**Snowpipe**

Snowpipe is Snowflake’s serverless, continuous data ingestion service that automatically loads new files from cloud storage like S3 or Azure Blob into Snowflake tables. Unlike batch COPY INTO jobs, Snowpipe ingests data in near real-time as soon as files arrive. Its supports structured and semi structued data. It will do automatic scaling depending on file arrival rate. In my Siemens project, we used Snowpipe with S3 and Streams to continuously load JSON logs into Snowflake for real-time reporting. This reduced manual scheduling and helped keep analytics dashboards always up to date.

create or replace storage integration int\_aws\_s3

type = external\_stage

storage\_provider = s3

enabled = true

storage\_aws\_role\_arn = 'arn:aws:iam::755743318453:role/snowflake-s3-full-access-role'

storage\_allowed\_locations = ('s3://snowflakes3datafiles/snowpipe/')

comment = 'This is the integration object for loading the files from AWS S3 to Snowflake' ;

alter storage integration int\_aws\_s3 set

storage\_allowed\_locations=('s3://snowflakes3datafiles/snowpipe/','s3://snowflakes3datafiles/snowpipe/file5/');

-- Describe Intergation Object

desc integration int\_aws\_s3 ;

-- Copy ARN and External ID in IAM Role at AWS

-- create file format - since it will be used in the external stage and copy command. so creating is better

create or replace file format file\_format\_csv

compression = none;

--describe file formats

describe file format file\_format\_csv ;

-- create a snowpipe folder at s3 bucket and add a file promotions1.csv

-- create external stage for loading data

create or replace stage ext\_stage\_ld

url = 's3://snowflakes3datafiles/snowpipe/'

storage\_integration = int\_aws\_s3

file\_format = file\_format\_csv ;

-- if error

-- copy storage\_aws\_external\_id in the snowflake-full-access role trust relationship

-- Show stages

show stages ;

-- list

list @ext\_stage\_ld ;

-- Create the table to load

create or replace TABLE PROMOTIONS

( P\_PROMO\_SK NUMBER(38,0), P\_PROMO\_ID VARCHAR(16), P\_START\_DATE\_SK NUMBER(38,0), P\_END\_DATE\_SK NUMBER(38,0),

P\_ITEM\_SK NUMBER(38,0), P\_COST NUMBER(15,2), P\_RESPONSE\_TARGET NUMBER(38,0), P\_PROMO\_NAME VARCHAR(50),

P\_CHANNEL\_DMAIL VARCHAR(1), P\_CHANNEL\_EMAIL VARCHAR(1), P\_CHANNEL\_CATALOG VARCHAR(1), P\_CHANNEL\_TV VARCHAR(1),

P\_CHANNEL\_RADIO VARCHAR(1), P\_CHANNEL\_PRESS VARCHAR(1), P\_CHANNEL\_EVENT VARCHAR(1), P\_CHANNEL\_DEMO VARCHAR(1),

P\_CHANNEL\_DETAILS VARCHAR(100), P\_PURPOSE VARCHAR(15), P\_DISCOUNT\_ACTIVE VARCHAR(1)

);

select \* from promotions ;

-- Copy command working

copy into promotions from @ext\_stage\_ld ;

-- select

select \* from promotions ;

-- truncate table and remove file promotions1.csv from S3

truncate table promotions ;

--create pipe

create or replace pipe pipe\_for\_promotions

auto\_ingest = TRUE

as

copy into promotions from @ext\_stage\_ld ;

--describe pipe

describe pipe pipe\_for\_promotions ;

-- copy notification\_channel value

-- Go to AWS console

-- Click on Bucket

-- Click on Properties

-- Scroll Down to Create Event Notification

-- click on create event notification

-- Give event name and prefix

-- Event Types - All object create events

-- Go to Destination

-- select SQS Queue

-- Paste the notification\_channel value under "Enter SQS queue ARN"

-- Click on Save

-- Remove file promotion1.csv from snowpipe

-- Keep Adding files to snowpipe folder and

-- Check the data

select \* from promotions order by p\_promo\_sk;

-- Add the promotions1.csv to snowpipe folder in the bucket

-- Check the status of pipe

select system$pipe\_status('pipe\_for\_promotions') ;

-- Refresh the pipe

alter pipe pipe\_for\_promotions refresh ;

-- Error message

select \* from table(validate\_pipe\_load(pipe\_name => 'pipe\_for\_promotions', start\_time=>dateadd(hour,-5,current\_timestamp()))) ;

-- Error in History

select \* from table(

information\_schema.copy\_history(table\_name => 'promotions', start\_time=> dateadd(hour,-5,current\_timestamp()) )

);

---------------------------------------------------------------------------

-- More on Pipes

desc pipe pipe\_for\_promotions ;

show pipes ;

show pipes like '%promotions%' ;

show pipes in database training ;

show pipes in schema demo;

show pipes like '%promotion%' in schema demo ;

-- Pipes Do not support PURGE COPY Option

-- You need use remove ..to delete the staged files

show stages ;

list @ext\_stage\_ld ;

remove @ext\_stage\_ld/promotions5.csv ;

delete from promotions where p\_promo\_sk between 2001 and 2500 ;

-- How to change the referenced stage url, storage integration and encryption

-- Pause, modify the stage, resume pipe

alter pipe pipe\_for\_promotions set pipe\_execution\_paused = true ;

select system$pipe\_status('pipe\_for\_promotions') ;

alter stage ext\_stage\_ld

set url = 's3://snowflakes3datafiles/snowpipe/' ;

list @ext\_stage\_ld ;

alter pipe pipe\_for\_promotions set pipe\_execution\_paused = false ;

select system$pipe\_status('pipe\_for\_promotions') ;

-- Hence You need to recreate it,

-- Pause, modify the stage, recreate pipe, check status

create or replace pipe pipe\_for\_promotions

auto\_ingest = TRUE as

copy into promotions from @ext\_stage\_ld;

select system$pipe\_status('pipe\_for\_promotions') ;

-- upload promotions5 in file5 folder

-- check the output

select \* from promotions order by 1 desc;

-- how can you change pipe. You can not alter the pipe, you need to recreate

-- for example if the table is renamed, and there is an existing pipe.

-- So we need to recreate the pipe.

alter table promotions rename to raw\_promotions ;

-- truncate the table

truncate table raw\_promotions ;

-- First pause the pipe

alter pipe pipe\_for\_promotions set pipe\_execution\_paused = true ;

--check the status of pipe if it is paused

select system$pipe\_status('pipe\_for\_promotions');

--create the pipe again

create or replace pipe pipe\_for\_promotions

auto\_ingest = true

as

copy into raw\_promotions from @ext\_stage\_ld ;

--check the pipe for files sent

alter pipe pipe\_for\_promotions refresh ;

-- The metadata says that the files are already copied

select system$pipe\_status('pipe\_for\_promotions') ;

-- Data is successfully loaded via refresh

select \* from raw\_promotions order by 1 desc;

show stages ;

-- cleanup

-- remove folder snowpipe ;

-- remove files from snowpipe ;

-- delete event in AWS console ;

drop integration int\_aws\_s3 ;

drop table raw\_promotions ;

drop file format file\_format\_csv ;

drop stage ext\_stage\_ld ;

drop pipe pipe\_for\_promotions ;

-------------------------------------------------------------------------------

--Troubleshooting the Pipe

-- Check the status of pipe

select system$pipe\_status('pipe\_for\_promotions') ;

-- Introduce an error in the promotions4.csv

-- Create a file promotions4error.csv

select \* from promotions order by 1 desc;

-- Error message more generic

select \* from table(validate\_pipe\_load(pipe\_name => 'pipe\_for\_promotions', start\_time=>dateadd(hour,-2,current\_timestamp()))) ;

-- Error in History - more specific

select \* from table(

information\_schema.copy\_history(table\_name => 'raw\_promotions', start\_time=> dateadd(hour,-2,current\_timestamp())))

order by 3 desc;

-- Lets try to refresh and see, if it loads the file once we fixed the error

alter pipe pipe\_for\_promotions refresh ;

-- Two ways to load the files --

-- 1. Manually by copy command OR

copy into promotions from @ext\_stage\_ld files = ('/promotions4err.csv');

select \* from raw\_promotions order by 1 desc;

--OR 2 Rename the File and replace it. Place the New file and wait for Pipe to load

delete from raw\_promotions where p\_promo\_sk between 2000 and 2500 ;

-- check pipe

-- Now remove the error File .

-- Demonstrate how the Pipe refresh will work, if files are already available to load one

-- OR 3 truncate the table and recreate the pipe again and add the errored file again

-- and refresh it

-- Recreating Pipe steps

desc pipe pipe\_for\_promotions ;

-- arn:aws:sqs:us-east-2:079828672386:sf-snowpipe-AIDARFFRI36BLT5VFBWHR-PvH1rcdaWXkPWUDeVlu4sw

-- pause, check status, recreate, pause, check config, resume pipe, check status

truncate table promotions ;

alter pipe pipe\_for\_promotions set pipe\_execution\_paused = true ;

select system$pipe\_status('pipe\_for\_promotions') ;

-- Recreate Pipe

create or replace pipe pipe\_for\_promotions auto\_ingest = TRUE as

copy into promotions from @ext\_stage\_ld;

desc pipe pipe\_for\_promotions ;

select system$pipe\_status('pipe\_for\_promotions') ;

select \* from raw\_promotions order by 1 desc ;

alter pipe pipe\_for\_promotions refresh ;

-- Error message more generic

select \* from table(validate\_pipe\_load(pipe\_name => 'pipe\_for\_promotions', start\_time=>dateadd(hour,-2,current\_timestamp()))) ;

-- Error in History - more specific

select \* from table(

information\_schema.copy\_history(table\_name => 'raw\_promotions', start\_time=> dateadd(hour,-1,current\_timestamp())))

order by 3 desc;

-- Now place the promotion5.csv and see that pipe is working correctly

-- Summary, an errored file shows error correctly.

-- Refresh doesnot load it

-- you can either manually load it or rename a file and load it

-- Refresh functionality how it works for Pipe..by adding an extra file.