**Streams for CDC**

In Snowflake, a Stream is used for **Change Data Capture (CDC)**. Its a medaitor between the source table and traget table and also it will capture changes through offset. Streams records row-level DML changes like (INSERT, UPDATE, DELETE) made to a table. We can then consume the stream to process only incremental data instead of reprocessing the entire dataset. we have to consume these changes within 14 days or else the data will stale(disabled). There are 3 types of streams available in snowflake

**Standard Stream or delta** → it is a default stream and it will Tracks all changes like (INSERT, UPDATE, DELETE) including truncates as well.

**Append only stream** - It will capture only inserts

**Insert-only Stream** → Tracks external tables only if new file is added at cloud stroage stream captures. Retention time is minimum 14 days and maximum 90 days

For example, if we used a stream on staging tables to capture only new or updated rows, which we then transformed and loaded into reporting tables. This reduced processing time and Depending on the use case, we can use **standard streams** for full CDC or **insert-only streams** for append-only data.

-- Create a table to demonstrate

create or replace table raw\_applicant\_staging

( id number,

first\_name varchar,

last\_name varchar,

sex varchar,

ethinicity varchar,

ssn varchar,

street\_address varchar,

education\_level varchar,

years\_of\_experience number,

job\_id number

);

select \* from raw\_applicant\_staging ;

-- job\_details --

create or replace table job\_details (

job\_id number,

name varchar,

city varchar,

state varchar,

education\_level varchar

) ;

delete from job\_details ;

insert into job\_details values (10, 'Painter', 'Raleigh', 'NC','High School') ;

insert into job\_details values (20, 'Software Engineer', 'Raleigh', 'NC','Masters') ;

insert into job\_details values (30, 'Data Architect', 'Raleigh', 'NC','Undergrad') ;

insert into job\_details values (40, 'Vice President', 'Raleigh', 'NC','Masters') ;

insert into job\_details values (50, 'Associate', 'Raleigh', 'NC','Masters') ;

--

select \* from job\_details ;

-- candidates

create or replace table candidates (

id number,

first\_name varchar,

last\_name varchar,

sex varchar,

ethinicity varchar,

ssn varchar,

street\_address varchar,

candidate\_education\_level varchar,

years\_of\_experience number,

job\_id number,

job\_name varchar,

job\_city varchar,

job\_state varchar,

required\_education\_level varchar,

status varchar,

comments varchar

) ;

-- populate the table candidates

delete from candidates ;

insert into candidates

values (100000, 'James', 'Schwartz', 'M', 'American','342-76-9087','5676 Washington Street','High School', 5,10, 'Painter', 'Raleigh', 'NC','High School','SHORTLISTED','The education level of candidate matched with job') ;

-- check the data

select \* from candidates ;

-- create stream

create or replace stream strm\_applicant1 on table raw\_applicant\_staging ;

drop stream strm\_applicant ;

desc stream strm\_applicant1 ;

-- stale\_after => Post 14 days

show streams ;

show parameters like '%extension%retention%' in database ;

show parameters like '%retention%' in account; -- 1

show parameters like '%extension%' in account; -- 14

-- Lets insert some data in raw-applicant\_staging table

insert into raw\_applicant\_staging values (111111, 'James', 'Schwartz', 'M', 'American','342-76-9087','5676 Washington Street','High School', 5,10) ;

insert into raw\_applicant\_staging values (222222, 'Jessica', 'Escobar', 'F', 'Hispanic','456-93-5629','3234 WateringCan Drive','Undergrad', 4, 10) ;

insert into raw\_applicant\_staging values (333333, 'Ben', 'Hardy', 'M', 'American','876-98-3245','6578 Historic Circle','Masters', 6, 30) ;

insert into raw\_applicant\_staging values (444444, 'Anjali', 'Singh', 'F', 'Indian American','435-87-6532','8978 Autumn Day Drive','Masters', 8,20) ;

insert into raw\_applicant\_staging values (555555, 'Dean', 'Tracy', 'M', 'African','767-34-7656','2343 India Street','Undergrad', 2,50) ;

-- check data in staging

select \* from raw\_applicant\_staging ;

-- check stream

select \* from strm\_applicant1; -- Observe extra columns

-- Pull the records which are inserted

select \* from strm\_applicant1 where metadata$action = 'INSERT' and metadata$isupdate = 'FALSE';

-- Update the data in staging

update raw\_applicant\_staging set job\_id = 10 where id = 555555 ;

-- check stream

select \* from raw\_applicant\_staging;

select \* from strm\_applicant1; -- Observe extra columns

-- Pull the records which are inserted

delete from raw\_applicant\_staging where id = 555555 ;

--

select \* from strm\_applicant1;

-- off set is not getting reset.since we are not consuming..

-- lets delete the data from staging..and see consumption of stream

delete from raw\_applicant\_staging ;

--staging cleaned

select \* from raw\_applicant\_staging ;

-- stream no records

select \* from strm\_applicant1;

-- INSERT Staging

insert into raw\_applicant\_staging values (111111, 'James', 'Schwartz', 'M', 'American','342-76-9087','5676 Washington Street','High School', 5,10) ;

insert into raw\_applicant\_staging values (222222, 'Jessica', 'Escobar', 'F', 'Hispanic','456-93-5629','3234 WateringCan Drive','Undergrad', 4, 10) ;

insert into raw\_applicant\_staging values (333333, 'Ben', 'Hardy', 'M', 'American','876-98-3245','6578 Historic Circle','Masters', 6, 30) ;

insert into raw\_applicant\_staging values (444444, 'Anjali', 'Singh', 'F', 'Indian American','435-87-6532','8978 Autumn Day Drive','Masters', 8,20) ;

insert into raw\_applicant\_staging values (555555, 'Dean', 'Tracy', 'M', 'African','767-34-7656','2343 India Street','Undergrad', 2,50) ;

--consume Stream

select \* from candidates ; -- Target

-- check data of job details

select \* from job\_details ;

-- Consume streams -- starting with insert

merge into candidates tgt

using

(select

id ,

first\_name ,

last\_name ,

sex ,

ethinicity ,

ssn ,

street\_address ,

str.education\_level as ed\_level,

years\_of\_experience ,

jdtl.job\_id ,

name ,

city ,

state ,

jdtl.education\_level as jreq\_level ,

case when str.education\_level = jdtl.education\_level then 'SHORTLISTED ' else 'STAGED' end as status,

case when str.education\_level = jdtl.education\_level then 'The education level of candidate matched with job'

else 'Application received from candidate'

end as comments,

str.metadata$action,

str.metadata$isupdate

from strm\_applicant1 str inner join job\_details jdtl on (str.job\_id = jdtl.job\_id)

) src

on tgt.id = src.id

-- insert clause

when not matched and src.metadata$action = 'INSERT' and metadata$isupdate = 'FALSE'

then insert

values (

id ,

first\_name ,

last\_name ,

sex ,

ethinicity ,

ssn ,

street\_address,

ed\_level,

years\_of\_experience ,

job\_id ,

name ,

city ,

state ,

jreq\_level,

status,

comments

)

;

-- check

select \* from candidates ;

-- check if stream is consumed

select \* from strm\_applicant1 ;

select \* from raw\_applicant\_staging;

--- Update Operation on raw

update raw\_applicant\_staging set job\_id = 10 where id = 555555 ;

select \* from strm\_applicant1 ;

-- update education level in second update

update raw\_applicant\_staging set education\_level = 'High School' where id = 555555 ;

select \* from strm\_applicant1 ;

-- Consume streams -- update

merge into candidates tgt

using

(select

id ,

first\_name ,

last\_name ,

sex ,

ethinicity ,

ssn ,

street\_address ,

str.education\_level as ed\_level,

years\_of\_experience ,

jdtl.job\_id ,

name ,

city ,

state ,

jdtl.education\_level as jreq\_level ,

case when str.education\_level = jdtl.education\_level then 'SHORTLISTED ' else 'STAGED' end as status,

case when str.education\_level = jdtl.education\_level then 'The education level of candidate matched with job'

else 'Application recived from candidate'

end as comments,

str.metadata$action,

str.metadata$isupdate

from strm\_applicant1 str inner join job\_details jdtl on (str.job\_id = jdtl.job\_id)

) src

on tgt.id = src.id

-- insert clause

when not matched and src.metadata$action = 'INSERT' and metadata$isupdate = 'FALSE'

then insert

values (

id ,

first\_name ,

last\_name ,

sex ,

ethinicity ,

ssn ,

street\_address,

ed\_level,

years\_of\_experience ,

job\_id ,

name ,

city ,

state ,

jreq\_level,

status,

comments

)

--update

when matched and src.metadata$action = 'INSERT' and metadata$isupdate = 'TRUE'

then update set

tgt.job\_id = src.job\_id,

tgt.candidate\_education\_level = src.ed\_level,

tgt.required\_education\_level = src.jreq\_level,

tgt.status = src.status,

tgt.comments = src.comments

;

-- check target table

select \* from candidates ;

select \* from raw\_applicant\_staging;

--- DELETE

delete from raw\_applicant\_staging where id = 555555 ;

select \* from strm\_applicant1 ;

-- Consume

-- Consume streams -- update

create task tsk\_str\_has\_data

WAREHOUSE = 'COMPUTE\_WH'

when system$stream\_has\_data('strm\_applicant1')

AS

merge into candidates tgt

using

(select

id ,

first\_name ,

last\_name ,

sex ,

ethinicity ,

ssn ,

street\_address ,

str.education\_level as ed\_level,

years\_of\_experience ,

jdtl.job\_id ,

name ,

city ,

state ,

jdtl.education\_level as jreq\_level ,

case when str.education\_level = jdtl.education\_level then 'SHORTLISTED ' else 'STAGED' end as status,

case when str.education\_level = jdtl.education\_level then 'The education level of candidate matched with job'

else 'Application recived from candidate'

end as comments,

str.metadata$action,

str.metadata$isupdate

from strm\_applicant1 str inner join job\_details jdtl on (str.job\_id = jdtl.job\_id)

) src

on tgt.id = src.id

-- insert clause

when not matched and src.metadata$action = 'INSERT' and metadata$isupdate = 'FALSE'

then insert

values (

id ,

first\_name ,

last\_name ,

sex ,

ethinicity ,

ssn ,

street\_address,

ed\_level,

years\_of\_experience ,

job\_id ,

name ,

city ,

state ,

jreq\_level,

status,

comments

)

--update

when matched and src.metadata$action = 'INSERT' and metadata$isupdate = 'TRUE'

then update

set tgt.job\_id = src.job\_id,

tgt.candidate\_education\_level = src.ed\_level,

tgt.required\_education\_level = src.jreq\_level,

tgt.status = src.status,

tgt.comments = src.comments

-- delete

when matched and src.metadata$action = 'DELETE' and metadata$isupdate = 'FALSE'

then delete

;

show tasks like 'tsk%';

alter task tsk\_str\_has\_data resume;

-- check target table

select \* from candidates ;

insert into raw\_applicant\_staging values (555555, 'Dean', 'Tracy', 'M', 'African','767-34-7656','2343 India Street','Undergrad', 2,50);

select \* from raw\_applicant\_staging;

select \* from strm\_applicant1;

alter task tsk\_str\_has\_data suspend;

show tasks;

-- Types of Streams

-- Standard, Append Only

create or replace table raw\_applicant\_staging

( id number,

first\_name varchar,

last\_name varchar,

sex varchar,

ethinicity varchar,

ssn varchar,

street\_address varchar,

education\_level varchar,

years\_of\_experience number,

job\_id number

);

select \* from raw\_applicant\_staging ;

-- Default Stream

create or replace stream strm\_applicant1 on table raw\_applicant\_staging ;

-- Append Only Stream

create or replace stream strm\_applicant2 on table raw\_applicant\_staging append\_only = true ;

--describe default stream

desc stream strm\_applicant1 ;

desc stream strm\_applicant2 ;

--show all streams

show streams ;

-- Lets insert some data in raw-applicant\_staging table

insert into raw\_applicant\_staging values (111111, 'James', 'Schwartz', 'M', 'American','342-76-9087','5676 Washington Street','High School', 5,10) ;

insert into raw\_applicant\_staging values (222222, 'Jessica', 'Escobar', 'F', 'Hispanic','456-93-5629','3234 WateringCan Drive','Undergrad', 4, 10) ;

insert into raw\_applicant\_staging values (333333, 'Ben', 'Hardy', 'M', 'American','876-98-3245','6578 Historic Circle','Masters', 6, 30) ;

insert into raw\_applicant\_staging values (444444, 'Anjali', 'Singh', 'F', 'Indian American','435-87-6532','8978 Autumn Day Drive','Masters', 8,20) ;

insert into raw\_applicant\_staging values (555555, 'Dean', 'Tracy', 'M', 'African','767-34-7656','2343 India Street','Undergrad', 2,50) ;

--check data in streams

select \* from strm\_applicant1 ;

select \* from strm\_applicant2 ;

-- delete the data in staging table

delete from raw\_applicant\_staging where id in (111111,222222,333333) ;

-- check data

select \* from strm\_applicant1 ; -- shows 2 rows

select \* from strm\_applicant2 ; -- shows 5 rows - does not capture the delete

--update the data

update raw\_applicant\_staging set education\_level = 'High School' where id = 555555 ;

-- check data

select \* from strm\_applicant1 ; -- shows 2 rows, and id 555555 updated the education level to High School

select \* from strm\_applicant2 ; -- shows 5 rows - does not capture the delete, no impact of update

-- truncate the data

truncate table raw\_applicant\_staging ;

-- check data

select \* from strm\_applicant1 ; -- No rows, reflects the truncate

select \* from strm\_applicant2 ; -- shows 5 rows - No effect of truncate

-- NET EFFECT of Default Stream,

insert into raw\_applicant\_staging values (666666, 'Sam', 'Hill', 'M', 'American','767-84-7346','2343 Historic Street','Masters', 2,50) ;

delete from raw\_applicant\_staging where id = 666666 ;

select \* from raw\_applicant\_staging ;

-- check data

select \* from strm\_applicant1 ; -- Insert and then delete is not reflected in the default stream. Since default stream provides the row level delta.

select \* from strm\_applicant2 ; -- The new row inserted is reflected.

-- How to advance offset in a stream using a false condition

create or replace table raw\_applicant\_temp as select \* from raw\_applicant\_staging where 1 = 2;

-- consume stream using false cindition

insert into raw\_applicant\_staging values (666666, 'Sam', 'Hill', 'M', 'American','767-84-7346','2343 Historic Street','Masters', 2,50) ;

insert into raw\_applicant\_temp

select id ,

first\_name ,

last\_name ,

sex ,

ethinicity ,

ssn ,

street\_address ,

education\_level,

years\_of\_experience ,

job\_id

from strm\_applicant1 where 1 = 2;

-- stream is consumed

select \* from strm\_applicant1 ;

-- can be consumed using a create temp table

insert into raw\_applicant\_staging values (666666, 'Sam', 'Hill', 'M', 'American','767-84-7346','2343 Historic Street','Masters', 2,50) ;

create or replace table raw\_applicant\_temp2 as

select id ,

first\_name ,

last\_name ,

sex ,

ethinicity ,

ssn ,

street\_address ,

education\_level,

years\_of\_experience ,

job\_id

from strm\_applicant1 where 1 = 2

-- stream is consumed

select \* from strm\_applicant1 ;

select \* from raw\_applicant\_temp2 ;