

Hive-宏(macro)的使用

描述

在编写HQL的过程，很多逻辑需要反复的使用。这是我们可以通过宏对这段重复使用的逻辑进行提炼。

宏的作用相当于编程过程中的函数，将一些常用的语句进行封装便于后续的重复多次使用。

显而易见，开发过程比UDF更加简单快捷。

同时UDF是Java编写的，代码中的堆变量内存回收不受开发者控制，同时UDF还是嵌套在HQL中执行的，对于规模较大的表，比较容易造成OOM。

Hive-0.12.0及以后版本支持宏的使用。

语法

创建宏

create temporary macro [if not exists] macro_name([col_name col_type, ...]) expression;

宏的参数可以为若干个，宏的返回值为expression的返回值。

```
create temporary macro fixed_number() 42; --42
create temporary macro if not exists simple_add(int x, int y) x + y;
create temporary macro nest_add(int x) x + fixed_number(); --macromacro
--macrocase-when
create temporary macro case_when()
case
  when ... then ...
  when ... then ...
  else ...
end;
```

删除宏

drop temporary macro [if exists] macro_name;

```
drop temporary macro fixed_number;
drop temporary if exists simple_add;
```

一些常用例子

空串和NULL相互转换

1.空串转NULL

业务需要：将空串""视为NULL，需要补足

使用场景：在使用nvl或者coalesce时，如果前一个参数为空串 ""，则无法取到后边的参数。

```
create temporary macro empty2null(string x) if(trim(x) = '', null, x);
coalesce(empty2null(x1),empty2null(x2),..., 'UNKNOWN');
```

2.NULL转空串

使用场景：当使用concat拼接两个字段时，只有其中一个为NULL，则输出也为NULL。

```
create temporary macro null2empty(x string) if(x is null, '', x);
concat(null2empty(str1), null2empty(str2));
```

3.判断NULL和空串

```
create temporary macro is_ne(x string) nvl(trim(x), '') = '';--NULLtruefalse trim(x)null''
```

4.数据倾斜解决方法

使用场景：当字段存在大量NULL值或者空串时，可能会引发数据倾斜问题，应该把key转化为随机字符串，使得该字段均匀分布到各个reduce中。

```
create temporary macro ne2rand(x string)
case
  when is_ne(x) then concat('hive',rand())
  else x
end;
```

有关日期的计算

1.月第一天

```
--
create temporary macro first_day(dt string) trunc(dt,'mm'); dtyyyy-dd-mm
--
create temporary macro first_day_last_month(dt string) trunc(add_months(dt, -1), 'mm');
```

2.月最后一天

```
--
create temporary macro last_day(dt string) last_day(dt);
--
create temporary macro last_day_last_month(dt string) last_day(add_months(dt, -1));
```

参考文献地址：

[宏的使用](#)

[宏的介绍-Hive官网](#)

[Hive中常被忽视的利器-宏](#)