

# RETAIL SALES DATA ANALYSIS PROJECT

## SNOWFLAKE CONTINUOUS DATA LOADING

- 1]. Create an AWS account in [aws.amazon.com](https://aws.amazon.com)
- 2]. After successful account creation and activation, you can use the AWS service.
- 3]. Go to the Console home and search for S3 (Simple Storage Service) and click on it.

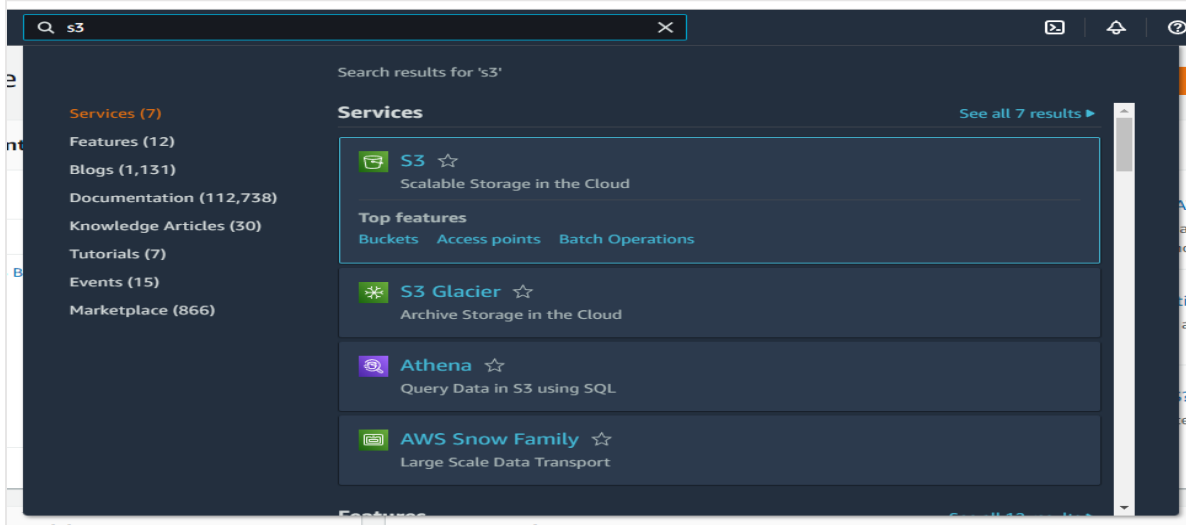


Figure 1: s3 bucket searching

### 4]. Create S3 bucket

Bucket name: **sk-retail-bucket**

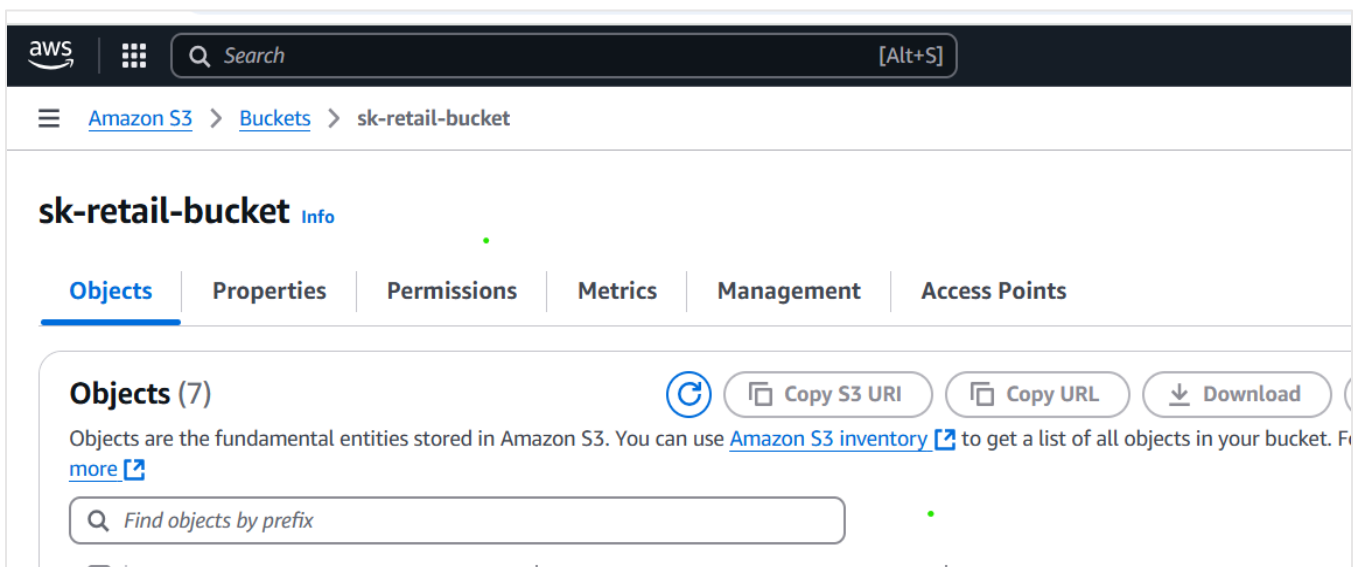


Figure 2: S3 bucket creation

## 5]. Create a folder inside the bucket ( e.g. DEMOGRAPHIC)

**Create folder** [Info](#)

Use folders to group objects in buckets. When you create a folder, S3 creates an object using the name that you specify followed by a slash (/). This object then appears as folder on

**ⓘ Your bucket policy might block folder creation**  
If your bucket policy prevents uploading objects without specific tags, metadata, or access control list (ACL) grantees, you will not be able to create a folder using this configuration to upload an empty folder and specify the appropriate settings.

**Folder**

**Folder name**

DEMOGRAPHIC/

Folder names can't contain "/". [See rules for naming](#)

**Server-side encryption** [Info](#)

Server-side encryption protects data at rest.

**ⓘ** The following encryption settings apply only to the folder object and not to sub-folder objects.

**Server-side encryption**

☒ **Don't specify an encryption key**  
The bucket settings for default encryption are used to encrypt the folder object when storing it in Amazon S3.

☐ **Specify an encryption key**  
The specified encryption key is used to encrypt the folder object before storing it in Amazon S3.

Figure 3: Folder creation inside bucket

For this project we created 7 folders in our bucket:

- 1) CAMPAIGN\_DSC
- 2) CAMPAIGN
- 3) COUPON\_REDEMPT
- 4) COUPON
- 5) DEMOGRAPHIC
- 6) PRODUCT
- 7) TRANSACTION

**sk-retail-bucket** [Info](#)

[Objects](#) | [Properties](#) | [Permissions](#) | [Metrics](#) | [Management](#) | [Access Points](#)

**Objects (7)** [Refresh](#) [Copy S3 URI](#) [Copy URL](#) [Download](#) [Open](#)

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

<input type="checkbox"/>	Name	Type	Last modified	Size
<input type="checkbox"/>	<a href="#">CAMPAIGN_DESC/</a>	Folder	-	
<input type="checkbox"/>	<a href="#">CAMPAIGN/</a>	Folder	-	
<input type="checkbox"/>	<a href="#">COUPON_REDEMPT/</a>	Folder	-	
<input type="checkbox"/>	<a href="#">COUPON/</a>	Folder	-	
<input type="checkbox"/>	<a href="#">DEMOGRAPHIC/</a>	Folder	-	
<input type="checkbox"/>	<a href="#">PRODUCT/</a>	Folder	-	
<input type="checkbox"/>	<a href="#">TRANSACTION/</a>	Folder	-	

Figure 4: All folders

6]. Once the S3 bucket and folders are created, search and select the IAM (Identity and Access Management) service from the AWS console.

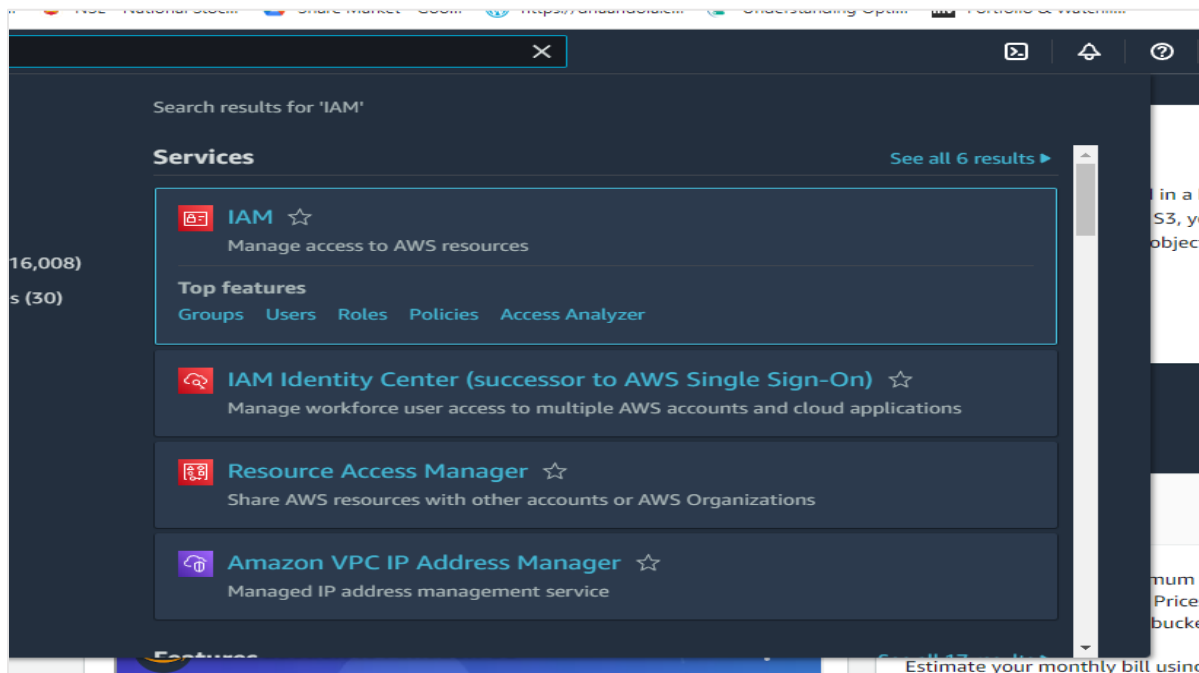


Figure 5: IAM role in AWS

7]. Click on the Policies from IAM Dashboard

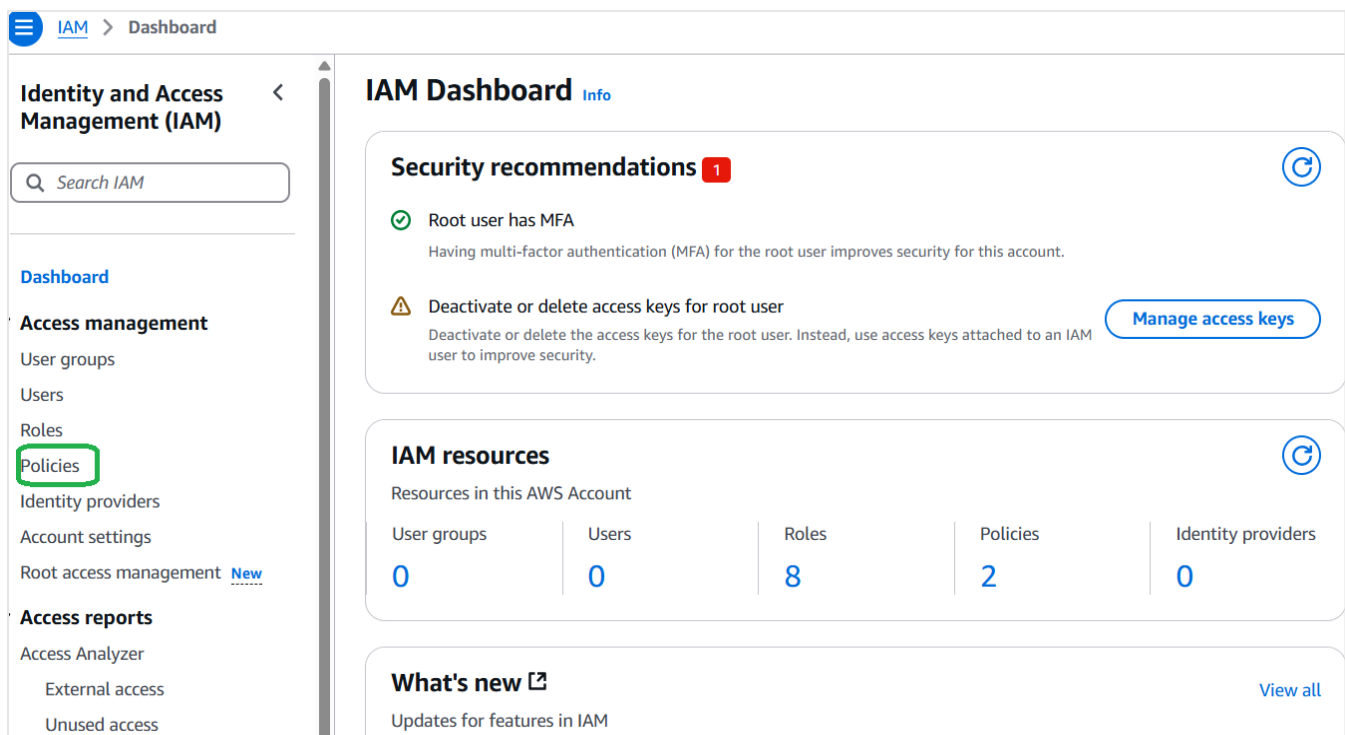


Figure 6: Search for policy

## 8]. Create IAM policy for the bucket by clicking on the “Create Policy” button

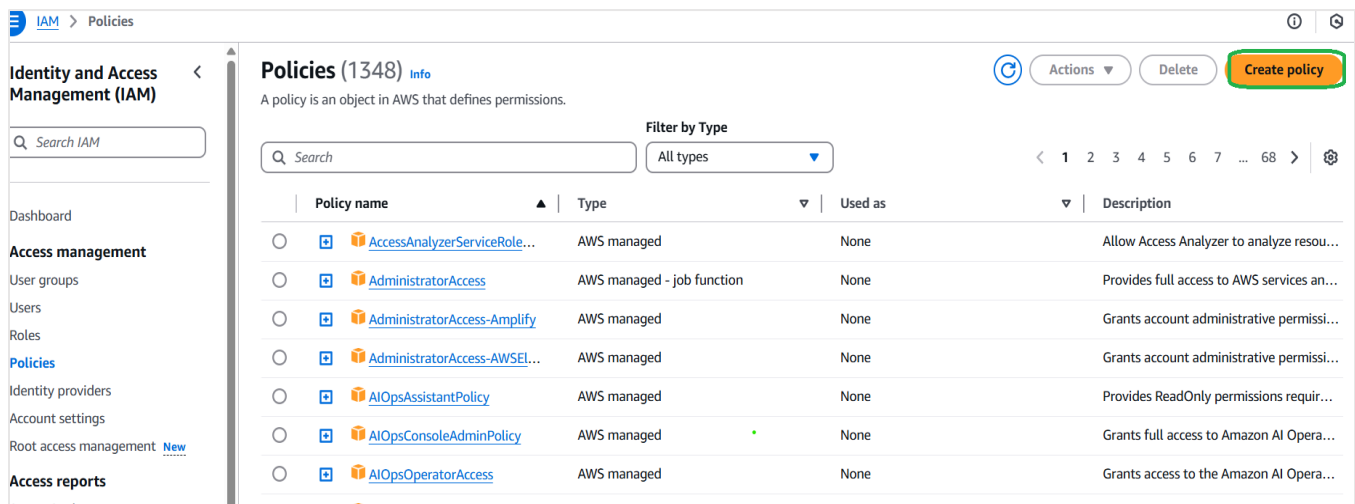


Figure 7: Create policy

## 9]. Click on the JSON tab and replace the existing text with the text given in the reference

Document click given [link](#) for documents.

[[Option 1: Configuring a Snowflake Storage Integration to Access Amazon S3 | Snowflake Documentation](#)]

After clicking on the above link you will get following doc then just copy the code.

(It is under the step no. 8 from the document)

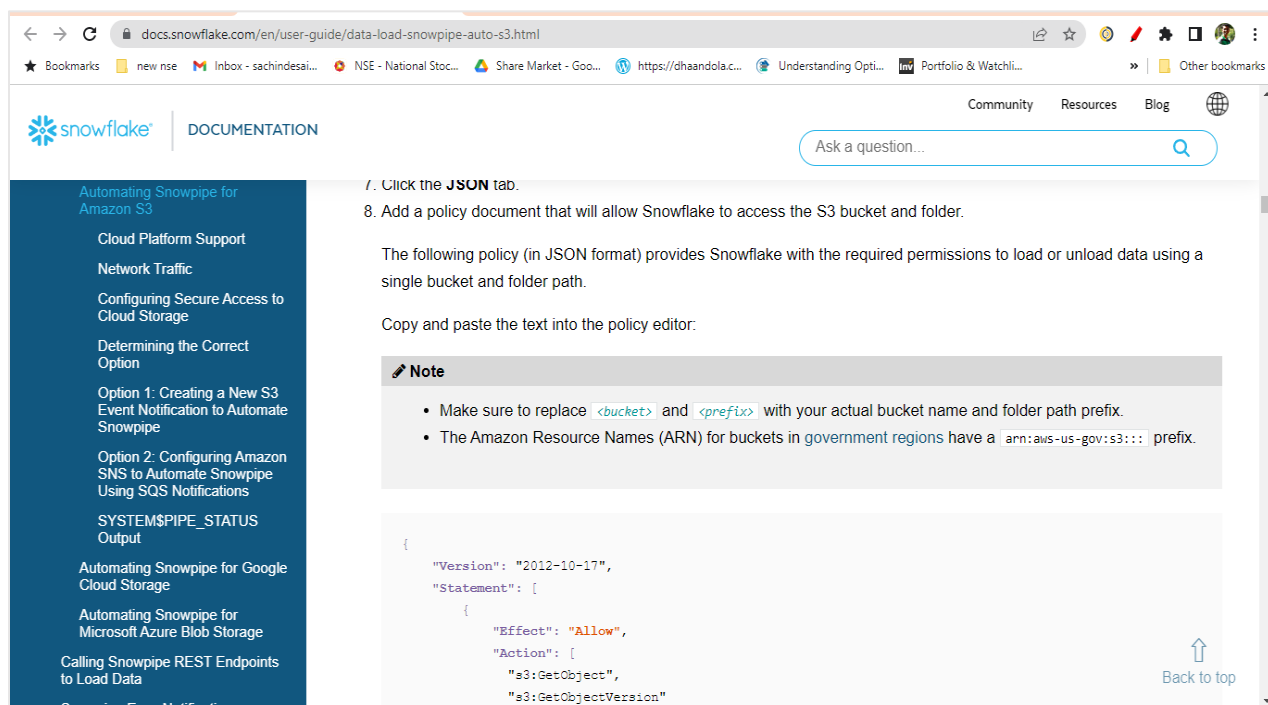


Figure 8: Policy permission script

10]. Replace the <bucket> and <prefix > with your actual bucket name and folder path.

Also set the S3: prefix to “\*”

"s3: prefix": [

“\*”

```
1 {
2   "Version": "2012-10-17",
3   "Statement": [
4     {
5       "Effect": "Allow",
6       "Action": [
7         "s3:PutObject",
8         "s3:GetObject",
9         "s3:GetObjectVersion",
10        "s3:DeleteObject",
11        "s3:DeleteObjectVersion"
12      ],
13      "Resource": "arn:aws:s3::sk-retail-bucket/*"
14    },
15    {
16      "Effect": "Allow",
17      "Action": "s3:ListBucket",
18      "Resource": "arn:aws:s3::sk-retail-bucket"
19    }
20  ]
21 }
```

Figure 9: Policy JSON script

11]. Click Next then skip the Add Tags. Enter the policy name → Click Create Policy.

Your policy will get created. (eg: RETAIL\_POLICY)

Identity and Access Management (IAM)

Search IAM

Dashboard

Access management

- User groups
- Users
- Roles
- Policies
- Identity providers
- Account settings
- Root access management

Access reports

- Access Analyzer
- External access
- Unused access
- Analyzer settings

## RETAIL\_POLICY

FOR SPECIFIC PERMISSION

**Policy details**

Type	Creation time	Edited time	ARN
Customer managed	March 29, 2025, 20:13 (UTC+05:30)	March 29, 2025, 20:13 (UTC+05:30)	arn:aws:iam::123456789012:policy/RETAIL_POLICY

**Permissions** | Entities attached | Tags | Policy versions (1) | Last Accessed

**Permissions defined in this policy**

Permissions defined in this policy document specify which actions are allowed or denied. To define permissions for an IAM identity (user, user group, or role), you must specify the service, action, and resource.

Search

**Allow (1 of 439 services)**

Service	Access level	Resource	Request condition
S3	Limited: List, Read, Write	Multiple	None

Figure 10: Created Policy

## 12]. Create IAM Role. Click on Create Role

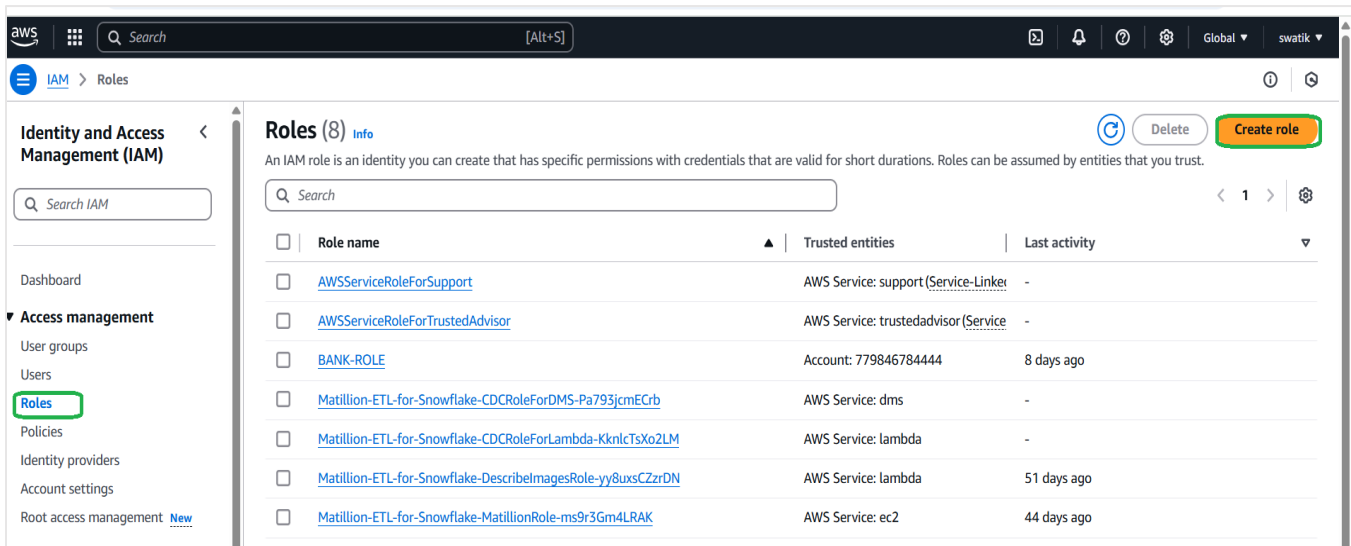


Figure 11: Creating role

## 13]. Select AWS Account from Trusted Entity Type.

You will get your account number selected by default when you select AWS account.

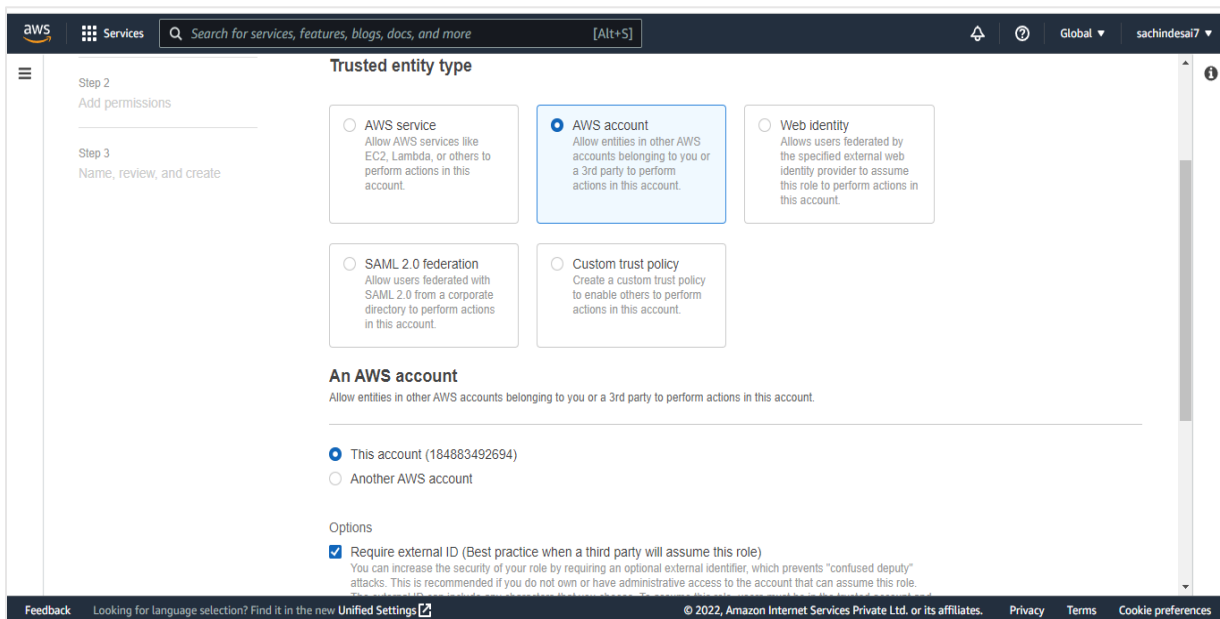


Figure 12: AWS account selection

## 14] Check Require external ID and enter 000 (as currently we are not having it) and click next (Optional)

Allow entities in other AWS accounts belonging to you or a 3rd party to perform actions in this account.

☒ This account (184883492694)  
☐ Another AWS account

Options

☒ **Require external ID** (Best practice when a third party will assume this role)  
 You can increase the security of your role by requiring an optional external identifier, which prevents "confused deputy" attacks. This is recommended if you do not own or have administrative access to the account that can assume this role. The external ID can include any characters that you choose. To assume this role, users must be in the trusted account and provide this exact external ID. [Learn more](#)

External ID

0000

**Important:** The console does not support using an external ID with the Switch Role feature. If you select this option, entities in the trusted account must use the API, CLI, or a custom federation proxy to make cross-account iam:AssumeRole calls. [Learn more](#)

☐ **Require MFA**  
 Requires that the assuming entity use multi-factor authentication.

Cancel Next

Figure 13: Provide external id

**15]. On the next page, Select the IAM policy that you have created. Linked policy to the IAM role.**

**Add permissions** [Info](#)

**Permissions policies (1/1047)** [Info](#)

Choose one or more policies to attach to your new role.

Search  Filter by Type Customer managed 2 matches

<input type="checkbox"/>	Policy name	Type	Description
<input type="checkbox"/>	<a href="#">CZ-BANK-POLICY</a>	Customer managed	-
<input checked="" type="checkbox"/>	<a href="#">RETAIL_POLICY</a>	Customer managed	FOR SPECIFIC PERMISSION

► **Set permissions boundary - optional**

Cancel Previous Next

Figure 14: Linked role to policy

16]. On the next page Enter any unique name to the role you are creating. The description is optional.

Click on the Create Role (Skip the Add Tags).

Click on the role that you have created. It will show you the summary page.

You will get the following window

Note down the Role ARN, which we will need when we create the 'Storage Integration'.

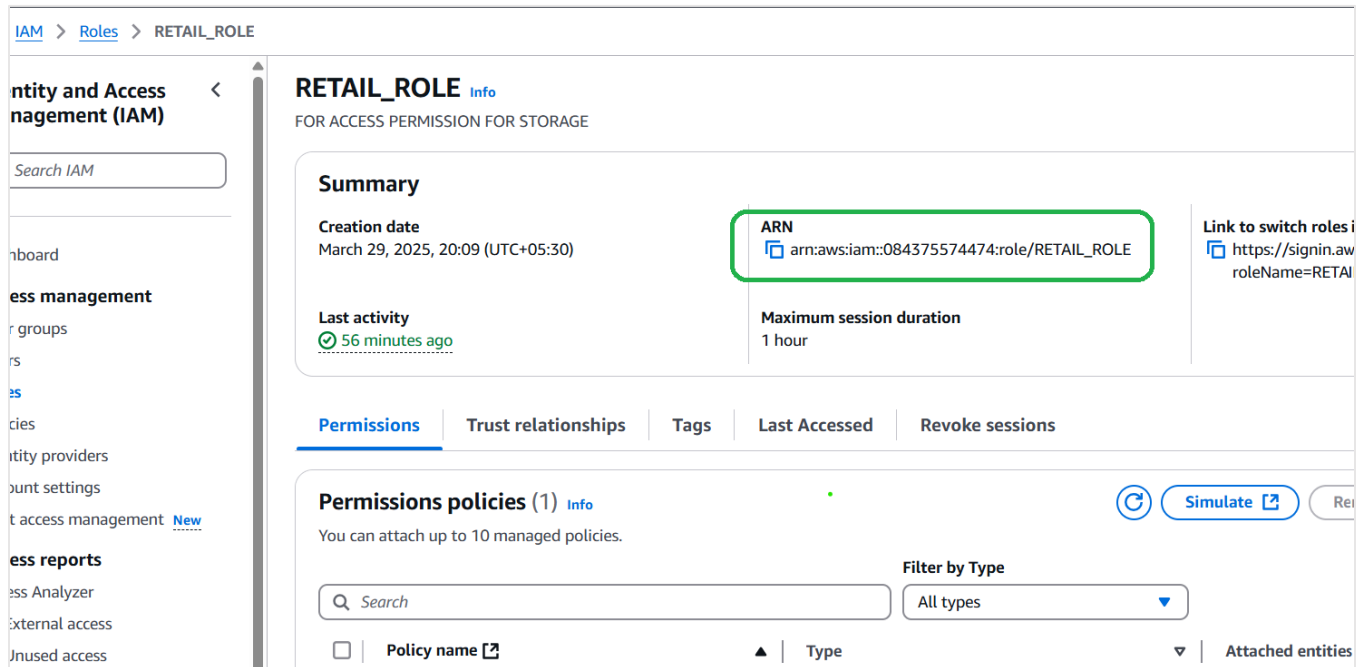


Figure 15: Retail role created

## 17]. Login to the Snowflake Account.

Create Cloud Storage Integration in Snowflake and map S3 user/role with it (STORAGE\_AWS\_ROLE\_ARN).

-- CREATE STORAGE INTEGRATION

CREATE OR REPLACE STORAGE INTEGRATION s3\_int\_retail

TYPE = EXTERNAL\_STAGE

STORAGE\_PROVIDER = S3

ENABLED = TRUE

STORAGE\_AWS\_ROLE\_ARN = 'arn:aws:iam::084375574474:role/RETAIL\_ROLE'

STORAGE\_ALLOWED\_LOCATIONS = ('s3://sk-retail-bucket/');



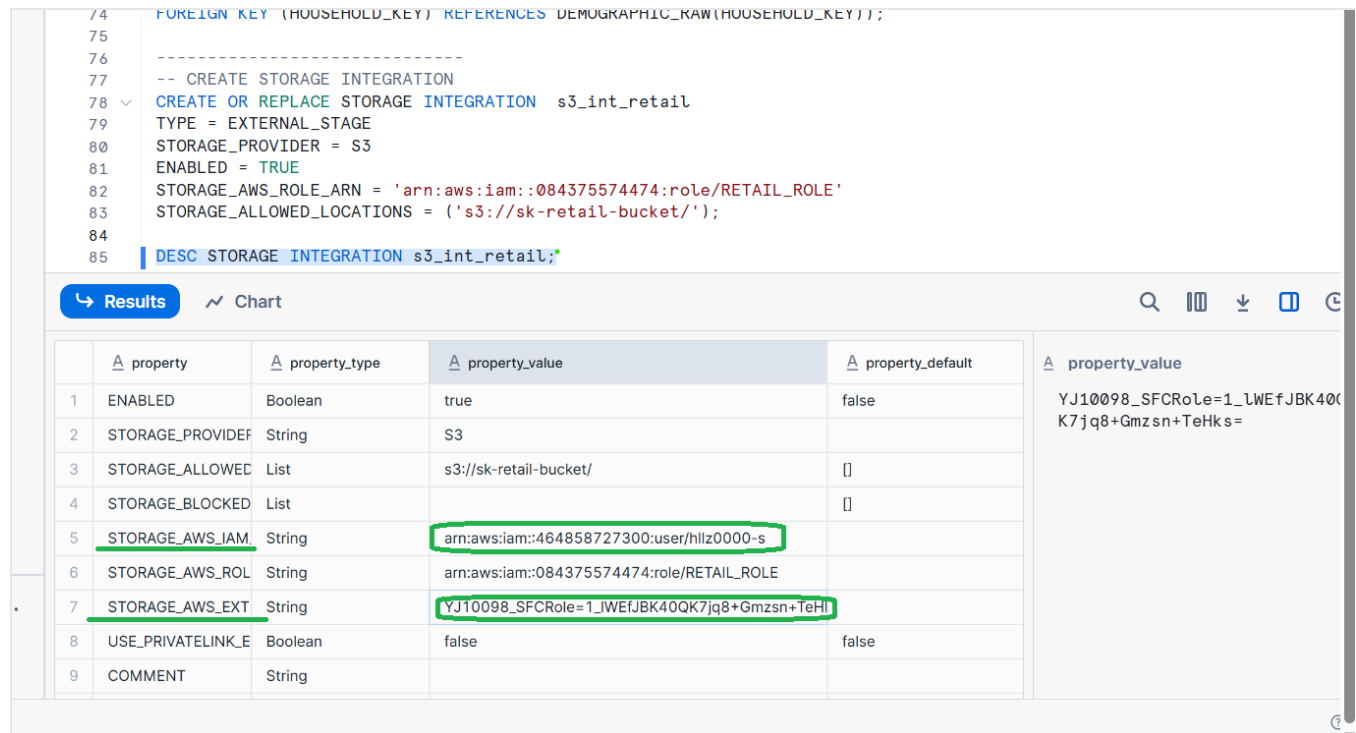
- s3\_int\_retail → cloud storage integration name
- STORAGE\_AWS\_ROLE\_ARN → is the Amazon Resource Name (ARN) of the role you created
- STORAGE\_ALLOWED\_LOCATIONS → is the name of a S3 bucket that stores your data files (eg: sk-retail-bucket)

## 18]. In Snowflake worksheet run command

-- Desc integration integration\_name;

**DESC STORAGE INTEGRATION s3\_int\_retail;**

And note down the STORAGE\_AWS\_IAM\_USER\_ARN and STORAGE\_AWS\_EXTERNAL\_ID from the result set



```

74 FOREIGN KEY (HOUSEHOLD_KEY) REFERENCES DEMOGRAPHIC_RAW (HOUSEHOLD_KEY));
75
76 -----
77 -- CREATE STORAGE INTEGRATION
78 CREATE OR REPLACE STORAGE INTEGRATION s3_int_retail
79 TYPE = EXTERNAL_STAGE
80 STORAGE_PROVIDER = S3
81 ENABLED = TRUE
82 STORAGE_AWS_ROLE_ARN = 'arn:aws:iam::084375574474:role/RETAIL_ROLE'
83 STORAGE_ALLOWED_LOCATIONS = ('s3://sk-retail-bucket/');
84
85 DESC STORAGE INTEGRATION s3_int_retail;

```

	property	property_type	property_value	property_default	property_value
1	ENABLED	Boolean	true	false	YJ10098_SFCRole=1_LWEfJBK40K7jq8+Gmzsn+TeHks=
2	STORAGE_PROVIDER	String	S3		
3	STORAGE_ALLOWED	List	s3://sk-retail-bucket/	[]	
4	STORAGE_BLOCKED	List		[]	
5	STORAGE_AWS_IAM	String	arn:aws:iam::464858727300:user/hlz0000-s		
6	STORAGE_AWS_ROLE	String	arn:aws:iam::084375574474:role/RETAIL_ROLE		
7	STORAGE_AWS_EXTERNAL_ID	String	YJ10098_SFCRole=1_LWEfJBK40K7jq8+Gmzsn+TeHks=		
8	USE_PRIVATELINK_ENDPOINT	Boolean	false	false	
9	COMMENT	String			

Figure 16: AWS external id

## 19]. Grant the IAM User Permissions to Access Bucket Objects

1. Log into the AWS Management Console.
2. From the home dashboard, choose Identity & Access Management (IAM)
3. Select the role you created
4. Click Trust Relationships -> Edit trust relationship
5. Replace the value of "AWS": with the AWS\_IAM\_USER\_ARN String you got using DESC INTEGRATION command and, value of "sts:ExternalId": with AWS\_EXTERNAL\_ID String
6. Click Update Policy

## Edit trust policy

```
1 {
2   "Version": "2012-10-17",
3   "Statement": [
4     {
5       "Effect": "Allow",
6       "Principal": {
7         "AWS": "arn:aws:iam::464858727300:user/h11z0000-s"
8       },
9       "Action": "sts:AssumeRole",
10      "Condition": {
11        "StringEquals": {
12          "sts:ExternalId": "YJ10098_SFCRole=1_lWEfJBK40QK7jq8+Gmzsn+TeHks="
13        }
14      }
15    }
16  ]
17 }
```

Figure 17: edit trust policy

**20]. Create Snowflake file format. This file format will be used at the time of Stage creation.**

**Create File Format**

Name \*

Schema Name

Format Type

Compression Method

Column separator

Row separator

Header lines to skip

Field optionally enclosed by

Null String

☐ Trim space before and after

[Show SQL](#)

Figure 18: File format creation

Or create csv format via SQL code in snowflake snowsight

```
CREATE OR REPLACE FILE FORMAT RETAILCSV
```

```
TYPE = CSV
```

```
FIELD_DELIMITER = ','
```

```
FIELD_OPTIONALLY_ENCLOSED_BY = ''
```

```
SKIP_HEADER = 1;
```

## 21]. Create a stage in snowflake pointing to your S3 bucket:

```
CREATE OR REPLACE STAGE RETAIL_STAGE
```

```
URL = 's3://sk-retail-bucket/' -- (Name of your bucket)
```

```
FILE_FORMAT = RETAILCSV
```

```
STORAGE_INTEGRATION = s3_int_retail;
```

--- we can see previously created stage

```
SHOW STAGES;
```

## 22]. Create a SNOWPIPE with Auto-Ingest Enabled

```
CREATE OR REPLACE PIPE SNOWPIPE_DEMOGRAPHIC_RAW
```

```
AUTO_INGEST = TRUE
```

```
AS COPY INTO RETAIL.RETAIL_SCHEMA.DEMOGRAPHIC_RAW --table NAME  
DEMOGRAPHIC_RAW that you created in snowflake)
```

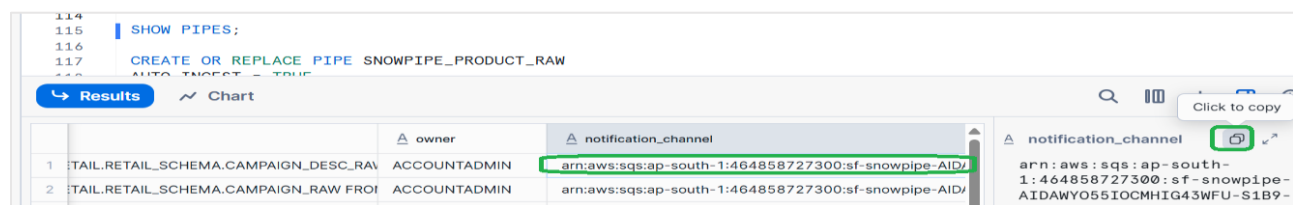
```
FROM @RETAIL_STAGE/DEMOGRAPHIC -----s3 bucket subfolder name
```

```
FILE_FORMAT = RETAILCSV;
```

## 23]. After creating snowpipe, get 'Notification Channel' value

Run command

```
Show pipes;
```



	owner	notification_channel	notification_channel
1	TAIL.RETAIL_SCHEMA.CAMPAIGN_DESC_RAW	ACCOUNTADMIN	arn:aws:sqs:ap-south-1:464858727300:sf-snowpipe-AIDv
2	TAIL.RETAIL_SCHEMA.CAMPAIGN_RAW FRO	ACCOUNTADMIN	arn:aws:sqs:ap-south-1:464858727300:sf-snowpipe-AIDv

Figure 19: Copied event notification ARN

Or Go to Database ☐ Pipes

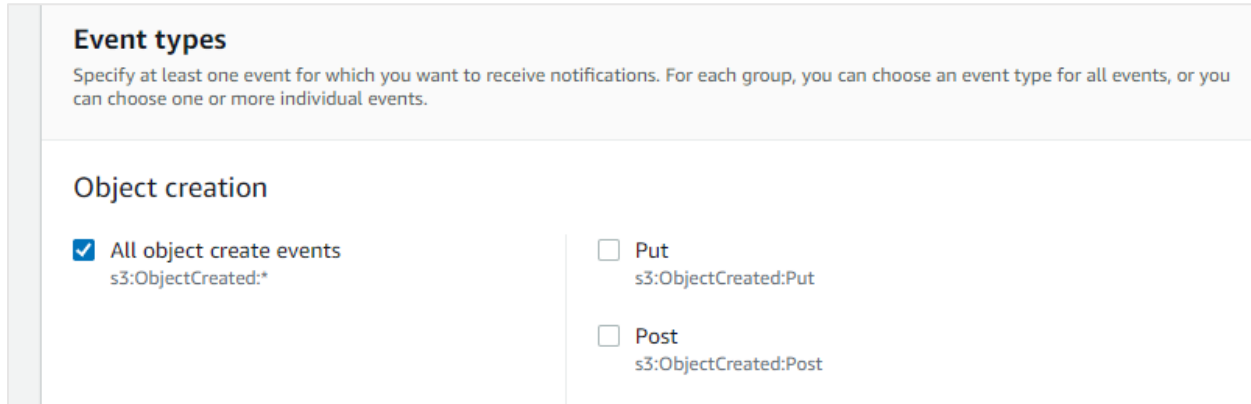
## 24]. Create an event on S3 bucket.

Go to your S3 bucket that you have created. Click on Properties tab and scroll down to

Event Notification -> Click Create Event Notification

Enter any name for the Notification.

Check All Object create Events



**Event types**

Specify at least one event for which you want to receive notifications. For each group, you can choose an event type for all events, or you can choose one or more individual events.

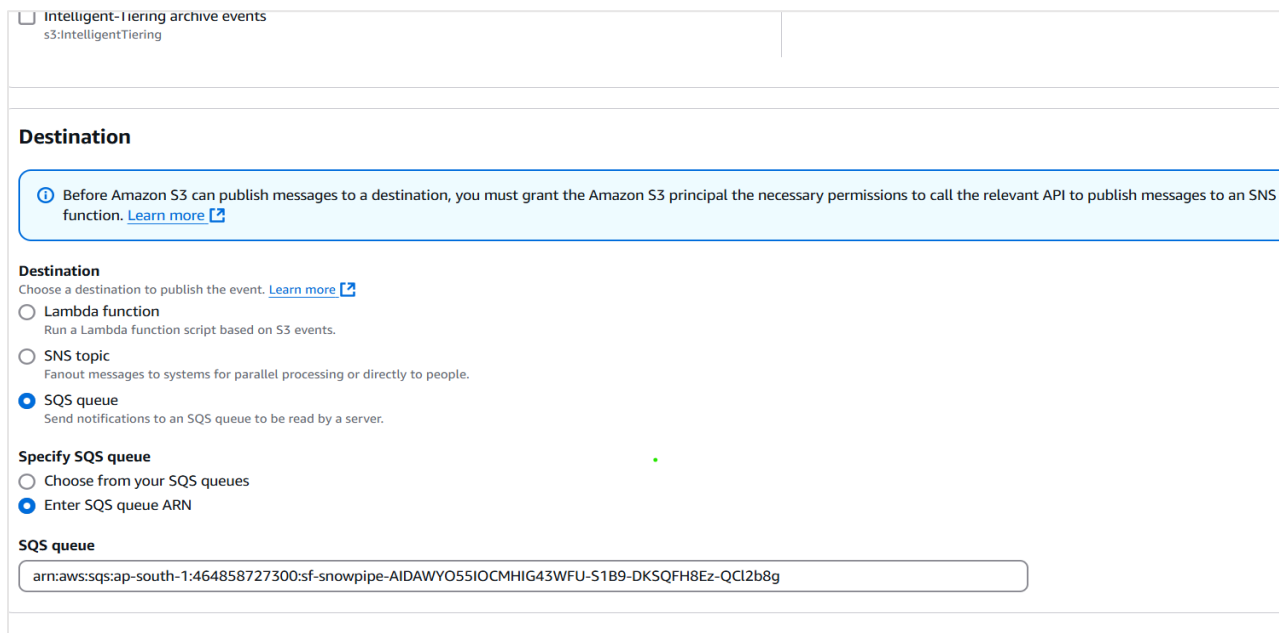
**Object creation**

- ☒ All object create events  
s3:ObjectCreated:\*
- ☐ Put  
s3:ObjectCreated:Put
- ☐ Post  
s3:ObjectCreated:Post

Figure 20: Create Event notification

Scroll down to Destination

Select SQS Queue ➔ Select Enter SQS Queue ARN ➔ And paste that 'Notification Channel' under SQS Queue



☐ Intelligent-Tiering archive events  
s3:IntelligentTiering

**Destination**

Before Amazon S3 can publish messages to a destination, you must grant the Amazon S3 principal the necessary permissions to call the relevant API to publish messages to an SNS function. [Learn more](#)

**Destination**

Choose a destination to publish the event. [Learn more](#)

- ☐ Lambda function  
Run a Lambda function script based on S3 events.
- ☐ SNS topic  
Fanout messages to systems for parallel processing or directly to people.
- ☒ SQS queue  
Send notifications to an SQS queue to be read by a server.

**Specify SQS queue**

- ☐ Choose from your SQS queues
- ☒ Enter SQS queue ARN

**SQS queue**

arn:aws:sqs:ap-south-1:464858727300:sf-snowpipe-AIDAWYO55IOCMHIG43WFU-S1B9-DK5QFH8Ez-QCl2b8g

Figure 21: Paste notification ARN in SQS

Now you are ready to load the file to s3 bucket.

**25]. Following are some snowpipe command which will help you to check snowpipe status**

```
SELECT SYSTEM$PIPE_STATUS(' SNOWPIPE_DEMOGRAPHIC_RAW ');
```

-- REFRESH DATA IN IN BUCKET--

```
ALTER PIPE SNOWPIPE_DEMOGRAPHIC_RAW REFRESH;
```

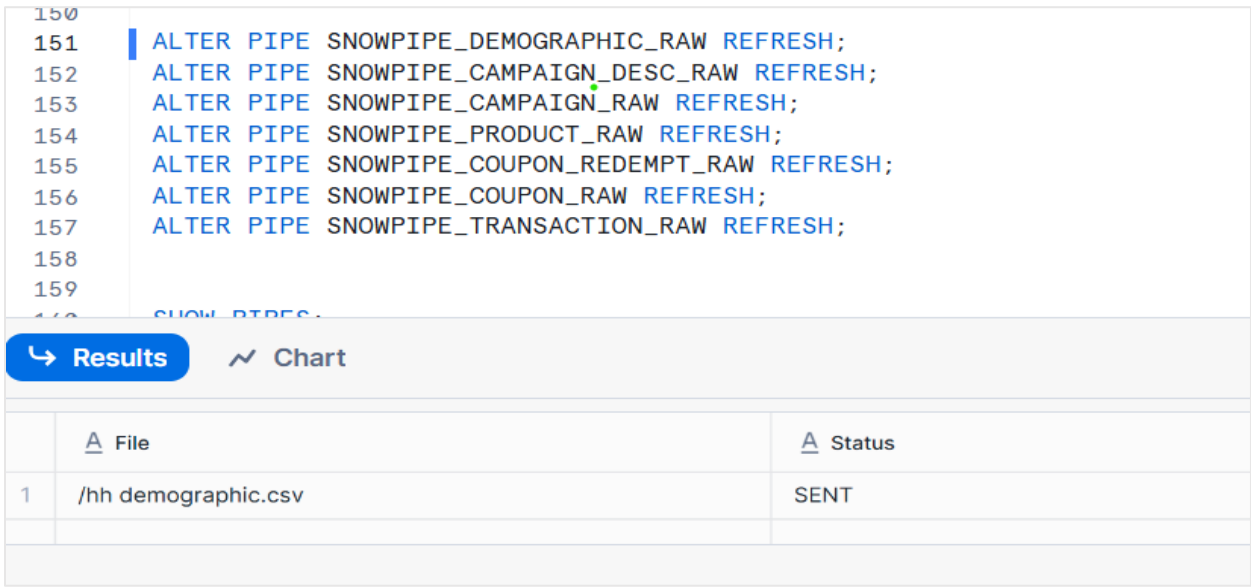


Figure 22: refresh PIPE

Show Total records.

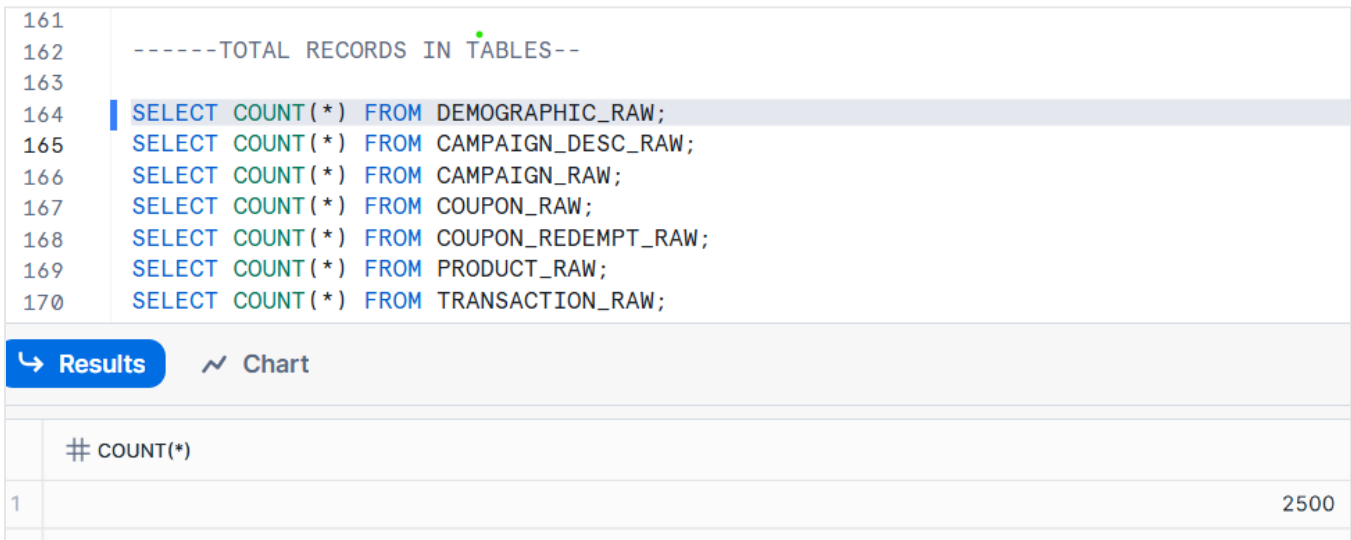


Figure 23: Total records

In this we successfully load all data in the Snowflake.

