MSBD5014 report

ZHAO Zhongyu
20753550

December 2020

1 Introduction

Continuous query evaluation on a database is a great challenge in many real time analytical systems. A large number of analytical queries are based on acyclic foreign-key joins, it is necessary to maintain the query results of any acyclic foreign-key joins under updates in short time.
This project implemented the algorithms[1] and build a program on top of Flink. It process update sequence as a data stream and give the query answers in real time as new data arrives as a stream. The code has been uploaded to github : https://github.com/waitzzy/MSBD5014$_AJU_FLINK/tree/master$.

2 Problem statement

The project is about TPC-H Q4 and figure 1 shows the sql query. It contains 2 tables: orders and lineitem. TPC-H data generator is used to generate *orders.tbl*and *lineitem.tbl*. Then we read them as two data streams which are the update sequences, then feed them to AJU flink program. The program should find orders which order dates between 1993-07-01 and 1997-10-01 and at least one of it's lineitems's commitdate should before receiptdate. Then it counts their order priorities respectively and print the order priorities and their counts as update information arrives in real time.

3 Algorithm description

First, we should build the foreign-key graph of TPC-H Q4 as the right part of figure 1. The parent relation in TPC-H Q4 is lineitem and child relation is orders, the foreign-key is *order_key*.
We use $L(R)$ to represent the set of live tuples in R and $N(R)$ to represent the set of non-live tuples in R. We also set some indexes for these two relations where $PK(R)$ represent the primary key of R and x_1 represent the primary key:

1. $I(L(R))$: An index on $L(R)$ where the primary key as the key.

```
====
---- QUERY: TPCH-Q4
# Q4 - Order Priority Checking Query
select
  o_orderpriority,
  count(*) as order_count
from
  orders
where
  o_orderdate >= '1993-07-01'
  and o_orderdate < '1993-10-01'
  and exists (
    select
      *
    from
      lineitem
    where
      l_orderkey = o_orderkey
      and l_commitdate < l_receiptdate
    )
group by
  o_orderpriority
order by
  o_orderpriority
```

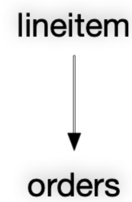


Figure 1: TPC-H Q4

2. $I(N(R))$: An index on $N(R)$ where the primary key as the key.

3. For a non-leaf R(lineitem), a counter $s(t)$ for each tuple $t \in R$ that is equal to the number of children of R on which t is alive. This counter is stored together with the tuple in $I(N(R))$.

4. $I(R, R_c)$,for non-leaf R and each child $R_c \in C(R)$:An index on $\pi_{x1,PK(R_c)}R$ using $PK(R_c)$ as the key.

When a new tuple t is inserted into relation R, we need to find out whether t is alive or not. If R is a leaf, t is alive by definition. Otherwise, we need to compute $s(t)$. To do so, for each child $R_c\ inC(R)$, we look up $I(L(Rc))$ with key $\pi_{PK(R_c)}t$. If there is a tuple in $I(L(Rc))$ with key $\pi_{PK(R_c)}t$,that means t is alive on R_c. After that, we will check the assertion keys by find whether there are more than one tuple in R_i that can join with t_j through any path. If it is not and $s(t)$ is equal to the number of children of R, it means t is alive and we add it to $I(L(R))$; otherwise we add it to $I(N(R))$. Meanwhile, we also need to update the index $I(R, Rc)$.
If t is not alive, then it will not affect other tuples' status nor the join result. If it is alive, then it will increase the $s(t_p)$ value for each tuple $t_p \in R_p$ that joins with t. These tuples can be found by probing the index $I(R, Rp)$. After that, we can do the insert update job of the corresponding sql query.
To delete a tuple t from R, we first check if it is alive. If it is non-alive, then we simply delete from $I(N(R))$. Otherwise, it will affect the status of tuples in R_p. More precisely, it will decrement the $s(t_p)$ value for every $t_p \in R_p$ that joins with t. If a t_p is currently alive, this will make it non-alive, which may trigger further updates recursively in a bottom-up fashion. When a live tuple in the root relation becomes non-alive or deleted, its corresponding join result will be deleted. Finally, we also need to update the index $I(R_p, R)$.

4 Some important implementation details

4.1 Data structures

We use *Relationunit* class to represent the orders relation and lineitem relation. It contains the indexes showed in section 3 which are implemented by hash table. It also contains the relation in foreign-key graph and assertion key. Those two classes are managed by $RelationsManager$. It is responsible for initialize the relations and their relationships. In TPC-H Q4, the root relation is lineitem and the leaf relation is orders. The lineitem is orders's parent. As figure 2 shows, they are not unconstrained pair, thus, there is no need to set assertion keys.

The original tuple in these two tables are stored in $OrdersTuple$ and $LineitemTuple$.

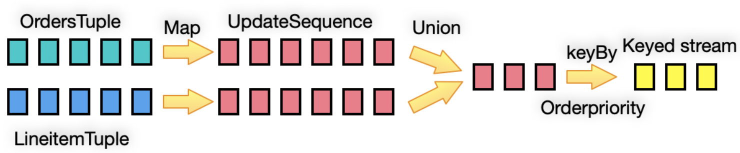


Figure 2: The data stream of TPC-H Q4

As figure 3 shows, They are two data streams. We convert them into $UpdateAction$ which contains the update information. Then we use can connect them as one update sequence data stream. Then we can use *orderpriority* to group the data stream and get the keyed stream

The result of TPC-H Q4 and $RelationManager$ are stored in $ValueState$ managed by Flink in a distributed and fault-tolerant fashion.

4.2 Algorithm

The main algorithm was implemented through $KeyedProcessFunction$ which can be thought of as a FlatMapFunction with access to keyed state. It handles events by being invoked for each event received in the input stream. Thus, when a new update tuple arrive, it will get access to the corresponding states, if the result state and relation state are null, it should be initialize first.

4.2.1 Insert update

We first put these tuple to the corresponding $RelationUnit$. If the relation is lineitem, we set the counter to be 0. Then we search the *orderkey* in orders relation to find whether there already exists a tuple which contains the same *orderkey*. The counter will be 1(equals to the number of lineitem in TPC-H Q4) and the new linetem tuple will be considered as live tuple if there exists. Since the order relation is the leaf relation and all tuples are alive. Thus, the order tuple and live lineitem tuple will be put into live tuples.

1. For lineitem tuple, we will first judge whether it's commitdate before receiptdate, then traverse all order tuples which have the same *orderkey* and statistic the priority of orders tuple which *orderdate* between 1993-07-01 and 1993-10-01 and add it to the result state respectively.

2. For order tuple, we will traverse all lineitem tuples and remove tuples which have same *orderkey* from non-live hash table and put it to live hash table. if order tuple and lineitem tuple satisfy the previous conditions, we add 1 to the corresponding priority counts. Since in TPC-H Q4 lineitem only has one child orders, so it is not needed to check the following influence.

If it is a non-live lineitem tuple, we put it to non-live tuple indexes directly. In the end, we update the state of relations and the state of TPC-H Q4 result.

4.2.2 Delete update

If the update action is delete. We first check if the tuple is alive. If it is non-alive, then we simply delete from non-live tuple indexes. Otherwise, it will affect some lineitem tuples. The delete live tuple should be removed from live tuples first. Then:

1. For lineitem tuples, we will traverse the order tuples which have same *orderkey* and substract them if they have been counted in the previous insert action.

2. For orders tuples, we should substract all the counters $s(t)$ by 1 for each corresponding lineitem tuple. Then remove these lineitem tuples from live tuple indexes to non-live tuple indexes. Since in TPC-H Q4 lineitem only has one child orders, so it is not needed to check the following influence.

In the end, we update the state of relations and the state of TPC-H Q4 result.

5 Experiment

5.1 Insert

We use TPC-H data generation generate some data as data source and figure 3 shows the update action data streams. There is only one pair satisfies the search conditions of TPC-H Q4 query marked with red line. It has the same order key 2536 and the order priority of the order tuple is 3-MEDIUM.

```
orderkey='2511', relationName='lineitem', tupleData=LineitemTuple{l_orderkey=2511, l_partkey=163035, l_suppkey=8068, l_linenumber=1, l_quantity=33.0, l_extendedprice=36234.99, l_discount=0.01, l_
orderkey='2511', relationName='lineitem', tupleData=LineitemTuple{l_orderkey=2511, l_partkey=113627, l_suppkey=6139, l_linenumber=2, l_quantity=50.0, l_extendedprice=82031.0, l_discount=0.1, l_ta
orderkey='2511', relationName='lineitem', tupleData=LineitemTuple{l_orderkey=2511, l_partkey=125399, l_suppkey=424, l_linenumber=3, l_quantity=30.0, l_extendedprice=42731.7, l_discount=0.09, l_ta
orderkey='2511', relationName='lineitem', tupleData=LineitemTuple{l_orderkey=2511, l_partkey=174154, l_suppkey=1706, l_linenumber=4, l_quantity=39.0, l_extendedprice=47897.85, l_discount=0.01, l_
orderkey='2511', relationName='lineitem', tupleData=LineitemTuple{l_orderkey=2511, l_partkey=5035, l_suppkey=5036, l_linenumber=5, l_quantity=46.0, l_extendedprice=43241.38, l_discount=0.06, l_ta
orderkey='2511', relationName='lineitem', tupleData=LineitemTuple{l_orderkey=2511, l_partkey=18301, l_suppkey=3304, l_linenumber=6, l_quantity=18.0, l_extendedprice=21947.4, l_discount=0.0, l_tax
orderkey='2536', relationName='lineitem', tupleData=LineitemTuple{l_orderkey=2536, l_partkey=154339, l_suppkey=6855, l_linenumber=1, l_quantity=30.0, l_extendedprice=41799.9, l_discount=0.01, l_t
orderkey='2537', relationName='lineitem', tupleData=LineitemTuple{l_orderkey=2537, l_partkey=160759, l_suppkey=760, l_linenumber=1, l_quantity=17.0, l_extendedprice=30935.75, l_discount=0.06, l_t
orderkey='2537', relationName='lineitem', tupleData=LineitemTuple{l_orderkey=2537, l_partkey=73114, l_suppkey=3115, l_linenumber=2, l_quantity=11.0, l_extendedprice=11958.21, l_discount=0.04, l_t
orderkey='2537', relationName='lineitem', tupleData=LineitemTuple{l_orderkey=2537, l_partkey=12914, l_suppkey=2915, l_linenumber=3, l_quantity=20.0, l_extendedprice=36538.2, l_discount=0.08, l_ta
orderkey='2543', relationName='orders', tupleData=OrdersTuple{o_orderkey=2543, o_custkey=39820, o_orderstatus='O', o_totalprice=150223.76, o_orderdate='1998-06-01', o_orderpriority='1-URGENT', o_
orderkey='2537', relationName='lineitem', tupleData=LineitemTuple{l_orderkey=2537, l_partkey=44356, l_suppkey=9365, l_linenumber=4, l_quantity=8.0, l_extendedprice=10402.8, l_discount=0.01, l_tax
orderkey='2568', relationName='orders', tupleData=OrdersTuple{o_orderkey=2568, o_custkey=76321, o_orderstatus='F', o_totalprice=142651.54, o_orderdate='1992-08-04', o_orderpriority='5-LOW', o_cle
orderkey='2538', relationName='lineitem', tupleData=LineitemTuple{l_orderkey=2538, l_partkey=109335, l_suppkey=6866, l_linenumber=1, l_quantity=21.0, l_extendedprice=28230.93, l_discount=0.03, l_
orderkey='2569', relationName='orders', tupleData=OrdersTuple{o_orderkey=2569, o_custkey=65491, o_orderstatus='O', o_totalprice=219822.18, o_orderdate='1996-03-25', o_orderpriority='4-NOT SPECIFI
orderkey='2538', relationName='lineitem', tupleData=LineitemTuple{l_orderkey=2538, l_partkey=192089, l_suppkey=4609, l_linenumber=2, l_quantity=11.0, l_extendedprice=12991.88, l_discount=0.09, l_
orderkey='2570', relationName='orders', tupleData=OrdersTuple{o_orderkey=2570, o_custkey=147247, o_orderstatus='O', o_totalprice=33841.14, o_orderdate='1995-10-12', o_orderpriority='3-MEDIUM', o_
orderkey='2538', relationName='lineitem', tupleData=LineitemTuple{l_orderkey=2538, l_partkey=31715, l_suppkey=4219, l_linenumber=3, l_quantity=39.0, l_extendedprice=64221.69, l_discount=0.0, l_ta
orderkey='2571', relationName='orders', tupleData=OrdersTuple{o_orderkey=2571, o_custkey=77473, o_orderstatus='F', o_totalprice=32484.92, o_orderdate='1995-05-25', o_orderpriority='4-NOT SPECIFIE
orderkey='2572', relationName='orders', tupleData=OrdersTuple{o_orderkey=2572, o_custkey=69028, o_orderstatus='F', o_totalprice=48427.49, o_orderdate='1994-07-15', o_orderpriority='2-HIGH', o_cle
orderkey='2511', relationName='orders', tupleData=OrdersTuple{o_orderkey=2511, o_custkey=67750, o_orderstatus='O', o_totalprice=264492.95, o_orderdate='1998-05-03', o_orderpriority='3-MEDIUM', o_
orderkey='2573', relationName='orders', tupleData=OrdersTuple{o_orderkey=2573, o_custkey=50341, o_orderstatus='F', o_totalprice=332188.99, o_orderdate='1994-10-01', o_orderpriority='3-MEDIUM', o_
orderkey='2536', relationName='orders', tupleData=OrdersTuple{o_orderkey=2536, o_custkey=10432, o_orderstatus='F', o_totalprice=43450.99, o_orderdate='1993-09-20', o_orderpriority='3-MEDIUM', o_c
orderkey='2537', relationName='orders', tupleData=OrdersTuple{o_orderkey=2537, o_custkey=108973, o_orderstatus='O', o_totalprice=87786.42, o_orderdate='1998-03-14', o_orderpriority='1-URGENT', o_
orderkey='2538', relationName='orders', tupleData=OrdersTuple{o_orderkey=2538, o_custkey=147554, o_orderstatus='F', o_totalprice=135335.68, o_orderdate='1994-01-24', o_orderpriority='1-URGENT', o
orderkey='2569', relationName='lineitem', tupleData=LineitemTuple{l_orderkey=2569, l_partkey=177627, l_suppkey=145, l_linenumber=3, l_quantity=21.0, l_extendedprice=35797.02, l_discount=0.06, l_t
orderkey='2569', relationName='lineitem', tupleData=LineitemTuple{l_orderkey=2569, l_partkey=165931, l_suppkey=964, l_linenumber=4, l_quantity=27.0, l_extendedprice=53917.11, l_discount=0.0, l_ta
orderkey='2569', relationName='lineitem', tupleData=LineitemTuple{l_orderkey=2569, l_partkey=195745, l_suppkey=5746, l_linenumber=5, l_quantity=35.0, l_extendedprice=64425.9, l_discount=0.07, l_t
orderkey='2570', relationName='lineitem', tupleData=LineitemTuple{l_orderkey=2570, l_partkey=127294, l_suppkey=4831, l_linenumber=1, l_quantity=27.0, l_extendedprice=35674.83, l_discount=0.07, l_
orderkey='2571', relationName='lineitem', tupleData=LineitemTuple{l_orderkey=2571, l_partkey=16954, l_suppkey=4458, l_linenumber=1, l_quantity=18.0, l_extendedprice=33677.1, l_discount=0.09, l_ta
orderkey='2572', relationName='lineitem', tupleData=LineitemTuple{l_orderkey=2572, l_partkey=59529, l_suppkey=4540, l_linenumber=1, l_quantity=7.0, l_extendedprice=10419.64, l_discount=0.1, l_tax
```

Figure 3: The data stream of update sequence

Figure 4 shows the result of this data stream. The order priority is 3-MEDIUM which is exactly the priority of order 2536. The count is 1 means there is only one tuple satisfies the condition.

Then we use some big data source, and the update sequence shown in figure

```
1> Q4SelectResultTuple{o_orderpriority='null', order_count='0'}
1> Q4SelectResultTuple{o_orderpriority='null', order_count='0'}
1> Q4SelectResultTuple{o_orderpriority='null', order_count='0'}
1> Q4SelectResultTuple{o_orderpriority='null', order_count='0'}
1> Q4SelectResultTuple{o_orderpriority='null', order_count='0'}
1> Q4SelectResultTuple{o_orderpriority='null', order_count='0'}
1> Q4SelectResultTuple{o_orderpriority='3-MEDIUM', order_count='1'}
1> Q4SelectResultTuple{o_orderpriority='null', order_count='0'}
1> Q4SelectResultTuple{o_orderpriority='null', order_count='0'}
1> Q4SelectResultTuple{o_orderpriority='null', order_count='0'}
```

Figure 4: The result of TPC-H Q4

5. There is more than one order tuples satisfy the query condition.

The figure 6 shows the result of this large data source. There are total 11 order

```
tupleData=OrdersTuple{o_orderkey=1740, o_custkey=47074, o_orderstatus='O', o_totalprice=105308.35, o_orderdate='1997-05-21', o_orderpriority='5-LOW', o_clerk='Clerk#000000081', o_shippriority=0,
tupleData=OrdersTuple{o_orderkey=1741, o_custkey=71557, o_orderstatus='O', o_totalprice=188074.27, o_orderdate='1996-07-25', o_orderpriority='1-URGENT', o_clerk='Clerk#000000848', o_shippriority=
tupleData=OrdersTuple{o_orderkey=1742, o_custkey=75059, o_orderstatus='F', o_totalprice=202977.22, o_orderdate='1993-08-17', o_orderpriority='3-MEDIUM', o_clerk='Clerk#000000240', o_shippriority=
tupleData=OrdersTuple{o_orderkey=1743, o_custkey=4300, o_orderstatus='O', o_totalprice=145093.76, o_orderdate='1997-02-24', o_orderpriority='1-URGENT', o_clerk='Clerk#000000125', o_shippriority=
tupleData=OrdersTuple{o_orderkey=1768, o_custkey=101798, o_orderstatus='F', o_totalprice=64676.04, o_orderdate='1992-08-21', o_orderpriority='1-URGENT', o_clerk='Clerk#000000238', o_shippriority=
tupleData=OrdersTuple{o_orderkey=1769, o_custkey=2207, o_orderstatus='F', o_totalprice=74034.13, o_orderdate='1994-06-10', o_orderpriority='4-NOT SPECIFIED', o_clerk='Clerk#000000007', o_shippric
tupleData=OrdersTuple{o_orderkey=1770, o_custkey=31039, o_orderstatus='P', o_totalprice=241421.1, o_orderdate='1995-05-22', o_orderpriority='2-HIGH', o_clerk='Clerk#000000336', o_shippriority=0,
tupleData=OrdersTuple{o_orderkey=1771, o_custkey=115480, o_orderstatus='F', o_totalprice=13607.75, o_orderdate='1994-11-30', o_orderpriority='2-HIGH', o_clerk='Clerk#000000519', o_shippriority=0,
tupleData=OrdersTuple{o_orderkey=1800, o_custkey=94520, o_orderstatus='F', o_totalprice=232031.41, o_orderdate='1993-09-03', o_orderpriority='1-URGENT', o_clerk='Clerk#000000292', o_shippriority=
', tupleData=LineitemTuple{l_orderkey=1740, l_partkey=116500, l_suppkey=1523, l_linenumber=1, l_quantity=35.0, l_extendedprice=53077.5, l_discount=0.07, l_tax=0.0, l_returnflag='N', l_linestatus='
', tupleData=LineitemTuple{l_orderkey=1743, l_partkey=62059, l_suppkey=2060, l_linenumber=1, l_quantity=27.0, l_extendedprice=27568.35, l_discount=0.0, l_tax=0.0, l_returnflag='N', l_linestatus='
tupleData=OrdersTuple{o_orderkey=1801, o_custkey=95890, o_orderstatus='O', o_totalprice=132865.83, o_orderdate='1995-05-26', o_orderpriority='4-NOT SPECIFIED', o_clerk='Clerk#000000629', o_shipp
', tupleData=LineitemTuple{l_orderkey=1740, l_partkey=6528, l_suppkey=6529, l_linenumber=2, l_quantity=39.0, l_extendedprice=55946.28, l_discount=0.0, l_tax=0.0, l_returnflag='N', l_linestatus='O'
', tupleData=LineitemTuple{l_orderkey=1743, l_partkey=20412, l_suppkey=5417, l_linenumber=2, l_quantity=37.0, l_extendedprice=49299.17, l_discount=0.01, l_tax=0.01, l_returnflag='N', l_linestatus=
tupleData=OrdersTuple{o_orderkey=1802, o_custkey=21733, o_orderstatus='O', o_totalprice=29831.87, o_orderdate='1996-11-07', o_orderpriority='2-HIGH', o_clerk='Clerk#000000196', o_shippriority=0,
', tupleData=LineitemTuple{l_orderkey=1743, l_partkey=51209, l_suppkey=8725, l_linenumber=3, l_quantity=35.0, l_extendedprice=40607.0, l_discount=0.06, l_tax=0.01, l_returnflag='N', l_linestatus='
', tupleData=LineitemTuple{l_orderkey=1741, l_partkey=140501, l_suppkey=502, l_linenumber=1, l_quantity=19.0, l_extendedprice=29288.5, l_discount=0.0, l_tax=0.0, l_returnflag='N', l_linestatus='O'
', tupleData=LineitemTuple{l_orderkey=1743, l_partkey=67855, l_suppkey=362, l_linenumber=4, l_quantity=16.0, l_extendedprice=29165.6, l_discount=0.04, l_tax=0.06, l_returnflag='N', l_linestatus='
', tupleData=LineitemTuple{l_orderkey=1741, l_partkey=185235, l_suppkey=7754, l_linenumber=2, l_quantity=33.0, l_extendedprice=43567.59, l_discount=0.04, l_tax=0.05, l_returnflag='N', l_linestatus
', tupleData=LineitemTuple{l_orderkey=1768, l_partkey=84931, l_suppkey=2456, l_linenumber=1, l_quantity=33.0, l_extendedprice=63225.69, l_discount=0.06, l_tax=0.0, l_returnflag='R', l_linestatus='
', tupleData=LineitemTuple{l_orderkey=1741, l_partkey=32814, l_suppkey=324, l_linenumber=3, l_quantity=13.0, l_extendedprice=22708.53, l_discount=0.03, l_tax=0.02, l_returnflag='N', l_linestatus='
', tupleData=LineitemTuple{l_orderkey=1768, l_partkey=30930, l_suppkey=3434, l_linenumber=2, l_quantity=3.0, l_extendedprice=5582.79, l_discount=0.07, l_tax=0.01, l_returnflag='R', l_linestatus='F
```

Figure 5: The data stream of TPC-H Q4

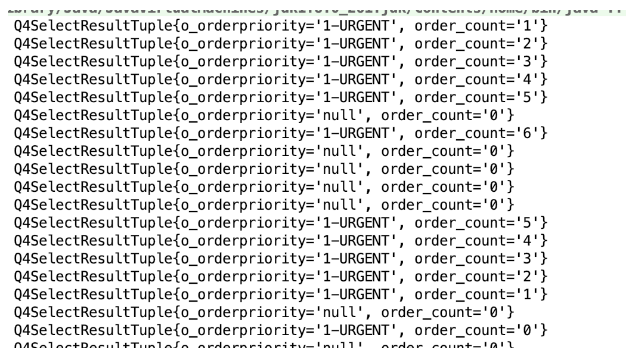
tuples satisfy the condition and they have 2 order priority type 1-URGENT and 3-MEDIUM. The count of order priority of 1-URGENT is 6 and the number of 3-MEDIUM is 5.

```
1> Q4SelectResultTuple{o_orderpriority='3-MEDIUM', order_count='1'}
1> Q4SelectResultTuple{o_orderpriority='3-MEDIUM', order_count='2'}
1> Q4SelectResultTuple{o_orderpriority='null', order_count='0'}
1> Q4SelectResultTuple{o_orderpriority='3-MEDIUM', order_count='3'}
1> Q4SelectResultTuple{o_orderpriority='3-MEDIUM', order_count='4'}
1> Q4SelectResultTuple{o_orderpriority='3-MEDIUM', order_count='5'}
1> Q4SelectResultTuple{o_orderpriority='null', order_count='0'}
1> Q4SelectResultTuple{o_orderpriority='null', order_count='0'}
1> Q4SelectResultTuple{o_orderpriority='null', order_count='0'}
1> Q4SelectResultTuple{o_orderpriority='null', order_count='0'}
1> Q4SelectResultTuple{o_orderpriority='null', order_count='0'}
1> Q4SelectResultTuple{o_orderpriority='null', order_count='0'}
1> Q4SelectResultTuple{o_orderpriority='null', order_count='0'}
1> Q4SelectResultTuple{o_orderpriority='null', order_count='0'}
1> Q4SelectResultTuple{o_orderpriority='null', order_count='0'}
1> Q4SelectResultTuple{o_orderpriority='null', order_count='0'}
1> Q4SelectResultTuple{o_orderpriority='null', order_count='0'}
1> Q4SelectResultTuple{o_orderpriority='null', order_count='0'}
1> Q4SelectResultTuple{o_orderpriority='null', order_count='0'}
1> Q4SelectResultTuple{o_orderpriority='null', order_count='0'}
1> Q4SelectResultTuple{o_orderpriority='null', order_count='0'}
1> Q4SelectResultTuple{o_orderpriority='1-URGENT', order_count='1'}
1> Q4SelectResultTuple{o_orderpriority='null', order_count='0'}
1> Q4SelectResultTuple{o_orderpriority='1-URGENT', order_count='2'}
1> Q4SelectResultTuple{o_orderpriority='1-URGENT', order_count='3'}
1> Q4SelectResultTuple{o_orderpriority='1-URGENT', order_count='4'}
1> Q4SelectResultTuple{o_orderpriority='1-URGENT', order_count='5'}
1> Q4SelectResultTuple{o_orderpriority='1-URGENT', order_count='6'}
1> Q4SelectResultTuple{o_orderpriority='null', order_count='0'}
```

Figure 6: The result of TPC-H Q4

5.2 Delete

We also did the experiment of delete update action on the second data source. We first insert these data, then delete them. Figure 7 shows the result, the number of 1-URGENT increase as we insert data which satisfying the condition, then decrease to 0 as the delete update stream arrive.



```
Q4SelectResultTuple{o_orderpriority='1-URGENT', order_count='1'}
Q4SelectResultTuple{o_orderpriority='1-URGENT', order_count='2'}
Q4SelectResultTuple{o_orderpriority='1-URGENT', order_count='3'}
Q4SelectResultTuple{o_orderpriority='1-URGENT', order_count='4'}
Q4SelectResultTuple{o_orderpriority='1-URGENT', order_count='5'}
Q4SelectResultTuple{o_orderpriority='null', order_count='0'}
Q4SelectResultTuple{o_orderpriority='1-URGENT', order_count='6'}
Q4SelectResultTuple{o_orderpriority='null', order_count='0'}
Q4SelectResultTuple{o_orderpriority='null', order_count='0'}
Q4SelectResultTuple{o_orderpriority='null', order_count='0'}
Q4SelectResultTuple{o_orderpriority='null', order_count='0'}
Q4SelectResultTuple{o_orderpriority='1-URGENT', order_count='5'}
Q4SelectResultTuple{o_orderpriority='1-URGENT', order_count='4'}
Q4SelectResultTuple{o_orderpriority='1-URGENT', order_count='3'}
Q4SelectResultTuple{o_orderpriority='1-URGENT', order_count='2'}
Q4SelectResultTuple{o_orderpriority='1-URGENT', order_count='1'}
Q4SelectResultTuple{o_orderpriority='null', order_count='0'}
Q4SelectResultTuple{o_orderpriority='1-URGENT', order_count='0'}
```

Figure 7: The delete results of TPC-H Q4

6 Conclusion and future directions

As the experiment shows, this program can handle TPC-H Q4. It can deal with both delete action and insert action, maintaining the data in real time as new update data arrives as a stream.
However, the structure of TPC-H Q4 is too simple and it only has 2 tables. So it does not need to maintain assertion key and for Insert-Update and Delete-Updata, the lineitem only has one child order, thus, it does not need to perform the above 2 operations recursively. In the future, I want to try some more complex queries and find a good way to perform join operation.

References

[1] Wang, Qichen, and Ke Yi. "Maintaining Acyclic Foreign-Key Joins under Updates." Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data. 2020.