LOADING MULTIPLE CSV FILES FROM AWS S3 INTO SNOWFLAKE TABLES VIA MATILLION

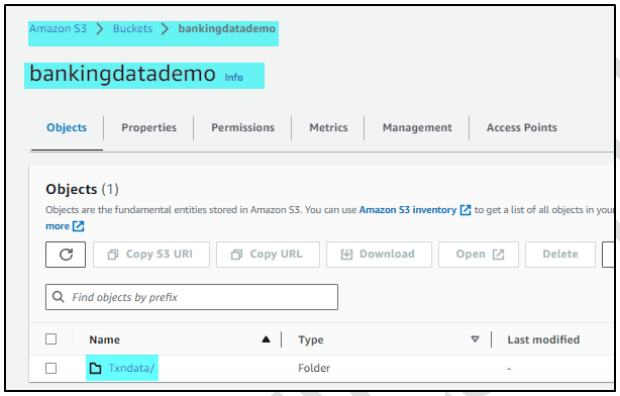
Uploading multiple files from S3 to Snowflake thru matillion involves below main steps:

1. Load files in S3
2. In matillion,
    a. Create a orchestration job
    b. create a table in snowflake thru matillion (need a table to update data)
    c. add S3 load (loads data from S3 to snowflake)
    d. add 'fixed iterator' (so we can iterate and work thru each file indivisually)
    e. run job
3. verify if data is loaded

Detailed steps:

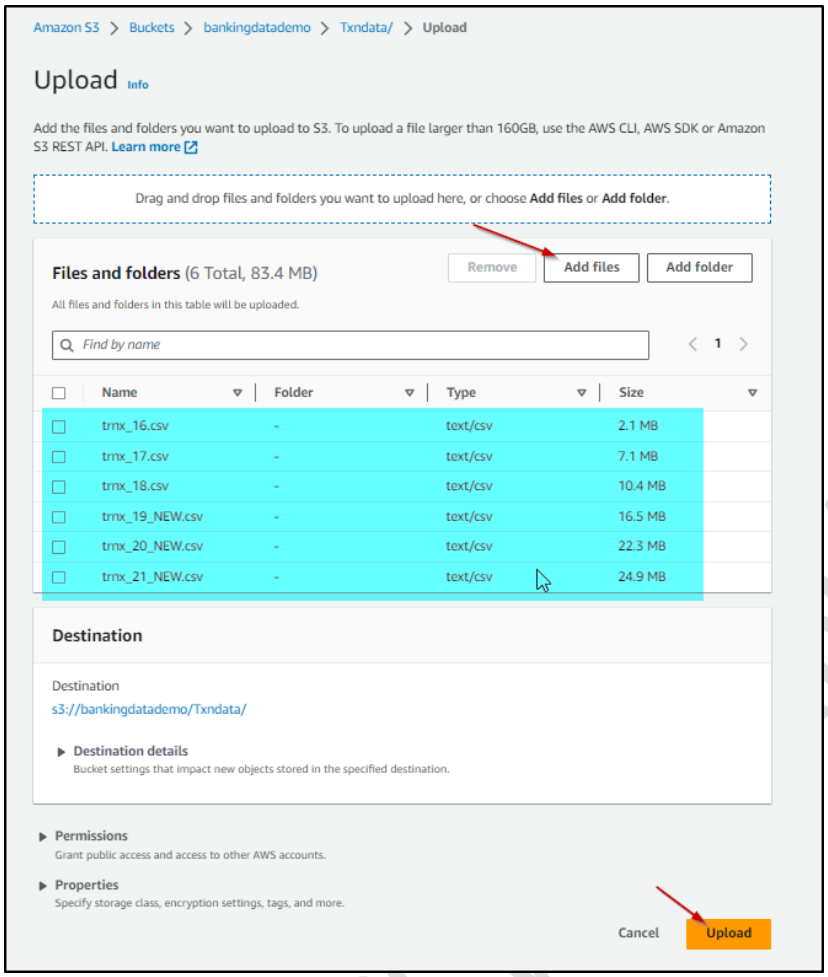
STEP1: Load files in S3:

1) Download the BANKING_DATA
2) Upload in S3:
    a. Create a S3 bucket (here: bankingdatademo)
    b. Create a folder for transaction data (here:)

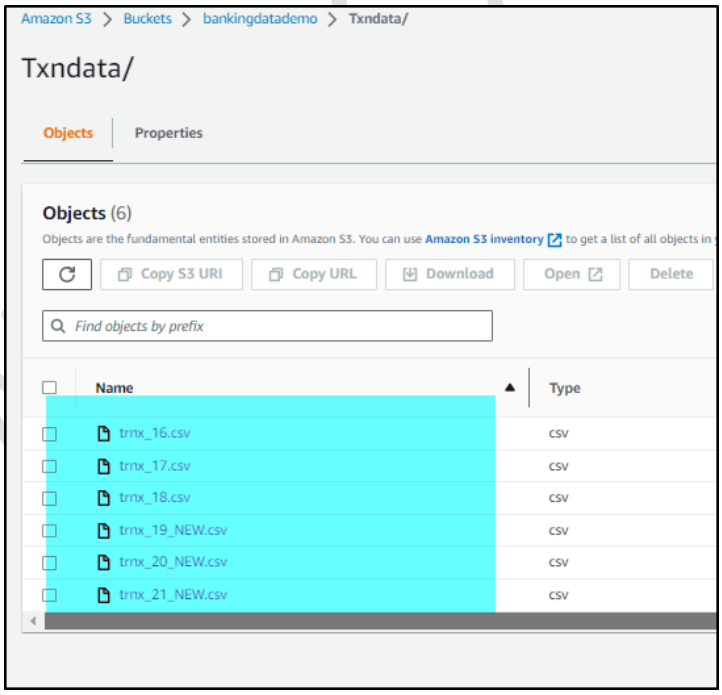


    c. Upload the files: Click on upload button > add files button > select files needed > click 'upload'

LOADING MULTIPLE CSV FILES FROM AWS S3 INTO SNOWFLAKE TABLES VIA MATILLION

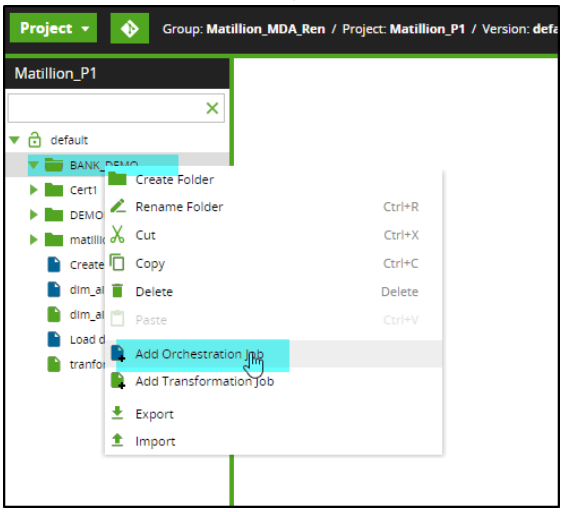


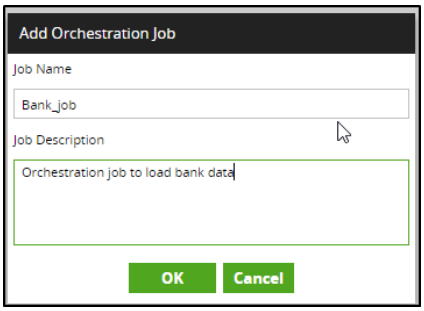
d. Once uploaded, you will be able to see the files in the s3 path:

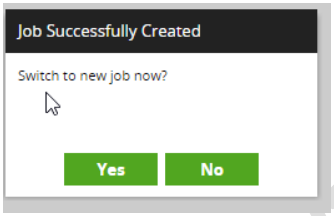


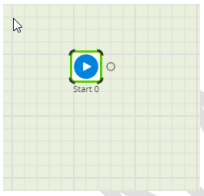
LOADING MULTIPLE CSV FILES FROM AWS S3 INTO SNOWFLAKE TABLES VIA MATILLION

STEP2a: Create a orchestration job





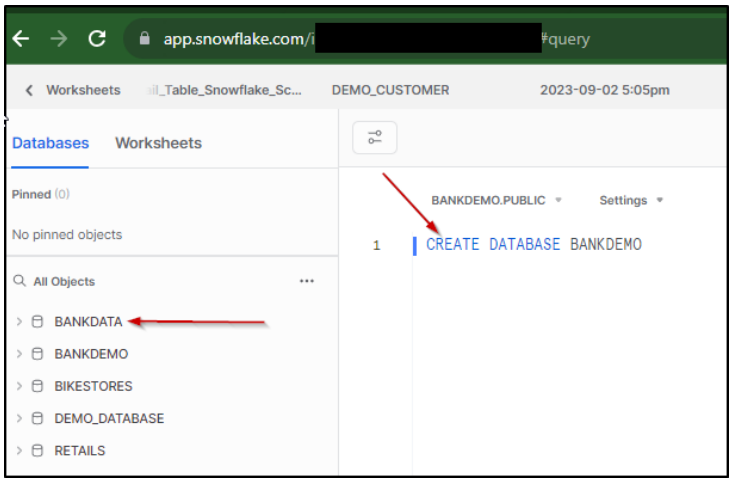




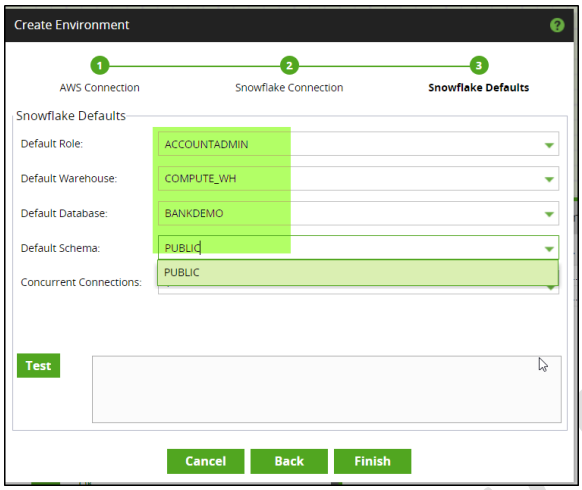
STEP2b: create a table in snowflake thru matillion (need a table to update data)

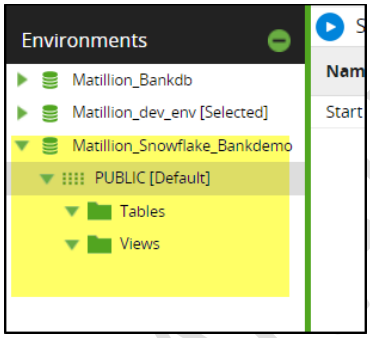
Tables need to be created in a database in snowflake. Either use an existing database or create database in snowflake. Here database ‘bankdata’ is created.

LOADING MULTIPLE CSV FILES FROM AWS S3 INTO SNOWFLAKE TABLES VIA MATILLION



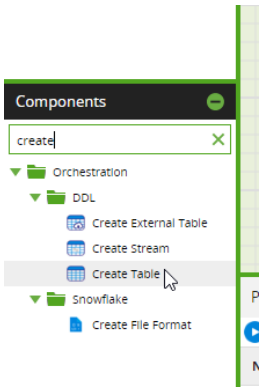
If a new database was created, create an environment to point to it.



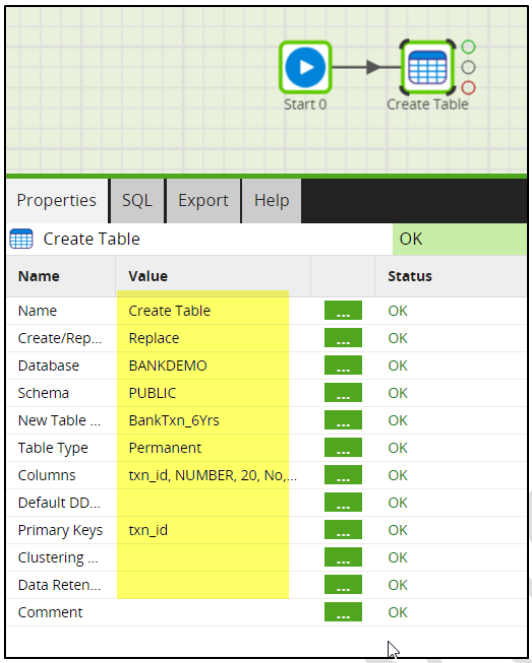


Add ‘create table’ component

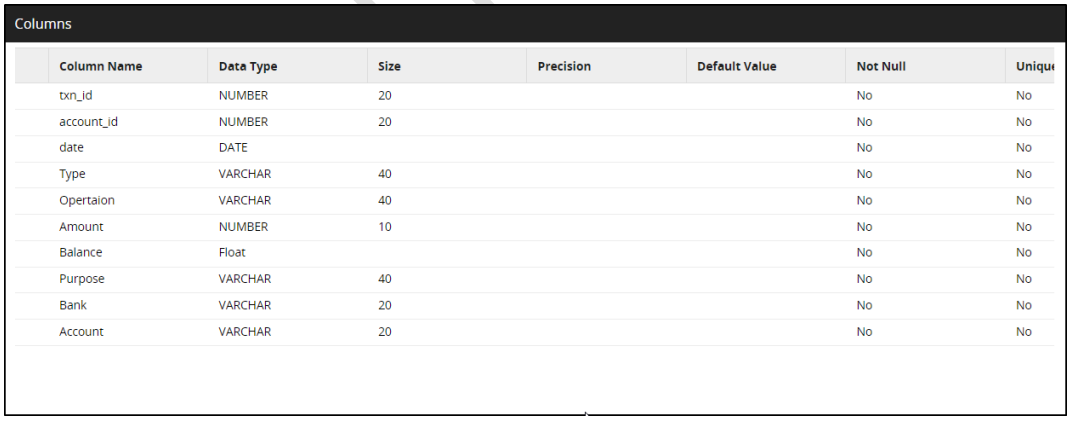
LOADING MULTIPLE CSV FILES FROM AWS S3 INTO SNOWFLAKE TABLES VIA MATILLION



Update properties accordingly:

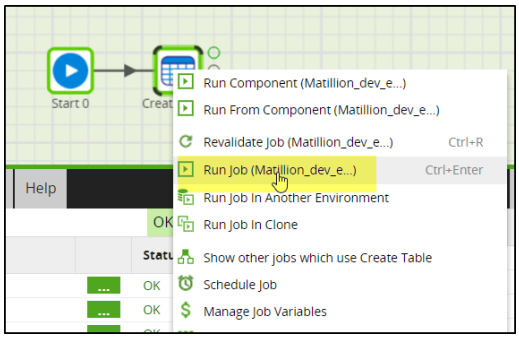


NOTE: Columns created here:



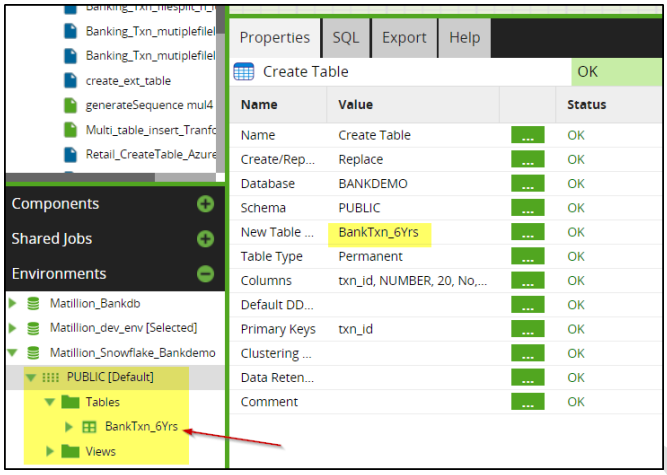
Right click on the create table component and run the job

LOADING MULTIPLE CSV FILES FROM AWS S3 INTO SNOWFLAKE TABLES VIA MATILLION



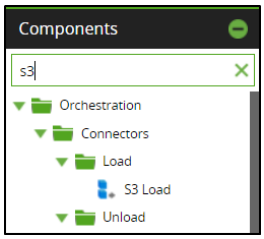
Job should run successfully and the new table should be listed in the environment as well in snowflake.

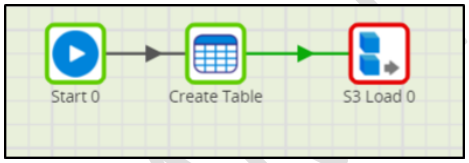
NOTE: If the table is not seen in the environment created, then refresh it.



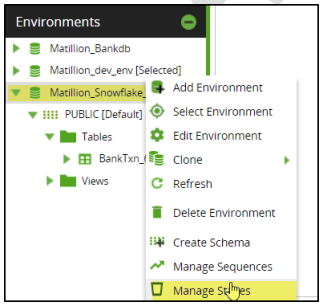
STEP2c: add S3 load (loads data from S3 to snowflake)

Create S3 load component



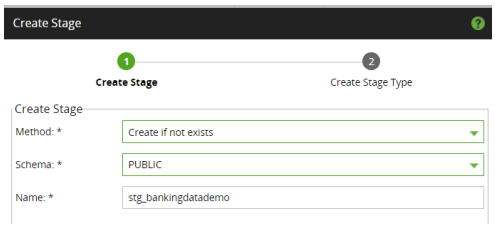


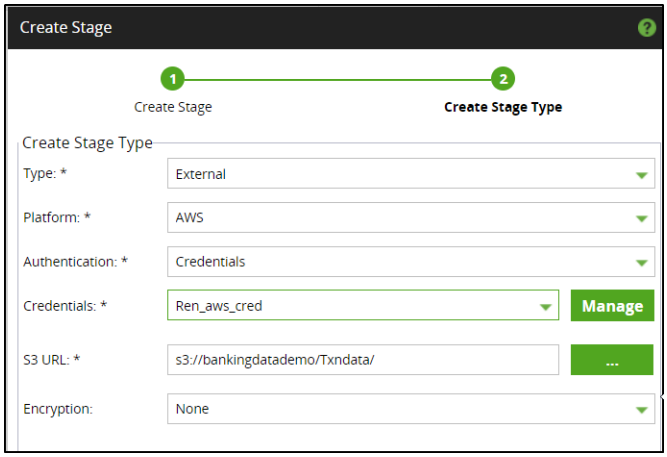
Create a stage : right click > manage stage

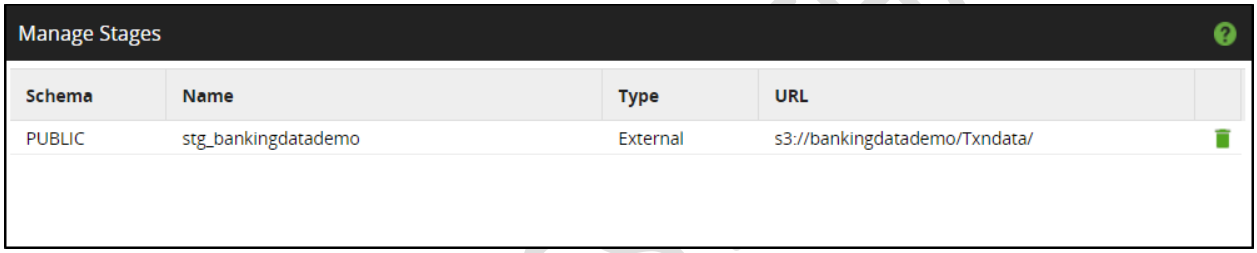


Enter appropriate values

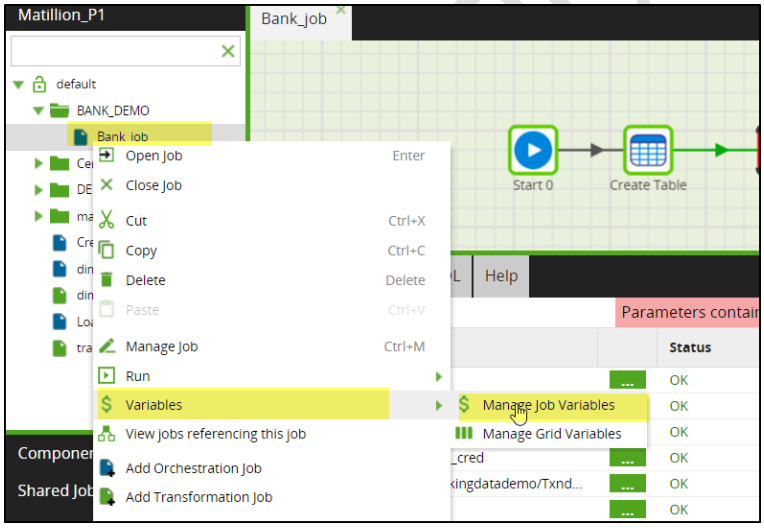
LOADING MULTIPLE CSV FILES FROM AWS S3 INTO SNOWFLAKE TABLES VIA MATILLION







Once stage is created, create a variable to hold the file name



Update values in each field:

Value here is the name of the first file in S3 that is expected to be read

LOADING MULTIPLE CSV FILES FROM AWS S3 INTO SNOWFLAKE TABLES VIA MATILLION

Manage Job Variables

Name	Type	Behaviour	Visibility	Value	Description
dictionary_name	Text	Copied	Public	Txndata	
file_name	Text	Shared	Public	trnx_16	

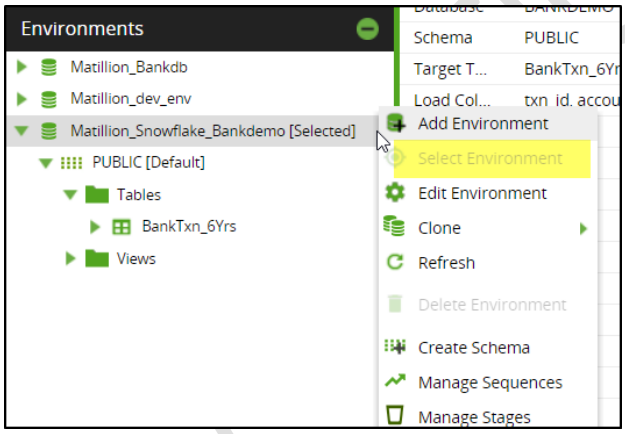
fill in the S3 load properties:

Properties | Export | SQL | Help

S3 Load — OK

Name	Value
Name	S3 Load
Stage	stg_bankingdatademo
Pattern	${dictionary_name}/${file_name}.csv
Encryption	None
Warehouse	COMPUTE_WH
Database	BANKDEMO
Schema	PUBLIC
Target Table	BankTxn_6Yrs
Load Columns	txn_id, account_id, date, Type, Opertaion, Amount, Balance, Purp
Format	[Custom]
File Type	CSV
Compression	AUTO
Record Delimiter	
Field Delimiter	,
Skip Header	1
Skip Blank Lines	False

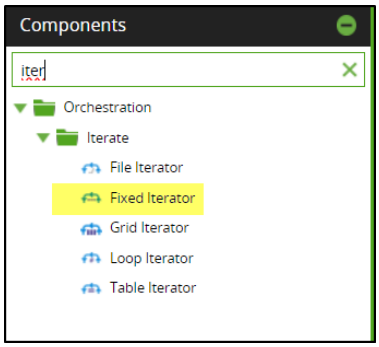
NOTE: If the newly created stage is not listing, verify the environment that is selected. To switch environments to newly created env, right click and 'select environment'

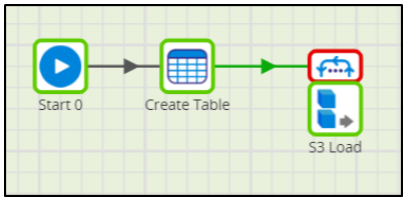


STEP2d: add 'fixed iterator' (so we can iterate and work thru each file individually)

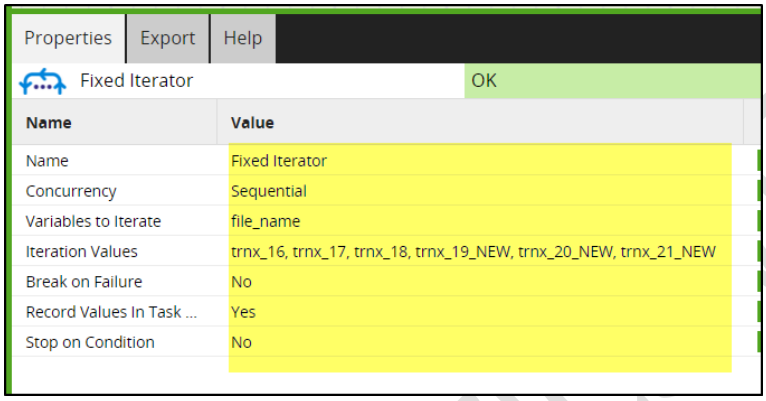
Add a 'fixed iterator ' component

LOADING MULTIPLE CSV FILES FROM AWS S3 INTO SNOWFLAKE TABLES VIA MATILLION

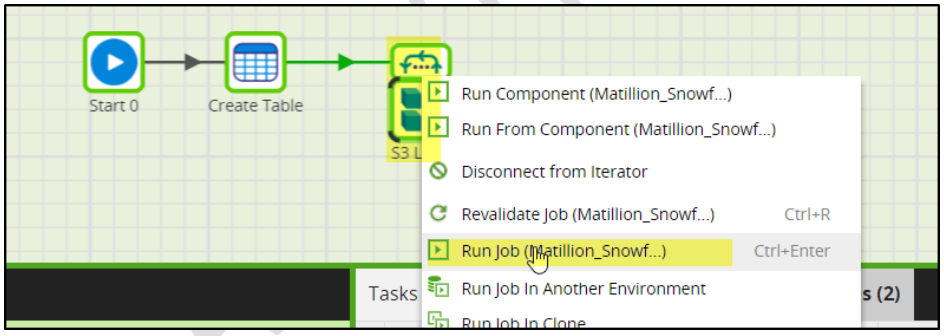




Update iterator properties:

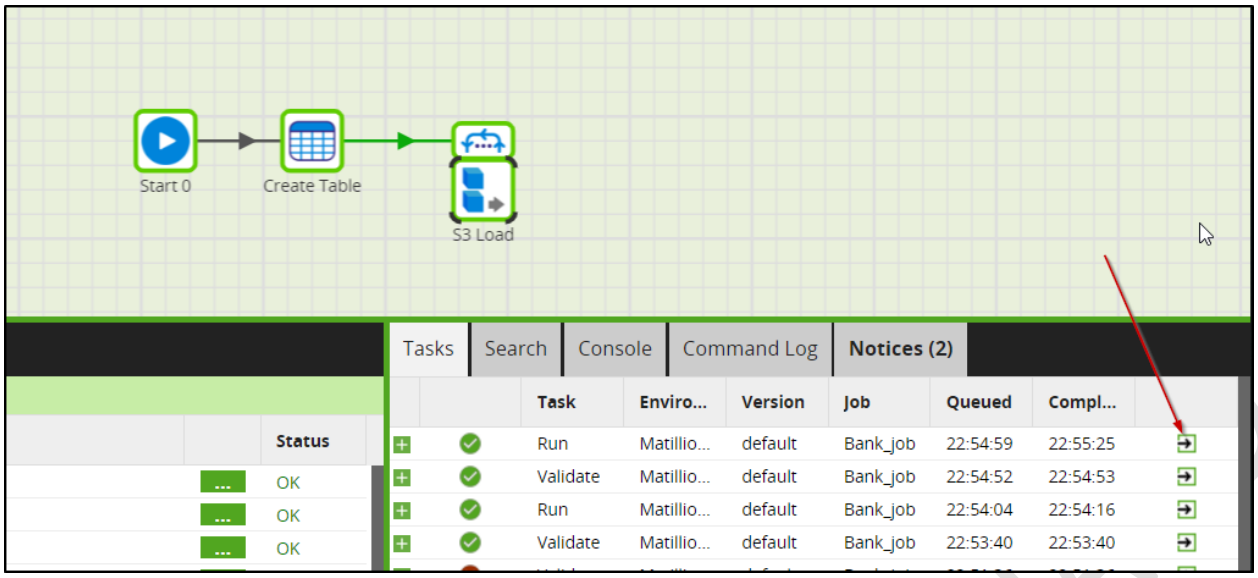


STEP2e: run job



Once the job is run, expand the job details

LOADING MULTIPLE CSV FILES FROM AWS S3 INTO SNOWFLAKE TABLES VIA MATILLION



Job has been run successfully. Row count and file names are listed.