支持事务的交表语句有改变

Hive1.2:

CREATE TABLE IF NOT EXISTS dc\_dw.fact\_fcv\_order\_detail\_d\_bak(  
  order\_detail\_no string comment'订单明细编码',  
  out\_time timestamp comment'销售时间',  
  mobile string comment'手机',  
  product\_code string comment'产品编码',  
  size\_no string comment'尺码',  
  data\_type string comment'数据类型 1:零售 2:优购',  
  data\_flag string comment'标示符1:插入 2:更新 3:删除 4:其他',  
  sort\_id int comment'排序ID',  
  update\_date date comment'更新日期',  
  dw\_update\_time timestamp comment'入库时间'  
) COMMENT '工厂可视定制化历史订单事实表' ROW FORMAT DELIMITED FIELDS TERMINATED BY '\001' STORED AS orc  
TBLPROPERTIES ('transactional'='true', 'orc.compress'='SNAPPY',   
'orc.create.index'='true','orc.stripe.size'='67108864');  
ALTER TABLE dc\_dw.fact\_fcv\_order\_detail\_d\_bak CLUSTERED BY(sort\_id) INTO 10 BUCKETS;

通过alter来添加CLUSTERED 以及CLUSTERED的数量

Hive2.1.1必须在建表的时候指定CLUSTERED 以及CLUSTERED的数量

CREATE TABLE IF NOT EXISTS dc\_dw.fact\_fcv\_order\_detail\_d\_bak(  
  order\_detail\_no string comment'订单明细编码',  
  out\_time timestamp comment'销售时间',  
  mobile string comment'手机',  
  product\_code string comment'产品编码',  
  size\_no string comment'尺码',  
  data\_type string comment'数据类型 1:零售 2:优购',  
  data\_flag string comment'标示符1:插入 2:更新 3:删除 4:其他',  
  sort\_id int comment'排序ID',  
  update\_date date comment'更新日期',  
  dw\_update\_time timestamp comment'入库时间'  
) COMMENT '工厂可视定制化历史订单事实表'

CLUSTERED BY(sort\_id) INTO 10 BUCKETS

ROW FORMAT DELIMITED FIELDS TERMINATED BY '\001' STORED AS orc  
TBLPROPERTIES ('transactional'='true', 'orc.compress'='SNAPPY',   
'orc.create.index'='true','orc.stripe.size'='67108864');

出现部分Sql语句运行保存的问题

Caused by: java.lang.RuntimeException: Map operator initialization failed

at org.apache.hadoop.hive.ql.exec.mr.ExecMapper.configure(ExecMapper.java:137)

... 22 more

Caused by: java.lang.NullPointerException

at org.apache.hadoop.hive.ql.exec.vector.VectorMapOperator.createAndInitPartitionContext(VectorMapOperator.java:365)

at org.apache.hadoop.hive.ql.exec.vector.VectorMapOperator.internalSetChildren(VectorMapOperator.java:535)

at org.apache.hadoop.hive.ql.exec.vector.VectorMapOperator.setChildren(VectorMapOperator.java:469)

at org.apache.hadoop.hive.ql.exec.mr.ExecMapper.configure(ExecMapper.java:106)

对于分区表，如果没有分区信息就会报错，如果没有数据就没有分区信息。

配置了set hive.vectorized.execution.enabled = true; set hive.vectorized.execution.reduce.enabled = true;

矢量查询(Vectorized query) 每次处理数据时会将1024行数据组成一个batch进行处理，而不是一行一行进行处理，这样能够显著提高执行速度。

开启矢量查询之后，对已分区表没有数据情况没有办法获取到分区的信息，所以报空指针

3.Hive Schema Version问题

如果元数据与Hive版本不同步升级，或用到Spark，因为Spark依赖的Hive默认是1.2.1。所以元数据也中的VERSION表， 会被改来改去，导致各种报错，例如：

MetaException(message:Hive Schema version 2.1.0 does not match metastore's schema version 1.2.0 Metastore is not upgraded or corrupt)

一种方法是设置

hive.metastore.schema.verification=false。

还有一种彻底的方法是把Hive启动时候检查Schema的功能屏蔽掉。

4 Hive新增保留字问题

Hive随着版本变迁，保留字越来越多。 保留字的详细情况可以参考：Hive各个版本的保留字

(https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DDL#LanguageManualDDL-Keywords,Non-reservedKeywordsandReservedKeywords)

保留字问题有以下两种：

新增预留字

新的版本Hive会预留更多保留字，如果SQL用到产生语法问题，可以通过参数设置 set hive.support.sql11.reserved.keywords=false来取消保留字校验， 达到不需要修改SQL，又能保证线上SQL稳定运行的兼容效果。

新增关键字

当SQL中含有Hive2.1中新增的关键字时候，只能修改SQL进行兼容了，需要将与关键字冲突的属性名用反引号包围。 新版本增加的关键字有：OFFSET、SUBQUERY、REWRITE、PRIMARY、FOREIGN、KEY、 REFERENCES、CONSTRAINT等。

5. hive.开头的参数如果不在Hive2.1预定义配置中（可能已经被移除），或者大小写不对，set语句会直接报错。

例如：set hive.auto.convert.JOIN=true，大小写不对，执行会报错。

解决方法：规范化语句，或者关闭强制校验参数

6. 建表时候不指定列的具体类型，在2.1会直接报错。

例如：

create table test\_table\_null\_no\_type as select null as id, "zhangsan" as name;

解决方法：规范化SQL，指定具体的列类型

7. 通过函数等运算产生的列，没有指定明确的别名，在2.1会偶尔报错。

例如：

select id, `\_c1` from (select id, get\_json\_object(deliver\_geojson,'$.features')

from tb) a

解决方法：规范化SQL，指定有业务含义的别名，不以下划线开头

8. 在使用windows函数的时候，当列名在多个表都存在的时候，不指定列所属表，在2.1会报错。

例如：

select row\_number() over(partition by user\_id order by a.log\_time desc) rank

from a join b on a.id = b.id;

其中user\_id在a表和b表中均存在。

解决方法：规范化SQL，指定列所属表名。

9. 在使用union all的时候，union起来的多个查询，含有orderBy、clusterBy、distributeBy、sortBy、limit语法在2.1会报错。

例如：

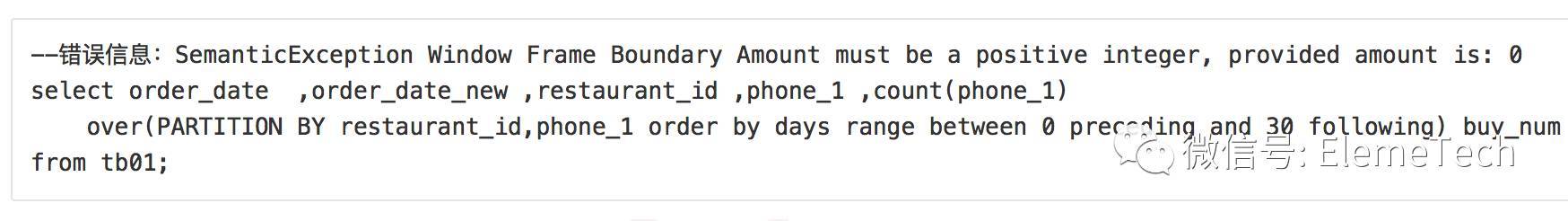
select 1 from default.dual limit 1

union all

select 2 from default.dual limit 1;

解决方法：只允许在最后一个语句中含有orderBy、clusterBy、distributeBy、sortBy、limit语句。

6. 使用windows函数时候，窗口的上限和下限必须大于0，否则在2.1会报错。



解决方法：规范化SQL，避免窗口函数上限和下限为0，例如将0 preceding修改为0 current row

7. case when语法中的数据类型不一致，在2.1会报错

解决方法：将case when里类型转换为一致。