hive不支持临时视图(temporary view)，只支持永久视图

```
-- 生成连续的5行数字比如从1到5 把生成的集合爆炸
select explode(sequence(1,100));

-- stack 函数[生成一个集合]
select stack(2,1,2,3);

--扩展横向迭代计算的4种方法
--创建视图表方式1
drop view  tast1;
drop view view1;
drop view view2;
create or replace temporary view tast1 as select stack(3,1,2,3)as c1;
select * from tast1;

--创建视图表方式二
create  view tast1(c1) as values (1),(2),(3);
select * from tast1;

--1.用子查询
select c1,c2,c1*c2 as c3 from(
select c1,c1+2 as c2 from tast1);
--2.用with as
with t1 as ( select c1,c1+2 as c2 from tast1)
,t2 as ( select c1,c2,c1*c2 as c3 from t1)
select * from t2;
--3用普通视图（永久视图-只保存逻辑不保存数据）--永久视图不能查询临时视图--支持重构
create or replace view view1 as select c1,c1+2 as c2 from tast1;
create view view2 as select c1,c2,c1*c2 as c3 from view1;
select * from view2;
--4用临时视图
create temporary view t_view1 as select c1,c1+2 as c2 from tast1; -- 中间结果可以不保存
create temporary view t_view2 as select c1,c2,c1*c2 as c3 from t_view1;
select *from t_view2;
--5cache table 缓存表 --spark的语法，pycharm不支持，可以运行
cache table cache1 as select c1,c1+2 as c2 from tast1;
cache table cache2 as select c1,c2,c1*c2 as c3 from cache1;
select * from cache2;

```

```sql
--查看所有的视图
show views;
--查看所有的表，也会显示视图
show tables;

create or replace view test2(c1,c2,c3,c4) as values
(1,1,6,1),
(1,2,23,null),
(1,3,8,null),
(1,4,4,null),
(2,1,32,1),
(2,2,9,null),
(2,3,15,null),
(2,4,8,null);

select * from test2;

--需求：计算c4的逻辑：当c2=1，则c4=1；否则c4=（上一个c4+当前的c3）/2
-- 错误方案
select c1, c2, c3, (lag(c4) over (partition by c1 order by c2)+c3)/2 as c4
from test2;
--先计算c2=2的情况
set spark.sql.shuffle.partitionby=4;
create or replace temporary view test_2 as
select c1,c2,c3,if(c2=2,(lag(c4)over(partition by c1 order by c2)+c3)/2,c4) as c4 from
test2;
select *from test_2;
--再计算c2=3的情况
create or replace temporary view test_3 as
select c1, c2, c3,if(c2=3,(lag(c4)over(partition by c1 order by c2)+c3)/2,c4) as c4
from test_2;
select * from test_2;
--再计算c2=4的情况
create or replace temporary view test_4 as
select c1, c2, c3,if(c2=4,(lag(c4)over(partition by c1 order by c2)+c3)/2,c4) as c4
from test_3;
select * from test_4;

substring/substr、to_date、date_add、datediff、date_format、concat、replace

```

```
cast(数据 as 类型)
power(字段1，字段2)返回字段1的字段2次方

--可以传入两个值不用over
least取最小值(两个列的最小值)，传入一个数组,greatest 取最大值


删除一个分区的数据。
alter table insurance_app.policy_result drop partition (month='2021-07');

查看hive表的建表语句
show create table policy_result;
--查看分区情况
show partitions insurance_app.policy_result;
--删除分区的数据
alter table insurance_app.policy_result drop partition (month='2021-12');

--计算相差几个月
month_between(开始时间，结束时间)
ceil()取整数，不会四舍五入


--分组后把某个字段的值形成一个arr集合，必须group by
--取arr可以用下标 比如 arr[0],arr[1]
collect_list（字段）as arr

spark不支持limit x,y ; 只支持limit x;

left/semi/anti/join
left 左表为主表，全部显示，右边匹配不上则为null
left semi 左表为主表，只显示匹配的上的左表的数据，和join差不多就是没有右表
left anti 显示匹配不上的数据，则右表没有匹配上的数据都显示

regexp_extract_all（str,'(字符)'） 匹配对应字符形成一个数组比如[1,1,1,1]
size（数组）计算数组有多少个
```

grouping

```
grouping_id():基于分组则为0，不基于分组则为1，ID为group by 按顺序填写的分组字段 并不是grouping sets里面的分组计算的
```

```
比如：id最大31(0001 1111)  id=15(0000 1111)按yearinfo分组 id=7(0000 0111)按yearinfo和quarterinfo分组
group by yearinfo, quarterinfo, monthinfo, dayinfo, hourinfo --用的是这个按二进制算
  grouping sets (
  yearinfo,
( yearinfo, area),
( yearinfo, quarterinfo, area)

```