PGPCC

Project Implementation

Building an Automated Business Process using Managed Services on a Public Cloud

Phase1 – Architecture

--Mahesh Jasti

Contents

Phase-1 Architecture

Phase-2 Implementation

Section 1: Objective of the Project

The objective of this project is to create an automated, event based real time process that does not have these limitations. Data should flow rapidly from the source to the destination.

Section 2: Scope of the Project

Create a solution architecture diagram based on the below points..

1.The customer uploads the invoice data to S3 bucket in a text format as per their guidelines and policies. This bucket will have a policy to auto delete any content that is more than 1 day old (24 hours).

2.An event will trigger in the bucket that will place a message in SNS topic.

3.A custom program running in EC2 will subscribe to the SNS topic and get the message placed by S3 event.

4.The program will use S3 API to read from the bucket, parse the content of the file and create a CSV record along with saving the original record in DynamoDB.

5.The program will use S3 API to write CSV record to destination S3 bucket as new S3 object.

6.Athena is used to query the CSV file (query to show aggregated expenses grouped by date).

Section 3: Implementation Architecture

The following picture outlines the implementation architecture for the project.
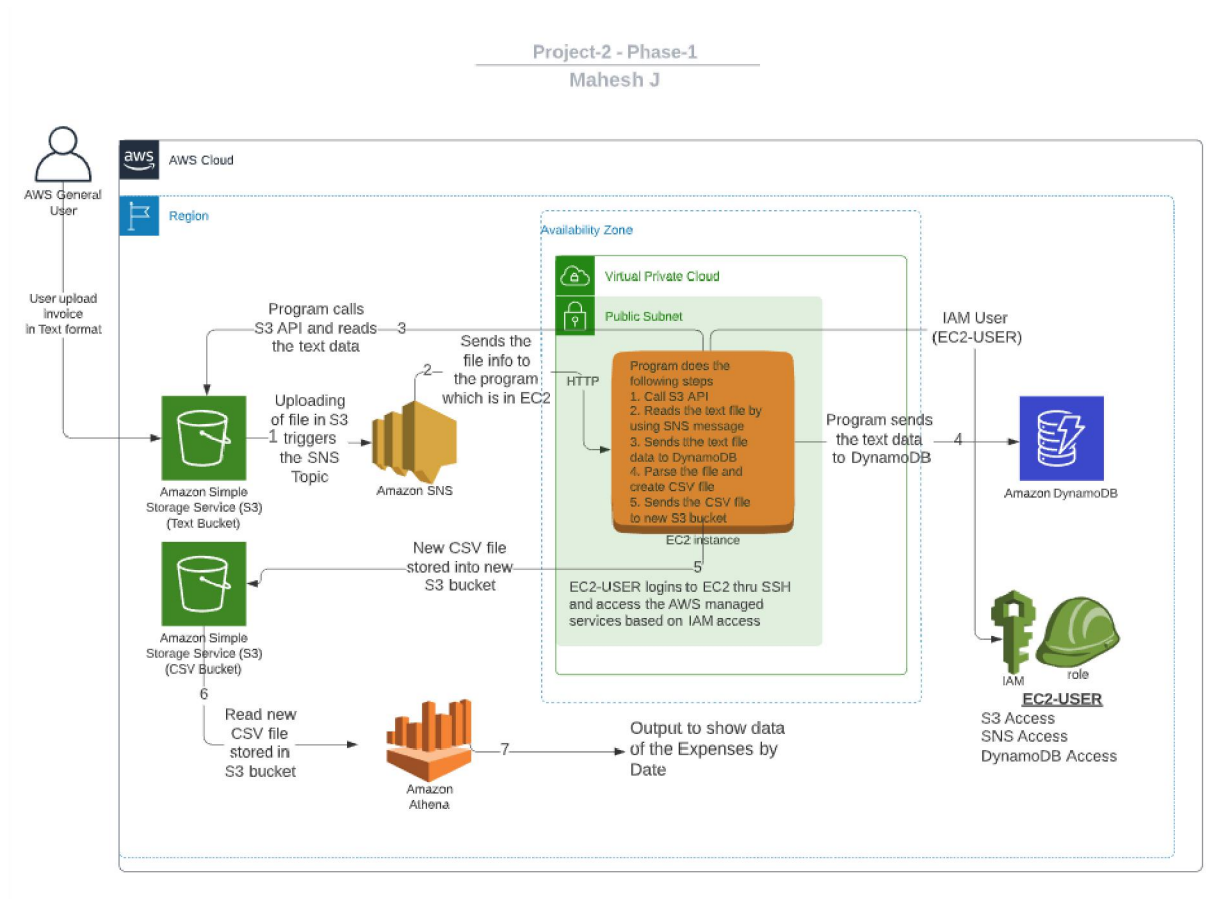
I have put the numbers from 1 to 7 to implement the flow. Below is the explanation for each step.

1. When user uploads the text file in S3 then it triggers the SNS topic.
2. SNS sends the notification to EC2 program.
3. EC2 program checks for the IAM access and then program calls the S3 api.
4. EC2 Program reads the data from text file and loads into dynamo DB.
5. EC2 program parses the data and creates the CSV file and store into new S3 bucket.
6. Athena reads the CSV file from the new S3 bucket.
7. Run the SQL query to show the result.