Databend

Databend

Stream 设计与实现

张祖前

Databend

Content

Databend

Part 1

背景介绍

Databend

什么是CDC

CDC即变更数据捕获，用于实时或近实时检测和捕获数据库中数据的变更（包括插入、更新和删除操作）。

它能够捕获源数据库中数据的变更事件，并将这些变更应用到目标数据库或数据仓库中，从而保持数据的一致性和同步。

方案	支持实时	侵入性	特点
时间戳	否	否	实现简单，变更延迟，难以捕获删除
触发器	是	否	实时精准捕获变更，性能影响
快照	否	否	实现简单，实时性差，资源消耗
日志	是	是	实时精准捕获变更，维护复杂

Databend

Databand Stream

启发自Snowflake SIGMOD'23 论文 What's the Difference? Incremental Processing with Change Queries in Snowflake

旨在提供一种更高效、简洁、且与云服务高度集成的数据变更捕获方法

Databend

Part 2

实现细节

自动变更追踪

column	默认值	说明
_origin_block_id	Null	该行第一次被添加到表时所属的数据块（block）的 ID
_origin_block_row_num	Null	该行第一次被添加时在数据块中的行号
_origin_version	Null	该行第一次移动时事务所基于的 table version
_row_version	0	该行的版本信息

自动变更追踪

- Insert
- Delete
- Compact
- Recluster
- Update

```
mysql> insert into t values(1, 1),(2, 2);
Query OK, 2 rows affected (0.08 sec)

mysql> select a, b, _origin_block_id,_origin_block_row_num,_origin_version,_row_version from t;
+------+------+------------------+----------------------+-----------------+--------------+
| a    | b    | _origin_block_id | _origin_block_row_num | _origin_version | _row_version |
+------+------+------------------+----------------------+-----------------+--------------+
|    1 |    1 |             NULL |                 NULL |            NULL |            0 |
|    2 |    2 |             NULL |                 NULL |            NULL |            0 |
+------+------+------------------+----------------------+-----------------+--------------+
2 rows in set (0.03 sec)
Read 2 rows, 274.00 B in 0.005 sec., 440.17 rows/sec., 58.89 KiB/sec.

mysql> update t set a = 0 where b = 1;
Query OK, 1 row affected (0.13 sec)

mysql> select a, b, _origin_block_id,_origin_block_row_num,_origin_version,_row_version from t;
+------+------+-----------------------------------------+----------------------+-----------------+--------------+
| a    | b    | _origin_block_id                        | _origin_block_row_num | _origin_version | _row_version |
+------+------+-----------------------------------------+----------------------+-----------------+--------------+
|    0 |    1 | -168313755496731310539518324155007614376 |                    0 |             166 |            1 |
|    2 |    2 | -168313755496731310539518324155007614376 |                    1 |             166 |            0 |
+------+------+-----------------------------------------+----------------------+-----------------+--------------+
2 rows in set (0.05 sec)
Read 2 rows, 101.00 B in 0.008 sec., 251.07 rows/sec., 12.38 KiB/sec.

mysql> insert into t values(3, 3), (4,4);
Query OK, 2 rows affected (0.08 sec)

mysql> optimize table t compact;
Query OK, 4 rows affected (0.11 sec)

mysql> select a, b, _origin_block_id,_origin_block_row_num,_origin_version,_row_version from t;
+------+------+-----------------------------------------+----------------------+-----------------+--------------+
| a    | b    | _origin_block_id                        | _origin_block_row_num | _origin_version | _row_version |
+------+------+-----------------------------------------+----------------------+-----------------+--------------+
|    0 |    1 | -168313755496731310539518324155007614376 |                    0 |             166 |            1 |
|    2 |    2 | -168313755496731310539518324155007614376 |                    1 |             166 |            0 |
|    3 |    3 |  103622614062755752182794218109030081480 |                    0 |             173 |            0 |
|    4 |    4 |  103622614062755752182794218109030081480 |                    1 |             173 |            0 |
+------+------+-----------------------------------------+----------------------+-----------------+--------------+
4 rows in set (0.06 sec)
Read 4 rows, 197.00 B in 0.006 sec., 615.71 rows/sec., 29.61 KiB/sec.
```

变更捕获

$$\Delta_{\text{insert}} = \{\forall r \in \text{A} : \neg\big(r.origin\_version \text{ IS NOT NULL} \wedge \big((r.origin\_version < v_0) \vee (r.origin\_version \geq v_0 \wedge r.origin\_block\_id \in \text{delete\_block\_ids})\big)\big)\}$$

$$\Delta_{\text{update}} = \{\forall r \in \text{A} \cup \text{D} : \big(A.r.row\_id = D.r.row\_id \wedge A.row\_version > D.row\_version\big)\}$$

$$\Delta_{\text{delete}} = \{r \in \text{D} \wedge r \notin \text{A}\}$$

Stream

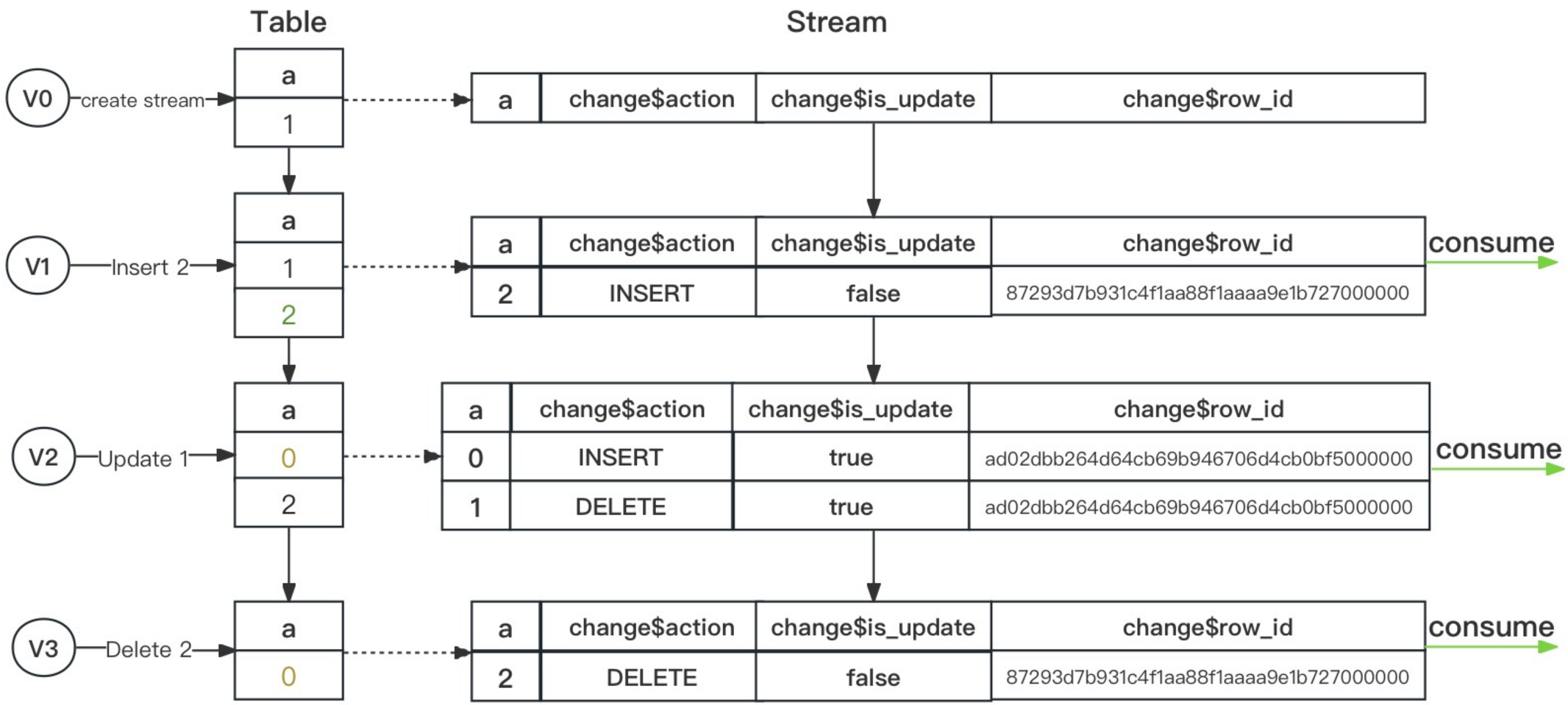
- 创建 Stream： CREATE STREAM [IF NOT EXISTS] [<database_name>.]<stream_name> ON TABLE [<database_name>.]<table_name>

  [AT (STREAM => <stream_name>)] [append_only = true | false] [COMMENT = '<comment>']

- Append only Stream

- Standard Stream

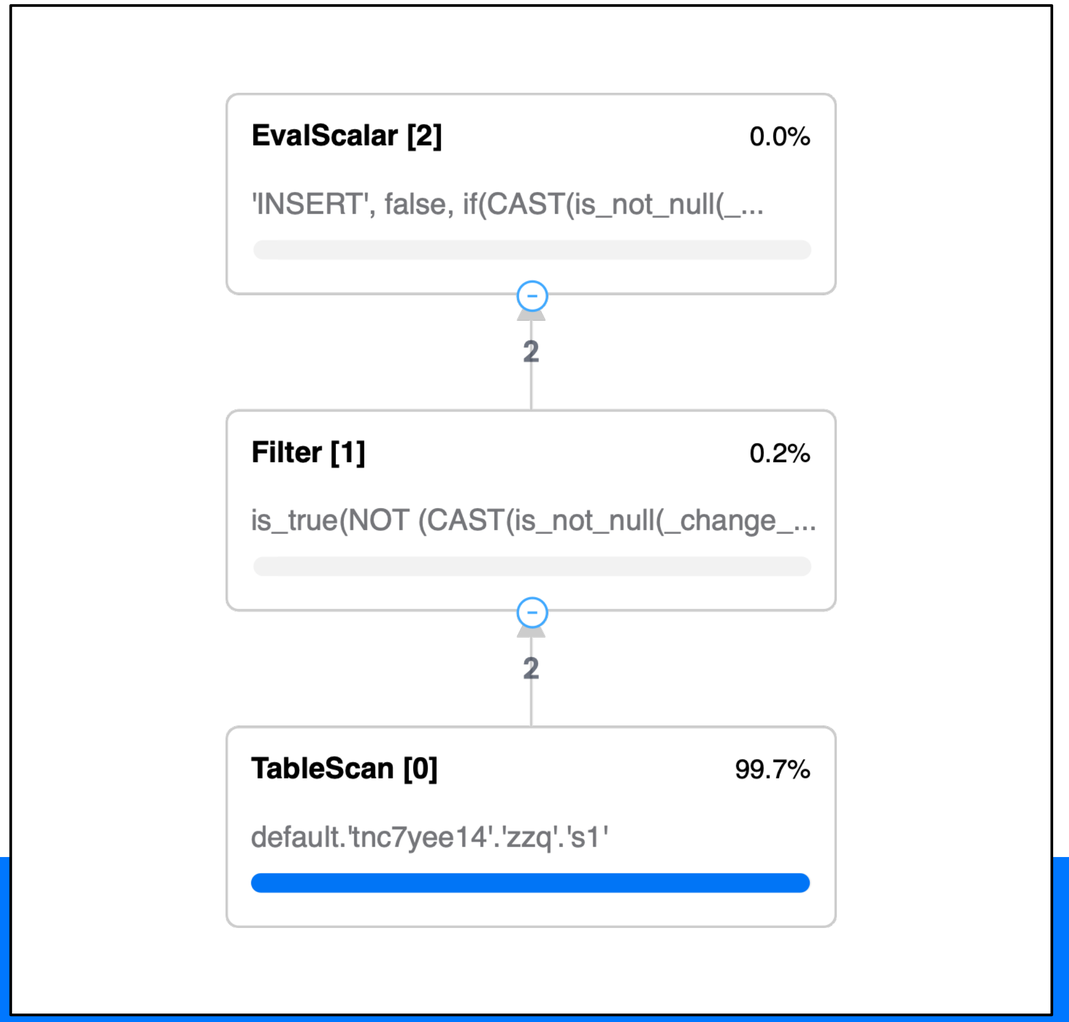
Stream Columns



Append only Stream

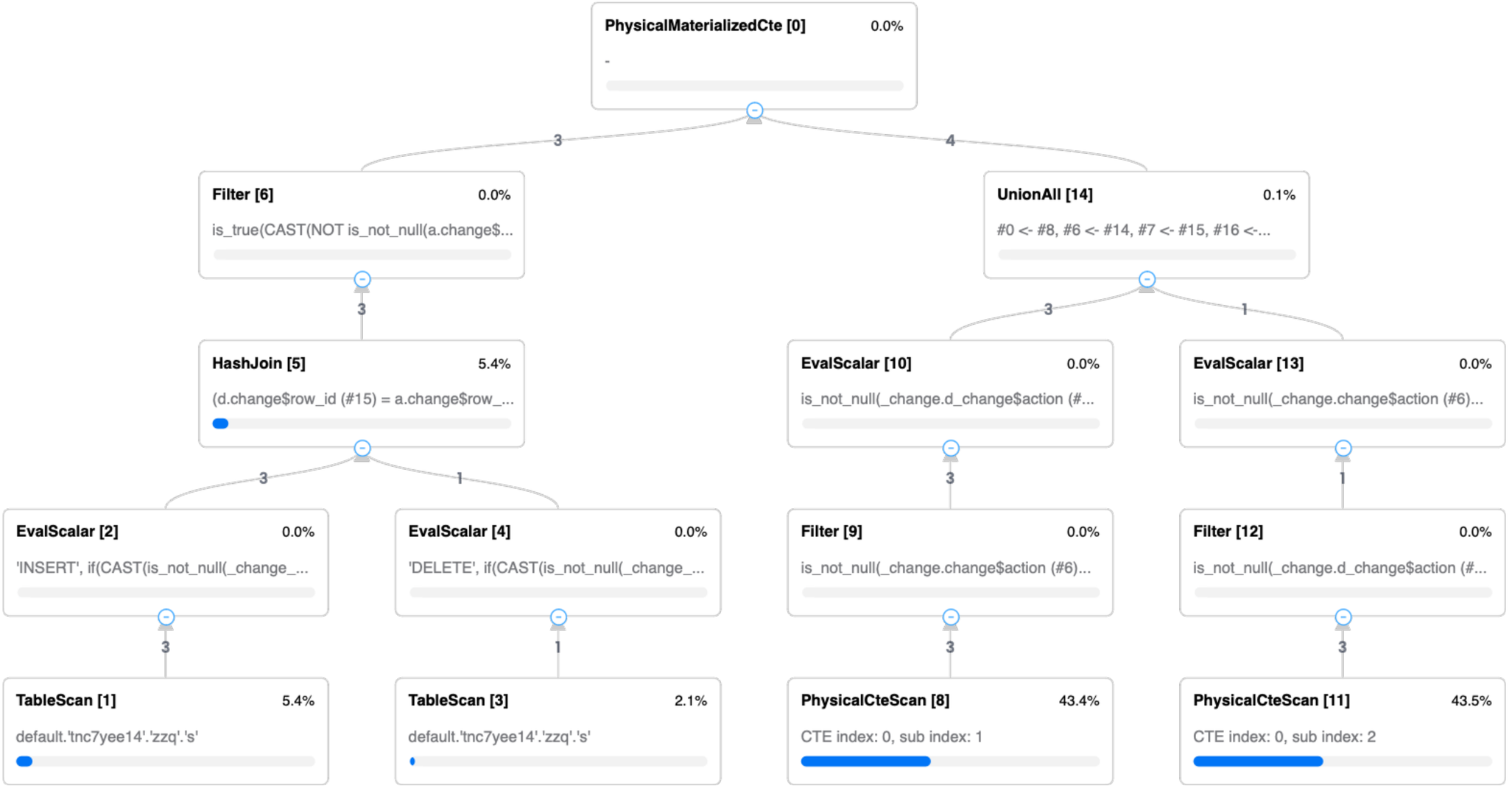
01 只追踪和记录新添加到表中的行

02 忽略周期内多次更新



Databend

Standard Stream



Databend

Part 3

最佳实践

Databend

最佳实践

生命周期管理

整合数据处理流程

测试和验证

Databend

Thank you!