**APACHE HIVE**

**I. Hive là gì?**

Hive là phần mềm cung cấp giao diện và các câu lệnh có cú pháp gần giống với SQL giúp ta dễ dàng tổng hợp, truy vấn và phân tích dữ liệu Hadoop - việc hoàn toàn không hề dễ dàng đặc biệt là với những người chưa từng làm quen với map-reduce.

Cụ thể Hive định nghĩa ra một ngôn ngữ truy vấn đơn giản có cú pháp gần giống với SQL, được biết với tên HiveSQL (hoặc HQL).

Ngoài ra ngôn ngữ này còn cho phép nhúng các mappers và reducers cho chính mình viết ra để thực thi các phân tích phức tạp.

**II. Lịch sử của Hive**

Tháng 9 năm 2008, Hive chỉ là một dự án con của Apache Hadoop.

Hiện nay Apache Hive đã trở thành dự án mã nguồn mở quan trọng, được vận hành và xây dựng bởi các tình nguyện viên tại Apache Software Foundation.

**III. Hive dùng để làm gì?**

Hive là một kho dữ liệu (data warehouse)  xử lý các dữ liệu dạng cấu trúc trên nền tảng hadoop. Sử dụng hive để tổng hợp, tạo truy vấn và phân tích dữ liệu một cách dễ dàng mà không cần phải hiểu nhiều về MapReduce. Hive cung cấp một giao diện web để tương tác với hdfs. Tương tác command line.

Điểm hay của HIVE là hỗ trợ truy xuất giống SQL đến dữ liệu có cấu trúc, được biết với tên HiveSQL (hoặc HQL) cũng như phân tích big data với MapReduce. Hive không được xây dựng để hồi đáp nhanh các câu truy vấn nhưng nó được xây dựng cho các ứng dụng khai thác dữ liệu (data mining). Các ứng dụng khai thác dữ liệu có thể mất nhiều phút đến nhiều giờ để phân tích dữ liệu và HIVE được dùng chủ yếu.

**IV. Cơ chế và nguyên lý**

**1. Cách thức hoạt động Hive trong hadoop ecosystem**

**1.1. Hoạt động**

Khi người dùng nhập ‘Query’ ở HIVE UI

Bước 1: các câu truy vấn được gửi đến Driver

Bước 2: Driver tạo ra 1 session cho truy vấn này và gửi query tới compiler để nhận execute plan (trong mỗi câu truy vấn trên hive đều phải có execution plan nó giống như là mã số để thực hiện câu lệnh vậy)

Bước 3: gửi yêu cầu lấy dữ liệu cần lấy của câu lệnh dưới metadata ở metastore

Bước 4: Metastore gửi lại dữ liệu cho compiler

Bước 5: Plan được sinh ra bởi compiler cần thiết để thực thi query sẽ được gửi lại tới phần thực thi (Execute engine)

Bước 6: Nhận lấy các execution plan từ driver và bắt đầu thực thi

Bước 6.1, 6.2, 6.3: xuống phần HDFS của hadoop ecosystem để thực thi quá trình thông qua mapreduceJob (đã được tìm hiểu qua bài hadoop mỗi câu lệnh khi muốn thực hiện ở mức HDFS phải có MapreduceJob)

Ở đây nó sẽ yêu cấu lấy các metadata cần thiết được phân tích ở bước 3, 4 và từ đó nó sẽ gọi Mapreduce thực thi công việc

Bước 7, 8, 9: Khi output được sinh ra, nó sẽ được ghi dưới dạng 1 temporary file, file này sẽ cung cấp các thông tin cần thiết cho các stages khác của plan

**1.2. Mô hình dữ liệu trong hive**

|  |  |
| --- | --- |
| Integer | Tinyint – 1 byte integer  Smallint – 2 byte integer  Int – 4 byte integer  Bigint – 8 byte integer |
| Boolean | True/false |
| Floating point number | Float – single precision  Double – double precision |
| String | sequence of characters in a specified character set |
| Others | structs  Maps  Arrays |

**1.3. Tables**

Trong hive thì có 2 loại tạo Table

1. Tạo table bình thường

Khi bạn tạo mới 1 tables thì Hive  sẽ chuyển các dữ liệu này tới tới kho dữ liệu của nó (warehouse directory)

CREATE TABLE managed\_table(dummy STRING);

LOAD DATA INPATH '/user/sample\_data/salary/salary.csv’INTO table managed\_tabl;

1. Tạo với external table

Với từ khóa EXTERNAL, Hive hiểu rằng nó không trực tiếp quản lý các dữ liệu này bởi vậy nó sẽ không thực hiện việc di chuyển chúng tới Hive’s data warehouse. Trên thực tế là Hive không thực hiện việc kiểm tra xem dữ liệu trên External localtion có tồn tại hay không. Đây là 1 tiện ích khác hữu dụng, nó cho phép ta có thể thực hiện việc khởi tạo dữ liệu khi  tạo 1 table trên  Hive. Và khi dữ liệu việc drop 1 external table thì Hive chỉ thực hiện việc xóa đi các metadata của nó.

create external table salary

(

salary\_id string,

employee\_id string,

payment double,

date string

)

ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' location '/user/sample\_data/salary/salary.csv'

**V. Tính năng**

Công cụ cho phép dễ dàng thực hiện tác vụ như trích xuất, vận chuyển và lưu trữ dữ liệu.

Cơ chế để xử lý cho nhiều định dạng dữ liệu khác nhau.

Truy cập tới dữ liệu dạng files được lưu trữ trực tiếp ở trong Apache HDFS hoặc đối với nhiều hệ thống lưu trữ dữ liệu khác như Apache HBase.

Thực hiện query thông qua MapReduce.

**VI. HiveQL**

Where:

|  |  |
| --- | --- |
| SQL | HiveSQL |
| SELECT \*  FROM table  WHERE rec1=”value1” AND rec2=”value2”; | SELECT \*  FROM TABLE  WHERE rec1 = "value1" AND rec2 = "value2"; |

Order By:

--->Tăng dần:

|  |  |
| --- | --- |
| SQL | HiveSQL |
| SELECT col1, col2  FROM table  ORDER BY col2; | SELECT col1, col2  FROM table  ORDER BY col2; |

--->Giảm dần:

|  |  |
| --- | --- |
| SQL | HiveSQL |
| SELECT col1, col2  FROM table  ORDER BY col2 DESC; | SELECT col1, col2  FROM table  ORDER BY col2 DESC; |

Group By:

|  |  |
| --- | --- |
| SQL | HiveSQL |
| SELECT owner, COUNT(\*)  FROM table  GROUP BY owner; | SELECT owner, COUNT(\*)  FROM table  GROUP BY owner; |

Joins:

|  |  |
| --- | --- |
| SQL | HiveSQL |
| SELECT pet.name, comment  FROM pet INNER JOIN event WHERE pet.name = event.name;  [LEFT JOIN]  [RIGHT JOIN]  [FULL OUTER JOIN] | SELECT pet.name, comment  FROM pet JOIN event ON (pet.name = event.name);  [LEFT JOIN]  [RIGHT JOIN]  [FULL OUTER JOIN] |

**APACHE PIG**

**I. Pig Latin**

**1. Định danh**

Định danh bao gồm tên của quan hệ (bí danh), fields, các biến... Trong Pig, định danh bắt đầu bằng chữ cái và có thể được theo sau bởi bất kỳ số lượng chữ cái, chữ số, hoặc dấu gạch dưới.

Định danh hợp lệ:

A

A123

abc\_123\_BeX\_

Định danh không hợp lệ:

\_A123

abc\_$

A!B

**2. Quan hệ, Bags, Tuples, Fields**

Các câu lệnh Pig Latin làm việc với các quan hệ. Một quan hệ có thể được định nghĩa như sau:

* Một quan hệ là 1 bag.
* Một bag là một tập hợp của các tuples.
* Một tuple là một tập các field.
* Một field là một phần dữ liệu.

Một quan hệ Pig là một bag của tuples. Một quan hệ Pig cũng tương tự như một bảng trong một cơ sở dữ liệu quan hệ, nơi các tuple trong bag tương ứng với các hàng trong một bảng. Không giống như một bảng quan hệ, quan hệ Pig không đòi hỏi rằng tất cả các tuple chứa cùng một số field hoặc các trường trong cùng một vị trí (cột) có cùng loại.

Cũng lưu ý rằng mối quan hệ đang có thứ tự có nghĩa là không có đảm bảo rằng các bộ được xử lý theo bất kỳ thứ tự cụ thể. Hơn nữa, chế biến có thể được song song trong trường hợp các bộ không được xử lý theo bất kỳ tổng đặt hàng.

Apache Pig cung cấp nhiều chức năng được xây dựng sẵn cụ thể là **eval, load, store, math, string, bag** và **tuple** chức năng.

**Chức năng Eval**

Đưa ra dưới đây là danh sách các **eval** chức năng được cung cấp bởi Apache Pig.

|  |  |
| --- | --- |
| **SN** | **Mô tả chức năng** |
| 1 | [**AVG ()**](https://www.tutorialspoint.com/apache_pig/apache_pig_avg.htm)  Để tính toán mức trung bình của các giá trị số trong một bag. |
| 2 | [**BagToString ()**](https://www.tutorialspoint.com/apache_pig/apache_pig_bagtostring.htm)  Để nối các yếu tố của một bag vào một string. Trong khi liên kết, chúng ta có thể đặt một dấu phân cách giữa các giá trị (không bắt buộc). |
| 3 | [**CONCAT ()**](https://www.tutorialspoint.com/apache_pig/apache_pig_concat.htm)  Để nối hai hoặc nhiều biểu thức của cùng loại. |
| 4 | [**COUNT()**](https://www.tutorialspoint.com/apache_pig/apache_pig_count.htm)  Để có được số phần tử trong một bag, bằng cách đếm số lượng các tuples trong một chiếc túi. |
| 5 | [**COUNT\_STAR ()**](https://www.tutorialspoint.com/apache_pig/apache_pig_count_star.htm)  Nó cũng tương tự như **COUNT ()**. Nó được sử dụng để có được số phần tử trong một bag. |
| 6 | [**DIFF ()**](https://www.tutorialspoint.com/apache_pig/apache_pig_diff.htm)  Để so sánh hai bags (fields) trong một tuple. |
| 7 | [**IsEmpty ()**](https://www.tutorialspoint.com/apache_pig/apache_pig_isempty.htm)  Để kiểm tra xem một bag hoặc map có trống hay không. |
| 8 | [**MAX ()**](https://www.tutorialspoint.com/apache_pig/apache_pig_max.htm)  Để tính giá trị cao nhất cho một cột (giá trị số hoặc chararrays) trong một bag đơn cột. |
| 9 | [**MIN ()**](https://www.tutorialspoint.com/apache_pig/apache_pig_min.htm)  Để có được giá trị nhỏ nhất (số hoặc chararray) cho một cột nào đó trong một bag đơn cột. |
| 10 | [**PluckTuple ()**](https://www.tutorialspoint.com/apache_pig/apache_pig_plucktuple.htm)  Sử dụng **PluckTuple ()**, chúng ta có thể xác định một tiền tố string và lọc các cột trong một quan hệ bắt đầu bằng tiền tố nhất định. |
| 11 | [**SIZE()**](https://www.tutorialspoint.com/apache_pig/apache_pig_size.htm)  Để tính số phần tử dựa trên bất kỳ kiểu dữ liệu Pig. |
| 12 | [**SUBTRACT ()**](https://www.tutorialspoint.com/apache_pig/apache_pig_subtract.htm)  Để nhập 2 bags. Phải mất hai bags đưa vào và trả về một bag chứa các tuples của cả 2 bags đầu. |
| 13 | [**SUM ()**](https://www.tutorialspoint.com/apache_pig/apache_pig_sum.htm)  Để có được tổng cộng các giá trị số của một cột trong một bag đơn cột. |
| 14 | [**Tokenize ()**](https://www.tutorialspoint.com/apache_pig/apache_pig_tokenize.htm)  Để chia một string (một nhóm từ) trong một tuple đơn và trả lại một bag chứa đầu ra của hoạt động phân chia. |

Các **load** và **store** các chức năng trong Apache Pig được sử dụng để xác định cách dữ liệu truyền tải và lưu trữ của Pig. Dưới đây là danh sách các chức năng tải và lưu trữ sẵn trong Pig.

|  |  |
| --- | --- |
| **SN** | **Mô tả chức năng** |
| 1 | [**PigStorage ()**](https://www.tutorialspoint.com/apache_pig/apache_pig_pigstore.htm)  Để load và store các file có cấu trúc. |
| 2 | [**TextLoader ()**](https://www.tutorialspoint.com/apache_pig/apache_pig_textloader.htm)  Để load dữ liệu phi cấu trúc vào Pig. |
| 3 | [**BinStorage ()**](https://www.tutorialspoint.com/apache_pig/apache_pig_binstorage.htm)  Để load và store dữ liệu vào Pig, sử dụng định dạng máy có thể đọc được. |
| 4 | [**Handling Compression**](https://www.tutorialspoint.com/apache_pig/apache_pig_handling_compression.htm)  Để load và store dữ liệu nén. |

Đưa ra dưới đây là danh sách các chức năng Bag và tuple.

|  |  |
| --- | --- |
| **SN** | **Mô tả chức năng** |
| 1 | [**TOBAG ()**](https://www.tutorialspoint.com/apache_pig/apache_pig_tobag.htm)  Để chuyển đổi hai hoặc nhiều biểu thức vào một bag. |
| 2 | [**TOP()**](https://www.tutorialspoint.com/apache_pig/apache_pig_top.htm)  Để có được top **N** tuples của một quan hệ. |
| 3 | [**TOTUPLE ()**](https://www.tutorialspoint.com/apache_pig/apache_pig_totuple.htm)  Để chuyển đổi một hoặc nhiều biểu thức thành một tuple. |
| 4 | [**TOMAP ()**](https://www.tutorialspoint.com/apache_pig/apache_pig_tomap.htm)  Để chuyển đổi các cặp key-value vào một map. |

Chúng ta có các chức năng string trong Apache Pig như sau:

|  |  |
| --- | --- |
| **SN** | **Chức năng & Mô tả** |
| 1 | [**ENDSWITH (string, testAgainst)**](https://www.tutorialspoint.com/apache_pig/apache_pig_endswith.htm)  Để xác minh xem một chuỗi cho trước kết thúc, bằng một chuỗi test cụ thể. |
| 2 | [**STARTSWITH (string, substring)**](https://www.tutorialspoint.com/apache_pig/apache_pig_startswith.htm)  Để xác minh xem chuỗi đầu tiên bắt đầu với chuỗi thứ hai. |
| 3 | [**SUBSTRING (string, startIndex, stopIndex)**](https://www.tutorialspoint.com/apache_pig/apache_pig_substring.htm)  Trả về một chuỗi con từ một chuỗi cho trước. StartIndex và StopIndex cho vào vị trí bắt đầu và kết thúc của chuỗi. |
| 4 | [**EqualsIgnoreCase (string1, string2)**](https://www.tutorialspoint.com/apache_pig/apache_pig_equalsignorecase.htm)  Để so sánh hai chuỗi có giống nhau không. |
| 5 | [**IndexOf (string, 'character', startIndex)**](https://www.tutorialspoint.com/apache_pig/apache_pig_indexof.htm)  Trả lại vị trí xuất hiện đầu tiên của một character trong một chuỗi. |
| 6 | [**LAST\_INDEX\_OF (string, 'character')**](https://www.tutorialspoint.com/apache_pig/apache_pig_last_index_of.htm)  Trả về chỉ số của sự xuất hiện cuối cùng của một character cho trước trong một chuỗi, đi từ đầu chuỗi. |
| 7 | [**LCFIRST (expression)**](https://www.tutorialspoint.com/apache_pig/apache_pig_lcfirst.htm)  Chuyển đổi ký tự đầu tiên trong một chuỗi chữ thường. |
| số 8 | [**UCFIRST (expression)**](https://www.tutorialspoint.com/apache_pig/apache_pig_ucfirst.htm)  Trả về một chuỗi với ký tự đầu tiên chuyển đổi sang chữ hoa. |
| 9 | [**UPPER (expression)**](https://www.tutorialspoint.com/apache_pig/apache_pig_upper.htm)  UPPER (expression) Trả về một chuỗi chuyển đổi sang chữ hoa. |
| 10 | [**LOWER (expression)**](https://www.tutorialspoint.com/apache_pig/apache_pig_lower.htm)  Chuyển đổi tất cả các ký tự trong một chuỗi chữ thường. |
| 11 | [**REPLACE (string, 'oldChar', 'newChar');**](https://www.tutorialspoint.com/apache_pig/apache_pig_replace.htm)  Để thay thế character tồn tại trong một chuỗi với các character mới. |
| 12 | [**STRSPLIT (string, regex, limit)**](https://www.tutorialspoint.com/apache_pig/apache_pig_strsplit.htm)  Chức năng này nhận một chuỗi để được chia, một biểu thức chính quy, và một giá trị số nguyên xác định giới hạn (số lượng chuỗi con của chuỗi nên được chia). |
| 13 | [**STRSPLITTOBAG (string, regex, limit)**](https://www.tutorialspoint.com/apache_pig/apache_pig_strsplittobag.htm)  Tương tự như **STRSPLIT ()**, nó chia tách các chuỗi bằng dấu phân cách nhất định () và trả về kết quả trong một bag. |
| 14 | [**TRIM (expression)**](https://www.tutorialspoint.com/apache_pig/apache_pig_trim.htm)  Trả về một bản sao của một chuỗi với khoảng trắng ở đầu và đuôi được gỡ bỏ. |
| 15 | [**LTRIM (expression)**](https://www.tutorialspoint.com/apache_pig/apache_pig_ltrim.htm)  Trả về một bản sao của một chuỗi với khoảng trắng đầu được gỡ bỏ. |
| 16 | [**RTRIM (expression)**](https://www.tutorialspoint.com/apache_pig/apache_pig_rtrim.htm)  Trả về một bản sao của một chuỗi với dấu khoảng trắng cuối được loại bỏ. |

Apache Pig cung cấp Date and Time chức năng sau:

|  |  |
| --- | --- |
| **SN** | **Chức năng & Mô tả** |
| 1 | [**ToDate(milliseconds)**](https://www.tutorialspoint.com/apache_pig/apache_pig_todate.htm)  Hàm này trả về một đối tượng ngày-thời gian theo các thông số nhất định. Các thay thế khác cho chức năng này là ToDate(iosstring), ToDate(userstring, format), ToDate(userstring, format, timezone) |
| 2 | [**CurrentTime()**](https://www.tutorialspoint.com/apache_pig/apache_pig_currenttime.htm)  trả về đối tượng ngày-thời gian của thời điểm hiện tại. |
| 3 | [**GetDay (datetime)**](https://www.tutorialspoint.com/apache_pig/apache_pig_getday.htm)  Trả lại ngày trong một tháng kể từ ngày đối tượng thời gian. |
| 4 | [**GetHour (datetime)**](https://www.tutorialspoint.com/apache_pig/apache_pig_gethour.htm)  Trả về giờ của một ngày từ các đối tượng ngày thời gian. |
| 5 | [**GetMilliSecond (datetime)**](https://www.tutorialspoint.com/apache_pig/apache_pig_getmillisecond.htm)  Trả về phần nghìn giây của một giây từ đối tượng ngày thời gian. |
| 6 | [**GetMinute (datetime)**](https://www.tutorialspoint.com/apache_pig/apache_pig_getminute.htm)  Trả về phút của một giờ từ đối tượng ngày thời gian. |
| 7 | [**GetMonth (datetime)**](https://www.tutorialspoint.com/apache_pig/apache_pig_getmonth.htm)  Trả về tháng của một năm kể từ ngày đối tượng thời gian. |
| 8 | [**GetSecond (datetime)**](https://www.tutorialspoint.com/apache_pig/apache_pig_getsecond.htm)  Trả về giây của một phút từ đối tượng ngày thời gian. |
| 9 | [**GetWeek (datetime)**](https://www.tutorialspoint.com/apache_pig/apache_pig_getweek.htm)  Trả về tuần của một năm kể từ ngày đối tượng thời gian. |
| 10 | [**GetWeekYear (datetime)**](https://www.tutorialspoint.com/apache_pig/apache_pig_getweekyear.htm)  Trả về tuần năm từ đối tượng ngày thời gian. |
| 11 | [**GetYear (datetime)**](https://www.tutorialspoint.com/apache_pig/apache_pig_getyear.htm)  Trả năm từ các đối tượng ngày thời gian. |
| 12 | [**AddDuration (datetime, duration)**](https://www.tutorialspoint.com/apache_pig/apache_pig_addduration.htm)  Trả về kết quả của một đối tượng ngày thời gian cùng với các đối tượng thời gian. |
| 13 | [**SubtractDuration (datetime, duration)**](https://www.tutorialspoint.com/apache_pig/apache_pig_subtractduration.htm)  Trừ các đối tượng Thời gian từ đối tượng Date-Time và trả về kết quả. |
| 14 | [**DaysBetween (datetime1, datetime2)**](https://www.tutorialspoint.com/apache_pig/apache_pig_daysbetween.htm)  Trả về số ngày giữa hai đối tượng ngày thời gian. |
| 15 | [**HoursBetween (datetime1, datetime2)**](https://www.tutorialspoint.com/apache_pig/apache_pig_hoursbetween.htm)  Trả về số giờ giữa hai đối tượng ngày thời gian. |
| 16 | [**MilliSecondsBetween (datetime1, datetime2)**](https://www.tutorialspoint.com/apache_pig/apache_pig_millisecondsbetween.htm)  Trả về số mili giây giữa hai ngày thời gian. |
| 17 | [**MinutesBetween (datetime1, datetime2)**](https://www.tutorialspoint.com/apache_pig/apache_pig_minutesbetween.htm)  Trả về số phút giữa hai ngày thời gian. |
| 18 | [**MonthsBetween (datetime1, datetime2)**](https://www.tutorialspoint.com/apache_pig/apache_pig_monthsbetween.htm)  Trả về số tháng giữa hai ngày thời gian. |
| 19 | [**SecondsBetween (datetime1, datetime2)**](https://www.tutorialspoint.com/apache_pig/apache_pig_secondsbetween.htm)  Trả về số giây giữa hai ngày thời gian. |
| 20 | [**WeeksBetween (datetime1, datetime2)**](https://www.tutorialspoint.com/apache_pig/apache_pig_weeksbetween.htm)  Trả về số tuần giữa hai ngày thời gian. |
| 21 | [**YearsBetween (datetime1, datetime2)**](https://www.tutorialspoint.com/apache_pig/apache_pig_yearsbetween.htm)  Trả về số năm giữa hai đối tượng ngày thời gian. |

Chúng ta có các chức năng Math sau trong Apache Pig

|  |  |
| --- | --- |
| **SN** | **Chức năng & Mô tả** |
| 1 | [**ABS (expression)**](https://www.tutorialspoint.com/apache_pig/apache_pig_abs.htm)  Để có được giá trị tuyệt đối của một biểu thức. |
| 2 | [**ACOS (expression)**](https://www.tutorialspoint.com/apache_pig/apache_pig_acos.htm)  Để có được arc cosine của một biểu thức. |
| 3 | [**ASIN (expression)**](https://www.tutorialspoint.com/apache_pig/apache_pig_asin.htm)  Để có được arc sine của một biểu thức. |
| 4 | [**ATAN (expression)**](https://www.tutorialspoint.com/apache_pig/apache_pig_atan.htm)  Chức năng này được sử dụng để có được arc tan của một biểu thức. |
| 5 | [**CBRT (expression)**](https://www.tutorialspoint.com/apache_pig/apache_pig_cbrt.htm)  Chức năng này được sử dụng để có được căn bậc 3 của một biểu thức. |
| 6 | [**CEIL (expression)**](https://www.tutorialspoint.com/apache_pig/apache_pig_ceil.htm)  Chức năng này được sử dụng để nhận được giá trị của một biểu thức làm tròn lên đến số nguyên gần nhất. |
| 7 | [**COS (expression)**](https://www.tutorialspoint.com/apache_pig/apache_pig_cos.htm)  Chức năng này được sử dụng để có được cosin lượng giác của một biểu thức. |
| 8 | [**COSH (expression)**](https://www.tutorialspoint.com/apache_pig/apache_pig_cosh.htm)  Chức năng này được sử dụng để có được những cosin hyperbolic của một biểu thức. |
| 9 | [**EXP (expression)**](https://www.tutorialspoint.com/apache_pig/apache_pig_exp.htm)  Chức năng này được sử dụng để có được số của Euler lũy thừa của x. |
| 10 | [**FLOOR (expression)**](https://www.tutorialspoint.com/apache_pig/apache_pig_floor.htm)  Để có được giá trị của một biểu thức làm tròn xuống số nguyên gần nhất. |
| 11 | [**LOG (expression)**](https://www.tutorialspoint.com/apache_pig/apache_pig_log.htm)  Để có được logarit (cơ số e) của một biểu thức. |
| 12 | [**LOG10 (expression)**](https://www.tutorialspoint.com/apache_pig/apache_pig_log10.htm)  Để có được logarit cơ số 10 của một biểu thức. |
| 13 | [**RANDOM ()**](https://www.tutorialspoint.com/apache_pig/apache_pig_random.htm)  Để có được một số ngẫu nhiên giả (loại double) lớn hơn hoặc bằng 0,0 và nhỏ hơn 1,0. |
| 14 | [**ROUND (expression)**](https://www.tutorialspoint.com/apache_pig/apache_pig_round.htm)  Để có được giá trị của một biểu thức làm tròn đến một số integer (nếu loại kết quả là float) hoặc làm tròn đến một long (nếu loại kết quả là double). |
| 15 | [**SIN (expression)**](https://www.tutorialspoint.com/apache_pig/apache_pig_sin.htm)  Để có được sin của một biểu thức. |
| 16 | [**SINH (expression)**](https://www.tutorialspoint.com/apache_pig/apache_pig_sinh.htm)  Để có được sin hyperbolic của một biểu thức. |
| 17 | [**SQRT (expression)**](https://www.tutorialspoint.com/apache_pig/apache_pig_sqrt.htm)  Để có được căn bậc hai dương của một biểu thức. |
| 18 | [**TAN (expression)**](https://www.tutorialspoint.com/apache_pig/apache_pig_tan.htm)  Để có được tan của một góc. |
| 19 | [**TANH (expression)**](https://www.tutorialspoint.com/apache_pig/apache_pig_tanh.htm)  Để có được tan hyperbolic của một biểu thức. |

**Các nguồn tham khảo:**

<https://pig.apache.org/docs>

<https://www.tutorialspoint.com/apache_pig>

<https://kipalog.com/posts/Data-Analysis-voi-Pig-Latin-Programming>