**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.

**Note:**

 To get the global sum value, we need to perform a **Group All** operation, and calculate the average value using the AVG function.

 To get the sum value of a group, we need to group it using the **Group By** operator and proceed with the average function.

**Syntax**

Given below is the syntax of the **sum()** function.

SUM(expression)

**Example**

Assume that we have a file named **employee.txt** in the HDFS directory **/pig\_data/** as shown below.

**employee.txt**

1,John,2007-01-24,250

2,Ram,2007-05-27,220

3,Jack,2007-05-06,170

3,Jack,2007-04-06,100

4,Jill,2007-04-06,220

5,Zara,2007-06-06,300

5,Zara,2007-02-06,350

And we have loaded this file into Pig with the schema name **employee\_data** as shown below.

employee\_data = LOAD 'hdfs://localhost:9000/pig\_data/ employee.txt' USING PigStorage(',')as (id:int, name:chararray, workdate:chararray, daily\_typing\_pages:int);

**Calculating the Sum of All GPA**

To demonstrate the **SUM()** function, let’s try to calculate the total number of pages typed daily of all the employees. We can use the Apache Pig’s built-in function **SUM()** (case sensitive) to calculate the sum of the numerical values. Let us group the schema **employee\_data** using the **Group All** operator, and store the result in the schema named **employee\_group** as shown below.

employee\_group = Group employee\_data all;

It will produce a schema as shown below.

Dump employee\_group;

(all,{(5,Zara,2007-02-06,350),(5,Zara,2007-06-06,300),(4,Jill,2007-04-

06,220),(3,Jack,2007-04-06,100),(3,Jack,2007-05-06,170),(2,Ram,2007-05-

27,220),(1,John,2007-01-24,250)})

Let us now calculate the global sum of the pages typed daily.

student\_workpages\_sum = foreach employee\_group Generate ([employee\_data.name,](http://employee_data.name/)employee\_data.daily\_typing\_pages),SUM(employee\_data.daily\_t yping\_pages);

**Verification**

Verify the relation **student\_workpages\_sum** using the **DUMP** operator as shown below.

Dump student\_workpages\_sum;

**Output**

It will produce the following output, displaying the contents of the relation

**student\_workpages\_sum** as follows.

(({ (Zara), (Zara), (Jill) ,(Jack) , (Jack) , (Ram) , (John) },

{ (350) , (300) , (220