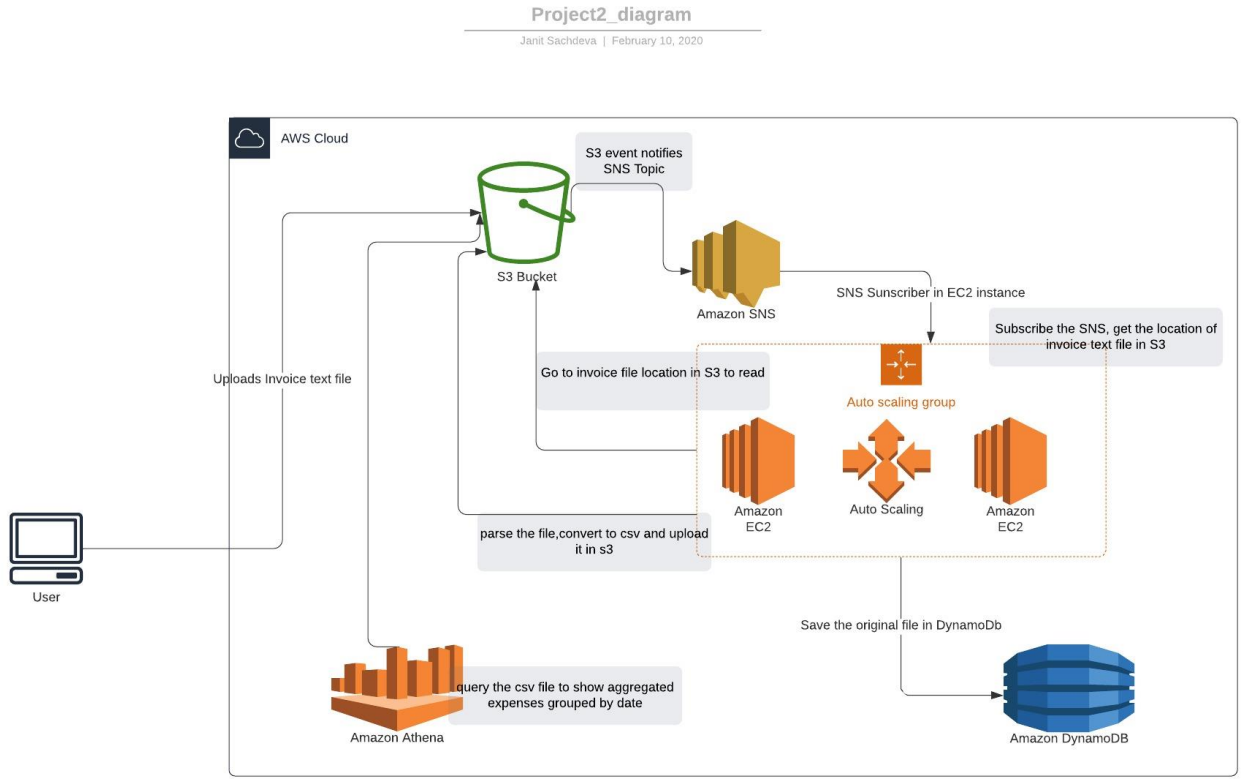
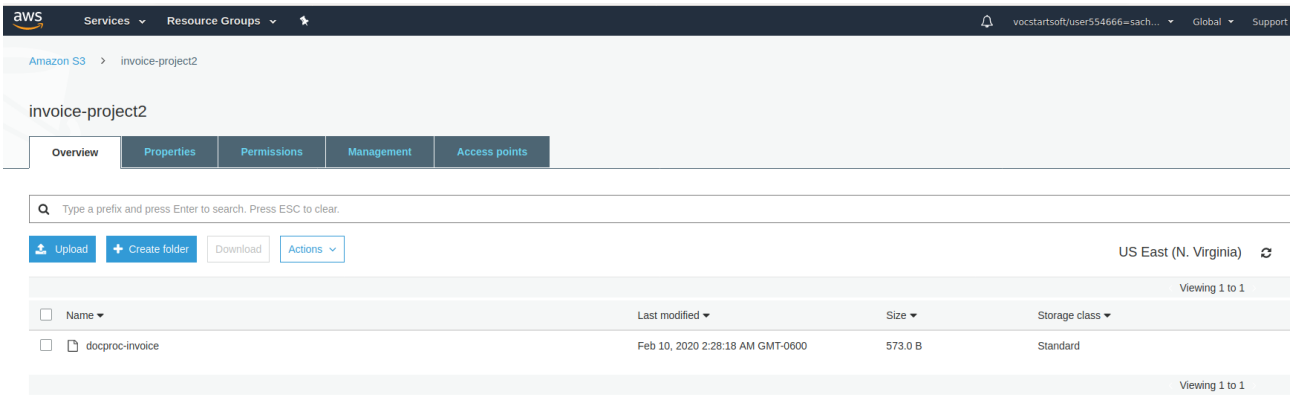
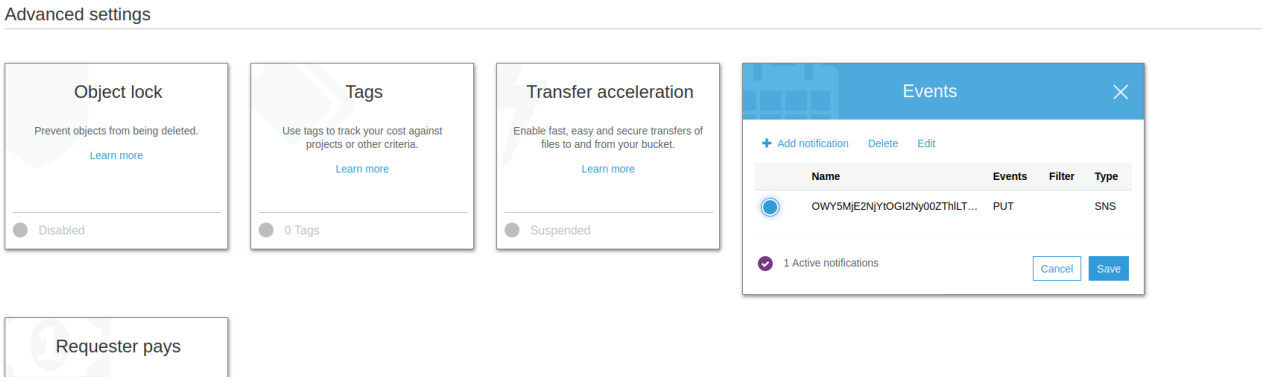
PROJECT 2

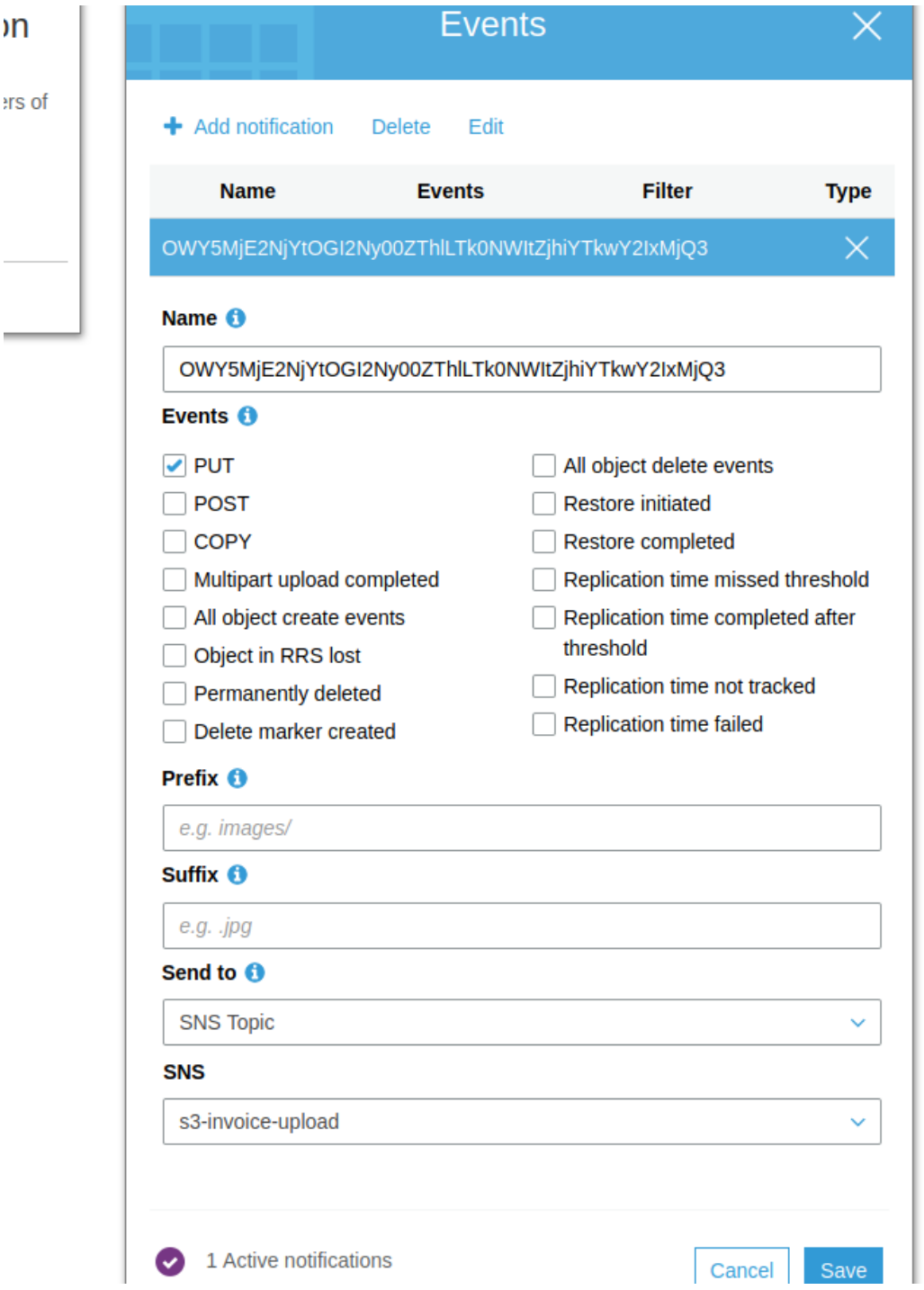


Steps to be followed:

1. Create a S3 bucket named-**invoice-project2**
2. Upload docproc-invoice file on s3 bucket.
3. Go to properties—Events
4. Create an SNS topic so that event would be triggered when file is uploaded.

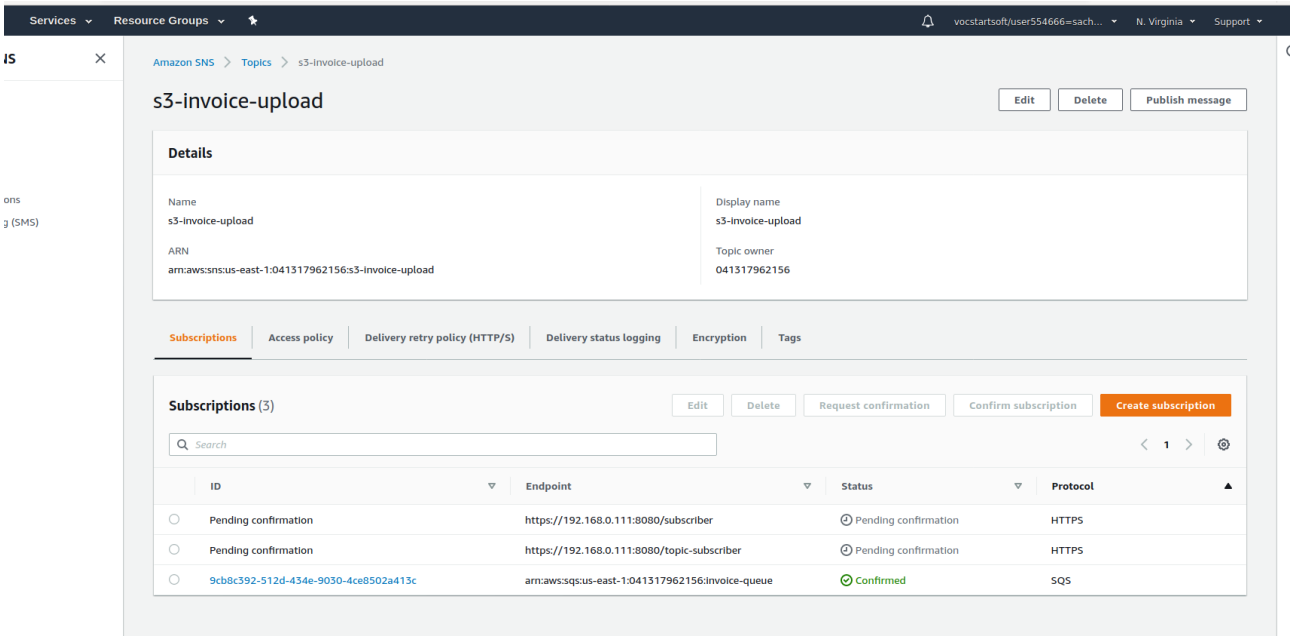

aws
Services
Resource Groups
vocstartsoft/user554666=sach...
Global
Support
Amazon S3
invoice-project2
invoice-project2
Overview
Properties
Permissions
Management
Access points
Type a prefix and press Enter to search. Press ESC to clear.
Upload
Create folder
Download
Actions
US East (N. Virginia)
Viewing 1 to 1
Name
Last modified
Size
Storage class
docproc-invoice
Feb 10, 2020 2:28:18 AM GMT-0600
573.0 B
Standard
Viewing 1 to 1


Advanced settings
Object lock
Prevent objects from being deleted.
Learn more
Disabled
Tags
Use tags to track your cost against projects or other criteria.
Learn more
0 Tags
Transfer acceleration
Enable fast, easy and secure transfers of files to and from your bucket.
Learn more
Suspended
Events
Add notification
Delete
Edit
Name
Events
Filter
Type
OWY5MjE2NjYtOGI2Ny00ZThlLT...
PUT
SNS
1 Active notifications
Cancel
Save
Requester pays



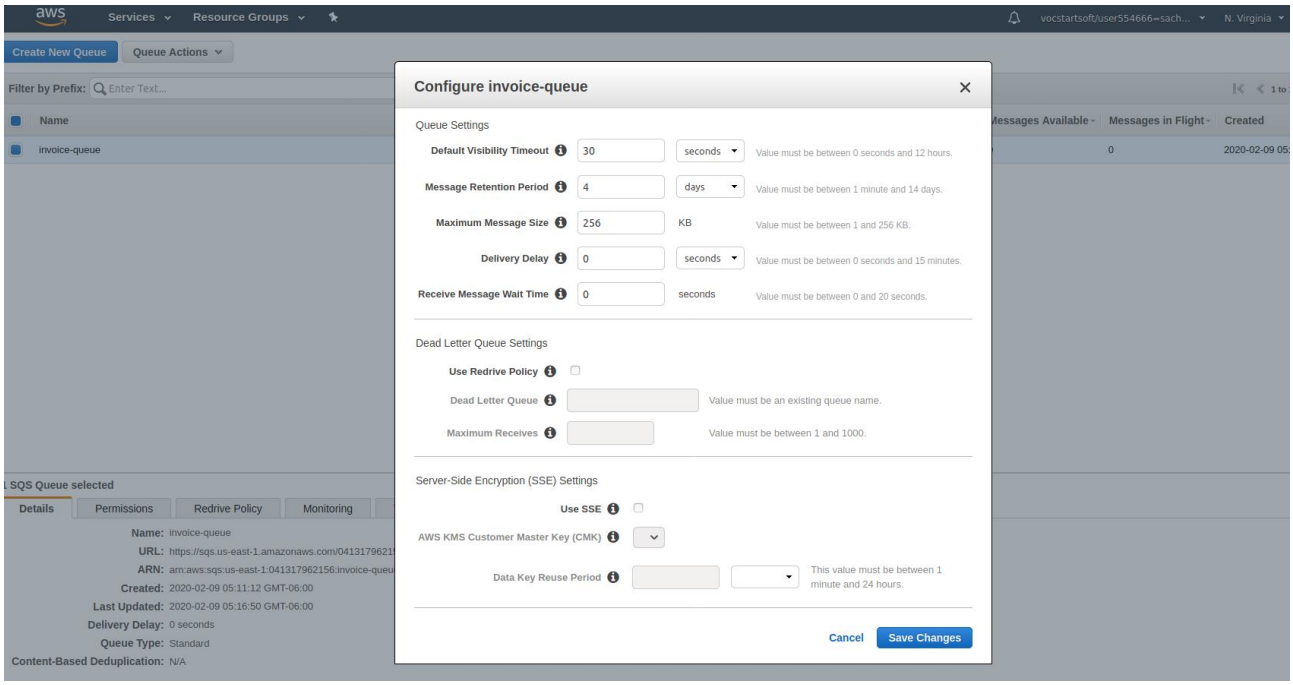
5. SNS topic name- s3-invoice-upload

6. Go to SNS

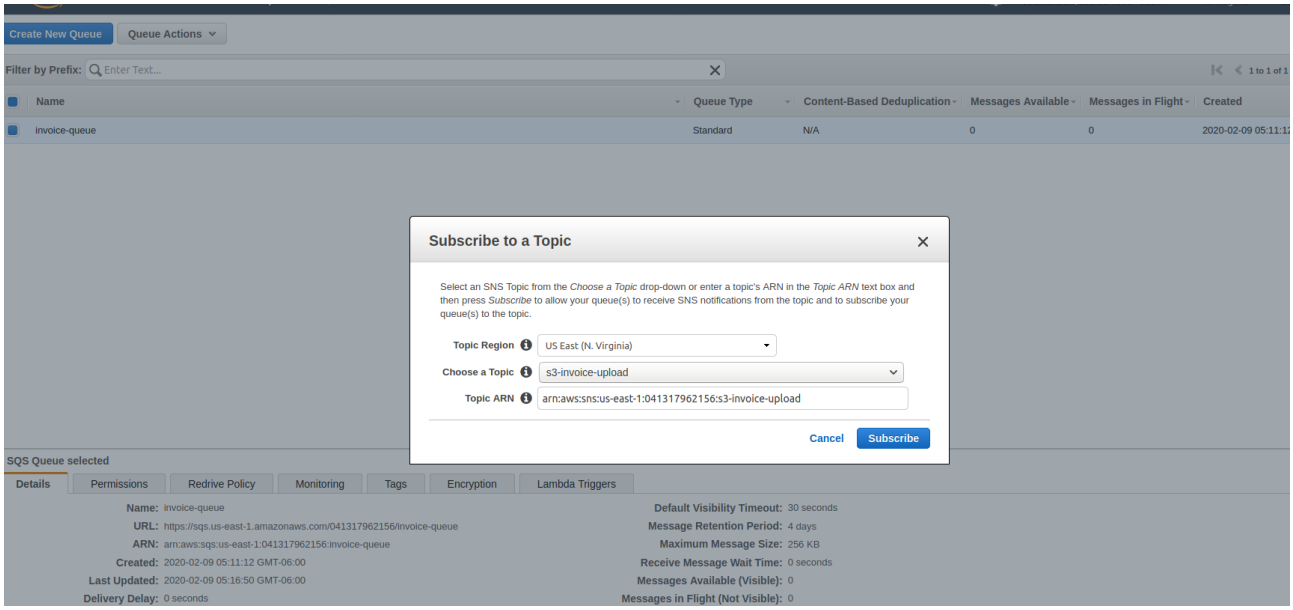


7. Create a HTTP subscription or SQS subscription.

8. Create a SQS queue- **invoice-queue**

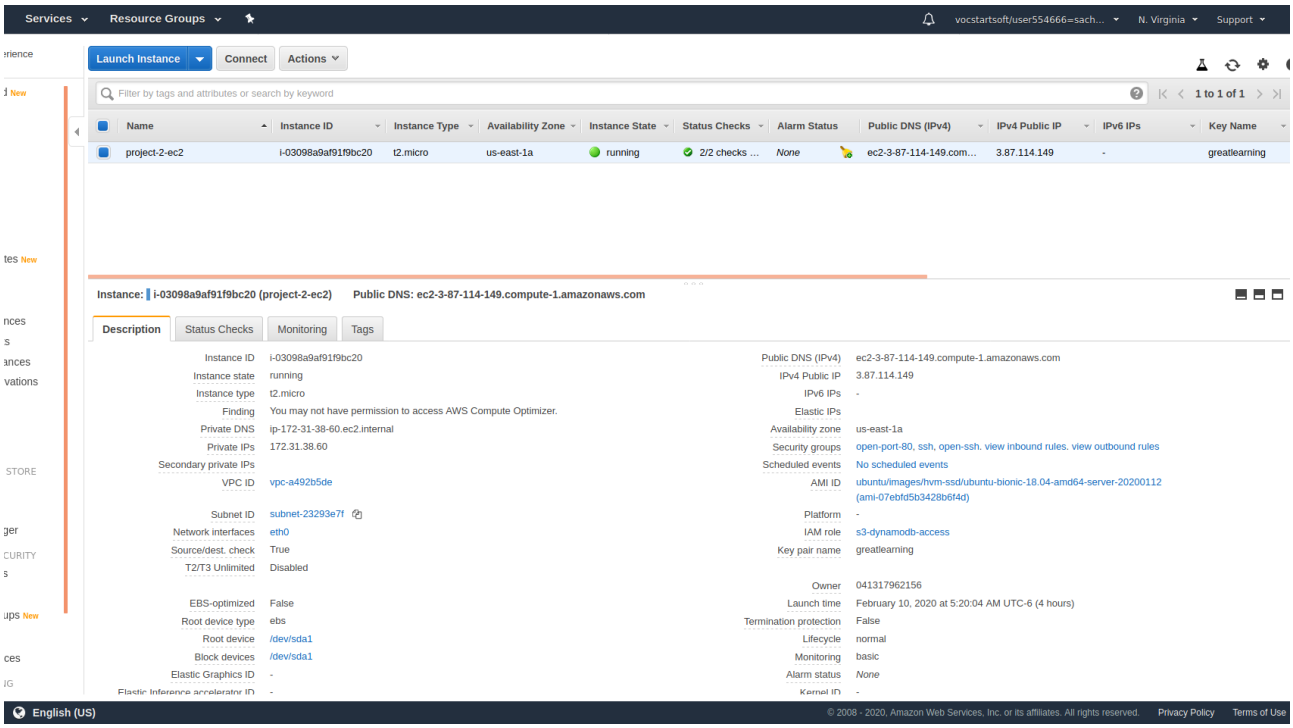


9. Subscribe to the SNS topic



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

11. Create an ec2 instance and assign access to cloudwatch, s3, sns, sqs, dynamodb.



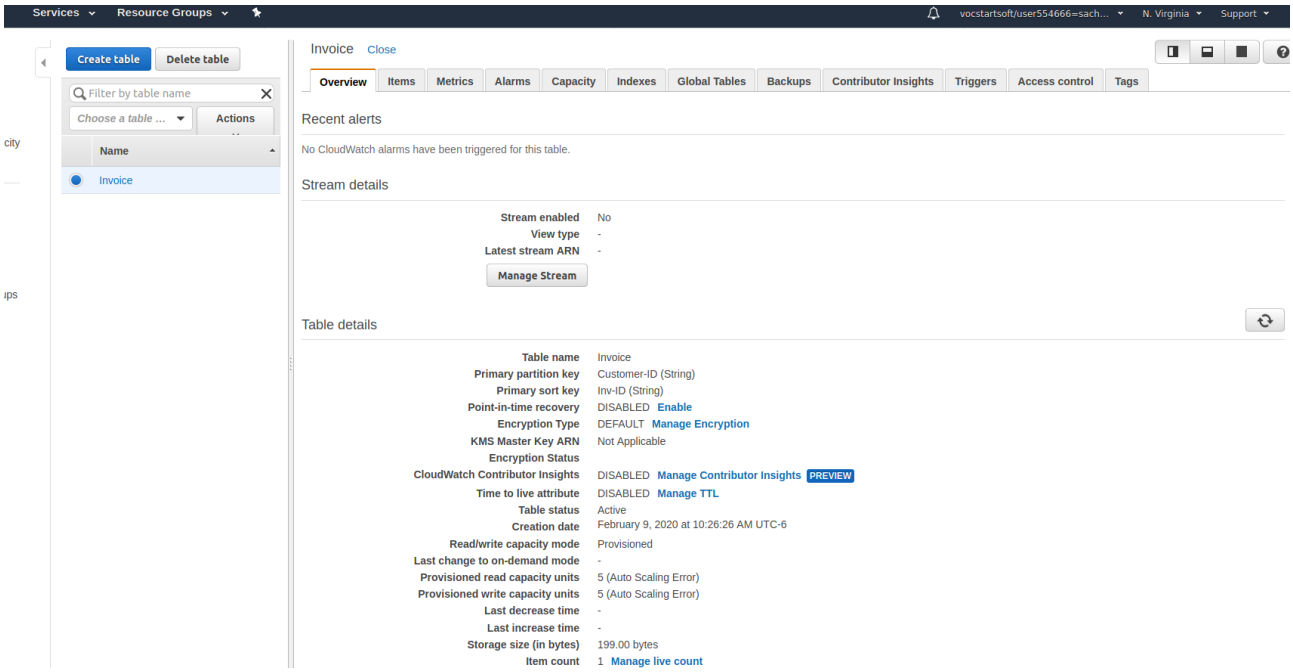
12. A custom program running in EC2 will listen to SQS and get the message placed by S3 event to SNS.

13. 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.

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

15. Code is written in java and has been attached along with the document submission.

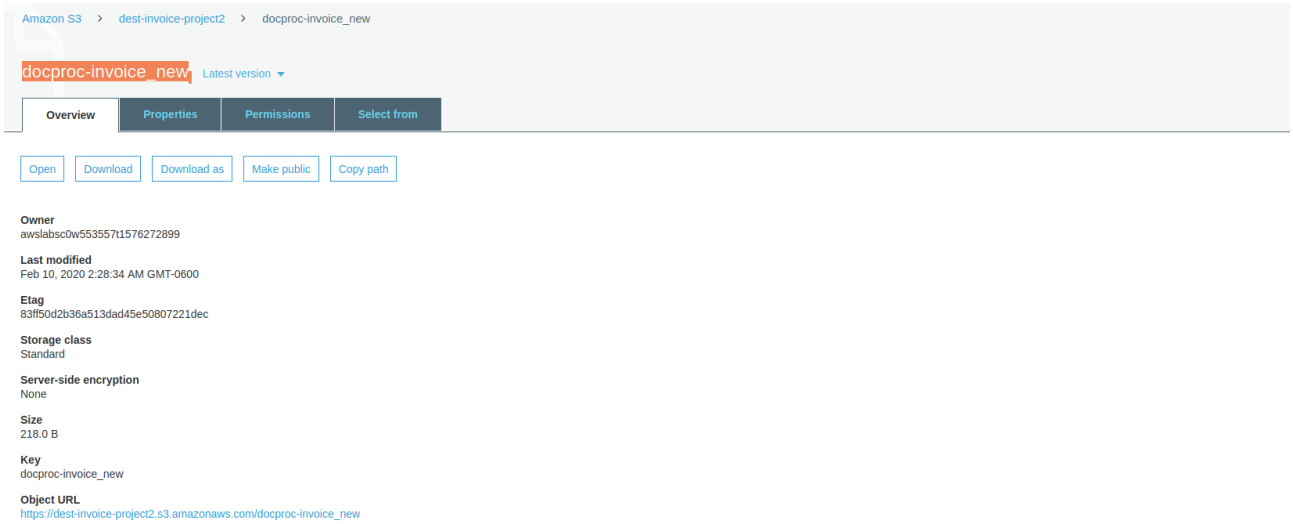
16. Create a dynamoDb table-**Invoice.**
Data will be written via java program.



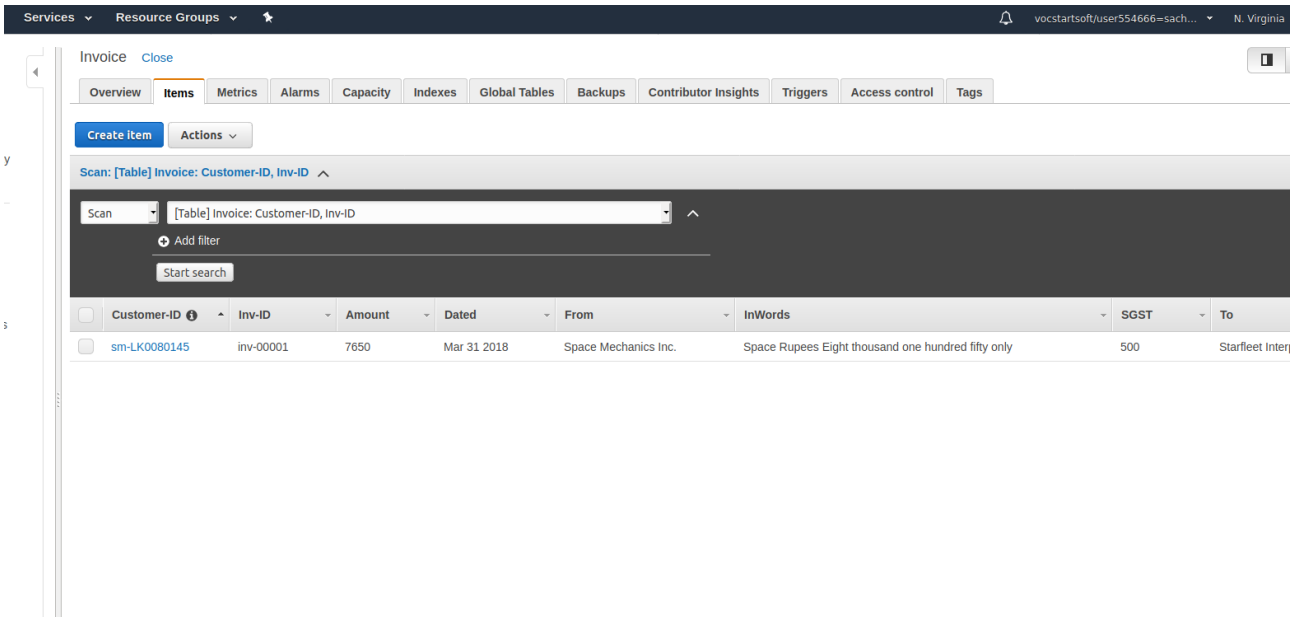
17. After running the java program

Results:

- Destination bucket is created- **dest-invoice-project2**
- docproc-invoice_new file is uploaded which is in csv format.



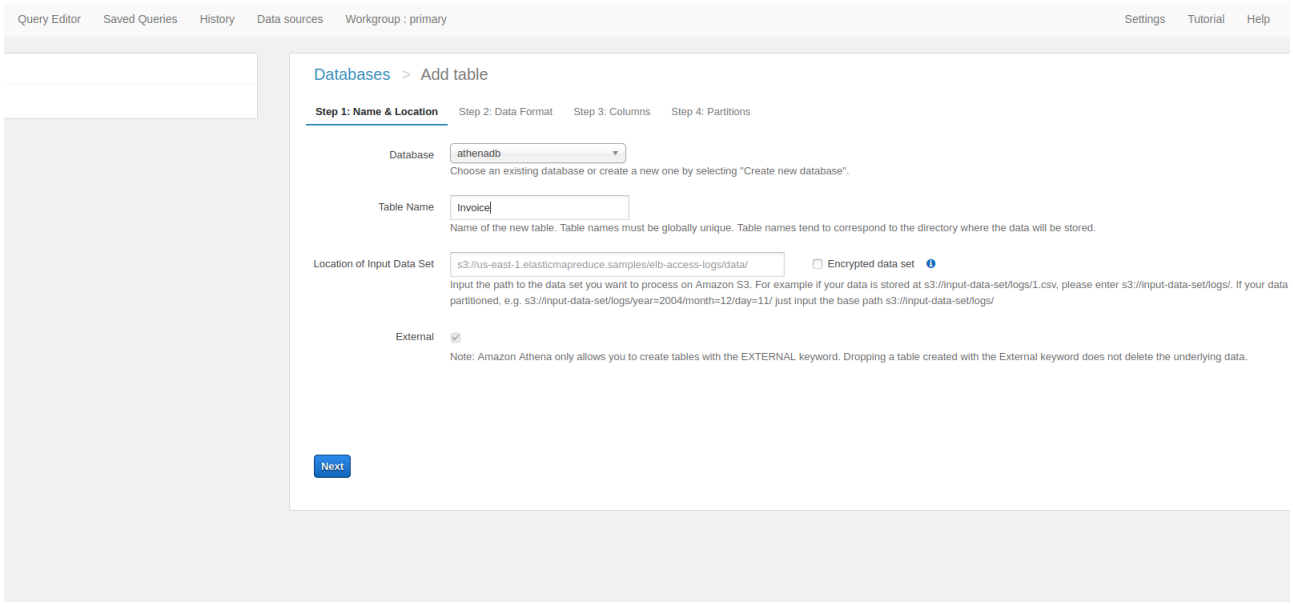
- Items are inserted in Dynamodb table-**Invoice**

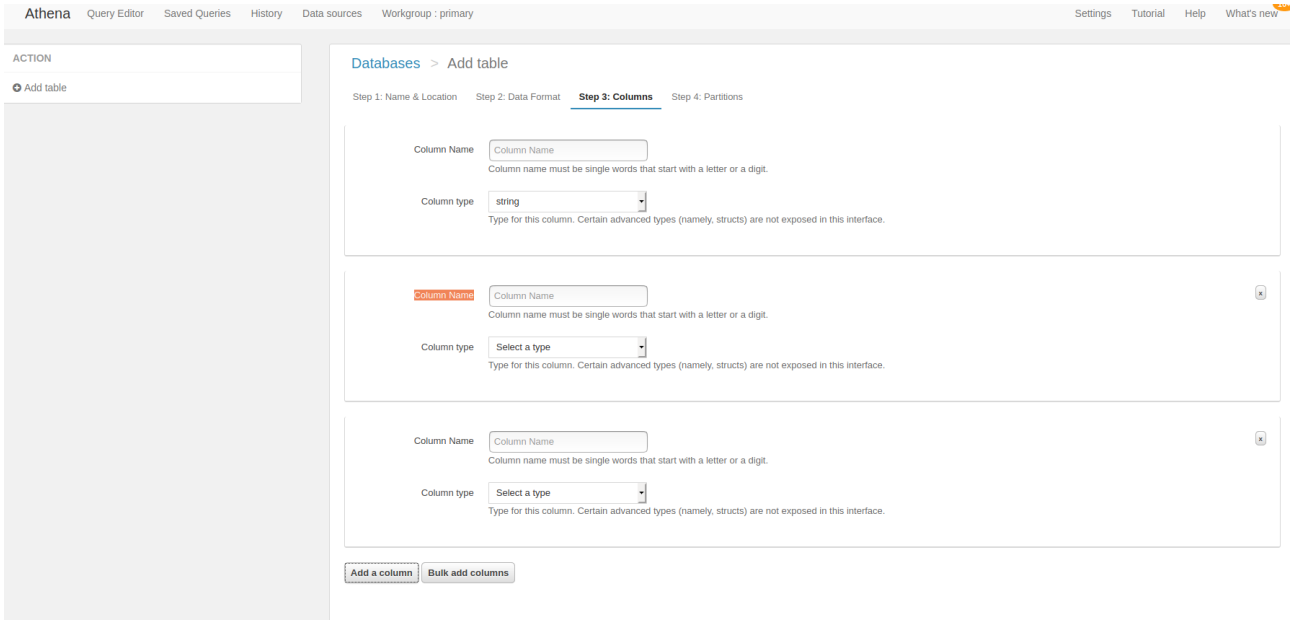


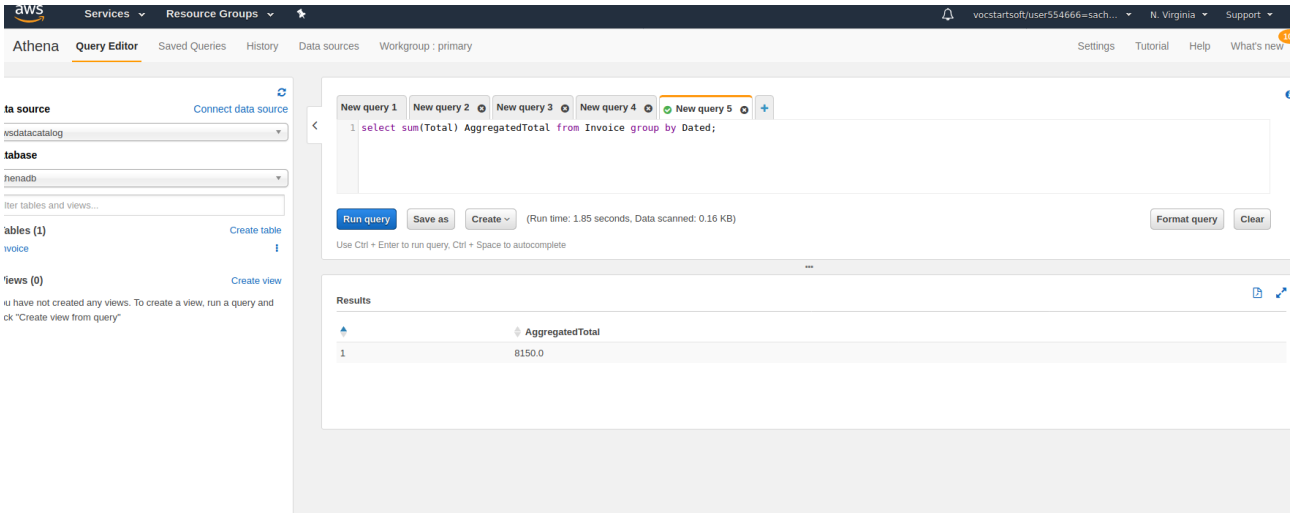
18. Athena

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

Create a table Invoice based on csv file uploaded in S3 bucket.




Athena
Query Editor
Saved Queries
History
Data sources
Workgroup : primary
Settings
Tutorial
Help
What's new
ACTION
Add table
Databases > Add table
Step 1: Name & Location
Step 2: Data Format
Step 3: Columns
Step 4: Partitions
Column Name
Column Name
Column name must be single words that start with a letter or a digit.
Column type
string
Type for this column. Certain advanced types (namely, structs) are not exposed in this interface.
Column Name
Column Name
Column name must be single words that start with a letter or a digit.
Column type
Select a type
Type for this column. Certain advanced types (namely, structs) are not exposed in this interface.
Column Name
Column Name
Column name must be single words that start with a letter or a digit.
Column type
Select a type
Type for this column. Certain advanced types (namely, structs) are not exposed in this interface.
Add a column
Bulk add columns


aws
Services
Resource Groups
vocstartsoft/user554666=sach...
N. Virginia
Support
Athena
Query Editor
Saved Queries
History
Data sources
Workgroup : primary
Settings
Tutorial
Help
What's new
ta source
Connect data source
tabase
Create table
Create view
New query 1
New query 2
New query 3
New query 4
New query 5
1 select sum(Total) AggregatedTotal from Invoice group by Dated;
Run query
Save as
Create
(Run time: 1.85 seconds, Data scanned: 0.16 KB)
Format query
Clear
Use Ctrl + Enter to run query, Ctrl + Space to autocomplete
Results
AggregatedTotal
1
8150.0