

Hive中grouping函数问题

```
1  -- Hive 2.3版本之前不支持分组的grouping函数，但支持GROUPING__ID函数
2  -- GROUPING__ID 的计算规则
3  -- step1: 将group by后的字段倒序组合：【cookieid,city,daystr】
4  -- step2: 将是否分组的结果构建0和1并合并得到结果
5      -- 注意：Hive2.3版本之前，1代表基于该维度，0代表不基于该维度
6      -- 情况1: 基于cookie, 基于city, 基于daystr      = 111 = 7
7      -- 情况2: 不基于cookie, 基于city, 基于daystr    = 011 = 3
8      -- 情况3: 不基于cookie, 基于city, 不基于daystr  = 010 = 2
9      -- 情况4: 不基于cookie, 不基于city, 基于daystr  = 001 = 1
```

- 举个栗子

```
1  select
2      daystr,
3      city,
4      cookieid,
5      count(*) as cnt,
6      GROUPING__ID
7  from test.t_cookie
8  group by daystr,city,cookieid
9  grouping sets (daystr,city,(daystr,city),
    (daystr,city,cookieid));
```

- 结果

daystr	city	cookieid	cnt	grouping__id
2015-03-10	<null>	<null>	4	1
2015-03-12	<null>	<null>	1	1
2015-04-12	<null>	<null>	2	1
2015-04-13	<null>	<null>	3	1
2015-04-15	<null>	<null>	2	1
2015-04-16	<null>	<null>	2	1
<null>	上海	<null>	9	2
<null>	北京	<null>	5	2
2015-03-10	上海	<null>	3	3
2015-03-10	北京	<null>	1	3
2015-03-12	上海	<null>	1	3
2015-04-12	上海	<null>	2	3
2015-04-13	上海	<null>	1	3
2015-04-13	北京	<null>	2	3
2015-04-15	上海	<null>	2	3
2015-04-16	北京	<null>	2	3
2015-03-10	上海	cookie1	1	7
2015-03-10	上海	cookie2	1	7
2015-03-10	上海	cookie3	1	7
2015-03-10	北京	cookie5	1	7
2015-03-12	上海	cookie7	1	7
2015-04-12	上海	cookie3	1	7
2015-04-12	上海	cookie5	1	7