**Calculating the Number of Tuples**

We can use the built-in function **COUNT()** (case sensitive) to calculate the number of tuples in a relation. Let us group the schema **student\_details** using the **Group All** operator, and store the result in the schema named **student\_group\_all** as shown below.

student\_group\_all = Group student\_details All;

It will produce a schema as shown below.

Dump student\_group\_all;

(all,{(8,Bharathi,Nambiayar,24,9848022333,Chennai,72),(7,Komal,Nayak,24,9848022

334,trivendram,83),(6,Archana,Mishra,23,9848022335,Chennai,87),(5,Trupthi,Mohan thy,23,9848022336,Bhuwaneshwar,75),(4,Preethi,Agarwal,21,9848022330,Pune,93),(3

,Rajesh,Khanna,22,9848022339,Delhi,90),(2,siddarth,Battacharya,22,9848022338,Ko

lkata,78),(1,Rajiv,Reddy,21,9848022337,Hyderabad,89)})

Let us now calculate number of tuples/records in the relation.

student\_count = foreach student\_group\_all Generate

COUNT(student\_details.gpa);

**Verification**

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

Dump student\_count;

**Output**

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

**student\_count**.

8

**COUNT\_STAR**

The **COUNT\_STAR()** function of Pig Latin is similar to the **COUNT()** function. It is used to get the number of elements in a bag. While counting the elements, the COUNT\_STAR() function includes the NULL values.

**Note:**

 To get the global count value (total number of tuples in a bag), we need to perform a **Group All** operation, and calculate the average value using the AVG function.

 To get the count value of a group (Number of tuples in 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 COUNT\_STAR function.

COUNT\_STAR(expression)

**Example**

Assume that we have a file named **student\_details.txt** in the HDFS directory

**/pig\_data/** as shown below. This file contains an empty record.

**student\_details.txt**

, , , , , , ,

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,87

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

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

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

student\_details = LOAD 'hdfs://localhost:9000/pig\_data/student\_data.txt' USING PigStorage(',')as (id:int, firstname:chararray, lastname:chararray, age:int, phone:chararray, city:chararray, gpa:int);

**Calculating the Number of Tuples**

We can use the built-in function **COUNT\_STAR()** to calculate the number of tuples in a relation. Let us group the schema **student\_details** using the **Group All** operator, and store the result in the schema named **student\_group\_all** as shown below.

**student\_group\_all** = Group **student\_details** All;

It will produce a schema as shown below.

Dump student\_group\_all;

(all,{(8,Bharathi,Nambiayar,24,9848022333,Chennai,72),(7,Komal,Nayak,24,9848022

334,trivendram,83),(6,Archana,Mishra,23,9848022335,Chennai,87),(5,Trupthi,Mohan thy,23,9848022336,Bhuwaneshwar,75),(4,Preethi,Agarwal,21,9848022330,Pune,93),(3

,Rajesh,Khanna,22,9848022339,Delhi,90),(2,siddarth,Battacharya,22,9848022338,Ko

lkata,78),(1,Rajiv,Reddy,21,9848022337,Hyderabad,89),( , , , , , ,)})

Let us now calculate the number of tuples/records in the relation.

student\_count = foreach student\_group\_all Generate

COUNT\_STAR(student\_details.gpa);

**Verification**

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

Dump student\_count;

**Output**

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

**student\_count**.

9

Since we have used the function COUNT\_STAR, it included the null tuple and returned 9.