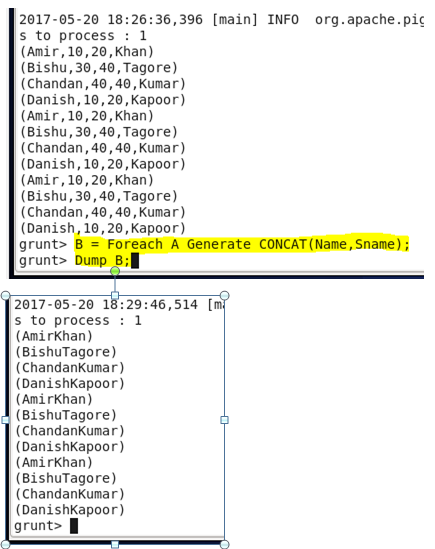
1. Concat

Used for concatenating two strings

A = Load 'Aggregation.txt' As (Name:chararray, Salary:int, temp:int, Sname:chararray);

B = Foreach A Generate CONCAT(Name,Sname);

Dump B;



1. Tokenize

Used for splitting the lines in tokens, output is a bag of words present in the sentence or input value.

A = Load 'helloworld' As (line:chararray);

B = Foreach A Generate TOKENIZE(line,' ') As words;

Dump B;

C:\Users\Shabbir\Pictures\TOKENIZE.PNG

1. Sum

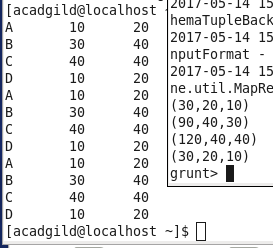
It is an aggregate function, used for Computing the sum of the numeric values in a single-column bag. SUM requires a preceding GROUP ALL statement for global sums and a GROUP BY statement for group sums.

A = Load 'Aggregation.txt' As (Name:chararray, Salary:int, temp:int);

B = Group A By Name;

C = Foreach B Generate SUM(A.Salary),MIN(A.temp), MAX(A.Salary);

Dump C;



1. Min

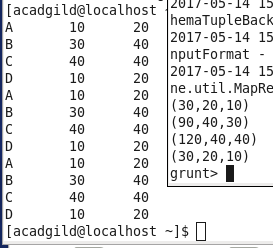
Computes the minimum of the numeric values or chararrays in a single-column bag. MIN requires a preceding GROUP… ALL statement for global minimums and a GROUP … BY statement for group minimums.

A = Load 'Aggregation.txt' As (Name:chararray, Salary:int, temp:int);

B = Group A By Name;

C = Foreach B Generate SUM(A.Salary),MIN(A.temp), MAX(A.Salary);

Dump C;



1. Max

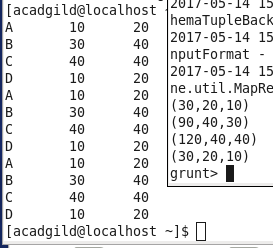
Computes the maximum of the numeric values or chararrays in a single-column bag. MAX requires a preceding GROUP ALL statement for global maximums and a GROUP BY statement for group maximums.

A = Load 'Aggregation.txt' As (Name:chararray, Salary:int, temp:int);

B = Group A By Name;

C = Foreach B Generate SUM(A.Salary),MIN(A.temp), MAX(A.Salary);

Dump C;



1. Limit

It displays limited no of column from the produced result

A = Load 'Aggregation.txt' As (Name:chararray, Salary:int, temp:int);

B = Group A By Name;

C = Foreach B Generate SUM(A.Salary),MIN(A.temp), MAX(A.Salary);

Dump C;

Store C into 'AF.txt';

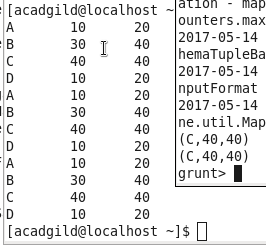
D = DISTINCT A;

Dump D;

E = ORDER A BY Salary DESC;

F = LIMIT E 2;

Dump F;



1. Store

It is used to Store the result/relation in a particular directory

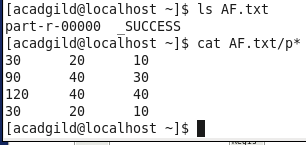
A = Load 'Aggregation.txt' As (Name:chararray, Salary:int, temp:int);

B = Group A By Name;

C = Foreach B Generate SUM(A.Salary),MIN(A.temp), MAX(A.Salary);

Dump C;

Store C into 'AF.txt';



1. Distinct

Used for eliminating duplicate values from a relation

A = Load 'Aggregation.txt' As (Name:chararray, Salary:int, temp:int);

B = Group A By Name;

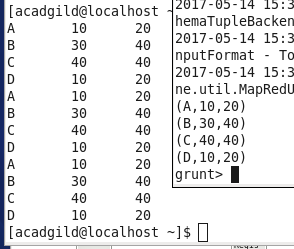
C = Foreach B Generate SUM(A.Salary),MIN(A.temp), MAX(A.Salary);

Dump C;

Store C into 'AF.txt';

D = DISTINCT A;

Dump D;



1. Flatten

Used on top of TOKENIZE, produces tuple of each word in a new line.

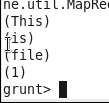
A = Load 'helloworld' As (line:chararray);

B = Foreach A Generate TOKENIZE(line,' ') As words;

Dump B;

C = Foreach B Generate flatten(words);

Dump C;



1. IsEmpty

Used for checking the bag or map is empty.

A = load 'ssn.txt' using PigStorage() as (ssn:long);

B = load 'students.txt' using PigStorage() as (ssn:long, name:chararray);

X = JOIN A by ssn LEFT OUTER, B by ssn;

Y = filter X by IsEmpty(B);