1. HIVE 调试。针对复杂SQL,如果不确定哪里出错了，可以打开HIVE的调试模式：

hive -hiveconf hive.root.logger=DEBUG,console

2. 通过hivevar传递变量给HIVE时，要特别注意变量中的引号。例如，针对如下SQL：

SELECT item\_id

FROM app.app\_srd\_jwriter\_sku\_da

WHERE dt='${hivevar:dt}'

and source ${hivevar:SKU\_SOURCE\_WHERE}

hive -d SKU\_SOURCE\_WHERE=${SKU\_SOURCE\_WHERE}

如果令：

SKU\_SOURCE\_WHERE='="jdmarket"'

SQL将执行报错：

FAILED: ParseException line 27:26 character '<EOF>' not supported here

18/08/06 11:07:10 ERROR ql.Driver: FAILED: ParseException line 27:26 character '<EOF>' not supported here

而且很难查找原因。

通过开启调试模式，可以看到最终执行的SQL如下：

SELECT item\_id

FROM app.app\_srd\_jwriter\_sku\_da

WHERE dt='2018-08-01'

and source ="jdmarket

SKU\_SOURCE\_WHERE中的第二个引号丢失了！原因未知。

解决办法也很简单：

SKU\_SOURCE\_WHERE="='jdmarket'"

根源可能还是在于HIVE处理单引号和双引号有所不同，也未细究。

3.像array字段插入默认空值：

set hive.exec.dynamic.partition.mode=nonstrict;

insert overwrite table app.app\_srd\_jwriter\_ypzj\_material\_sentence2 partition(source)

select a.id

,a.sentence

,a.words

,a.third\_cate\_cd

,a.probability

,a.article\_id

,a.index article\_index

,**array(cast (null as float)) product\_words**

,**array(cast (null as float)) scene\_words**

,a.ppl

,b.embeddings

,a.source

from

(

select md5(sentence) id, \*

from app.app\_srd\_jwriter\_ypzj\_material\_sentence

) a

join(

select md5(sentence) id, \*

from tmp.app\_srd\_jwriter\_ypzj\_material\_sentence\_embeddings

) b

on( a.id = b.id)

;

4.分区重命名:

alter table app\_srd\_jwriter\_ypzj\_inventory\_desc\_new\_da partition(dt='2019-02-02', model='skipvector\_similarity\_v0.1') rename to partition(dt='2018-11-14', model='skipvector\_similarity\_v0.1');