Apache Hive Built In Function

Apache Hive provides several types of built-in function which are used to perform various operations.

The following are the list of built-in function provided by Apache Hive.

1. Mathematical Functions
2. Collection Functions
3. Date Functions
4. Conditional Functions
5. String Functions

Let us see each built-in function in detail.

1. Mathematical Functions

Mathematical Functions are used to perform mathematical operations.

The following is the list of Mathematical Functions supported in Apache Hive.

ReturnType	Name	Description
DOUBLE	round(DOUBLE a)	It will return the rounded BIGINT value of a.
DOUBLE	round(DOUBLE a, INT d)	It will return a rounded to d decimal places.
DOUBLE	bround(DOUBLE a)	It will return the rounded BIGINT value of a using HALF_EVEN rounding mode (as of Hive 1.3.0, 2.0.0).
DOUBLE	bround(DOUBLE a, INT d)	It will return a rounded to d decimal places using HALF_EVEN rounding mode (as of Hive 1.3.0, 2.0.0).

DOUBLE	bround(DOUBLE a, INT d)	It will return a rounded to d decimal places using HALF_EVEN rounding mode (as of Hive 1.3.0, 2.0.0).
BIGINT	floor(DOUBLE a)	It will return the maximum BIGINT value that is equal to or less than a.
BIGINT	ceil(DOUBLE a), ceiling(DOUBLE a)	This function returns a min BIGINT value that is greater than or similar to a.
DOUBLE	rand(), rand(INT seed)	It will return a random number (that changes from row to row) that is distributed uniformly from 0 to 1.
DOUBLE	exp(DOUBLE a), exp(DECIMAL a)	It will return ea where e is the base of the natural logarithm. Decimal version added in Hive 0.13.0.
DOUBLE	ln(DOUBLE a), ln(DECIMAL a)	It will return the natural logarithm of the argument a. Decimal version added in Hive 0.13.0.
DOUBLE	log10(DOUBLE a), log10(DECIMAL a)	It will return the base-10 logarithm of the argument a. Decimal version added in Hive 0.13.0.
DOUBLE	log2(DOUBLE a), log2(DECIMAL a)	It will return the base-2 logarithm of the argument a. Decimal version added in Hive 0.13.0.
DOUBLE	log(DOUBLE base, DOUBLE a)	It will return the base-base logarithm of the argument a. Decimal versions added in Hive 0.13.0.
	log(DECIMAL base, DECIMAL a)	
DOUBLE	pow(DOUBLE a, DOUBLE p), power(DOUBLE a, DOUBLE p)	It will return ap.
DOUBLE	sqrt(DOUBLE a), sqrt(DECIMAL a)	It will return the square root of a. Decimal version added in Hive 0.13.0.
STRING	bin(BIGINT a)	It will return the number in binary format.
STRING	hex(BIGINT h) hex(STRING h) hex(BINARY h)	If the argument is an INT or binary, hex returns the number as a STRING in hexadecimal format.
BINARY	unhex(STRING a)	Inverse of hex.
STRING	conv(BIGINT number1, INT from_base, INT to_base), conv(STRING number1, INT from_base, INT to_base)	It will converts a number from a given base to another.
DOUBLE	abs(DOUBLE a)	It will return the absolute value.
INT or DOUBLE	pmod(INT a, INT b), pmod(DOUBLE a, DOUBLE b)	It will return the positive value of a mod b.
DOUBLE	sin(DOUBLE a), sin(DECIMAL a)	It will return the sine of a (a is in radians). Decimal version added in Hive 0.13.0.
DOUBLE	asin(DOUBLE a), asin(DECIMAL a)	It will return the arc sin of a if -1<=a<=1 or NULL otherwise. Decimal version added in Hive 0.13.0.
DOUBLE	cos(DOUBLE a), cos(DECIMAL a)	It will return the cosine of a (a is in radians). Decimal version added in Hive 0.13.0.
DOUBLE	acos(DOUBLE a), acos(DECIMAL a)	It will return the arccosine of a if -1<=a<=1 or NULL otherwise. Decimal version added in Hive 0.13.0.

DOUBLE	atan(DOUBLE a), atan(DECIMAL a)	It will return the arctangent of a. Decimal version added in Hive 0.13.0.
DOUBLE	degrees(DOUBLE a), degrees(DECIMAL a)	Converts the value of a from radians to degrees. Decimal version added in Hive 0.13.0.
DOUBLE	radians(DOUBLE a), radians(DOUBLE a)	Converts the value of a from degrees to radians. Decimal version added in Hive 0.13.0.
INT or DOUBLE	positive(INT a), positive(DOUBLE a)	It will return a.
INT or DOUBLE	negative(INT a), negative(DOUBLE a)	It will return -a.
DOUBLE or INT	sign(DOUBLE a), sign(DECIMAL a)	It will return the sign of a as '1.0' (if a is positive) or '-1.0' (if a is negative), '0.0' otherwise. The decimal version will return INT instead of DOUBLE. Decimal version added in Hive 0.13.0.
DOUBLE	e()	It will return the value of e.
DOUBLE	pi()	It will return the value of pi.
BIGINT	factorial(INT a)	It will return the factorial of an (as of Hive 1.2.0). Valid a is [0..20].
DOUBLE	cbrt(DOUBLE a)	It will return the cube root of a double value (as of Hive 1.2.0).
INT	shiftleft(TINYINT|SMALLINT|INT a, INT b)	Bitwise left shift (as of Hive 1.2.0). Shifts a b position to the left.
BIGINT	shiftleft(BIGINT a, INT b)	It will return int for tinyint, smallint, and int a. Return bigint for bigint a.

INT	shiftright(TINYINT|SMALLINT|INT a, INT b)	Bitwise right shift (as of Hive 1.2.0). Shifts a b position to the right.
BIGINT	shiftright(BIGINT a, INT b)	It will return int for tinyint, smallint, and int a. Return bigint for bigint a.
INT	shiftrightunsigned(TINYINT|SMALLINT|INT a, INT b),	Bitwise unsigned right shift (as of Hive 1.2.0). Shifts a b position to the right.
BIGINT	shiftrightunsigned(BIGINT a, INT b)	It will return int for tinyint, smallint, and int a. Return bigint for bigint a.
T	greatest(T v1, T v2, ...)	It will return the greatest value of the list of values (as of Hive 1.1.0). Fixed to return NULL when one or more arguments are NULL, and strict type restriction relaxed, consistent with ">" operator (as of Hive 2.0.0).
T	least(T v1, T v2, ...)	It will return the least value of the list of values (as of Hive 1.1.0). Fixed to return NULL when one or more arguments are NULL, and strict type restriction relaxed, consistent with "<" operator (as of Hive 2.0.0).
INT	width_bucket(NUMERIC expr, NUMERIC min_value, NUMERIC max_value, INT num_buckets)	It will return an integer between 0 and num_buckets+1 by mapping expr into the ith equally sized bucket. Buckets are made by dividing [min_value, max_value] into equally sized regions. If expr < min_value, return 1, if expr > max_value return num_buckets+1.

2. Collection Functions

These built-in functions are used for collections. A collection represents the grouping of elements and returning a single or array of elements depends on the return type mentioned in the function name.

The following list of Collection Functions is supported in Apache Hive.

ReturnType	Name	Description
int	size(Map<K.V>)	It will return the number of elements in the map type.
int	size(Array<T>)	It will return the number of elements in the array type.
array <K>	map_keys(Map <K.V>)	It will return an unordered array containing the keys of the input map.
array <V>	map_values(Map <K.V>)	It will return an unordered array containing the values of the input map.
boolean	array_contains(Array <T>, value)	It will return TRUE if the array contains value.
array <t>	sort_array(Array <T>)	Perform sorting of input array in ascending order according and return the result.

3. Date Functions

Dates built-in functions are used to operate on date data types such as adding the number of days to the date or other similar operations.

The following are the lists of Date Functions supported in Apache Hive.

ReturnType	Name	Description
string	from_unixtime(bigint unixtime[, string format])	This function is used to convert the UNIX epoch seconds into the string of the current system timestamp.
bigint	unix_timestamp()	This function fetches the Unix current timestamp in seconds.
bigint	unix_timestamp(string date)	It will convert time string in format yyyy-MM-dd HH:mm:ss to Unix timestamp (in seconds).
bigint	unix_timestamp(string date, string pattern)	This function converts the time string into a provided pattern.
int	year(string date)	It returns the year part of a date or a timestamp string: year("2020-01-01 00:00:00") = 2020, year("2020-01-01") = 2020.
int	quarter(date or timestamp or string)	It will return the quarter of the year for a date, timestamp, or string in the range 1 to 4.
int	month(string date)	This function returns the month of a date or a timestamp string.
int	day(string date) dayofmonth(date)	This function returns the day of a date or a timestamp string.
int	hour(string date)	This function returns the hour of a timestamp.
int	minute(string date)	It will return the minute of the timestamp.

int	second(string date)	It will return the second of the timestamp.
int	weekofyear(string date)	This function returns the week number from a timestamp string.
string	add_months(string startdate_column, int nummonths_column, output_date_format)	It will return the date that is num_months after startdate_column (as of Hive 1.1.0). startdate_column is a string, date or timestamp. nummonths_column is an integer.
string	last_day(string date)	It will return the last day of the month to which the date belongs.
string	next_day(string startdate_column, string dayofweek_column)	This function returns the first date that is later than the start date.
string	trunc(string date, string format)	This function returns the date truncated to a unit that is mentioned in the format.
double	months_between(date1, date2)	It will return number of months between dates date1 and date2.
string	date_format(date/timestamp/string ts, string fmt)	This function converts the date or timestamp or string into a value of the string.

4. Conditional Functions

These built-in functions work on test conditions. If the test condition is true then it will return true.

The following is the list of Conditional Functions supported in Apache Hive.

ReturnType	Name	Description
T	if(boolean testCondition, T valueTrue, T valueFalseOrNull)	It will return valueTrue when testCondition is true, returns valueFalseOrNull otherwise.
boolean	isnull(a)	It will return true if a is NULL and false otherwise.
boolean	isnotnull (a)	It will return true if a is not NULL and false otherwise.
T	nvl(T value, T default_value)	It will return default value if value is null else returns value (as of HIve 0.11).
T	COALESCE(T v1, T v2, ...)	It will return the first v that is not NULL, or NULL if all v's are NULL.
T	CASE a WHEN b THEN c [WHEN d THEN e]* [ELSE f] END	When a = b, returns c; when a = d, returns e; else returns f.
T	CASE WHEN a THEN b [WHEN c THEN d]* [ELSE e] END	When a = true, returns b; when c = true, returns d; else returns e.
T	nullif(a, b)	It will return NULL if a=b; otherwise returns a (as of Hive 2.3.0).
void	assert_true(boolean condition)	It will throw an exception if 'condition' is not true, otherwise return null (as of Hive 0.8.0). For example, select assert_true (2<1).

5. String Functions

String built-in functions are used to perform different operations like reversing sting, converting into upper and lower case, removing spaces, and so on.

Below are the lists of String Functions supported in Hive.

ReturnType	Name	Description
int	ascii(string strdata)	This function returns the numeric value of the first character of strdata.
string	base64(binary bin)	The function will convert the argument passed into a base 64 string.
int	character_length(string str_data)	This function returns the number of UTF-8 characters that contain str_data.
string	chr(bigint|double y)	This function returns the ASCII character that has a binary equivalent to Y.
string	concat(string|binary Y, string|binary Z...)	This function returns the string or bytes by performing concatenation of strings.
array	context_ngrams(array>, array, int E, int F)	This function returns the top set of tokenized sentences.
string	concat_ws(string SEP, string Y, string Z...)	This function is similar to Concat() but also provide custom separator SEP.
string	concat_ws(string SEP, array)	It is similar to concat_ws() but it also takes an array of strings.
string	decode(binary bin, string charset)	This function uses the provides character set to decodes the first argument into the string.
binary	encode(string src, string charset)	This function uses the provides character set to encodes the first argument into the binary.
int	find_in_set(string strdata, string strList_data)	This function returns the first appearance of strdata from strList_data in which strList_data is the comma-delimited string.
int	length(string ZY)	This function will return the length of a provided string.
int	locate(string substr_data, string str_data[, int pos_data])	This function returns the first occurrence position of substr_data in the string str_data that too after pos_data position.
string	lower(string O) lcase(string O)	This function will convert all characters of O in lower case.
string	ltrim(string H)	This function will return the output by trimming the space from the beginning of H.
string	printf(String format, Obj... args)	It will return the input formatted according to do printf-style format strings.
string	repeat(string str, int n)	Repeats str n times.
string	replace(string A, string OLD, string NEW)	It will return the string A with all non-overlapping occurrences of OLD replaced with NEW .
string	reverse(string A)	It will return the reversed string.
string	rpad(string strdata3, int lendata3, string paddata3)	It will return str, right-padded with pad to a length of len. If str is longer than len, the return value is shortened to len characters. In the case of an empty pad string, the return value is null.

string	rtrim(string R)	It will return the string resulting from trimming spaces from the end(right-hand side) of R.
string	space(int n)	It will return a string of n spaces.
array	split(string str, string pat)	Splits str around pat (pat is a regular expression).
string	trim(string L)	It will return the string resulting from trimming spaces from both ends of L.
binary	unbase64(string str)	Converts the argument from a base 64 string to BINARY.
string	upper(string P) ucase(string Q)	It will return the string resulting from converting all characters of P to upper case.
string	initcap(string P)	It will return the string, with the first letter of each word in uppercase, all other letters in lowercase.
int	levenshtein(string I, string J)	It will return the Levenshtein distance between two strings.
string	soundex(string H)	It will return the soundex code of the string.