PGPCC

Project Implementation

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

Phase 2– Implementation

--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.

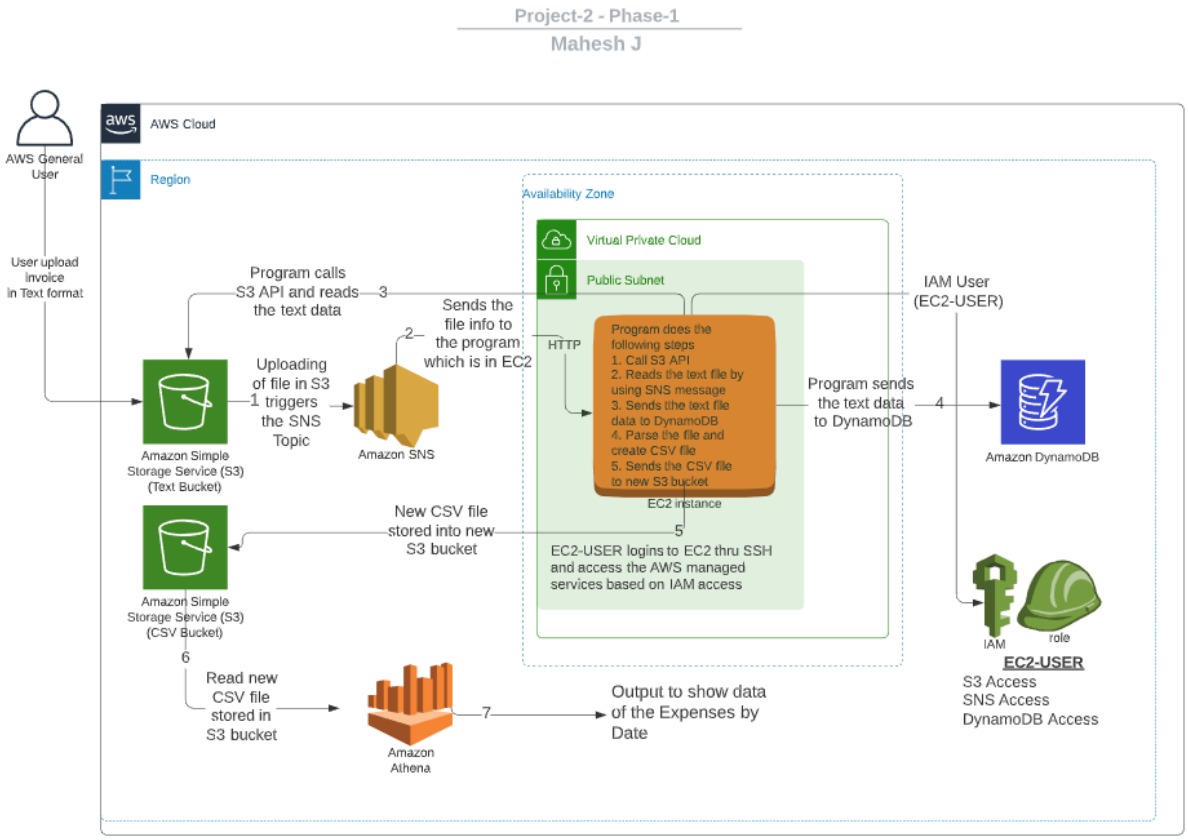
4.The program will use S3 API to read from the bucket, parse the content of the file and create a CSV record along with saving the original record in DynamoDB.

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.

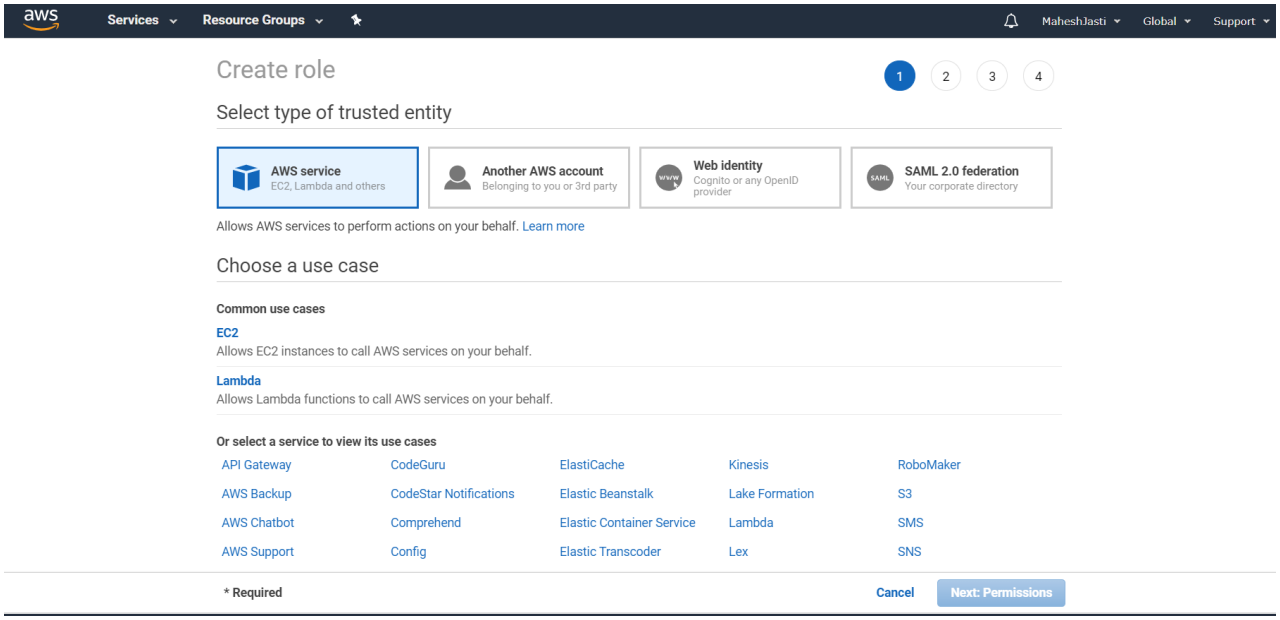
3. EC2 program checks for the IAM access and then program calls the S3 api.
4. EC2 Program reads the data from text file and loads into dynamo DB.
5. EC2 program parses the data and creates the CSV file and store into new S3 bucket.
6. Athena reads the CSV file from the new S3 bucket.
7. Run the SQL query to show the result.

Section 4: High level steps to implement the project

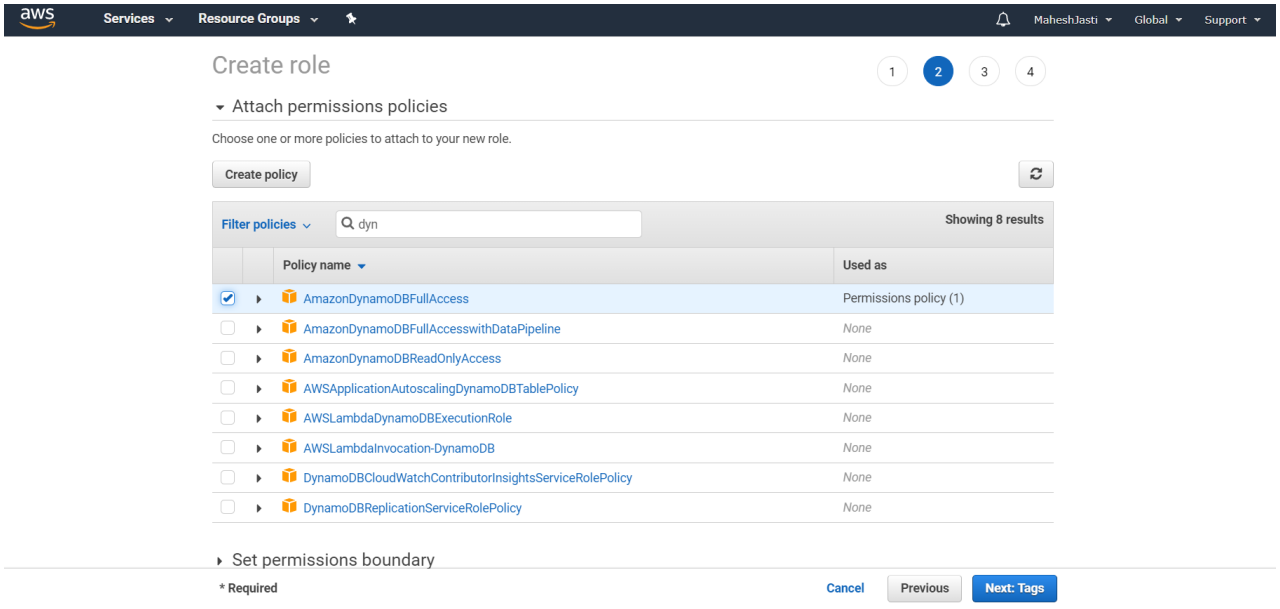
1. Create the IAM role for EC2 Instance (S3 and DynamoDB)
2. Create the VPC and Subnet and initiate an EC2 instance
3. Login to the instance and setup the softwares and program as per the given instructions
4. Setup S3 and SNS event trigger
5. Verify DynamoDB after input file is loaded into S3
6. Setup the Athena by using S3 target CSV file

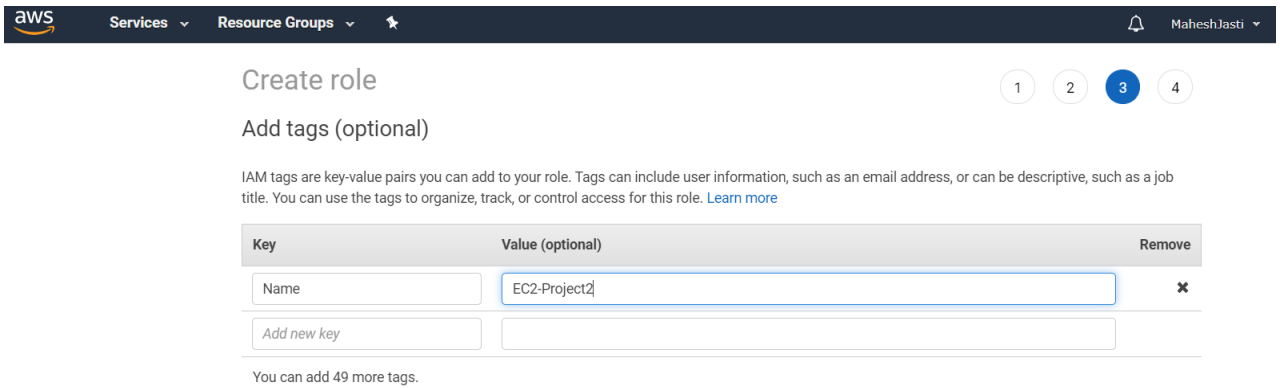
Section 5: Screenshots from AWS console

IAM Role: Create a role in IAM to have EC2 access the S3 and Dynamo DB.

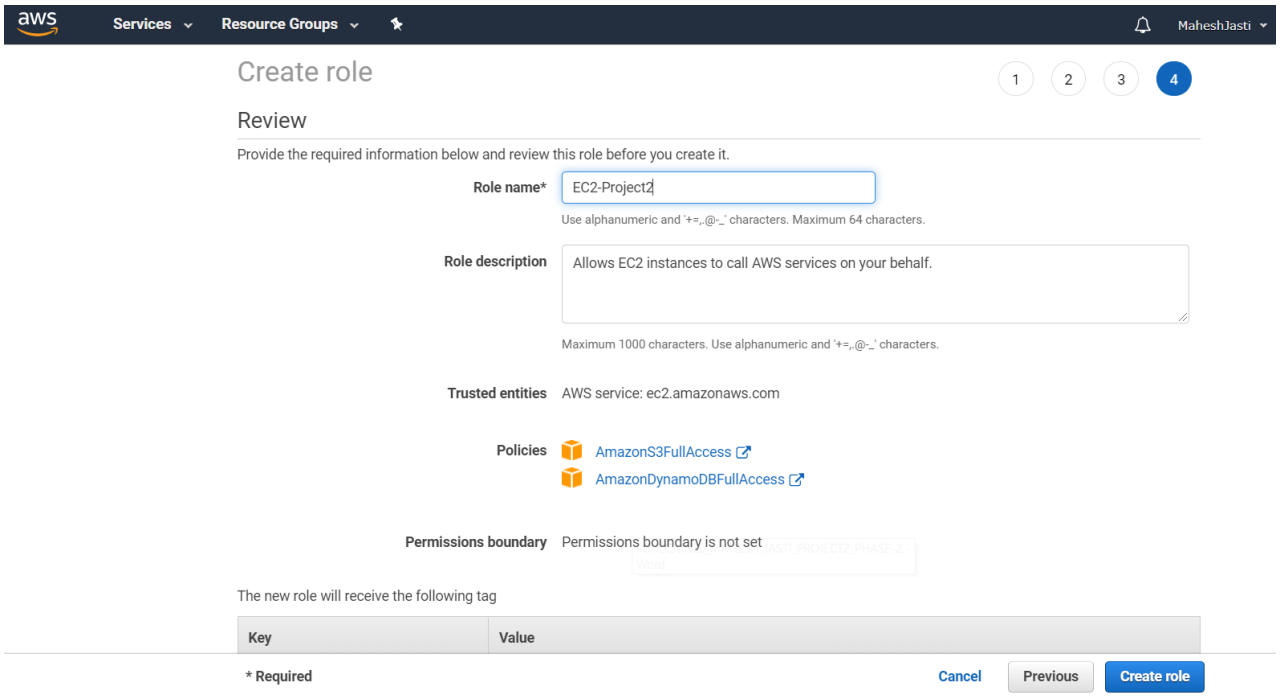


Select S3 full access and DynamoDB full access..

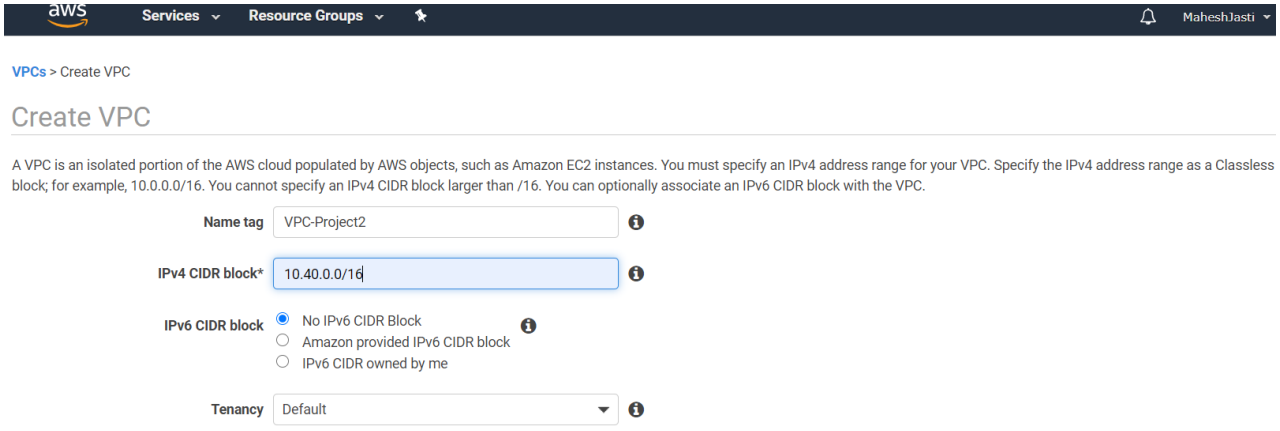


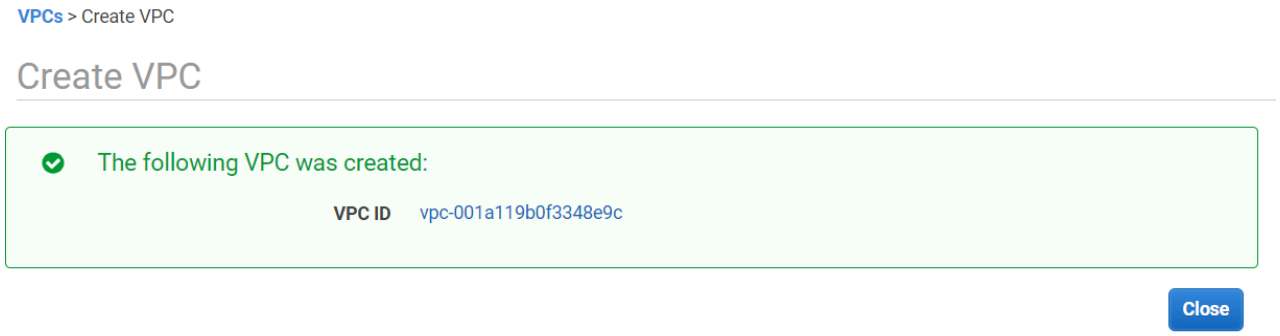


Enter the tag name and Role name and click on Create Role button.

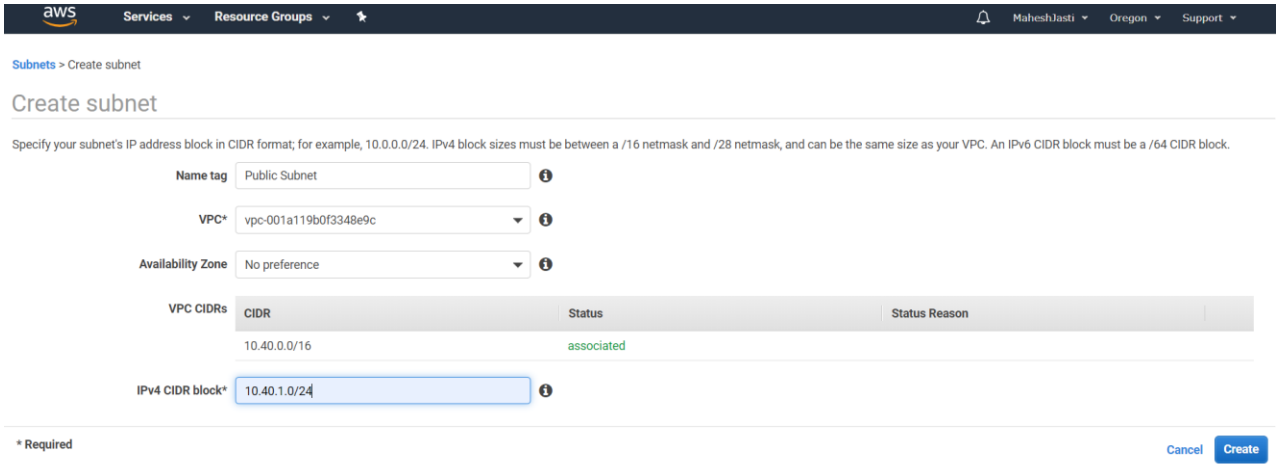


Create a VPC

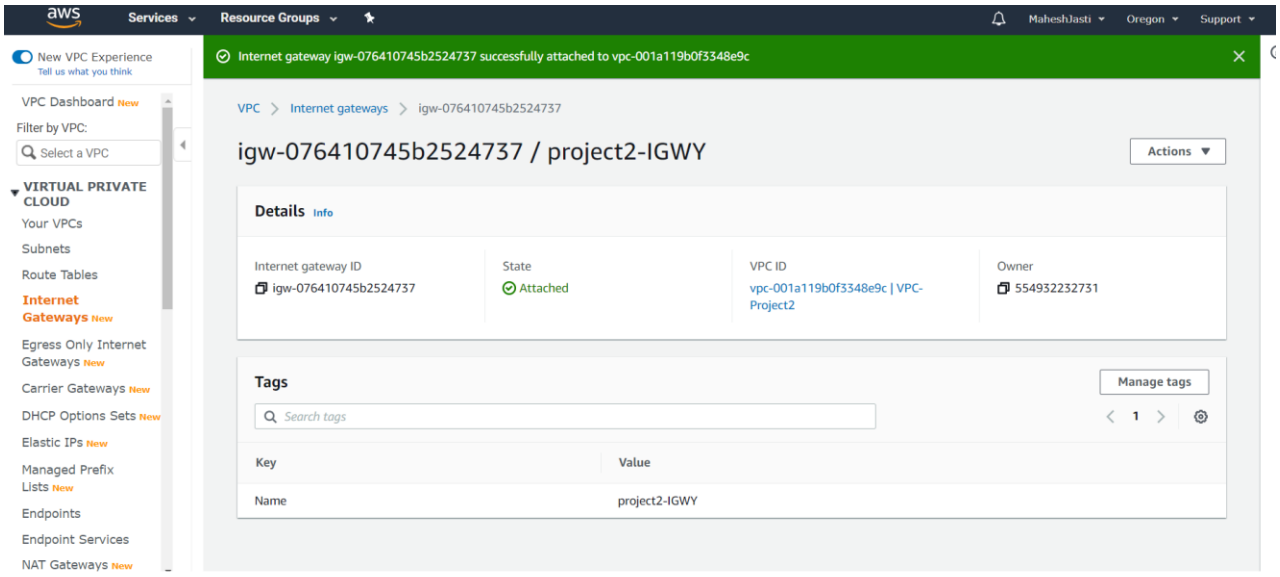




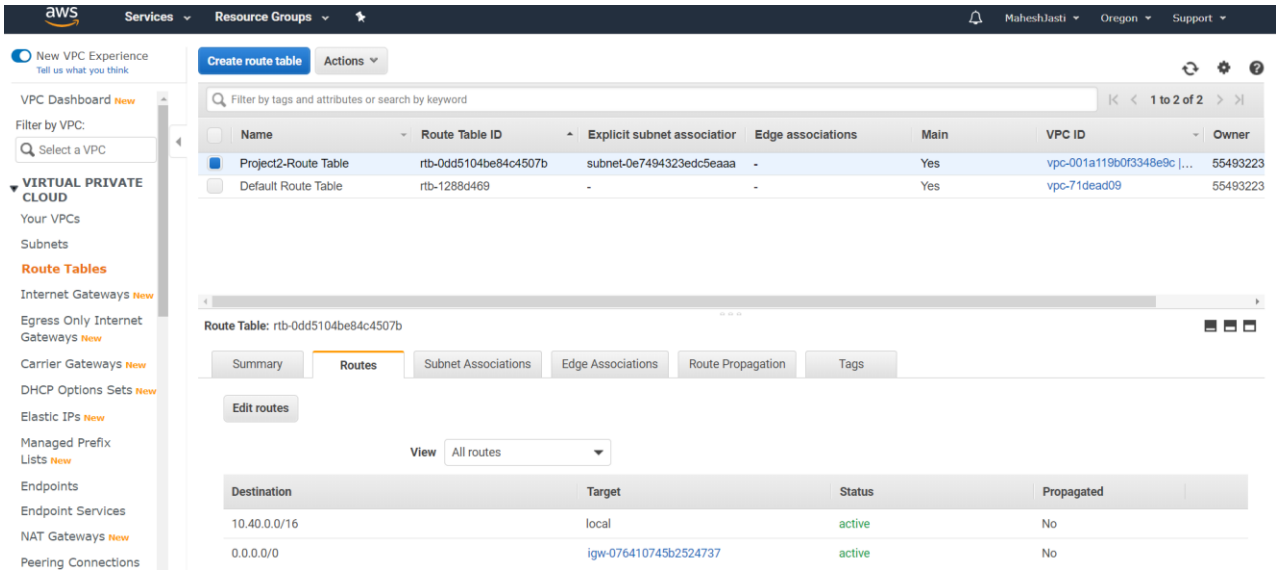
Now create subnets.

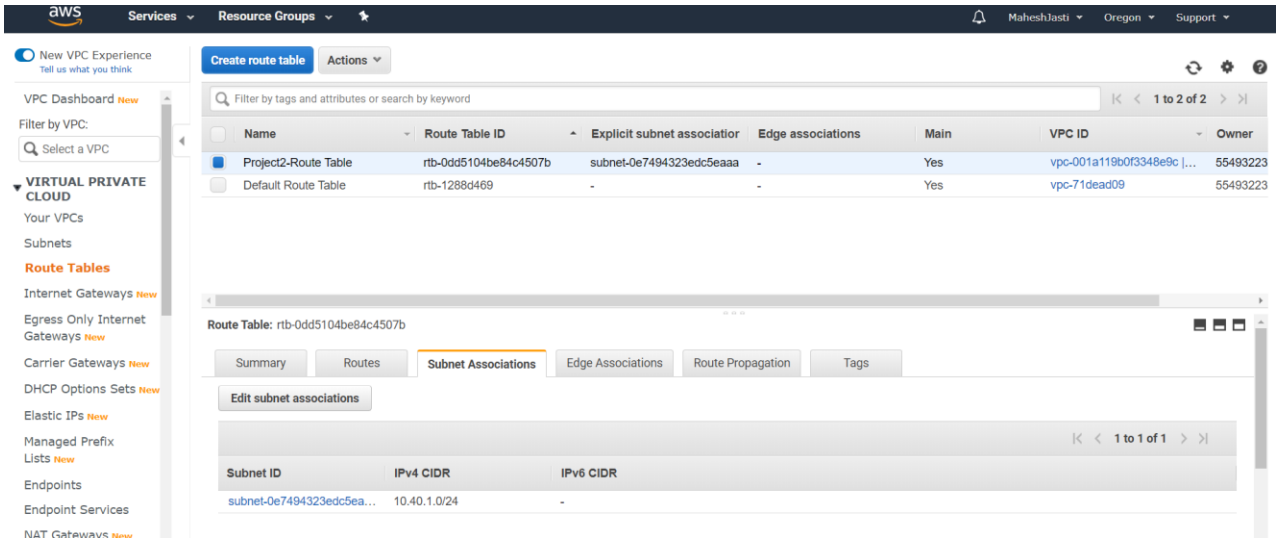


Create Internet Gateway and assign to the new VPC.

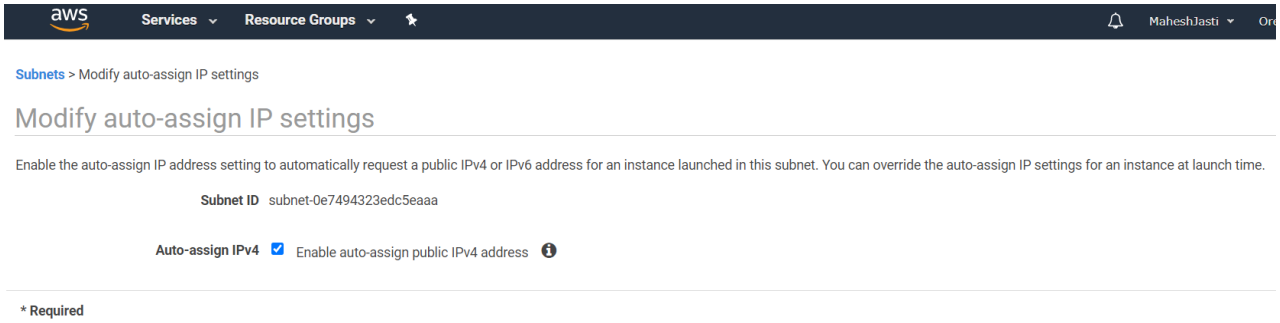


Edit the route table and add the Internet Gateway and also associate the Subnet..

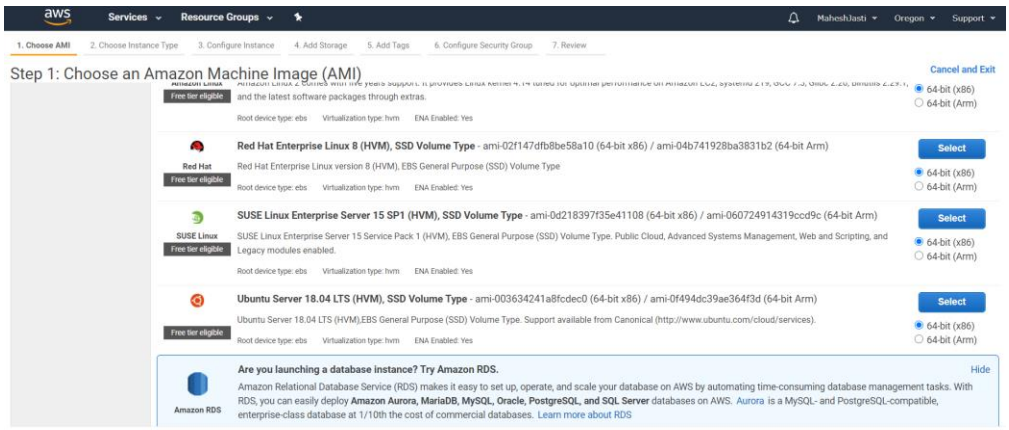




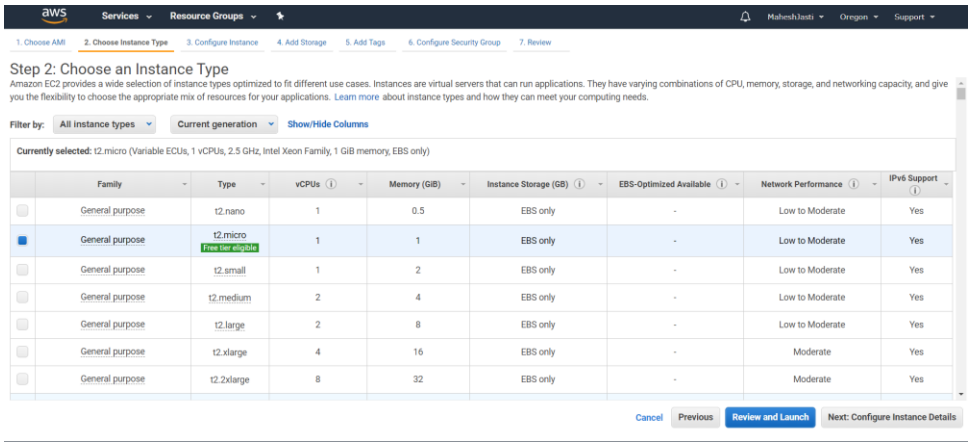
Enable auto-assign public IPV4 address



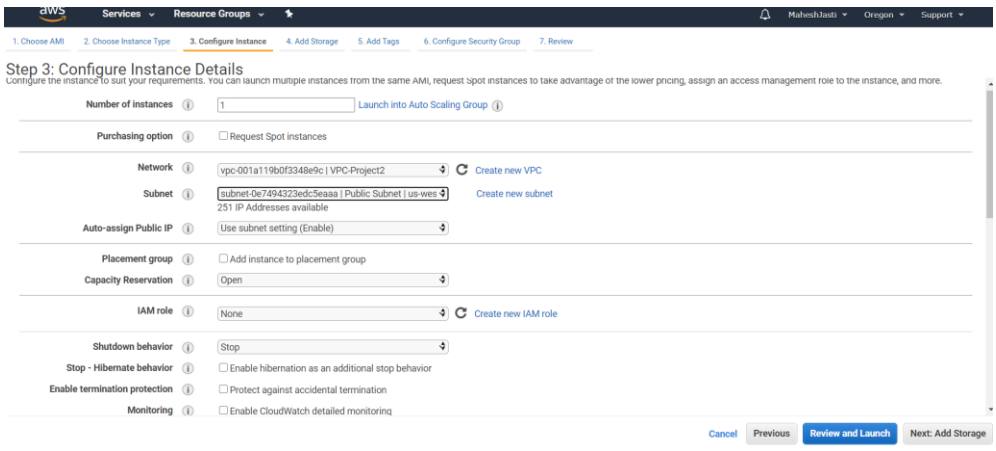
Now create the EC2 instance… (Ubuntu Instance)..



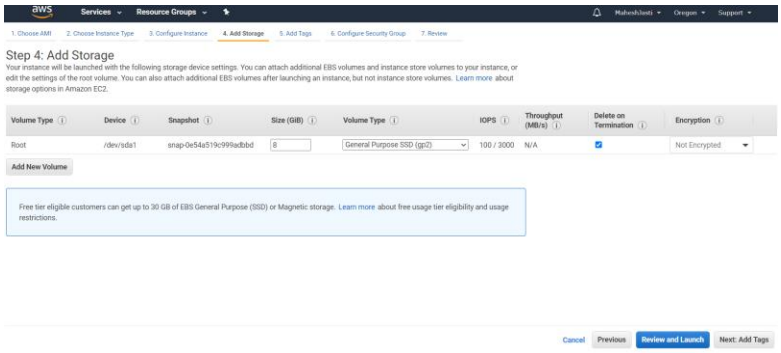
Select t2 micro instance and click “Configure Instance details” button..



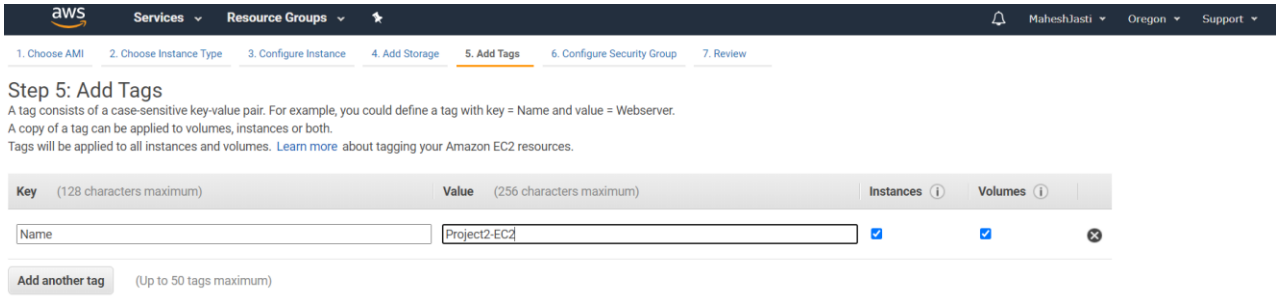
Select the VPC and Subnet and click on “Add Storage” button.



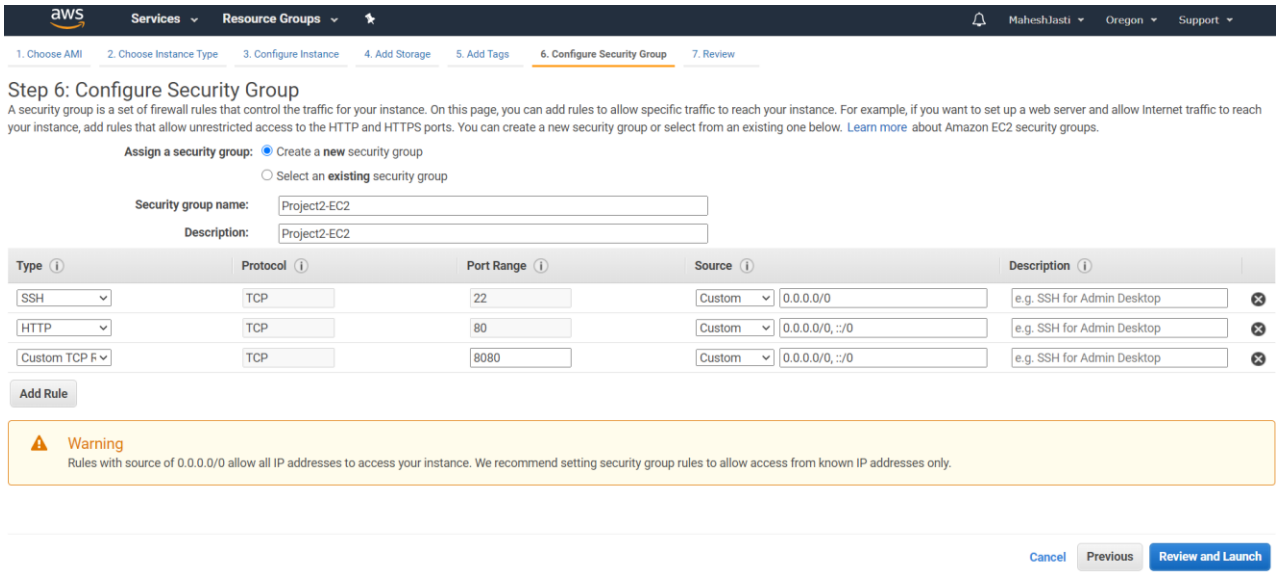
No changes to Add Storage screen.. Move to Tags screen..



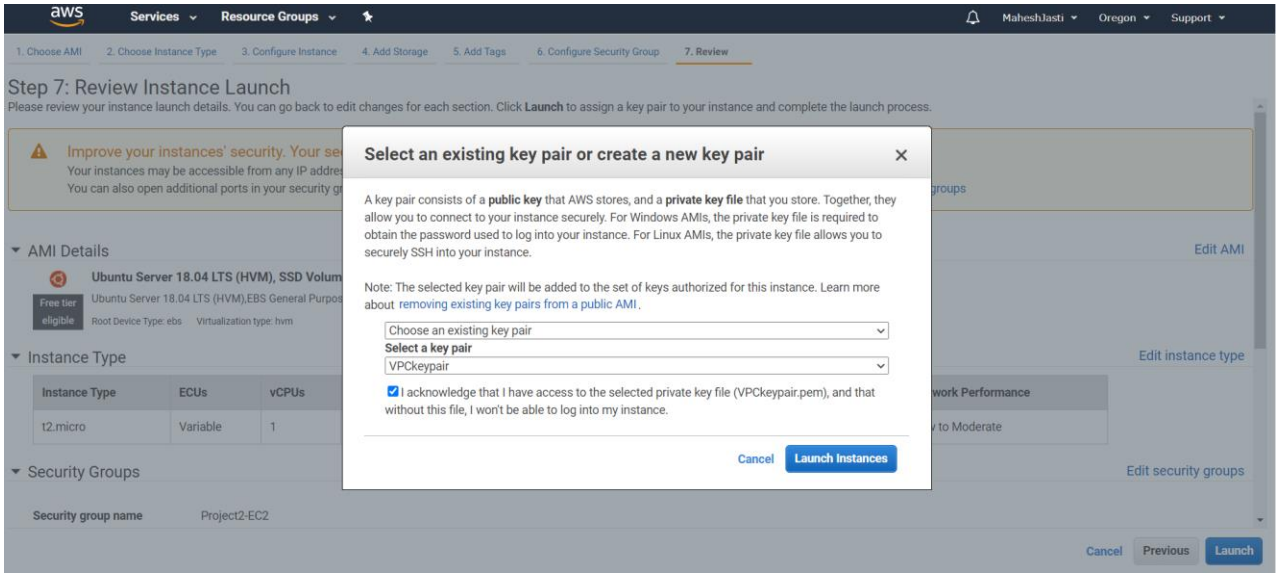
Add tag..



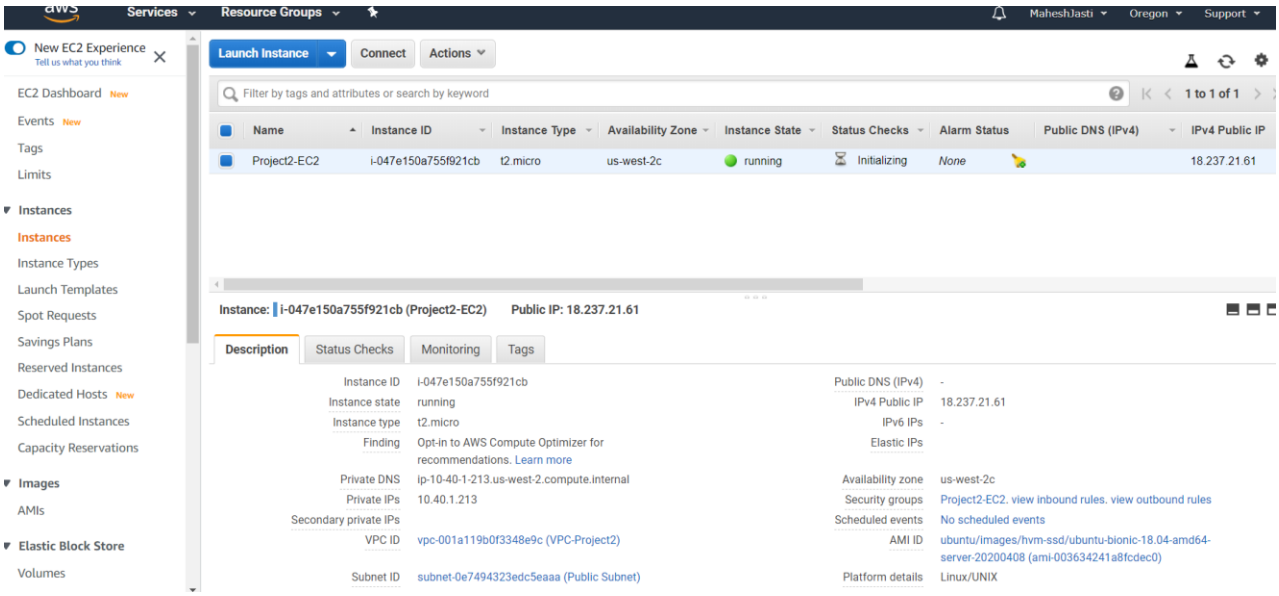
Add the security groups as below… then click review and launch..



Select the key pair and click on “Launch Instances” button

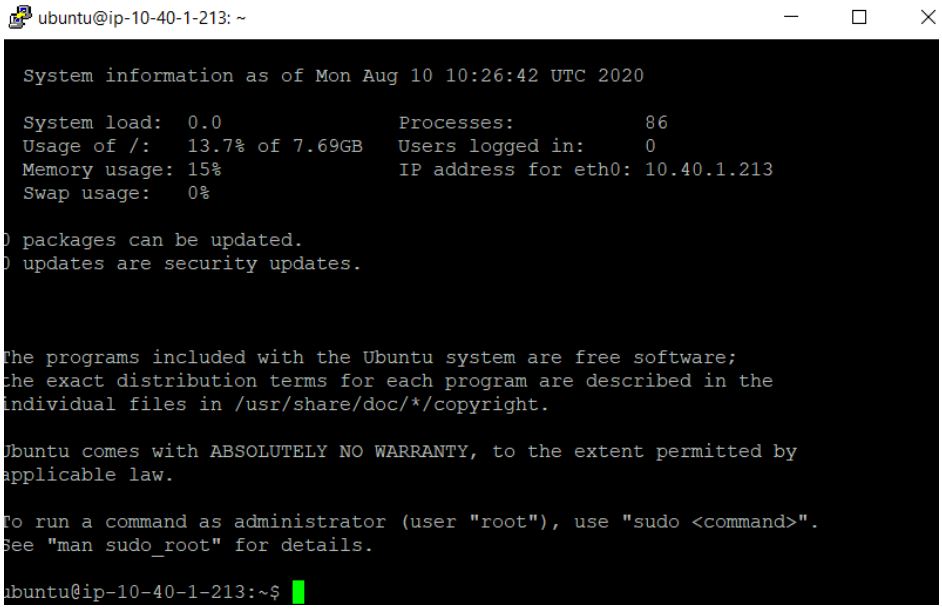


EC2 instance has been created with public IP as 18.237.21.61



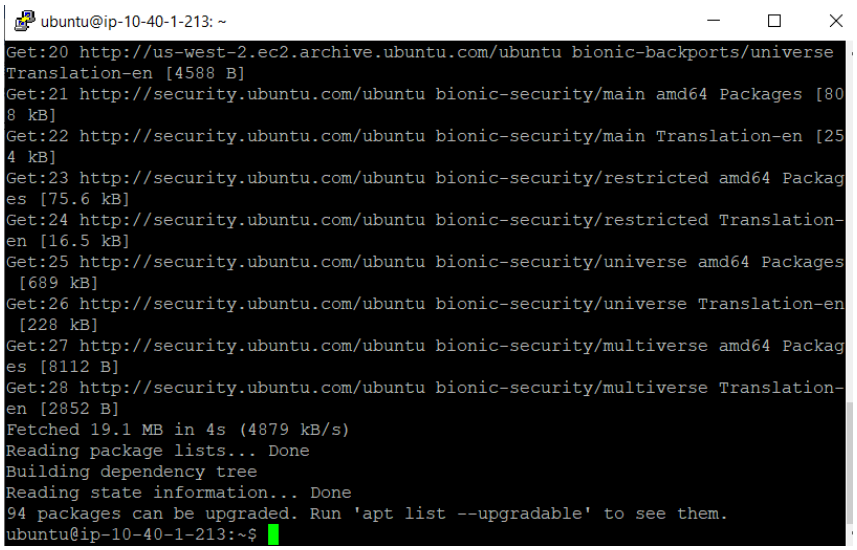
SSH to EC2 Instance and copy the program into the EC2 instance..

SSH thru putty..

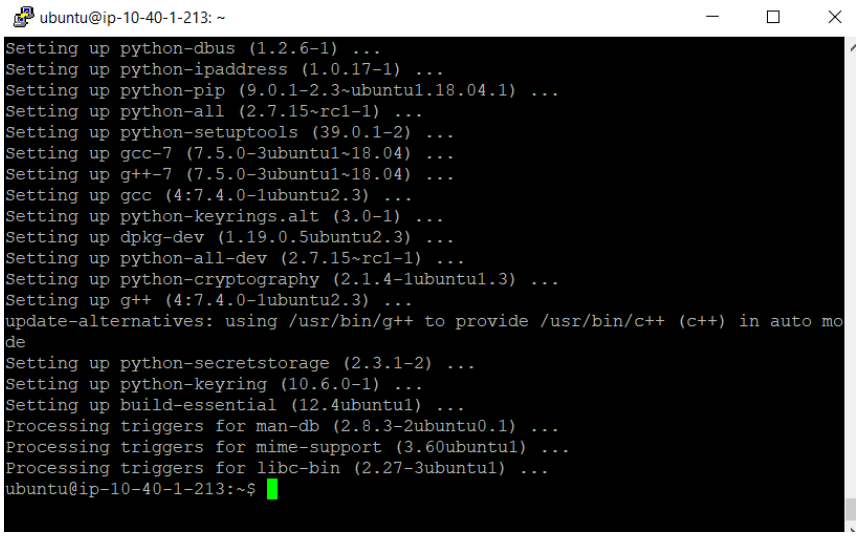


Execute the below commands..

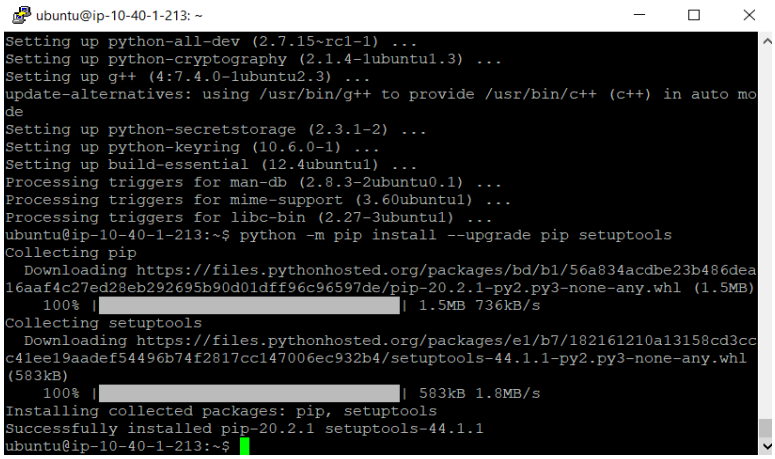
sudo apt update



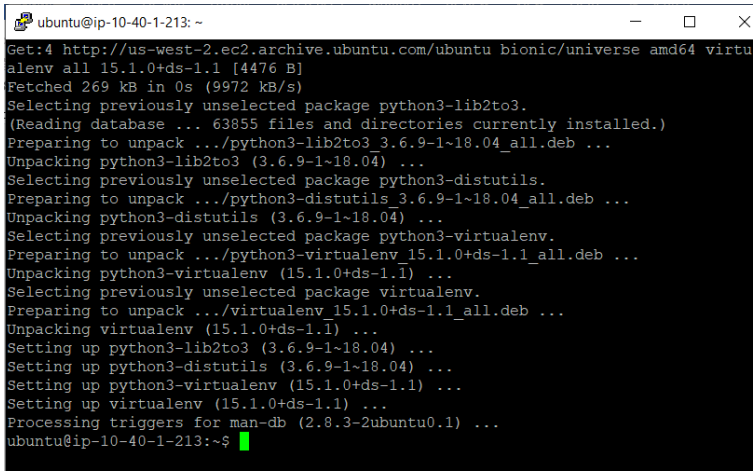
sudo apt install python-pip -y



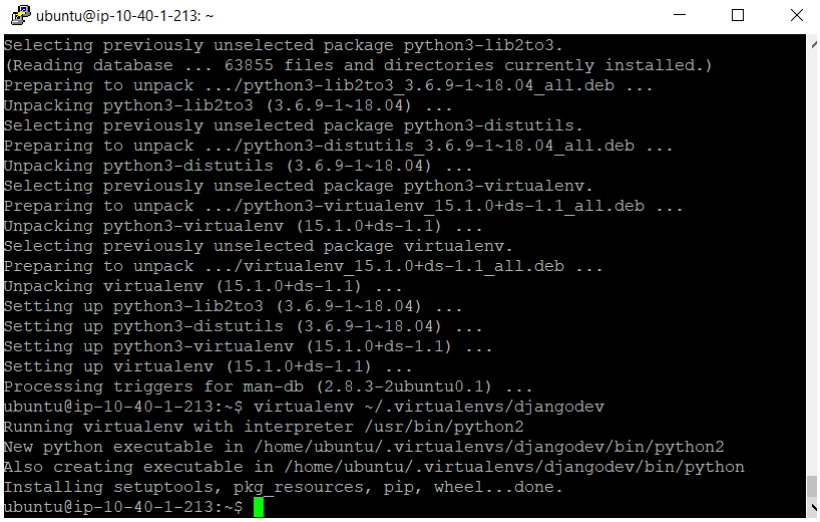
python -m pip install --upgrade pip setuptools



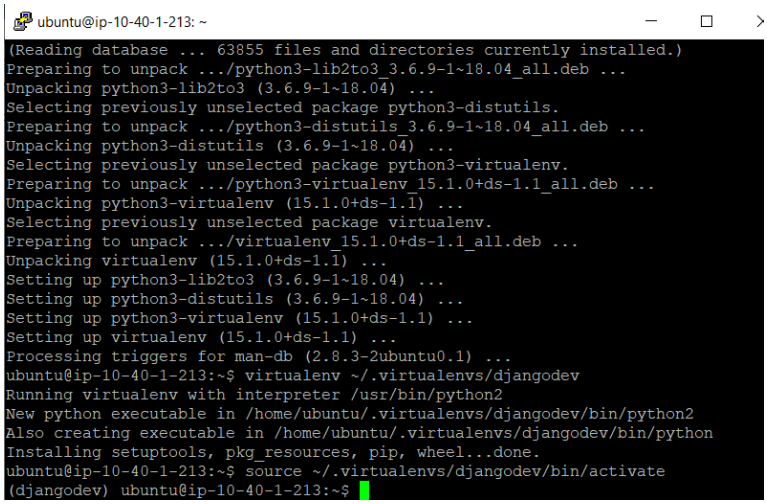
sudo apt install virtualenv -y



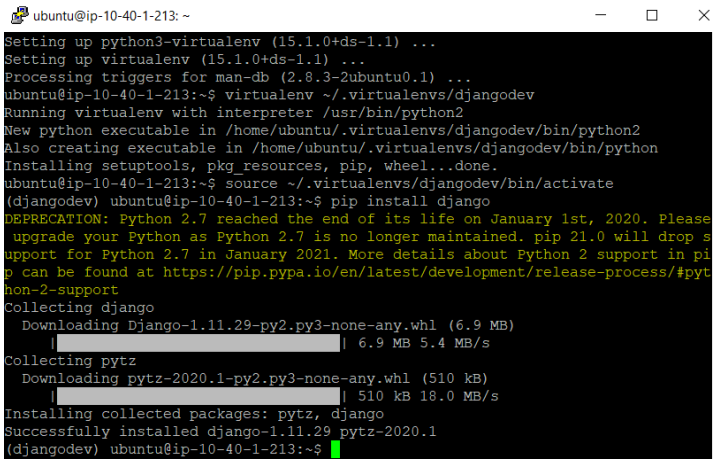
virtualenv ~/.virtualenvs/djangodev



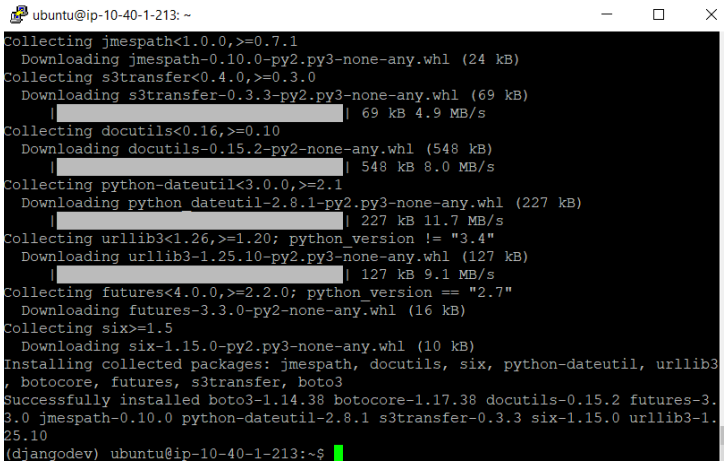
source ~/.virtualenvs/djangodev/bin/activate



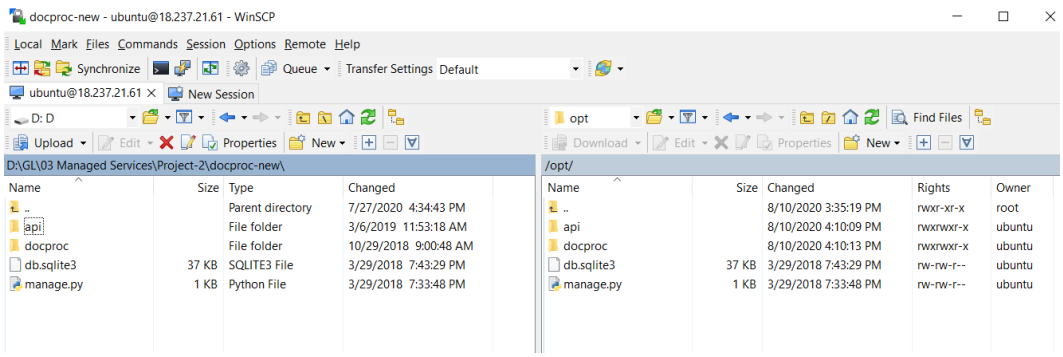
pip install Django



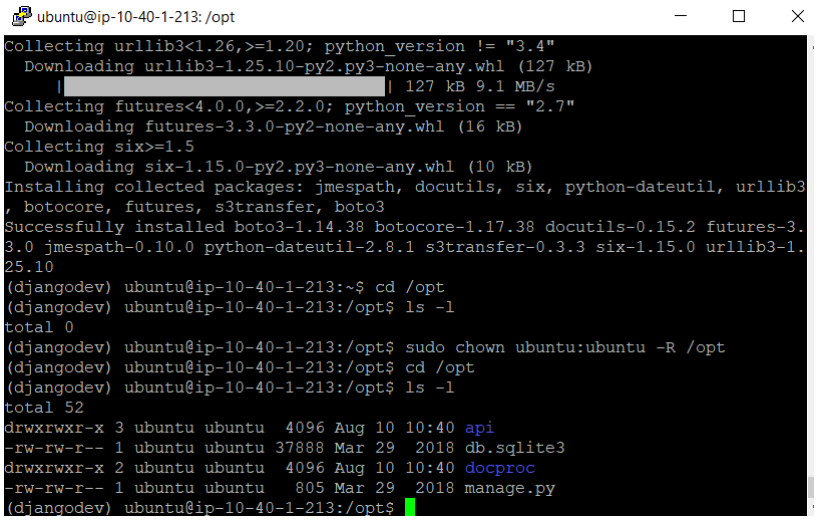
pip install boto3



Move the files to Ubuntu instance. I have used the winscp..



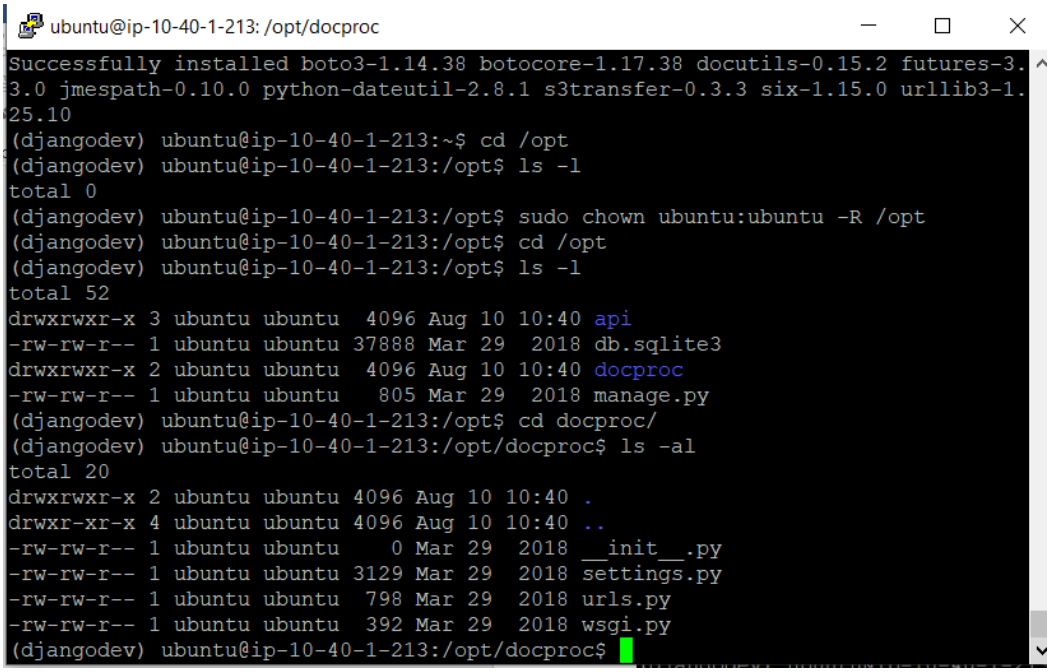
All the files has been moved to opt folder..



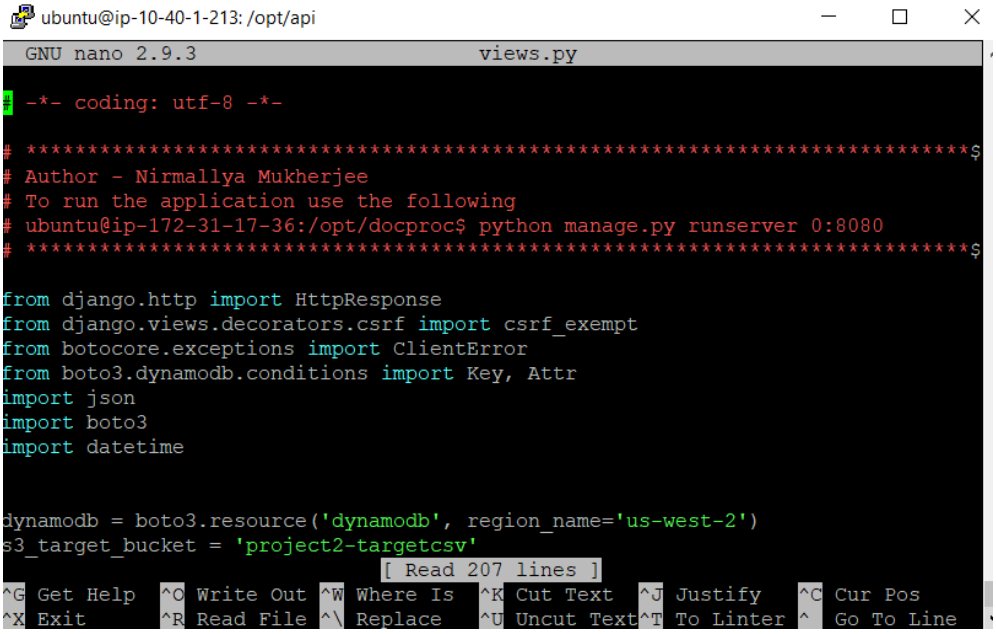
Execute the following 2 commands

Cd docproc/

Ls -al

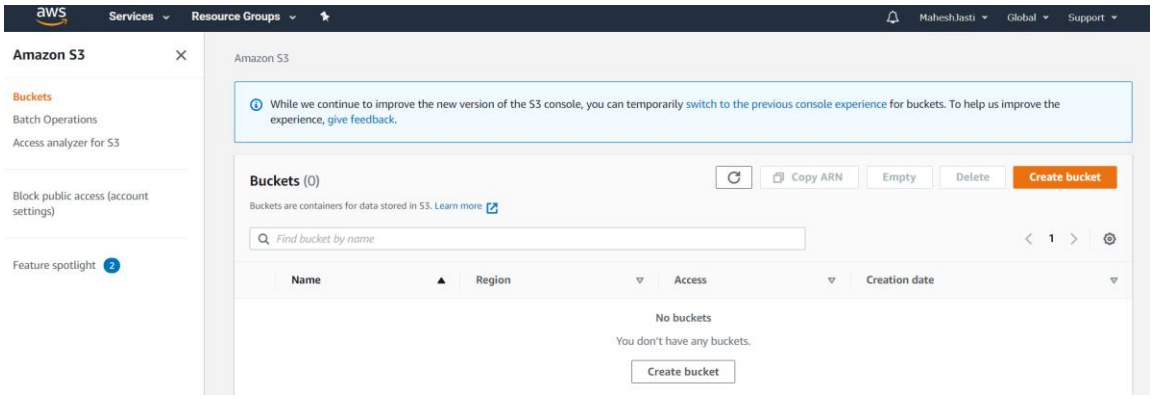


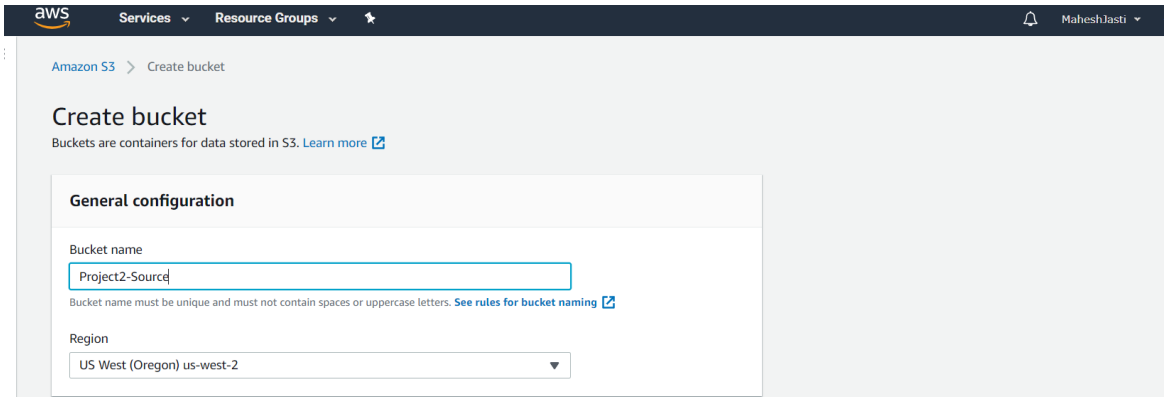
Update the Dynamo Region and S3 Target bucket name in views.py file..



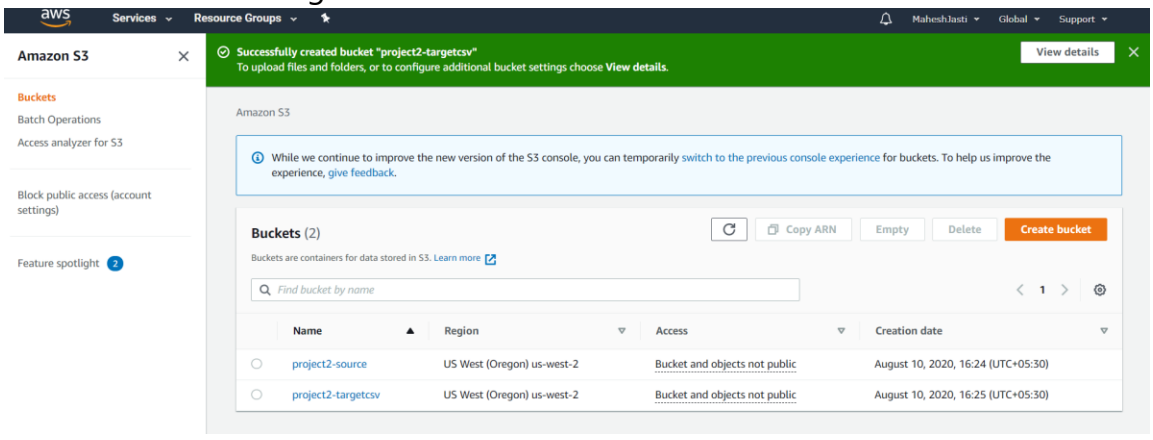
No create a S3 buckets for input and target..

Project2-source – Input Bucket

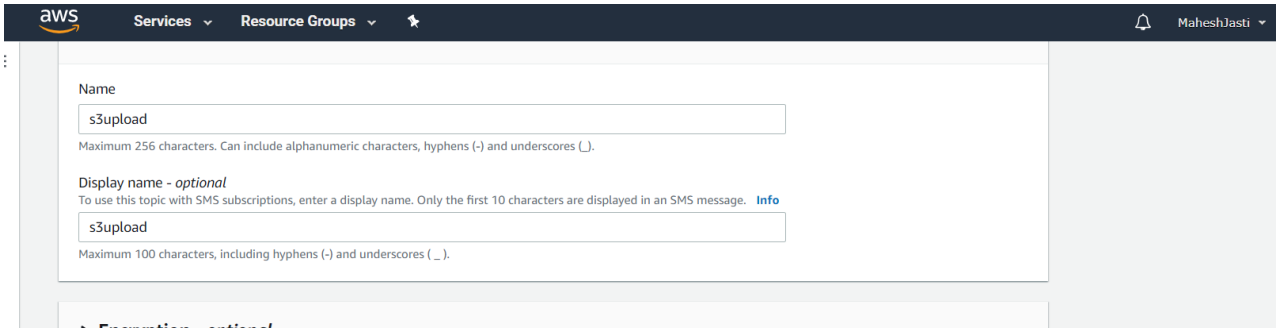




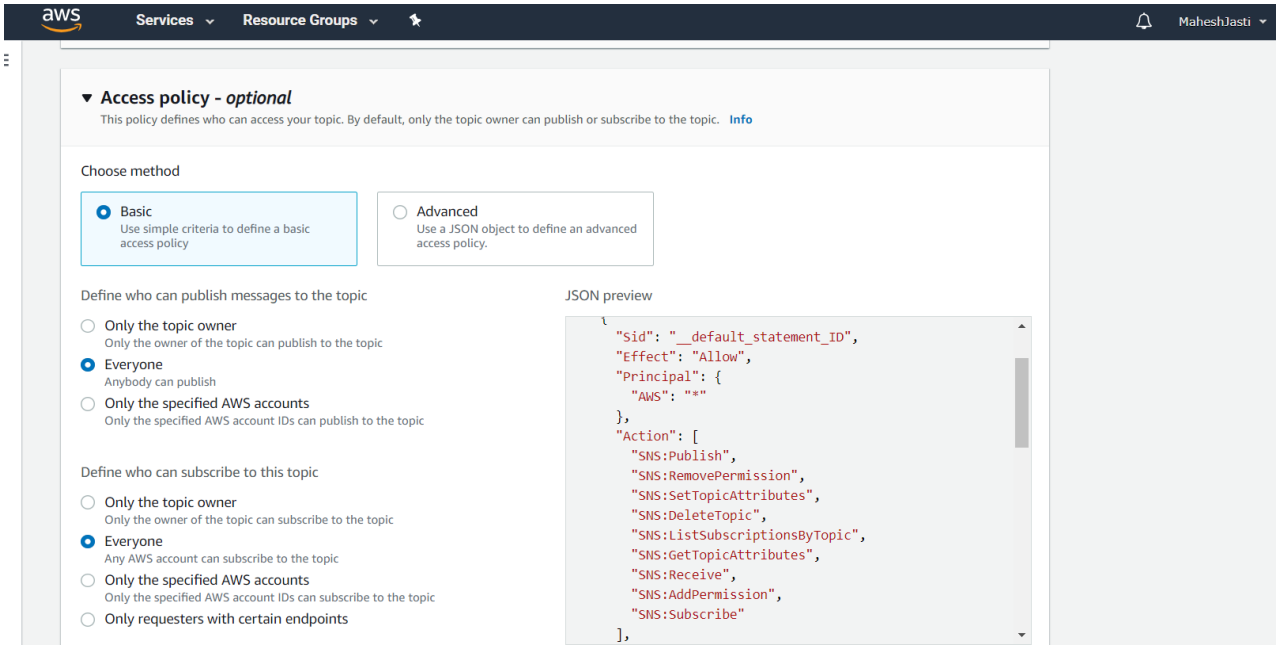
S3 for Source and Target



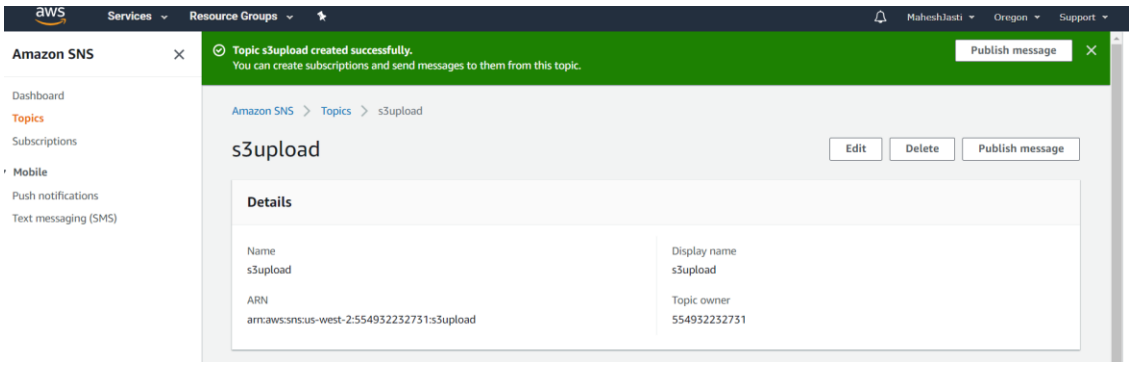
Setup SNS for S3



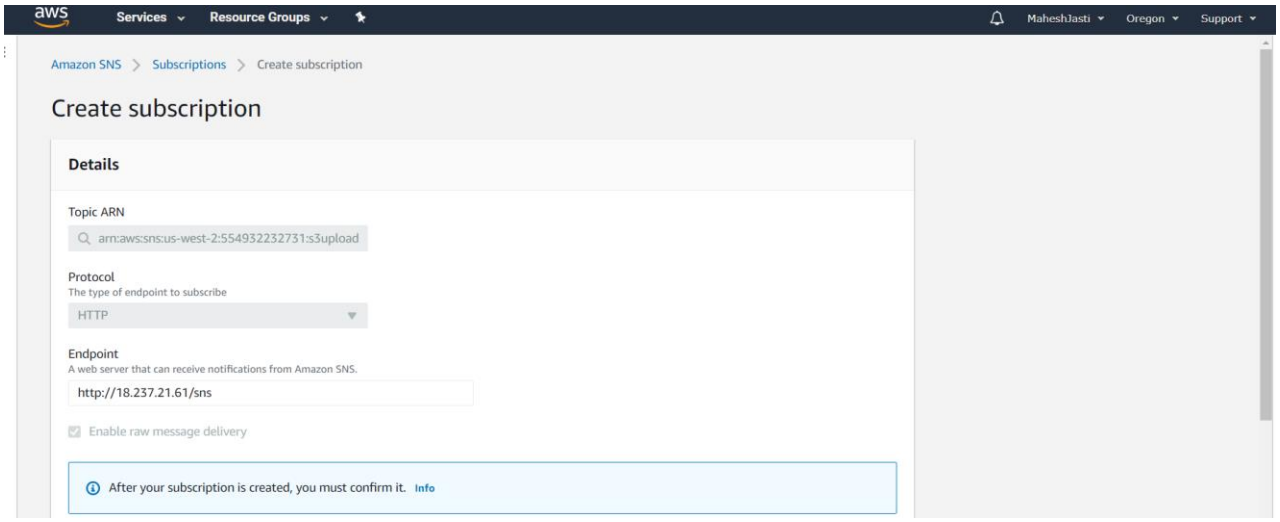
Setup access policy to “everyone”

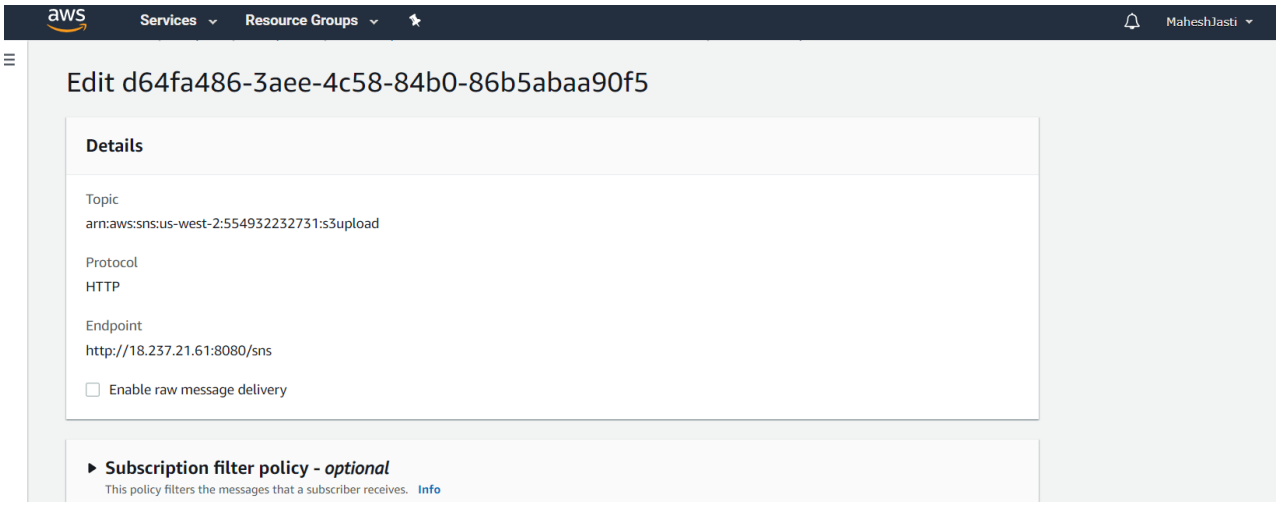


Topic is created

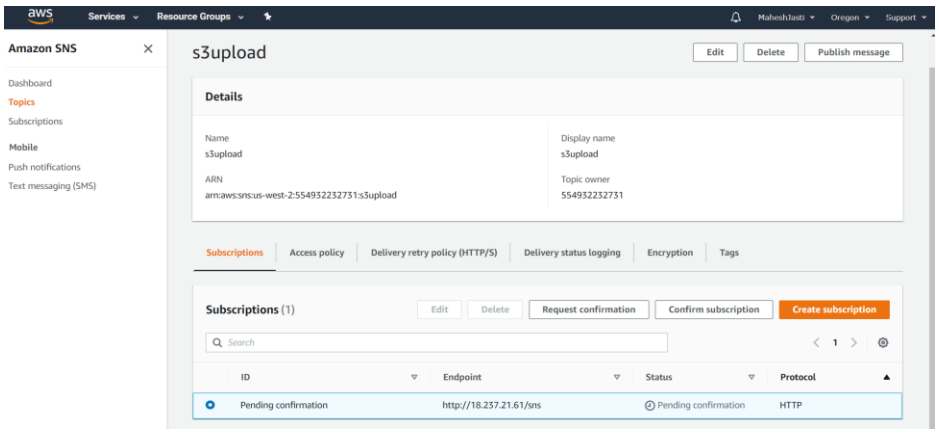


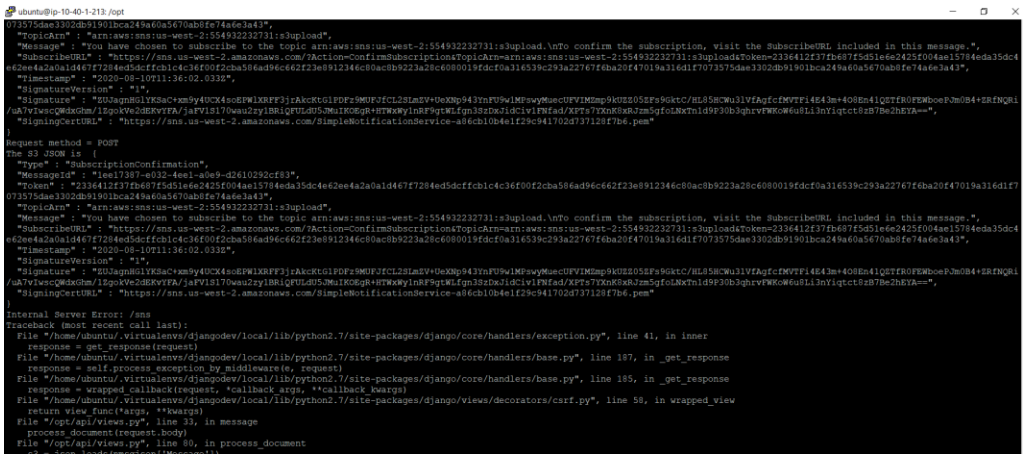
Create a subscription..





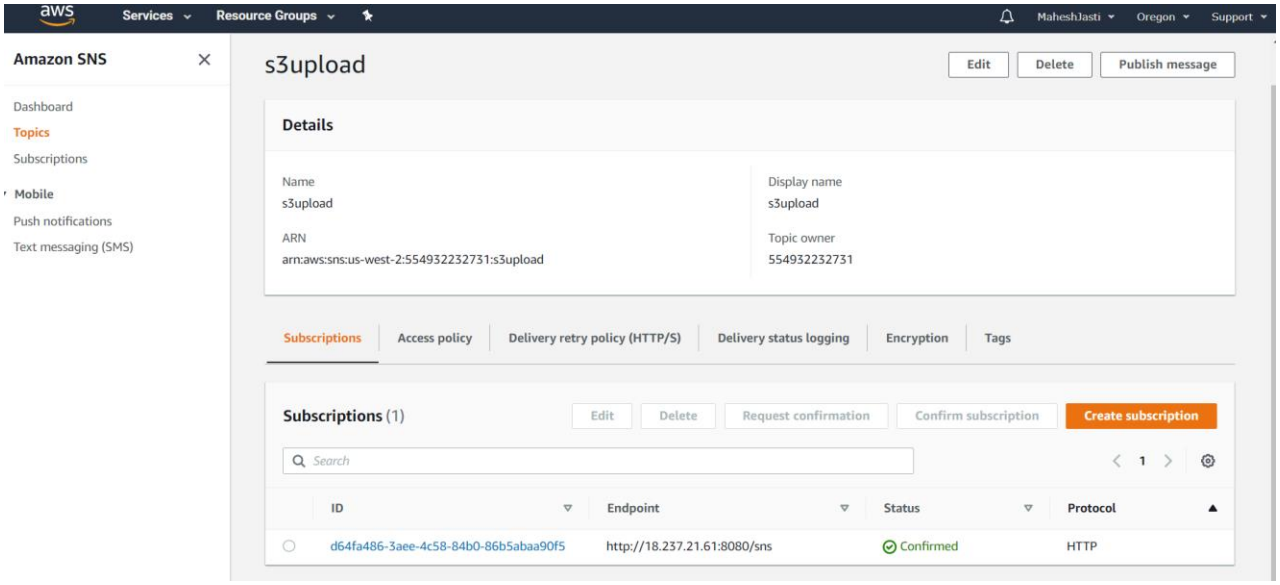
Click on the request confirmation



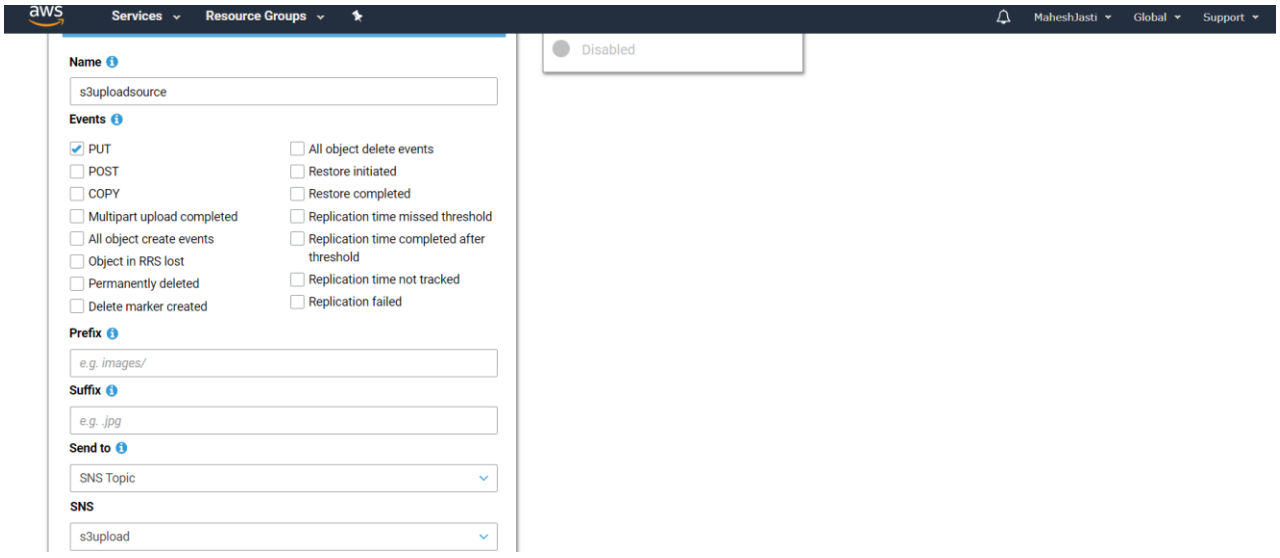


Now copy the url and open in a new browser

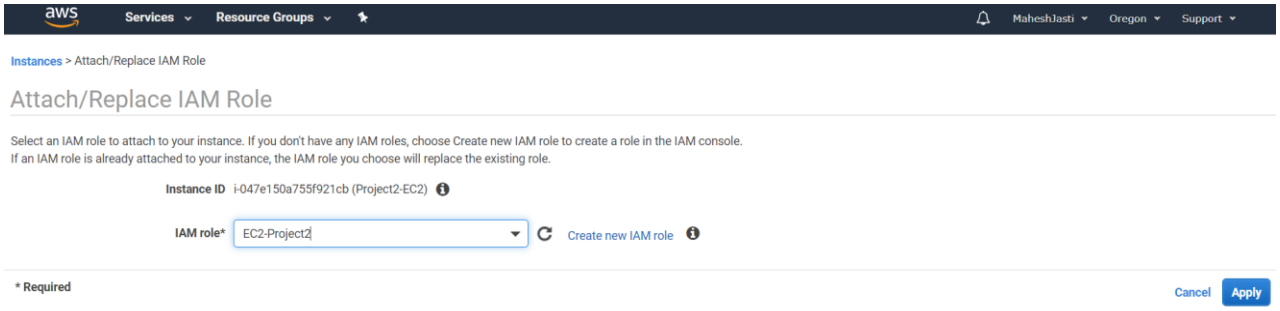
Once you refresh the page your subscription is confirmed.



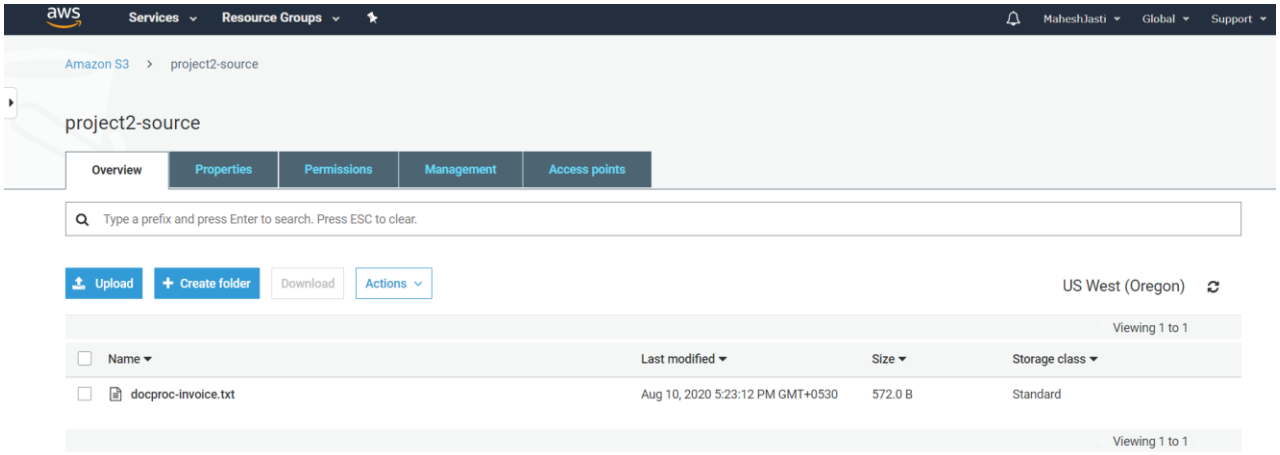
Setup event trigger for S3 source bucket (project2-source bucket)



Now attach the EC2-Project2 role to the EC2 instance..



Now upload a file in S3 porject2-source bucket.

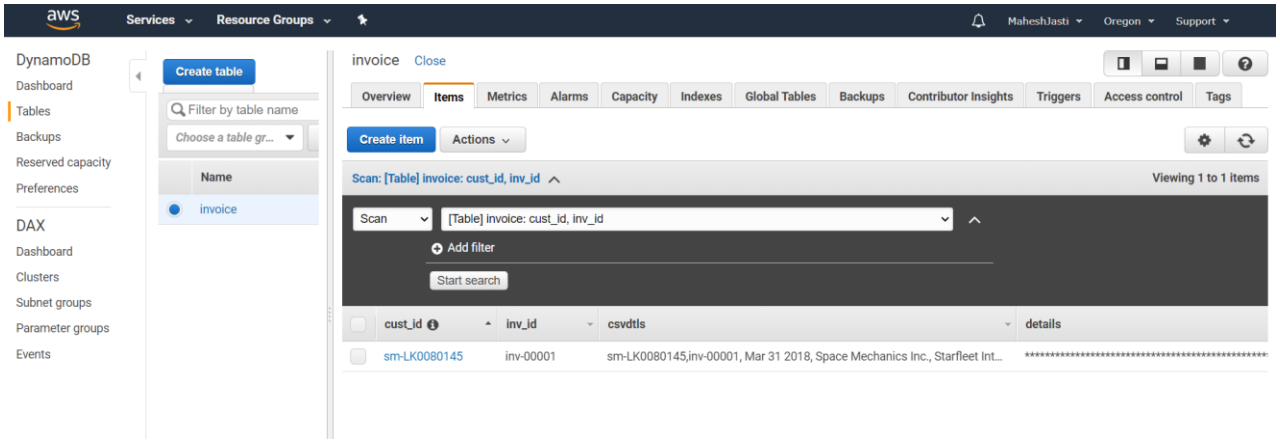


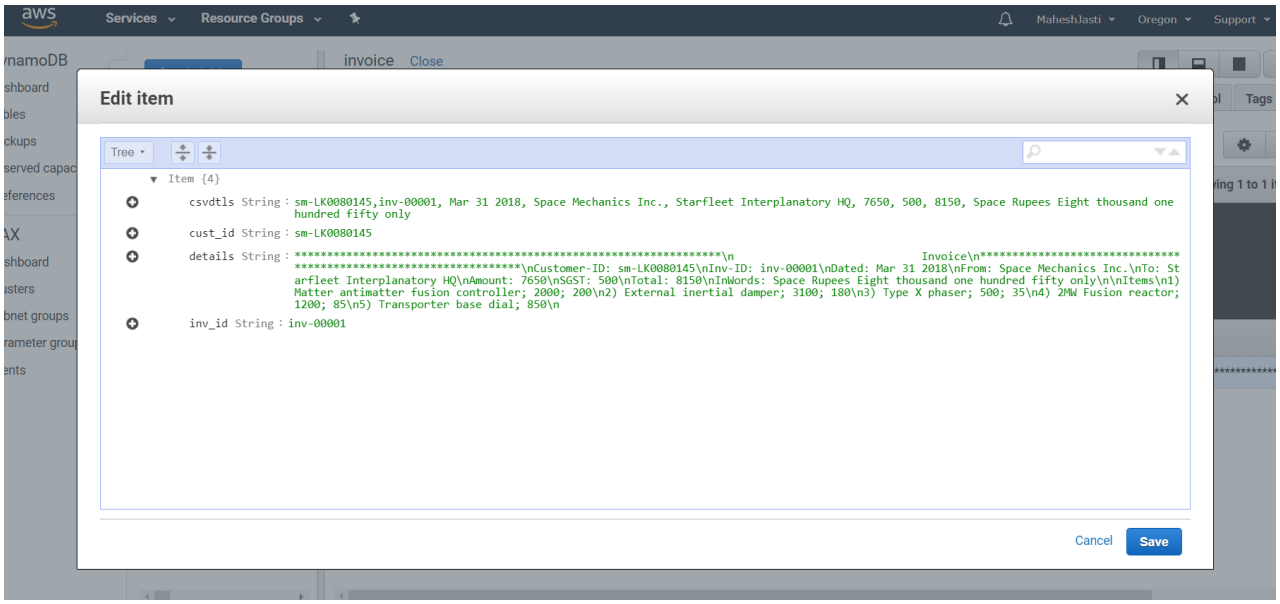
The program has been processed and the data is parsed and loaded into the dynamo table..

ubuntu@ip-10-40-1-213: /opt

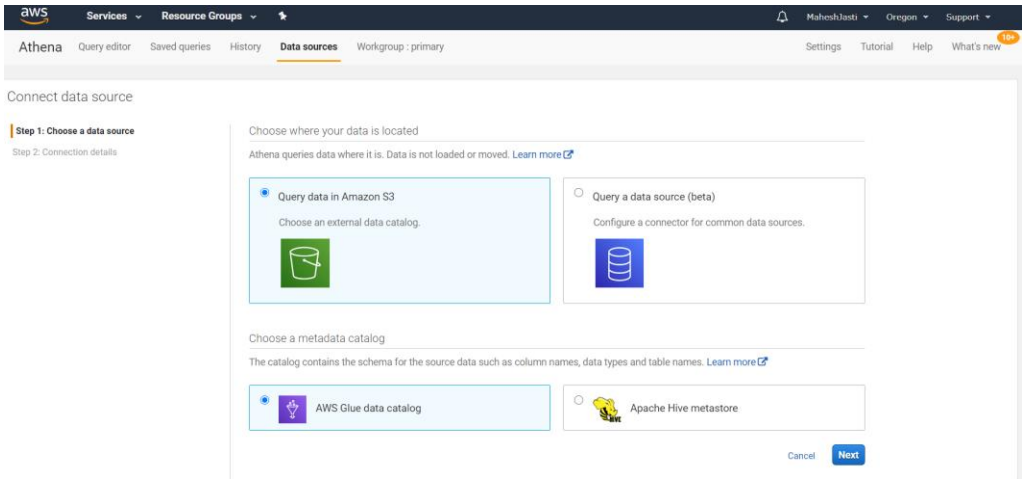
```
bscriptionArn=arn:aws:sns:us-west-2:554932232731:s3upload:d64fa486-3aee-4c58-84b
0-86b5abaa90f5"
}
Will read the file docproc-invoice.txt from the bucket project2-source
Line->  **********************************************************************
Line->                              Invoice
Line->  **********************************************************************
Line->  Customer-ID: sm-LK0080145
  Found Customer-ID  sm-LK0080145
Line->  Inv-ID: inv-00001
  Found Invoice-ID  inv-00001
Line->  Dated: Mar 31 2018
Line->  From: Space Mechanics Inc.
Line->  To: Starfleet Interplanatory HQ
Line->  Amount: 7650
Line->  SGST: 500
Line->  Total: 8150
Line->  InWords: Space Rupees Eight thousand one hundred fifty only
Line->
Line->  Items
Line->  1) Matter antimatter fusion controller; 2000; 200
Line->  2) External inertial damper; 3100; 180
Line->  3) Type X phaser; 500; 35
Line->  4) 2MW Fusion reactor; 1200; 85
Line->  5) Transporter base dial; 850
  Found dated   Mar 31 2018
  Found fromcust   Space Mechanics Inc.
  Found tocust   Starfleet Interplanatory HQ
  Found amt   7650
  Found sgst   500
  Found tot   8150
  Found words   Space Rupees Eight thousand one hundred fifty only
CSV -> sm-LK0080145,inv-00001, Mar 31 2018, Space Mechanics Inc., Starfleet Inte
rplanatory HQ, 7650, 500, 8150, Space Rupees Eight thousand one hundred fifty on
ly

**************************************************************************
Creating table invoice
 DONE

**************************************************************************
Inserting data in the table
  Done.
Written new s3 file 77772933_sm-LK0080145_inv-00001.csv
Uploading S3 object content sm-LK0080145,inv-00001, Mar 31 2018, Space Mechanics
 Inc., Starfleet Interplanatory HQ, 7650, 500, 8150, Space Rupees Eight thousand
 one hundred fifty only
Done
```

The items are loaded into dynamo db

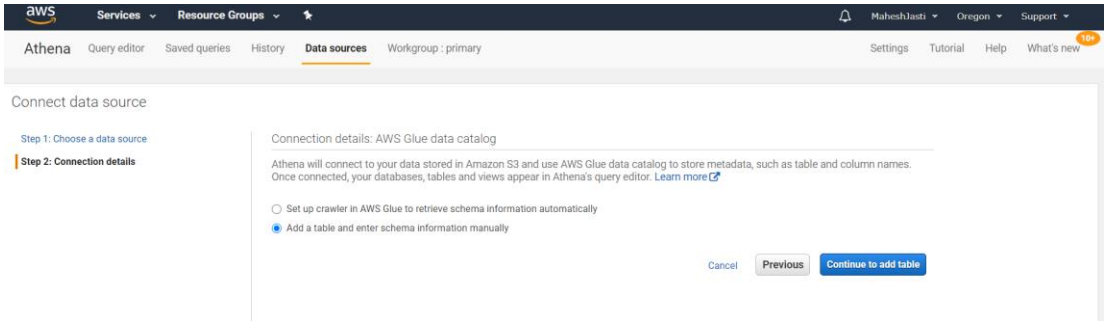




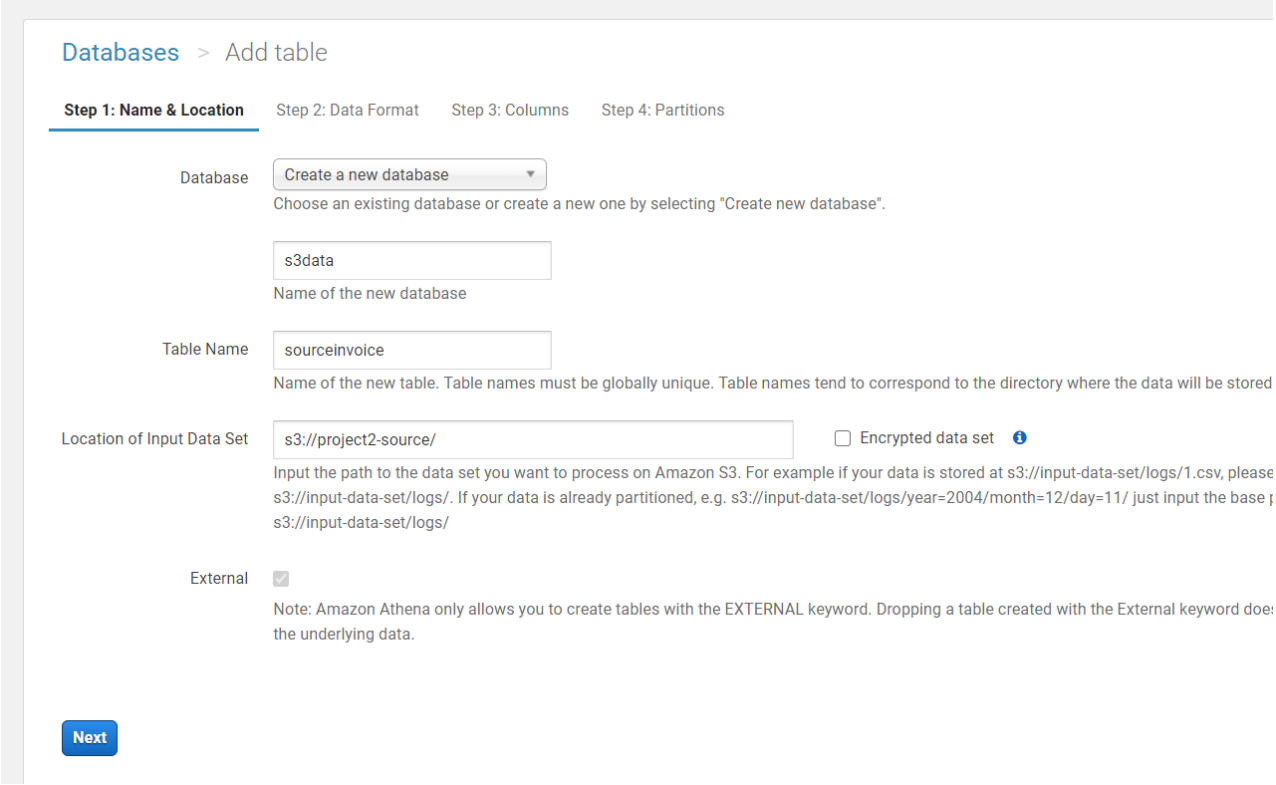
Go to Athena window.

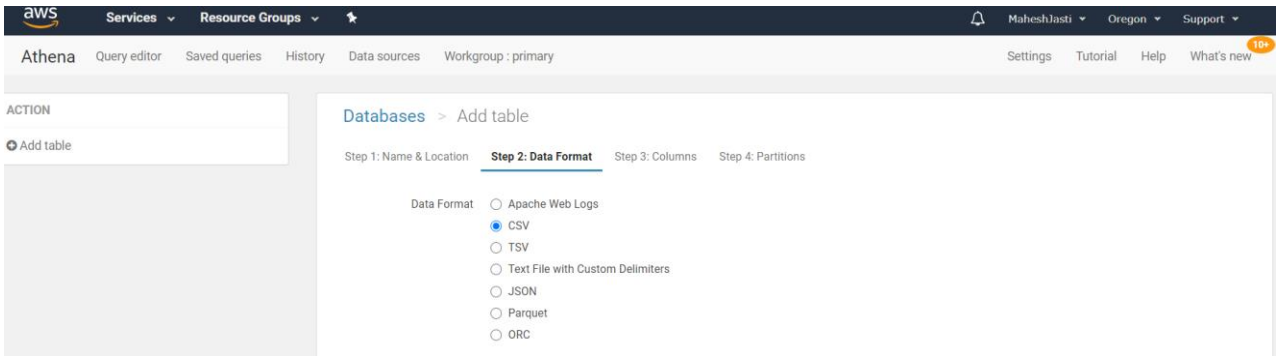


Add a table

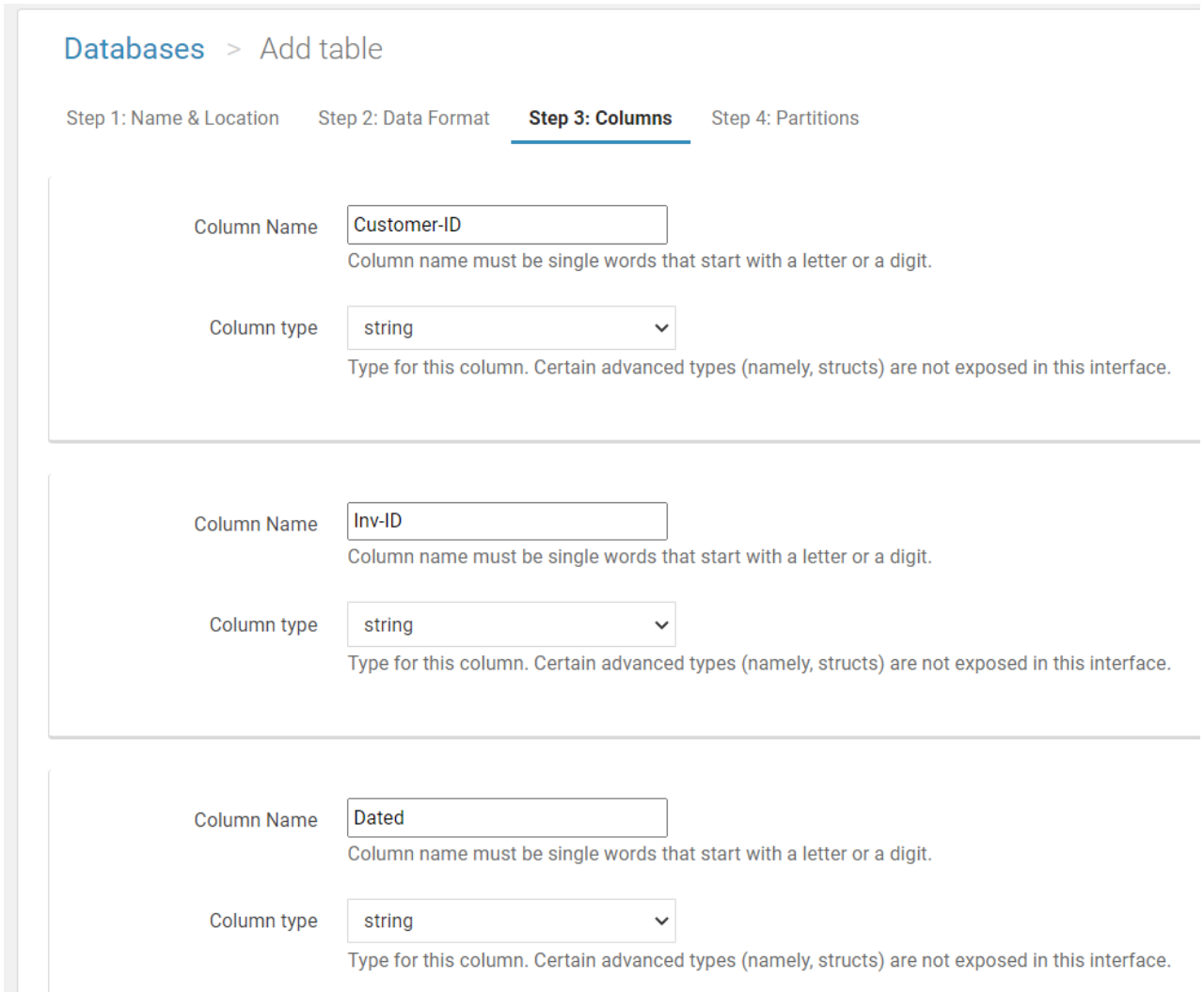


Enter the required details

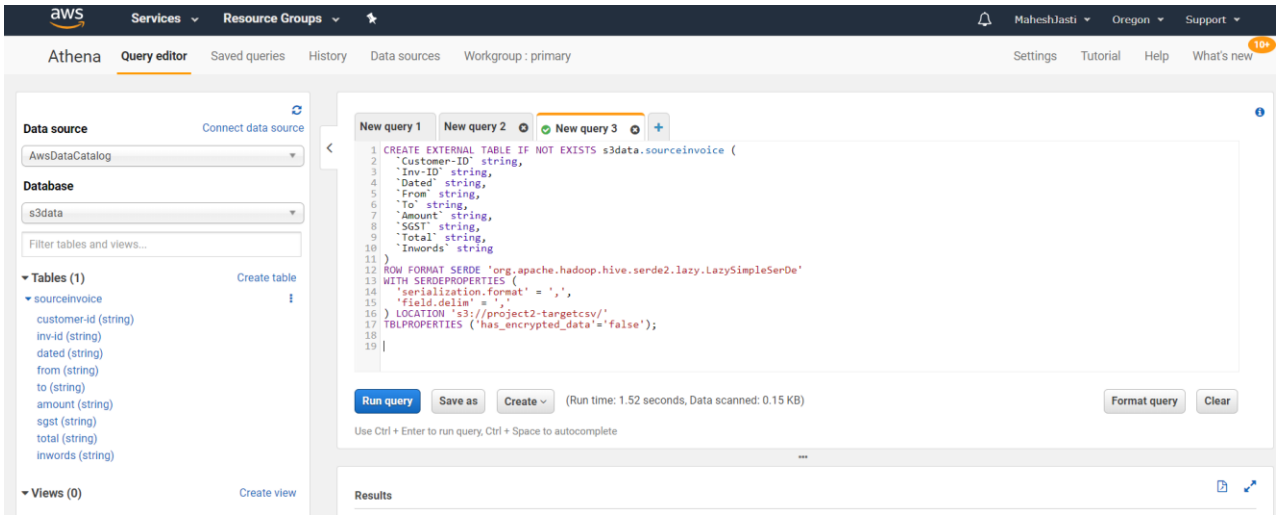




Add the columns as per the input text doc..

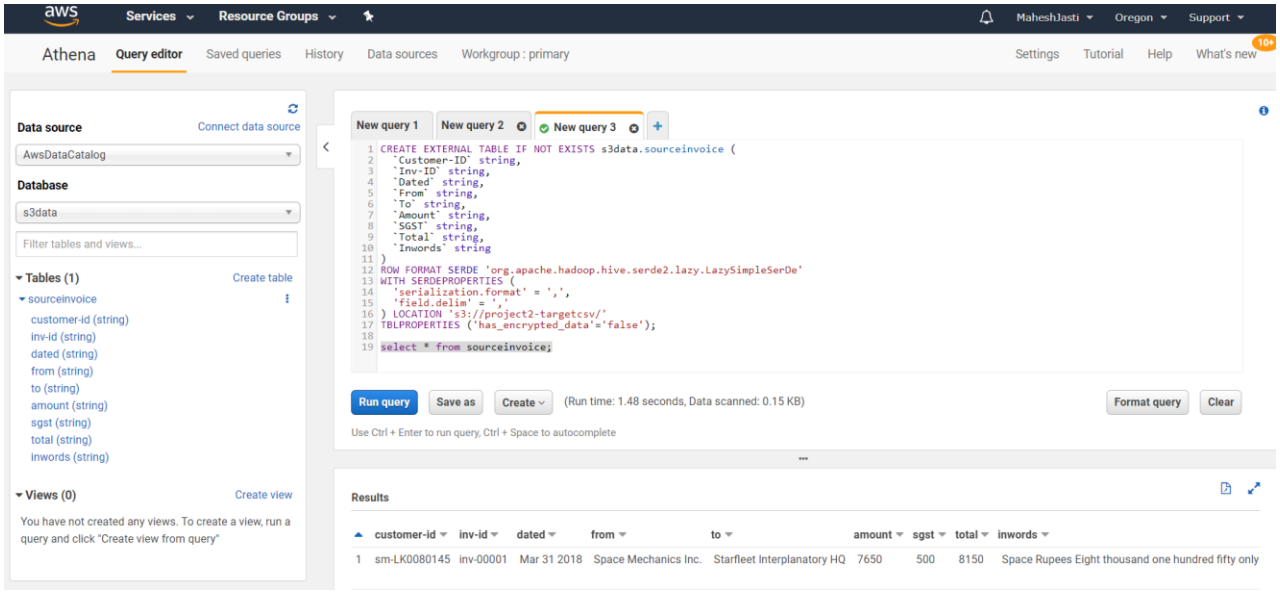


Add all the other columns and click on Create Table..



Now write the below query and click on Run Query by selecting ..

Select * from sourceinvoice



Section 6: Lessons learnt / Observations

1. After creating VPC, we need to create the IGWY and setup route table
2. If we need public IP, we need to select the auto assign IP address
3. Set up SNS topic with HTTP endpoint
4. A custom TCP security rule should be added for 8080 port (web server)
5. While creating table in Athena, we should store the query in some s3 location otherwise
we get the error as "No output location provided. An output location is required either through the Workgroup result configuration setting or as an API input. (Service: AmazonAthena; Status Code: 400; Error Code: InvalidRequestException)"
6. Remember to use 8080 while check in browser http://<public-ip>:8080/sns
7. While creating subscription, we have to give http://<public-ip>:8080/sns