AWS S3 Archival Using Snowflake External Function

Data Sheet | LumenData

Problem Statement

1. S3 Files after being loaded into RAW tables Snowflake should be archived into Archive folder and the original source file should be deleted.
2. As Snowflake runs python in Sandbox, any external connections to AWS S3 or Rest API connections are not allowed. We are using External Functions provided by Snowflake.

Key Components

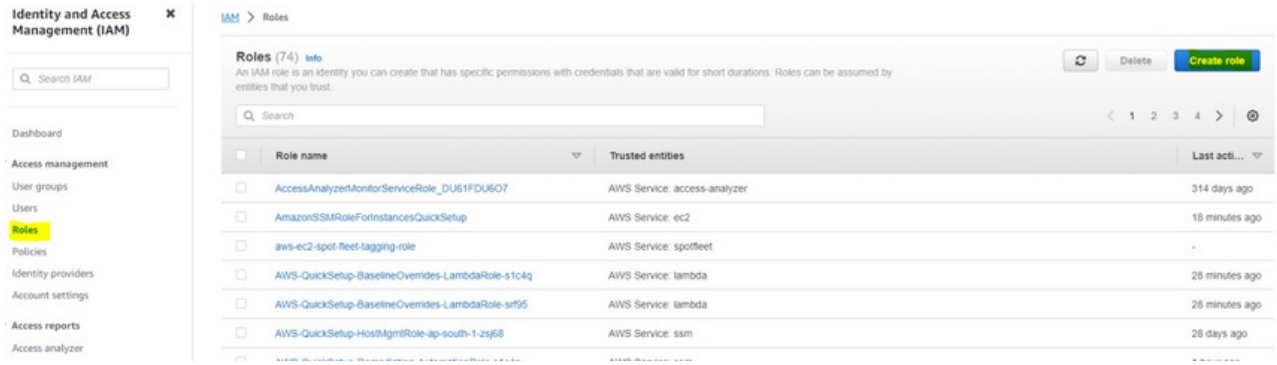
AWS IAM Roles
Snowflake External Functions
AWS Lambda
AWS API Gateway

Steps:

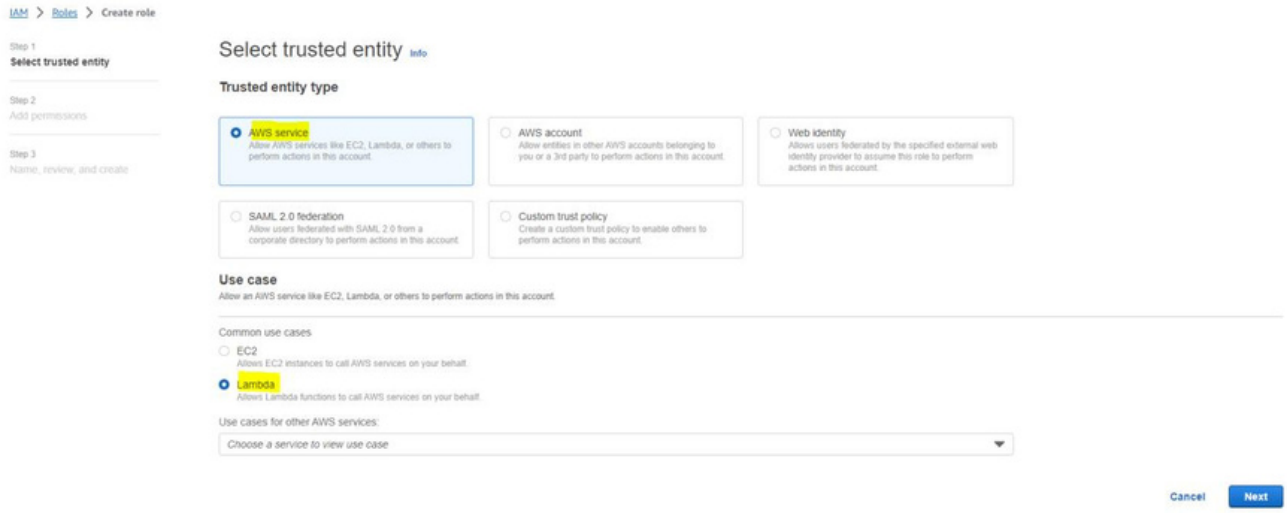
1. **Configuring AWS Lambda Role in AWS IAM**

LUMENDATA
Masters in Data Management

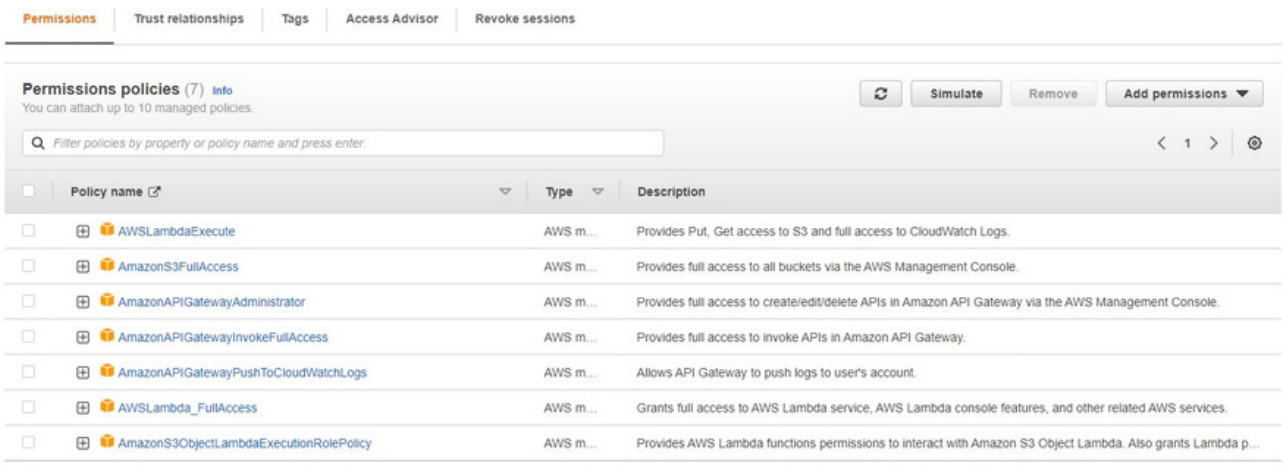
a) Navigate to IAM in AWS and Click on Create Role



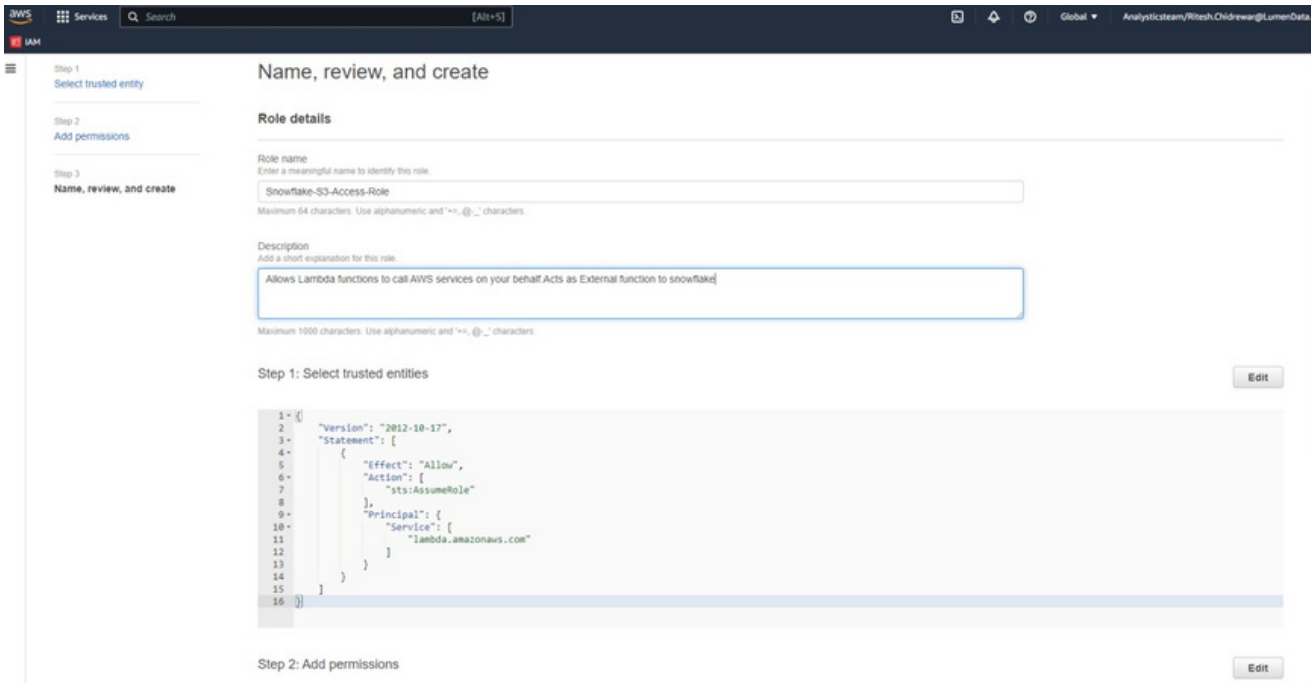
b) Select AWS service as Lambda



c) Attach the below policies to **Snowflake-S3-Access-Role**



d) Provide the Role name and Description (Optional) and click on Create Role.

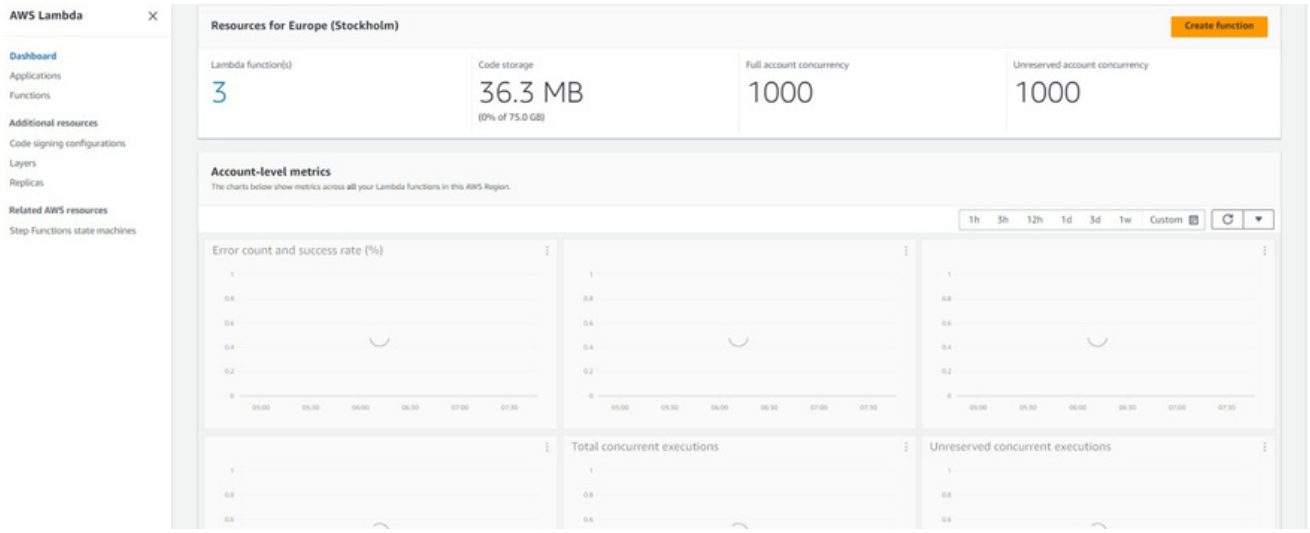


e) Make sure to note the Role Names and ARNs in notepad throughout the process.

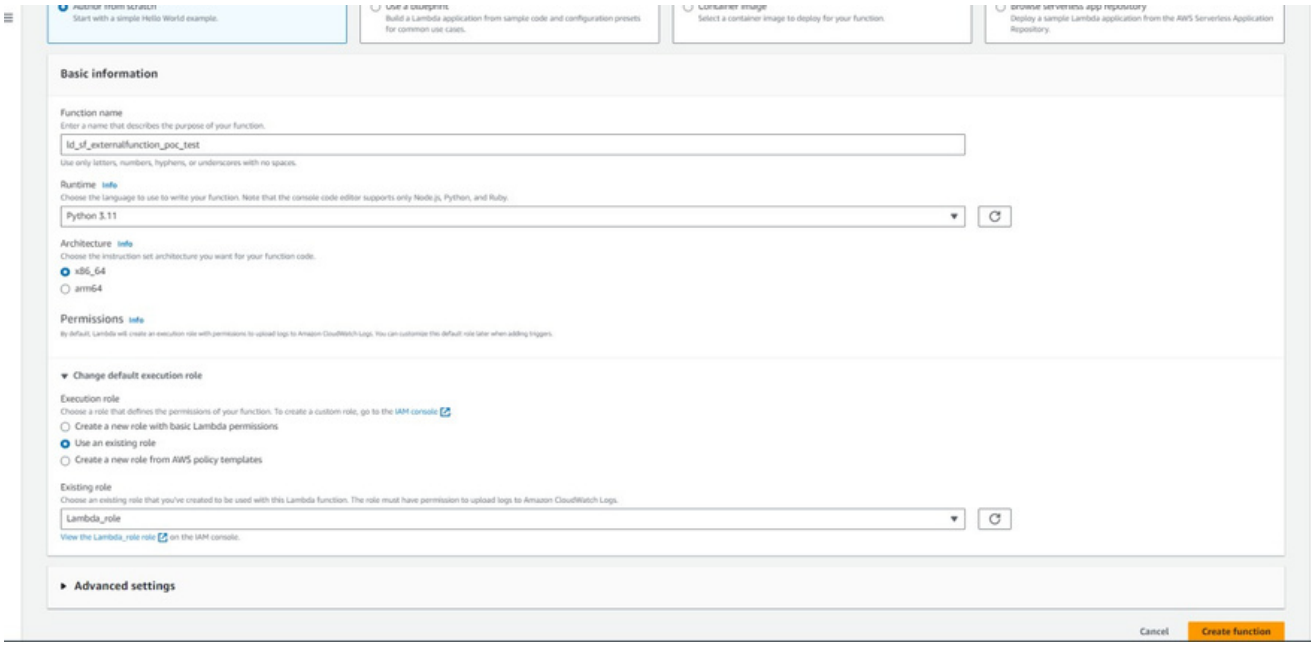
Lambda Service Role: **Snowflake-S3-Access-Role**

2. Configuring AWS Lambda Function

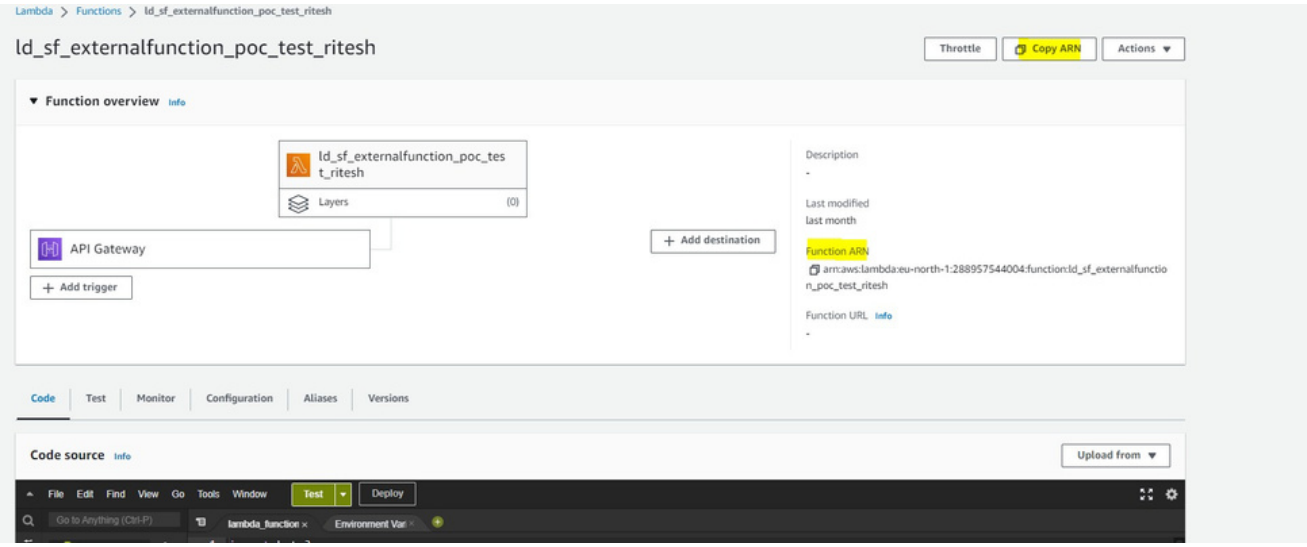
a) Navigate to AWS Lambda Service and Create a new function



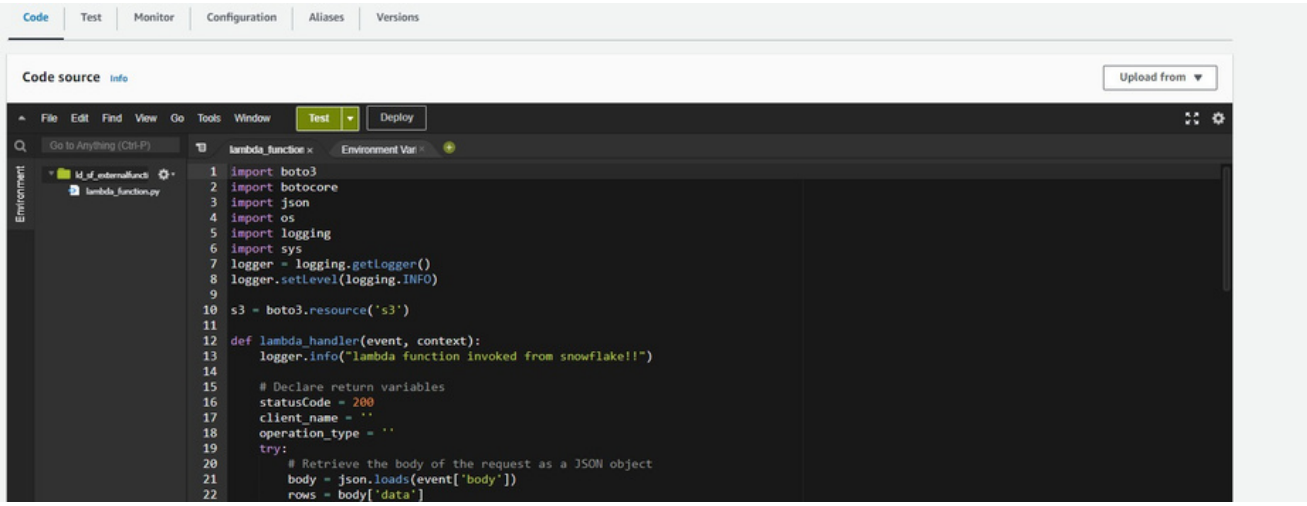
b) Provide the Function Name, Runtime as Python (Latest available), Architecture, Default Execution role (The role we created above for lambda) select from the dropdown.



c) As for us, we have already created one of the lambda functions. Please note the ARN of the Function



d) Code editor for lambda is available.



e) Snowflake Data Transfer Format

i. Snowflake transfers data to the lambda function in a specific format (JSON).

Ref link: [Snowflake-Data-Format]
(IMP)Please make sure you pass and retrieve the data in the above-mentioned format or else Snowflake won't be able to parse the result/data.

Ex. {
"body": "{\"data\" : [[0, \"client-name \", \"archive\"]]}"
}

ii. (IMP)Snowflake wraps the data into JSON which has JSONArray data inside it.

Below are the two rows with respective values. First array Value i.e., 0 corresponds to 1 st row and 1 represents the second row.

At the lambda side, we have to fetch the body object and iterate through each row to access the data.

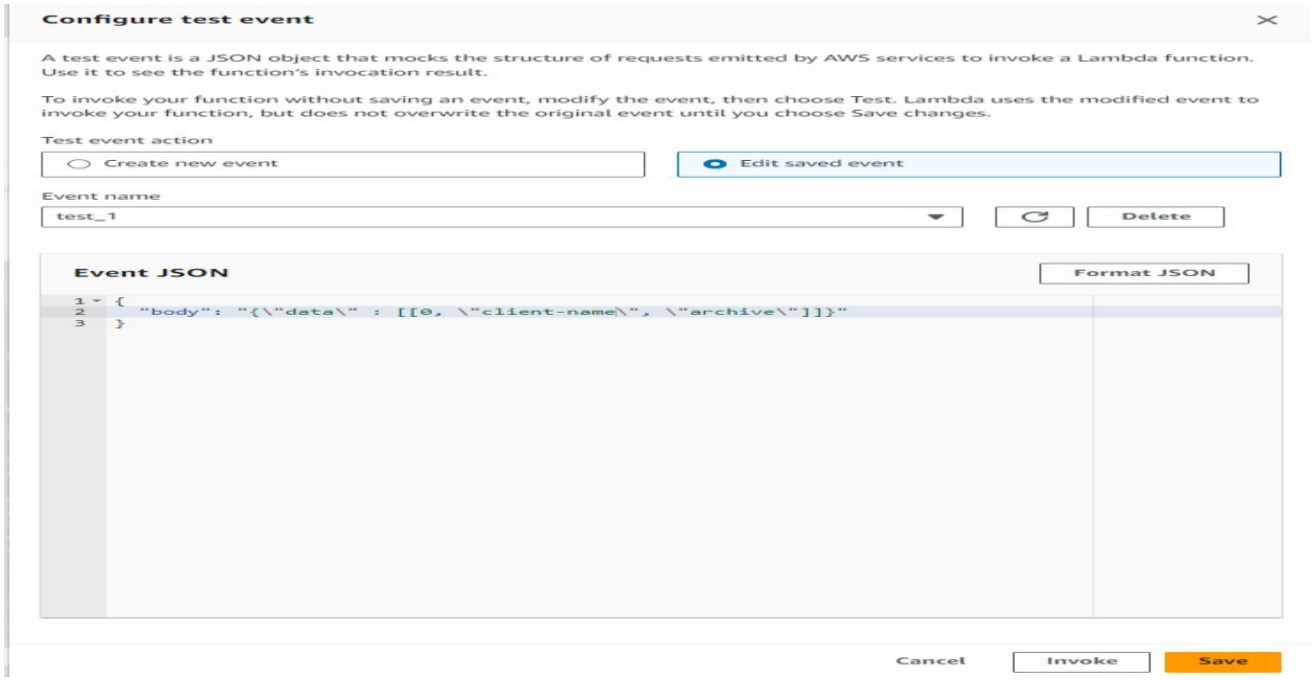
Ex: Snowflake data format- {“data”: [[0,”arg1”,arg2”], [1,”arg3”,”arg4”]]}

This entire structure is wrapped into JSON string while transferring data through Snowflake-Api Below is the final format.

Ex: {
“body”: “{“data”: [[0,”arg1”,arg2”],[1,”arg3”,”arg4”]]}”
}

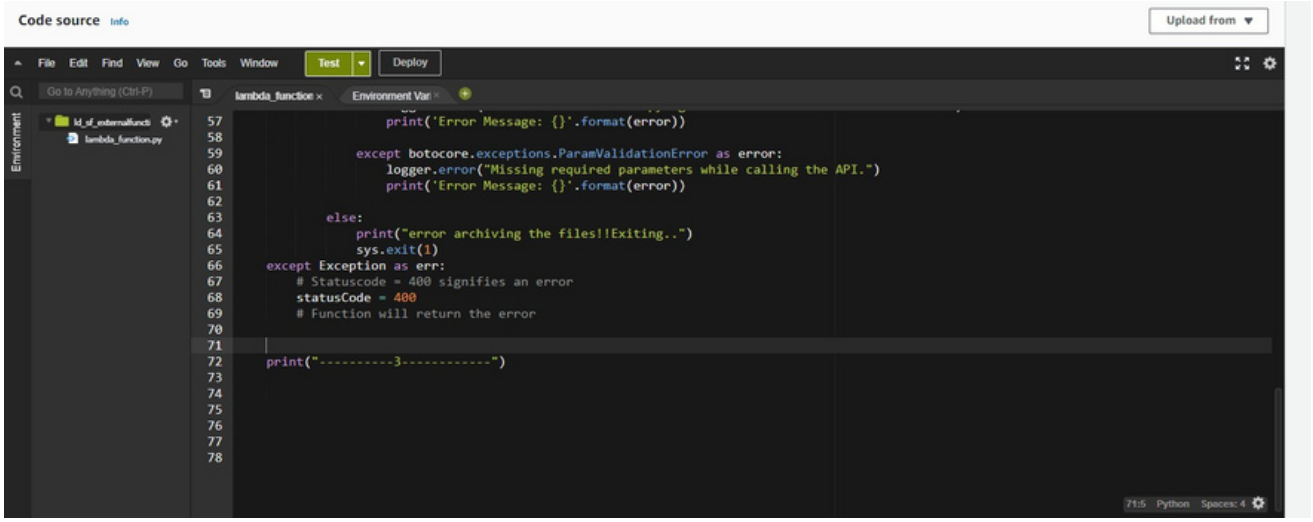
f) Configure the test event in the lambda function.

We provided the client’s name and the operation type(archive) to be performed on S3 file.

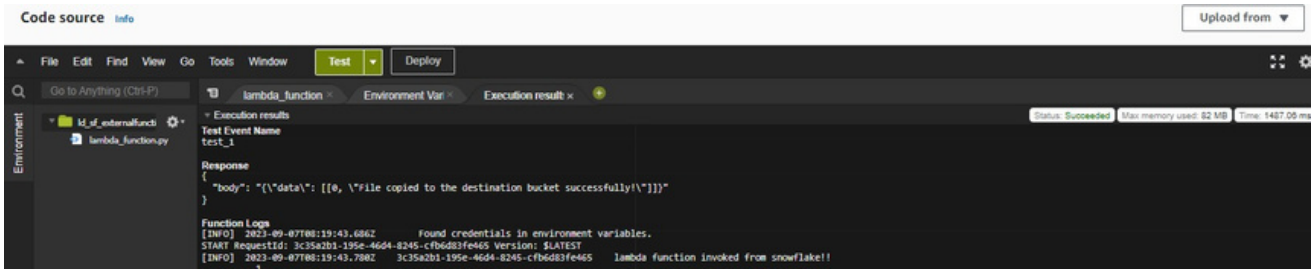


g) Sample S3 archival code

Once your code is ready, you can test it with the Test Event configured in the previous steps.



Response received after successful Copy with test_event



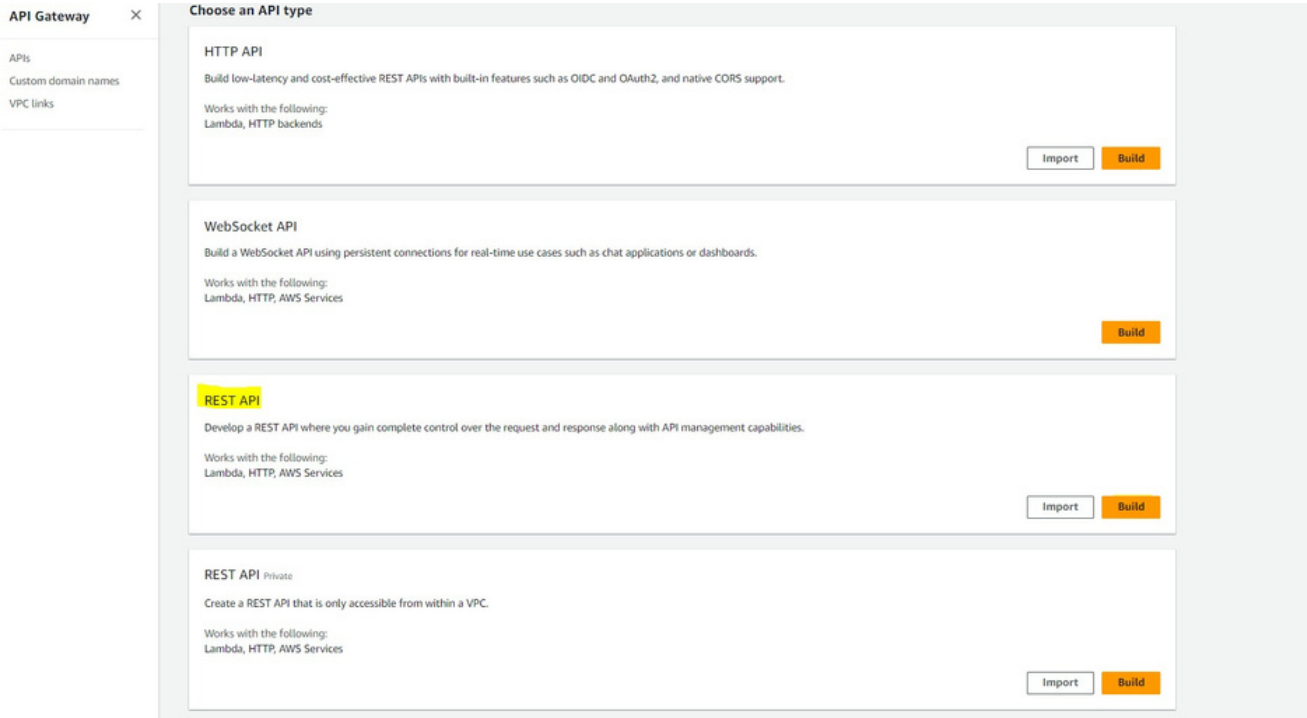
h) Now we have tested the code successfully, we can add the function name and ARN to our notepad documentation.

Lambda Service Role: Snowflake-S3-Access-Role
Lambda Functionld_sf_externalfunction_poc
Lambda Function ARN: arn:aws:lambda:eu-north-1:288957544004:function:ld_sf_externalfunction_poc_test

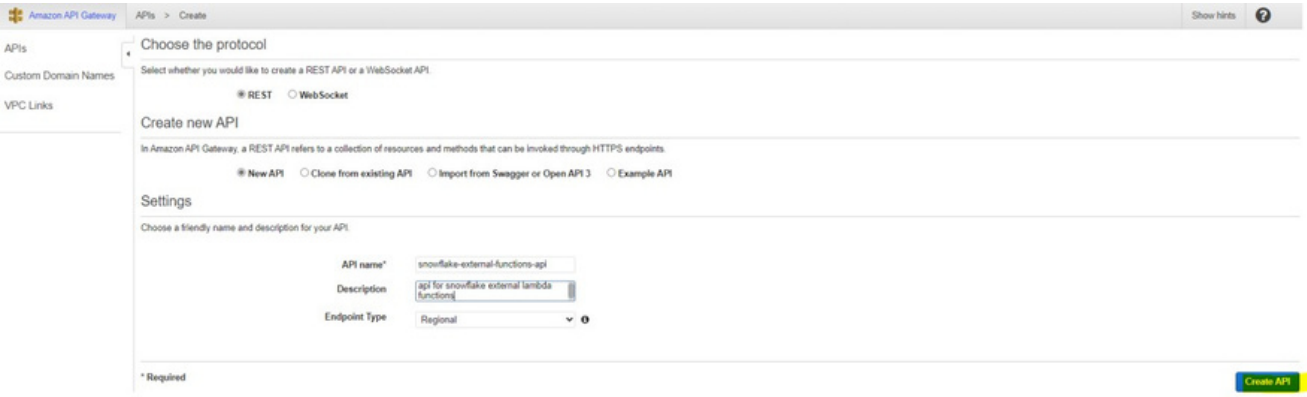
3. Configuring AWS API Gateway

a) Create the API Gateway to expose the Lambda function created.

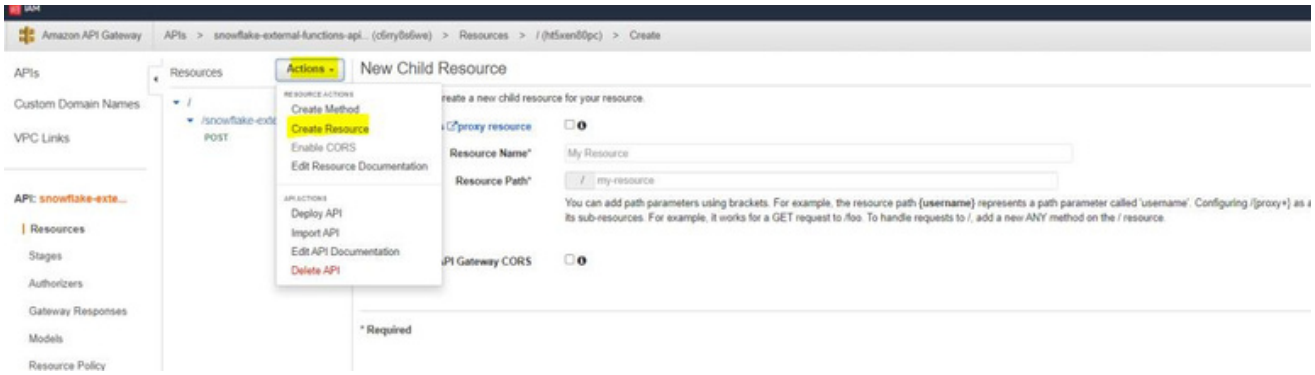
i. Select REST API and click on build.



ii. Provide the API name and Description. Click on Create API.

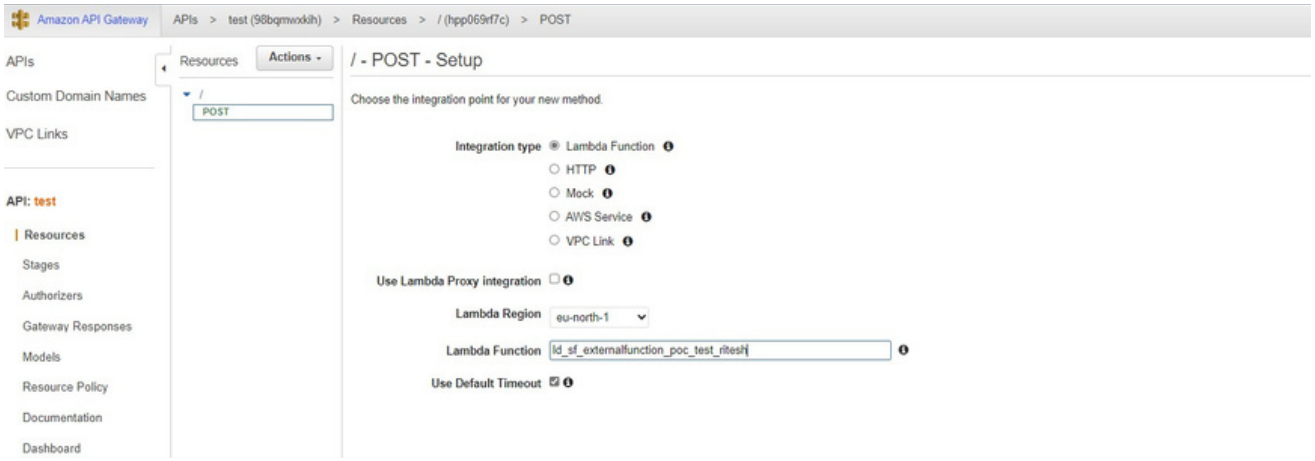


iii. Create a Resource and Method. Provide the resource name and Click on Create Resource.

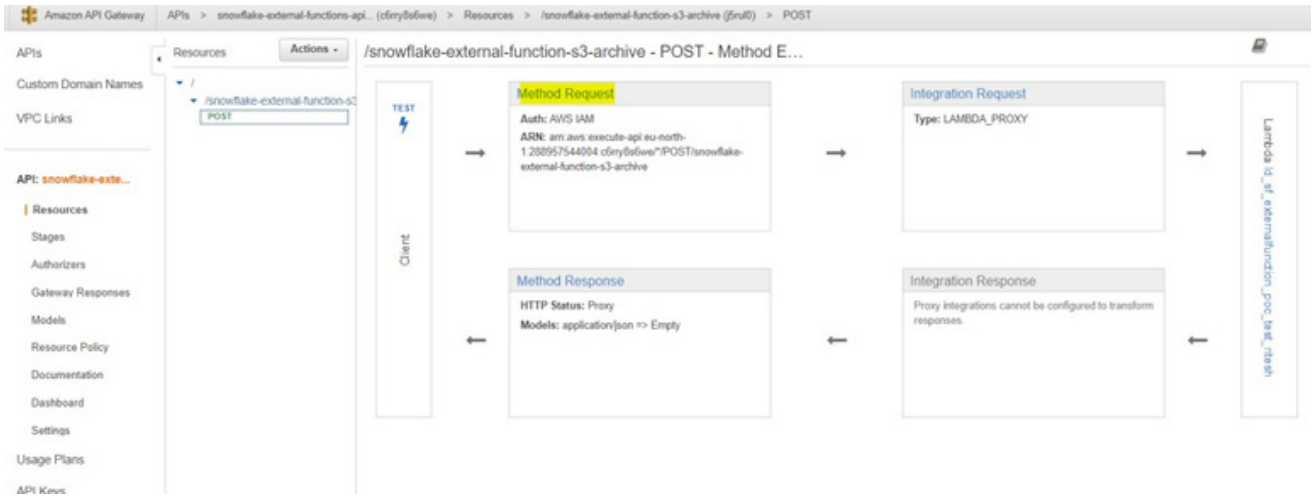


iv. Create Method.

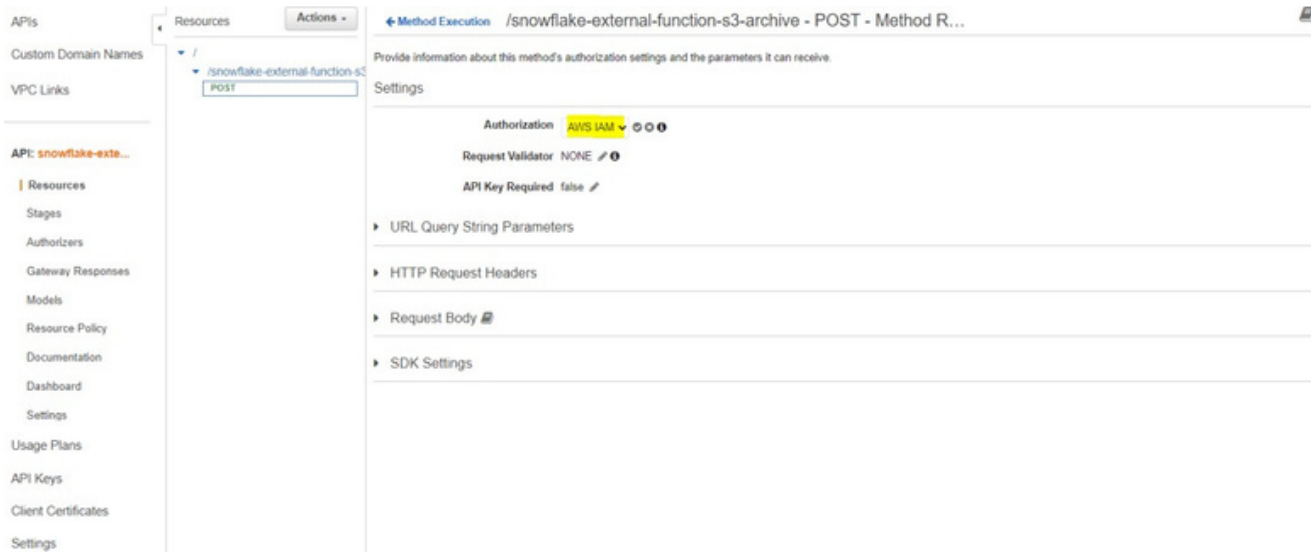
After creating the method, a warning will pop up. Click on Ok.



Once the Method is created, Click on Method Request



Inside Method request, select Authorization as IAM



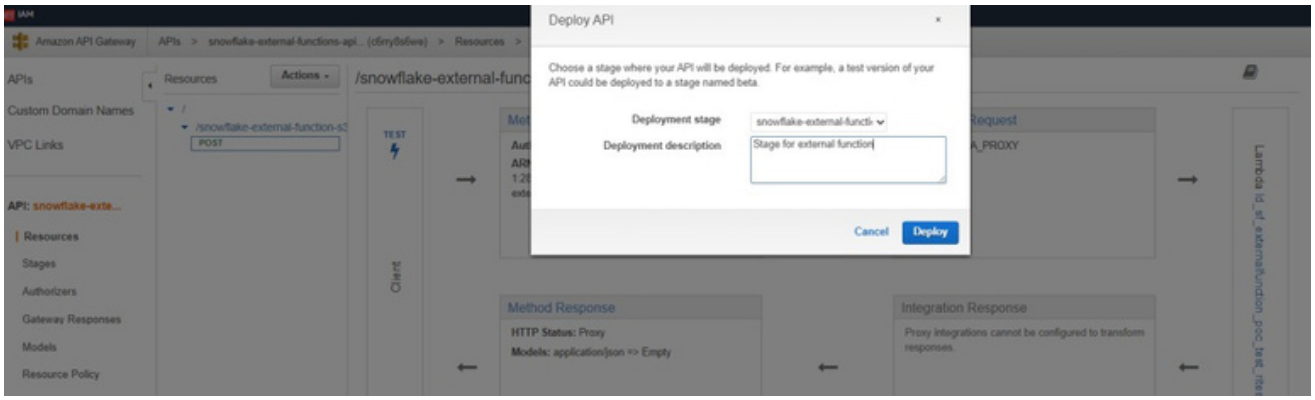
Once authorization is set, return to Method overview screen, and you can see the ARN populated.

Note: Add this ARN to the notepad Documentation

Lambda Service Role: Snowflake-S3-Access-Role
Lambda Functionld_sf_externalfunction_poc
Lambda Function ARN: arn:aws:lambda:eu-north1:288957544004:function:ld_sf_externalfunction_poc_te st
Lambda Method ARN : arn:aws:execute-api:eu-north1:*******:******/*/POST/snowflake-external-functions3-archive

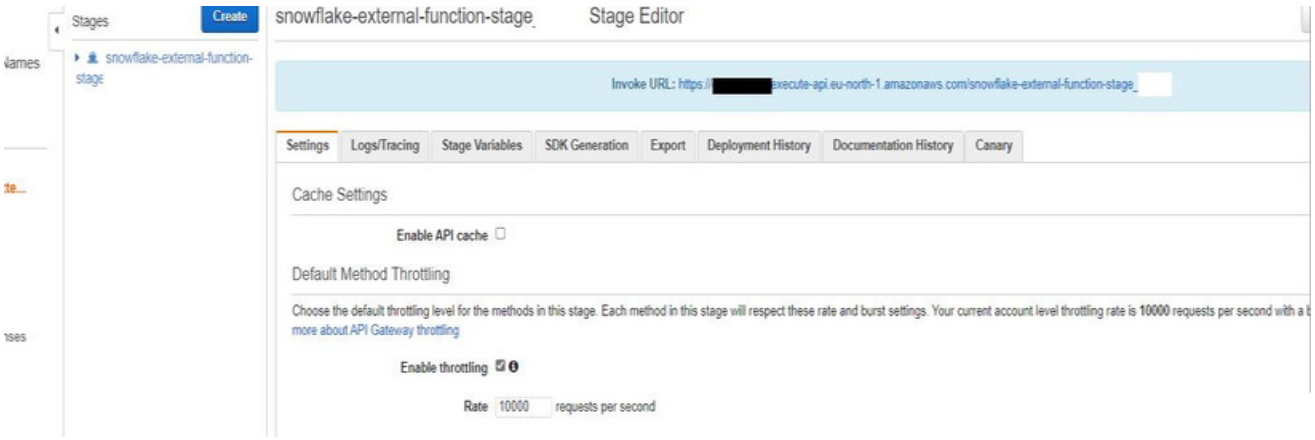
b) Deploy API created to Stage

i. Provide the Deployment stage name and description



ii. Once deployed, we get the Stage invoke URL. This is the URL through which the API is accessed by the authorized roles.

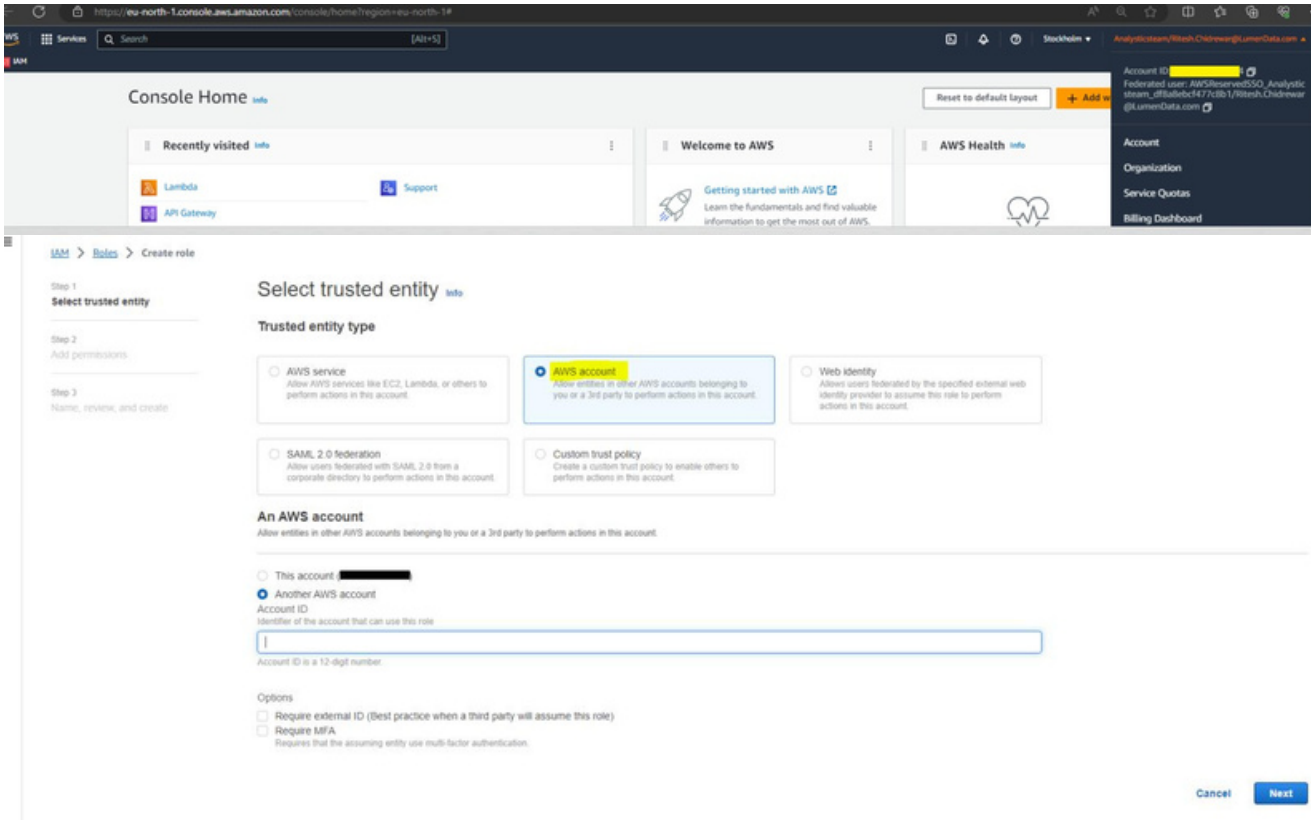
Note this in notepad documentation.



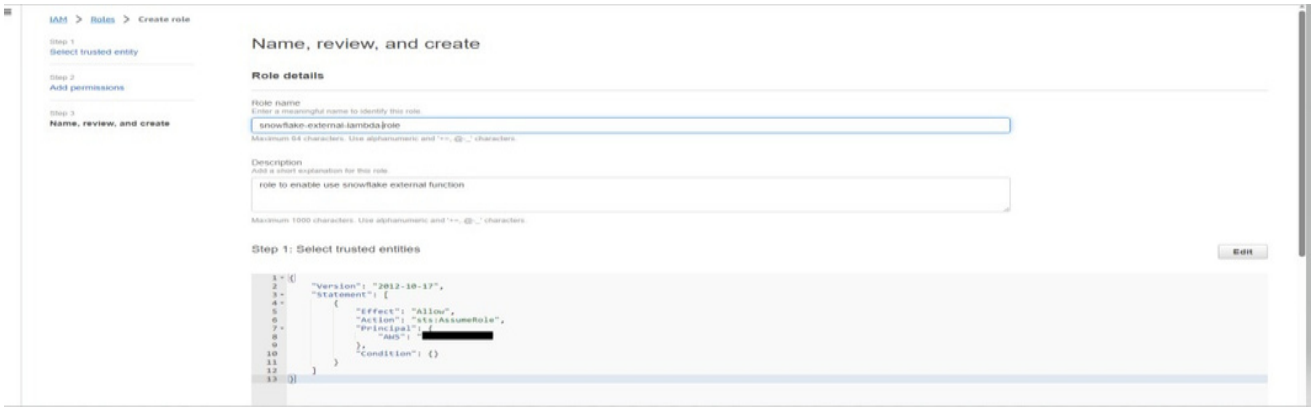
4. Configuring another AWS IAM role to enable handshake between Snowflake and API Integration.

Follow the same steps for IAM creation as mentioned above. Also, attach the same policies

a) For IAM creation, select Another AWS Account. Copy the account ID as shown in the screenshot.



b) Provide the Role name and Description.



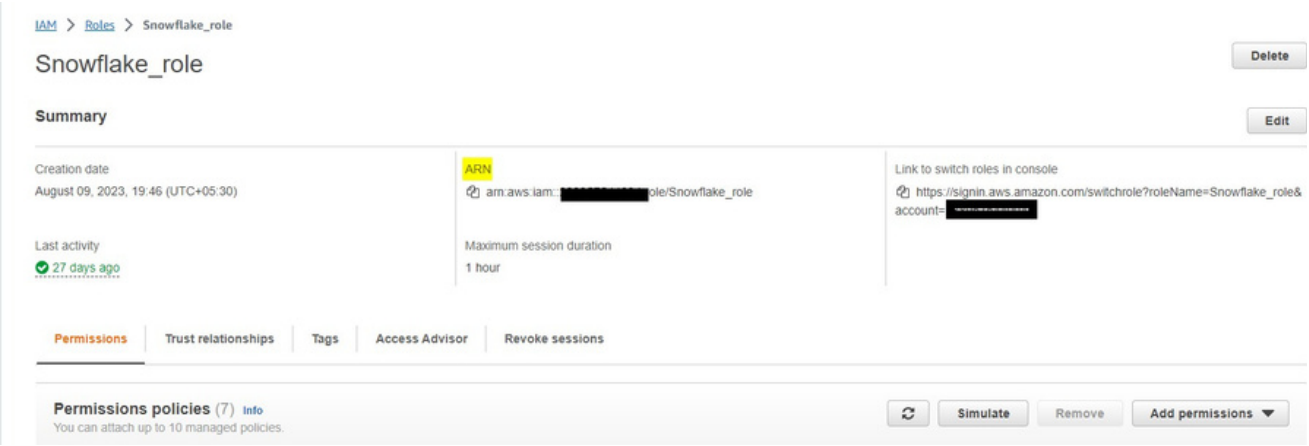
Note Role name in our Notepad Documentation.

Lambda Service Role: Snowflake-S3-Access-Role
Lambda Functionld_sf_externalfunction_poc
Lambda Function ARN: arn:aws:lambda:eu-north-1:*******:function:ld_sf_externalfunction_poc_test
Lambda Method ARN : arn:aws:execute-api:eu-north1:*******:******/*/POST/snowflake-external-function-s3-archive

Account ID: 012345678910 IAM

Account Role: snowflake-external-lambda-role

c) Retrieve the Role ARN for the above role that you've created.
Sample screenshot



Note ARN in our Notepad Documentation.

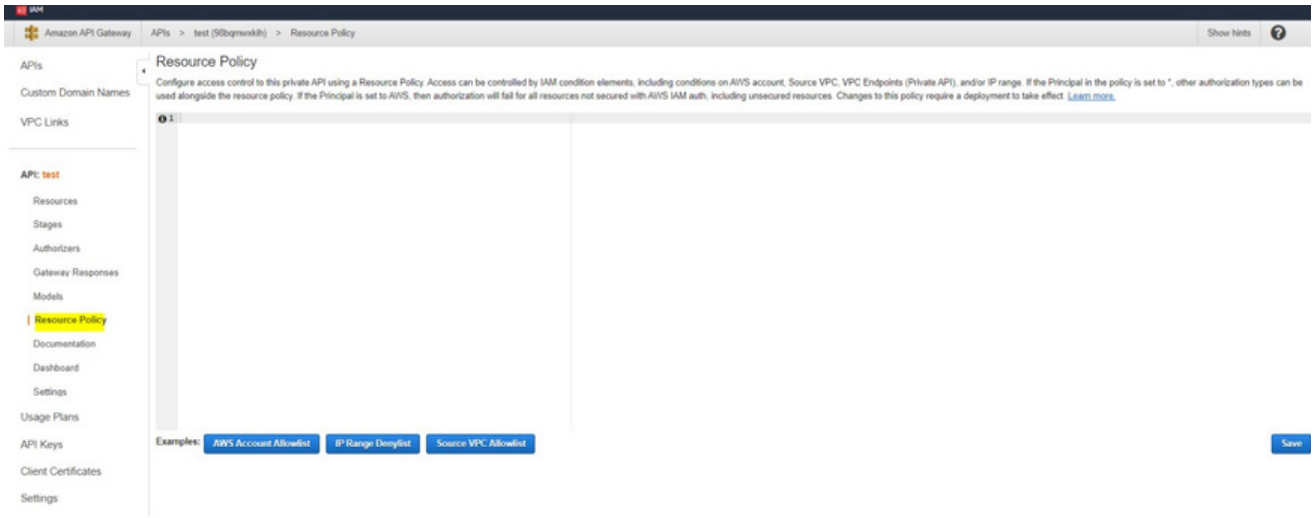
Lambda Service Role: Snowflake-S3-Access-Role
Lambda Functionld_sf_externalfunction_poc
Lambda Function ARN: arn:aws:lambda:eu-north-1:*******:function:ld_sf_externalfunction_poc_test

Lambda Method ARN : arn:aws:execute-api:eu-north1:*******:******/*/POST/snowflake-external-function-s3-archive

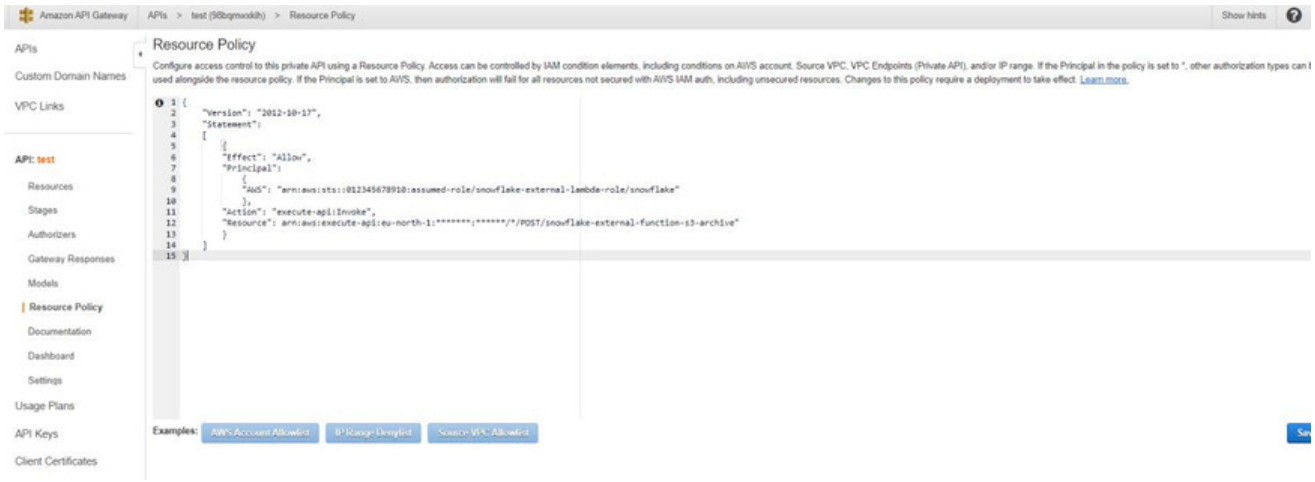
Account ID: 012345678910 IAM
Account Role: snowflake-external-lambda-role
IAM Account Role ARN:
arn:aws:iam::012345678910:role/snowflake-external-lambdarole

d) Updating the Resource policy in API Gateway

To authorize the role created, we have to update the resource policy in the API Gateway.



Replace the account_id, IAM account Role, and Method ARN with the values from our documentation.



5. Snowflake Configuration and Setup

a) Attached below is the doc for snowflake commands and queries used for setting up the connection and creation of External function.

i. Create the API Integration object in Snowflake.

Once the DESCRIBE query is executed, we get the below Snowflake credentials .

Copy **API_AWS_IAM_USER_ARN and API_AWS_EXTERNAL_ID** in our notepad documentation sheet.

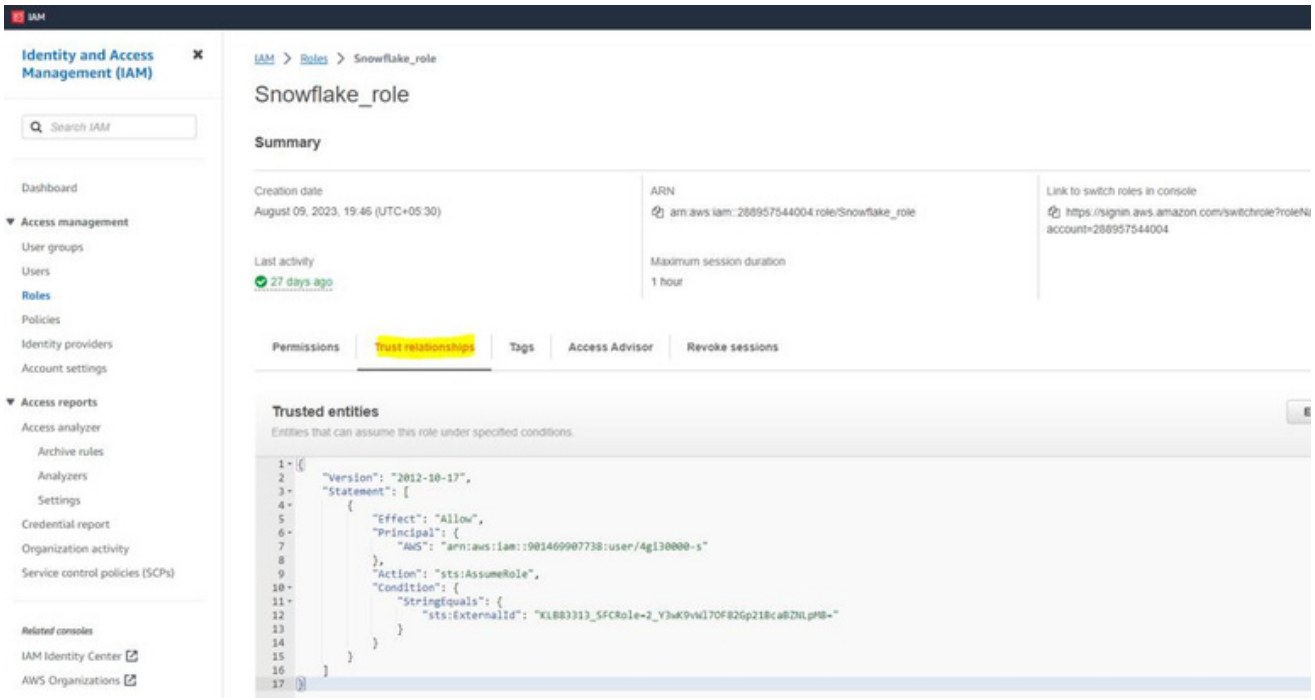
Results | Chart

	property	property_type	property_value	property_default
1	ENABLED	Boolean	true	false
2	API_KEY	String		
3	API_PROVIDER	String	[redacted]	
4	API_AWS_IAM_USER_ARN	String		
5	API_AWS_ROLE_ARN	String		
6	API_AWS_EXTERNAL_ID	String		
7	API_ALLOWED_PREFIXES	List		[]
8	API_BLOCKED_PREFIXES	List		[]
9	COMMENT	String		

Lambda Service Role: Snowflake-S3-Access-Role
Lambda Functionld_sf_externalfunction_poc
Lambda Function ARN: arn:aws:lambda:eu-north-1:*******:function:ld_sf_externalfunction_poc_test
Lambda Method ARN : arn:aws:execute-api:eu-north1:*******:******/*/POST/snowflake-external-functions3-archive
Account ID: 012345678910
IAM Account Role: snowflake-external-lambda-role
IAM Account Role ARN: arn:aws:iam::012345678910:role/snowflake-externallambda-role

API Integration Name:
ld_sf_aws_external_function_api_integration
api_aws_iam_user_arn:
arn:aws:iam::109876543210:user/xxxx-x-xxxxxxxx
api_aws_external_id:
DEMO_SFCRole=0_x0xxxxxxxxxxxxxx+xx0xx0xxx=

ii. Update the Trust Relationship in AWS IAM Role



iii. Once the trust relationship is updated and saved, return to Snowflake and create the external function. (Attached doc above)

iv. Test the function and check the output and file moved successfully to the archive location in S3.

```
--Test/call external function using below query
select LD.PUBLIC.ld_sf_aws_external_function('client_name','archive');
```

ABOUT LUMENDATA:

LumenData is a leading provider of Enterprise Data Management, Cloud & Analytics solutions. We help businesses navigate their data visualization and analytics anxieties and enable them to accelerate their innovation journeys. Founded in 2008, with locations in multiple countries, LumenData is privileged to serve over 100 leading companies, including KwikTrip, Versant Health, US Food & Drug Administration, US Department of Labor, Cummins Engine, BCG, and others. LumenData is SOC2 certified and has instituted extensive controls to protect client data, including adherence to GDPR and CCPA regulations.

Get in touch to discuss how we can facilitate data-driven transformation for your organization.

MEET OUR AUTHORS

Ritesh Chidrewar
Senior Consultant

Ankit Kumar
Technical Lead