Streams

Sujith Nair
Cloud Data Architect

Snowflake Snowpro Certified

How does a stream become empty ? How can you purge a stream ?

Stream capture changes like INSERT,UPDATES and DELETES.

Stream will become empty or loose rows when they are consumed by
- Insert into the target table. Insert into customer as select * from customer_stream;
- Create table from the stream Create table customer as select * from customer_stream;
- Re-create the stream Create or replace stream customer_stream on table customer ;
- Use merge with streams to update the target table;

```
merge into CUSTOMER C
using CUSTOMER_STREAM S
  on  f.id = s.id
when matched
   and S.METADATA$ACTION ='INSERT'
   and S.METADATA$ISUPDATE ='TRUE'
   then update
   set c.name = s.name,
      when matched
   and S.METADATA$ACTION ='DELETE'
   and S.METADATA$ISUPDATE ='FALSE'
   then delete
WHEN NOT MATCHED THEN INSERT(ID) values('1')
```

Can you join a stream and table ?

Yes, you can join a table and stream object to select rows that have changed and insert into the target table

```
Insert into CUSTOMER
SELECT  s.id, s.name
FROM CUSTOMER_STREAM S
       JOIN CUSTOMER C
ON s.id=c.id ;

merge into CUSTOMER C
using CUSTOMER_STREAM S
  on  f.id = s.id
when matched
   and S.METADATA$ACTION ='INSERT'
   and S.METADATA$ISUPDATE ='TRUE'
   then update
   set c.name = s.name,
      when matched
   and S.METADATA$ACTION ='DELETE'
   and S.METADATA$ISUPDATE ='FALSE'
   then delete
WHEN NOT MATCHED THEN INSERT(ID) values('1')
```

In what scenario did you use streams on your project ?

- Use Case :We receive IOT data in Cloud storage(s3 or Azure Cloud storage) every few mins, we want to ingest this data to snowflake as soon as possible. After data has been loaded to snowflakes raw layer, the data needs to be inserted into the final tables in the snowflake schema immediately so that data is available to the business.

- We ingest data from s3 into snowflake raw tables with snowpipe
- Append only Streams are defined on the raw tables
- Tasks poll the stream for data ,Streams are consumed and data is loaded to the target snowflake tables.

Use case: You have large files of around 1 GB that is being added to cloud storage. There is no specific schedule for the receipt of these files They could come any time and multiple times on the same day. The data in the file needs to be ingested immediately (near real time) and be made available to business for consumption.

Why not Snowpipe ?
- File size close to 1 GB, not suitable for snowpipe

- Create External tables which will give us select access to the files .
- Define a stream on the external table.
- As files are added to the cloud directory the rows from the file will be available in the stream.
- Tasks poll the stream for data ,Streams are consumed and data is loaded to the target snowflake tables

What is change tracking on a table? When should it be used instead of streams ?

Change tracking is an alternate method of capturing changes to a table. It is different from streams and offers a few advantages over streams.
We can enable it with the

Advantages of change tracking over streams:
- We can select changes between a time period.
- Inserting data does not cause the rows to get emptied

alter table customer set change_tracking =TRUE

Change tracking cannot be used for external tables
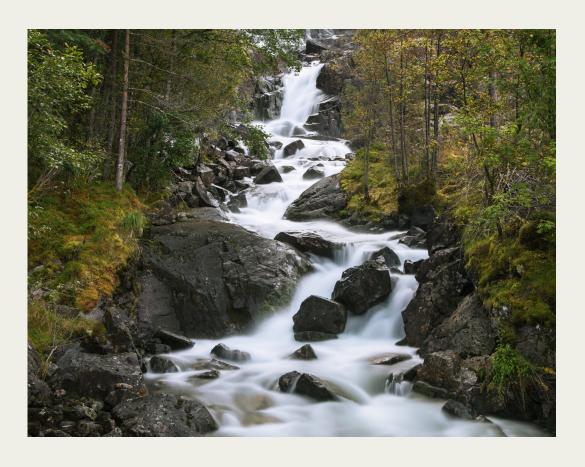
Steam object creation is a method of capturing change data capture(CDC) from a table or view.

The changes that are captured are DELETE ,INSERT and UPDATE

CREATE STREAM mystream ON TABLE mytable;

There are 3 types of streams

- Default/Standard (Insert/Update & Delete)
- Append Only (Insert for tables)
- Insert only(Applicable to external tables)

Thank you!

Learn2CloudData Solutions