Se

**Serverless S3 File Downloader Using Lambda and API Gateway**

**DE**

Document Version 1.0

Problem Statement | Value Addition

Serverless S3 File Downloader Using Lambda, and API Gateway

TYPE

Internal Project

VERSION

1.0

DEPARTMENT

Data Engineering

CATEGORY

Technology

By using simple steps, we can create a convenient API to securely fetch files directly from Amazon S3.

**WHAT IS INCLUDED**

Managing and securely sharing files stored in Amazon S3 can be cumbersome, requiring direct navigation through the S3 console and lacking efficient bulk download options. The absence of a streamlined solution poses challenges in user accessibility, security, and scalability.

**PROBLEM STATEMENT**

# 

**VALUE ADDITION**

The proposed serverless S3 file downloader enhances user experience by eliminating the need for direct S3 console navigation, providing efficient bulk download options, and ensuring security through IAM roles and pre-signed URLs. Leveraging AWS Lambda and API Gateway offers scalability, cost-effectiveness, and seamless integration. This concept adds significant value by simplifying file management, improving security controls, and optimizing the overall efficiency of handling S3 files.

Contents

[**Purpose of the Document:** 3](#_Toc156399724)

[**Description:** 3](#_Toc156399725)

[**Objective:** 3](#_Toc156399726)

[**Technology Stack:** 3](#_Toc156399727)

* [**AWS S3:** 3](#_Toc156399728)
* [**Supported Storage Tiers for Direct Downloads**: 3](#_Toc156399729)
* [**Limitations for Deep Archive and Flexible Retrieval Tiers:** 4](#_Toc156399730)
* [**AWS IAM (Identity and Access Management):** 4](#_Toc156399731)
* [**AWS Lambda:** 4](#_Toc156399732)
* [**AWS API Gateway:** 4](#_Toc156399733)

[**Architecture:** 5](#_Toc156399734)

[**Implementation:** 5](#_Toc156399735)

* [**S3 BUCKET CREATION** 5](#_Toc156399736)
* [**IAM ROLE CREATION** 6](#_Toc156399737)
* [**LAMBDA FUNCTION** 8](#_Toc156399738)
* [**API GATEWAY** 11](#_Toc156399739)

[**Output**: 16](#_Toc156399740)

[**Advantages:** 16](#_Toc156399741)

[**Disadvantages:** 16](#_Toc156399742)

[**Conclusion:** 17](#_Toc156399743)

## **Purpose of the Document:**

This document outlines the design and implementation of a serverless S3 file downloader using Python, Lambda, and API Gateway. The purpose is to address the identified challenges by providing a user-friendly, secure, and scalable solution for managing and downloading files from S3. The document serves as a guide for understanding the architecture, security considerations, and advantages of implementing this serverless approach.

## **Description:**

Streamline your S3 file management with our secure, serverless downloader! Built on AWS Lambda, Python, and API Gateway, it prioritizes security, cost-effectiveness, and user-friendliness. Here is how it works:

**Scales effortlessly, costs less**: Lambda handles computing needs flexibly, charging only for actual usage. S3 ensures data integrity, while dedicated IAM roles and pre-signed URLs protect access. No more fixed server costs!

**User-friendly features, bulk operations:** Easily download files without S3 console navigation. Search and download multiple files at once—a feature S3 itself lacks.

**Secure file sharing, temporary links:** Integrates seamlessly with pre-signed URLs for secure sharing without exposing your bucket. Customize expiration times and permissions for granular control.

**Wide range of uses:** Perfect for secure document sharing, public file downloads, and API-driven file delivery.

**Built-in security, API Gateway, and Lambda:** Benefit from the robust security features of API Gateway and Lambda to safeguard your sensitive files.

Experience a more secure, efficient, and cost-effective way to manage, share, and download your S3 files with our serverless downloader and pre-signed URLs!

**Objective:**

The objective of this architecture is to build a serverless file download system where users can download files from an Amazon S3 bucket through an API Gateway endpoint triggered by an AWS (Amazon Web Services) Lambda function. This approach eliminates the need for provisioning and managing servers, offering a cost-effective and scalable solution.

## **Technology Stack:**

**AWS S3:** Simple Storage Service for storing and retrieving files. Lambda functions can interact with S3 to download files.

## **Supported Storage Tiers for Direct Downloads**:

* + Standard
  + Intelligent-Tiering
  + Infrequent Access (IA)
  + One Zone-Infrequent Access (One Zone-IA)
  + Glacier Instant Retrieval

## **Limitations for Deep Archive and Flexible Retrieval Tiers:**

1. Direct downloads not supported:
   * Unlike files in Standard and other accessible tiers, files stored in Deep Archive and Flexible Retrieval cannot be directly downloaded through the current architecture. Downloading them requires additional steps.
2. Retrieval process and considerations:

Deep Archive:

* + Accessing Files: You need to create a separate task (retrieval job) to get your files. This creates extra cost and takes time (usually hours).
  + Downloading process: After the job is done, you will need to set up a separate Lambda function or background process to download the file. This adds extra complexity to your workflow.

Flexible Retrieval:

* + Access Options: You have two choices for getting your files: Standard (3-5 hours) and Bulk (5-12 hours). Both costs more than the standard tier downloads.
  + Starting a retrieval: Like Deep Archive, you need to initiate these retrievals outside of your current Lambda function flow using the same tools.

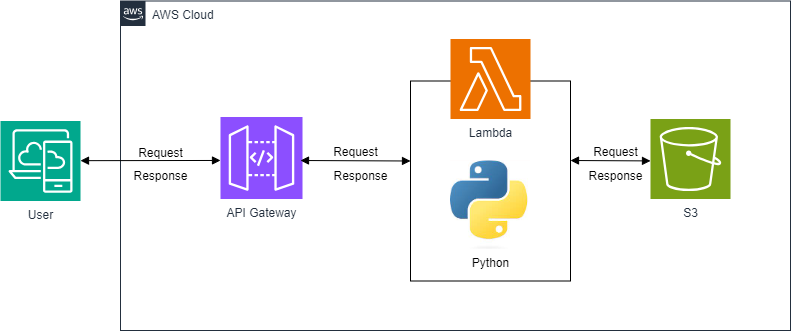
1. Performance and cost implications:
   * Downloading from Deep Archive and Flexible Retrieval adds extra steps and potential delays compared to other tiers, impacting responsiveness.
   * The retrieval process incurs additional costs, which can become significant for frequent downloads or large files.
2. Increased complexity and management:
   * Integrating retrieval jobs and potentially additional Lambda functions adds complexity to your architecture and requires separate management and monitoring.

**AWS IAM (Identity and Access Management):** Identity and Access Management for securing and managing access to AWS services.

**AWS Lambda:** Serverless compute service for executing code in response to events. Handles the file download logic.

**AWS API Gateway:** Fully managed service for creating and managing RESTful APIs. Servers as the HTTP endpoint for clients to trigger Lambda functions.

## **Architecture:**

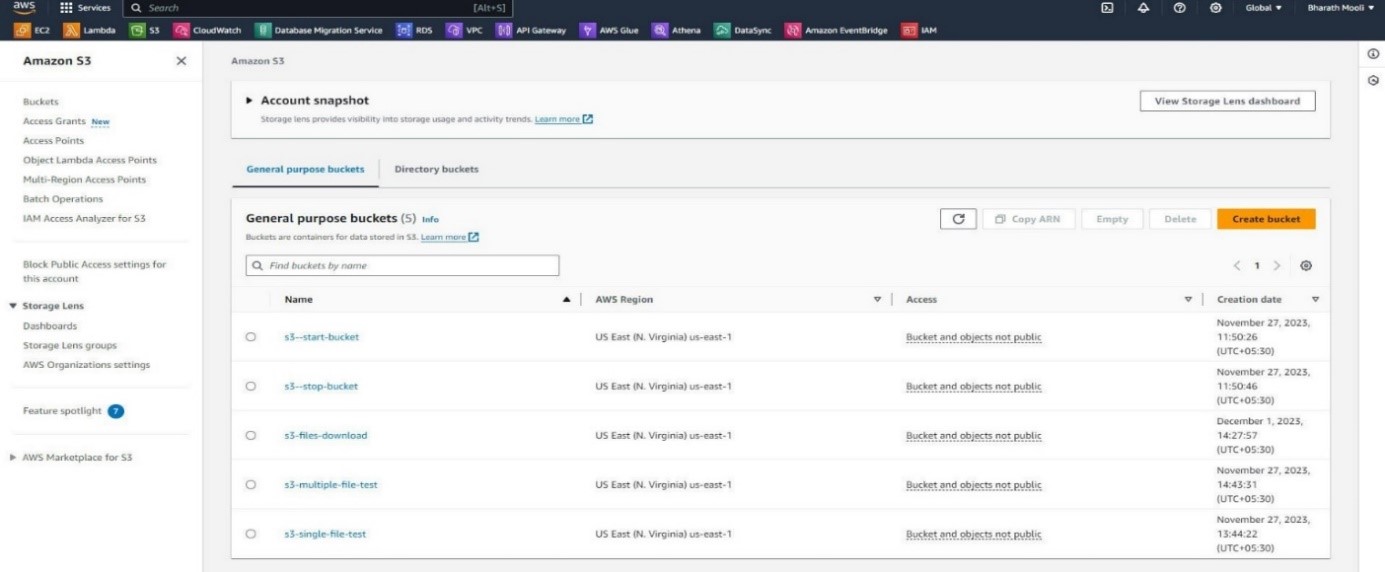


## **Implementation:**

## **S3 BUCKET CREATION**

1. **Creating a Bucket:**

* Click on the **Create bucket** button.
* **Enter a unique bucket name** (must be lowercase and between 3-63 characters long, without special characters or underscores).
* Choose a **region** where you want your bucket to reside. This will determine the physical location of your data.
* (Optional) Configure any additional settings like **versioning** or **server-side encryption**.
* Click **Create bucket**.

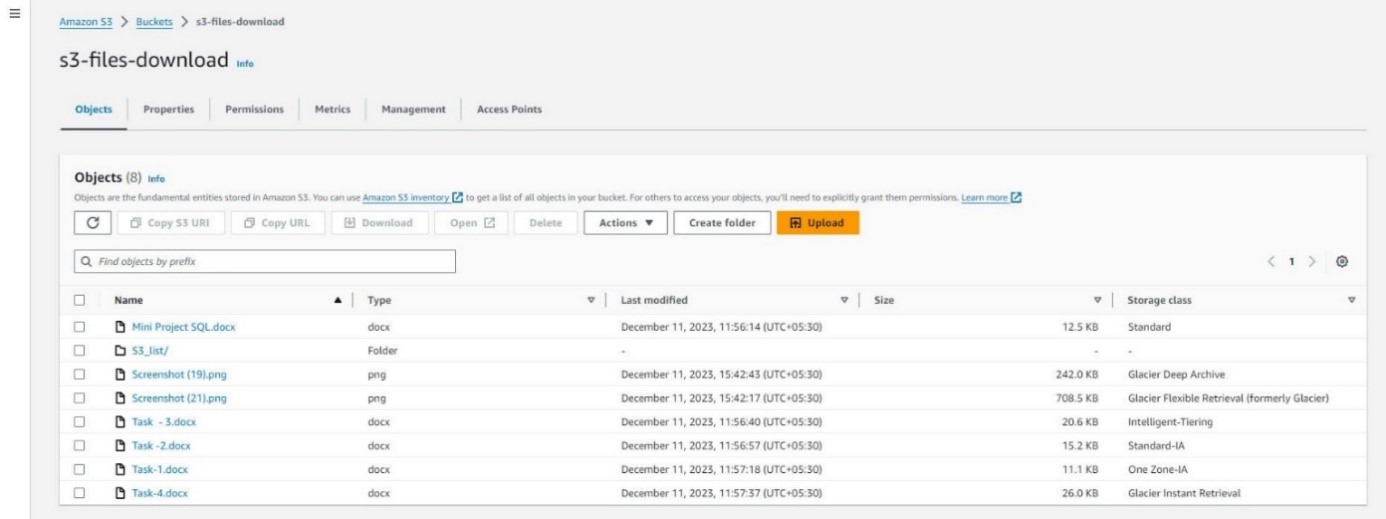


1. **Selecting the Bucket:**

* Once created, your new bucket will appear on the list.
* Click on the checkbox next to your newly created bucket to select it.

1. **Uploading Files:**

* Click on the **Upload** button.
* You can either **drag and drop files** from your computer or click **choose files** to select them manually.
* Click on **Upload**.

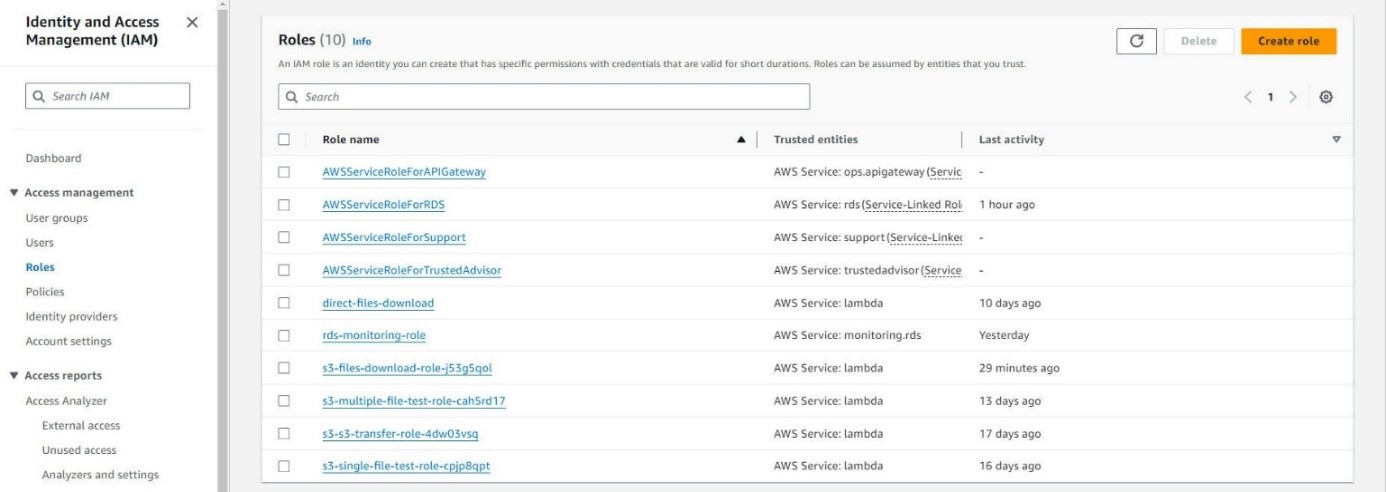


## **IAM ROLE CREATION**

**Creating the IAM Role:**

**1. Navigate to the IAM console:**

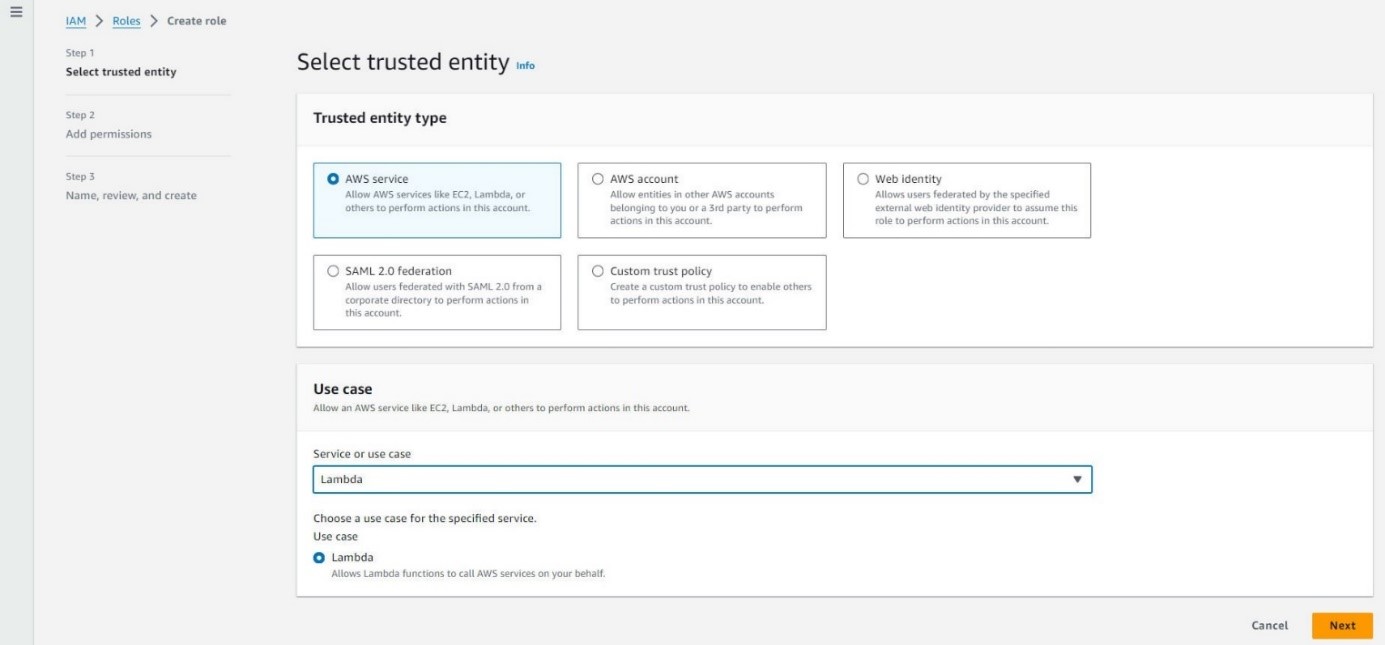
* Open the AWS Management Console and go to the Identity and Access Management (IAM) service.
* Role's tab: Click on the "Roles" tab in the navigation pane.



**2. Create role:** Click on the "Create role" button.

**3. Trusted entity type:** Select "AWS service" as the trusted entity type.

**4. Use case or service:** Choose "Lambda" from the service list.

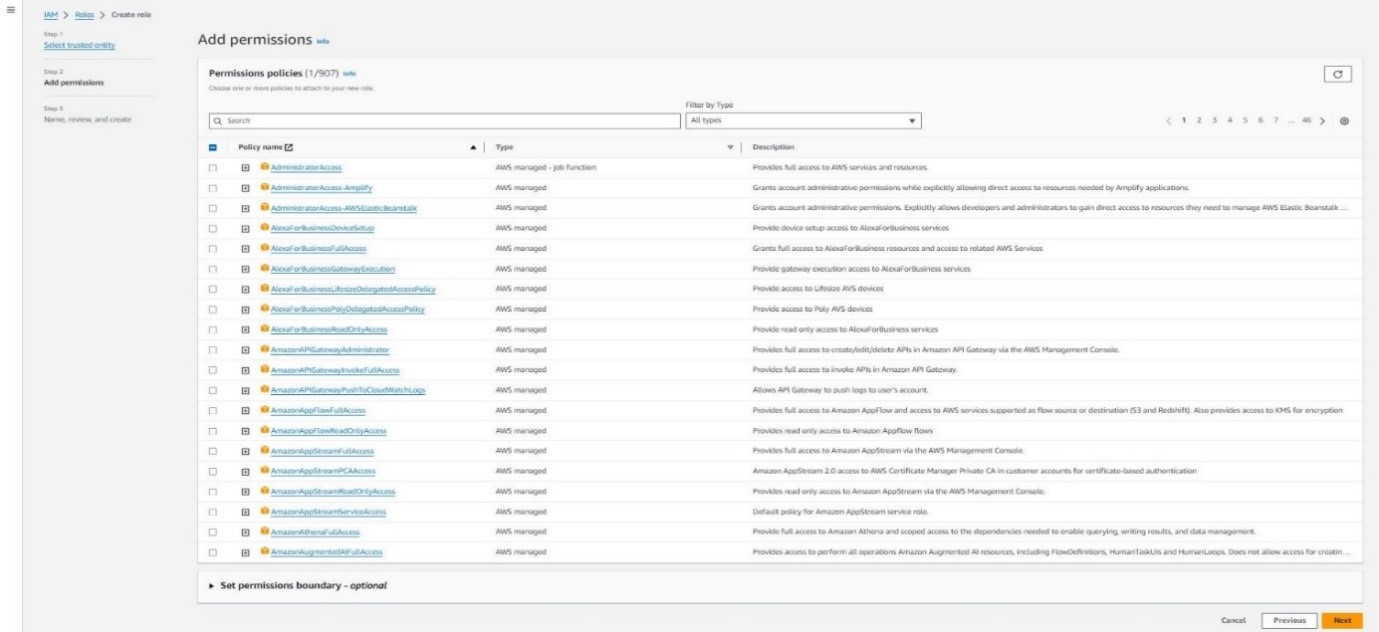


**5. Add permissions:** Click on the "Add permissions" tab.

**6. Choose permissions:** Select the following IAM policies:

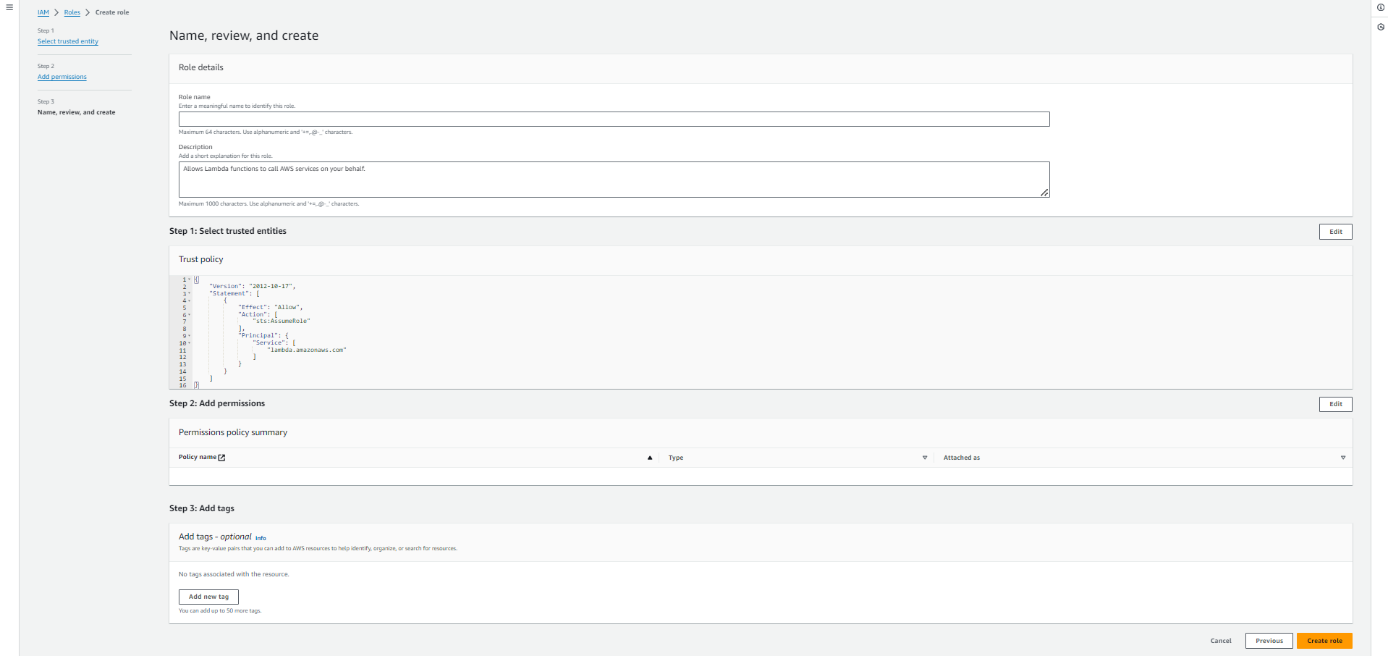
* **AWSLambdaBasicExecutionRole:** Allows the Lambda function to run code.
* **AmazonS3FullAccess:** Grants full access to read and write files in your S3 bucket.
* **AmazonAPIGatewayInvokeAccess:** Allows API Gateway to invoke your Lambda function.

**7. Next:** Click on "Next" to proceed.



**8. Role name and description:** Provide a name for your IAM role and an optional description.

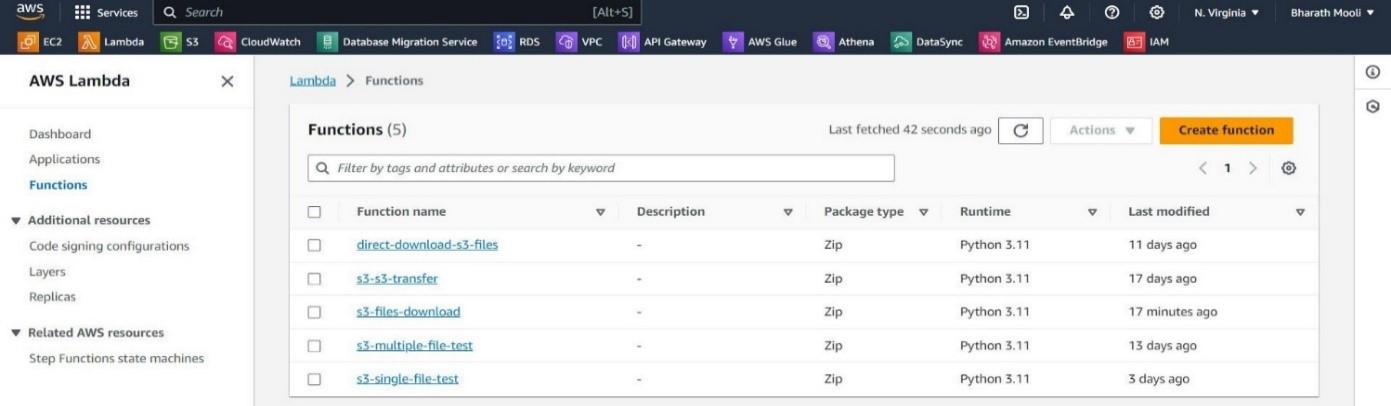
**9. Create role:** Click on "Create role" to create the IAM role.



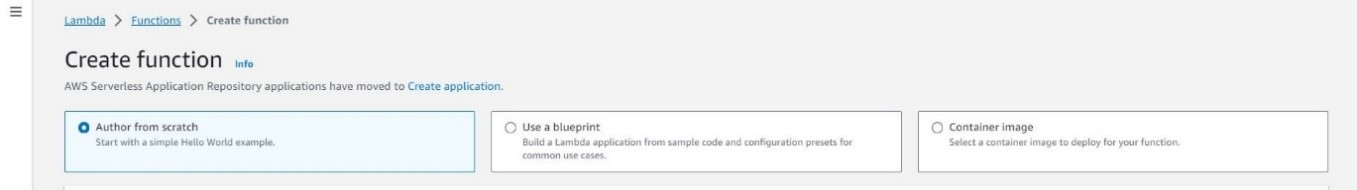
## **LAMBDA FUNCTION**

1. **Navigate to the Lambda service:**

* Open the AWS Management console and go to the Lambda service.

**2. Create a new function:**

* Click on the "Create function" button.
* Select "Author from scratch" as the creation method.



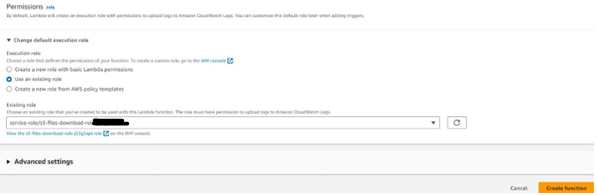
**3. Configure basic information:**

* **Function name:** Provide a descriptive name for your Lambda function.
* **Runtime:** Choose "Python 3.11" as the runtime environment for your function.
* **Architecture:** Select "x86\_64" as the architecture.



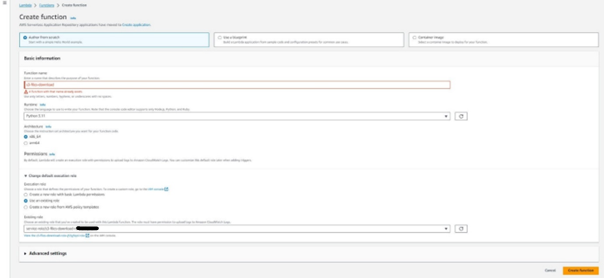
1. **Assign IAM role:**

* In the "Permissions" section, choose the IAM role you created earlier from the "Existing IAM roles" list.
* This role will grant your function the necessary permissions to access resources and perform its tasks.



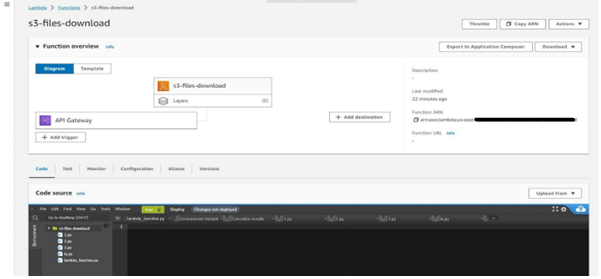
**5. Create the function:**

* Click on "Create function" to create the Lambda function based on your chosen settings.



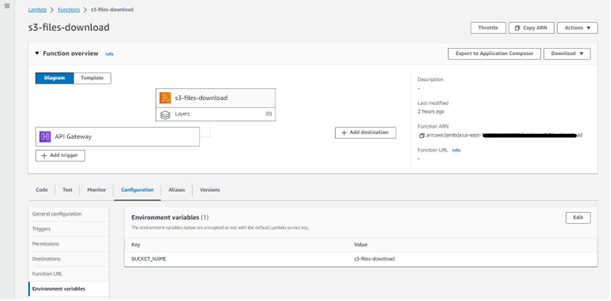
**6. Copy and deploy code:**

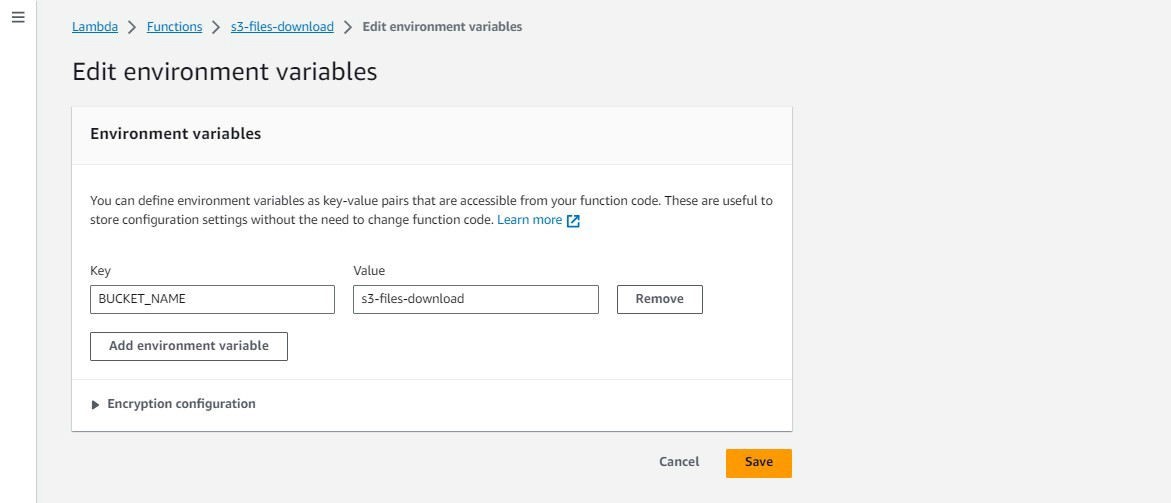
* Once the function is created, you will be redirected to the code editor.
* Copy and paste the Python code you want your function to execute into the editor.
* Click on "Deploy" to upload the code and deploy the function to AWS.



**Setting Environment Variables:**

1. **Navigate to the Lambda function:** Go to the Lambda service in the console and select your function.
2. **Configuration:** Click on the "Configuration" tab.
3. **Environment variables:** Click on the "Environment variables" section.
4. **Add variable:** Click on "Add environment variable".
5. **Name:** Enter "BUCKET\_NAME" as the name of the environment variable.
6. **Value:** Enter the name of your S3 bucket as the value of the environment variable.
7. **Save:** Click on "Save" to update the environment variables.

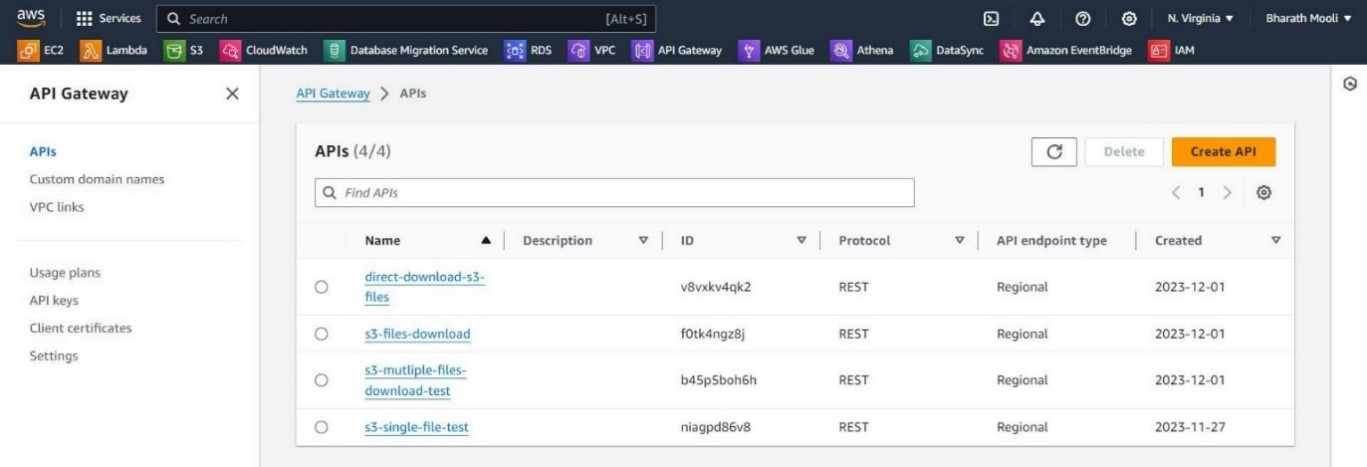




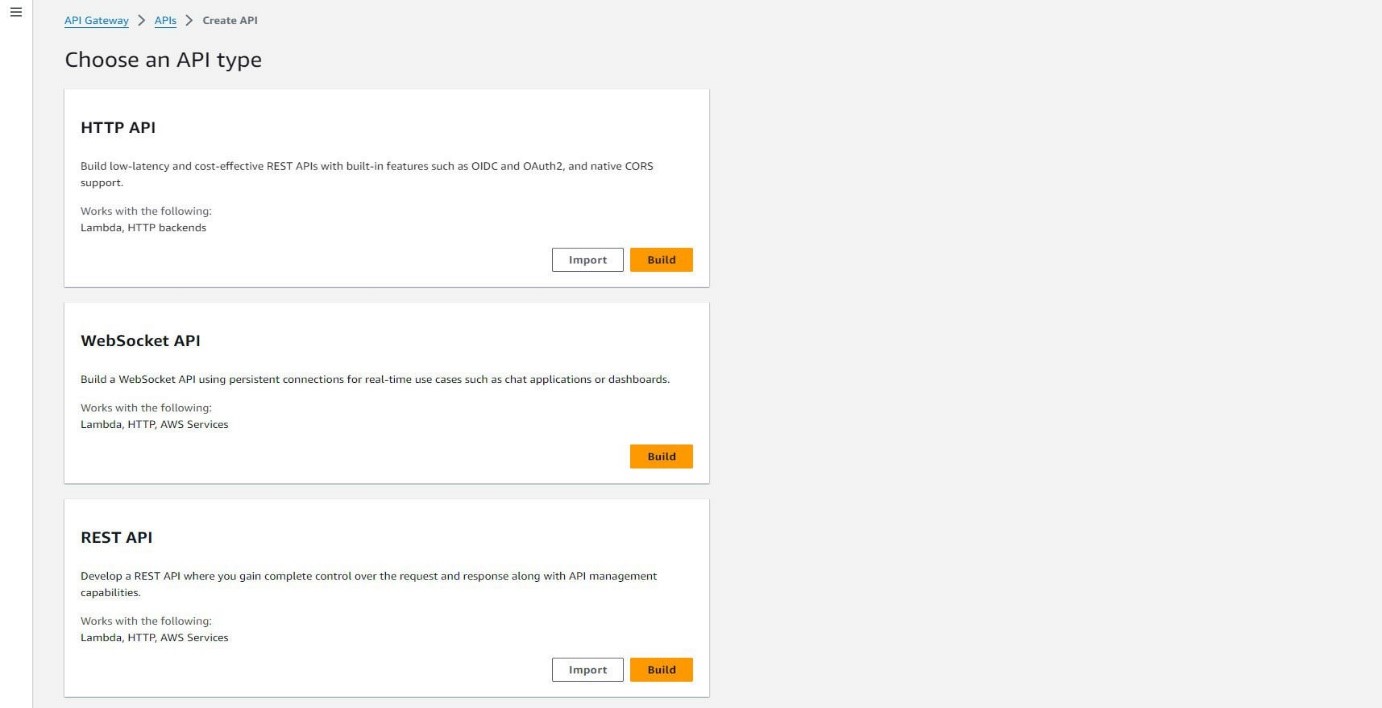
## **API GATEWAY**

1. **Creating the API:**

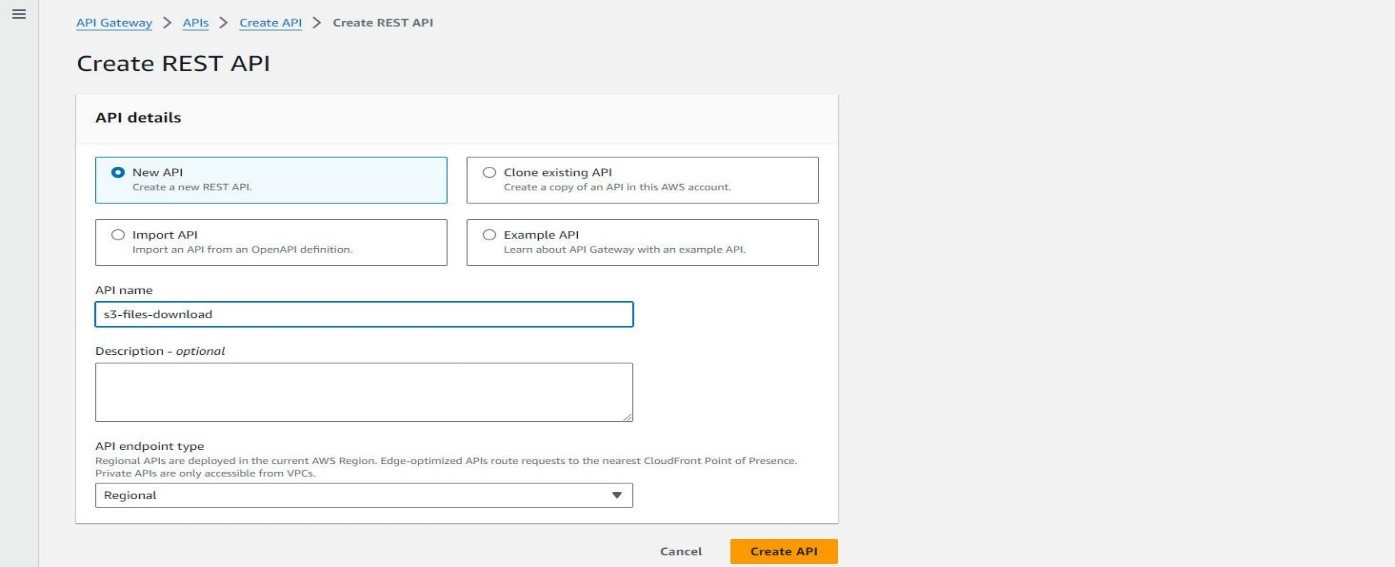
* **Navigate to API Gateway:** Open the AWS Management Console and navigate to the API Gateway service.

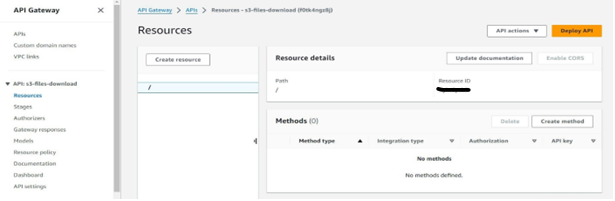


* **Choose REST API and Build:** Select the "Create API" button and choose "REST API" as the type.
* Click "Build" to proceed.



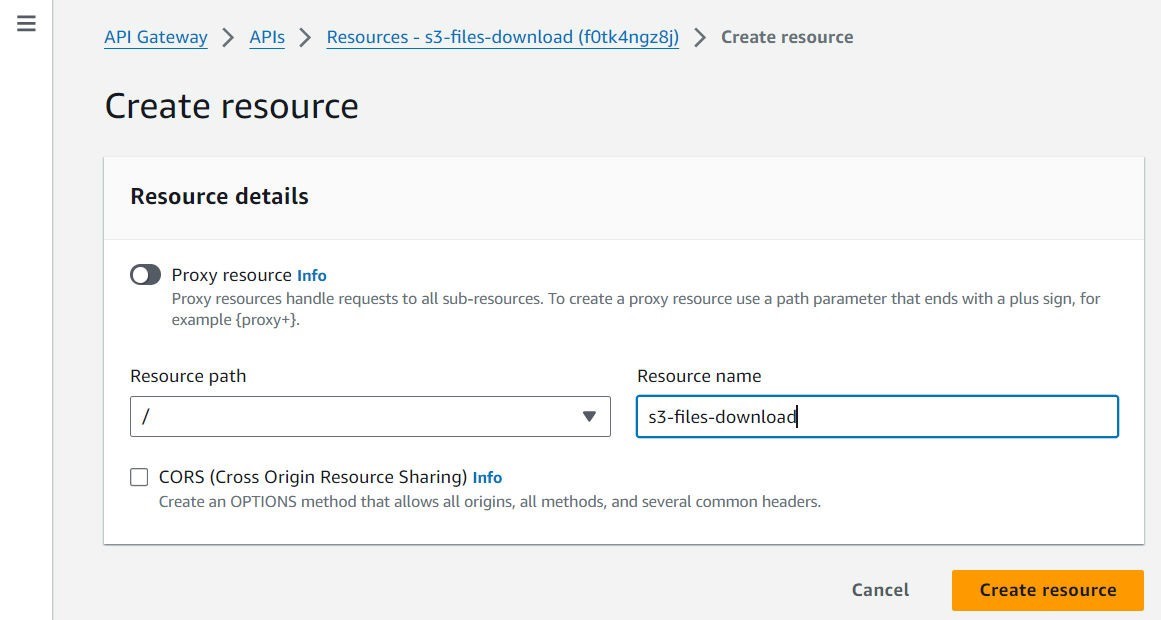
* **New API and Name:** Enter a descriptive name for your API. Choose "REGIONAL" for the API endpoint type, indicating it will be accessible from any region within your AWS account.
* Click "Create API" to finalize.



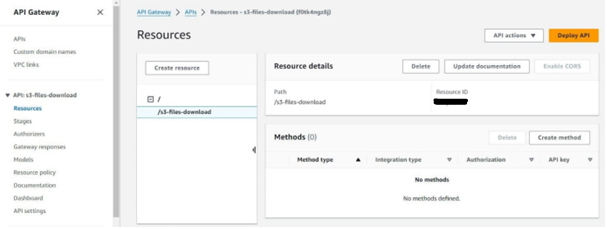


1. **Adding a Resource and Method:**

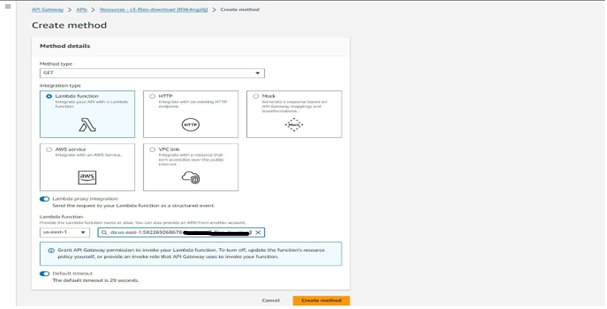
* **Create Resource:** Click the "Create Resource" button. Enter a path name for your resource, which will be part of the API URL.



* **Create Method:** Click on the “Create Method” next to the newly created resource.

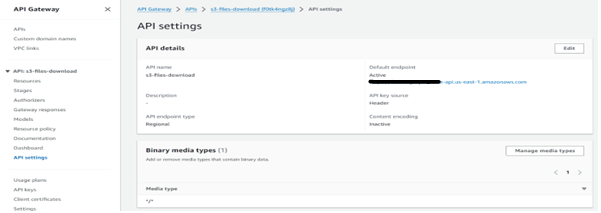


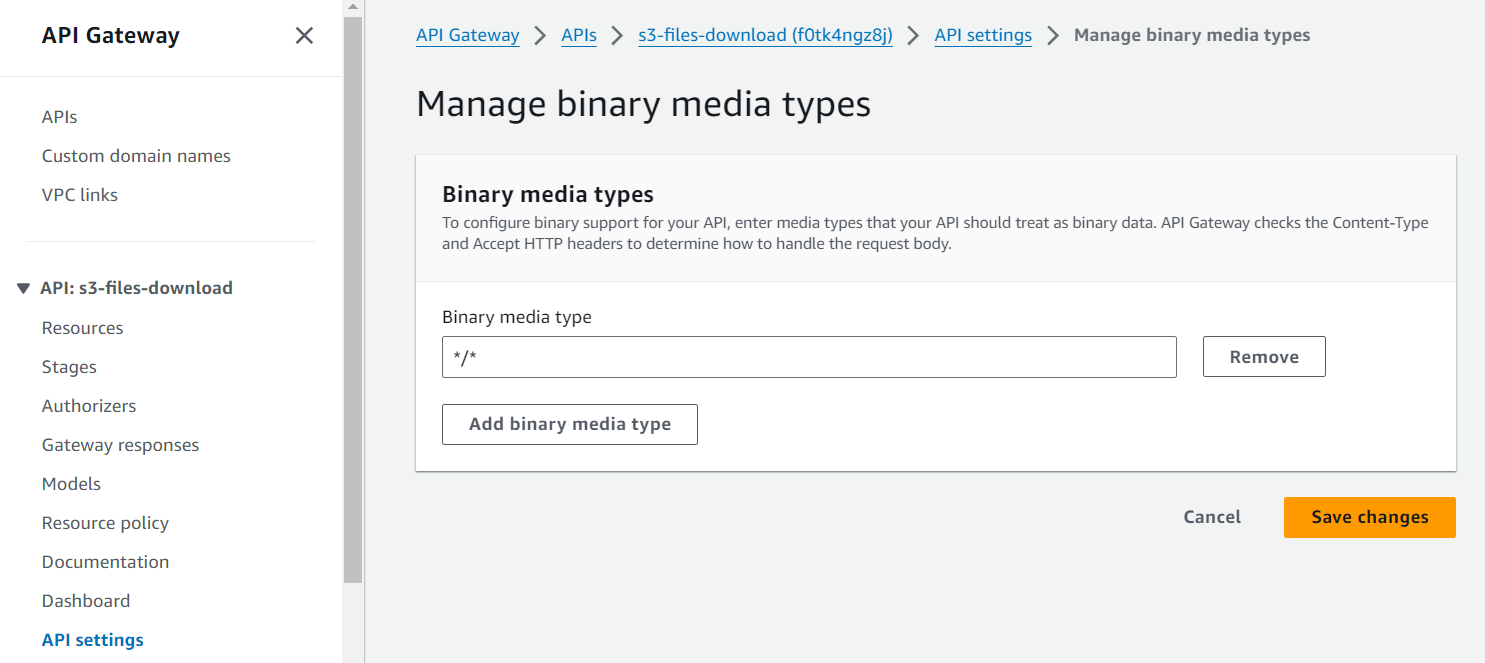
* Select "Method Type" as "GET" method, as it will be used to access your binary content.
* **Lambda Integration:** Select "Lambda Function" as the integration type. This means your API method will trigger a Lambda function to process the request.
* **Lambda Proxy Integration:** Choose "Lambda Proxy integration" for optimal handling of requests and responses, especially with binary data.
* **Select Lambda Function:** Select the Lambda function where your code resides. This code will handle the logic for accessing and processing the binary media.
* **Create Method:** Click "Create Method" to finalize the configuration of your API method.



1. **Enabling Binary Media Types:**

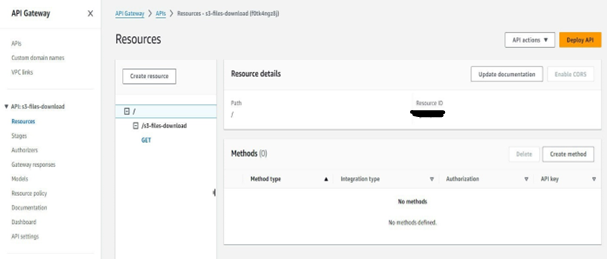
* **API Settings:** Go to the "Settings" tab of your API.



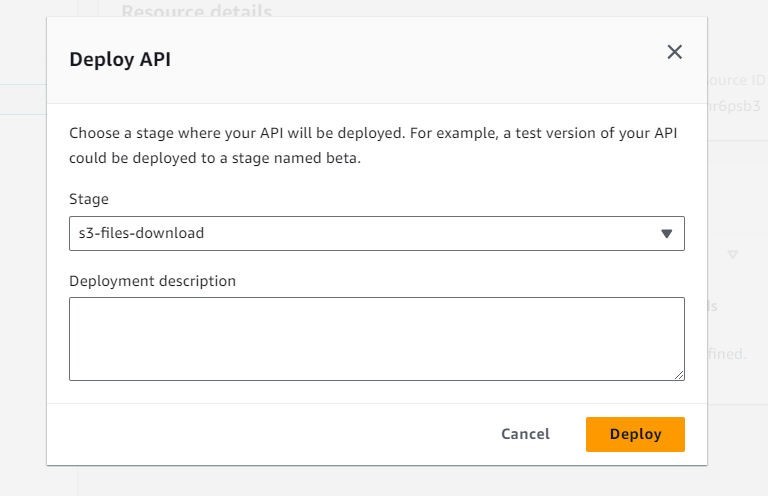


1. **Deploying the API:**

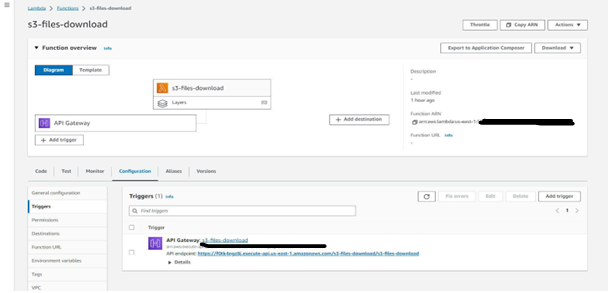
* **Resource Tab:** Go back to the "Resources" tab of your API.



* **Deploy API:** Click the "Actions" menu for your resource and select "Deploy API".
* **Stage and Stage Name:** Choose a stage for deployment. Enter a descriptive name for the stage.
* **API URL Generation:** Click "Deploy". This will create a deployment of your API and generate a URL for accessing it.



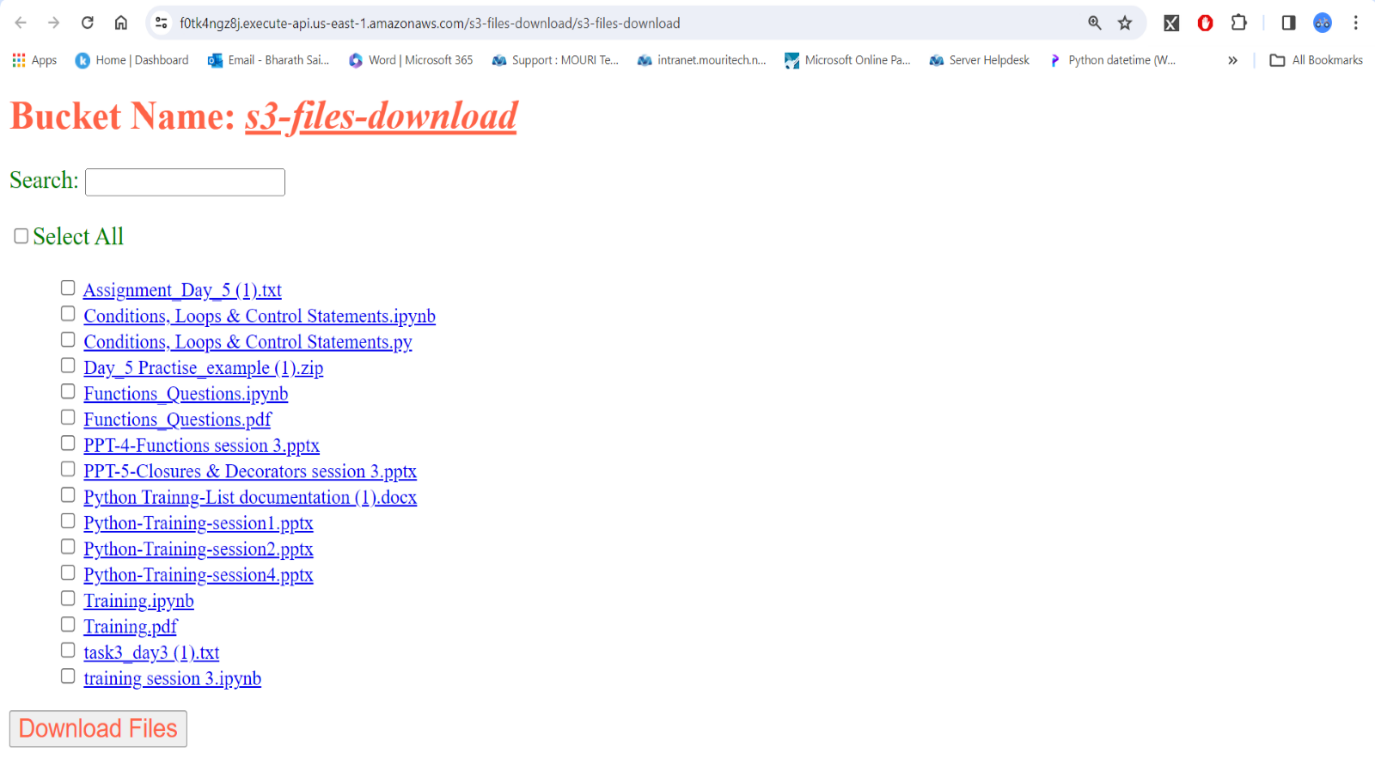
* **Navigate to the Lambda function:** Go to the Lambda service in the console and select your function.
* **Configuration:** Click on the "Configuration" tab.
* Click on the "Triggers" section.
* Copy the API endpoint.



**Instead of manually navigating the S3 console, the API offers a streamlined directory view of all files in the bucket. Say goodbye to tedious file searches and slow, one-by-one downloads! Our downloader lets you instantly search for the specific files you need and download them all in bulk, a time-saving feature not available in the standard AWS S3 console. Spend less time navigating menus and more time focusing on what matters.**

## **Output**:

**Link:** [**https://f0tk4ngz8j.execute-api.us-east-1.amazonaws.com/s3-files-download/s3-files-download**](https://f0tk4ngz8j.execute-api.us-east-1.amazonaws.com/s3-files-download/s3-files-download)



## **Advantages:**

* **Cost-efficiency:** Serverless architecture eliminates costs associated with idle servers, as you only pay for Lambda execution time.
* **Scalability:** Handles traffic spikes automatically, without manual intervention.
* **Simplified management:** AWS manages infrastructure, freeing you to focus on application logic.
* **Security:** API Gateway and Lambda offer built-in security features, protecting sensitive files.
* **Flexibility:** Integrates with other AWS services for enhanced functionality.
* **Faster deployments:** Rapidly deploy updates without server provisioning delays.

## **Disadvantages:**

* **Cold start latency:** Lambda functions might experience delays for infrequent downloads.
* **Limited execution time:** Default 15-minute limit for Lambda functions might not suffice for exceptionally large files.
* **Vendor lock-in:** Reliance on AWS services can make migration to other platforms challenging.
* **Cost considerations for high download volumes:** Pay-per-execution model can become expensive for frequent, high-volume downloads.

## **Conclusion:**

A serverless S3 file downloader using API Gateway and Lambda offers a cost effective, scalable, and minimal maintenance solution. While there are some limitations, careful design and consideration of these constraints lead to a robust and efficient system. This technology stack leverages AWS services to provide a fully managed and scalable architecture for file downloading.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*THE END \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**Understand**

Survey #64, 4th Floor SBR Surya Pearl, Sector III Madhapur, Hitech City,  
Hyderabad-500081, Telangana, India

**Web & Email**

www.mouritech.com  
info@mouritech.com

**Tel & Fax**

Tel: +91 406 725 4100