

# Hive内置函数速查表

## 关系运算符

名称	语法	描述
=	A=B	等值比较：如果表达式A与表达式B相等，则为TRUE；否则为FALSE
<>	A <> B	不等值比较：如果表达式A为NULL，或者表达式B为NULL，返回NULL；如果表达式A与表达式B不相等，则为TRUE；否则为FALSE
<	A < B	小于比较：如果表达式A为NULL，或者表达式B为NULL，返回NULL；如果表达式A小于表达式B，则为TRUE；否则为FALSE
<=	A <= B	小于等于比较：如果表达式A为NULL，或者表达式B为NULL，返回NULL；如果表达式A小于或者等于表达式B，则为TRUE；否则为FALSE
>	A > B	大于比较：如果表达式A为NULL，或者表达式B为NULL，返回NULL；如果表达式A大于表达式B，则为TRUE；否则为FALSE
>=	A >= B	大于等于比较：如果表达式A为NULL，或者表达式B为NULL，返回NULL；如果表达式A大于或者等于表达式B，则为TRUE；否则为FALSE
IS NULL	A IS NULL	空值判断: 如果表达式A的值为NULL，则为TRUE；否则为FALSE
IS NOT NULL	A IS NOT NULL	非空判断：如果表达式A的值为NULL，则为FALSE；否则为TRUE
LIKE	A LIKE B	LIKE比较：如果字符串A或者字符串B为NULL，则返回NULL；如果字符串A符合表达式B 的正则语法，则为TRUE；否则为FALSE
RLIKE	A RLIKE B	RLIKE比较如果字符串A或者字符串B为NULL，则返回NULL；如果字符串A符合JAVA正则表达式B的正则语法，则为TRUE；否则为FALSE。

名称	语法	描述
REGEXP	A REGEXP B	功能与RLIKE相同

## 数学运算符

名称	语法	描述
+	A + B	相加：返回A与B相加的结果
-	A - B	相减：返回A与B相减的结果
*	A * B	相乘：返回A与B相乘的结果
/	A / B	相除：返回A除以B的结果。结果的数值类型为double
%	A % B	取余：返回A除以B的余数
&	A & B	位与操作：返回A和B按位进行与操作的结果
	A   B	位或操作：返回A和B按位进行或操作的结果
^	A ^ B	位异或操作：返回A和B按位进行异或操作的结果
~	A ~ B	位取反操作：返回A按位取反操作的结果

## 逻辑运算符

名称	语法	描述
AND	A AND B	逻辑与操作：如果A和B均为TRUE，则为TRUE；否则为FALSE。如果A为NULL或B为NULL，则为NULL
OR	A OR B	逻辑或操作：如果A为TRUE，或者B为TRUE，或者A和B均为TRUE，则为TRUE；否则为FALSE
NOT	NOT A	逻辑非操作：如果A为FALSE，或者A为NULL，则为TRUE；否则为FALSE

## 数值计算

名称	语法	描述
round	round(double a)	取整函数：返回double类型的整数值部分（遵循四舍五入）
floor	floor(double a)	向下取整函数：返回等于或者小于该double变量的最大的整数
ceil	ceil(double a)	向上取整函数：返回等于或者大于该double变量的最小的整数
rand	rand()	取随机数函数：返回一个0到1范围内的随机Double数值
exp	exp(double a)	自然指数函数：返回自然对数e的a次方
log10	log10(double a)	以10为底对数函数：返回以10为底的a的对数
log2	log2(double a)	以2为底对数函数：返回以2为底的a的对数
log	log(double base, double a)	对数函数：返回以base为底的a的对数
pow	pow(double a, double p)	幂运算函数：返回a的p次幂
power	power(double a, double p)	幂运算函数：返回a的p次幂,与pow功能相同
sqrt	sqrt(double a)	开平方函数：返回a的平方根
bin	bin(BIGINT a)	二进制函数：返回a的二进制代码表示
hex	hex(BIGINT a)	十六进制函数：如果变量是int类型，那么返回a的十六进制表示；如果变量是string类型，则返回该字符串的十六进制表示
unhex	unhex(string a)	反转十六进制函数：返回该十六进制字符串所代码的字符串

名称	语法	描述
conv	conv(BIGINT num, int from_base, int to_base)	进制转换函数：将数值num从from_base进制转化到to_base进制
abs	abs(double a) abs(int a)	绝对值函数：返回数值a的绝对值
pmod	pmod(int a, int b)	正取余函数：返回正的a除以b的余数
sin	sin(double a)	正弦函数：返回a的正弦值
asin	asin(double a)	反正弦函数：返回a的反正弦值
cos	cos(double a)	余弦函数：返回a的余弦值
acos	acos(double a)	反余弦函数：返回a的反余弦值

## 日期函数

名称	语法	描述
from_unixtime	from_unixtime(bigint unixtime[, string format])	UNIX时间戳转日期函数：转化UNIX时间戳（从1970-01-01 00:00:00 UTC到指定时间的秒数）到当前时区的时间格式
unix_timestamp	unix_timestamp()	获取当前UNIX时间戳函数：获得当前时区的UNIX时间戳
unix_timestamp	unix_timestamp(string date)	日期转UNIX时间戳函数：转换格式为"yyyy-MM-dd HH:mm:ss"的日期到UNIX时间戳，如果转化失败，则返回0。
unix_timestamp	unix_timestamp(string date, string pattern)	指定格式日期转UNIX时间戳函数：转换pattern格式的日期到UNIX时间戳。如果转化失败，则返回0。
to_date	to_date(string timestamp)	日期时间转日期函数：返回日期时间字段中的日期部分。
year	year(string date)	日期转年函数：返回日期中的年。
month	month (string date)	日期转月函数：返回日期中的月份。
day	day(string date)	日期转日函数：返回日期中的天。
hour	hour(string date)	日期转小时函数：返回日期中的小时。
minute	minute(string date)	日期转分钟函数：返回日期中的分钟。
second	second(string date)	日期转秒函数：返回日期中的秒。

名称	语法	描述
weekofyear	weekofyear (string date)	日期转周函数：返回日期在当前的周数。
datediff	datediff(string enddate, string startdate)	日期比较函数：返回结束日期减去开始日期的天数。
date_add	date_add(string startdate, int days)	日期增加函数：返回开始日期startdate增加days天后的日期。
date_sub	date_sub (string startdate, int days)	日期减少函数：返回开始日期startdate减少days天后的日期。

## 条件函数



名称	语法	描述
if	if(boolean testCondition, T valueTrue, T valueFalseOrNull)	If函数：当条件testCondition为TRUE时，返回valueTrue；否则返回valueFalseOrNull
COALESCE	COALESCE(T v1, T v2, ...)	非空查找函数：返回参数中的第一个非空值；如果所有值都为NULL，那么返回NULL
CASE	CASE a WHEN b THEN c [WHEN d THEN e]* [ELSE f] END	条件判断函数：如果a等于b，那么返回c；如果a等于d，那么返回e；否则返回f
CASE	CASE WHEN a THEN b [WHEN c THEN d]* [ELSE e] END	条件判断函数：如果a为TRUE,则返回b；如果c为TRUE，则返回d；否则返回e

## 字符串函数

名称	语法	描述
length	length(string A)	字符串长度函数：返回字符串A的长度
reverse	reverse(string A)	字符串反转函数：返回字符串A的反转结果
concat	concat(string A, string B...)	字符串连接函数：返回输入字符串连接后的结果，支持任意个输入字符串
concat_ws	concat_ws(string SEP, string A, string B...)	带分隔符字符串连接函数：返回输入字符串连接后的结果，SEP表示各个字符串间的分隔符
substr	substr(string A, int start)	字符串截取函数：返回字符串A从start位置到结尾的字符串
substring	substring(string A, int start)	字符串截取函数：返回字符串A从start位置到结尾的字符串
upper	upper(string A)	字符串转大写函数：返回字符串A的大写格式
ucase	ucase(string A)	字符串转大写函数：返回字符串A的大写格式
lower	lower(string A)	字符串转小写函数：返回字符串A的小写格式
lcase	lcase (string A)	字符串转小写函数：返回字符串A的小写格式
trim	trim(string A)	去空格函数：去除字符串两边的空格
ltrim	ltrim(string A)	左边去空格函数：去除字符串左边的空格

名称	语法	描述
rtrim	rtrim(string A)	右边去空格函数：去除字符串右边的空格
regexp_replace	regexp_replace(string A, string B, string C)	正则表达式替换函数：将字符串A中的符合java正则表达式B的部分替换为C
regexp_extract	regexp_extract(string subject, string pattern, int index)	正则表达式解析函数：将字符串subject按照pattern正则表达式的规则拆分，返回index指定的字符
parse_url	parse_url(string urlString, string partToExtract [, string keyToExtract])	URL解析函数：返回URL中指定的部分。partToExtract的有效值为：HOST, PATH, QUERY, REF, PROTOCOL, AUTHORITY, FILE, and USERINFO
get_json_object	get_json_object(string json_string, string path)	json解析函数：解析json的字符串json_string,返回path指定的内容。如果输入的json字符串无效，那么返回NULL
space	space(int n)	空格字符串函数：返回长度为n的字符串
repeat	repeat(string str, int n)	重复字符串函数：返回重复n次后的str字符串
ascii	ascii(string str)	首字符ascii函数：返回字符串str第一个字符的ascii码
lpad	lpad(string str, int len, string pad)	左补足函数：将str进行用pad进行左补足到len位
rpadd	rpadd(string str, int len, string pad)	右补足函数：将str进行用pad进行右补足到len位

名称	语法	描述
split	split(string str, string pat)	分割字符串函数：按照pat字符串分割str，会返回分割后的字符串数组
find_in_set	find_in_set(string str, string strList)	集合查找函数：返回str在strlist第一次出现的位置，strlist是用逗号分割的字符串。如果没有找该str字符，则返回0

## 聚合函数

名称	语法	描述
count	count(*), count(expr), count(DISTINCT expr[, expr_.])	个数统计函数： count(*)统计检索出的 行的个数，包括NULL 值的行；count(expr) 返回指定字段的非空值 的个数； count(DISTINCT expr[, expr_.])返回指 定字段的不同的非空值 的个数
sum	sum(col), sum(DISTINCT col)	总和统计函数： sum(col)统计结果集中 col的相加的结果； sum(DISTINCT col)统 计结果中col不同值相 加的结果
avg	avg(col), avg(DISTINCT col)	平均值统计函数： avg(col)统计结果集中 col的平均值； avg(DISTINCT col)统 计结果中col不同值相 加的平均值
min	min(col)	最小值统计函数：统计 结果集中col字段的最 小值
max	max(col)	最大值统计函数：统计 结果集中col字段的最 大值

名称	语法	描述
var_pop	var_pop(col)	非空集合总体变量函数：统计结果集中col非空集合的总体变量（忽略null）
var_samp	var_samp (col)	非空集合样本变量函数：统计结果集中col非空集合的样本变量（忽略null）
stddev_pop	stddev_pop(col)	总体标准偏离函数：该函数计算总体标准偏离，并返回总体变量的平方根，其返回值与VAR_POP函数的平方根相同
stddev_samp	stddev_samp (col)	样本标准偏离函数：该函数计算样本标准偏离
percentile	percentile(BIGINT col, p)	中位数函数：求准确的第pth个百分位数，p必须介于0和1之间，但是col字段目前只支持整数，不支持浮点数类型

名称	语法	描述
percentile_approx	percentile_approx(DOUBLE col, p [, B])	近似中位数函数：求近似的第pth个百分位数，p必须介于0和1之间，返回类型为double，但是col字段支持浮点类型。参数B控制内存消耗的近似精度，B越大，结果的准确度越高。默认为10,000。当col字段中的distinct值的个数小于B时，结果为准确的百分位数
histogram_numeric	histogram_numeric(col, b)	直方图：以b为基准计算col的直方图信息

## 复合类型构造函数

名称	语法	描述
map	map (key1, value1, key2, value2, ...)	Map类型构建：根据输入的key和value对构建map类型
struct	struct(val1, val2, val3, ...)	Struct类型构建：根据输入的参数构建结构体struct类型
array	array(val1, val2, ...)	Array类型构建：根据输入的参数构建数组array类型

## 复合类型操作函数

名称	语法	描述
map	M[key]	map类型访问：返回map类型M中，key值为指定值的value值
struct	S.x	struct类型访问：返回结构体S中的x字段
array	A[n]	array类型访问：返回数组A中的第n个变量值

## 高级函数

名称	语法	描述
explode	explode(Array/Map)	将array或者map中每个元素单独生成一行，就是内置表生成函数（UDTF），UDTF将为每个输入行生成零个或多个输出行。
lateral view	lateral view udtf(expression) tableAlias as columnAlias (,columnAlias)*	lateral view其实就是用来和想类似explode这种UDTF函数联用的，lateral view会将UDTF生成的结果放到一个虚拟表中，然后这个虚拟表会和输入行进行join来达到连接UDTF外的select字段的目的。
reflect	reflect(class,method [,arg1 [,arg2..]])	reflect函数可以支持在sql中调用java中的自带函数，秒杀一切udf函数

## 窗口函数和分析函数



名称	语法	描述
OVER	OVER()	指定分析函数工作的数据窗口大小，这个数据窗口大小可能会随着行的变化而变化
CURRENT ROW	CURRENT ROW()	表示当前行
PRECEDING	n PRECEDING	表示往前n行数据
FOLLOWING	n FOLLOWING	表示往后n行数据
UNBOUNDED	UNBOUNDED	表示起点
UNBOUNDED PRECEDING	UNBOUNDED PRECEDING	表示从前面的起点
UNBOUNDED FOLLOWING	UNBOUNDED FOLLOWING	表示到后面的终点
LAG	LAG(col,n)	表示往前第n行数据
LEAD	LEAD(col,n)	表示往后第n行数据
RANK()	RANK()	表示返回数据项在分组中的排名，排名相等会在名次中留下空位
DENSE_RANK()	DENSE_RANK()	表示返回数据项在分组中的排名，排名相等会在名次中不会留下空位
NTILE()	NTILE(int n)	把有序分区中的行分发到指定数据的组中，各个组有编号，编号从1开始，对于每一行，NTILE返回此行所属的组的编号
ROW_NUMBER()	ROW_NUMBER()	为每条记录返回一个数字