List of Apache Pig Built in Functions

Let’s discuss various Apache Pig Built in Functions namely eval, load, store, math, string, bag, and tuple, one by one in depth.

a. Eval Functions

Eval Functions is the first types of Pig Built in Functions. Here are the Pig Eval functions, offered by Apache Pig.  
**i. AVG()**

* **AVG Syntax**

AVG(expression)  
We use AVG(), to compute the average of the numerical values within a bag.

* **AVG Example**

In this example, the average GPA for each Employee is computed  
A = LOAD ‘Employee.txt’ AS (name:chararray, term:chararray, gpa:float);  
DUMP A;  
(johny,fl,3.9F)  
(johny,wt,3.7F)  
(johny,sp,4.0F)  
(johny,sm,3.8F)  
(Mariya,fl,3.8F)  
(Mariya,wt,3.9F)  
(Mariya,sp,4.0F)  
(Mariya,sm,4.0F)  
B = GROUP A BY name;  
DUMP B;  
(johny,{(johny,fl,3.9F),(johny,wt,3.7F),(johny,sp,4.0F),(johny,sm,3.8F)})  
(Mariya,{(Mariya,fl,3.8F),(Mariya,wt,3.9F),(Mariya,sp,4.0F),(Mariya,sm,4.0F)})  
C = FOREACH B GENERATE A.name, AVG(A.gpa);  
DUMP C;  
({(johny),(johny),(johny),(johny)},3.850000023841858)  
({(Mariya),(Mariya),(Mariya),(Mariya)},3.925000011920929)

**ii. BagToString()**  
This function is used to concatenate the elements of a bag into a string. We can place a delimiter between these values (optional) while concatenating.

**iii. CONCAT()**

* **The syntax of CONCAT()**

CONCAT (expression, expression)  
We use this Pig Function to concatenate two or more expressions of the same type.

* **Example of CONCAT()**

In this example, fields f1, an underscore string literal, f2 and f3 are concatenated.  
X = LOAD ‘data’ as (f1:chararray, f2:chararray, f3:chararray);  
DUMP X;  
(apache,open,source)  
(hadoop,map,reduce)  
(pig,pig,latin)  
Y = FOREACH X GENERATE CONCAT(f1, ‘\_’, f2,f3);  
DUMP Y;  
(apache\_opensource)  
(hadoop\_mapreduce)  
(pig\_piglatin)

**iv. COUNT()**

* **The syntax of COUNT()**

COUNT(expression)  
While counting the number of tuples in a bag, we use it to get the number of elements in a bag.

* **Example of COUNT()**

In this example, we count the tuples in the bag:  
X = LOAD ‘data’ AS (f1:int,f2:int,f3:int);  
DUMP X;  
(1,2,3)  
(4,2,1)  
(8,3,4)  
(4,3,3)  
(7,2,5)  
(8,4,3)  
Y = GROUP X BY f1;  
DUMP B;  
(1,{(1,2,3)})  
(4,{(4,2,1),(4,3,3)})  
(7,{(7,2,5)})  
(8,{(8,3,4),(8,4,3)})  
A = FOREACH Y GENERATE COUNT(X);  
DUMP A;  
(1L)  
(2L)  
(1L)  
(2L)

**v. COUNT\_STAR()**

* **The syntax of COUNT\_STAR()**

COUNT\_STAR(expression)  
We can say it is similar to the COUNT() function. To get the number of elements in a bag, we use it.