Here is the code that has been included in the File path within Figure 10-9

poc/raw/@{item().SRC\_DB}/@{item().SRC\_SCHEMA}/@{item().DST\_TAB\_NAME}/@{formatDateTime(utcnow(),'yyyy-MM-dd')}/ @{item().SAT\_TAB\_NAME}

Below is the script used to create the control and audit tables in Snowflake. Be sure to run the script in the ETL schema of your ADF\_DB database. The script will create two tables: PIPELINE\_CTL and AUDIT\_TAB.

CREATE OR replace TABLE pipeline\_ctl

(

parameter\_id number(38,0) NOT NULL autoincrement,

server\_name varchar(500),

src\_type varchar(500),

src\_schema varchar(500),

src\_db varchar(500),

src\_tab\_name varchar(500),

dst\_type varchar(500),

dst\_tab\_name varchar(500),

include\_pipeline\_flag varchar(500),

process\_type varchar(500),

load\_snowflake varchar(500),

load\_frequency varchar(500),

dst\_folder varchar(500),

file\_type varchar(500),

lake\_dst\_folder varchar(500),

dst\_schema varchar(500),

distribution\_type varchar(500),

asql\_to\_lake\_pipeline\_date timestamp\_ntz(9),

asql\_to\_lake\_pipeline\_status varchar(500),

load\_snow\_etl\_pipeline\_date timestamp\_ntz(9),

load\_snow\_etl\_pipeline\_status varchar(500),

load\_snow\_curated\_pipeline\_date timestamp\_ntz(9),

load\_snow\_curated\_pipeline\_status varchar(500),

load\_delta\_pipeline\_date timestamp\_ntz(9),

load\_delta\_pipeline\_status varchar(500),

partition\_field varchar(500),

priority\_lane varchar(500),

spark\_flag varchar(500),

swim\_lane int,

PRIMARY KEY (parameter\_id)

);

CREATE OR replace TABLE audit\_tab

(

pipeline\_name varchar(100),

db\_name varchar(20),

sch\_name varchar(20),

table\_name varchar(50),

source\_count number(38,0),

adls\_count number(38,0),

snowflake\_count number(38,0),

load\_time timestamp\_ntz(9) DEFAULT CURRENT\_TIMESTAMP()

);

Here is the code that has been added to the Source Query section in Figure 10-20

SELECT \* FROM @item().SRC\_SCHEMA}.@item().DST\_TAB\_NAME}

Here is the code that is being used for the Base parameters’ values in the settings section of the Databricks notebook shown in Figure 10-28.

@item().DST\_TAB\_NAME}

@{item().DST\_SCHEMA}

raw/AdventureWorksLT2019/SALESLT/@{item().DST\_TAB\_NAME}/@{formatDateTime(utcnow(),'yyyy')}-@{formatDateTime(utcnow(),'MM')}-@{formatDateTime(utcnow(),'dd')}/

Within Databricks, the notebook would contain the following Scala code, which accepts the Parameters from the ADF copy activity dynamically and then passes it to a data frame which reads the parquet file based on the dynamic parameters and then writes it to the Snowflake table (Figure 10-29).

import org.apache.spark.sql.{SaveMode, SparkSession}

spark.conf.set(

"fs.azure.account.key.adl001.dfs.core.windows.net",

"ENTER-ACCOUNT-KEY-HERE"

)

val DST\_TAB\_NAME = dbutils.widgets.get("DST\_TAB\_NAME")

val DST\_SCHEMA = dbutils.widgets.get("DST\_SCHEMA")

val FOLDER\_PATH = dbutils.widgets.get("FOLDER\_PATH")

var options = Map(

"sfUrl" -> "px.east-us-2.azure.snowflakecomputing.com",

"sfUser" -> "USERNAME",

"sfPassword" -> "PW",

"sfDatabase" -> "ADVENTUREWORKS",

"sfSchema" -> DST\_SCHEMA,

"truncate\_table" -> "ON",

"usestagingtable" -> "OFF",

"sfWarehouse" -> "COMPUTE\_WH"

)

val df = spark.read.parquet("abfss://poc@gze2np1adl001.dfs.core.windows.net/"+FOLDER\_PATH+DST\_TAB\_NAME)

df.write

.format("snowflake")

.options(options)

.option("dbtable", DST\_TAB\_NAME)

.mode(SaveMode.Overwrite)

.save()

The following code will update the table to create the swim\_lane column.

UPDATE adf\_db.etl.pipeline\_ctl PL1  
SET    swim\_lane = 2  
FROM   (SELECT src\_tab\_name,  
               Row\_number()  
                 OVER (  
                   ORDER BY src\_tab\_name ) AS rn  
        FROM   adf\_db.etl.pipeline\_ctl) b  
WHERE  PL1.src\_tab\_name = b.src\_tab\_name  
       AND b.rn > 5

The code below can be used to create the log table in Snowflake DW:

CREATE OR replace TABLE audit\_tab

(

pipeline\_name varchar(100),

src\_db\_name varchar(20),

dest\_db\_name varchar(20),

sch\_name varchar(20),

table\_name varchar(50),

source\_count number(38,0),

adls\_count number(38,0),

snowflake\_count number(38,0),

load\_time timestamp\_ntz(9) DEFAULT CURRENT\_TIMESTAMP()

);