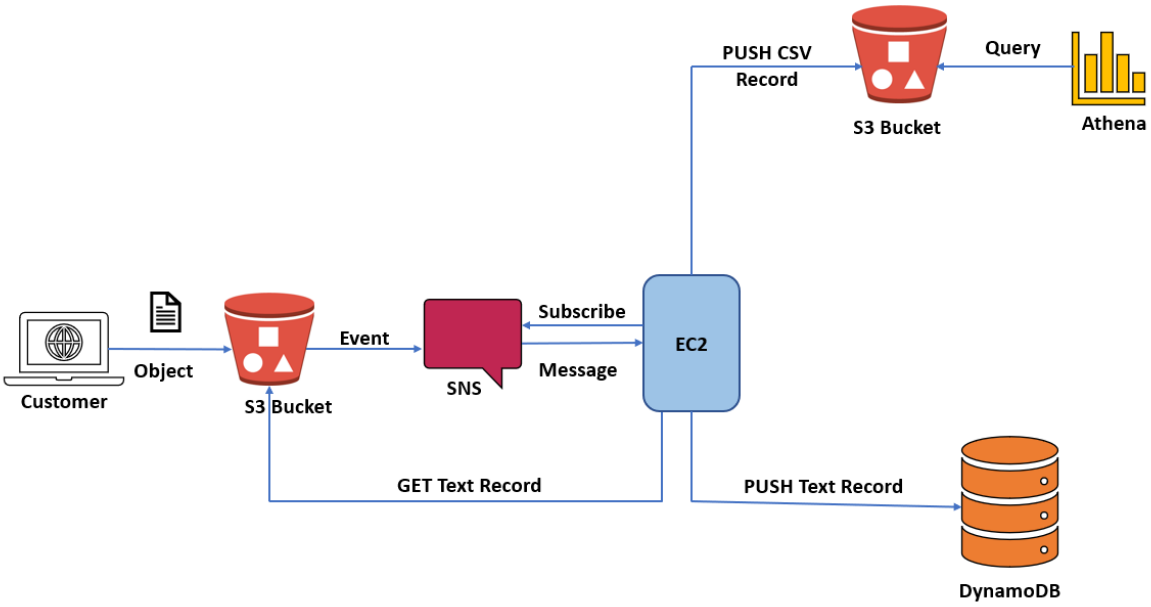
Creating an event triggered business process leveraging multiple managed services from AWS (EC2, S3, SNS, Athena, DynamoDB)

The objective of this project is to create an automated, event based real time process. Data should flow rapidly from the source to the destination. We will use and leverage multiple managed services available from a public cloud platform (AWS in this case) to achieve this.

ARCHITECTURE



IMPLEMTATION

The business process will be developed completely on the AWS public cloud using multiple managed services with some custom code for data transformation.

1. The customer uploads the invoice data to S3 bucket in a text format as per the organizations 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.

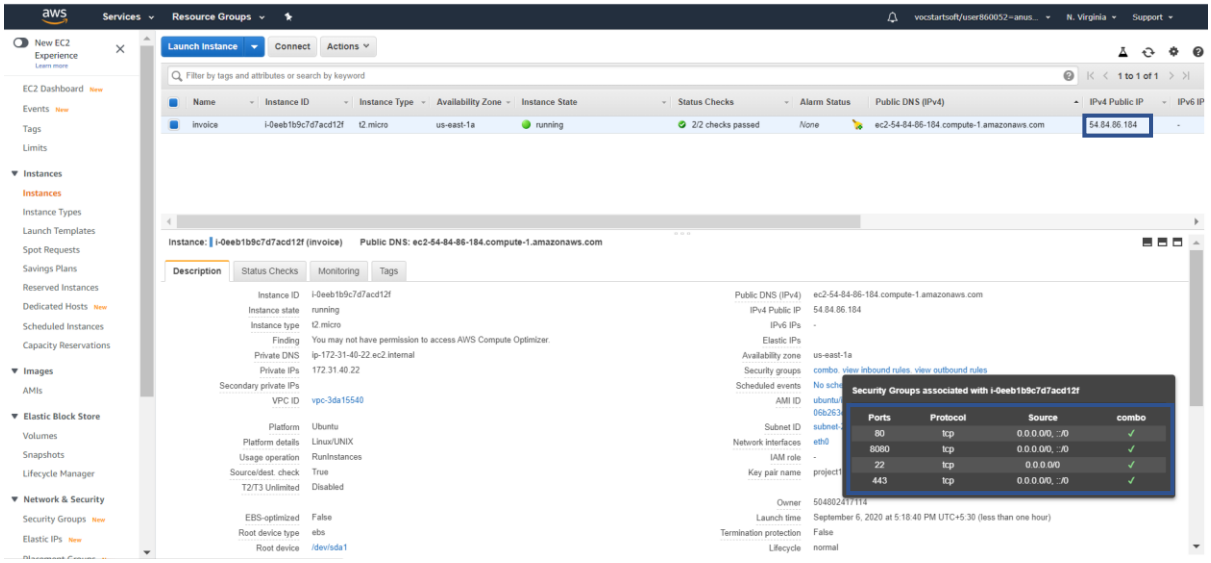
4. The program will use S3 API to read from the bucket, parse the content of the file and create a CSV record along with saving the original record in DynamoDB.

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

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

IMPLEMENTATION

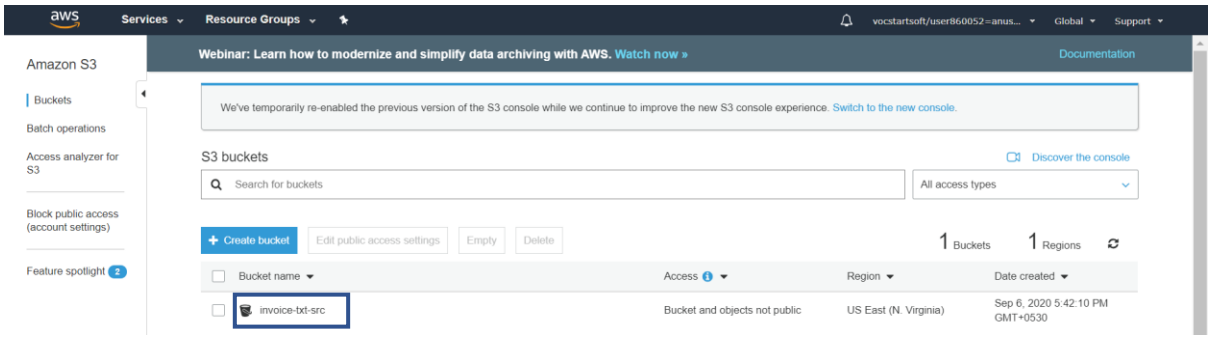
1) Create an Ubuntu EC2 instance and copy the custom code used for data transformation from a TEXT record to a CSV record

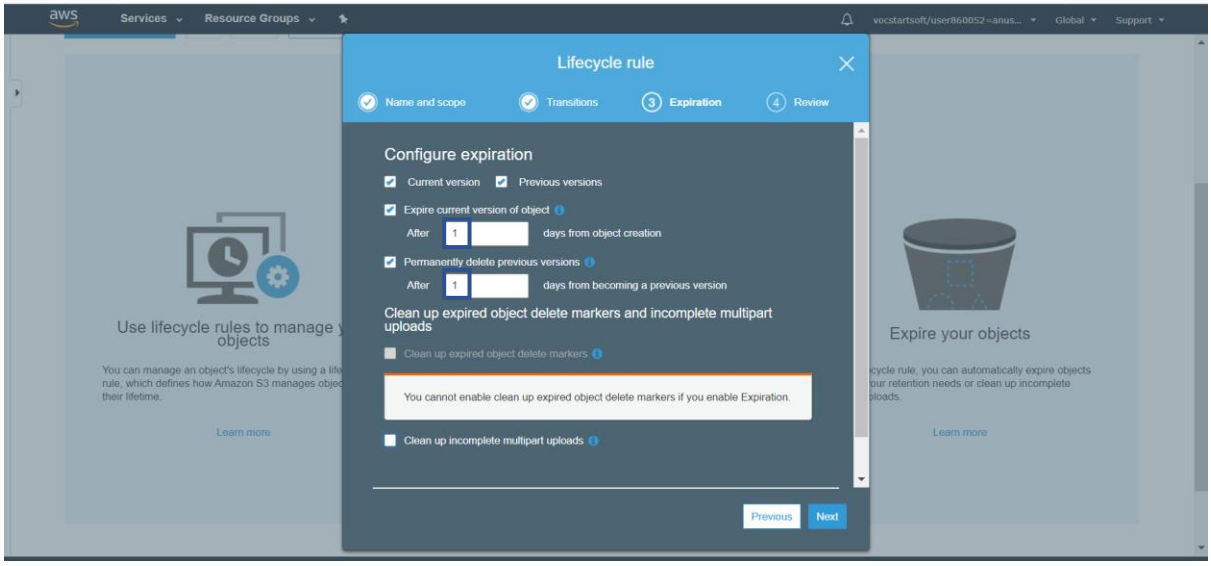


S3 Destination Bucket Name: invoice-csv-dest
DynamoDB Region: us-east-1

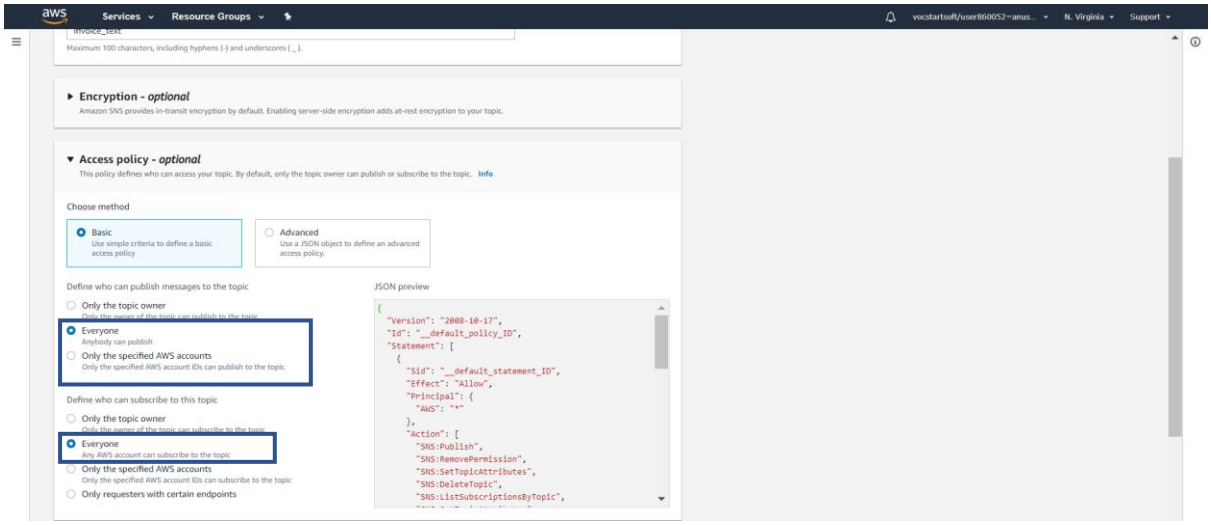
```
(djangoev) ubuntu@ip-172-31-40-22:/opt/docproc-new/api$ cat views.py | egrep -w "us-east-1|invoice-csv-dest"
dynamodb = boto3.resource('dynamodb', region_name='us-east-1')
s3_target_bucket = 'invoice-csv-dest'
(djangoev) ubuntu@ip-172-31-40-22:/opt/docproc-new/api$
```

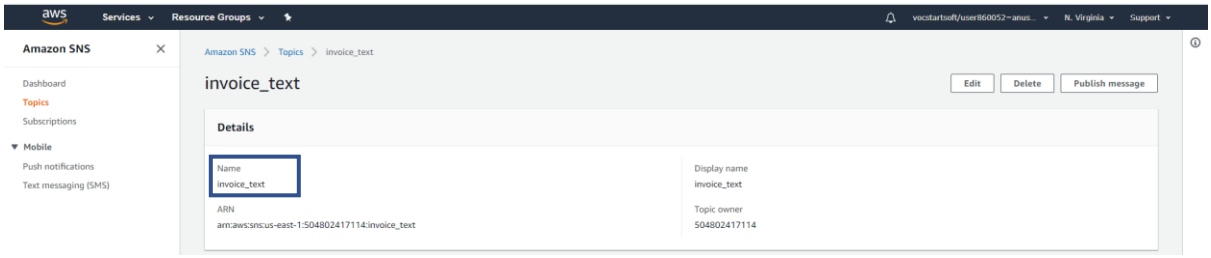
2) Create a source bucket in S3 and add a rule to auto delete any object that is more than 1 day old (24 hours).
S3 Source Bucket: invoice-txt-src



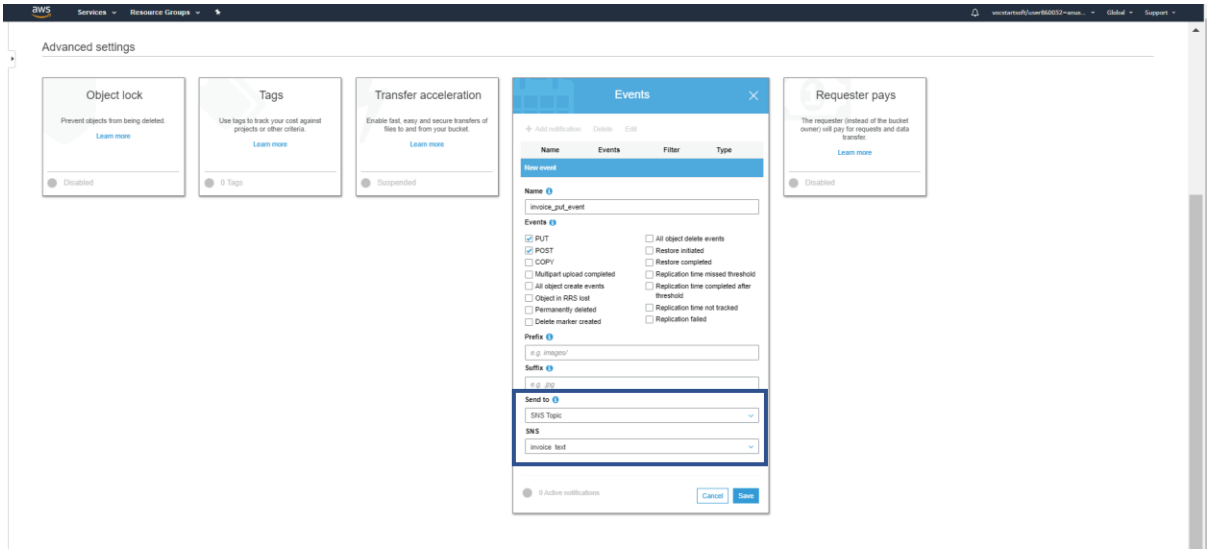


3) Create SNS topic
Topic: invoice_text

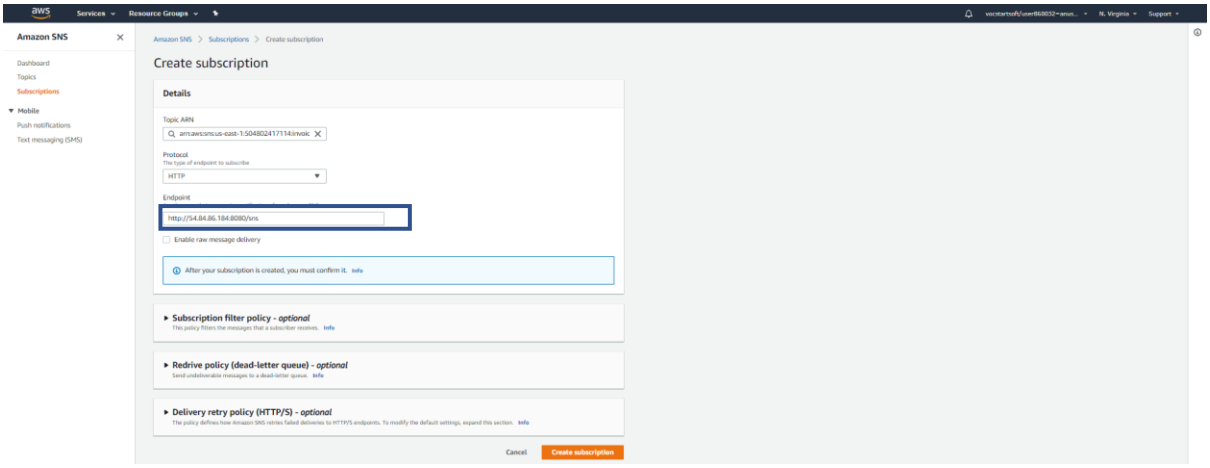


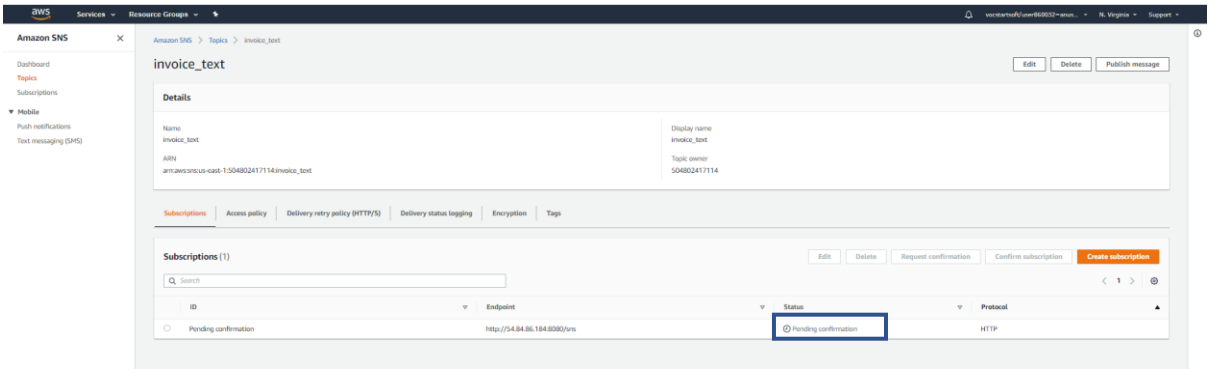


4) Create an event in S3 that will place a message in SNS topic
Event: invoice_put_event

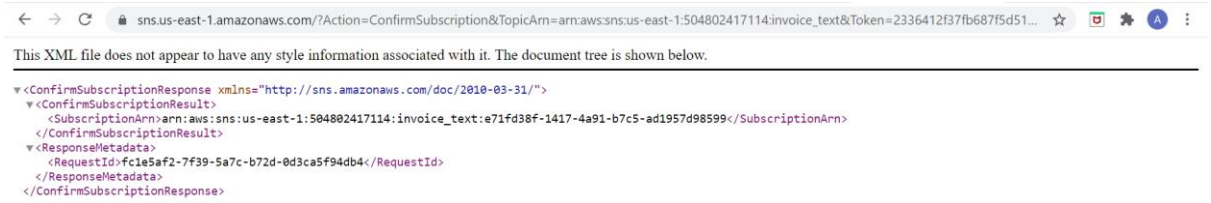


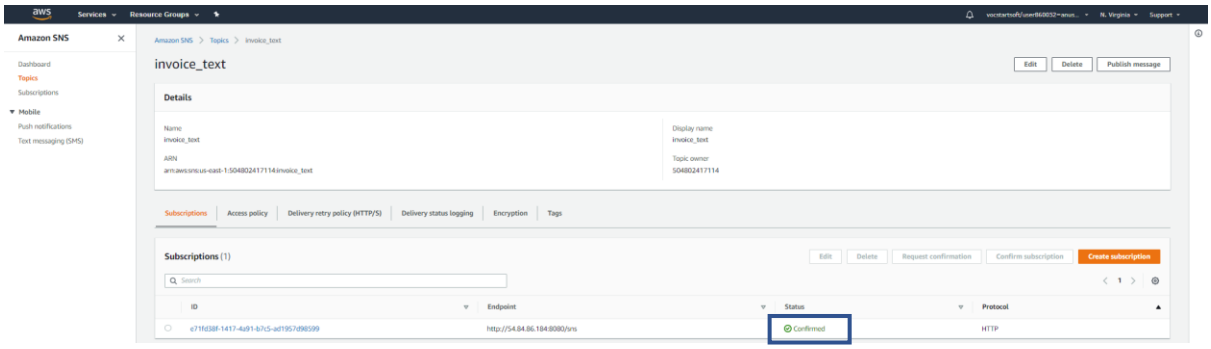
5) EC2 application must subscribe for the SNS topic



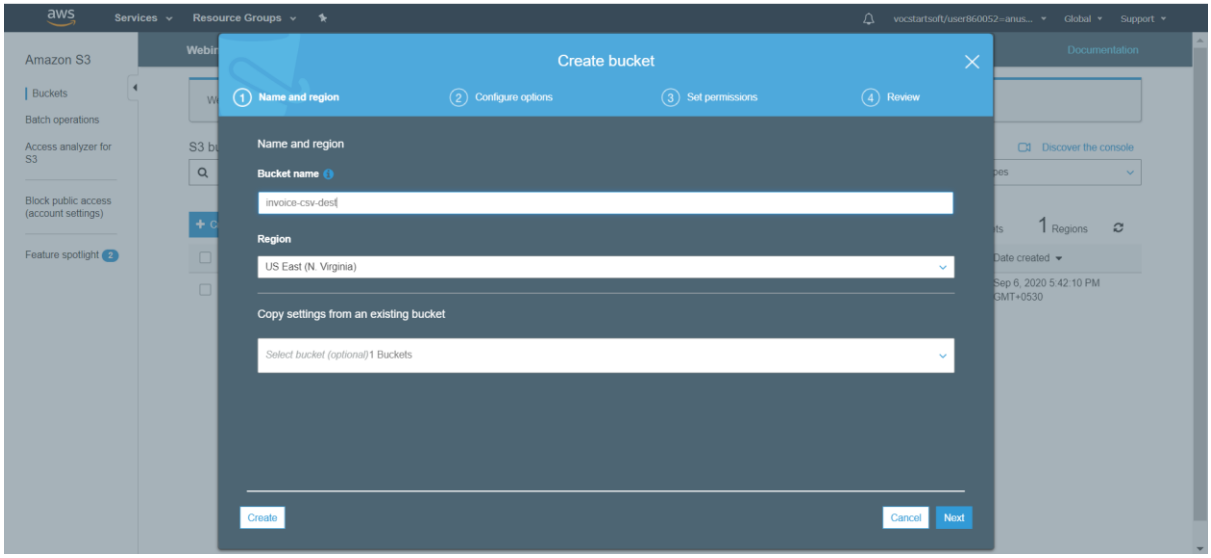



```
Request method = POST
The S3 JSON is  {
  "Type" : "SubscriptionConfirmation",
  "MessageId" : "8053eb4d-8c37-4ef1-a806-cc083676f5df",
  "Token" : "2336412f37fb687f5d51e6e2425f004ae03ffbe2e3ac3afcf1fc8c327c70e2a8e8300bc5860fa8dda3b1c4fd2b39d2841a9daff2d6241c849b57725907cdfcfccc1659fea57eef1dd8d63f71ab47c3afa15220ec4e92289f62166ca96b411602afa3b889234981001a0dd7120186cb37",
  "TopicArn" : "arn:aws:sns:us-east-1:504802417114:invoice_text",
  "Message" : "You have chosen to subscribe to the topic arn:aws:sns:us-east-1:504802417114:invoice_text.\nTo confirm the subscription, visit the SubscribeURL included in this message.",
  "SubscribeURL" : "https://sns.us-east-1.amazonaws.com/?Action=ConfirmSubscription&TopicArn=arn:aws:sns:us-east-1:504802417114:invoice_text&Token=2336412f37fb687f5d51e6e2425f004ae03ffbe2e3ac3afcf1fc8c327c70e2a8e8300bc5860fa8dda3b1c4fd2b39d2841a9daff2d6241c849b57725907cdfcfccc1659fea57eef1dd8d63f71ab47c3afa15220ec4e92289f62166ca96b411602afa3b889234981001a0dd7120186cb37",
  "Timestamp" : "2020-09-06T12:35:13.679Z",
  "SignatureVersion" : "1",
  "Signature" : "YPa7TuYEJgZJAzCeGZAmX6Q9qnLj67miSrzHDQDxKE33ZohxF/Yb6eqp6Lb95SuV9e4NoGjuIMS/6UUwc5faLaTpETapH/SjikAJBm6LZ6eb0m2dR5/WEymHyE6U6iQ+zhV9q/kSB80DsOiZNIk9cp8UxoBHxqSKHdwXV8TOKlGmp/NmPZUzrT86u9s51u0dR+3ievIL3wn3rQlvP/RwUSx9lms6xhsPPU2GU6oreQBjx0EgPVii6+BA+UqoEcjntXCdwVp+j06B8jZb6LbthniN/oAqIy6uQiyvIYHYwVbWHC+nCW0fRAqlDcZm4rQoTo1HmyF1IJO5mftr9lURCg==",
  "SigningCertURL" : "https://sns.us-east-1.amazonaws.com/SimpleNotificationService-a86cb10b4e1f29c941702d737128f7b6.pem"
}
```

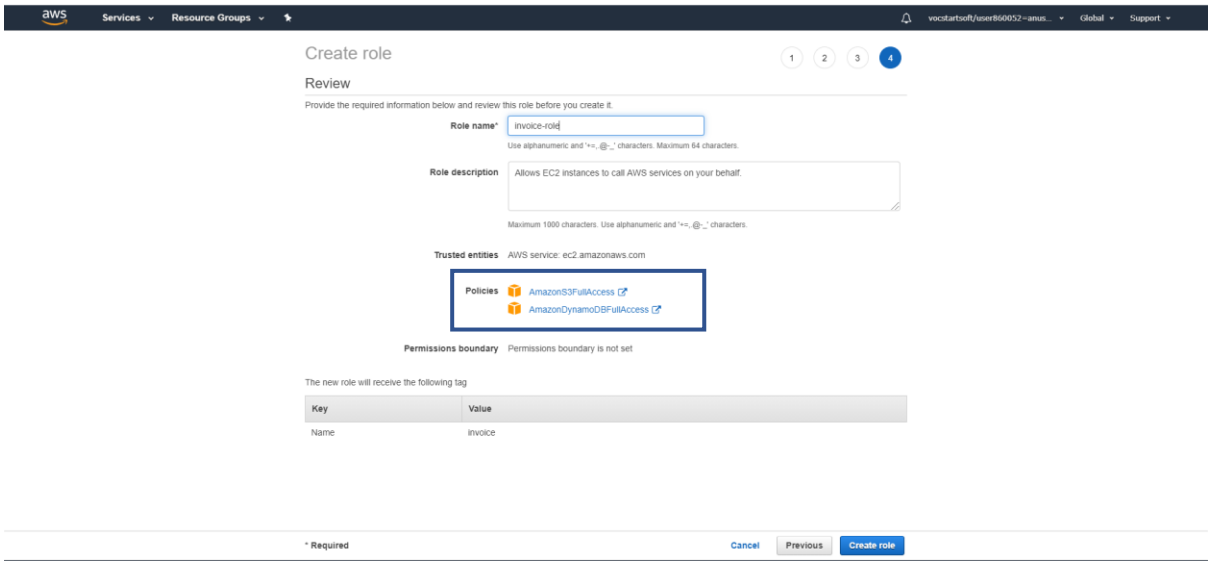


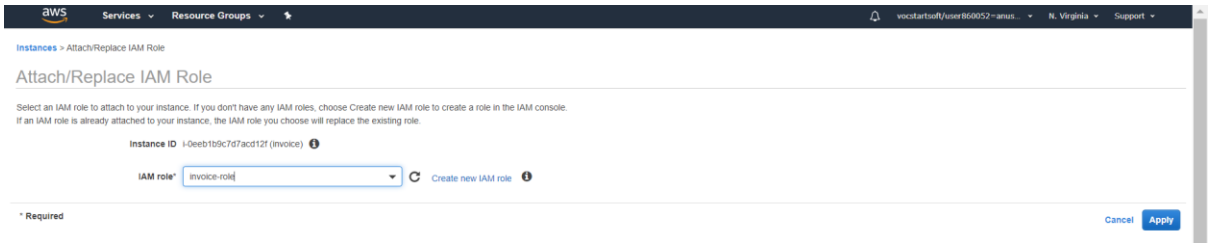


6) Create destination bucket in S3 which will hold the CSV records



7) Create an IAM role to add S3FullAccess and DynamoDBFullAccess, attach the role to the EC2 instance



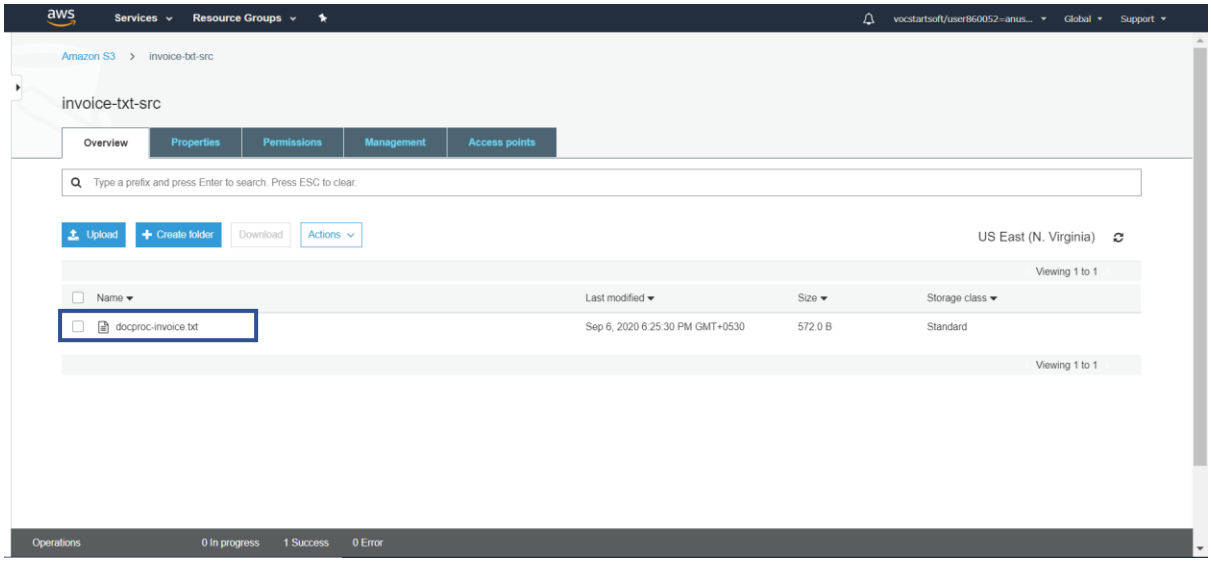


8) Upload a TEXT invoice document to S3 source bucket

docproc-invoice - Notepad

File Edit Format View Help

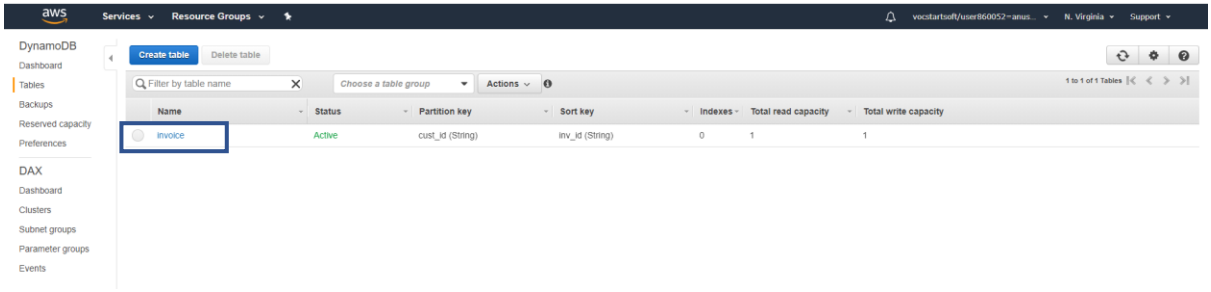
```
**************************************************************
                        Invoice
**************************************************************
Customer-ID: sm-LK0080145
Inv-ID: inv-00001
Dated: Mar 31 2018
From: Space Mechanics Inc.
To: Starfleet Interplanatory HQ
Amount: 7650
SGST: 500
Total: 8150
InWords: Space Rupees Eight thousand one hundred fifty only

Items
1) Matter antimatter fusion controller, 2000, 200
2) External inertial damper, 3100, 180
3) Type X phaser, 500, 35
4) 2MW Fusion reactor, 1200, 85
5) Transporter base dial, 850
```

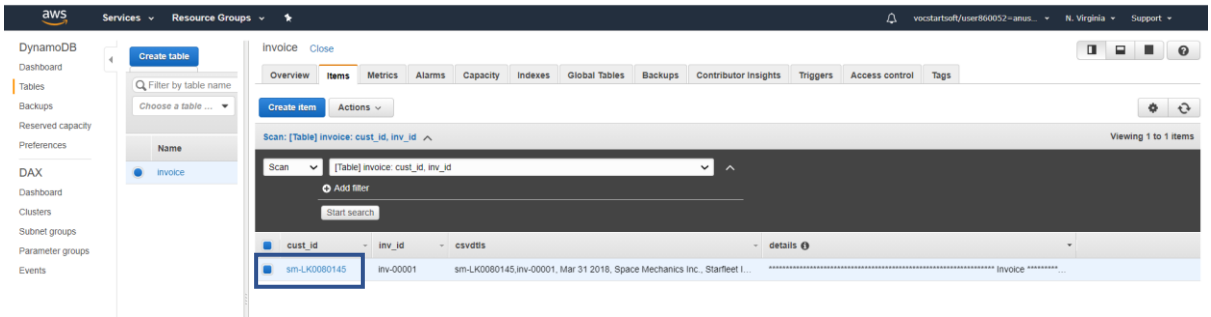


9) Navigate to DynamoDB and see the TEXT invoice record placed

A New table named invoice is created and a record is placed in the table

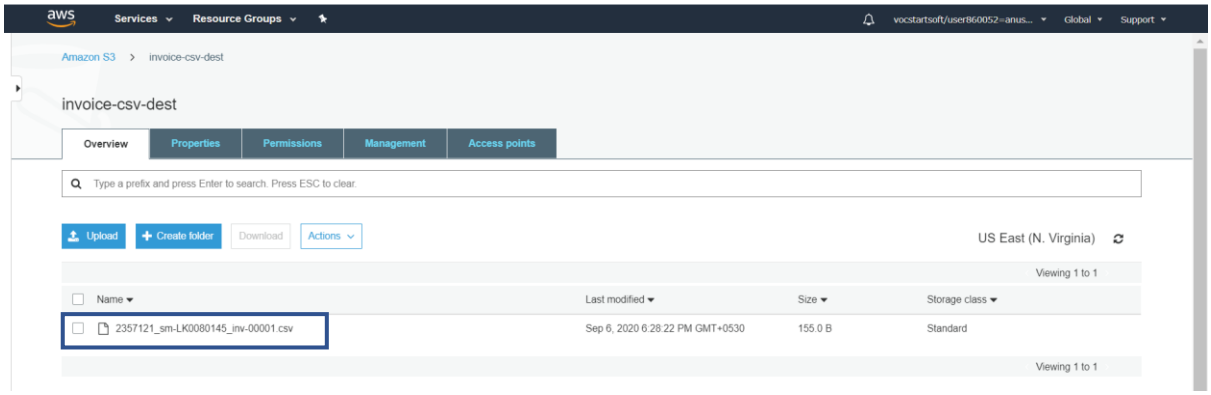
```
**********************************************************************
Creating table invoice
 Skipped due to exception  ResourceInUseException
 Reason  Table already exists: invoice

**********************************************************************
Inserting data in the table
  Done.
Written new s3 file 10297710_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
[06/Sep/2020 12:58:10] "POST /sns HTTP/1.1" 200 43
 DONE

**********************************************************************
```

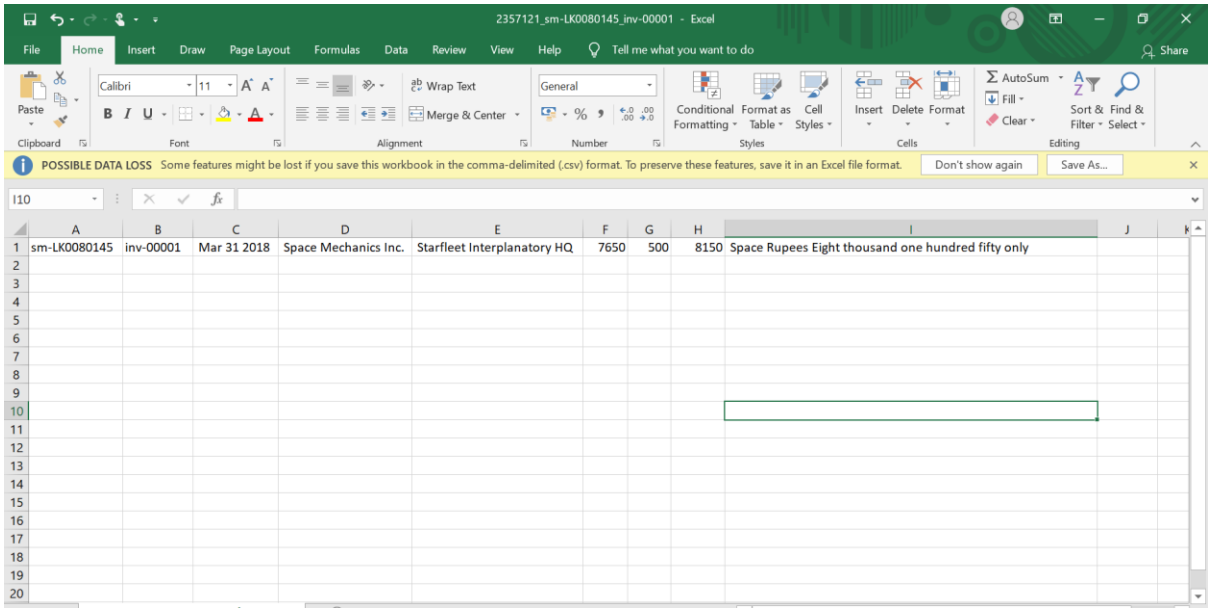




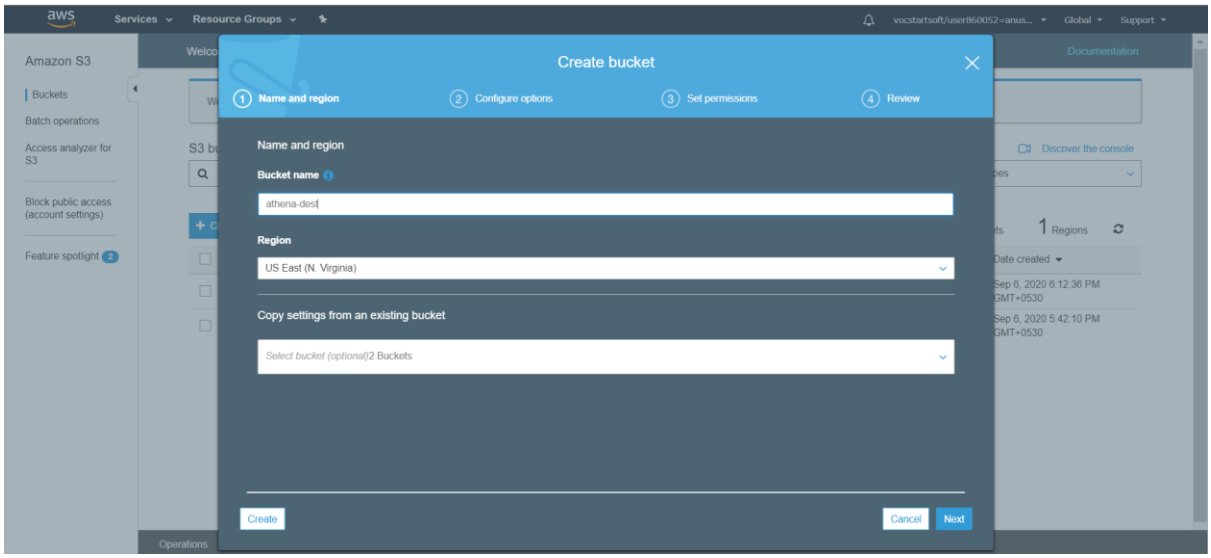
10) See the CSV record created in S3 Destination bucket

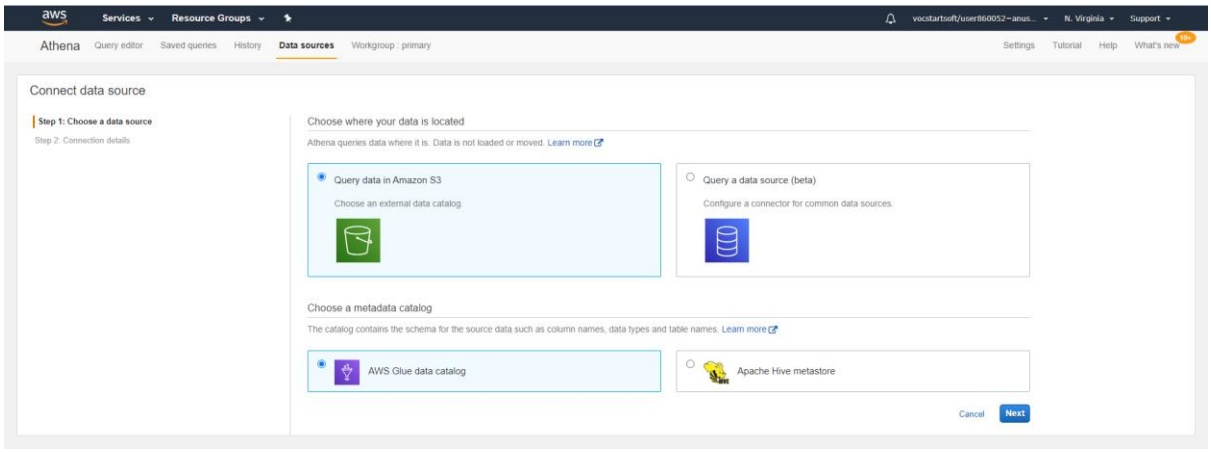
```
Will read the file docproc-invoice.txt from the bucket invoice-txt-src
Line->  ****************************************************************
Line->                              Invoice
Line->  ****************************************************************
Line->  Customer-ID: sm-LK0080145
  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 Interplanatory HQ, 7650, 500, 8150, Space Rupees Eight th
ousand one hundred fifty only
```

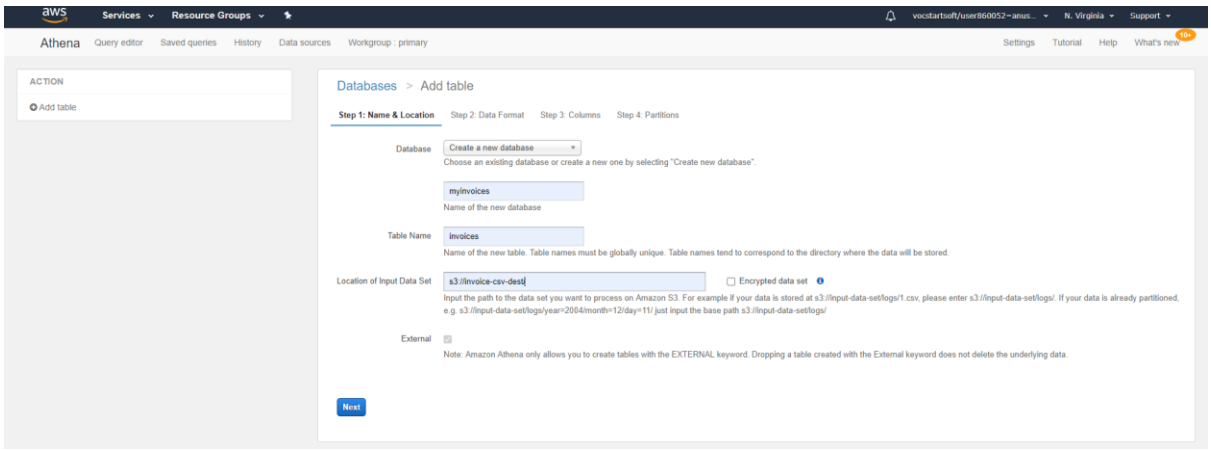


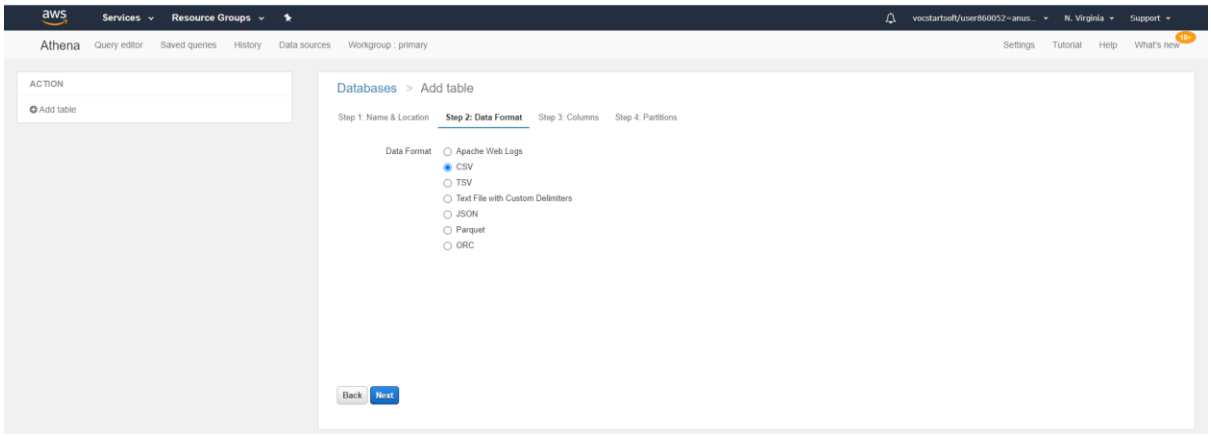


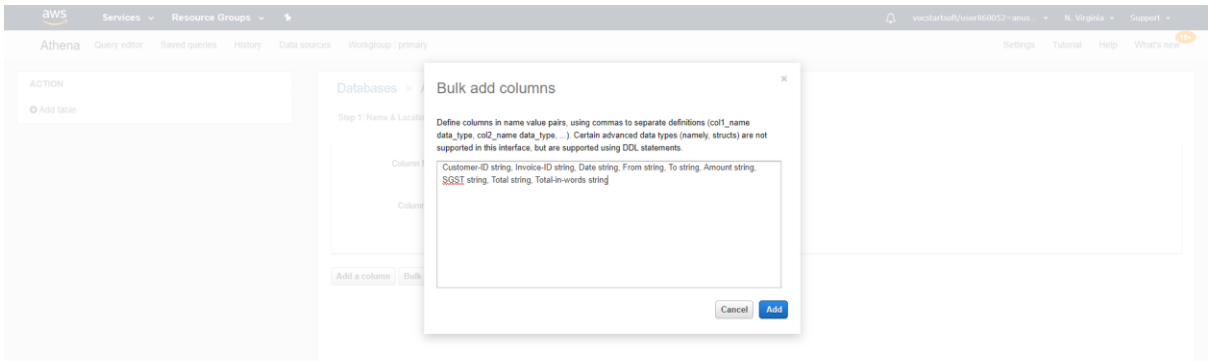
11)Configure Athena to query the CSV record in S3 destination bucket

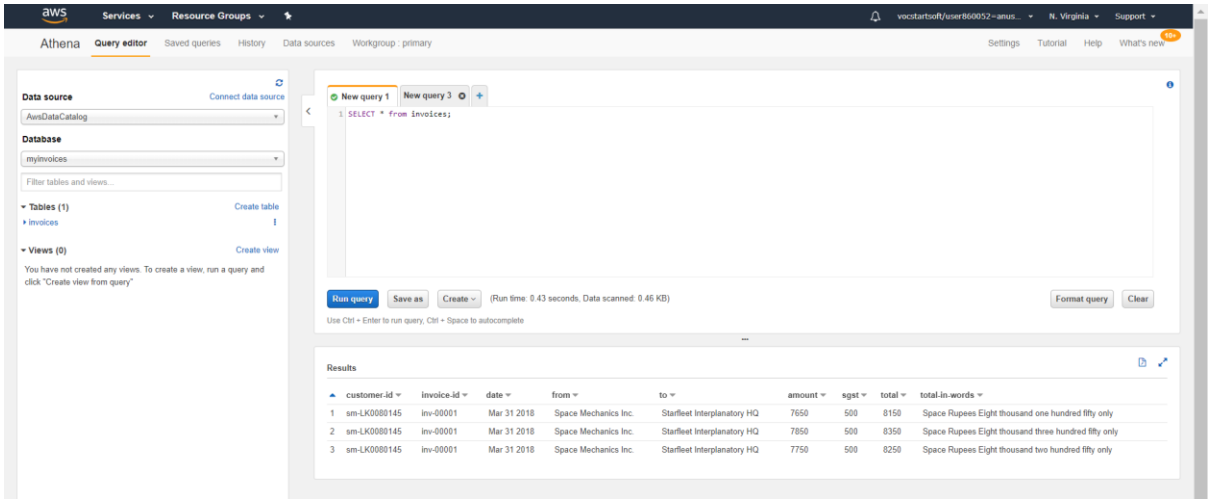

Create bucket
Name and region
Configure options
Set permissions
Review
Name and region
Bucket name
athena-dest
Region
US East (N. Virginia)
Copy settings from an existing bucket
Select bucket (optional) 2 Buckets
Create
Cancel
Next

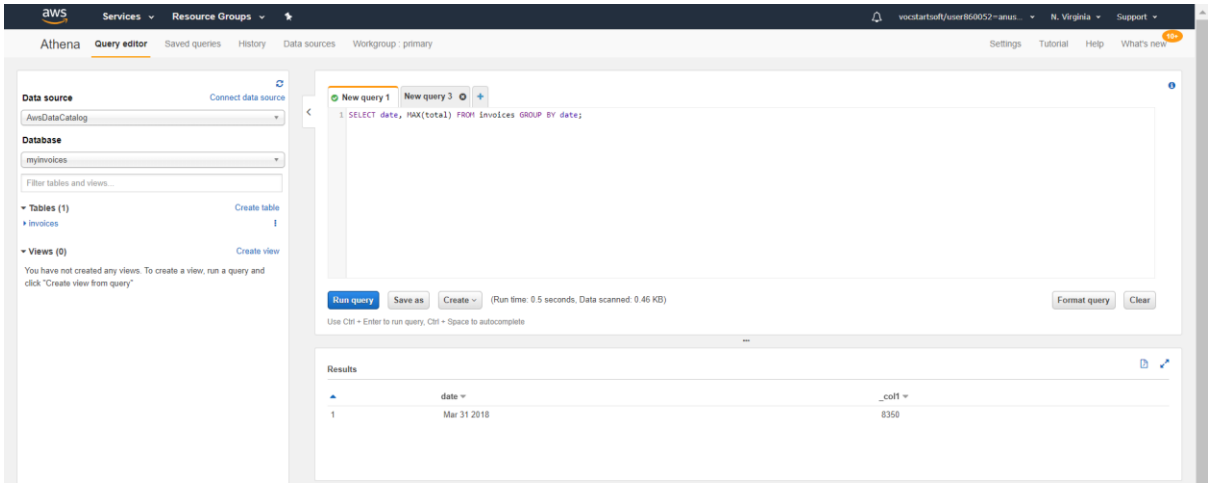

Athena
Query editor
Saved queries
History
Data sources
Workgroup : primary
Connect data source
Step 1: Choose a data source
Step 2: Connection details
Choose where your data is located
Athena queries data where it is. Data is not loaded or moved. Learn more
Query data in Amazon S3
Choose an external data catalog.
Query a data source (beta)
Configure a connector for common data sources.
Choose a metadata catalog
The catalog contains the schema for the source data such as column names, data types and table names. Learn more
AWS Glue data catalog
Apache Hive metastore
Cancel
Next


Athena
Query editor
Saved queries
History
Data sources
Workgroup : primary
ACTION
Add table
Databases > Add table
Step 1: Name & Location
Step 2: Data Format
Step 3: Columns
Step 4: Partitions
Database
Create a new database
Choose an existing database or create a new one by selecting "Create new database".
myinvoices
Name of the new database
Table Name
invoices
Name of the new table. Table names must be globally unique. Table names tend to correspond to the directory where the data will be stored.
Location of Input Data Set
s3://invoice-csv-dest/
Encrypted data set
Input the path to the data set you want to process on Amazon S3. For example if your data is stored at s3://input-data-set/logs/1.csv, please enter s3://input-data-set/logs/. If your data is already partitioned, e.g. s3://input-data-set/logs/year=2004/month=12/day=11/ just input the base path s3://input-data-set/logs/
External
Note: Amazon Athena only allows you to create tables with the EXTERNAL keyword. Dropping a table created with the External keyword does not delete the underlying data.
Next


Services
Resource Groups
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
Data Format
Apache Web Logs
CSV
TSV
Text File with Custom Delimiters
JSON
Parquet
ORC
Back
Next


Bulk add columns
Define columns in name value pairs, using commas to separate definitions (col1_name data_type, col2_name data_type, ...). Certain advanced data types (namely, structs) are not supported in this interface, but are supported using DDL statements.
Customer-ID string, Invoice-ID string, Date string, From string, To string, Amount string, SGST string, Total string, Total-in-words string
Cancel
Add


Services
Resource Groups
Athena
Query editor
Saved queries
History
Data sources
Workgroup : primary
Settings
Tutorial
Help
What's new
Data source
Connect data source
AwsDataCatalog
Database
myinvoices
Filter tables and views...
Tables (1)
Create table
invoices
Views (0)
Create view
You have not created any views. To create a view, run a query and click "Create view from query"
New query 1
New query 3
SELECT * from invoices;
Run query
Save as
Create
(Run time: 0.43 seconds, Data scanned: 0.46 KB)
Format query
Clear
Use Ctrl + Enter to run query, Ctrl + Space to autocomplete
Results
customer-id
invoice-id
date
from
to
amount
sgst
total
total-in-words
1 sm-LK0080145 inv-00001 Mar 31 2018 Space Mechanics Inc. Starfleet Interplanatory HQ 7650 500 8150 Space Rupees Eight thousand one hundred fifty only
2 sm-LK0080145 inv-00001 Mar 31 2018 Space Mechanics Inc. Starfleet Interplanatory HQ 7850 500 8350 Space Rupees Eight thousand three hundred fifty only
3 sm-LK0080145 inv-00001 Mar 31 2018 Space Mechanics Inc. Starfleet Interplanatory HQ 7750 500 8250 Space Rupees Eight thousand two hundred fifty only


aws
Services
Resource Groups
vocstartsoft/user860052=anus...
N. Virginia
Support
Athena
Query editor
Saved queries
History
Data sources
Workgroup : primary
Settings
Tutorial
Help
What's new
Data source
Connect data source
AwsDataCatalog
Database
myinvoices
Filter tables and views...
Tables (1)
Create table
invoices
Views (0)
Create view
You have not created any views. To create a view, run a query and click "Create view from query"
New query 1
New query 3
1 SELECT date, MAX(total) FROM invoices GROUP BY date;
Run query
Save as
Create
(Run time: 0.5 seconds, Data scanned: 0.46 KB)
Format query
Clear
Use Ctrl + Enter to run query, Ctrl + Space to autocomplete
Results
date
_col1
1
Mar 31 2018
8350