Project overview:

Need to build an API which can load the data to s3 with storing the metadata into DynamoDB. Once file loaded into s3 need to run the glue crawler which user can able to query on the Athena table. All those should be deployed using AWS SAM.

Prerequisite:

**Step 1: Create DynamoDB table**

Create a DynamoDB table to store the metadata with items as rolename as a partition key and requesttime as sort key.

**Step 2: Create Lambda function**

Create a lambda function which can store the meta data information such as role name and request time into DynamoDB table and will generate an pre-signed URL to upload files into s3. After that, we need to give below permissions associated with this lambda function.

* S3 full access.
* Glue full access.
* API gateway full access.
* Cloud watch access.

**Step 3: Create HTTP API Gateway**

Create an HTTP gate way in the AWS API gateway and add the above Lambda function in the integration part.

**Step 4: Create s3 bucket.**

Create an s3 bucket.

**Step 5: Create glue Crawler.**

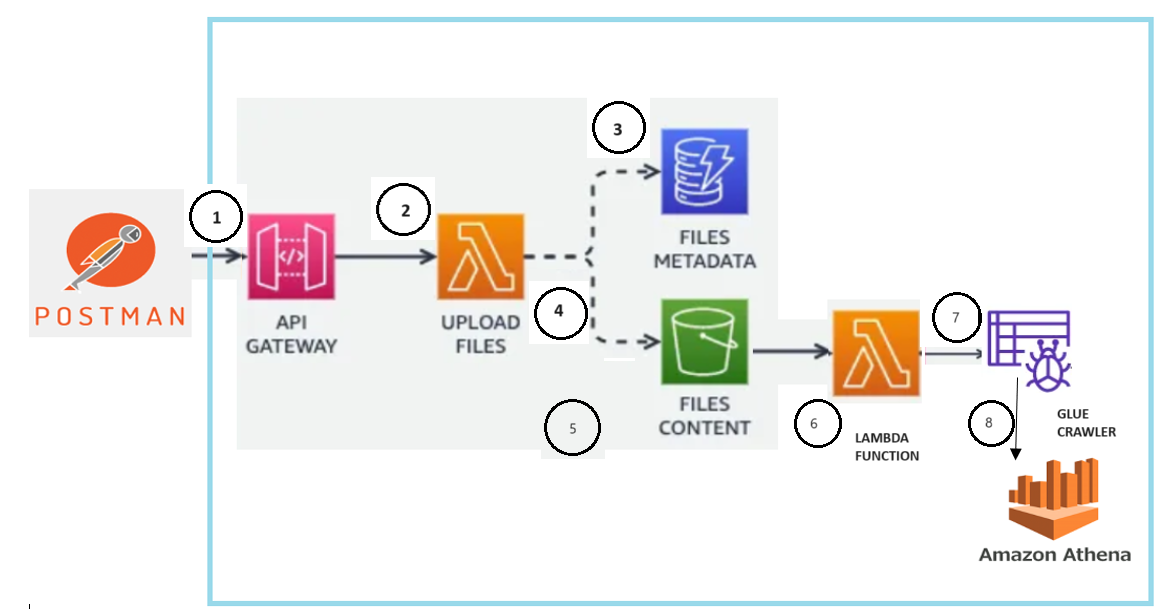
Create glue Crawler with data source where the API load the file into the S3.

**Step 6: Create lambda function**

Create a lambda function which can run the glue Crawler using boto3 and add the s3 path as event for the lambda.

**Step 7**: Need to install the postman for testing the application.

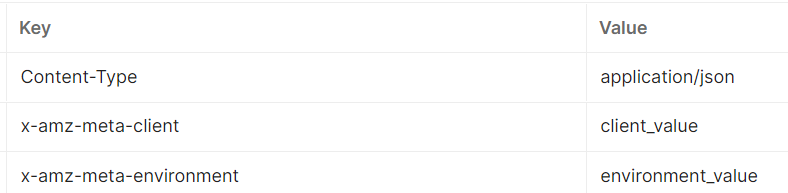
Architecture:





Flow of execution:

1. Post the end point URL in the postman with parameter of rolename. Then it will call the API gateway.
2. Once API gateway called, lambda will be triggered and then lambda function code will be executed.
3. Once lambda code starts execution, metadata (role name given in the postman and request time) will be captured into DyanmoDB.
4. Once metadata captured successfully to DynamoDB, pre-signed url will be returned to the postman.
5. Copy the returned URL and add the PUT method for the URL and upload the Json file in the binary. And also need to add below key values in the header.



1. Once file uploaded into the s3 a lambda will be triggered and lambda function code will be executed.
2. While executing the lambda function code glue crawler will be started and fetch the metadata of a file into glue data catalog table.
3. Now we can query the glue data catalog table using athena.

Deployment :

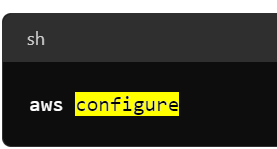
Prerequisite:

1. Need to install AWS CLI.
2. Need to install AWS SAM.
3. Need to install python.

**YOUR TITLE HERE**

Steps to deploy:

1. Once AWS CLI is installed, need to configure your AWS account with below command.



1. Clone the project from the git and update template.yaml file with proper iam roles and s3 bucket name.
2. We can start the deployment using aws sam using the command (sam build –guided).

Note: Need to give proper IAM role permissions for the CLI-user.