**Calculating the Average GPA**

We can use the built-in function **AVG** (case-sensitive) to calculate the average of a set of numerical values. Let’s 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;

This 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 global average GPA of all the students using the **AVG** function as shown below.

**student\_gpa\_avg** = foreach **student\_group\_all** Generate

(**student\_details**.firstname, **student\_details**.gpa), AVG(**student\_details**.gpa);

**Verification**

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

Dump student\_gpa\_avg;

**Output**

It will display the contents of the relation **student\_gpa\_avg** as follows.

(({(Bharathi),(Komal),(Archana),(Trupthi),(Preethi),(Rajesh),(siddarth),(Rajiv)

},

{ (72) , (83) , (87) , (75) , (93) , (90) , (78) , (89) }),83.375)

**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.

**Note:**

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

 To get the maximum 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 **Max()** function.

Max(expression)

**Example**

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

**/pig\_data/** as shown below.

**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 Maximum GPA**

We can use the built-in function **MAX** (case-sensitive) to calculate the maximum value from a set of given numerical values. 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;

This 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 global maximum of GPA, i.e., maximum among the GPA values of all the students using the **MAX** function as shown below.

**student\_gpa\_max** = foreach **student\_group\_all** Generate

(**student\_details**.firstname, **student\_details**.gpa), MAX(**student\_details**.gpa);

**Verification**

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

Dump student\_gpa\_max;

**Output**

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

**student\_gpa\_max**.

(({(Bharathi),(Komal),(Archana),(Trupthi),(Preethi),(Rajesh),(siddarth),(Rajiv)

} ,

{ (72) , (83) , (87) , (75) , (93) , (90) , (78) , (89) }) ,93)

**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.

**Note:**

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

 To get the minimum 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 **Min()** function.

MIN(expression)

**Example**

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

**/pig\_data/** as shown below.

**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 named **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 Minimum GPA**

We can use the built-in function **MIN()** (case sensitive) to calculate the minimum value from a set of given numerical values. 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 global minimum of GPA, i.e., minimum among the GPA values of all the students using the **MIN** function as shown below.

student\_gpa\_min = foreach student\_group\_all Generate

(student\_details.firstname, student\_details.gpa), MIN(student\_details.gpa);

**Verification**

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

Dump student\_gpa\_min;

**Output**

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

**student\_gpa\_min**.

(({(Bharathi),(Komal),(Archana),(Trupthi),(Preethi),(Rajesh),(siddarth),(Rajiv)

} ,

{ (72) , (83) , (87) , (75) , (93) , (90) , (78) , (89) }) ,**72**)