Hive内置函数大全

2017年08月01日 15:05:14

摘要

Hive内部提供了很多函数给开发者使用，包括数学函数，类型转换函数，条件函数，字符函数，聚合函数，表生成函数等等，这些函数都统称为内置函数。

目录

* 数学函数
* 集合函数
* 类型转换函数
* 日期函数
* 条件函数
* 字符函数
* 聚合函数
* 表生成函数

数学函数

| Return Type | **Name (Signature)** | **Description** |
| --- | --- | --- |
| DOUBLE | round(DOUBLE a) | Returns the rounded BIGINT value of a.  返回对a四舍五入的BIGINT值 |
| DOUBLE | round(DOUBLE a, INT d) | Returns a rounded to d decimal places.  返回DOUBLE型d的保留n位小数的DOUBLW型的近似值 |
| DOUBLE | bround(DOUBLE a) | Returns the rounded BIGINT value of a using HALF\_EVEN rounding mode (as of [Hive 1.3.0, 2.0.0](https://issues.apache.org/jira/browse/HIVE-11103)). Also known as Gaussian rounding or bankers' rounding. Example: bround(2.5) = 2, bround(3.5) = 4. 银行家舍入法（1~4：舍，6~9：进，5->前位数是偶：舍，5->前位数是奇：进） |
| DOUBLE | bround(DOUBLE a, INT d) | Returns a rounded to d decimal places using HALF\_EVEN rounding mode (as of [Hive 1.3.0, 2.0.0](https://issues.apache.org/jira/browse/HIVE-11103)). Example: bround(8.25, 1) = 8.2, bround(8.35, 1) = 8.4. 银行家舍入法,保留d位小数 |
| BIGINT | floor(DOUBLE a) | Returns the maximum BIGINT value that is equal to or less than a  向下取整，最数轴上最接近要求的值的左边的值  如：6.10->6   -3.4->-4 |
| BIGINT | ceil(DOUBLE a), ceiling(DOUBLE a) | Returns the minimum BIGINT value that is equal to or greater than a.  求其不小于小给定实数的最小整数如：ceil(6) = ceil(6.1)= ceil(6.9) = 6 |
| DOUBLE | rand(), rand(INT seed) | Returns a random number (that changes from row to row) that is distributed uniformly from 0 to 1. Specifying the seed will make sure the generated random number sequence is deterministic.  每行返回一个DOUBLE型随机数seed是随机因子 |
| DOUBLE | exp(DOUBLE a), exp(DECIMAL a) | Returns ea where e is the base of the natural logarithm. Decimal version added in [Hive 0.13.0](https://issues.apache.org/jira/browse/HIVE-6327).  返回e的a幂次方， a可为小数 |
| DOUBLE | ln(DOUBLE a), ln(DECIMAL a) | Returns the natural logarithm of the argument a. Decimal version added in [Hive 0.13.0](https://issues.apache.org/jira/browse/HIVE-6327).  以自然数为底d的对数，a可为小数 |
| DOUBLE | log10(DOUBLE a), log10(DECIMAL a) | Returns the base-10 logarithm of the argument a. Decimal version added in [Hive 0.13.0](https://issues.apache.org/jira/browse/HIVE-6327).  以10为底d的对数，a可为小数 |
| DOUBLE | log2(DOUBLE a), log2(DECIMAL a) | Returns the base-2 logarithm of the argument a. Decimal version added in [Hive 0.13.0](https://issues.apache.org/jira/browse/HIVE-6327).  以2为底数d的对数，a可为小数 |
| DOUBLE | log(DOUBLE base, DOUBLE a)  log(DECIMAL base, DECIMAL a) | Returns the base-base logarithm of the argument a. Decimal versions added in [Hive 0.13.0](https://issues.apache.org/jira/browse/HIVE-6327).  以base为底的对数，base 与 a都是DOUBLE类型 |
| DOUBLE | pow(DOUBLE a, DOUBLE p), power(DOUBLE a, DOUBLE p) | Returns ap.  计算a的p次幂 |
| DOUBLE | sqrt(DOUBLE a), sqrt(DECIMAL a) | Returns the square root of a. Decimal version added in [Hive 0.13.0](https://issues.apache.org/jira/browse/HIVE-6327).  计算a的平方根 |
| STRING | bin(BIGINT a) | Returns the number in binary format (see <http://dev.mysql.com/doc/refman/5.0/en/string-functions.html#function_bin>).  计算二进制a的STRING类型，a为BIGINT类型 |
| STRING | hex(BIGINT a) hex(STRING a) hex(BINARY a) | If the argument is an INT or binary, hex returns the number as a STRING in hexadecimal format. Otherwise if the number is a STRING, it converts each character into its hexadecimal representation and returns the resulting STRING. (See<http://dev.mysql.com/doc/refman/5.0/en/string-functions.html#function_hex>, BINARY version as of Hive [0.12.0](https://issues.apache.org/jira/browse/HIVE-2482).)  计算十六进制a的STRING类型，如果a为STRING类型就转换成字符相对应的十六进制 |
| BINARY | unhex(STRING a) | Inverse of hex. Interprets each pair of characters as a hexadecimal number and converts to the byte representation of the number. (BINARY version as of Hive [0.12.0](https://issues.apache.org/jira/browse/HIVE-2482), used to return a string.)  hex的逆方法 |
| STRING | conv(BIGINT num, INT from\_base, INT to\_base), conv(STRING num, INT from\_base, INT to\_base) | Converts a number from a given base to another (see <http://dev.mysql.com/doc/refman/5.0/en/mathematical-functions.html#function_conv>).  将GIGINT/STRING类型的num从from\_base进制转换成to\_base进制 |
| DOUBLE | abs(DOUBLE a) | Returns the absolute value.  计算a的绝对值 |
| INT or DOUBLE | pmod(INT a, INT b), pmod(DOUBLE a, DOUBLE b) | Returns the positive value of a mod b.  a对b取模 |
| DOUBLE | sin(DOUBLE a), sin(DECIMAL a) | Returns the sine of a (a is in radians). Decimal version added in [Hive 0.13.0](https://issues.apache.org/jira/browse/HIVE-6327).  求a的正弦值 |
| DOUBLE | asin(DOUBLE a), asin(DECIMAL a) | Returns the arc sin of a if -1<=a<=1 or NULL otherwise. Decimal version added in [Hive 0.13.0](https://issues.apache.org/jira/browse/HIVE-6327).  求d的反正弦值 |
| DOUBLE | cos(DOUBLE a), cos(DECIMAL a) | Returns the cosine of a (a is in radians). Decimal version added in [Hive 0.13.0](https://issues.apache.org/jira/browse/HIVE-6327).  求余弦值 |
| DOUBLE | acos(DOUBLE a), acos(DECIMAL a) | Returns the arccosine of a if -1<=a<=1 or NULL otherwise. Decimal version added in [Hive 0.13.0](https://issues.apache.org/jira/browse/HIVE-6327).  求反余弦值 |
| DOUBLE | tan(DOUBLE a), tan(DECIMAL a) | Returns the tangent of a (a is in radians). Decimal version added in [Hive 0.13.0](https://issues.apache.org/jira/browse/HIVE-6327).  求正切值 |
| DOUBLE | atan(DOUBLE a), atan(DECIMAL a) | Returns the arctangent of a. Decimal version added in [Hive 0.13.0](https://issues.apache.org/jira/browse/HIVE-6327).  求反正切值 |
| DOUBLE | degrees(DOUBLE a), degrees(DECIMAL a) | Converts value of a from radians to degrees. Decimal version added in [Hive 0.13.0](https://issues.apache.org/jira/browse/HIVE-6385).  奖弧度值转换角度值 |
| DOUBLE | radians(DOUBLE a), radians(DOUBLE a) | Converts value of a from degrees to radians. Decimal version added in [Hive 0.13.0](https://issues.apache.org/jira/browse/HIVE-6327).  将角度值转换成弧度值 |
| INT or DOUBLE | positive(INT a), positive(DOUBLE a) | Returns a.  返回a |
| INT or DOUBLE | negative(INT a), negative(DOUBLE a) | Returns -a.  返回a的相反数 |
| DOUBLE or INT | sign(DOUBLE a), sign(DECIMAL a) | Returns the sign of a as '1.0' (if a is positive) or '-1.0' (if a is negative), '0.0' otherwise. The decimal version returns INT instead of DOUBLE. Decimal version added in [Hive 0.13.0](https://issues.apache.org/jira/browse/HIVE-6246).  如果a是正数则返回1.0，是负数则返回-1.0，否则返回0.0 |
| DOUBLE | e() | Returns the value of e.  数学常数e |
| DOUBLE | pi() | Returns the value of pi.  数学常数pi |
| BIGINT | factorial(INT a) | Returns the factorial of a (as of Hive [1.2.0](https://issues.apache.org/jira/browse/HIVE-9857)). Valid a is [0..20]. 求a的阶乘 |
| DOUBLE | cbrt(DOUBLE a) | Returns the cube root of a double value (as of Hive [1.2.0](https://issues.apache.org/jira/browse/HIVE-9858)). 求a的立方根 |
| INT BIGINT | shiftleft(TINYINT|SMALLINT|INT a, INT b)  shiftleft(BIGINT a, INT b) | Bitwise left shift (as of Hive [1.2.0](https://issues.apache.org/jira/browse/HIVE-9859)). Shifts a b positions to the left.  Returns int for tinyint, smallint and int a. Returns bigint for bigint a.  按位左移 |
| INT  BIGINT | shiftright(TINYINT|SMALLINT|INT a, INTb)  shiftright(BIGINT a, INT b) | Bitwise right shift (as of Hive [1.2.0](https://issues.apache.org/jira/browse/HIVE-9859)). Shifts a b positions to the right.  Returns int for tinyint, smallint and int a. Returns bigint for bigint a.  按拉右移 |
| INT  BIGINT | shiftrightunsigned(TINYINT|SMALLINT|INTa, INT b),  shiftrightunsigned(BIGINT a, INT b) | Bitwise unsigned right shift (as of Hive [1.2.0](https://issues.apache.org/jira/browse/HIVE-9859)). Shifts a b positions to the right.  Returns int for tinyint, smallint and int a. Returns bigint for bigint a.  无符号按位右移（<<<） |
| T | greatest(T v1, T v2, ...) | Returns the greatest value of the list of values (as of Hive [1.1.0](https://issues.apache.org/jira/browse/HIVE-9402)). 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](https://issues.apache.org/jira/browse/HIVE-12082)). 求最大值 |
| T | least(T v1, T v2, ...) | Returns the least value of the list of values (as of Hive [1.1.0](https://issues.apache.org/jira/browse/HIVE-9402)). 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](https://issues.apache.org/jira/browse/HIVE-12082)). 求最小值 |

集合函数

| Return Type | **Name(Signature)** | **Description** |
| --- | --- | --- |
| int | size(Map<K.V>) | Returns the number of elements in the map type.  求map的长度 |
| int | size(Array<T>) | Returns the number of elements in the array type.  求数组的长度 |
| array<K> | map\_keys(Map<K.V>) | Returns an unordered array containing the keys of the input map.  返回map中的所有key |
| array<V> | map\_values(Map<K.V>) | Returns an unordered array containing the values of the input map.  返回map中的所有value |
| boolean | array\_contains(Array<T>, value) | Returns TRUE if the array contains value.  如该数组Array<T>包含value返回true。，否则返回false |
| array | sort\_array(Array<T>) | Sorts the input array in ascending order according to the natural ordering of the array elements and returns it (as of version [0.9.0](https://issues.apache.org/jira/browse/HIVE-2279)).  按自然顺序对数组进行排序并返回 |

类型转换函数

| Return Type | Name(Signature) | **Description** |
| --- | --- | --- |
| binary | binary(string|binary) | Casts the parameter into a binary.  将输入的值转换成二进制 |
| Expected "=" to follow "type" | cast(expr as <type>) | Converts the results of the expression expr to <type>. For example, cast('1' as BIGINT) will convert the string '1' to its integral representation. A null is returned if the conversion does not succeed. If cast(expr as boolean) Hive returns true for a non-empty string.  将expr转换成type类型 如：cast("1" as BIGINT) 将字符串1转换成了BIGINT类型，如果转换失败将返回NULL |

日期函数

| Return Type | Name(Signature) | **Description** |
| --- | --- | --- |
| string | from\_unixtime(bigint unixtime[, string format]) | Converts the number of seconds from unix epoch (1970-01-01 00:00:00 UTC) to a string representing the timestamp of that moment in the current system time zone in the format of "1970-01-01 00:00:00".  将时间的秒值转换成format格式（format可为“yyyy-MM-dd hh:mm:ss”,“yyyy-MM-dd hh”,“yyyy-MM-dd hh:mm”等等）如from\_unixtime(1250111000,"yyyy-MM-dd") 得到2009-03-12 |
| bigint | unix\_timestamp() | Gets current Unix timestamp in seconds.  获取本地时区下的时间戳 |
| bigint | unix\_timestamp(string date) | Converts time string in format yyyy-MM-dd HH:mm:ss to Unix timestamp (in seconds), using the default timezone and the default locale, return 0 if fail: unix\_timestamp('2009-03-20 11:30:01') = 1237573801  将格式为yyyy-MM-dd HH:mm:ss的时间字符串转换成时间戳  如unix\_timestamp('2009-03-20 11:30:01') = 1237573801 |
| bigint | unix\_timestamp(string date, string pattern) | Convert time string with given pattern (see [<http://docs.oracle.com/javase/tutorial/i18n/format/simpleDateFormat.html>]) to Unix time stamp (in seconds), return 0 if fail: unix\_timestamp('2009-03-20', 'yyyy-MM-dd') = 1237532400.  将指定时间字符串格式字符串转换成Unix时间戳，如果格式不对返回0 如：unix\_timestamp('2009-03-20', 'yyyy-MM-dd') = 1237532400 |
| string | to\_date(string timestamp) | Returns the date part of a timestamp string: to\_date("1970-01-01 00:00:00") = "1970-01-01".  返回时间字符串的日期部分 |
| int | year(string date) | Returns the year part of a date or a timestamp string: year("1970-01-01 00:00:00") = 1970, year("1970-01-01") = 1970.  返回时间字符串的年份部分 |
| int | quarter(date/timestamp/string) | Returns the quarter of the year for a date, timestamp, or string in the range 1 to 4 (as of Hive [1.3.0](https://issues.apache.org/jira/browse/HIVE-3404)). Example: quarter('2015-04-08') = 2.  返回当前时间属性哪个季度 如quarter('2015-04-08') = 2 |
| int | month(string date) | Returns the month part of a date or a timestamp string: month("1970-11-01 00:00:00") = 11, month("1970-11-01") = 11.  返回时间字符串的月份部分 |
| int | day(string date) dayofmonth(date) | Returns the day part of a date or a timestamp string: day("1970-11-01 00:00:00") = 1, day("1970-11-01") = 1.  返回时间字符串的天 |
| int | hour(string date) | Returns the hour of the timestamp: hour('2009-07-30 12:58:59') = 12, hour('12:58:59') = 12.  返回时间字符串的小时 |
| int | minute(string date) | Returns the minute of the timestamp.  返回时间字符串的分钟 |
| int | second(string date) | Returns the second of the timestamp.  返回时间字符串的秒 |
| int | weekofyear(string date) | Returns the week number of a timestamp string: weekofyear("1970-11-01 00:00:00") = 44, weekofyear("1970-11-01") = 44.  返回时间字符串位于一年中的第几个周内  如weekofyear("1970-11-01 00:00:00") = 44, weekofyear("1970-11-01") = 44 |
| int | datediff(string enddate, string startdate) | Returns the number of days from startdate to enddate: datediff('2009-03-01', '2009-02-27') = 2.  计算开始时间startdate到结束时间enddate相差的天数 |
| string | date\_add(string startdate, int days) | Adds a number of days to startdate: date\_add('2008-12-31', 1) = '2009-01-01'.  从开始时间startdate加上days |
| string | date\_sub(string startdate, int days) | Subtracts a number of days to startdate: date\_sub('2008-12-31', 1) = '2008-12-30'.  从开始时间startdate减去days |
| timestamp | from\_utc\_timestamp(timestamp, string timezone) | Assumes given timestamp is UTC and converts to given timezone (as of Hive [0.8.0](https://issues.apache.org/jira/browse/HIVE-2272)). For example, from\_utc\_timestamp('1970-01-01 08:00:00','PST') returns 1970-01-01 00:00:00.  如果给定的时间戳并非UTC，则将其转化成指定的时区下时间戳 |
| timestamp | to\_utc\_timestamp(timestamp, string timezone) | Assumes given timestamp is in given timezone and converts to UTC (as of Hive [0.8.0](https://issues.apache.org/jira/browse/HIVE-2272)). For example, to\_utc\_timestamp('1970-01-01 00:00:00','PST') returns 1970-01-01 08:00:00.  如果给定的时间戳指定的时区下时间戳，则将其转化成UTC下的时间戳 |
| date | current\_date | Returns the current date at the start of query evaluation (as of Hive [1.2.0](https://issues.apache.org/jira/browse/HIVE-5472)). All calls of current\_date within the same query return the same value.  返回当前时间日期 |
| timestamp | current\_timestamp | Returns the current timestamp at the start of query evaluation (as of Hive [1.2.0](https://issues.apache.org/jira/browse/HIVE-5472)). All calls of current\_timestamp within the same query return the same value.  返回当前时间戳 |
| string | add\_months(string start\_date, int num\_months) | Returns the date that is num\_months after start\_date (as of Hive [1.1.0](https://issues.apache.org/jira/browse/HIVE-9357)). start\_date is a string, date or timestamp. num\_months is an integer. The time part of start\_date is ignored. If start\_date is the last day of the month or if the resulting month has fewer days than the day component of start\_date, then the result is the last day of the resulting month. Otherwise, the result has the same day component as start\_date.  返回当前时间下再增加num\_months个月的日期 |
| string | last\_day(string date) | Returns the last day of the month which the date belongs to (as of Hive [1.1.0](https://issues.apache.org/jira/browse/HIVE-9358)). date is a string in the format 'yyyy-MM-dd HH:mm:ss' or 'yyyy-MM-dd'. The time part of date is ignored.  返回这个月的最后一天的日期，忽略时分秒部分（HH:mm:ss） |
| string | next\_day(string start\_date, string day\_of\_week) | Returns the first date which is later than start\_date and named as day\_of\_week (as of Hive[1.2.0](https://issues.apache.org/jira/browse/HIVE-9520)). start\_date is a string/date/timestamp. day\_of\_week is 2 letters, 3 letters or full name of the day of the week (e.g. Mo, tue, FRIDAY). The time part of start\_date is ignored. Example: next\_day('2015-01-14', 'TU') = 2015-01-20.  返回当前时间的下一个星期X所对应的日期 如：next\_day('2015-01-14', 'TU') = 2015-01-20  以2015-01-14为开始时间，其下一个星期二所对应的日期为2015-01-20 |
| string | trunc(string date, string format) | Returns date truncated to the unit specified by the format (as of Hive [1.2.0](https://issues.apache.org/jira/browse/HIVE-9480)). Supported formats: MONTH/MON/MM, YEAR/YYYY/YY. Example: trunc('2015-03-17', 'MM') = 2015-03-01.  返回时间的最开始年份或月份  如trunc("2016-06-26",“MM”)=2016-06-01  trunc("2016-06-26",“YY”)=2016-01-01   注意所支持的格式为MONTH/MON/MM, YEAR/YYYY/YY |
| double | months\_between(date1, date2) | Returns number of months between dates date1 and date2 (as of Hive [1.2.0](https://issues.apache.org/jira/browse/HIVE-9518)). If date1 is later than date2, then the result is positive. If date1 is earlier than date2, then the result is negative. If date1 and date2 are either the same days of the month or both last days of months, then the result is always an integer. Otherwise the UDF calculates the fractional portion of the result based on a 31-day month and considers the difference in time components date1 and date2. date1 and date2 type can be date, timestamp or string in the format 'yyyy-MM-dd' or 'yyyy-MM-dd HH:mm:ss'. The result is rounded to 8 decimal places. Example: months\_between('1997-02-28 10:30:00', '1996-10-30') = 3.94959677  返回date1与date2之间相差的月份，如date1>date2，则返回正，如果date1<date2,则返回负，否则返回0.0  如：months\_between('1997-02-28 10:30:00', '1996-10-30') = 3.94959677  1997-02-28 10:30:00与1996-10-30相差3.94959677个月 |
| string | date\_format(date/timestamp/string ts, string fmt) | Converts a date/timestamp/string to a value of string in the format specified by the date format fmt (as of Hive [1.2.0](https://issues.apache.org/jira/browse/HIVE-10276)). Supported formats are Java SimpleDateFormat formats –<https://docs.oracle.com/javase/7/docs/api/java/text/SimpleDateFormat.html>. The second argument fmt should be constant. Example: date\_format('2015-04-08', 'y') = '2015'.  date\_format can be used to implement other UDFs, e.g.:   * dayname(date) is date\_format(date, 'EEEE') * dayofyear(date) is date\_format(date, 'D')   按指定格式返回时间date 如：date\_format("2016-06-22","MM-dd")=06-22 |

条件函数

| Return Type | **Name(Signature)** | **Description** |
| --- | --- | --- |
| T | if(boolean testCondition, T valueTrue, T valueFalseOrNull) | Returns valueTrue when testCondition is true, returns valueFalseOrNull otherwise.  如果testCondition 为true就返回valueTrue,否则返回valueFalseOrNull ，（valueTrue，valueFalseOrNull为泛型） |
| T | nvl(T value, T default\_value) | Returns default value if value is null else returns value (as of HIve [0.11](https://issues.apache.org/jira/browse/HIVE-2288)).  如果value值为NULL就返回default\_value,否则返回value |
| T | COALESCE(T v1, T v2, ...) | Returns the first v that is not NULL, or NULL if all v's are NULL.  返回第一非null的值，如果全部都为NULL就返回NULL  如：COALESCE (NULL,44,55)=44/strong> |
| 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.  如果a=b就返回c,a=d就返回e，否则返回f  如CASE 4 WHEN 5  THEN 5 WHEN 4 THEN 4 ELSE 3 END 将返回4 |
| 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.  如果a=ture就返回b,c= ture就返回d，否则返回e  如：CASE WHEN  5>0  THEN 5 WHEN 4>0 THEN 4 ELSE 0 END 将返回5；CASE WHEN  5<0  THEN 5 WHEN 4<0 THEN 4 ELSE 0 END 将返回0 |
| boolean | isnull( a ) | Returns true if a is NULL and false otherwise.  如果a为null就返回true，否则返回false |
| boolean | isnotnull ( a ) | Returns true if a is not NULL and false otherwise.  如果a为非null就返回true，否则返回false |

字符函数

| Return Type | **Name(Signature)** | **Description** |
| --- | --- | --- |
| int | ascii(string str) | Returns the numeric value of the first  character of str.  返回str中首个ASCII字符串的整数值 |
| string | base64(binary bin) | Converts the argument from binary to a base 64 string (as of Hive [0.12.0](https://issues.apache.org/jira/browse/HIVE-2482))..  将二进制bin转换成64位的字符串 |
| string | concat(string|binary A, string|binary B...) | Returns the string or bytes resulting from concatenating the strings or bytes passed in as parameters in order. For example, concat('foo', 'bar') results in 'foobar'. Note that this function can take any number of input strings..  对二进制字节码或字符串按次序进行拼接 |
| array<struct<string,double>> | context\_ngrams(array<array<string>>, array<string>, int K, int pf) | Returns the top-k contextual N-grams from a set of tokenized sentences, given a string of "context". See [StatisticsAndDataMining](https://cwiki.apache.org/confluence/display/Hive/StatisticsAndDataMining) for more information..  与ngram类似，但context\_ngram()允许你预算指定上下文(数组)来去查找子序列，具体看[StatisticsAndDataMining](https://cwiki.apache.org/confluence/display/Hive/StatisticsAndDataMining)(这里的解释更易懂) |
| string | concat\_ws(string SEP, string A, string B...) | Like concat() above, but with custom separator SEP..  与concat()类似，但使用指定的分隔符喜进行分隔 |
| string | concat\_ws(string SEP, array<string>) | Like concat\_ws() above, but taking an array of strings. (as of Hive [0.9.0](https://issues.apache.org/jira/browse/HIVE-2203)).  拼接Array中的元素并用指定分隔符进行分隔 |
| string | decode(binary bin, string charset) | Decodes the first argument into a String using the provided character set (one of 'US-ASCII', 'ISO-8859-1', 'UTF-8', 'UTF-16BE', 'UTF-16LE', 'UTF-16'). If either argument is null, the result will also be null. (As of Hive [0.12.0](https://issues.apache.org/jira/browse/HIVE-2482).).  使用指定的字符集charset将二进制值bin解码成字符串，支持的字符集有：'US-ASCII', 'ISO-8859-1', 'UTF-8', 'UTF-16BE', 'UTF-16LE', 'UTF-16'，如果任意输入参数为NULL都将返回NULL |
| binary | encode(string src, string charset) | Encodes the first argument into a BINARY using the provided character set (one of 'US-ASCII', 'ISO-8859-1', 'UTF-8', 'UTF-16BE', 'UTF-16LE', 'UTF-16'). If either argument is null, the result will also be null. (As of Hive [0.12.0](https://issues.apache.org/jira/browse/HIVE-2482).).  使用指定的字符集charset将字符串编码成二进制值，支持的字符集有：'US-ASCII', 'ISO-8859-1', 'UTF-8', 'UTF-16BE', 'UTF-16LE', 'UTF-16'，如果任一输入参数为NULL都将返回NULL |
| int | find\_in\_set(string str, string strList) | Returns the first occurance of str in strList where strList is a comma-delimited string. Returns null if either argument is null. Returns 0 if the first argument contains any commas. For example, find\_in\_set('ab', 'abc,b,ab,c,def') returns 3..  返回以逗号分隔的字符串中str出现的位置，如果参数str为逗号或查找失败将返回0，如果任一参数为NULL将返回NULL回 |
| string | format\_number(number x, int d) | Formats the number X to a format like '#,###,###.##', rounded to D decimal places, and returns the result as a string. If D is 0, the result has no decimal point or fractional part. (As of Hive [0.10.0](https://issues.apache.org/jira/browse/HIVE-2694); bug with float types fixed in [Hive 0.14.0](https://issues.apache.org/jira/browse/HIVE-7257), decimal type support added in [Hive 0.14.0](https://issues.apache.org/jira/browse/HIVE-7279)).  将数值X转换成"#,###,###.##"格式字符串，并保留d位小数，如果d为0，将进行四舍五入且不保留小数 |
| string | get\_json\_object(string json\_string, string path) | Extracts json object from a json string based on json path specified, and returns json string of the extracted json object. It will return null if the input json string is invalid. NOTE: The json path can only have the characters [0-9a-z\_], i.e., no upper-case or special characters. Also, the keys \*cannot start with numbers.\* This is due to restrictions on Hive column names..  从指定路径上的JSON字符串抽取出JSON对象，并返回这个对象的JSON格式，如果输入的JSON是非法的将返回NULL,注意此路径上JSON字符串只能由数字 字母 下划线组成且不能有大写字母和特殊字符，且key不能由数字开头，这是由于Hive对列名的限制 |
| boolean | in\_file(string str, string filename) | Returns true if the string str appears as an entire line in filename..  如果文件名为filename的文件中有一行数据与字符串str匹配成功就返回true |
| int | instr(string str, string substr) | Returns the position of the first occurrence of substr in str. Returns null if either of the arguments are null and returns 0 if substr could not be found in str. Be aware that this is not zero based. The first character in str has index 1..  查找字符串str中子字符串substr出现的位置，如果查找失败将返回0，如果任一参数为Null将返回null，注意位置为从1开始的 |
| int | length(string A) | Returns the length of the string..  返回字符串的长度 |
| int | locate(string substr, string str[, int pos]) | Returns the position of the first occurrence of substr in str after position pos..  查找字符串str中的pos位置后字符串substr第一次出现的位置 |
| string | lower(string A) lcase(string A) | Returns the string resulting from converting all characters of B to lower case. For example, lower('fOoBaR') results in 'foobar'..  将字符串A的所有字母转换成小写字母 |
| string | lpad(string str, int len, string pad) | Returns str, left-padded with pad to a length of len..  从左边开始对字符串str使用字符串pad填充，最终len长度为止，如果字符串str本身长度比len大的话，将去掉多余的部分 |
| string | ltrim(string A) | Returns the string resulting from trimming spaces from the beginning(left hand side) of A. For example, ltrim(' foobar ') results in 'foobar '..  去掉字符串A前面的空格 |
| array<struct<string,double>> | ngrams(array<array<string>>, int N, int K, int pf) | Returns the top-k N-grams from a set of tokenized sentences, such as those returned by the sentences() UDAF. See [StatisticsAndDataMining](https://cwiki.apache.org/confluence/display/Hive/StatisticsAndDataMining) for more information..  返回出现次数TOP K的的子序列,n表示子序列的长度，具体看[StatisticsAndDataMining](https://cwiki.apache.org/confluence/display/Hive/StatisticsAndDataMining) (这里的解释更易懂) |
| string | parse\_url(string urlString, string partToExtract [, string keyToExtract]) | Returns the specified part from the URL. Valid values for partToExtract include HOST, PATH, QUERY, REF, PROTOCOL, AUTHORITY, FILE, and USERINFO. For example, parse\_url('http://facebook.com/path1/p.php?k1=v1&k2=v2#Ref1', 'HOST') returns 'facebook.com'. Also a value of a particular key in QUERY can be extracted by providing the key as the third argument, for example, parse\_url('http://facebook.com/path1/p.php?k1=v1&k2=v2#Ref1', 'QUERY', 'k1') returns 'v1'..  返回从URL中抽取指定部分的内容，参数url是URL字符串，而参数partToExtract是要抽取的部分，这个参数包含(HOST, PATH, QUERY, REF, PROTOCOL, AUTHORITY, FILE, and USERINFO,例如：parse\_url('http://facebook.com/path1/p.php?k1=v1&k2=v2#Ref1', 'HOST') ='facebook.com'，如果参数partToExtract值为QUERY则必须指定第三个参数key  如：parse\_url('http://facebook.com/path1/p.php?k1=v1&k2=v2#Ref1', 'QUERY', 'k1') =‘v1’ |
| string | printf(String format, Obj... args) | Returns the input formatted according do printf-style format strings (as of Hive[0.9.0](https://issues.apache.org/jira/browse/HIVE-2695))..  按照printf风格格式输出字符串 |
| string | regexp\_extract(string subject, string pattern, int index) | Returns the string extracted using the pattern. For example, regexp\_extract('foothebar', 'foo(.\*?)(bar)', 2) returns 'bar.' Note that some care is necessary in using predefined character classes: using '\s' as the second argument will match the letter s; '\\s' is necessary to match whitespace, etc. The 'index' parameter is the Java regex Matcher group() method index. See docs/api/java/util/regex/Matcher.html for more information on the 'index' or Java regex group() method..  抽取字符串subject中符合正则表达式pattern的第index个部分的子字符串，注意些预定义字符的使用，如第二个参数如果使用'\s'将被匹配到s,'\\s'才是匹配空格 |
| string | regexp\_replace(string INITIAL\_STRING, string PATTERN, string REPLACEMENT) | Returns the string resulting from replacing all substrings in INITIAL\_STRING that match the java regular expression syntax defined in PATTERN with instances of REPLACEMENT. For example, regexp\_replace("foobar", "oo|ar", "") returns 'fb.' Note that some care is necessary in using predefined character classes: using '\s' as the second argument will match the letter s; '\\s' is necessary to match whitespace, etc..  按照Java正则表达式PATTERN将字符串INTIAL\_STRING中符合条件的部分成REPLACEMENT所指定的字符串，如里REPLACEMENT这空的话，抽符合正则的部分将被去掉  如：regexp\_replace("foobar", "oo|ar", "") = 'fb.' 注意些预定义字符的使用，如第二个参数如果使用'\s'将被匹配到s,'\\s'才是匹配空格 |
| string | repeat(string str, int n) | Repeats str n times..  重复输出n次字符串str |
| string | reverse(string A) | Returns the reversed string..  反转字符串 |
| string | rpad(string str, int len, string pad) | Returns str, right-padded with pad to a length of len..  从右边开始对字符串str使用字符串pad填充，最终len长度为止，如果字符串str本身长度比len大的话，将去掉多余的部分 |
| string | rtrim(string A) | Returns the string resulting from trimming spaces from the end(right hand side) of A. For example, rtrim(' foobar ') results in ' foobar'..  去掉字符串后面出现的空格 |
| array<array<string>> | sentences(string str, string lang, string locale) | Tokenizes a string of natural language text into words and sentences, where each sentence is broken at the appropriate sentence boundary and returned as an array of words. The 'lang' and 'locale' are optional arguments. For example, sentences('Hello there! How are you?') returns ( ("Hello", "there"), ("How", "are", "you") )..  字符串str将被转换成单词数组，如：sentences('Hello there! How are you?') =( ("Hello", "there"), ("How", "are", "you") ) |
| string | space(int n) | Returns a string of n spaces..  返回n个空格 |
| array | split(string str, string pat) | Splits str around pat (pat is a regular expression)..  按照正则表达式pat来分割字符串str,并将分割后的数组字符串的形式返回 |
| map<string,string> | str\_to\_map(text[, delimiter1, delimiter2]) | Splits text into key-value pairs using two delimiters. Delimiter1 separates text into K-V pairs, and Delimiter2 splits each K-V pair. Default delimiters are ',' for delimiter1 and '=' for delimiter2..  将字符串str按照指定分隔符转换成Map，第一个参数是需要转换字符串，第二个参数是键值对之间的分隔符，默认为逗号;第三个参数是键值之间的分隔符，默认为"=" |
| string | substr(string|binary A, int start) substring(string|binary A, int start) | Returns the substring or slice of the byte array of A starting from start position till the end of string A. For example, substr('foobar', 4) results in 'bar' (see [<http://dev.mysql.com/doc/refman/5.0/en/string-functions.html#function_substr>])..  对于字符串A,从start位置开始截取字符串并返回 |
| string | substr(string|binary A, int start, int len) substring(string|binary A, int start, int len) | Returns the substring or slice of the byte array of A starting from start position with length len. For example, substr('foobar', 4, 1) results in 'b' (see [<http://dev.mysql.com/doc/refman/5.0/en/string-functions.html#function_substr>])..  对于二进制/字符串A,从start位置开始截取长度为length的字符串并返回 |
| string | substring\_index(string A, string delim, int count) | Returns the substring from string A before count occurrences of the delimiter delim (as of Hive [1.3.0](https://issues.apache.org/jira/browse/HIVE-686)). If count is positive, everything to the left of the final delimiter (counting from the left) is returned. If count is negative, everything to the right of the final delimiter (counting from the right) is returned. Substring\_index performs a case-sensitive match when searching for delim. Example: substring\_index('www.apache.org', '.', 2) = 'www.apache'..  截取第count分隔符之前的字符串，如count为正则从左边开始截取，如果为负则从右边开始截取 |
| string | translate(string|char|varchar input, string|char|varchar from, string|char|varchar to) | Translates the input string by replacing the characters present in the from string with the corresponding characters in the to string. This is similar to the translatefunction in [PostgreSQL](http://www.postgresql.org/docs/9.1/interactive/functions-string.html). If any of the parameters to this UDF are NULL, the result is NULL as well. (Available as of Hive [0.10.0](https://issues.apache.org/jira/browse/HIVE-2418), for string types)  Char/varchar support added as of [Hive 0.14.0](https://issues.apache.org/jira/browse/HIVE-6622)..  将input出现在from中的字符串替换成to中的字符串 如：translate("MOBIN","BIN","M")="MOM" |
| string | trim(string A) | Returns the string resulting from trimming spaces from both ends of A. For example, trim(' foobar ') results in 'foobar'.  将字符串A前后出现的空格去掉 |
| binary | unbase64(string str) | Converts the argument from a base 64 string to BINARY. (As of Hive [0.12.0](https://issues.apache.org/jira/browse/HIVE-2482).).  将64位的字符串转换二进制值 |
| string | upper(string A) ucase(string A) | Returns the string resulting from converting all characters of A to upper case. For example, upper('fOoBaR') results in 'FOOBAR'..  将字符串A中的字母转换成大写字母 |
| string | initcap(string A) | Returns string, with the first letter of each word in uppercase, all other letters in lowercase. Words are delimited by whitespace. (As of Hive [1.1.0](https://issues.apache.org/jira/browse/HIVE-3405).).  将字符串A转换第一个字母大写其余字母的字符串 |
| int | levenshtein(string A, string B) | Returns the Levenshtein distance between two strings (as of Hive [1.2.0](https://issues.apache.org/jira/browse/HIVE-9556)). For example, levenshtein('kitten', 'sitting') results in 3..  计算两个字符串之间的差异大小  如：levenshtein('kitten', 'sitting') = 3 |
| string | soundex(string A) | Returns soundex code of the string (as of Hive [1.2.0](https://issues.apache.org/jira/browse/HIVE-9738)). For example, soundex('Miller') results in M460..  将普通字符串转换成soundex字符串 |

聚合函数

| Return Type | **Name(Signature)** | **Description** |
| --- | --- | --- |
| BIGINT | count(\*), count(expr), count(DISTINCT expr[, expr...]) | count(\*) - Returns the total number of retrieved rows, including rows containing NULL values.  统计总行数，包括含有NULL值的行  count(expr) - Returns the number of rows for which the supplied expression is non-NULL.  统计提供非NULL的expr表达式值的行数  count(DISTINCT expr[, expr]) - Returns the number of rows for which the supplied expression(s) are unique and non-NULL. Execution of this can be optimized with [hive.optimize.distinct.rewrite](https://cwiki.apache.org/confluence/display/Hive/Configuration+Properties#ConfigurationProperties-hive.optimize.distinct.rewrite).  统计提供非NULL且去重后的expr表达式值的行数 |
| DOUBLE | sum(col), sum(DISTINCT col) | Returns the sum of the elements in the group or the sum of the distinct values of the column in the group.  sum(col),表示求指定列的和，sum(DISTINCT col)表示求去重后的列的和 |
| DOUBLE | avg(col), avg(DISTINCT col) | Returns the average of the elements in the group or the average of the distinct values of the column in the group.  avg(col),表示求指定列的平均值，avg(DISTINCT col)表示求去重后的列的平均值 |
| DOUBLE | min(col) | Returns the minimum of the column in the group.  求指定列的最小值 |
| DOUBLE | max(col) | Returns the maximum value of the column in the group.  求指定列的最大值 |
| DOUBLE | variance(col), var\_pop(col) | Returns the variance of a numeric column in the group.  求指定列数值的方差 |
| DOUBLE | var\_samp(col) | Returns the unbiased sample variance of a numeric column in the group.  求指定列数值的样本方差 |
| DOUBLE | stddev\_pop(col) | Returns the standard deviation of a numeric column in the group.  求指定列数值的标准偏差 |
| DOUBLE | stddev\_samp(col) | Returns the unbiased sample standard deviation of a numeric column in the group.  求指定列数值的样本标准偏差 |
| DOUBLE | covar\_pop(col1, col2) | Returns the population covariance of a pair of numeric columns in the group.  求指定列数值的协方差 |
| DOUBLE | covar\_samp(col1, col2) | Returns the sample covariance of a pair of a numeric columns in the group.  求指定列数值的样本协方差 |
| DOUBLE | corr(col1, col2) | Returns the Pearson coefficient of correlation of a pair of a numeric columns in the group.  返回两列数值的相关系数 |
| DOUBLE | percentile(BIGINT col, p) | Returns the exact pth percentile of a column in the group (does not work with floating point types). p must be between 0 and 1. NOTE: A true percentile can only be computed for integer values. Use PERCENTILE\_APPROX if your input is non-integral.  返回col的p%分位数 |

表生成函数

| Return Type | **Name(Signature)** | **Description** |
| --- | --- | --- |
| Array Type | explode(array<TYPE> a) | For each element in a, generates a row containing that element.  对于a中的每个元素，将生成一行且包含该元素 |
| N rows | explode(ARRAY) | Returns one row for each element from the array..  每行对应数组中的一个元素 |
| N rows | explode(MAP) | Returns one row for each key-value pair from the input map with two columns in each row: one for the key and another for the value. (As of Hive [0.8.0](https://issues.apache.org/jira/browse/HIVE-1735).).  每行对应每个map键-值，其中一个字段是map的键，另一个字段是map的值 |
| N rows | posexplode(ARRAY) | Behaves like explode for arrays, but includes the position of items in the original array by returning a tuple of (pos, value). (As of [Hive 0.13.0](https://issues.apache.org/jira/browse/HIVE-4943).).  与explode类似，不同的是还返回各元素在数组中的位置 |
| N rows | stack(INT n, v\_1, v\_2, ..., v\_k) | Breaks up v\_1, ..., v\_k into n rows. Each row will have k/n columns. n must be constant..  把M列转换成N行，每行有M/N个字段，其中n必须是个常数 |
| tuple | json\_tuple(jsonStr, k1, k2, ...) | Takes a set of names (keys) and a JSON string, and returns a tuple of values. This is a more efficient version of the get\_json\_object UDF because it can get multiple keys with just one call..  从一个JSON字符串中获取多个键并作为一个元组返回，与get\_json\_object不同的是此函数能一次获取多个键值 |
| tuple | parse\_url\_tuple(url, p1, p2, ...) | This is similar to the parse\_url() UDF but can extract multiple parts at once out of a URL. Valid part names are: HOST, PATH, QUERY, REF, PROTOCOL, AUTHORITY, FILE, USERINFO, QUERY:<KEY>..  返回从URL中抽取指定N部分的内容，参数url是URL字符串，而参数p1,p2,....是要抽取的部分，这个参数包含HOST, PATH, QUERY, REF, PROTOCOL, AUTHORITY, FILE, USERINFO, QUERY:<KEY> |
|  | inline(ARRAY<STRUCT[,STRUCT]>) | Explodes an array of structs into a table. (As of Hive [0.10](https://issues.apache.org/jira/browse/HIVE-3238).).  将结构体数组提取出来并插入到表中 |