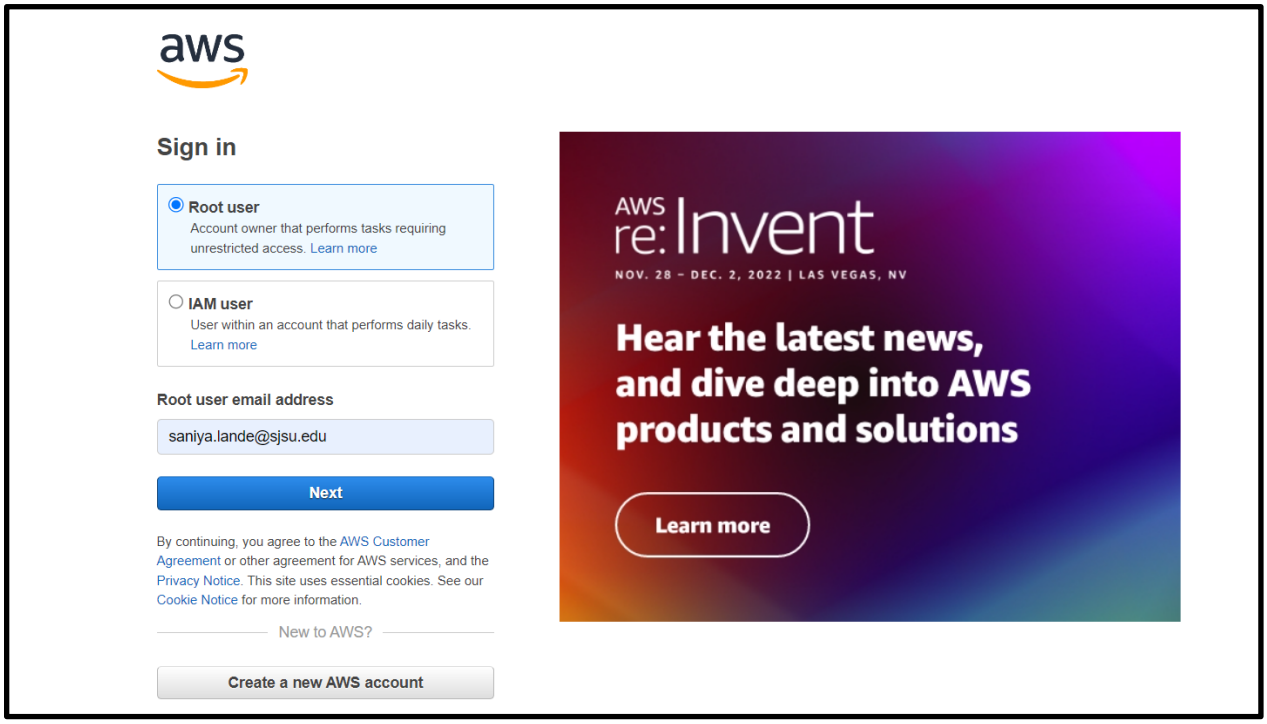
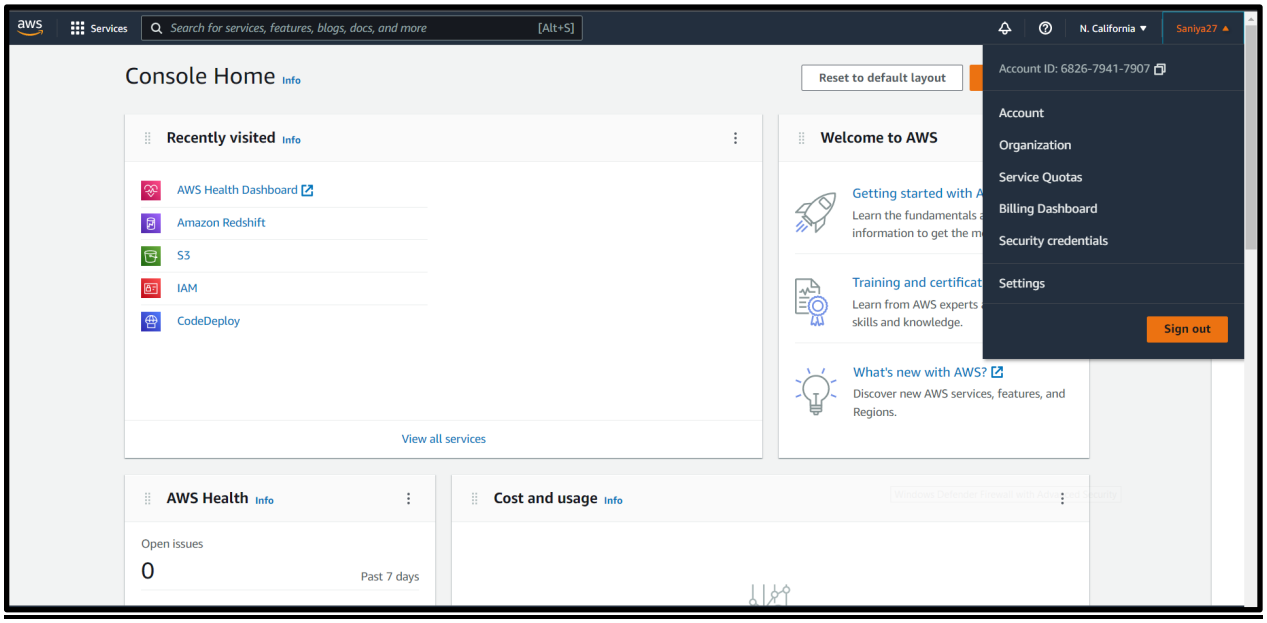
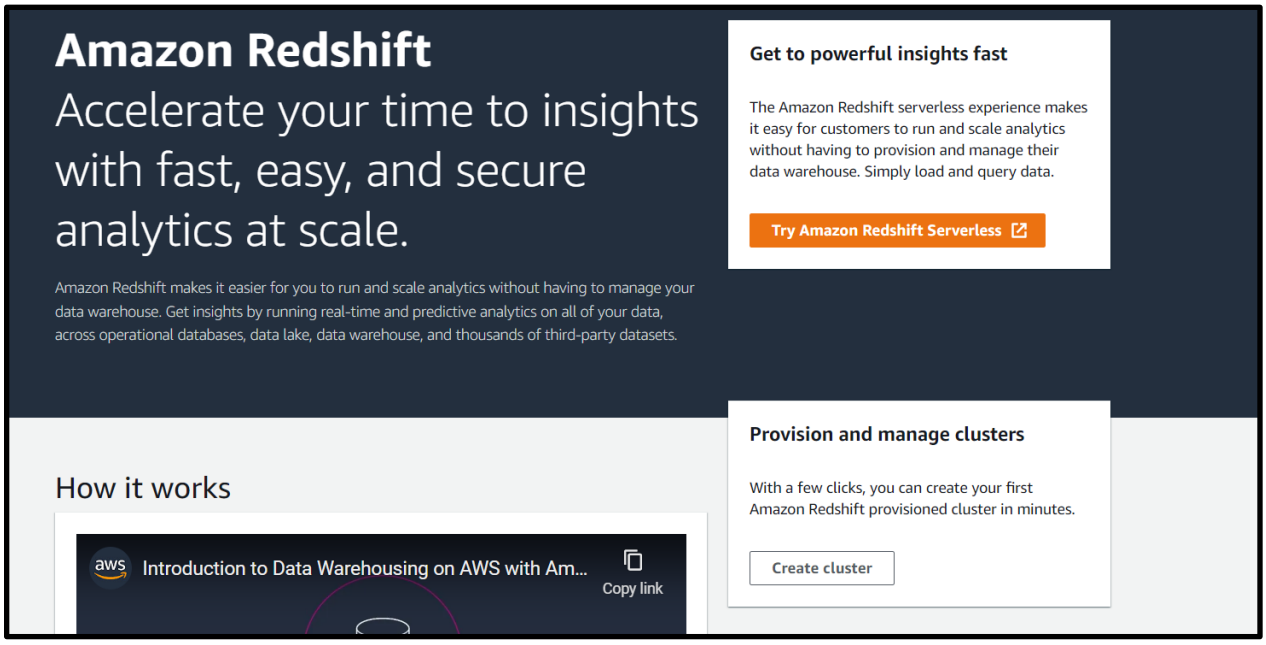
Homework-1

Step1: AWS Console page login - Account details

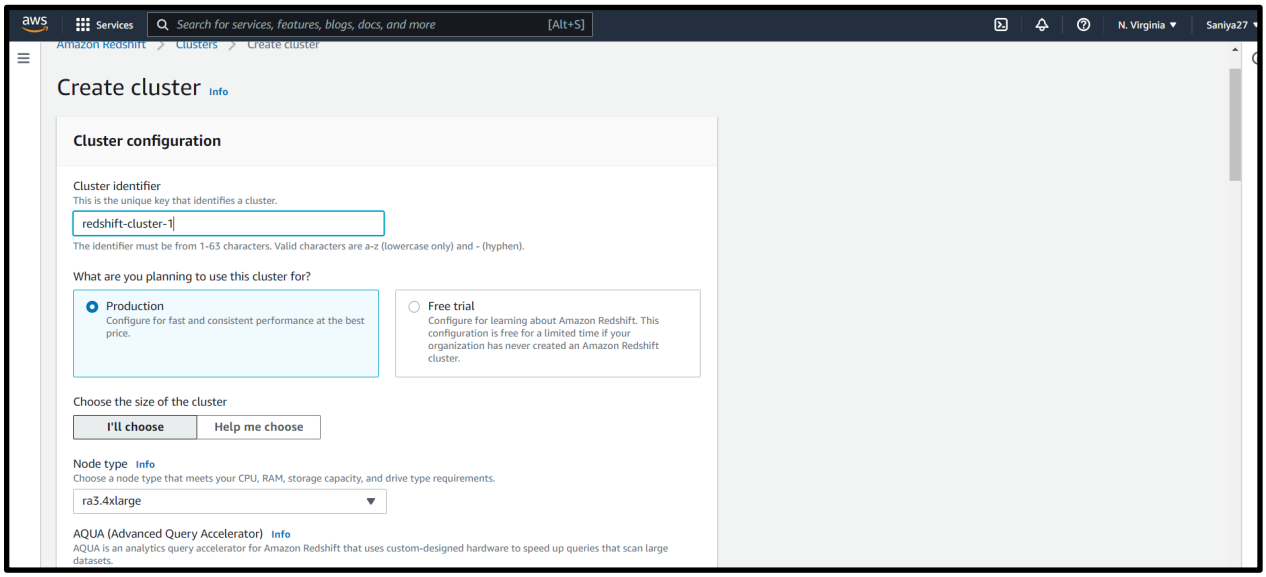




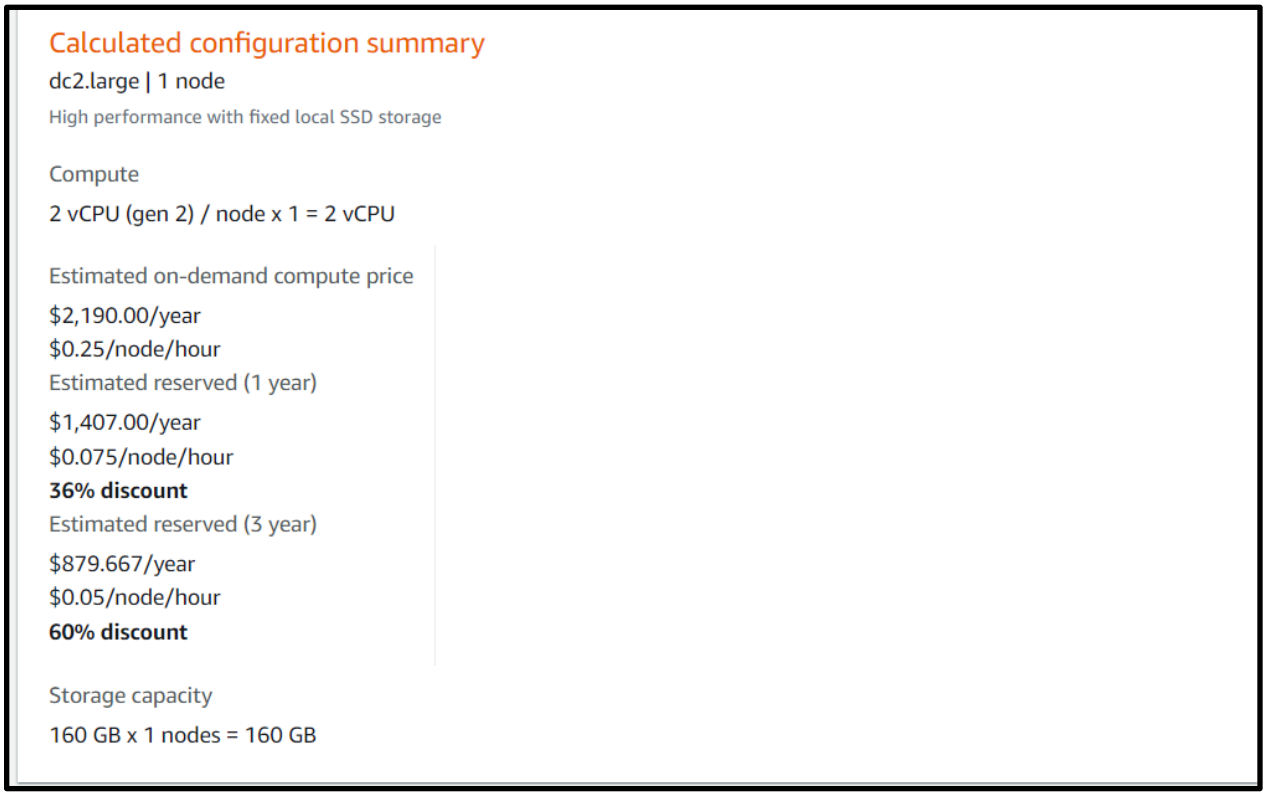
Step 2: Redshift Cluster creation



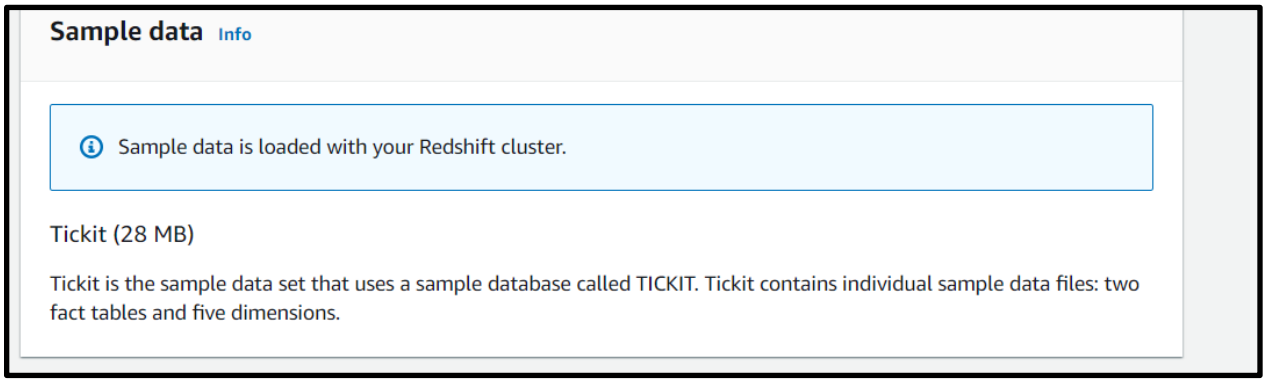
Step 3: Creating cluster in free trial



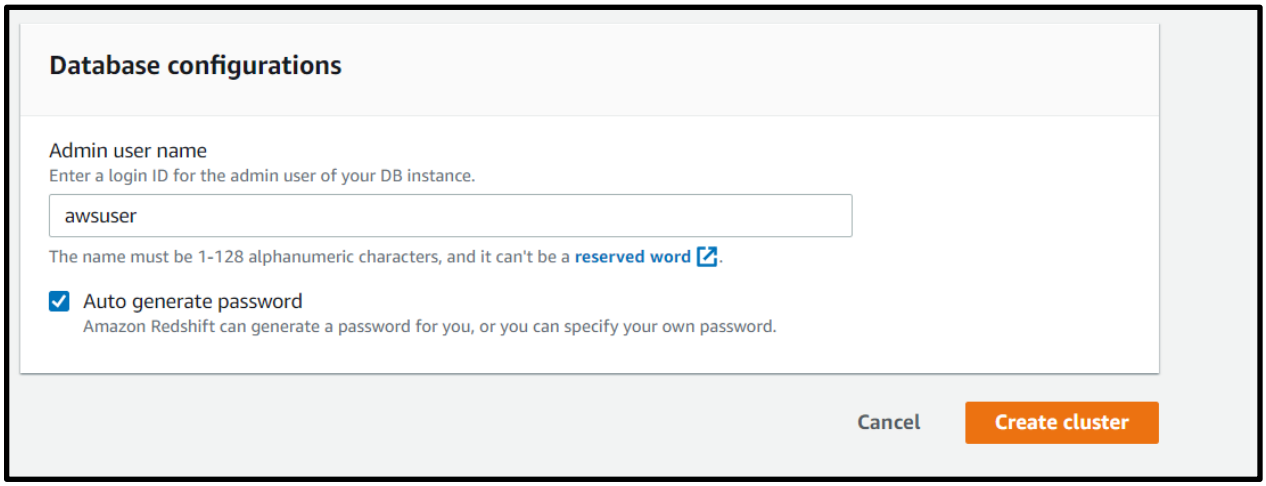
Step 4: Database Configuration



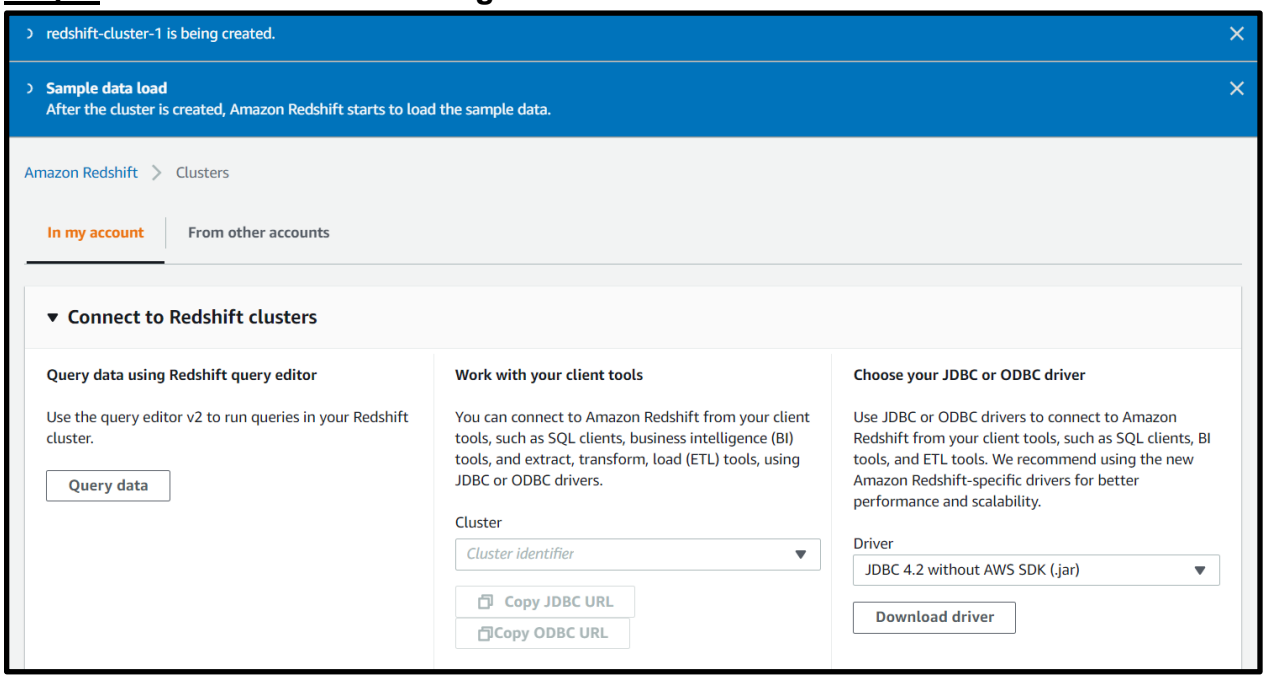
Step 5: Loading sample dataset into Amazon Redshift cluster



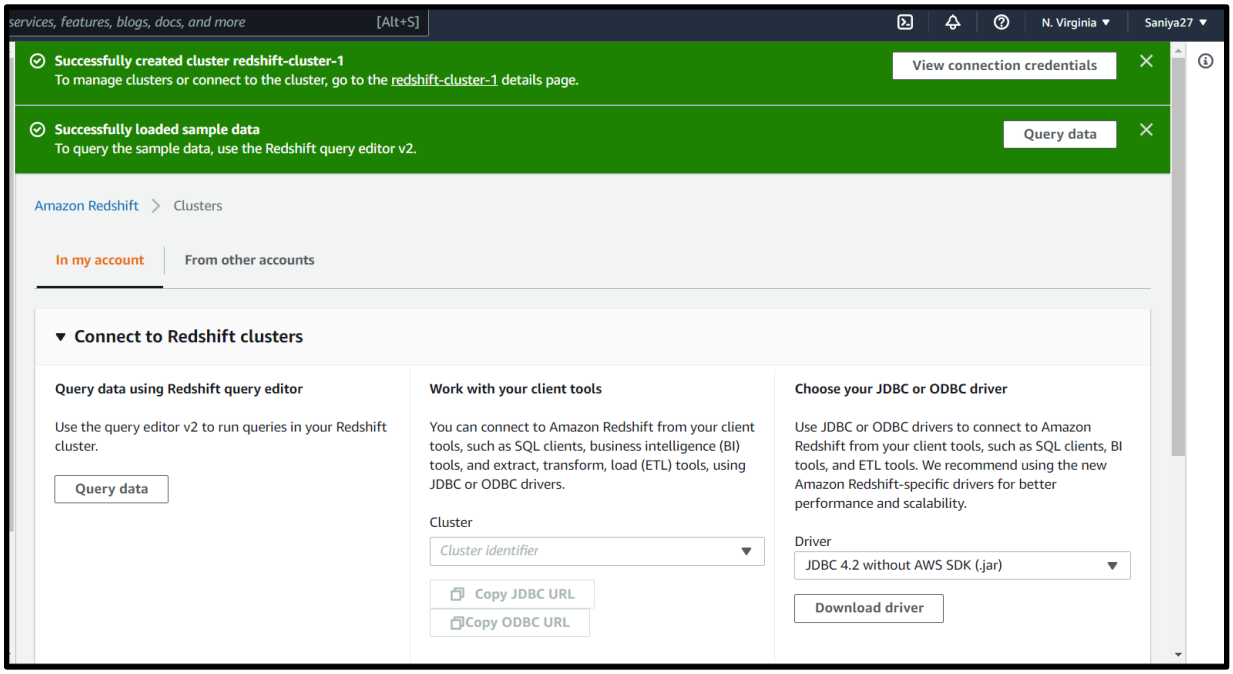
Step 6: Creating username and password for the redshift-cluster-1



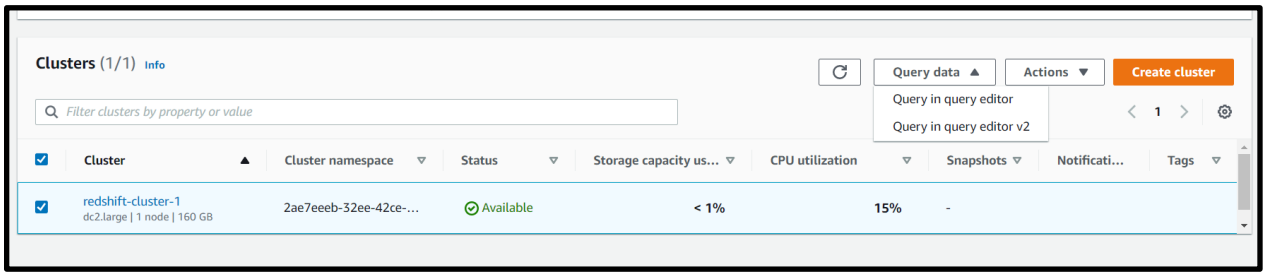
Step 7: Redshift-cluster-1 being created



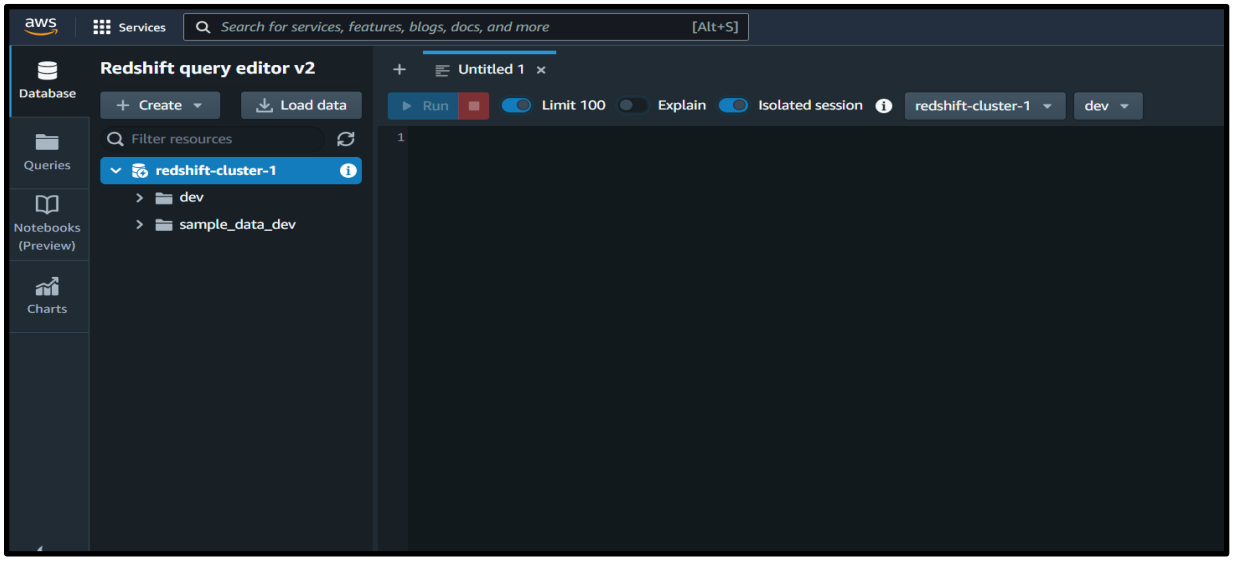
Step 8 : Cluster named Redshift-cluster-1 created and sample data successfully loaded



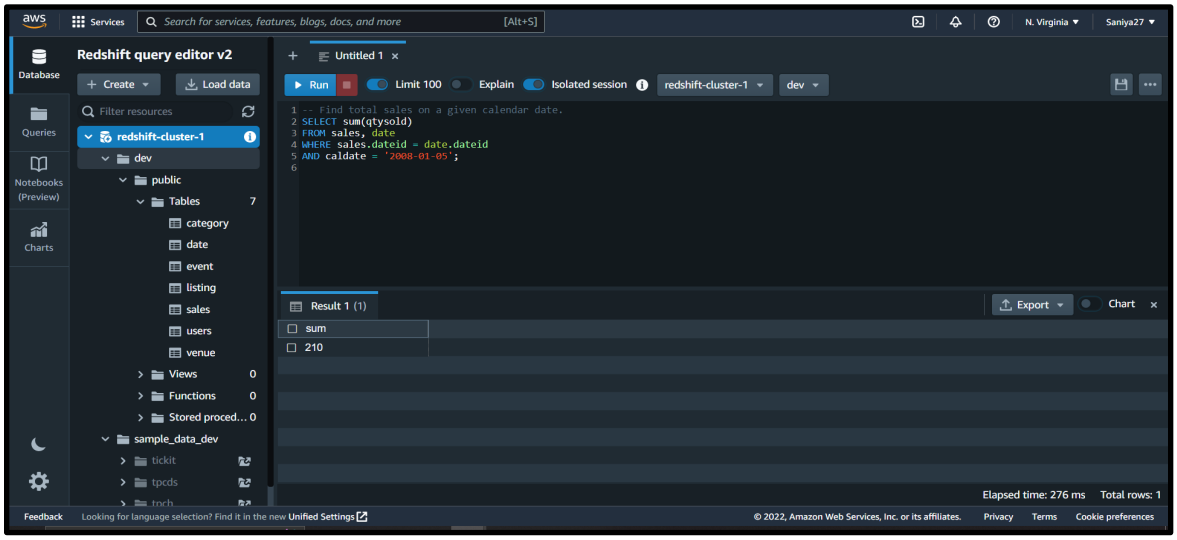
Step 9: Selecting query editor v2 for querying data in redshift-cluster-1

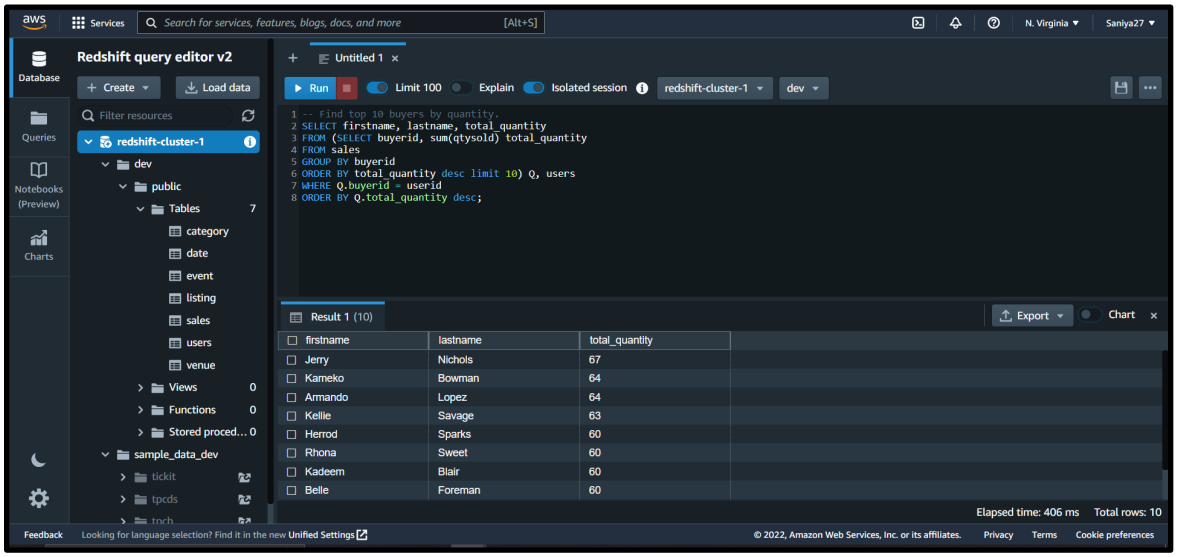


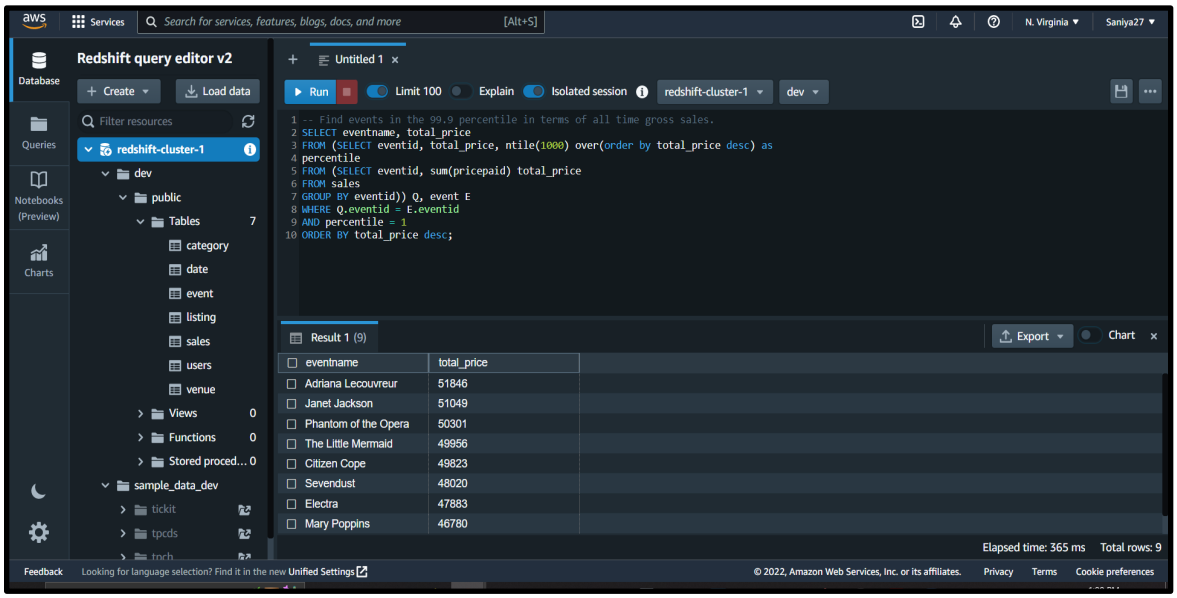
Step 10: Connecting to sample database from query editor v2



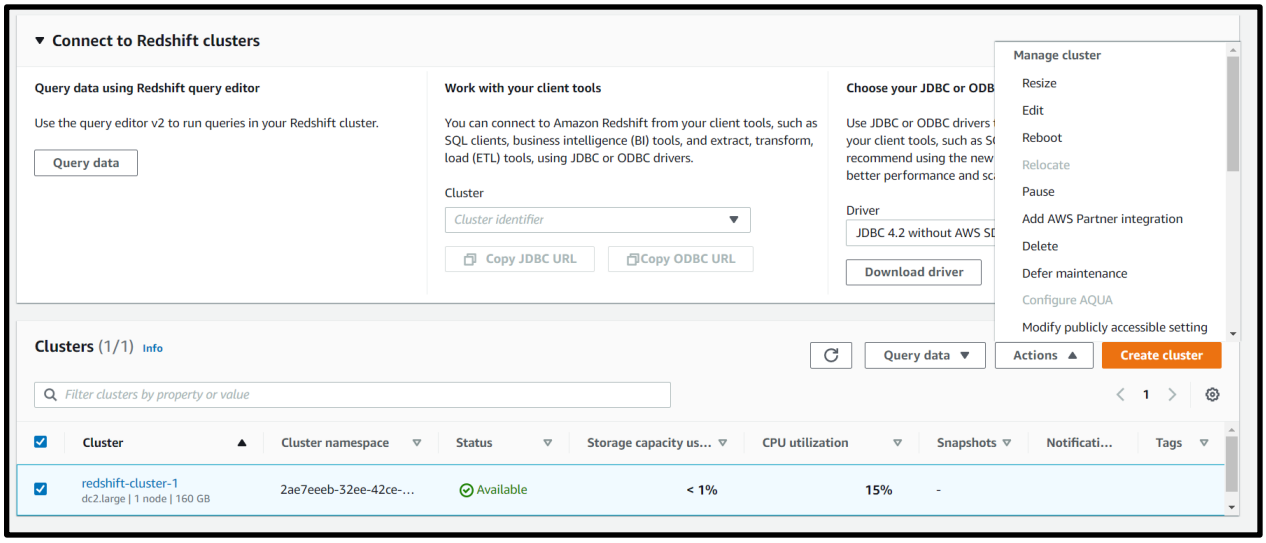
Step 11:Trying example Queries in query editor v2 as shown below



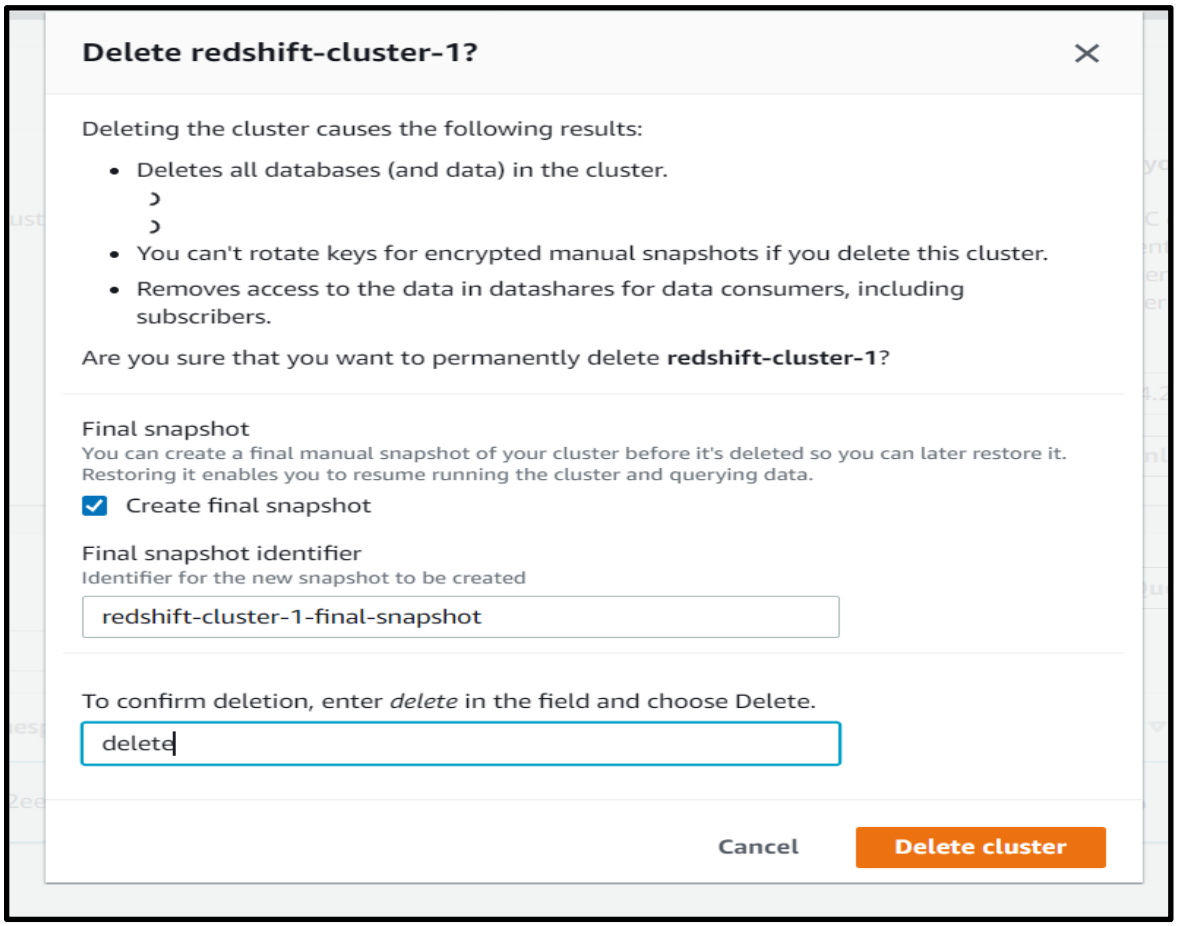




Step 12: Selecting the redshift-cluster-1 to delete the cluster



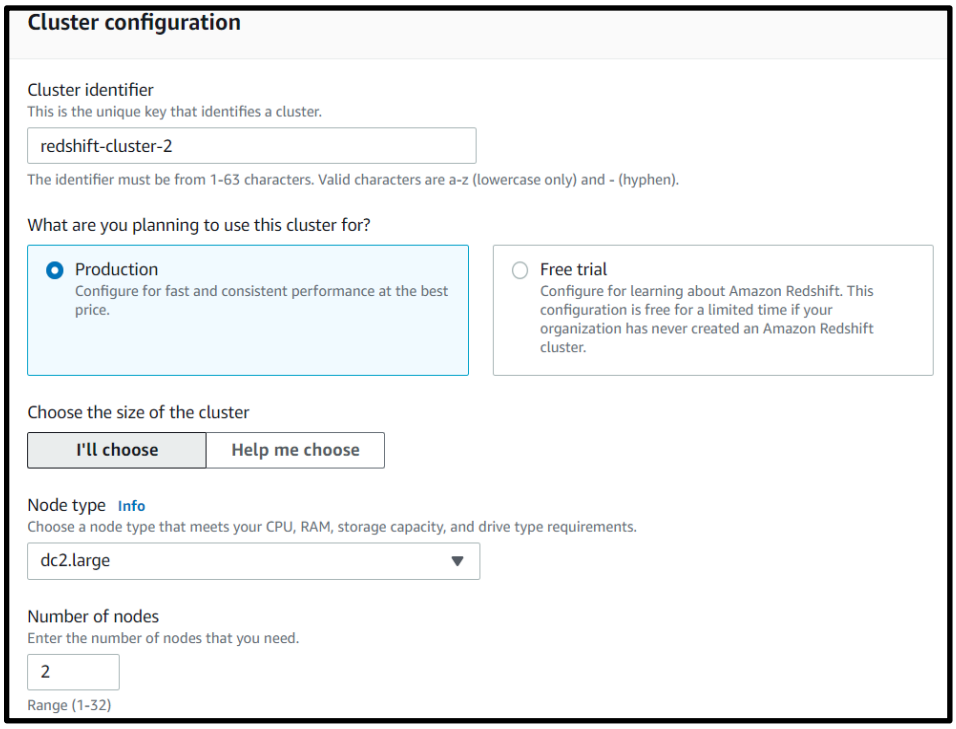
Step 13: Deleting redshift-cluster-1



Using Production tier - Bringing our own data into Amazon Redshift

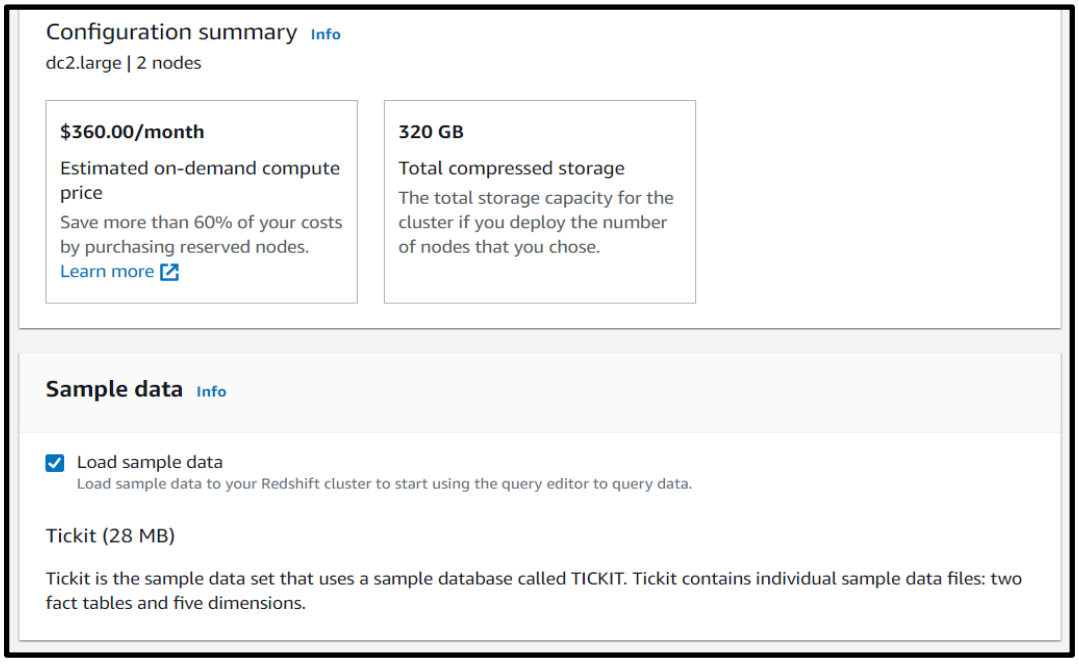
Step 1: Production Cluster creation

Selected dc2.large Node type and 2 number of nodes

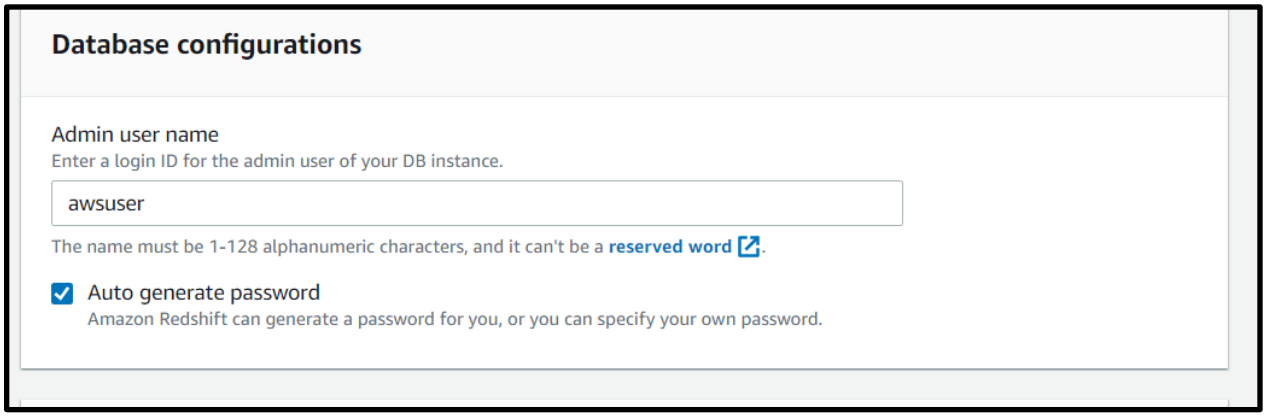


We have directly used query editor v2 to run queries on databases. After creating our cluster, we immediately ran queries using the Amazon Redshift console.

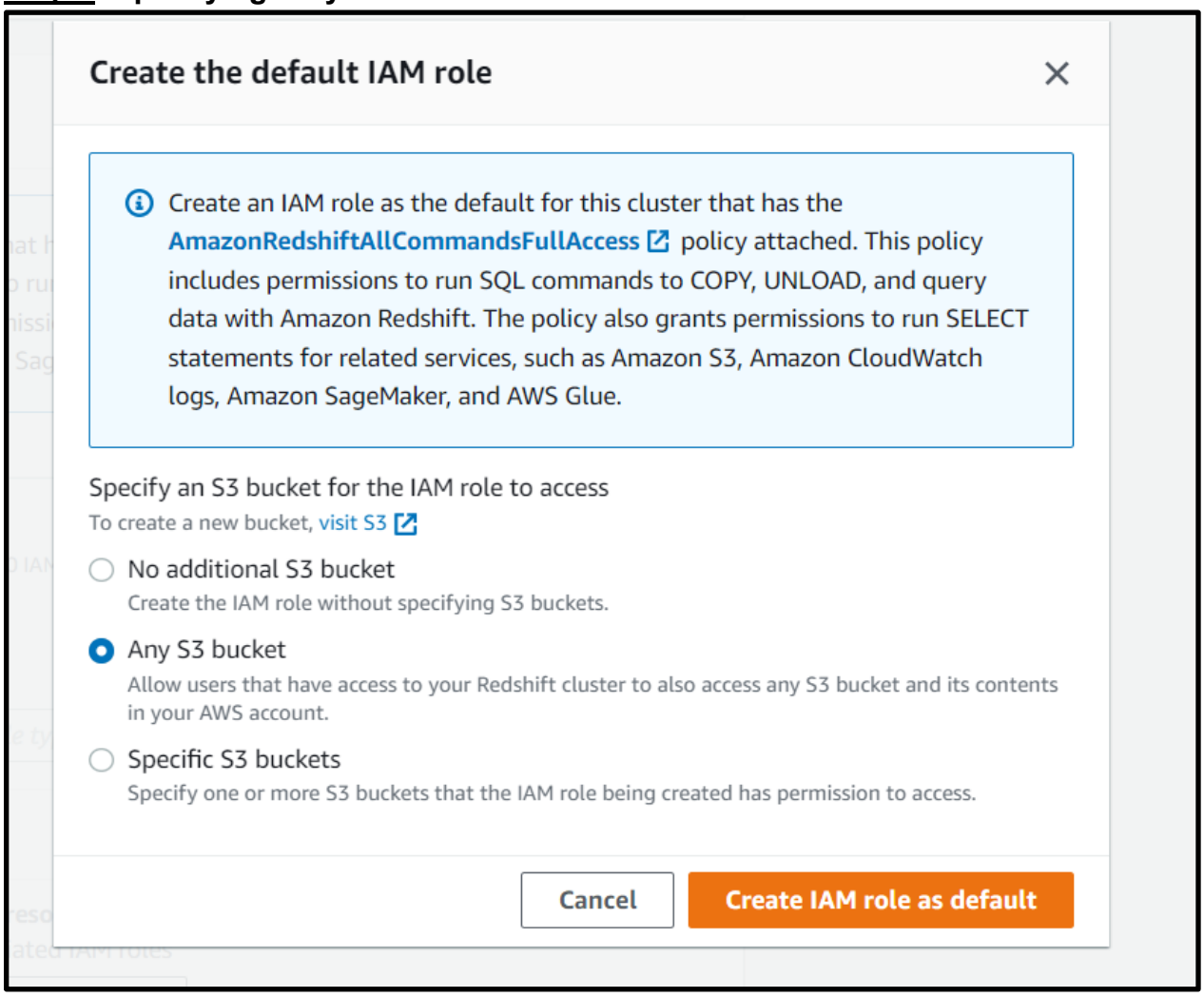
Step 2: Selected to load sample data to redshift-cluster-2

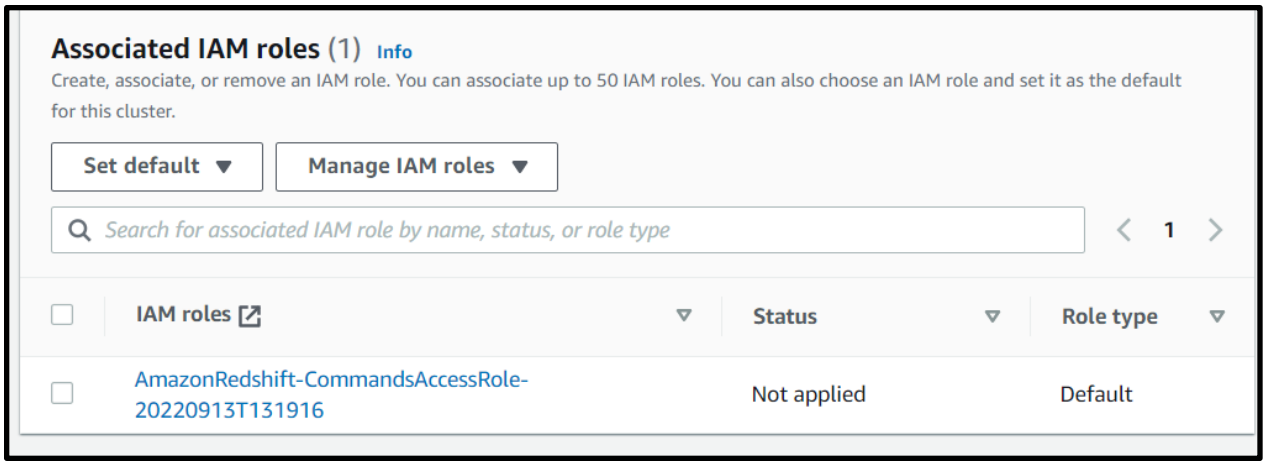


Step 3 : Database Configuration

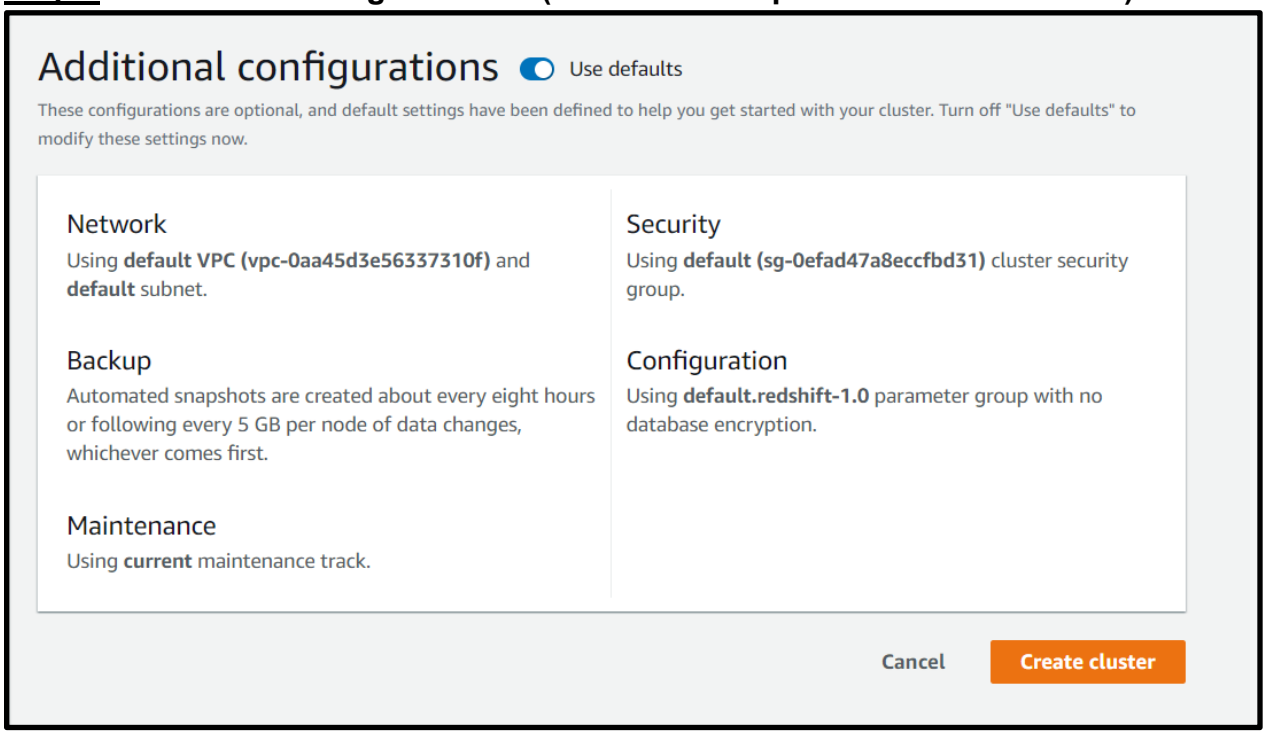


Step 4: Specifying 'Any S3 bucket' to create the default IAM role

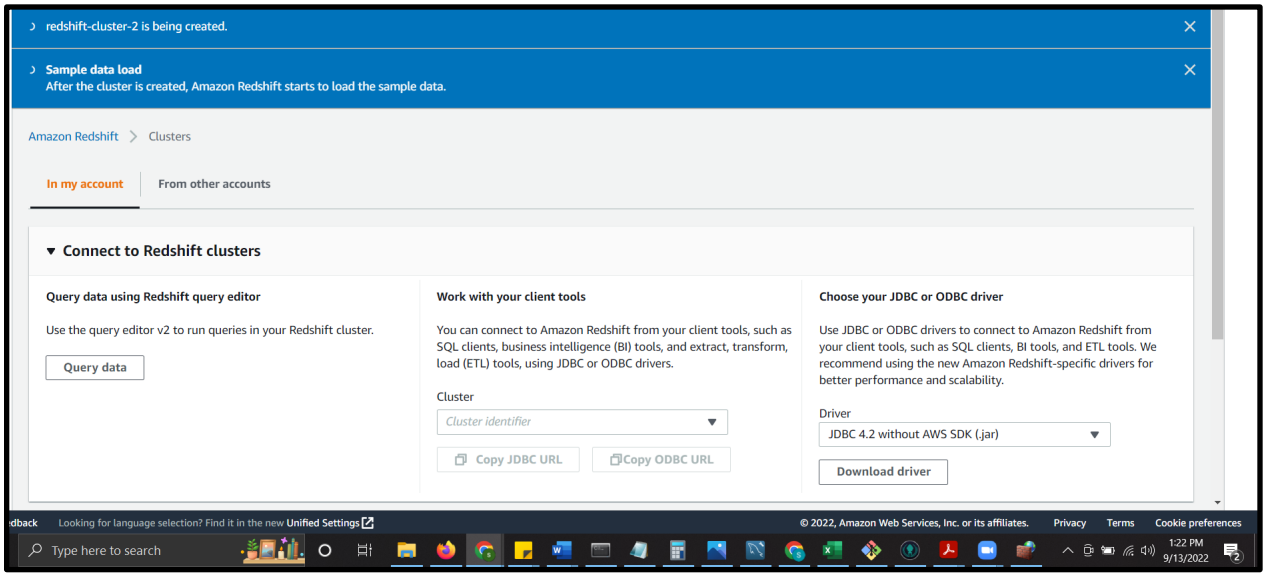




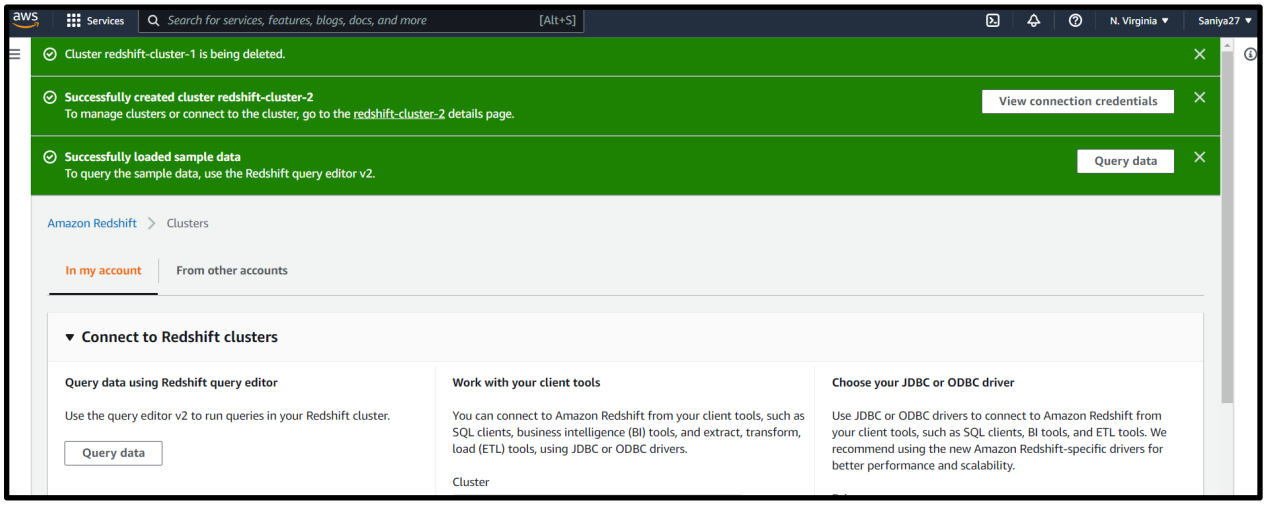
Step 5: Additional Configurations - (Used default options to create cluster)



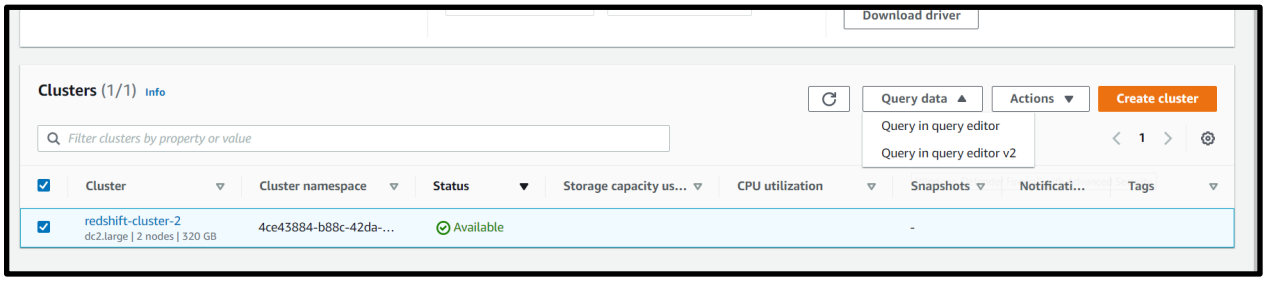
Step 6: redshift-cluster-2 being created



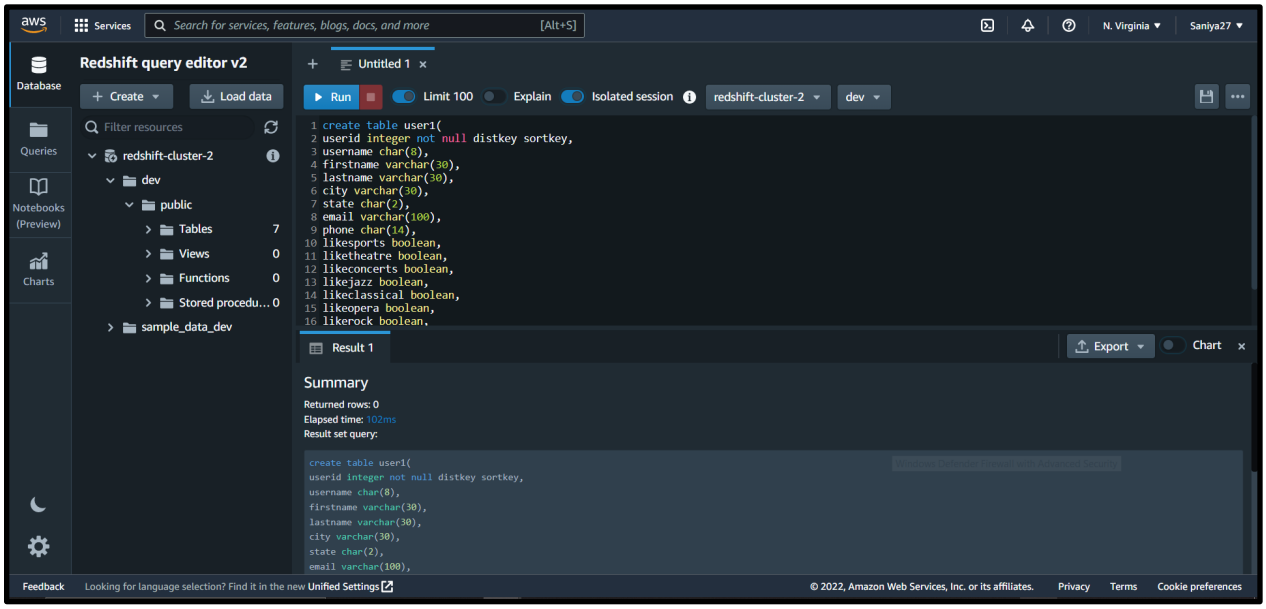
Step 7: Created redshift-cluster-2



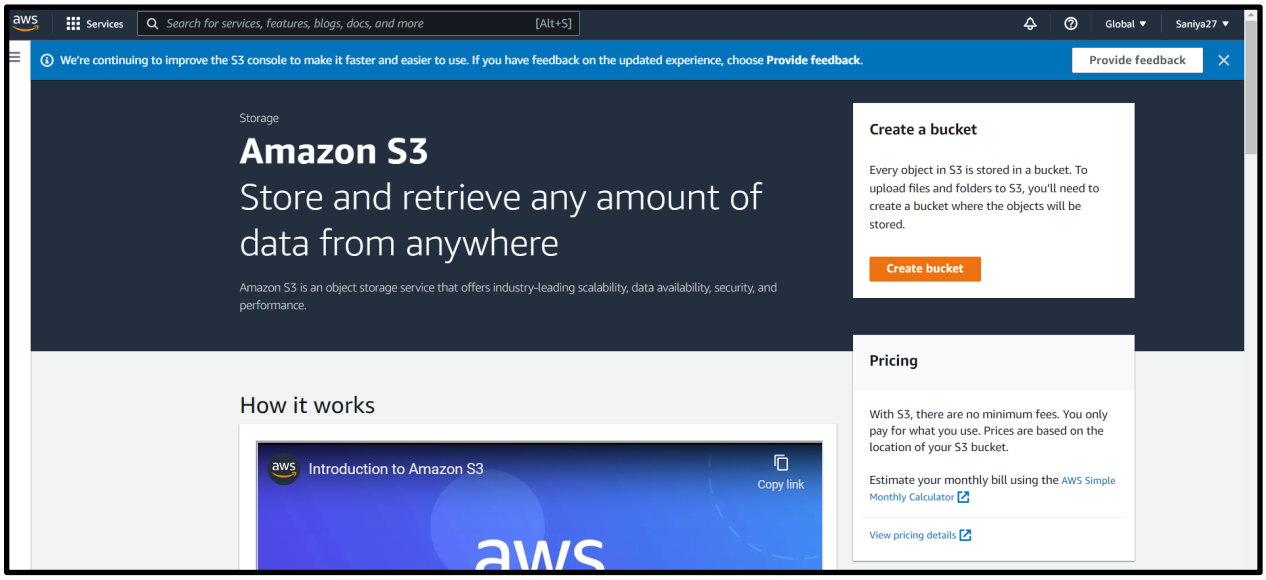
Step 8: Selected the cluster to create a table for data loading from S3 via query editor



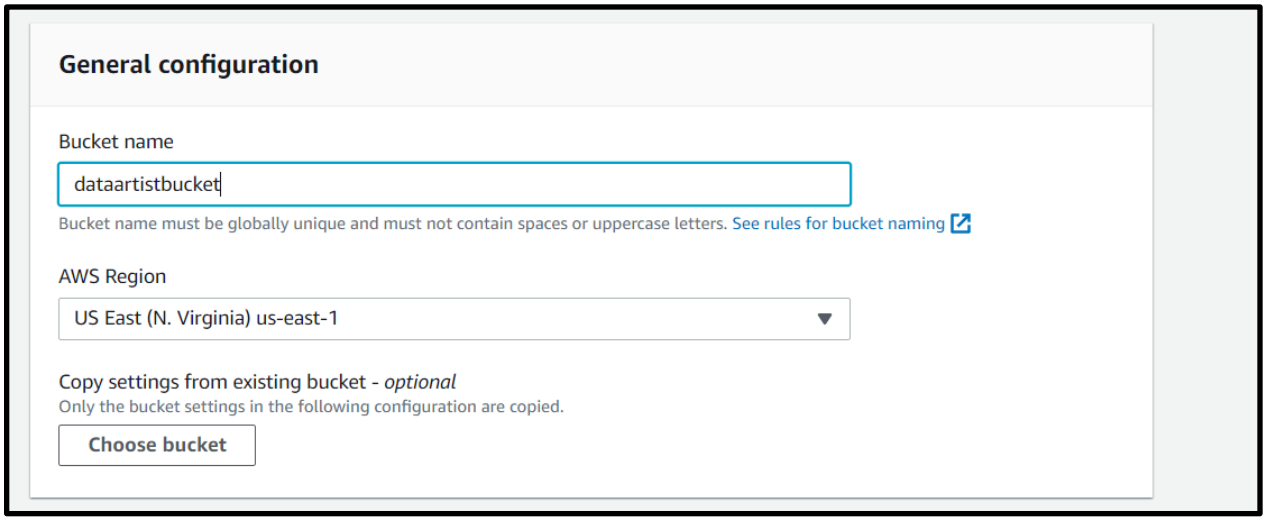
Step 9: Table Creation for loading the data from S3 bucket.



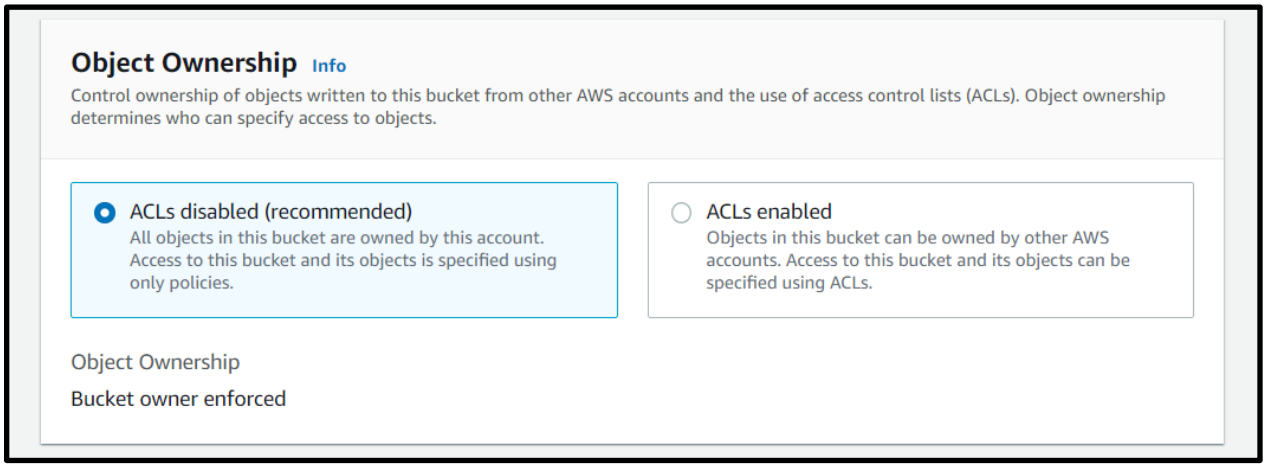
Step 10: Creating S3 bucket



Step 11: Naming the S3 bucket and specifying the AWS Region



Step 12: Configurations selected while creating a S3 bucket



Block Public Access settings for this bucket

Public access is granted to buckets and objects through access control lists (ACLs), bucket policies, access point policies, or all. In order to ensure that public access to this bucket and its objects is blocked, turn on Block all public access. These settings apply only to this bucket and its access points. AWS recommends that you turn on Block all public access, but before applying any of these settings, ensure that your applications will work correctly without public access. If you require some level of public access to this bucket or objects within, you can customize the individual settings below to suit your specific storage use cases. Learn more

- **Block *all* public access**
  Turning this setting on is the same as turning on all four settings below. Each of the following settings are independent of one another.
  - **Block public access to buckets and objects granted through *new* access control lists (ACLs)**
    S3 will block public access permissions applied to newly added buckets or objects, and prevent the creation of new public access ACLs for existing buckets and objects. This setting doesn't change any existing permissions that allow public access to S3 resources using ACLs.
  - **Block public access to buckets and objects granted through *any* access control lists (ACLs)**
    S3 will ignore all ACLs that grant public access to buckets and objects.
  - **Block public access to buckets and objects granted through *new* public bucket or access point policies**
    S3 will block new bucket and access point policies that grant public access to buckets and objects. This setting doesn't change any existing policies that allow public access to S3 resources.
  - **Block public and cross-account access to buckets and objects through *any* public bucket or access point policies**
    S3 will ignore public and cross-account access for buckets or access points with policies that grant public access to buckets and objects.

Bucket Versioning

Versioning is a means of keeping multiple variants of an object in the same bucket. You can use versioning to preserve, retrieve, and restore every version of every object stored in your Amazon S3 bucket. With versioning, you can easily recover from both unintended user actions and application failures. Learn more

Bucket Versioning

- (•) Disable
- ( ) Enable

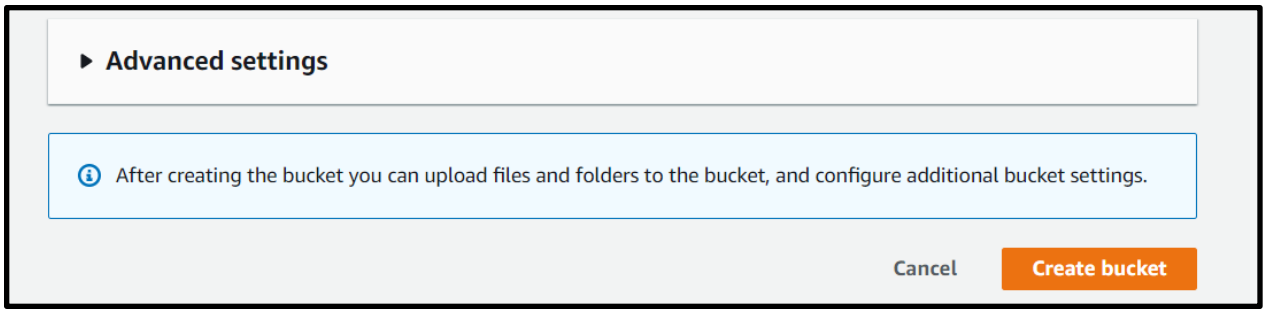
Your internet connection is unstable

Tags (0) - *optional*

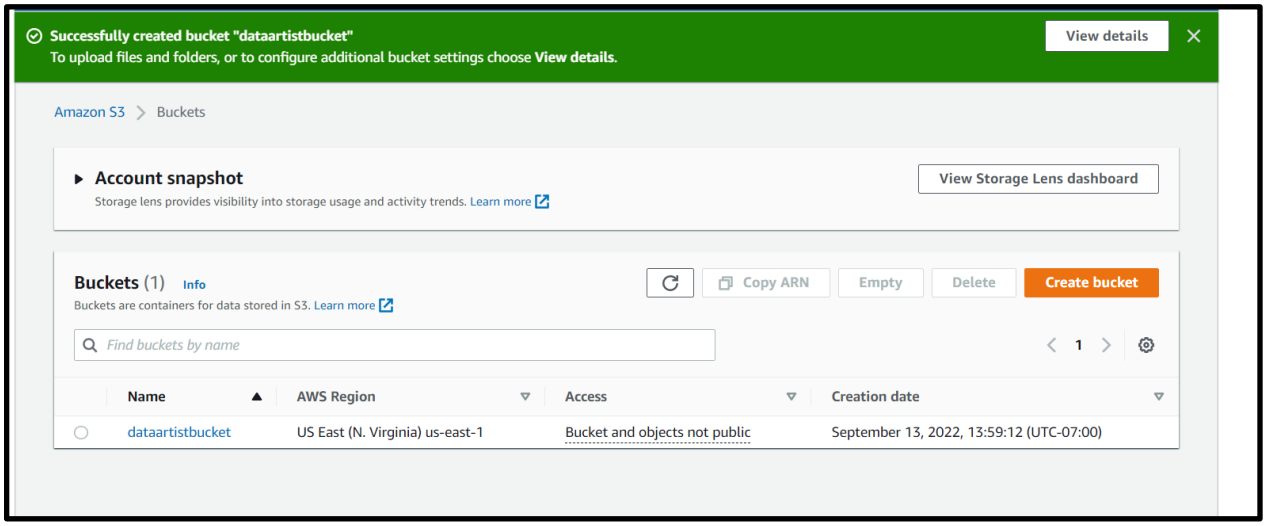
Track storage cost or other criteria by tagging your bucket. Learn more

No tags associated with this bucket.

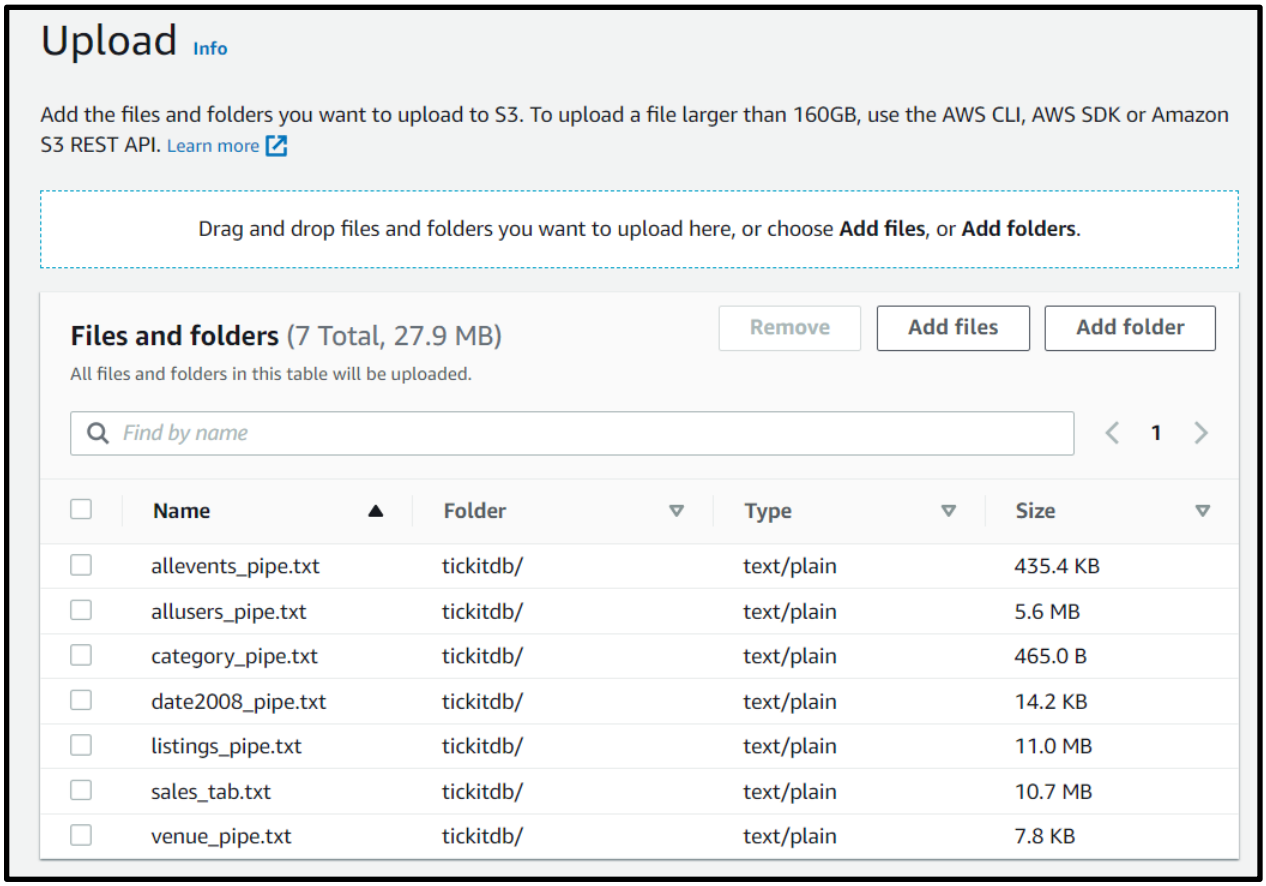
Add tag

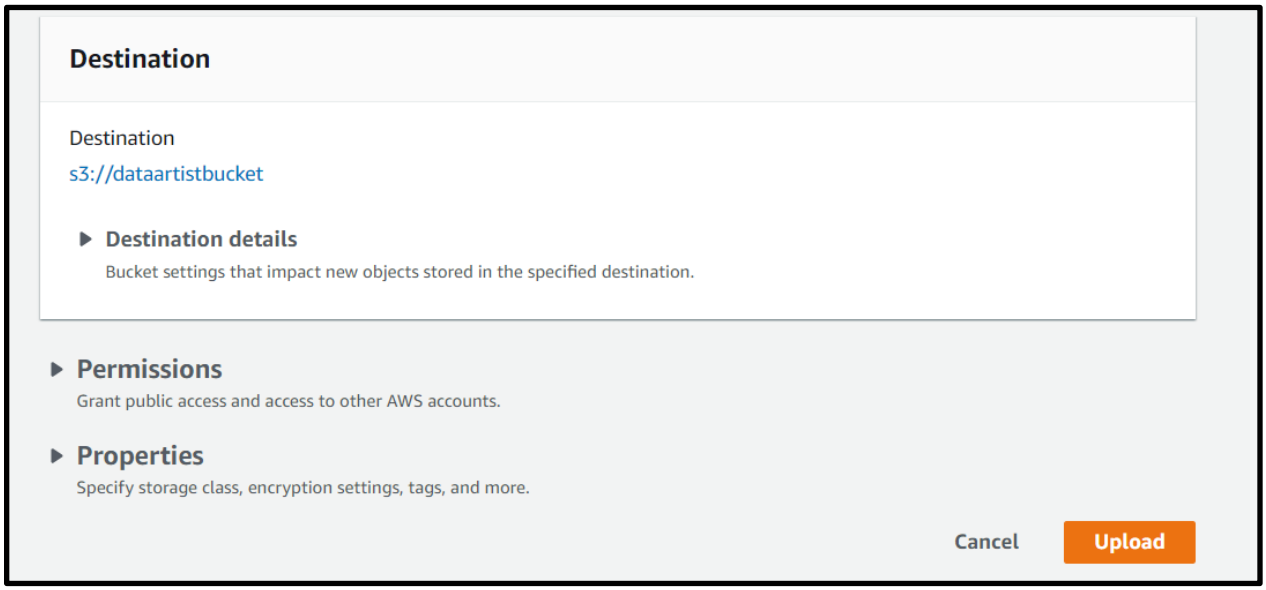


Step 13 : Successfully created S3 bucket named ‘dataartistbucket’

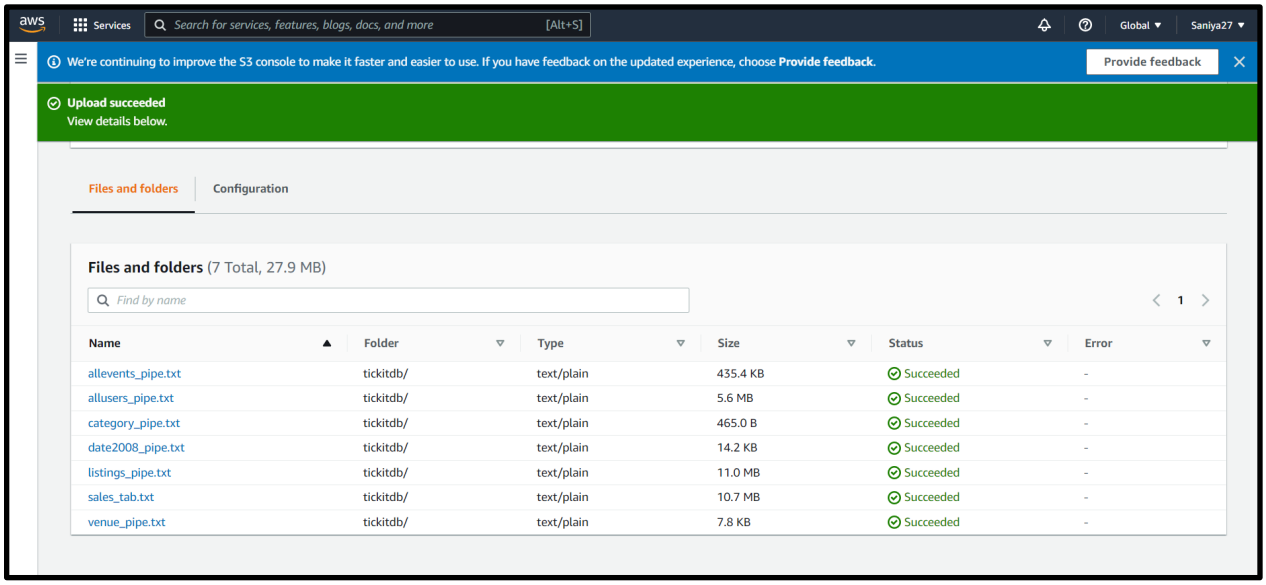


Step 14: Uploading the text files from ticketdb into the S3 bucket (downloaded ticketdb zip folder, unzipped it and loaded into S3)

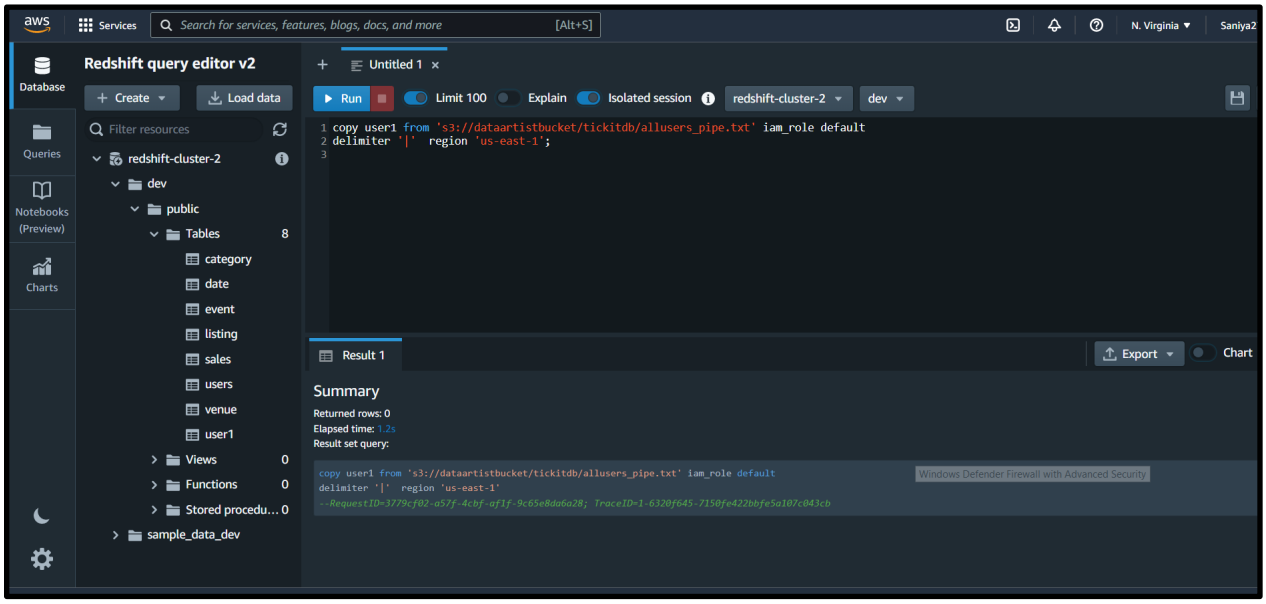


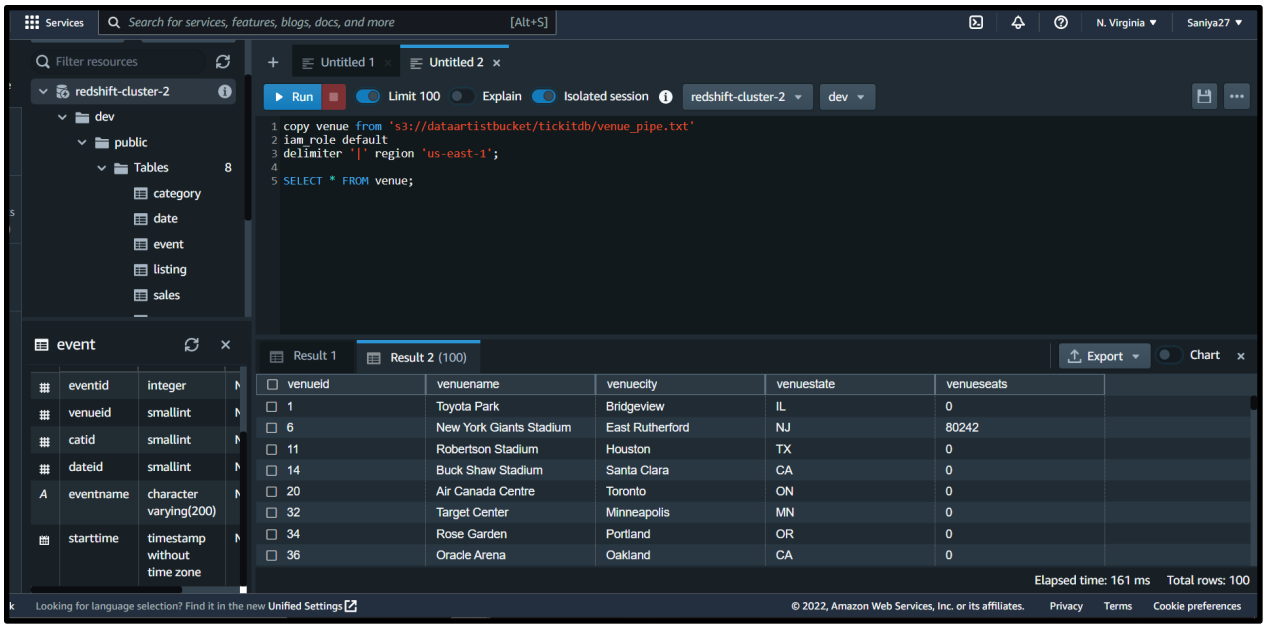


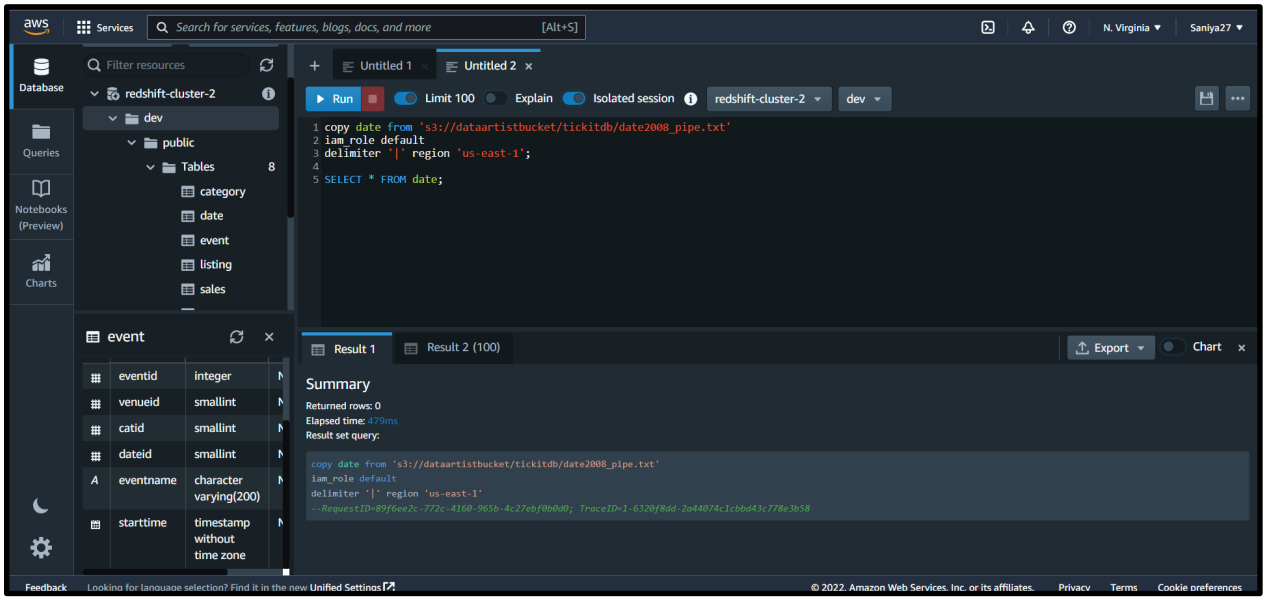
Step 15 : Successfully uploaded the data into S3 bucket.



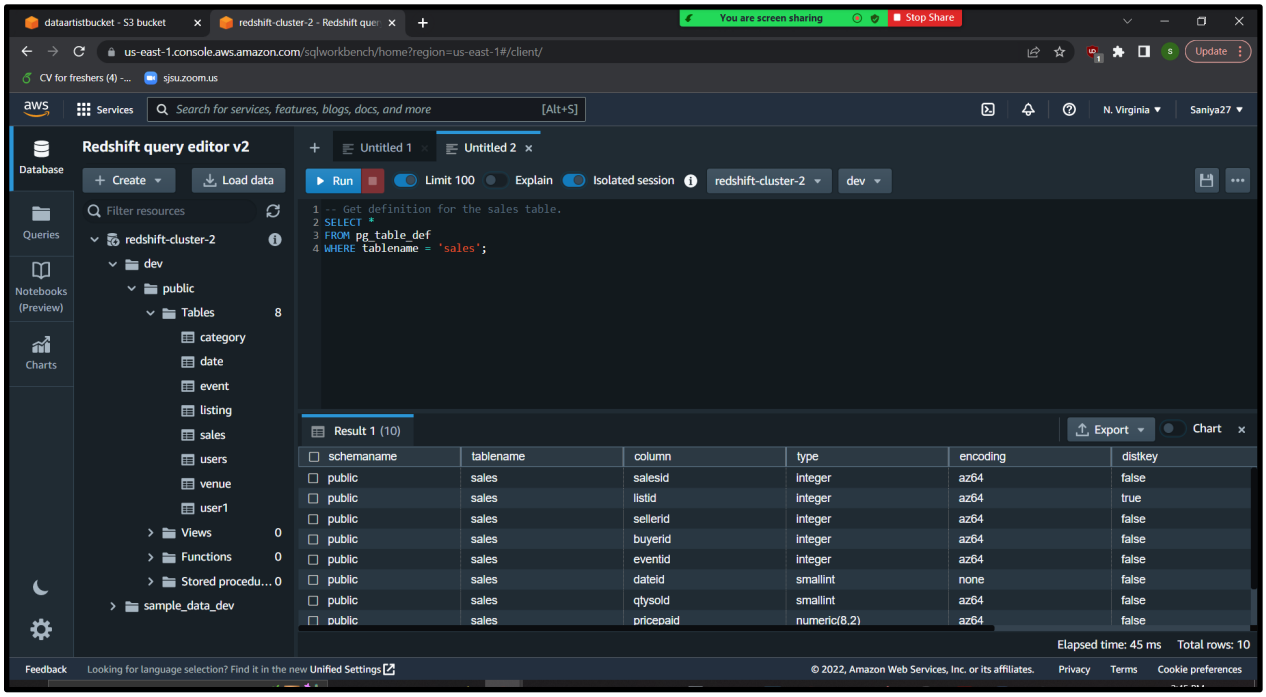
Step 16: Copying the data from S3 bucket into tables on redshift-cluster-2

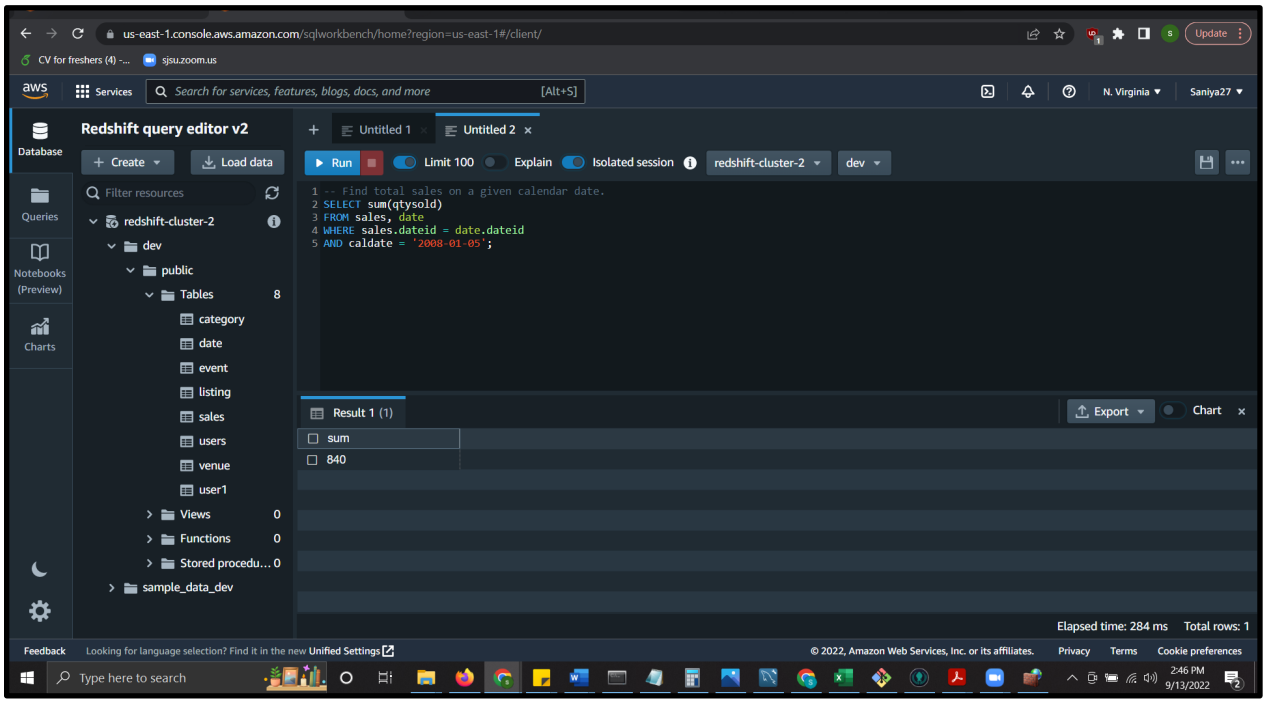


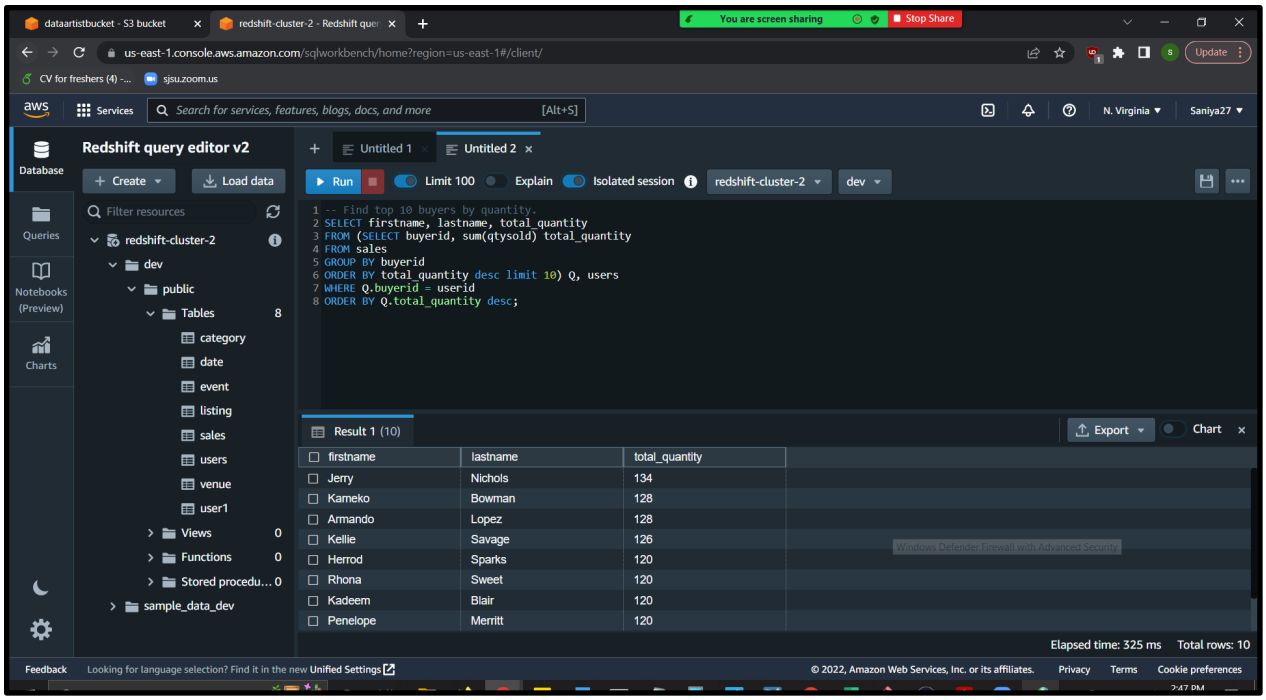

Services
Search for services, features, blogs, docs, and more [Alt+S]
N. Virginia
Saniya27
Filter resources
redshift-cluster-2
dev
public
Tables 8
category
date
event
listing
sales
Untitled 1
Untitled 2
Run
Limit 100
Explain
Isolated session
redshift-cluster-2
dev
1 copy venue from 's3://dataartistbucket/tickitdb/venue_pipe.txt'
2 iam_role default
3 delimiter '|' region 'us-east-1';
4
5 SELECT * FROM venue;
event
eventid integer
venueid smallint
catid smallint
dateid smallint
eventname character varying(200)
starttime timestamp without time zone
Result 1
Result 2 (100)
Export
Chart
venueid venuename venuecity venuestate venueseats
1 Toyota Park Bridgeview IL 0
6 New York Giants Stadium East Rutherford NJ 80242
11 Robertson Stadium Houston TX 0
14 Buck Shaw Stadium Santa Clara CA 0
20 Air Canada Centre Toronto ON 0
32 Target Center Minneapolis MN 0
34 Rose Garden Portland OR 0
36 Oracle Arena Oakland CA 0
Elapsed time: 161 ms
Total rows: 100
Looking for language selection? Find it in the new Unified Settings
© 2022, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

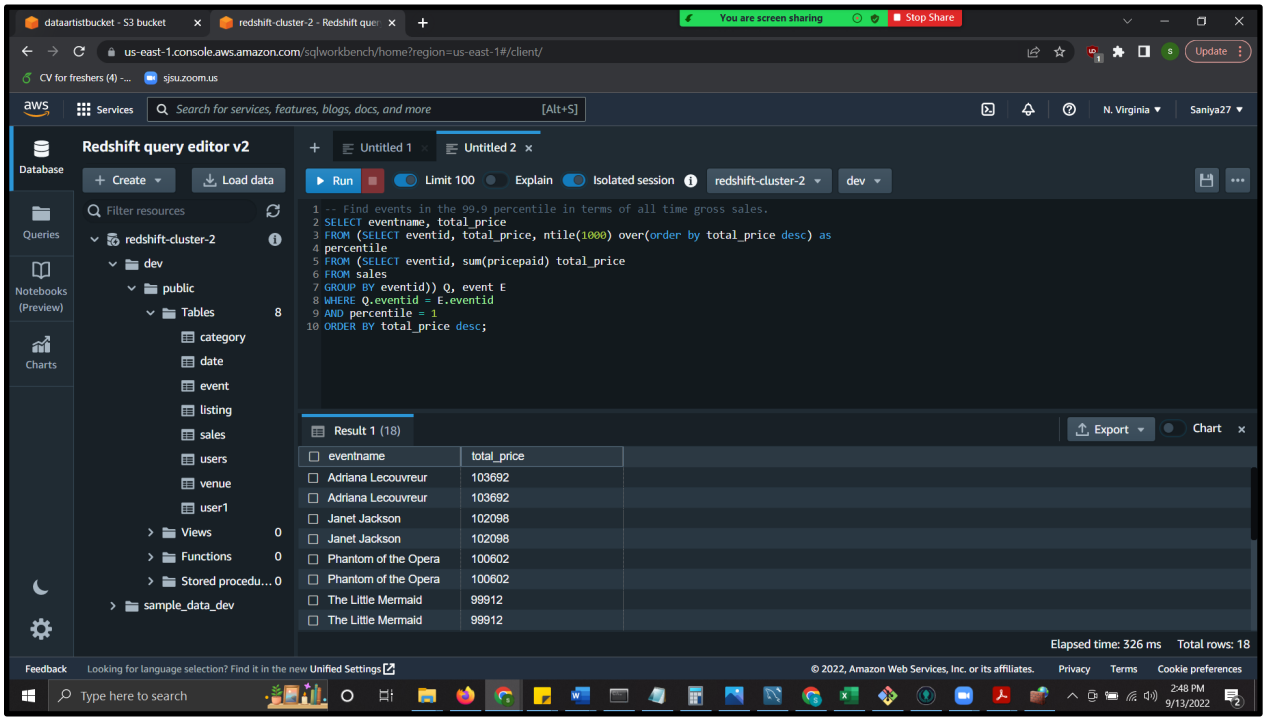

aws
Services
Search for services, features, blogs, docs, and more [Alt+S]
N. Virginia
Saniya27
Database
Queries
Notebooks (Preview)
Charts
Filter resources
redshift-cluster-2
dev
public
Tables 8
category
date
event
listing
sales
Untitled 1
Untitled 2
Run
Limit 100
Explain
Isolated session
redshift-cluster-2
dev
1 copy date from 's3://dataartistbucket/tickitdb/date2008_pipe.txt'
2 iam_role default
3 delimiter '|' region 'us-east-1';
4
5 SELECT * FROM date;
event
eventid integer
venueid smallint
catid smallint
dateid smallint
eventname character varying(200)
starttime timestamp without time zone
Result 1
Result 2 (100)
Export
Chart
Summary
Returned rows: 0
Elapsed time: 479ms
Result set query:
copy date from 's3://dataartistbucket/tickitdb/date2008_pipe.txt'
iam_role default
delimiter '|' region 'us-east-1'
--RequestID=89f6ee2c-772c-4160-965b-4c27ebf0b0d0; TraceID=1-6320f8dd-2a44074c1cbbd43c778e3b58
Feedback
Looking for language selection? Find it in the new Unified Settings
© 2022, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

Step 17 : Trying example queries using the query editor v2

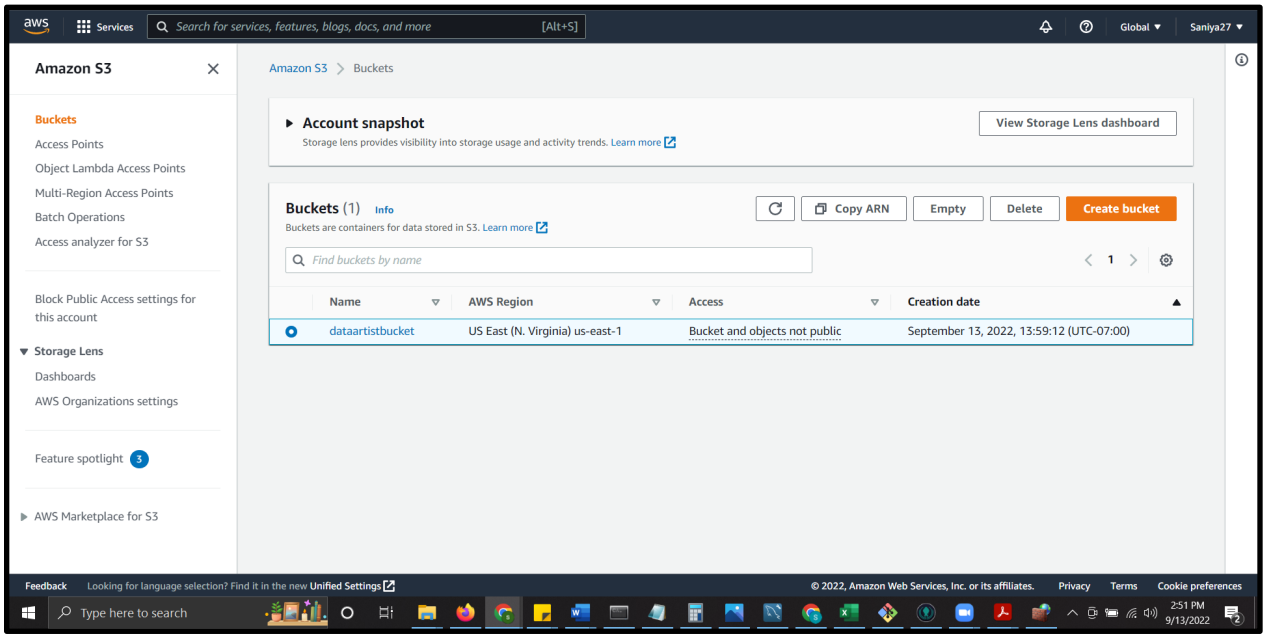


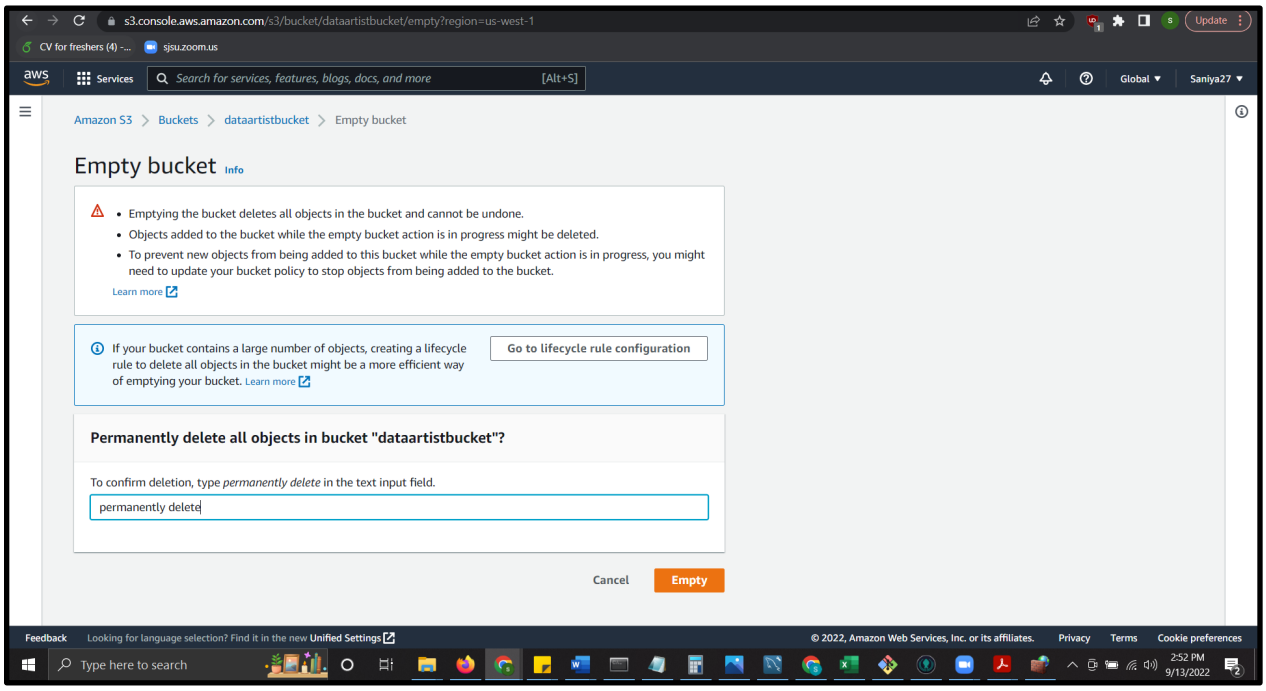


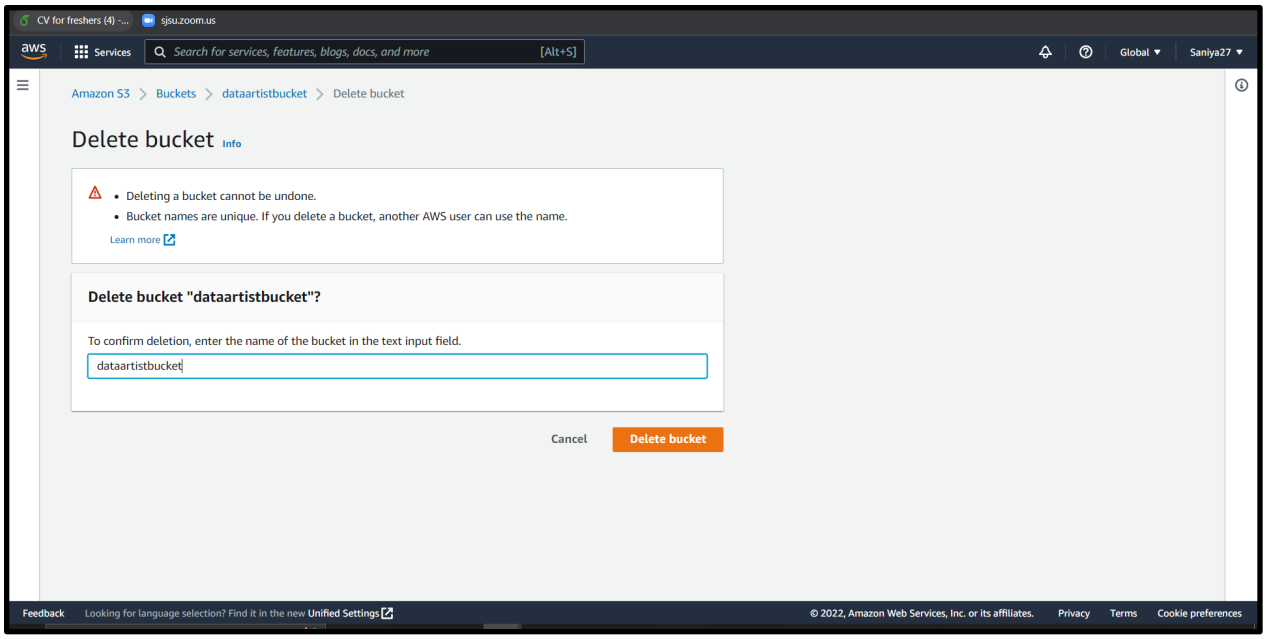

Redshift query editor v2
Untitled 1
Untitled 2
Run
Limit 100
Explain
Isolated session
redshift-cluster-2
dev
Create
Load data
Filter resources
redshift-cluster-2
dev
public
Tables 8
category
date
event
listing
sales
users
venue
user1
Views 0
Functions 0
Stored procedu... 0
sample_data_dev
-- Find top 10 buyers by quantity.
SELECT firstname, lastname, total_quantity
FROM (SELECT buyerid, sum(qtysold) total_quantity
FROM sales
GROUP BY buyerid
ORDER BY total_quantity desc limit 10) Q, users
WHERE Q.buyerid = userid
ORDER BY Q.total_quantity desc;
Result 1 (10)
Export
Chart
firstname lastname total_quantity
Jerry Nichols 134
Kameko Bowman 128
Armando Lopez 128
Kellie Savage 126
Herrod Sparks 120
Rhona Sweet 120
Kadeem Blair 120
Penelope Merritt 120
Elapsed time: 325 ms
Total rows: 10
Feedback
Looking for language selection? Find it in the new Unified Settings
© 2022, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

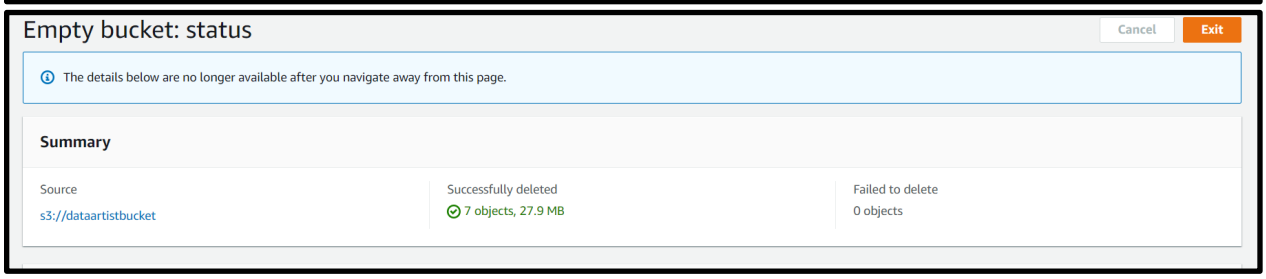

Redshift query editor v2
Untitled 1
Untitled 2
Run
Limit 100
Explain
Isolated session
redshift-cluster-2
dev
Create
Load data
Filter resources
redshift-cluster-2
dev
public
Tables 8
category
date
event
listing
sales
users
venue
user1
Views 0
Functions 0
Stored procedu... 0
sample_data_dev
-- Find events in the 99.9 percentile in terms of all time gross sales.
SELECT eventname, total_price
FROM (SELECT eventid, total_price, ntile(1000) over(order by total_price desc) as
percentile
FROM (SELECT eventid, sum(pricepaid) total_price
FROM sales
GROUP BY eventid)) Q, event E
WHERE Q.eventid = E.eventid
AND percentile = 1
ORDER BY total_price desc;
Result 1 (18)
Export
Chart
eventname total_price
Adriana Lecouvreur 103692
Adriana Lecouvreur 103692
Janet Jackson 102098
Janet Jackson 102098
Phantom of the Opera 100602
Phantom of the Opera 100602
The Little Mermaid 99912
The Little Mermaid 99912
Elapsed time: 326 ms
Total rows: 18
Feedback
Looking for language selection? Find it in the new Unified Settings
© 2022, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

Step 18 : Resetting our environment by deleting the cluster (associated S3 bucket & data folder) created









Step 19: Deletion of all created clusters