Academic year 2019-2020

Department of Computer Science and Engineering

**KARNATAKA LAW SOCIETY’S**

**GOGTE INSTITUTE OF TECHNOLOGY**

UDYAMBAG, BELAGAVI-590008



Course Project Report

**“Execution of Apache Pig MAX Function“**

Sem: 7

1. Abhishek Tadkod 2GI17CS002
2. Akshay Raichur 2GI17CS012
3. Chetana Bhat 2GI17CS031
4. Jayanth Apagundi 2GI17CS182

Guide

Prof. Arati Shapurkar

Gogte Institute of Technology

Belagavi.

Contents

Definition3

Syntax3

Example3

Verification 5

Program5

Output6

Definition:

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 maximum value using the MAX() function.
* To get the maximum value of a group, we need to group it using the **Group By** operator and proceed with the maximum function.

Syntax:

grunt> 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 the file is loaded into Pig with the relation name **student\_details** as shown below.

grunt> student\_details = LOAD 'pig\_data/student\_details.txt' USING PigStorage(',')

as (id:int, firstname:chararray, lastname:chararray, age:int, phone:chararray, city:chararray, gp

## Calculating the Maximum GPA:

We can use the built-in function **MAX()** to calculate the maximum value from a set of given numerical values. Let us group the relation **student\_details** using the **Group All** operator, and store the result in the relation named **student\_group\_all** as shown below.

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

This will produce a relation as shown below.

**grunt> 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)})

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

grunt> 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.

grunt > Dump student\_gpa\_max;

### Output:

It will produce the as shown, 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)

Program:

student\_details = LOAD '/root/Desktop/BIGDATA/students' USING PigStorage(',')as (id:int, firstname:chararray, lastname:chararray, age:int, phone:chararray, city:chararray, gpa:int);

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

student\_gpa\_max = foreach student\_group\_all Generate (student\_details.firstname, student\_details.gpa), MAX(student\_details.gpa);

Dump student\_gpa\_max;

Output:

