**BUILD A SNOWFLAKE BULK DATA PIPELINE USING AWS AND EXTERNAL STAGES**

CREATE WAREHOUSE ASSIGNMENT;

USE WAREHOUSE ASSIGNMENT;

CREATE DATABASE DEMO\_DB;

USE DATABASE DEMO\_DB;

USE ROLE ACCOUNTADMIN;

CREATE OR REPLACE TRANSIENT TABLE DEMO\_DB.PUBLIC.ASSIGNMENT

AS

SELECT \* FROM

"SNOWFLAKE\_SAMPLE\_DATA"."TPCDS\_SF100TCL"."CUSTOMER";

select \* from ASSIGNMENT limit 100;

--create csv format

CREATE OR REPLACE FILE FORMAT assign\_csv\_format

TYPE = 'CSV'

FIELD\_OPTIONALLY\_ENCLOSED\_BY = '"'

SKIP\_HEADER = 1;

--integration object

create or replace storage integration assignment

type = external\_stage

storage\_provider = s3

enabled = true

storage\_aws\_role\_arn = 'arn:aws:iam::211125354816:role/jsonrole'

storage\_allowed\_locations = ('s3://s3-bucket-92/jsondata/');

desc integration assignment;

--CREATE STAGE

create or replace stage DEMO\_DB.public.csv\_ext\_assignment\_stage

storage\_integration = assignment

url = 's3://s3-bucket-92/jsondata/'

file\_format = DEMO\_DB.public.assign\_csv\_format;

--copy into s3

COPY INTO @DEMO\_DB.public.csv\_ext\_assignment\_stage/jsonrole/

from DEMO\_DB.PUBLIC.ASSIGNMENT;

SELECT $1 C\_CUSTOMER\_SK,

$2 C\_CUSTOMER\_ID ,

$3 C\_CURRENT\_CDEMO\_SK ,

$4 C\_CURRENT\_HDEMO\_SK ,

$5 C\_CURRENT\_ADDR\_SK,

$6 C\_FIRST\_SHIPTO\_DATE\_SK ,

$7 C\_FIRST\_SALES\_DATE\_SK ,

$8 C\_SALUTATION ,

$9 C\_FIRST\_NAME ,

$10 C\_LAST\_NAME,

$11 C\_PREFERRED\_CUST\_FLAG ,

$12 C\_BIRTH\_DAY ,

$13 C\_BIRTH\_MONTH ,

$14 C\_BIRTH\_YEAR,

$16 C\_LOGIN ,

$17 C\_EMAIL\_ADDRESS ,

$18 C\_LAST\_REVIEW\_DATE

FROM @DEMO\_DB.public.csv\_ext\_assignment\_stage/jsonrole/ ---replace it with new stage

(file\_format => DEMO\_DB.public.assign\_csv\_format);

SELECT $1 C\_CUSTOMER\_SK,

$2 C\_CUSTOMER\_ID ,

$3 C\_CURRENT\_CDEMO\_SK ,

$4 C\_CURRENT\_HDEMO\_SK ,

$5 C\_CURRENT\_ADDR\_SK,

$6 C\_FIRST\_SHIPTO\_DATE\_SK ,

$7 C\_FIRST\_SALES\_DATE\_SK ,

$8 C\_SALUTATION ,

$9 C\_FIRST\_NAME ,

$10 C\_LAST\_NAME,

$11 C\_PREFERRED\_CUST\_FLAG ,

$12 C\_BIRTH\_DAY ,

$13 C\_BIRTH\_MONTH ,

$14 C\_BIRTH\_YEAR,

$16 C\_LOGIN ,

$17 C\_EMAIL\_ADDRESS ,

$18 C\_LAST\_REVIEW\_DATE

FROM @DEMO\_DB.public.csv\_ext\_assignment\_stage/jsonrole/ ---replace it with new stage

(file\_format => DEMO\_DB.public.assign\_csv\_format)

WHERE C\_CUSTOMER\_SK ='64596949';

SELECT $9 C\_FIRST\_NAME,$10 C\_LAST\_NAME,COUNT(\*)

FROM @DEMO\_DB.public.csv\_ext\_assignment\_stage/jsonrole/

(file\_format => DEMO\_DB.public.assign\_csv\_format)

GROUP BY $9,$10;

--creating view

CREATE OR REPLACE VIEW ASSIGN\_CUST\_DATA\_VIEW

AS

SELECT $1 C\_CUSTOMER\_SK,

$2 C\_CUSTOMER\_ID ,

$3 C\_CURRENT\_CDEMO\_SK ,

$4 C\_CURRENT\_HDEMO\_SK ,

$5 C\_CURRENT\_ADDR\_SK,

$6 C\_FIRST\_SHIPTO\_DATE\_SK ,

$7 C\_FIRST\_SALES\_DATE\_SK ,

$8 C\_SALUTATION ,

$9 C\_FIRST\_NAME ,

$10 C\_LAST\_NAME,

$11 C\_PREFERRED\_CUST\_FLAG ,

$12 C\_BIRTH\_DAY ,

$13 C\_BIRTH\_MONTH ,

$14 C\_BIRTH\_YEAR,

$16 C\_LOGIN ,

$17 C\_EMAIL\_ADDRESS ,

$18 C\_LAST\_REVIEW\_DATE

FROM @DEMO\_DB.public.csv\_ext\_assignment\_stage/jsonrole/

(file\_format => DEMO\_DB.public.assign\_csv\_format);

--

SELECT \* FROM ASSIGN\_CUST\_DATA\_VIEW;

Create or replace transient table CUST\_SNOWFLAKE\_TABLE

AS

SELECT \* FROM ASSIGNMENT limit 10000;

SELECT B.\*

FROM CUST\_SNOWFLAKE\_TABLE B

LEFT OUTER JOIN

ASSIGN\_CUST\_DATA\_VIEW A

ON

A.C\_CUSTOMER\_SK = B.C\_CUSTOMER\_SK;

COPY INTO @DEMO\_DB.public.csv\_ext\_assignment\_stage/Customer\_joined\_data/

from(

SELECT B.\*

FROM CUST\_SNOWFLAKE\_TABLE B

LEFT OUTER JOIN

ASSIGN\_CUST\_DATA\_VIEW A

ON

A.C\_CUSTOMER\_SK = B.C\_CUSTOMER\_SK

);