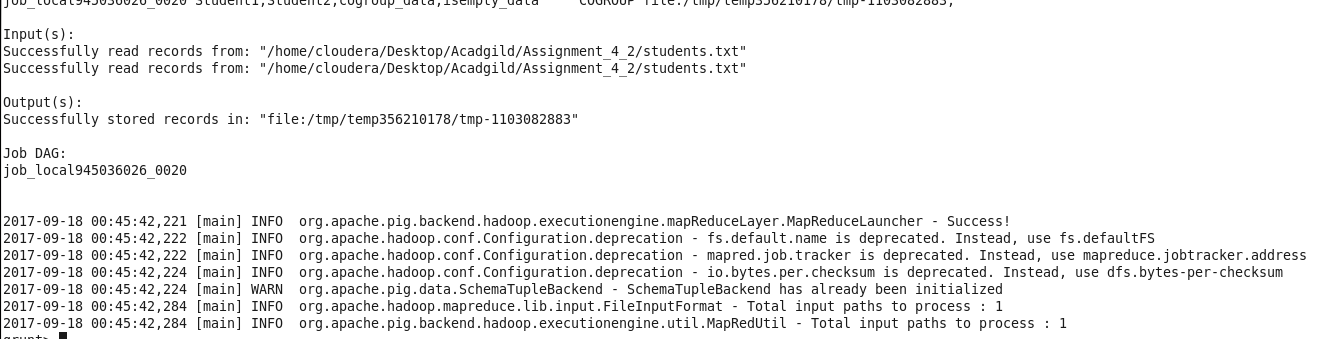


**10) IsEmpty :** The **IsEmpty()** function of Pig Latin is used to check if a bag or map is empty.



**OUTPUT**

**NO data is missing so there will be now rows.**



**GITHUB Repository will contain Document and Pig Scipts used in this document**