Date

loading parquet file

# Goals

In this assignment , we will load parquet file from s3 to snowflake table.

## Understand limitation while copying parquet file.

## Learn to parse parquet file.

# PREPARATION

Let’s upload data as parquet format from snowflake to s3. We will load this data again back to snowflake. During the process we will see different challenges.

Create file format,

CREATE OR REPLACE FILE FORMAT "DEMO\_DB"."PUBLIC".MY\_ PARQUET \_FORMAT

TYPE= PARQUET

COMPRESSIION = 'AUTO'

Upload data to s3 as parquet file,

COPY INTO @DEMO\_DB.PUBLIC.RAVEN\_POC\_STAGE/CUSTOMER\_PARQUET/

FROM

(

SELECT

C\_CUSTOMER\_SK ,

C\_CUSTOMER\_ID ,

C\_CURRENT\_CDEMO\_SK ,

C\_CURRENT\_HDEMO\_SK ,

C\_CURRENT\_ADDR\_SK ,

C\_FIRST\_SHIPTO\_DATE\_SK ,

C\_FIRST\_SALES\_DATE\_SK ,

C\_SALUTATION ,

C\_FIRST\_NAME ,

C\_LAST\_NAME ,

C\_PREFERRED\_CUST\_FLAG ,

C\_BIRTH\_DAY ,

C\_BIRTH\_MONTH ,

C\_BIRTH\_YEAR ,

C\_LOGIN ,

C\_EMAIL\_ADDRESS ,

C\_LAST\_REVIEW\_DATE

FROM DEMO\_DB.PUBLIC.CUSTOMER\_TEST

)

file\_format = DEMO\_DB.PUBLIC.MY\_PARQUET\_FORMAT

Header = true

Overwrite = true

# Query and load data to snowflake table.

**Let’s query uploaded data from snowflake,**

SELECT $1

FROM @DEMO\_DB.PUBLIC.RAVEN\_POC\_STAGE/CUSTOMER\_PARQUET1/

(file\_format => DEMO\_DB.PUBLIC.MY\_PARQUET\_FORMAT)

**Parse file,**

SELECT $1:C\_CUSTOMER\_SK ,

$1:C\_CUSTOMER\_ID ,

$1:C\_CURRENT\_CDEMO\_SK ,

$1:C\_CURRENT\_HDEMO\_SK ,

$1:C\_CURRENT\_ADDR\_SK ,

$1:C\_FIRST\_SHIPTO\_DATE\_SK ,

$1:C\_FIRST\_SALES\_DATE\_SK ,

$1:C\_SALUTATION ,

$1:C\_FIRST\_NAME ,

$1:C\_LAST\_NAME ,

$1:C\_PREFERRED\_CUST\_FLAG ,

$1:C\_BIRTH\_DAY ,

$1:C\_BIRTH\_MONTH ,

$1:C\_BIRTH\_YEAR ,

$1:C\_BIRTH\_COUNTRY ,

$1:C\_LOGIN ,

$1:C\_EMAIL\_ADDRESS ,

$1:C\_LAST\_REVIEW\_DATE

FROM @DEMO\_DB.PUBLIC.RAVEN\_POC\_STAGE/CUSTOMER\_PARQUET/

(file\_format => DEMO\_DB.PUBLIC.MY\_PARQUET\_FORMAT)

**Create table as below,**

create or replace TRANSIENT TABLE CUSTOMER\_PARQUET (

C\_CUSTOMER\_SK NUMBER(5,0),

C\_CUSTOMER\_ID VARCHAR(16),

C\_CURRENT\_CDEMO\_SK NUMBER(38,0),

C\_CURRENT\_HDEMO\_SK NUMBER(38,0),

C\_CURRENT\_ADDR\_SK NUMBER(38,0),

C\_FIRST\_SHIPTO\_DATE\_SK NUMBER(38,0),

C\_FIRST\_SALES\_DATE\_SK NUMBER(38,0),

C\_SALUTATION VARCHAR(10),

C\_FIRST\_NAME VARCHAR(20),

C\_LAST\_NAME VARCHAR(30),

C\_PREFERRED\_CUST\_FLAG VARCHAR(1),

C\_BIRTH\_DAY NUMBER(38,0),

C\_BIRTH\_MONTH NUMBER(38,0),

C\_BIRTH\_YEAR NUMBER(38,0),

C\_LOGIN VARCHAR(13),

C\_EMAIL\_ADDRESS VARCHAR(50),

C\_LAST\_REVIEW\_DATE VARCHAR(10)

);

**Try to copy uploaded data to CUSTOMER\_PARQUET table.**

COPY INTO CUSTOMER\_PARQUET

FROM

(

SELECT

$1:C\_CUSTOMER\_SK,

$1:C\_CUSTOMER\_ID,

$1:C\_CURRENT\_CDEMO\_SK,

$1:C\_CURRENT\_HDEMO\_SK,

$1:C\_CURRENT\_ADDR\_SK,

$1:C\_FIRST\_SHIPTO\_DATE\_SK,

$1:C\_FIRST\_SALES\_DATE\_SK,

$1:C\_SALUTATION,

$1:C\_FIRST\_NAME,

$1:C\_LAST\_NAME,

$1:C\_PREFERRED\_CUST\_FLAG,

$1:C\_BIRTH\_DAY,

$1:C\_BIRTH\_MONTH,

$1:C\_BIRTH\_YEAR,

$1:C\_LOGIN,

$1:C\_EMAIL\_ADDRESS,

$1:C\_LAST\_REVIEW\_DATE

FROM @DEMO\_DB.PUBLIC.RAVEN\_POC\_STAGE/CUSTOMER\_PARQUET/

)

file\_format = DEMO\_DB.PUBLIC.MY\_PARQUET\_FORMAT

ON\_ERROR = 'CONTINUE'

Can you mention for which column you are facing the error.

Column name : Can you

Can you reject bad records using ON\_ERROR copy options ?

Note your observation below,