**SNOWFLAKE CONTINUOUS DATA LOADING**

**How we can automatically ingest data from the s3 to Snowflake?**

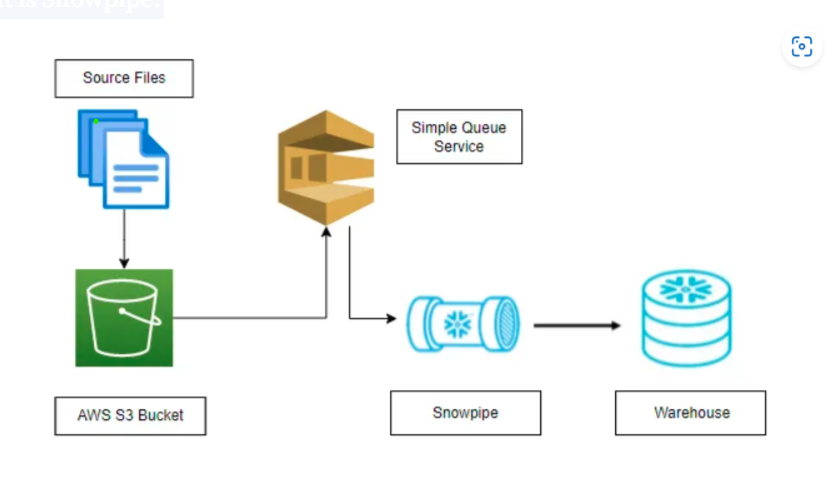
****

Figure : Snowpipe data auto ingest architecture

**What is Snowpipe?**

Snowpipe enables loading data from files as soon as they’re available in a stage. This means you can load data from files in micro-batches, making it available to users within minutes, rather than manually executing COPY statements on a schedule to load larger batches.

In this document we can see step by step process of data ingest from AWS s3 to snowflake data warehouse by using snowpipe.

1. Create an AWS account in aws.amazon.com (by root user)
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.

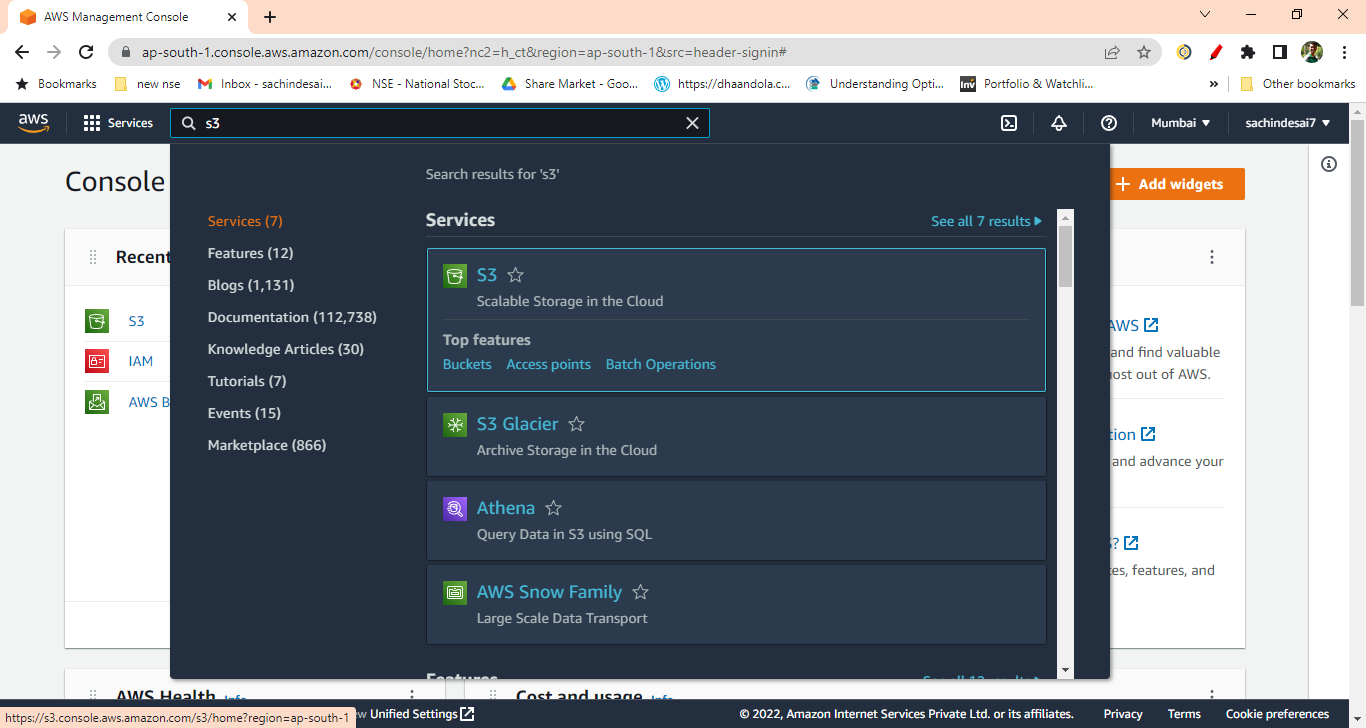


Figure : Search for s3 bucket

1. Create S3 bucket

Bucket name: **cz-bank-bucket**

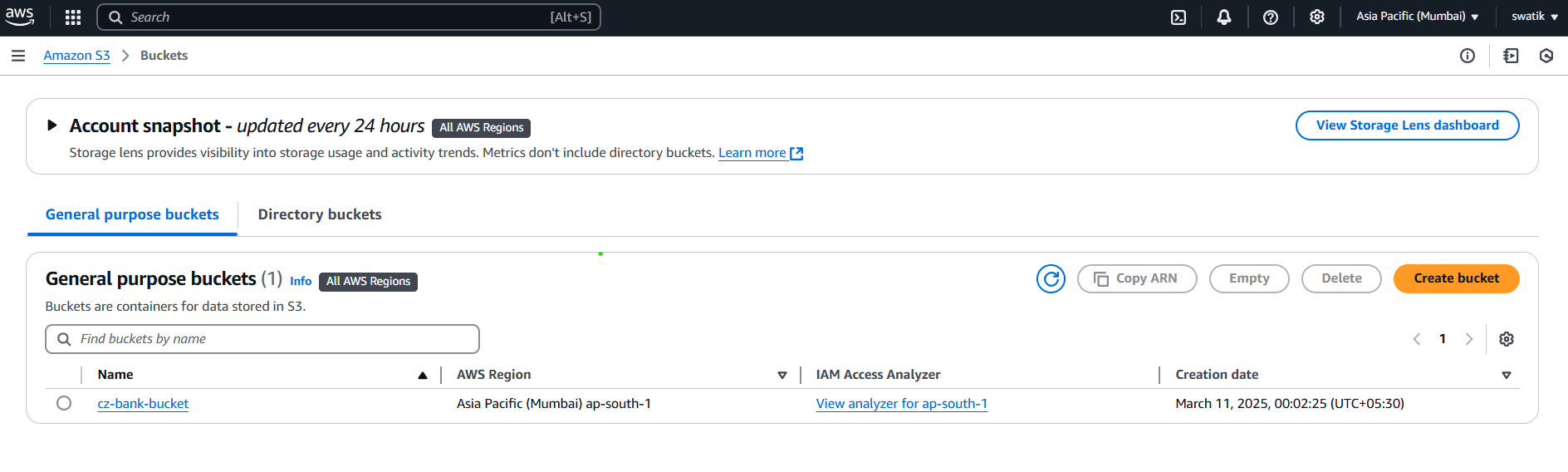
****

Figure : Create bucket

1. Create a folder inside the bucket ( e.g. ACCOUNT)

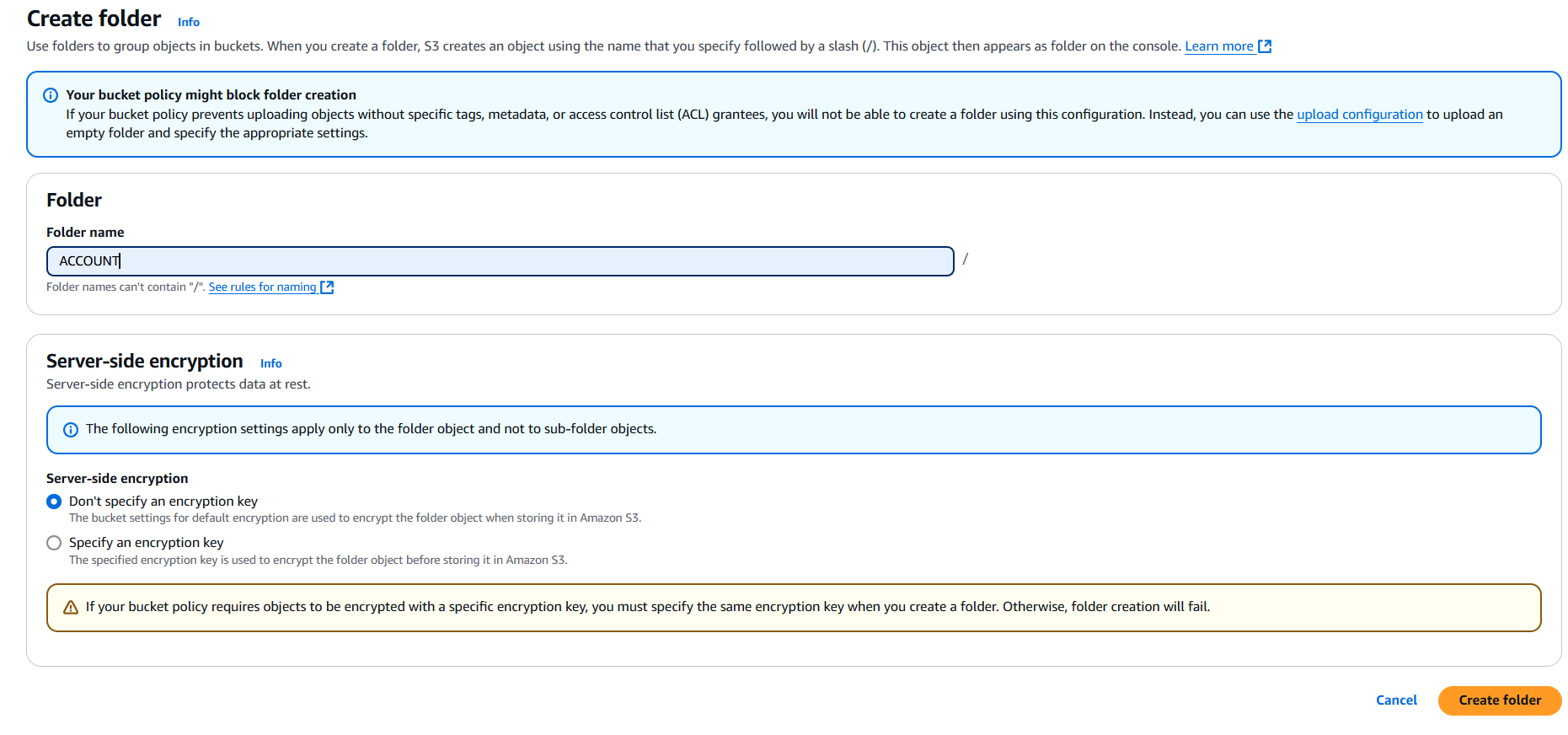


Figure : Create folder inside the bucket

For this project we created 8 folders in our bucket:

1. ACCOUNT
2. CARD
3. CLIENT
4. DISPOSITION
5. DISTRICT
6. LOAN
7. ORDER1
8. TRANSATIONS

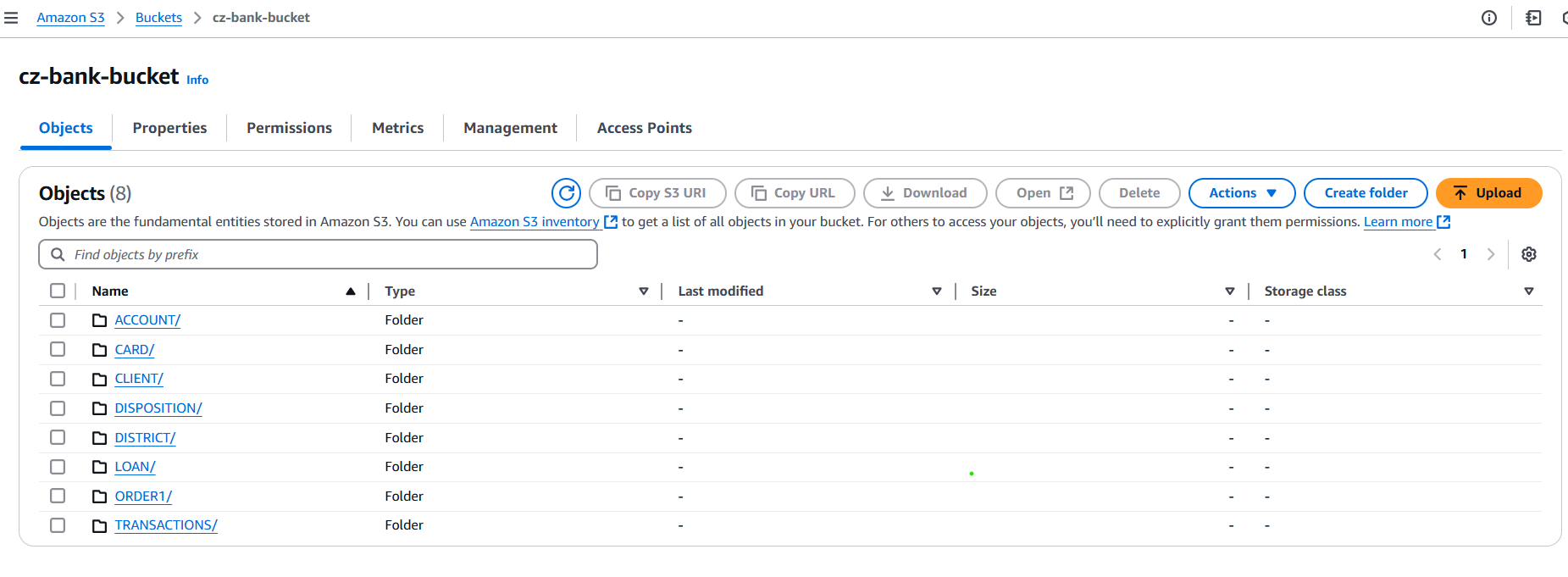


Figure : All folders in s3 buckets

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

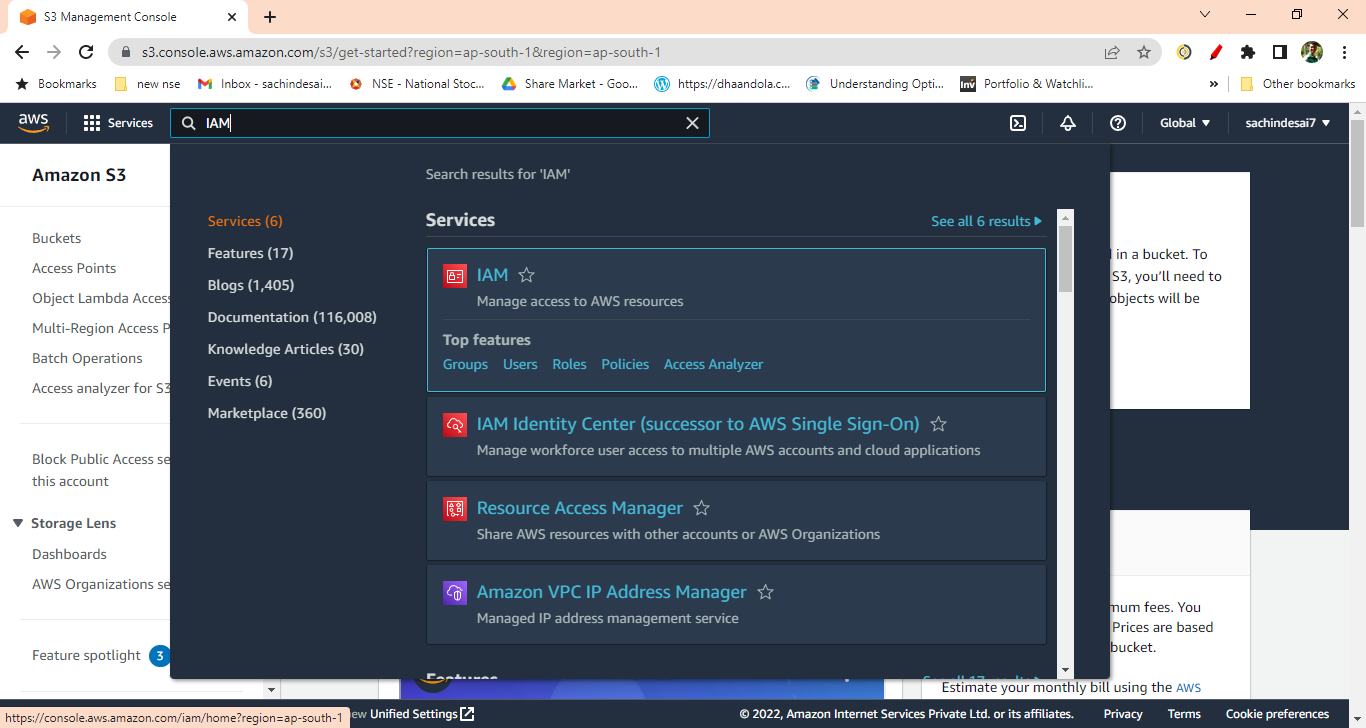


Figure : search for IAM service

1. Click on the Policies from IAM Dashboard.

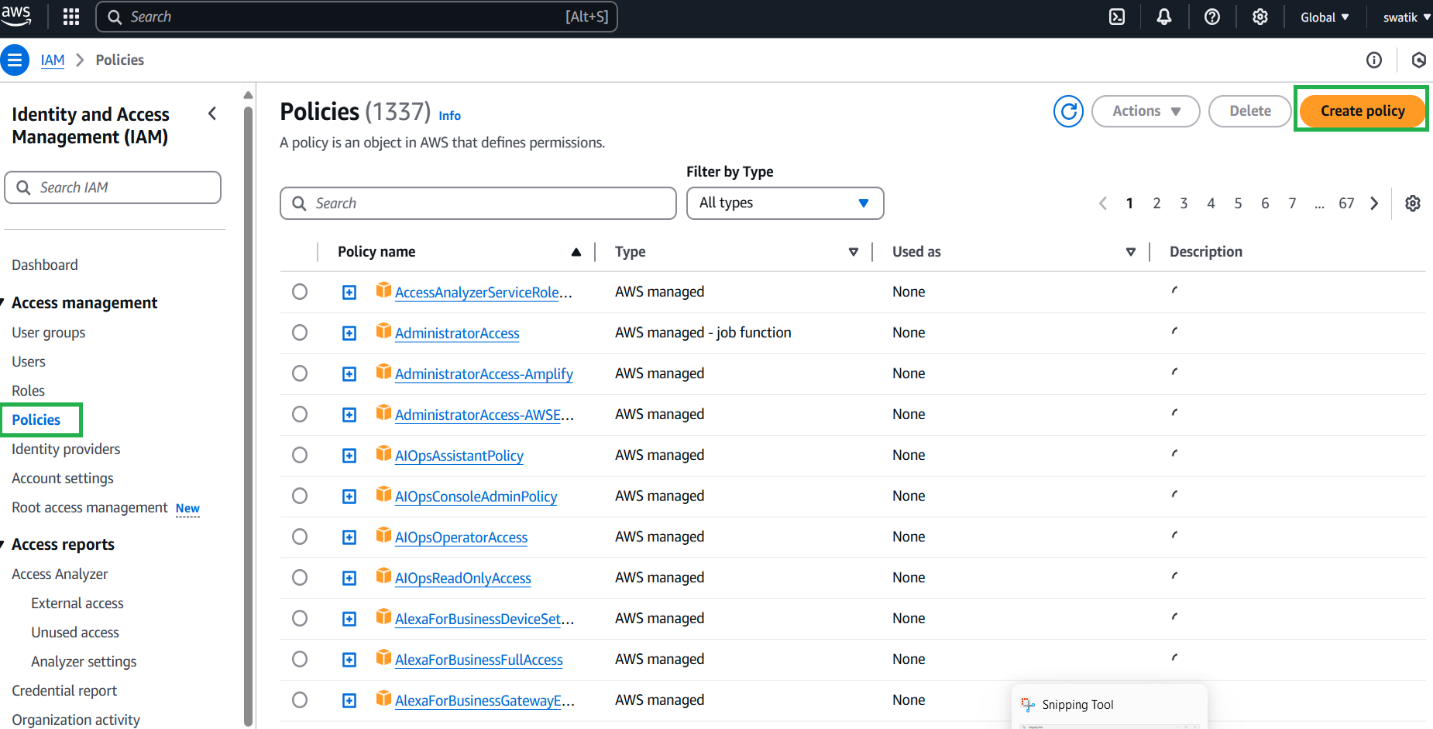


Figure : Create policy

1. Create IAM policy for the bucket by clicking on the “Create Policy” button. Then click on Json tab and then next.

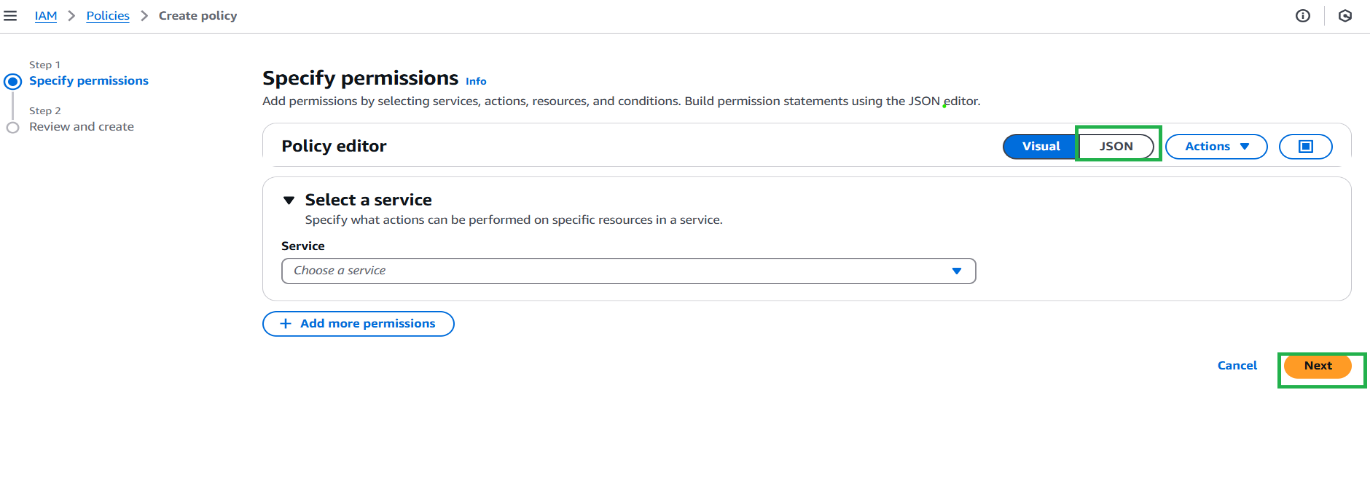


Figure : Click on JSON tab and next

1. After Clicking the JSON tab and replace the existing text with the text given in the reference

Document click given [link](https://docs.snowflake.com/en/user-guide/data-load-snowpipe-auto-s3.html) for documents.

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)

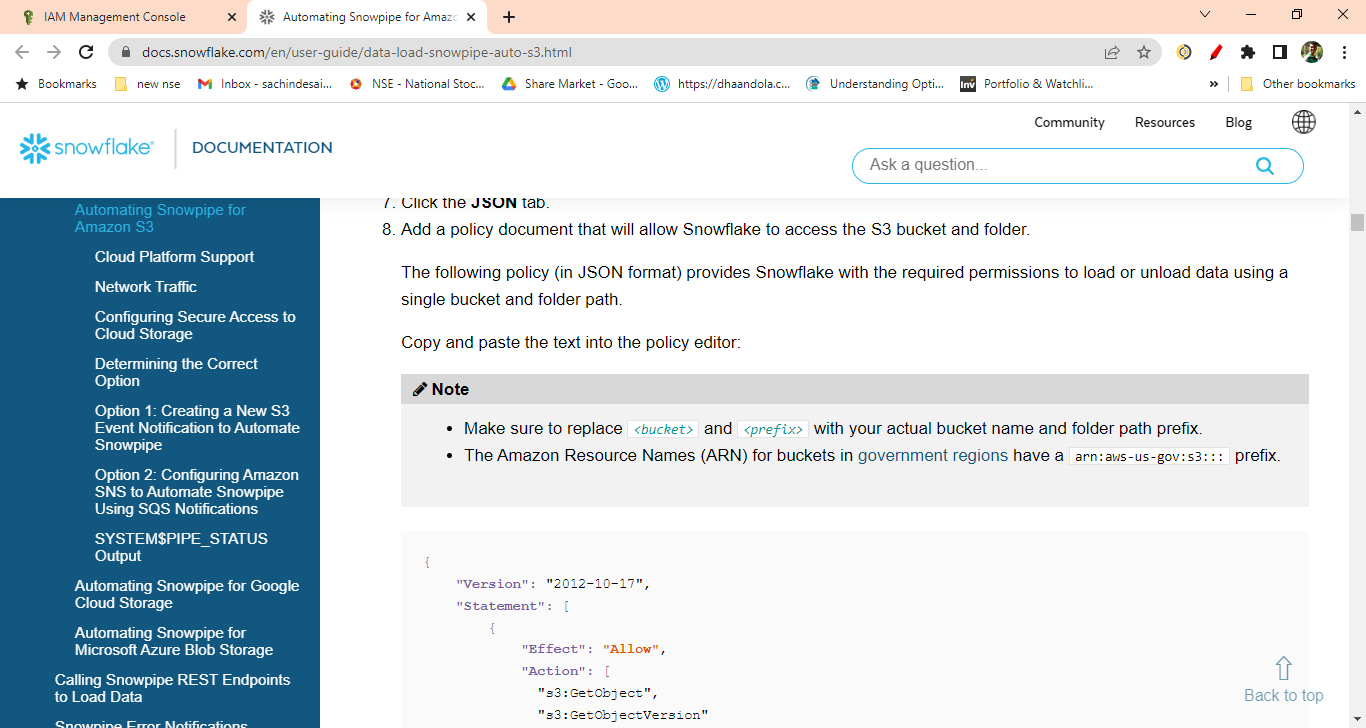


Figure : Reference document for policy

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

Also set the S3: prefix to “ \*”

"s3: prefix”: [

“\*"



Figure : Trust policy permission

1. Click Next then skip the Add Tags. Enter the policy name. Click Create Policy. Your policy will get created.

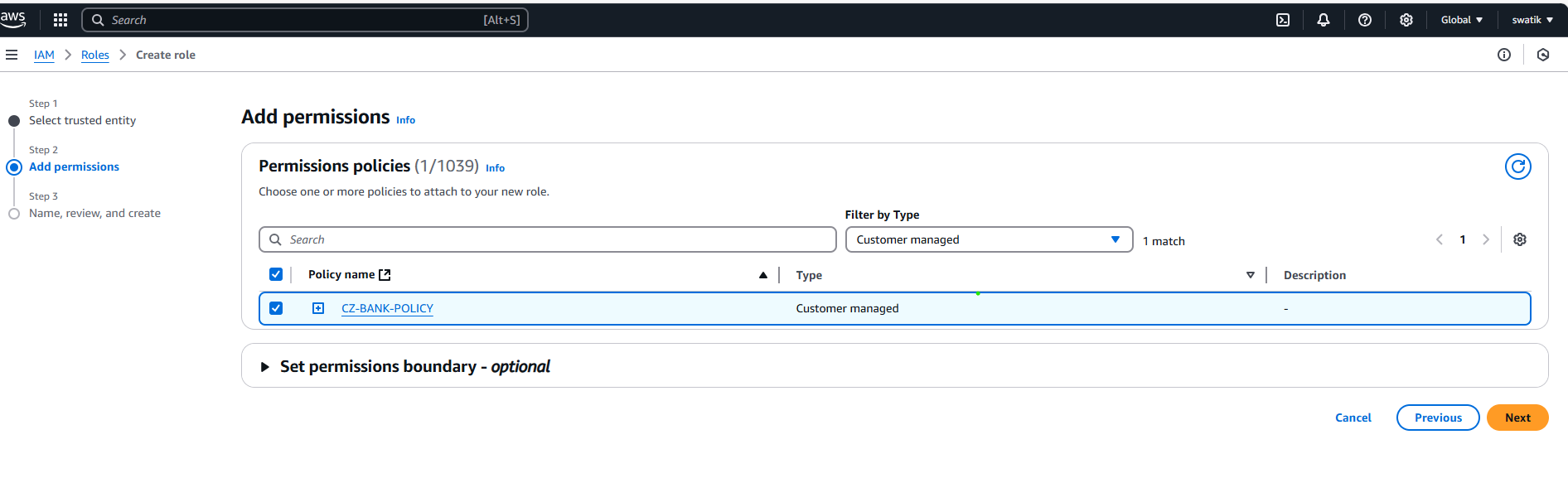


Figure : Policy created

1. After policy, Create IAM Role. Click on Create Role

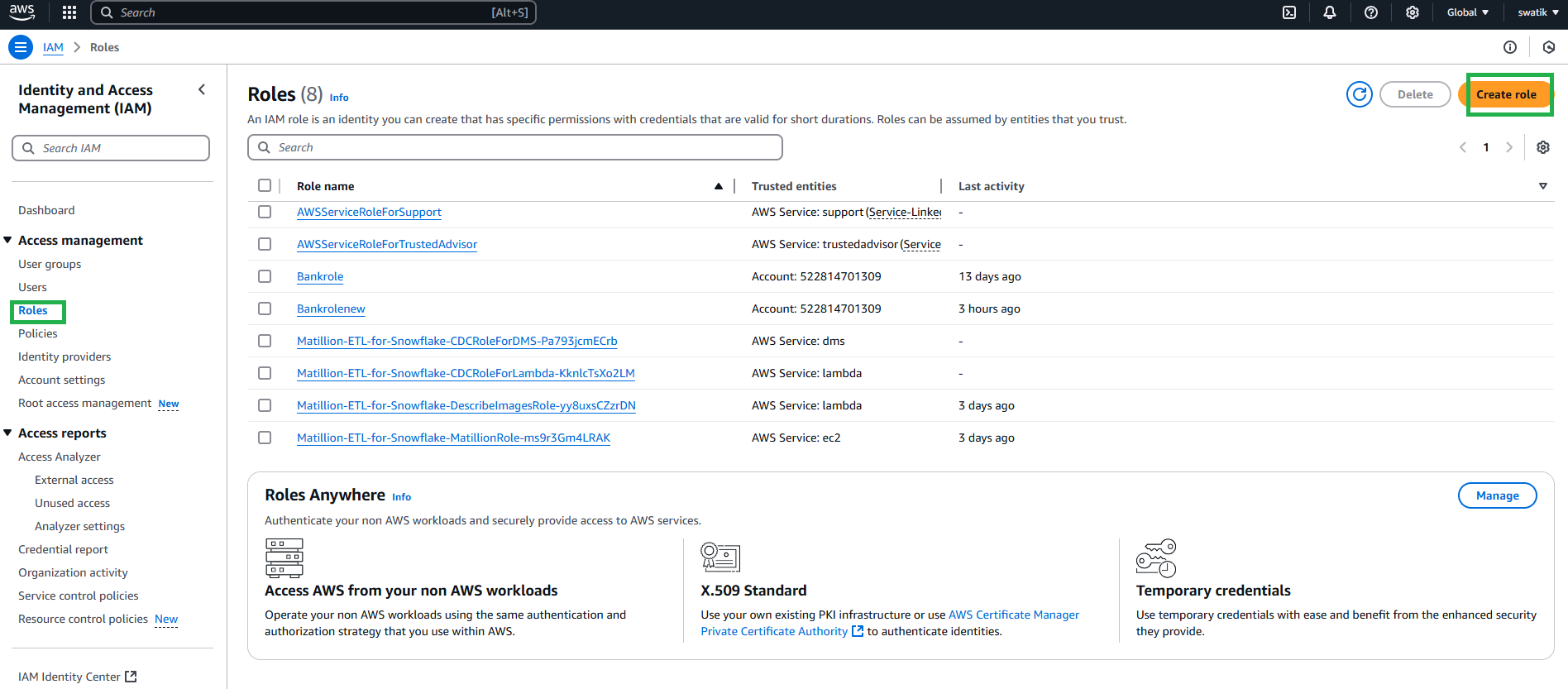


Figure : Create role

1. Select AWS Account from Trusted Entity Type. You will get your account number selected by default when you select AWS account. Check Require external ID and enter 0000 (as currently we are not having it) and click next

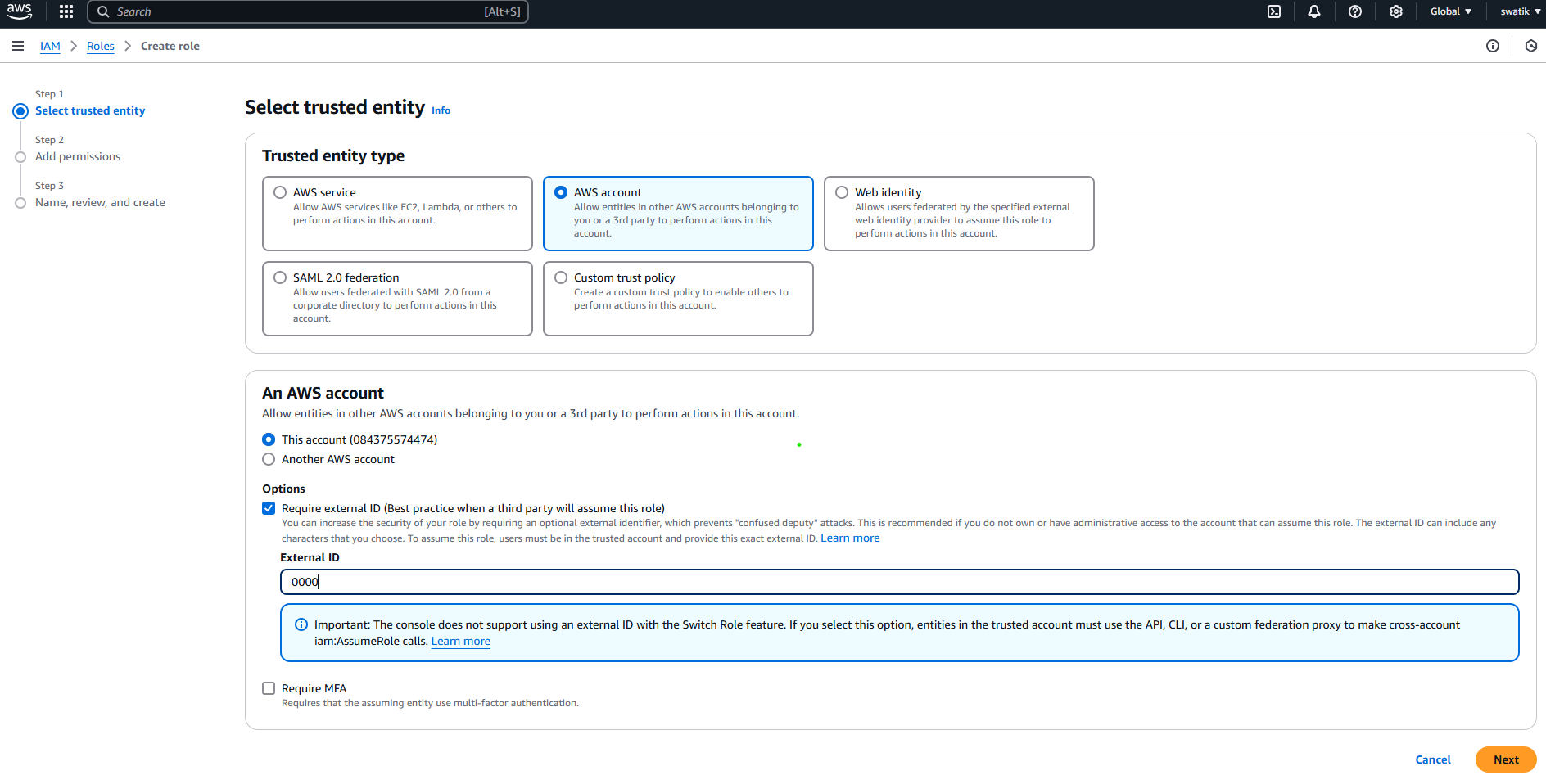


Figure : Select AWS account

1. On the next page, attached the IAM policy that you have created

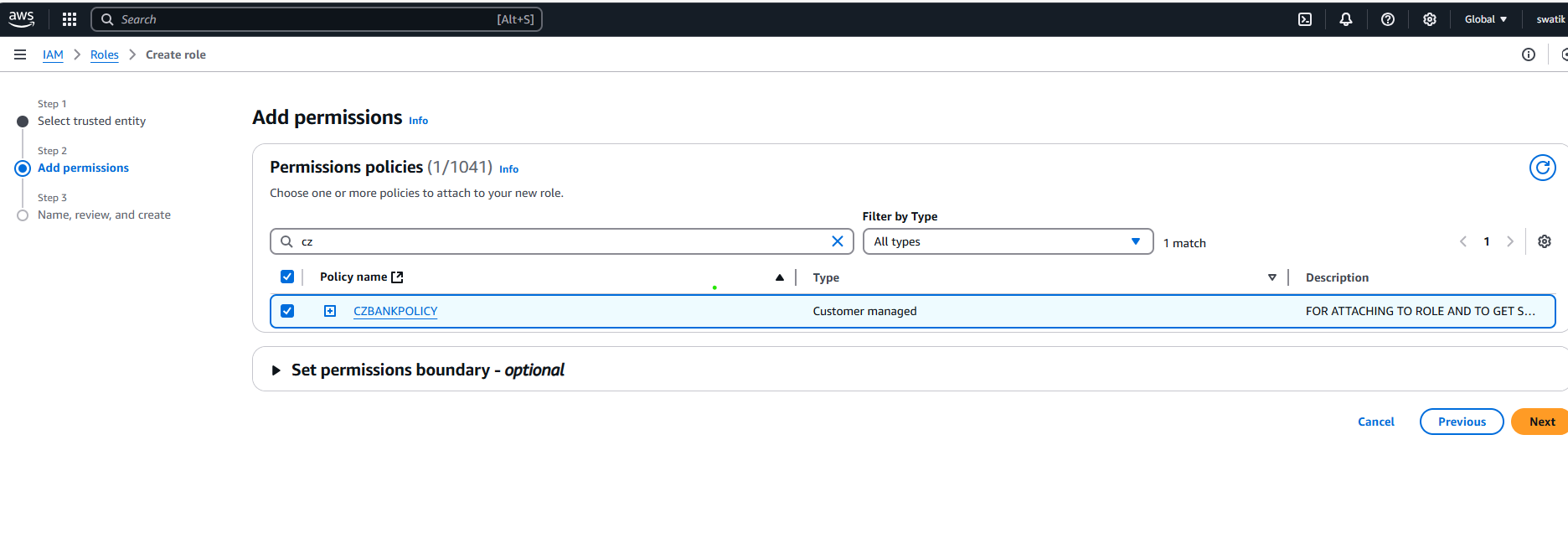


Figure : Attached policy to role

1. 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).

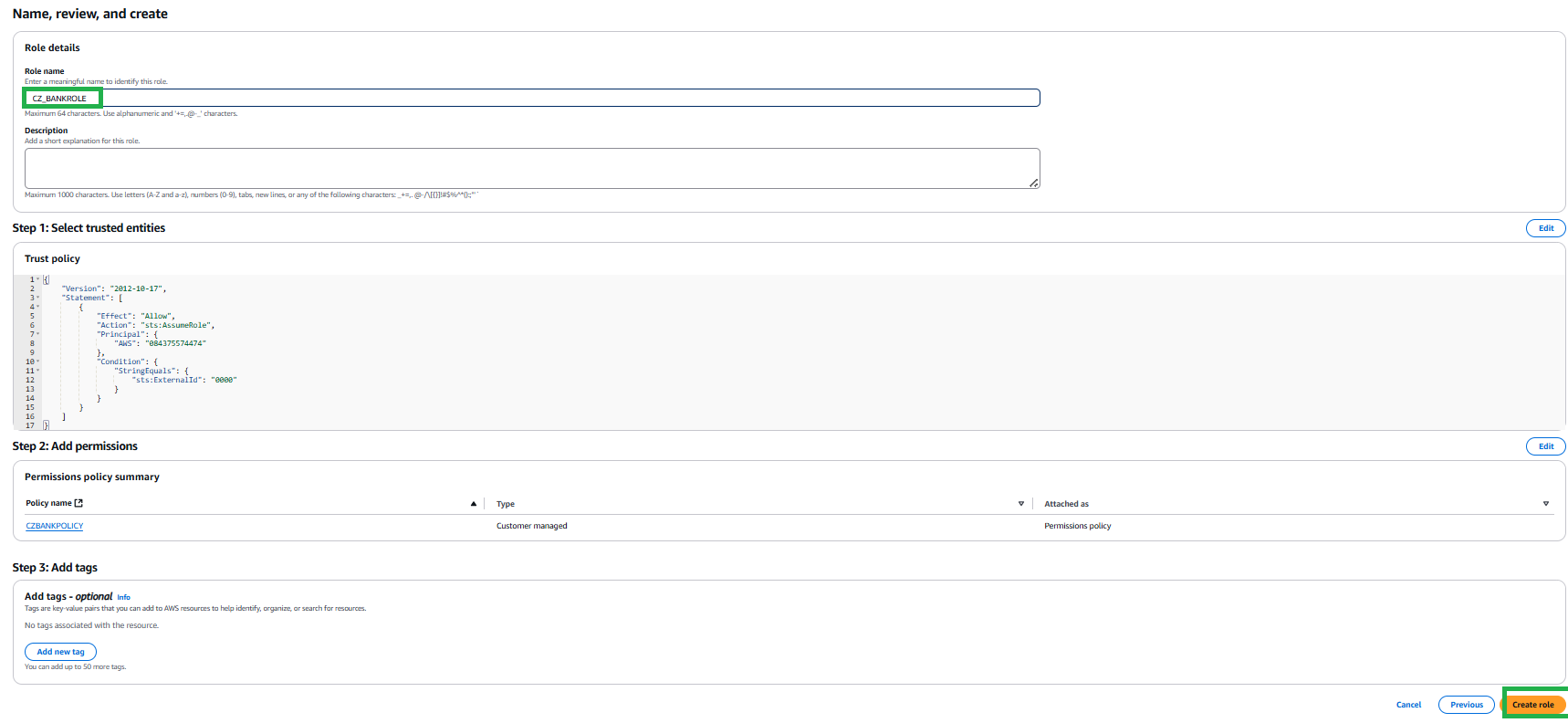


Figure : Give name and create role

1. Click on the role that you have created. It will show you the summary page. You will get the following window.

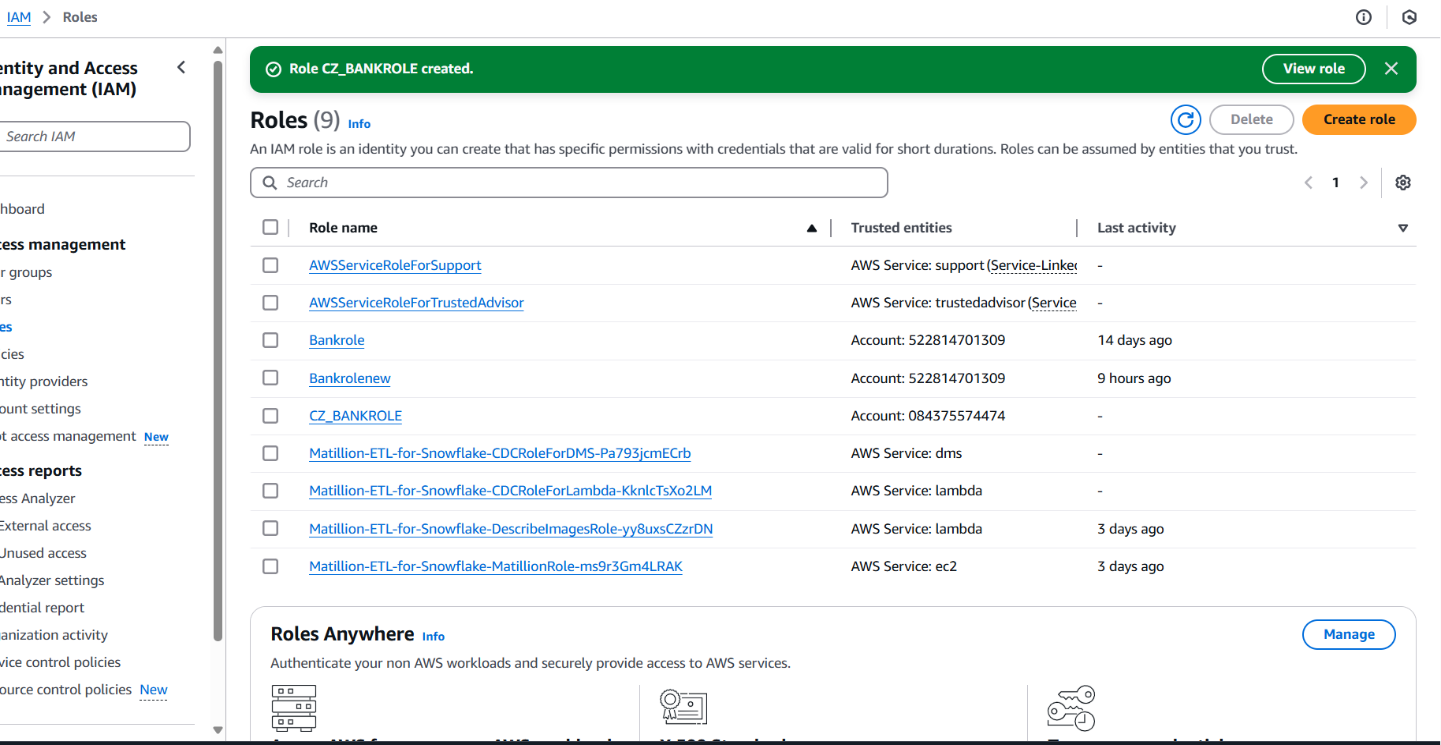


Figure : Role created

1. Note down the Role ARN, which we will need when we create the ‘Storage Integration’.

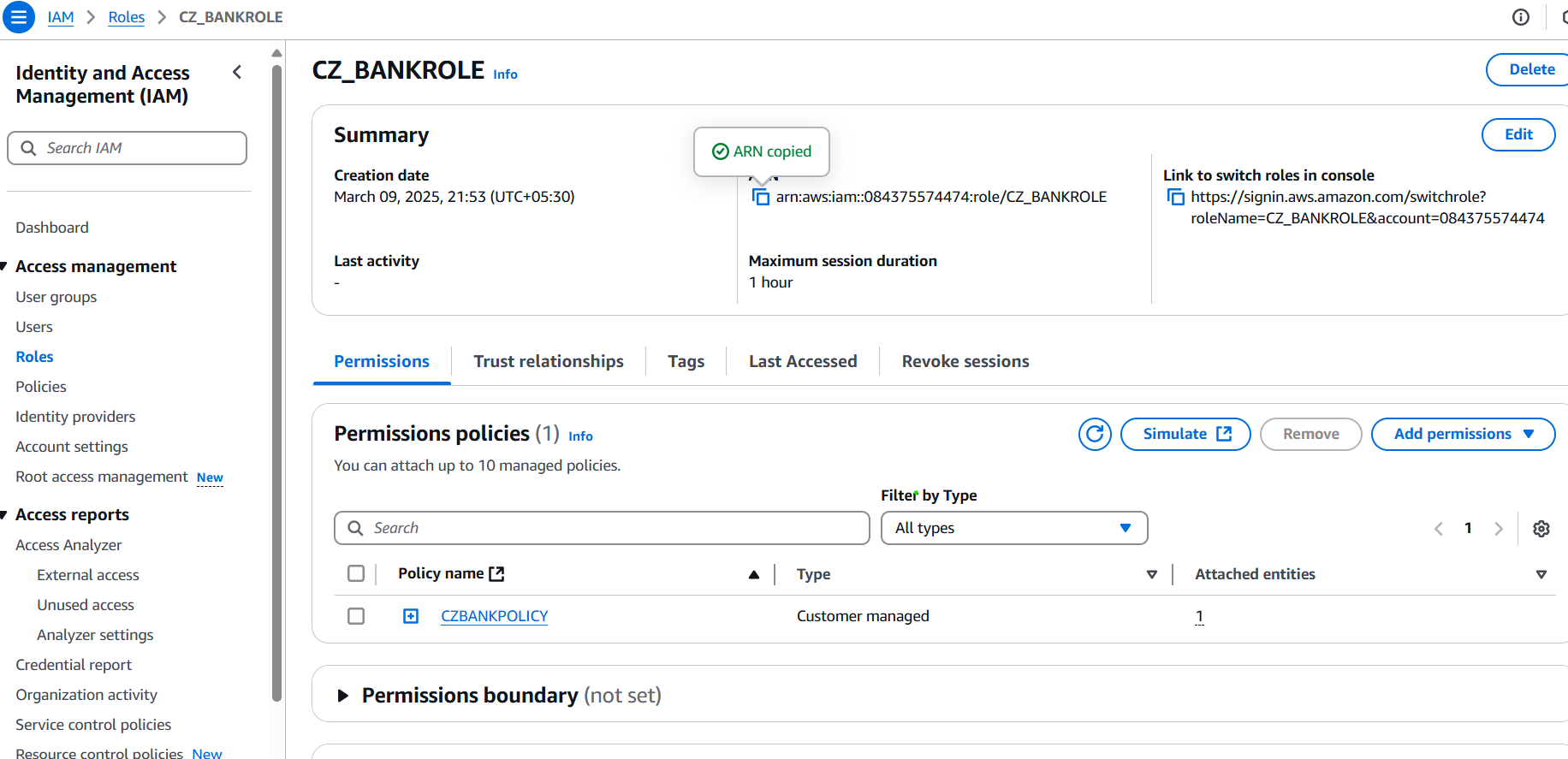


Figure : Note down Role ARN

1. Login to the Snowflake Account.

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

CREATE or REPLACE STORAGE INTEGRATION s3\_int

TYPE = EXTERNAL\_STAGE

STORAGE\_PROVIDER = S3

ENABLED = TRUE

STORAGE\_AWS\_ROLE\_ARN = 'arn:aws:iam::084375574474:role/BANK-ROLE'

STORAGE\_ALLOWED\_LOCATIONS = ('s3://cz-bank-bucket/');

1. In Snowflake worksheet run command

##Desc integration integration\_name;

DESC INTEGRATION s3\_int;

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

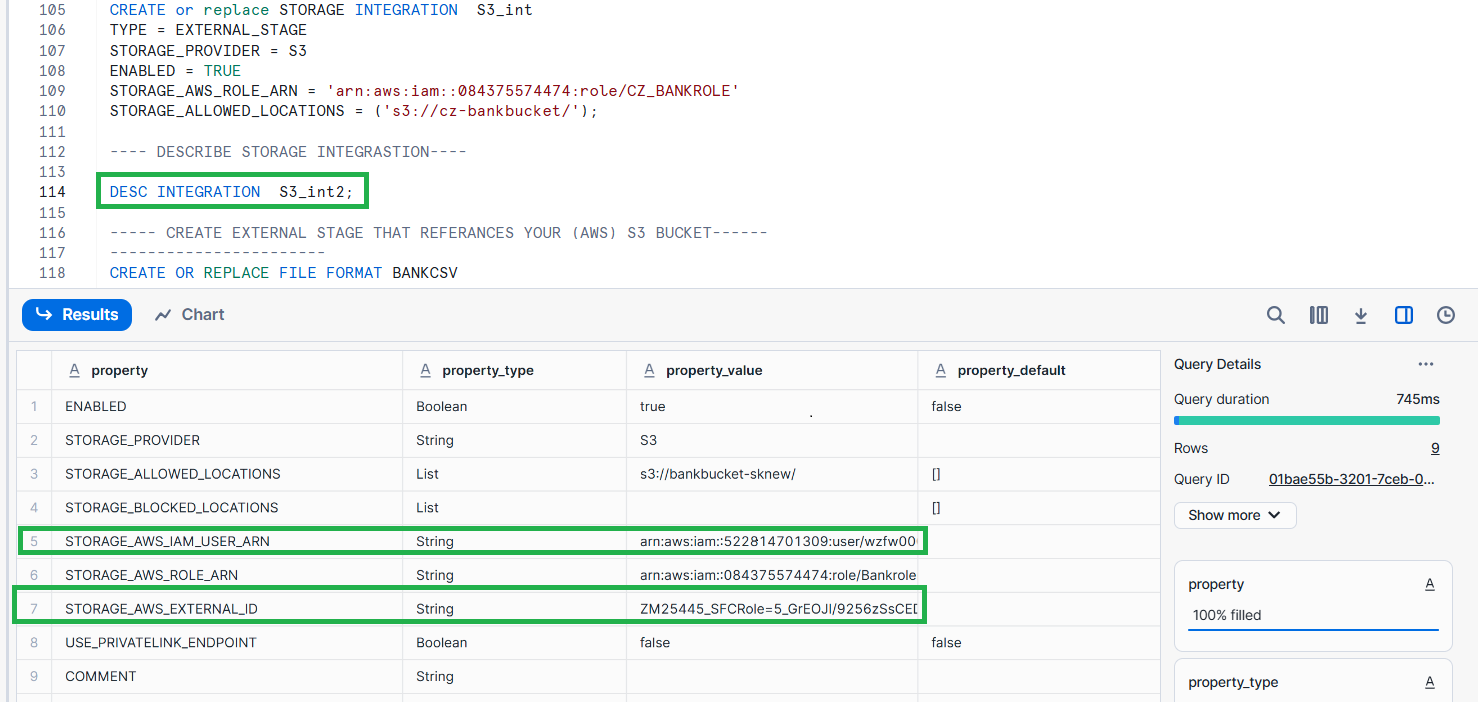


Figure : Note down IAM user and AWS externl\_id

1. Now go to the AWS Console, In IAM Role

* Select the role you created
* Click Trust Relationships -> Edit trust relationship
* 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
* Click Update Policy

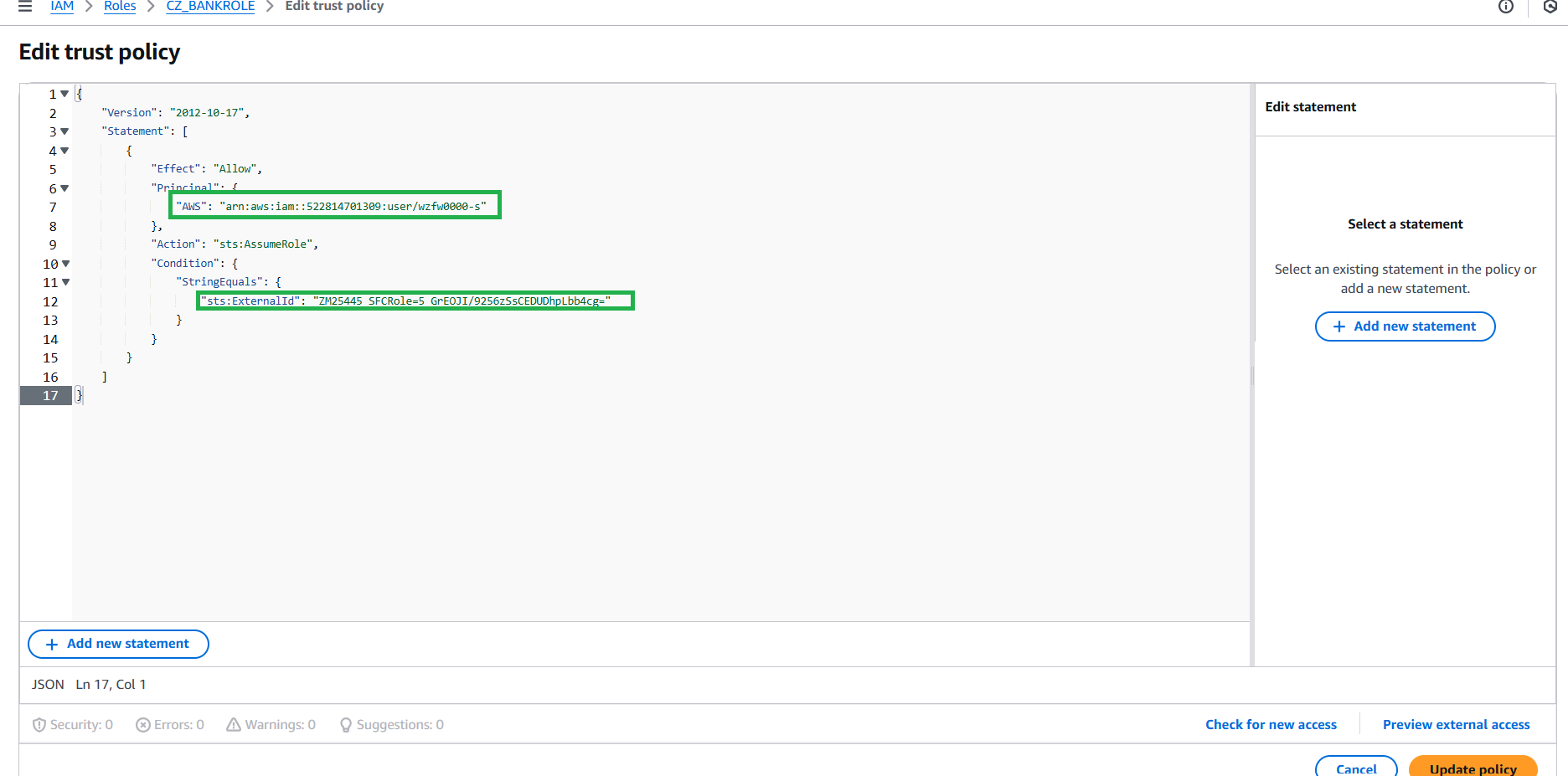


Figure : Edit trust policy

1. Create Snowflake file format. This file format will be used at the time of Stage creation.

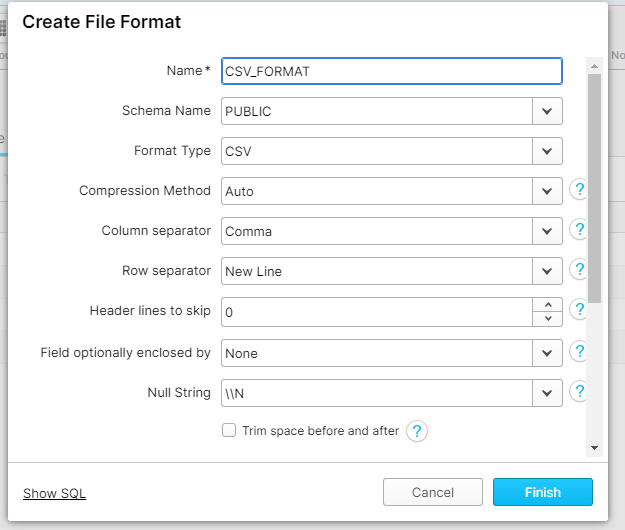


Figure : Create file format

OR we also create by using SQL code in snowflake.

CREATE OR REPLACE FILE FORMAT BANKCSV

TYPE = 'CSV'

FIELD\_DELIMITER = ','

SKIP\_HEADER = 1;

CREATE OR REPLACE FILE FORMAT BANKCSV

TYPE = CSV

FIELD\_OPTIONALLY\_ENCLOSED\_BY = ' " '

SKIP\_HEADER = 1;

1. Create a stage in snowflake pointing to your S3 bucket:

CREATE OR REPLACE STAGE BANKSTAGE

URL = 's3://cz-bank-bucket' -- (Name of your bucket)

FILE\_FORMAT = BANKCSV

STORAGE\_INTEGRATION = s3\_int;

1. Now Create auto-ingest Snowpipe.

CREATE OR REPLACE PIPE SNOWPIPE\_ACCOUNT

AUTO\_INGEST = TRUE

AS COPY INTO BANK.BANK\_SCHEMA.ACCOUNT --table name that you created in snowflake)

FROM @BANKSTAGE/ACCOUNT ---------s3 bucket subfolder name

FILE\_FORMAT = BANKCSV;

Create separated Snowpipe for each folders in s3 buckets.

1. After creating snowpipe, get ‘Notification Channel’ value

Run command

SHOW PIPES;

Or Go to Database 🡪 Pipes

Here also you will get the notification channel value.

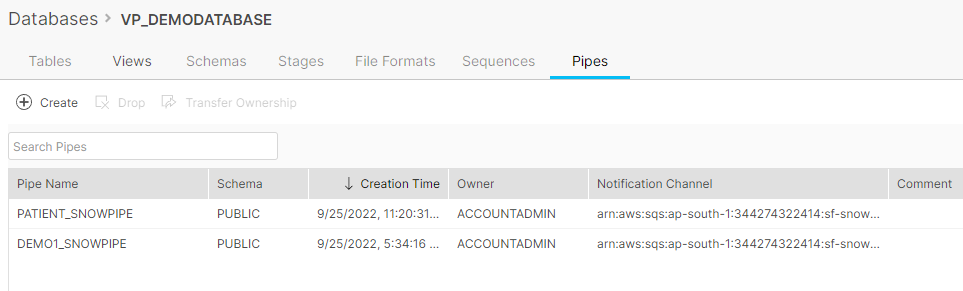


Figure : Pipe in snowflake database

1. This is the final step. 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.

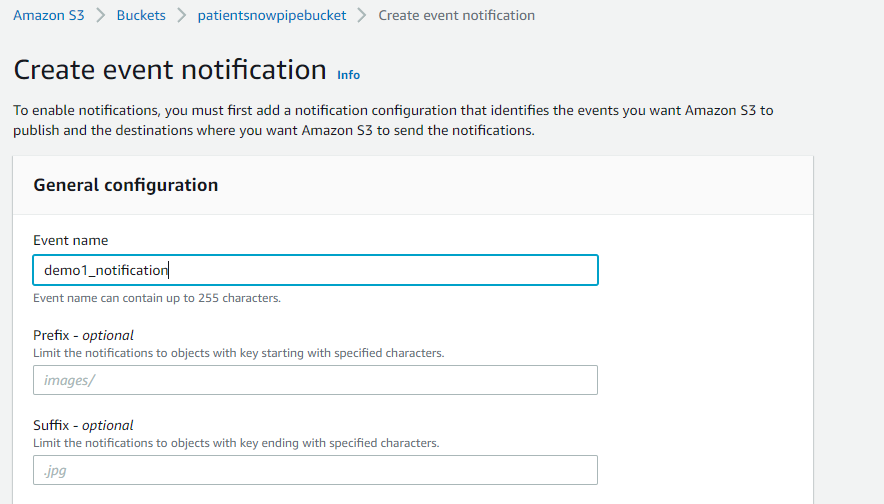


Figure : creating event notificatio

Check All Object create Events

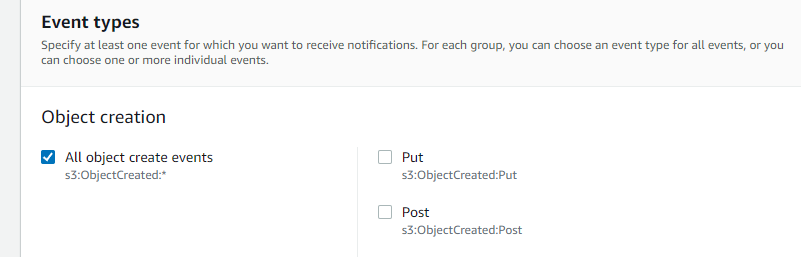


Figure : Object for event creation

Scroll down to Destination

**To create an SQS notification for the Snowpipe.**

**What is SQS?**

Amazon Simple Queue Service (SQS) lets you send, store, and receive messages between software components at any volume, without losing messages or requiring other services to be available.

As soon as a new file is uploaded to that folder in s3 it will send a notification to pipe. For that first run below command in snowflake

Select SQS Queue, Select Enter SQS Queue ARN. And paste that ‘Notification Channel’ under SQS Queue

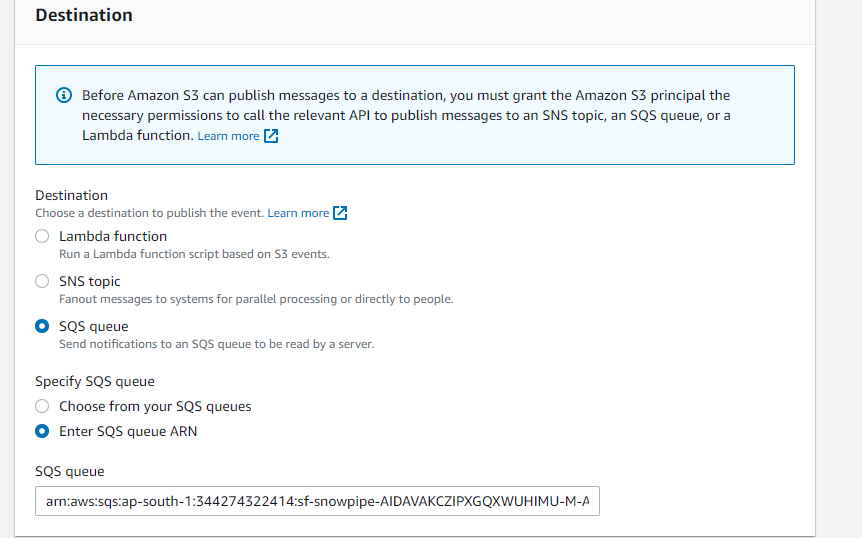


Figure : SQS event notification

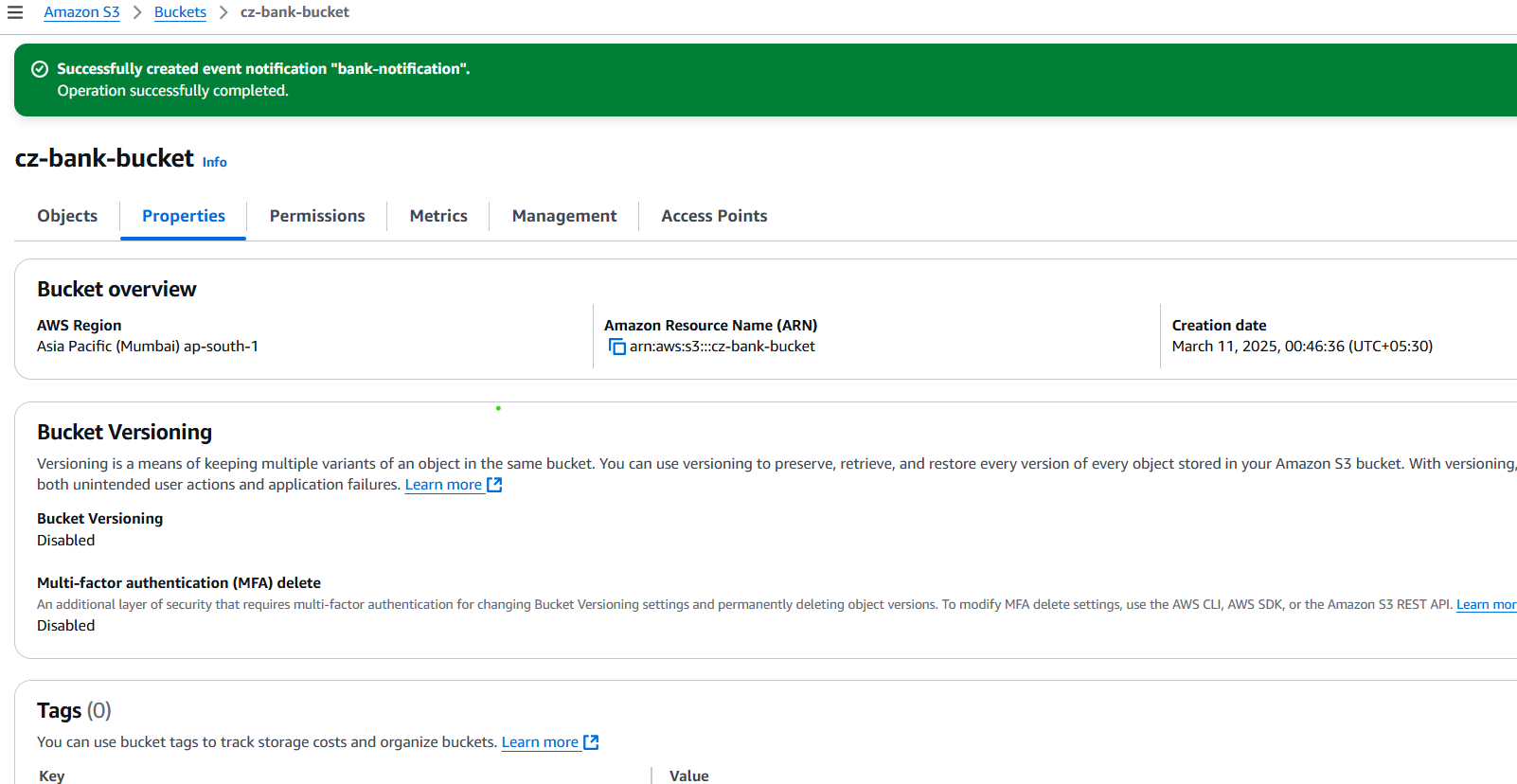


Figure : Event notification created

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

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

select SYSTEM$PIPE\_STATUS('patient\_snowpipe');

select \* from table(information\_schema.copy\_history(table\_name=>'tab\_patient', start\_time=>

dateadd(hours, -1, current\_timestamp())));

Thus, we can automatically ingest data into a warehouse from your S3 bucket to Snowflake