



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

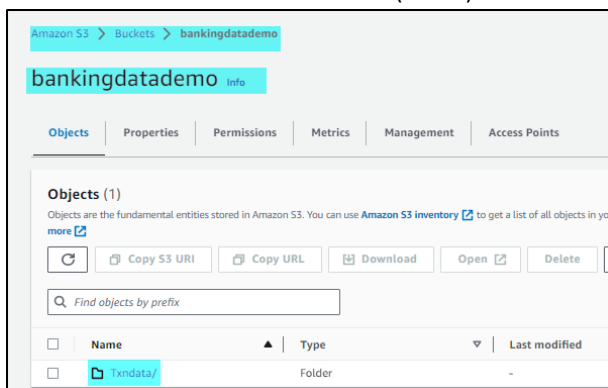
Uploading multiple files from S3 to Snowflake thru matillion involves below main steps:

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

Amazon S3 > Buckets > bankingdatademo > Txndata/ > Upload

Upload

Info

Add the files and folders you want to upload to S3. To upload a file larger than 160GB, use the AWS CLI, AWS SDK or Amazon S3 REST API. [Learn more](#)

Drag and drop files and folders you want to upload here, or choose [Add files](#) or [Add folder](#).

Files and folders (6 Total, 83.4 MB)

Remove

Add files

Add folder

All files and folders in this table will be uploaded.

Find by name

< 1 >

<input type="checkbox"/>	Name	Folder	Type	Size
<input type="checkbox"/>	trnx_16.csv	-	text/csv	2.1 MB
<input type="checkbox"/>	trnx_17.csv	-	text/csv	7.1 MB
<input type="checkbox"/>	trnx_18.csv	-	text/csv	10.4 MB
<input type="checkbox"/>	trnx_19_NEW.csv	-	text/csv	16.5 MB
<input type="checkbox"/>	trnx_20_NEW.csv	-	text/csv	22.3 MB
<input type="checkbox"/>	trnx_21_NEW.csv	-	text/csv	24.9 MB

Destination

Destination

s3://bankingdatademo/Txndata/

Destination details

Bucket settings that impact new objects stored in the specified destination.

Permissions

Grant public access and access to other AWS accounts.

Properties

Specify storage class, encryption settings, tags, and more.

Cancel

Upload

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

Amazon S3 > Buckets > bankingdatademo > Txndata/

Txndata/

Objects

Properties

Objects (6)

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in

Copy S3 URI

Copy URL

Download

Open

Delete

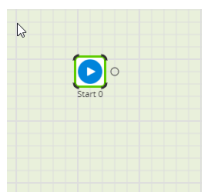
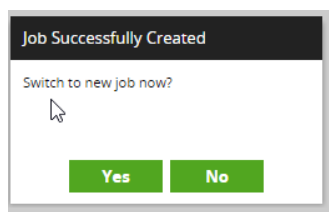
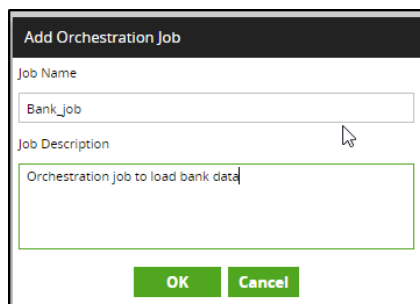
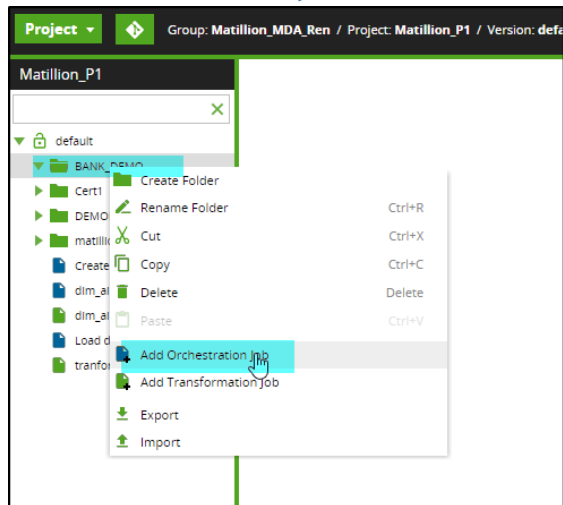
Find objects by prefix

<input type="checkbox"/>	Name	Type
<input type="checkbox"/>	trnx_16.csv	csv
<input type="checkbox"/>	trnx_17.csv	csv
<input type="checkbox"/>	trnx_18.csv	csv
<input type="checkbox"/>	trnx_19_NEW.csv	csv
<input type="checkbox"/>	trnx_20_NEW.csv	csv
<input type="checkbox"/>	trnx_21_NEW.csv	csv



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

STEP2a: Create a orchestration job

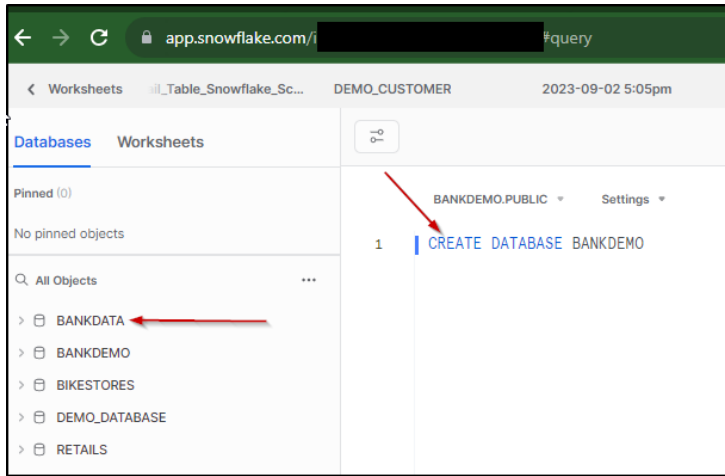


STEP2b: create a table in snowflake thru matillion (need a table to update data)

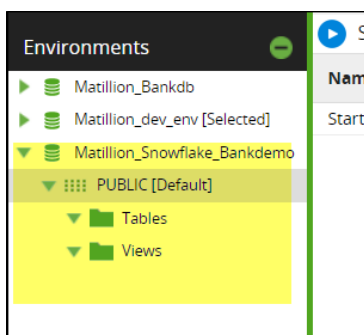
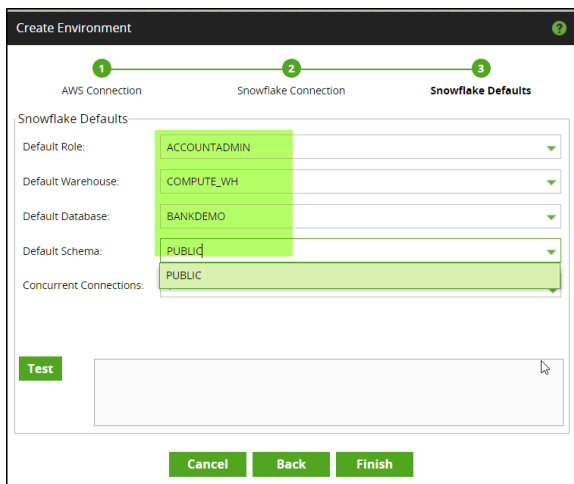
Tables need to be created in a database in snowflake. Either use an existing database or create database in snowflake. Here database 'bankdata' is created.



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



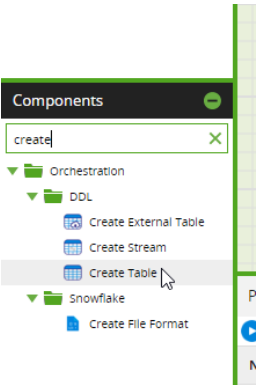
If a new database was created, create an environment to point to it.



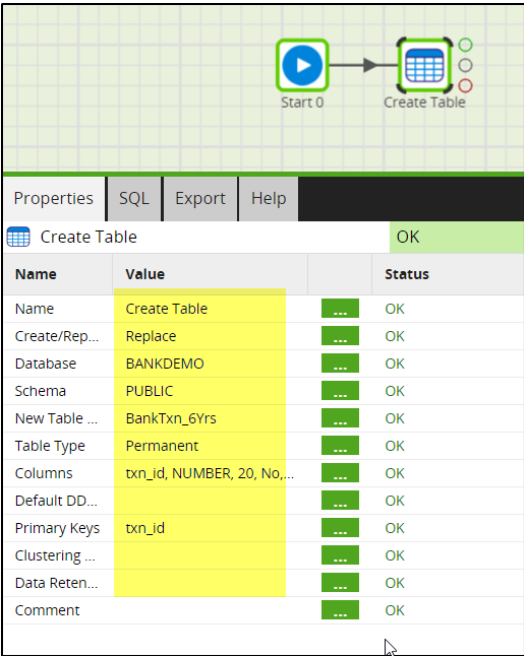
Add 'create table' component



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



Update properties accordingly:



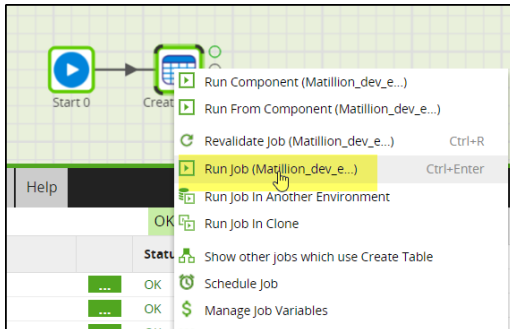
NOTE: Columns created here:

Columns							
Column Name	Data Type	Size	Precision	Default Value	Not Null	Unique	
txn_id	NUMBER	20			No	No	
account_id	NUMBER	20			No	No	
date	DATE				No	No	
Type	VARCHAR	40			No	No	
Opertaion	VARCHAR	40			No	No	
Amount	NUMBER	10			No	No	
Balance	Float				No	No	
Purpose	VARCHAR	40			No	No	
Bank	VARCHAR	20			No	No	
Account	VARCHAR	20			No	No	

Right click on the create table component and run the job

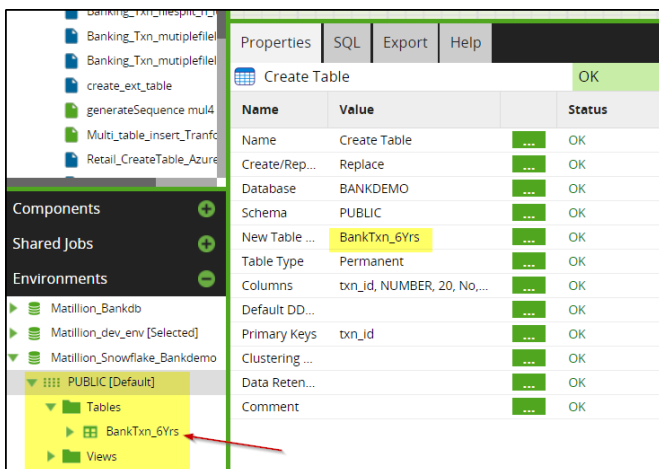


LOADING MULTIPLE CSV FILES FROM AWS S3 INTO SNOWFLAKE TABLES VIA MATILLION



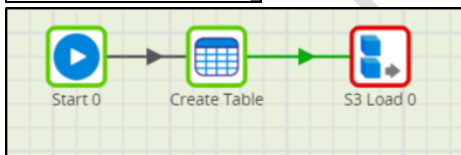
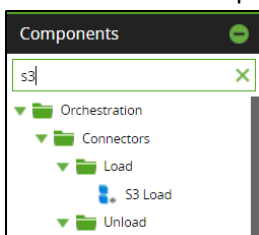
Job should run successfully and the new table should be listed in the environment as well in snowflake.

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

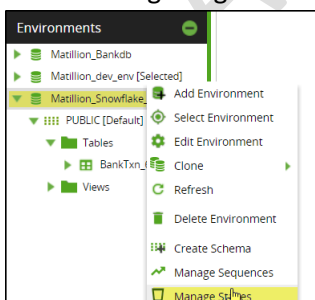


STEP2c: add S3 load (loads data from S3 to snowflake)

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

Create Stage

1

2

Create Stage

Create Stage Type

Method: *
Create if not exists

Schema: *
PUBLIC

Name: *
stg_bankingdatademo

Create Stage

1

2

Create Stage Type

Type: *
External

Platform: *
AWS

Authentication: *
Credentials

Credentials: *
Ren_aws_cred

S3 URL: *
s3://bankingdatademo/Txndata/

Encryption: *
None

Manage Stages

Schema	Name	Type	URL
PUBLIC	stg_bankingdatademo	External	s3://bankingdatademo/Txndata/

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

Matillion_P1

Bank Job

Bank Job

Start 0

Create Table

Parameters contain

Status

OK

OK

OK

OK

OK

OK

Update values in each field:
Value here is the name of the first file in S3 that is expected to be read



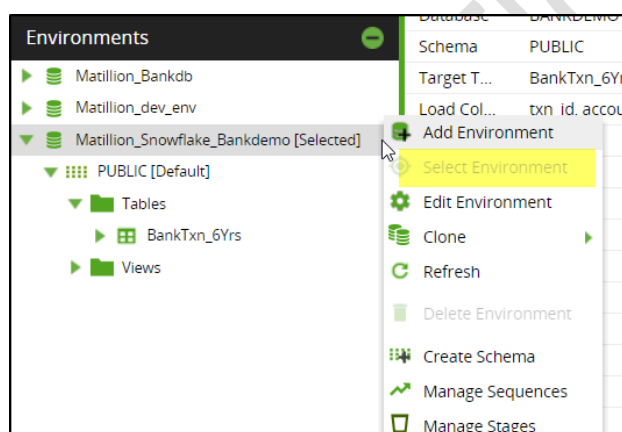
LOADING MULTIPLE CSV FILES FROM AWS S3 INTO SNOWFLAKE TABLES VIA MATILLION

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				
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'

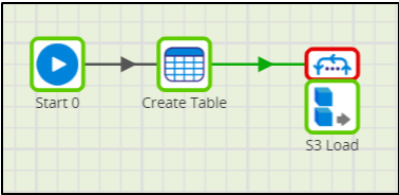
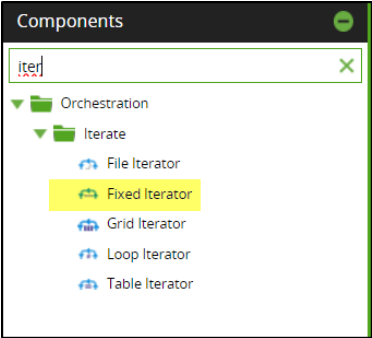


STEP2d: add 'fixed iterator' (so we can iterate and work thru each file individually)

Add a 'fixed iterator ' component



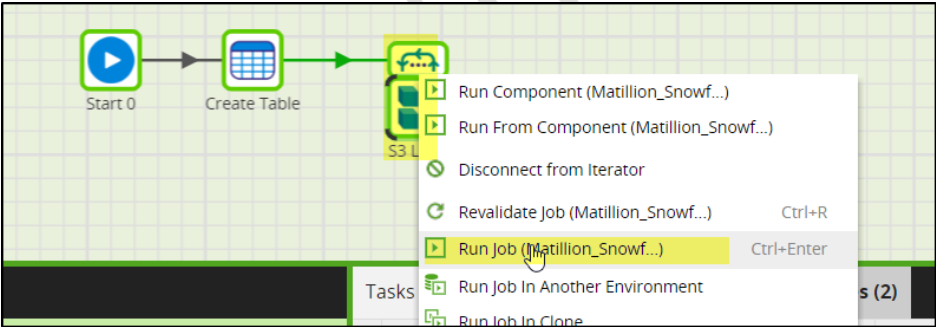
LOADING MULTIPLE CSV FILES FROM AWS S3 INTO SNOWFLAKE TABLES VIA MATILLION



Update iterator properties:

Properties		Export	Help
Fixed Iterator		OK	
Name	Value		
Name	Fixed Iterator		
Concurrency	Sequential		
Variables to Iterate	file_name		
Iteration Values	trnx_16, trnx_17, trnx_18, trnx_19_NEW, trnx_20_NEW, trnx_21_NEW		
Break on Failure	No		
Record Values In Task ...	Yes		
Stop on Condition	No		

STEP2e: run job



Once the job is run, expand the job details



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

Start 0

Create Table

S3 Load

Tasks

Search

Console

Command Log

Notices (2)

	Status	Task	Enviro...	Version	Job	Queued	Compl...	
	OK	Run	Matillio...	default	Bank_job	22:54:59	22:55:25	
	OK	Validate	Matillio...	default	Bank_job	22:54:52	22:54:53	
	OK	Run	Matillio...	default	Bank_job	22:54:04	22:54:16	
	OK	Validate	Matillio...	default	Bank_job	22:53:40	22:53:40	

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

Bank_job	Task - Bank_job							View Jobs
Environment: Matillion_Snowflake_ Version: default		Queued: 22:54:59		Duration: 25.7s				
Job	Component	Duration	Queued	Started	Completed	Row C...	Message	
Bank_job	Create Table	1.2s	22:54:59	22:54:59	22:55:00		Created table ["BANKDEMO"."PUBLIC"."BankTxn_6Yrs"]	
Bank_job	Fixed Iterator	24.5s	22:55:00	22:55:00	22:55:25		6 iterations generated.	
Bank_job	S3 Load	3.8s	22:55:00	22:55:00	22:55:04		file_name = trnx_16	
Bank_job	S3 Load	3.8s	22:55:00	22:55:00	22:55:04	28205		
Bank_job	S3 Load	3.6s	22:55:04	22:55:04	22:55:07		file_name = trnx_17	
Bank_job	S3 Load	3.6s	22:55:04	22:55:04	22:55:07	91628		
Bank_job	S3 Load	3.2s	22:55:07	22:55:07	22:55:11		file_name = trnx_18	
Bank_job	S3 Load	3.2s	22:55:07	22:55:07	22:55:11	133022		
Bank_job	S3 Load	4.4s	22:55:11	22:55:11	22:55:15		file_name = trnx_19_NEW	
Bank_job	S3 Load	4.4s	22:55:11	22:55:11	22:55:15	196779		
Bank_job	S3 Load	4.9s	22:55:15	22:55:15	22:55:20		file_name = trnx_20_NEW	
Bank_job	S3 Load	4.9s	22:55:15	22:55:15	22:55:20	284409		
Bank_job	S3 Load	4.7s	22:55:20	22:55:20	22:55:25		file_name = trnx_21_NEW	
Bank_job	S3 Load	4.7s	22:55:20	22:55:20	22:55:25	314532		