

参考文档

<https://tonydong.blog.csdn.net/article/details/105855073?spm=1001.2014.3001.5502> -- 解析窗口函数基本

<https://tonydong.blog.csdn.net/article/details/107518005?spm=1001.2014.3001.5502> -- 高级实战

涉及到语句知识点

with recursive ... as 递归

窗口函数-- 关键字是over

从定义上来讲，窗口函数包含了一个OVER子句，用于指定数据分析的窗口：

```
1 window_function ( expression, ... ) OVER (
2     PARTITION BY ...
3     ORDER BY ...
4     frame_clause
5 )
```

其中，window_function 是窗口函数的名称；expression 是参数，有些函数不需要参数；OVER子句包含三个选项：分区（PARTITION BY）、排序（ORDER BY）以及窗口大小（frame_clause）。

PARTITION BY选项用于定义分区，作用类似于 GROUP BY 的分组。如果指定了分区选项，窗口函数将会分别针对每个分区单独进行分析；如果省略分区选项，所有的数据作为一个整体进行分析。

ORDER BY选项用于指定分区内的排序方式，通常用于数据的排名分析。

聚合窗口函数 MySQL Oracle SQL Server PostgreSQL SQLite

SUM() OVER() ✓ ✓ ✓ ✓ ✓

COUNT() OVER() ✓ ✓ ✓ ✓ ✓

AVG() OVER() ✓ ✓ ✓ ✓ ✓

MAX() OVER() ✓ ✓ ✓ ✓ ✓

MIN() OVER() ✓ ✓ ✓ ✓ ✓

窗口函数 描述 MySQL Oracle SQL Server PostgreSQL SQLite

ROW_NUMBER() 为分区中的每行数据分配一个从 1 开始的序列号。 ✓ ✓ ✓ ✓ ✓

RANK() 计算每行数据在分区中的名次，排名可能产生跳跃。 ✓ ✓ ✓ ✓ ✓

DENSE_RANK() 计算每行数据在分区中的名次，排名不会产生跳跃。 ✓ ✓ ✓ ✓ ✓

PERCENT_RANK() 计算每行数据在分区中的相对排名，取值为 $(\text{rank} - 1) / (\text{rows} - 1)$ 。 ✓ ✓ ✓
✓ ✓

CUME_DIST() 计算每行数据在分区内的累积分布，取值范围大于 0 且小于等于 1。 ✓ ✓ ✓ ✓
✓

NTILE() 将分区内的数据分为 N 等份，计算每行数据所在的位置。 ✓ ✓ ✓ ✓ ✓

FIRST_VALUE() 返回窗口内第一行对应的数据。 ✓ ✓ ✓ ✓ ✓

LAST_VALUE() 返回窗口内最后一行对应的数据。 ✓ ✓ ✓ ✓ ✓

LAG() 返回分区中在当前行之前第 N 行对应的数据。 ✓ ✓ ✓ ✓ ✓

LEAD() 返回分区中在当前行之后第 N 行对应的数据。 ✓ ✓ ✓ ✓ ✓

NTH_VALUE() 返回窗口内第 N 行对应的数据。 ✓ ✓ ✗ ✓ ✓