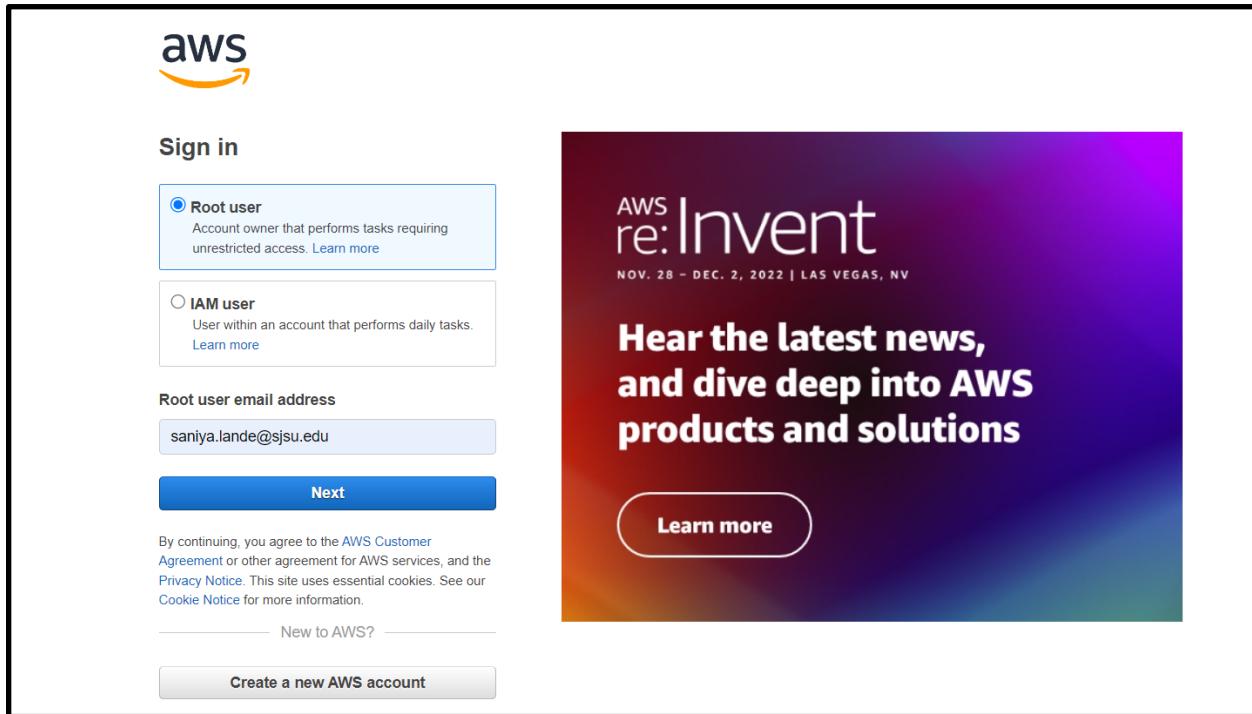


Homework-1

Step1: AWS Console page login - Account details



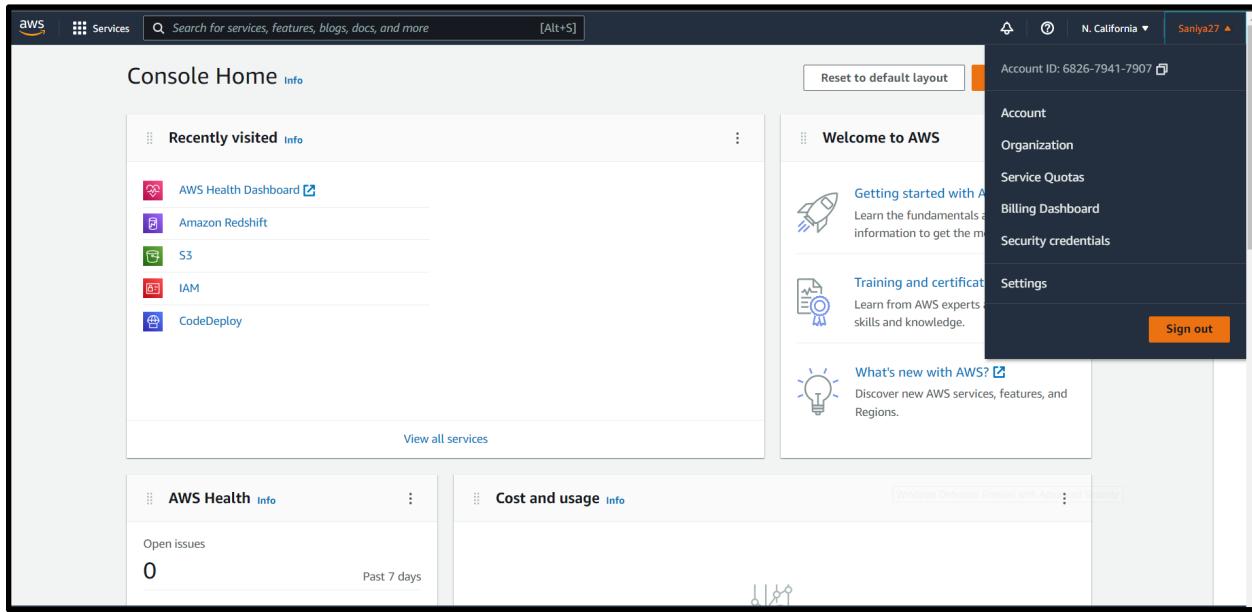
The image shows the AWS sign-in page and a promotional banner for AWS re:Invent 2022.

AWS Sign-in Page:

- The AWS logo is at the top left.
- The title "Sign in" is centered above the input fields.
- Root user:** Selected radio button. Description: "Account owner that performs tasks requiring unrestricted access." [Learn more](#).
- IAM user:** Unselected radio button. Description: "User within an account that performs daily tasks." [Learn more](#).
- A text input field for "Root user email address" containing "saniya.lande@sjsu.edu".
- A blue "Next" button.
- Below the "Next" button:
 - Text: "By continuing, you agree to the [AWS Customer Agreement](#) or other agreement for AWS services, and the [Privacy Notice](#). This site uses essential cookies. See our [Cookie Notice](#) for more information."
 - A horizontal line with "New to AWS?" in the center.
 - A grey "Create a new AWS account" button.

AWS re:Invent 2022 Banner:

- Background image for the banner.
- Text: "AWS re:Invent NOV. 28 – DEC. 2, 2022 | LAS VEGAS, NV".
- Text: "Hear the latest news, and dive deep into AWS products and solutions".
- A blue "Learn more" button.



The image shows the AWS Console Home page.

Header:

- The AWS logo and "Services" menu icon are on the left.
- A search bar with placeholder "Search for services, features, blogs, docs, and more" and a keyboard shortcut "[Alt+S]".
- User information: "Account ID: 6826-7941-7907" with a copy icon, "N. California" region, and a "Saniya27" profile icon.

Left Sidebar:

- "Recently visited" section with links to AWS Health Dashboard, Amazon Redshift, S3, IAM, and CodeDeploy.
- A "View all services" button.

Right Sidebar:

- "Welcome to AWS" section with links to "Getting started with AWS", "Training and certificates", and "What's new with AWS?".
- "Account" and "Organization" sections.
- "Service Quotas", "Billing Dashboard", and "Security credentials" links.
- "Settings" and "Sign out" buttons.

Bottom Navigation:

- "AWS Health" and "Cost and usage" cards.
- A "Windows Defender Firewall with Advanced Security" card.

Step 2: Redshift Cluster creation

Amazon Redshift

Accelerate your time to insights with fast, easy, and secure analytics at scale.

Amazon Redshift makes it easier for you to run and scale analytics without having to manage your data warehouse. Get insights by running real-time and predictive analytics on all of your data, across operational databases, data lake, data warehouse, and thousands of third-party datasets.

How it works

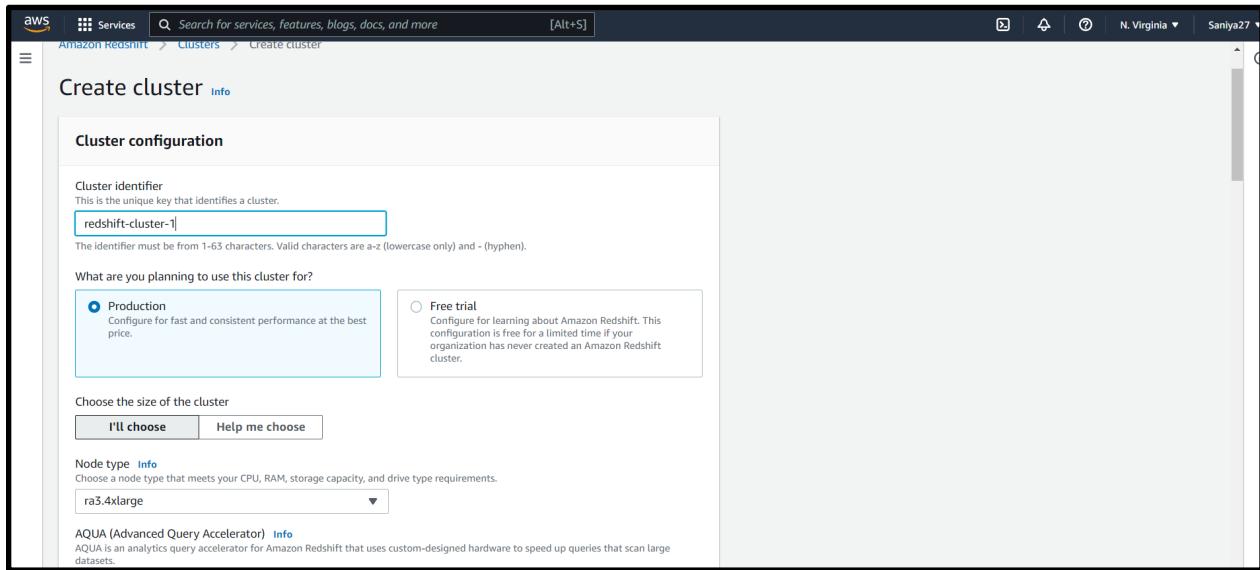
A screenshot showing a video thumbnail for 'Introduction to Data Warehousing on AWS with Am...'. Below the thumbnail is a 'Copy link' button.

Provision and manage clusters

With a few clicks, you can create your first Amazon Redshift provisioned cluster in minutes.

[Create cluster](#)

Step 3: Creating cluster in free trial



The screenshot shows the 'Create cluster' wizard on the AWS Management Console. The 'Cluster configuration' step is displayed. Key fields include:

- Cluster identifier:** redshift-cluster-1 (highlighted with a blue border)
- What are you planning to use this cluster for?**
 - Production**: Configure for fast and consistent performance at the best price.
 - Free trial**: Configure for learning about Amazon Redshift. This configuration is free for a limited time if your organization has never created an Amazon Redshift cluster.
- Choose the size of the cluster:** I'll choose (highlighted with a blue border)
- Node type:** r5.4xlarge (highlighted with a blue border)

Step 4: Database Configuration

Calculated configuration summary

dc2.large | 1 node

High performance with fixed local SSD storage

Compute

2 vCPU (gen 2) / node x 1 = 2 vCPU

Estimated on-demand compute price

\$2,190.00/year

\$0.25/node/hour

Estimated reserved (1 year)

\$1,407.00/year

\$0.075/node/hour

36% discount

Estimated reserved (3 year)

\$879.667/year

\$0.05/node/hour

60% discount

Storage capacity

160 GB x 1 nodes = 160 GB

Step 5: Loading sample dataset into Amazon Redshift cluster

Sample data [Info](#)

 Sample data is loaded with your Redshift cluster.

Ticket (28 MB)

Ticket is the sample data set that uses a sample database called TICKIT. Ticket contains individual sample data files: two fact tables and five dimensions.

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

Database configurations

Admin user name
Enter a login ID for the admin user of your DB instance.

The name must be 1-128 alphanumeric characters, and it can't be a [reserved word](#).

Auto generate password
Amazon Redshift can generate a password for you, or you can specify your own password.

[Cancel](#) [Create cluster](#)

Step 7: Redshift-cluster-1 being created

redshift-cluster-1 is being created. [X](#)

Sample data load [X](#)
After the cluster is created, Amazon Redshift starts to load the sample data.

Amazon Redshift > Clusters

In my account [From other accounts](#)

▼ Connect to Redshift clusters

Query data using Redshift query editor Use the query editor v2 to run queries in your Redshift cluster. Query data	Work with your client tools You can connect to Amazon Redshift from your client tools, such as SQL clients, business intelligence (BI) tools, and extract, transform, load (ETL) tools, using JDBC or ODBC drivers. Cluster <input type="text" value="Cluster identifier"/> Copy JDBC URL Copy ODBC URL	Choose your JDBC or ODBC driver Use JDBC or ODBC drivers to connect to Amazon Redshift from your client tools, such as SQL clients, BI tools, and ETL tools. We recommend using the new Amazon Redshift-specific drivers for better performance and scalability. Driver <input type="text" value="JDBC 4.2 without AWS SDK (jar)"/> Download driver
---	---	---

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

The screenshot shows the Amazon Redshift console under the 'Clusters' section. At the top, there are two green success notifications: one for creating the cluster 'redshift-cluster-1' and another for loading sample data. Below these, the 'Connect to Redshift clusters' section is visible, featuring three main sections: 'Query data using Redshift query editor', 'Work with your client tools', and 'Choose your JDBC or ODBC driver'. The 'Query data' button is highlighted.

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

The screenshot shows the 'Clusters' page with one cluster listed: 'redshift-cluster-1'. The cluster is in an 'Available' state with 15% CPU utilization. The 'Actions' dropdown menu is open, showing options like 'Query in query editor' and 'Query in query editor v2'.

Step 10: Connecting to sample database from query editor v2

The screenshot shows the 'Redshift query editor v2' interface. On the left sidebar, under the 'Database' section, 'redshift-cluster-1' is selected, revealing its sub-folders 'dev' and 'sample_data_dev'. The main query editor window is titled 'Untitled 1' and contains a single digit '1'.

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

The screenshot shows the AWS Redshift query editor v2 interface. The left sidebar displays the database schema with a cluster named 'redshift-cluster-1' and a schema named 'dev'. The 'Tables' section under 'dev/public' contains tables like category, date, event, listing, sales, users, and venue. A sample data schema 'sample_data_dev' is also listed. The main editor window shows a query to find total sales on a given calendar date:

```

1 -- Find total sales on a given calendar date.
2 SELECT sum(qtysold)
3 FROM sales, date
4 WHERE sales.dateid = date.dateid
5 AND caldate = '2008-01-05';
6

```

The results table shows a single row with a value of 210. The status bar at the bottom indicates an elapsed time of 276 ms and total rows of 1.

The screenshot shows the AWS Redshift query editor v2 interface. The left sidebar displays the database schema with a cluster named 'redshift-cluster-1' and a schema named 'dev'. The 'Tables' section under 'dev/public' contains tables like category, date, event, listing, sales, users, and venue. A sample data schema 'sample_data_dev' is also listed. The main editor window shows a query to find top 10 buyers by quantity:

```

1 -- Find top 10 buyers by quantity.
2 SELECT firstname, lastname, total_quantity
3 FROM (SELECT buyerid, sum(qtysold) total_quantity
4 FROM sales
5 GROUP BY buyerid
6 ORDER BY total_quantity desc limit 10) Q, users
7 WHERE Q.buyerid = user.id
8 ORDER BY Q.total_quantity desc;

```

The results table lists 10 buyers with their first name, last name, and total quantity. The status bar at the bottom indicates an elapsed time of 406 ms and total rows of 10.

The screenshot shows the AWS Redshift query editor v2 interface. The left sidebar displays the database schema with a cluster named 'redshift-cluster-1' and a schema named 'dev'. The 'Tables' section under 'dev/public' contains tables like category, date, event, listing, sales, users, and venue. A sample data schema 'sample_data_dev' is also listed. The main editor window shows a query to find events in the 99.9 percentile in terms of all time gross sales:

```

1 -- Find events in the 99.9 percentile in terms of all time gross sales.
2 SELECT evenname, total_price
3 FROM (SELECT eventid, total_price, ntile(1000) over(order by total_price desc) as
4 percentile
5 FROM (SELECT eventid, sum(pricepaid) total_price
6 FROM sales
7 GROUP BY eventid) Q, event E
8 WHERE Q.eventid = E.eventid
9 AND percentile = 1
10 ORDER BY total_price desc;

```

The results table lists 9 events with their even name and total price. The status bar at the bottom indicates an elapsed time of 365 ms and total rows of 9.

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

The screenshot shows the 'Clusters' section of the Amazon Redshift console. On the left, there's a sidebar with 'Connect to Redshift clusters' options like 'Query data using Redshift query editor' and 'Work with your client tools'. The main area lists one cluster:

Cluster	Cluster namespace	Status	Storage capacity us...	CPU utilization	Snapshots	Notifications	Tags
redshift-cluster-1 dc2.large 1 node 160 GB	2ae7eeeb-32ee-42ce-...	Available	< 1%	15%	-	-	-

A context menu on the right side of the cluster row provides options like 'Manage cluster', 'Resize', 'Edit', 'Reboot', 'Relocate', 'Pause', 'Add AWS Partner integration', 'Delete', 'Defer maintenance', 'Configure AQUA', and 'Modify publicly accessible setting'. The 'Delete' option is highlighted.

Step 13: Deleting redshift-cluster-1

The screenshot shows a confirmation dialog box titled 'Delete redshift-cluster-1?'. It contains the following text and options:

Deleting the cluster causes the following results:

- Deletes all databases (and data) in the cluster.
- You can't rotate keys for encrypted manual snapshots if you delete this cluster.
- Removes access to the data in datashares for data consumers, including subscribers.

Are you sure that you want to permanently delete **redshift-cluster-1**?

Final snapshot
You can create a final manual snapshot of your cluster before it's deleted so you can later restore it. Restoring it enables you to resume running the cluster and querying data.

Create final snapshot

Final snapshot identifier
Identifier for the new snapshot to be created

To confirm deletion, enter *delete* in the field and choose Delete.

Cancel **Delete cluster**

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

Cluster configuration

Cluster identifier
This is the unique key that identifies a cluster.

The identifier must be from 1-63 characters. Valid characters are a-z (lowercase only) and - (hyphen).

What are you planning to use this cluster for?

Production
Configure for fast and consistent performance at the best price.

Free trial
Configure for learning about Amazon Redshift. This configuration is free for a limited time if your organization has never created an Amazon Redshift cluster.

Choose the size of the cluster

Node type [Info](#)
Choose a node type that meets your CPU, RAM, storage capacity, and drive type requirements.
 ▾

Number of nodes
Enter the number of nodes that you need.

Range (1-32)

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

Configuration summary [Info](#)
dc2.large | 2 nodes

\$360.00/month Estimated on-demand compute price Save more than 60% of your costs by purchasing reserved nodes. Learn more	320 GB Total compressed storage The total storage capacity for the cluster if you deploy the number of nodes that you chose.
--	---

Sample data [Info](#)

Load sample data
Load sample data to your Redshift cluster to start using the query editor to query data.

Ticket (28 MB)
Ticket is the sample data set that uses a sample database called TICKIT. Ticket contains individual sample data files: two fact tables and five dimensions.

Step 3 : Database Configuration

Database configurations

Admin user name
Enter a login ID for the admin user of your DB instance.

awsuser

The name must be 1-128 alphanumeric characters, and it can't be a [reserved word](#).

Auto generate password
Amazon Redshift can generate a password for you, or you can specify your own password.

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

Create the default IAM role

ⓘ Create an IAM role as the default for this cluster that has the [AmazonRedshiftAllCommandsFullAccess](#) policy attached. This policy includes permissions to run SQL commands to COPY, UNLOAD, and query data with Amazon Redshift. The policy also grants permissions to run SELECT statements for related services, such as Amazon S3, Amazon CloudWatch logs, Amazon SageMaker, and AWS Glue.

Specify an S3 bucket for the IAM role to access
To create a new bucket, [visit S3](#)

No additional S3 bucket
Create the IAM role without specifying S3 buckets.

Any S3 bucket
Allow users that have access to your Redshift cluster to also access any S3 bucket and its contents in your AWS account.

Specific S3 buckets
Specify one or more S3 buckets that the IAM role being created has permission to access.

Cancel **Create IAM role as default**

Associated IAM roles (1) Info

Create, associate, or remove an IAM role. You can associate up to 50 IAM roles. You can also choose an IAM role and set it as the default for this cluster.

[Set default ▾](#) [Manage IAM roles ▾](#)

Search for associated IAM role by name, status, or role type < 1 >

<input type="checkbox"/>	IAM roles 	Status	Role type
<input type="checkbox"/>	AmazonRedshift-CommandsAccessRole-20220913T131916	Not applied	Default

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

Additional configurations Use defaults

These configurations are optional, and default settings have been defined to help you get started with your cluster. Turn off "Use defaults" to modify these settings now.

Network Using default VPC (vpc-0aa45d3e56337310f) and default subnet .	Security Using default (sg-0efad47a8eccfb31) cluster security group.
Backup Automated snapshots are created about every eight hours or following every 5 GB per node of data changes, whichever comes first.	Configuration Using default.redshift-1.0 parameter group with no database encryption.
Maintenance Using current maintenance track.	

[Cancel](#) [Create cluster](#)

Step 6: redshift-cluster-2 being created

The screenshot shows the AWS Redshift console with the following details:

- Cluster Status:** redshift-cluster-2 is being created.
- Sample data load:** After the cluster is created, Amazon Redshift starts to load the sample data.
- Navigation:** Amazon Redshift > Clusters
- Account Selection:** In my account
- Connect to Redshift clusters:**
 - Query data using Redshift query editor:** Use the query editor v2 to run queries in your Redshift cluster. A "Query data" button is present.
 - Work with your client tools:** You can connect to Amazon Redshift from your client tools, such as SQL clients, business intelligence (BI) tools, and extract, transform, load (ETL) tools, using JDBC or ODBC drivers. A "Cluster identifier" dropdown menu is shown, along with "Copy JDBC URL" and "Copy ODBC URL" buttons.
 - Choose your JDBC or ODBC driver:** Use JDBC or ODBC drivers to connect to Amazon Redshift from your client tools, such as SQL clients, BI tools, and ETL tools. We recommend using the new Amazon Redshift-specific drivers for better performance and scalability. A "Driver" dropdown menu is set to "JDBC 4.2 without AWS SDK (.jar)" and a "Download driver" button is available.

At the bottom, the browser toolbar includes a search bar, various icons, and the date/time: 9/13/2022 1:22 PM.

Step 7: Created redshift-cluster-2

The screenshot shows the AWS Redshift console with the following details:

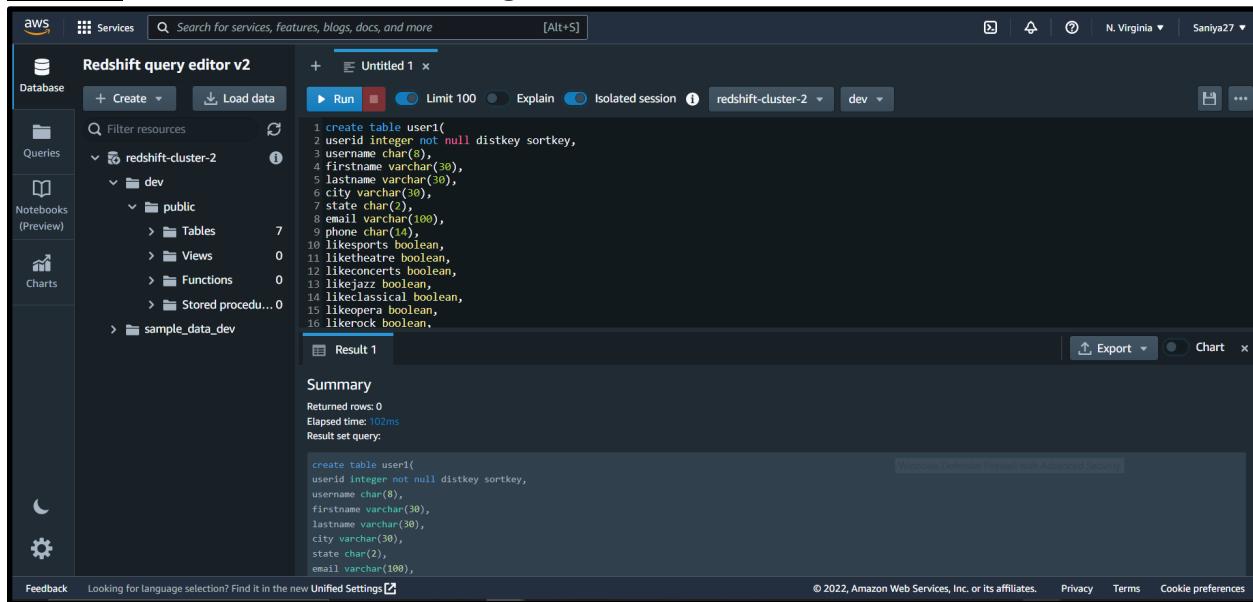
- Success Message:** Cluster redshift-cluster-1 is being deleted. Successfully created cluster redshift-cluster-2. To manage clusters or connect to the cluster, go to the redshift-cluster-2 details page.
- Sample Data Message:** Successfully loaded sample data. To query the sample data, use the Redshift query editor v2. A "Query data" button is present.
- Navigation:** Amazon Redshift > Clusters
- Account Selection:** In my account
- Connect to Redshift clusters:**
 - Query data using Redshift query editor:** Use the query editor v2 to run queries in your Redshift cluster. A "Query data" button is present.
 - Work with your client tools:** You can connect to Amazon Redshift from your client tools, such as SQL clients, business intelligence (BI) tools, and extract, transform, load (ETL) tools, using JDBC or ODBC drivers. A "Cluster" dropdown menu is shown.
 - Choose your JDBC or ODBC driver:** Use JDBC or ODBC drivers to connect to Amazon Redshift from your client tools, such as SQL clients, BI tools, and ETL tools. We recommend using the new Amazon Redshift-specific drivers for better performance and scalability.

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

The screenshot shows the AWS Redshift console with the following details:

- Clusters:** Clusters (1/1) Info
- Filter:** Filter clusters by property or value
- Columns:** Cluster, Cluster namespace, Status, Storage capacity us..., CPU utilization, Snapshots, Notifications, Tags
- Selected Cluster:** redshift-cluster-2 (dc2.large | 2 nodes | 320 GB) - Available
- Actions:** Create cluster, Download driver

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



The screenshot shows the AWS Redshift query editor v2 interface. In the left sidebar, under 'Database' > 'redshift-cluster-2' > 'dev' > 'Tables', there are 7 tables listed. A new table named 'user1' is being created in the 'dev' database. The SQL code for the table creation is as follows:

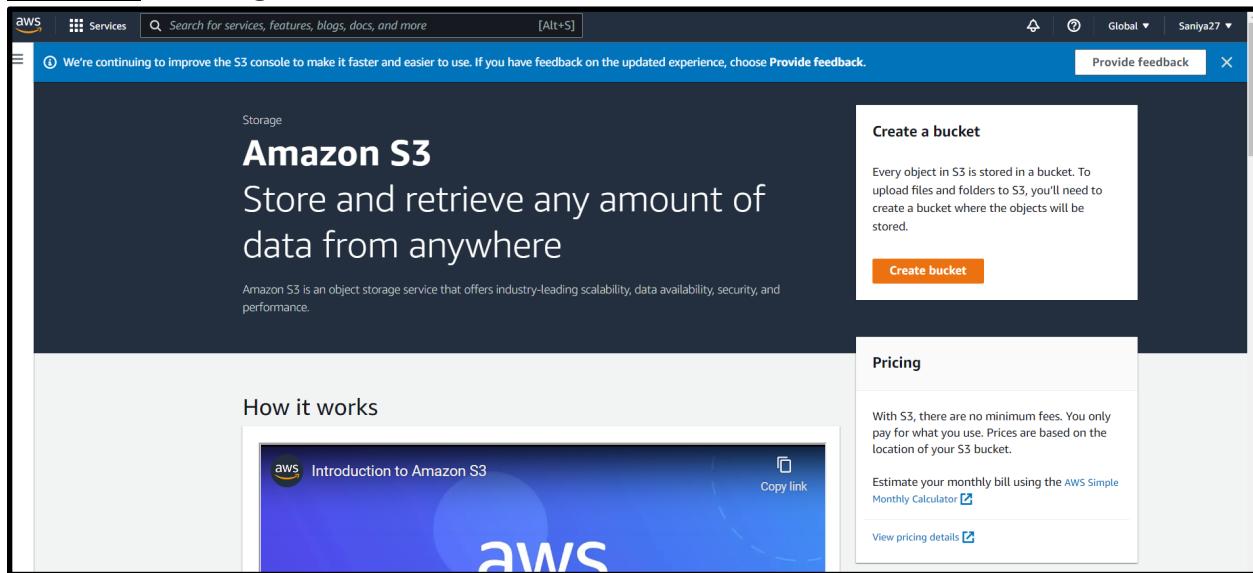
```
1 create table user1(
2     userid integer not null distkey sortkey,
3     username char(8),
4     firstname varchar(30),
5     lastname varchar(30),
6     city varchar(30),
7     state char(2),
8     email varchar(100),
9     phone char(14),
10    likesports boolean,
11    liketheatre boolean,
12    likeconcerts boolean,
13    likejazz boolean,
14    likeclassical boolean,
15    likeopera boolean,
16    likerock boolean,
```

The 'Result 1' tab shows the summary of the query execution:

Summary
Returned rows: 0
Elapsed time: 102ms
Result set query:

```
create table user1(
    userid integer not null distkey sortkey,
    username char(8),
    firstname varchar(30),
    lastname varchar(30),
    city varchar(30),
    state char(2),
    email varchar(100),
```

Step 10: Creating S3 bucket



The screenshot shows the AWS S3 console. On the left, there's a 'Storage' sidebar. The main area features the 'Amazon S3' heading and the sub-headline 'Store and retrieve any amount of data from anywhere'. Below this, a paragraph explains what Amazon S3 is. On the right, a 'Create a bucket' wizard is open. It contains a brief description of what an S3 bucket is and a large orange 'Create bucket' button. To the right of the wizard, there's a 'Pricing' section with a note about no minimum fees and a link to the AWS Simple Monthly Calculator. At the bottom, there's a 'View pricing details' link.

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

General configuration

Bucket name

Bucket name must be globally unique and must not contain spaces or uppercase letters. See [rules for bucket naming](#)

AWS Region

Copy settings from existing bucket - *optional*
Only the bucket settings in the following configuration are copied.

[Choose bucket](#)

Step 12: Configurations selected while creating a S3 bucket

Object Ownership [Info](#)

Control ownership of objects written to this bucket from other AWS accounts and the use of access control lists (ACLs). Object ownership determines who can specify access to objects.

ACLs disabled (recommended)
All objects in this bucket are owned by this account.
Access to this bucket and its objects is specified using only policies.

ACLs enabled
Objects in this bucket can be owned by other AWS accounts. Access to this bucket and its objects can be specified using ACLs.

Object Ownership
Bucket owner enforced

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

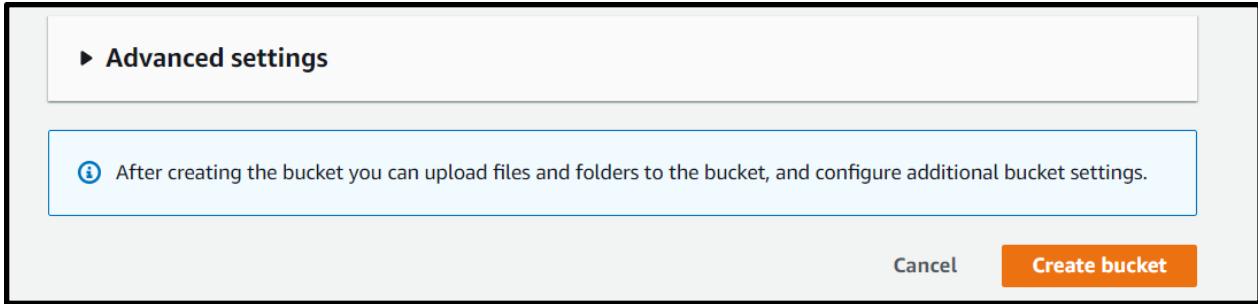
Your internet connection is unstable

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'

The screenshot shows the AWS S3 Buckets page. At the top, a green banner says 'Successfully created bucket "dataartistbucket"' and 'To upload files and folders, or to configure additional bucket settings choose View details.' Below the banner, the page title is 'Amazon S3 > Buckets'. There's an 'Account snapshot' section with a 'View Storage Lens dashboard' button. The main area shows a table of buckets:

Name	AWS Region	Access	Creation date
dataartistbucket	US East (N. Virginia) us-east-1	Bucket and objects not public	September 13, 2022, 13:59:12 (UTC-07:00)

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

Upload [Info](#)

Add the files and folders you want to upload to S3. To upload a file larger than 160GB, use the AWS CLI, AWS SDK or Amazon S3 REST API. [Learn more](#)

Drag and drop files and folders you want to upload here, or choose **Add files**, or **Add folders**.

Files and folders (7 Total, 27.9 MB)						
All files and folders in this table will be uploaded.						
<input type="text"/> Find by name						
<input type="checkbox"/>	Name	Folder	Type	Size		
<input type="checkbox"/>	allevents_pipe.txt	ticketdb/	text/plain	435.4 KB		
<input type="checkbox"/>	allusers_pipe.txt	ticketdb/	text/plain	5.6 MB		
<input type="checkbox"/>	category_pipe.txt	ticketdb/	text/plain	465.0 B		
<input type="checkbox"/>	date2008_pipe.txt	ticketdb/	text/plain	14.2 KB		
<input type="checkbox"/>	listings_pipe.txt	ticketdb/	text/plain	11.0 MB		
<input type="checkbox"/>	sales_tab.txt	ticketdb/	text/plain	10.7 MB		
<input type="checkbox"/>	venue_pipe.txt	ticketdb/	text/plain	7.8 KB		

Destination

Destination
<s3://dataartistbucket>

► **Destination details**
Bucket settings that impact new objects stored in the specified destination.

► **Permissions**
Grant public access and access to other AWS accounts.

► **Properties**
Specify storage class, encryption settings, tags, and more.

[Cancel](#) [Upload](#)

Step 15 : Successfully uploaded the data into S3 bucket.

The screenshot shows the AWS S3 console interface. At the top, a green banner displays the message "Upload succeeded". Below this, the "Files and folders" tab is selected. A table lists seven files: allevents_pipe.txt, allusers_pipe.txt, category_pipe.txt, date2008_pipe.txt, listings_pipe.txt, sales_tab.txt, and venue_pipe.txt. All files are listed as "Succeeded" with a size of 27.9 MB. The table has columns for Name, Folder, Type, Size, Status, and Error.

Name	Folder	Type	Size	Status	Error
allevents_pipe.txt	tickitdb/	text/plain	435.4 KB	Succeeded	-
allusers_pipe.txt	tickitdb/	text/plain	5.6 MB	Succeeded	-
category_pipe.txt	tickitdb/	text/plain	465.0 B	Succeeded	-
date2008_pipe.txt	tickitdb/	text/plain	14.2 KB	Succeeded	-
listings_pipe.txt	tickitdb/	text/plain	11.0 MB	Succeeded	-
sales_tab.txt	tickitdb/	text/plain	10.7 MB	Succeeded	-
venue_pipe.txt	tickitdb/	text/plain	7.8 KB	Succeeded	-

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

The screenshot shows the Redshift query editor v2 interface. On the left, the sidebar displays the database structure under "redshift-cluster-2" (dev) / public / Tables, including categories like category, date, event, listing, sales, users, and user1. The main area shows a query editor window with the following SQL code:

```
copy user1 from 's3://dataartistbucket/tickitdb/allusers_pipe.txt' iam_role default  
delimiter '|' region 'us-east-1';
```

The "Result" tab at the bottom shows the execution summary: "Returned rows: 0" and "Elapsed time: 1.2s". The result set query is also displayed.

Services Search for services, features, blogs, docs, and more [Alt+S] N. Virginia Saniya27

Filter resources redshift-cluster-2 Untitled 1 Untitled 2 Run Limit 100 Explain Isolated session redshift-cluster-2 dev

```
copy venue from 's3://dataartistbucket/ticketdb/venue_pipe.txt'
iam_role default
delimiter '|' region 'us-east-1';
SELECT * FROM venue;
```

Result 1 Result 2 (100) Export Chart

venueid	venuename	venucity	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 Filter resources redshift-cluster-2 Untitled 1 Untitled 2 Run Limit 100 Explain Isolated session redshift-cluster-2 dev

```
copy date from 's3://dataartistbucket/ticketdb/date2008_pipe.txt'
iam_role default
delimiter '|' region 'us-east-1';
SELECT * FROM date;
```

Result 1 Result 2 (100) Export Chart

Summary
Returned rows: 0
Elapsed time: 479ms
Result set query:

```
copy date from 's3://dataartistbucket/ticketdb/date2008_pipe.txt'
iam_role default
delimiter '|' region 'us-east-1'
--RequestID=89f9ee2c-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

The screenshot shows the AWS Redshift Query Editor v2 interface. On the left, the sidebar displays the database structure under 'redshift-cluster-2' / 'dev' / 'public'. A table named 'sales' is selected, showing 8 columns: salesid, listid, sellerid, buyerid, eventid, dateid, qtsold, and pricepaid. The main pane shows the query:

```
1 -- Get definition for the sales table.
2 SELECT *
3 FROM pg_table_def
4 WHERE tablename = 'sales';
```

The results pane displays the schema definition for the 'sales' table:

schemaname	tablename	column	type	encoding	distkey
public	sales	salesid	integer	az64	false
public	sales	listid	integer	az64	true
public	sales	sellerid	integer	az64	false
public	sales	buyerid	integer	az64	false
public	sales	eventid	integer	az64	false
public	sales	dateid	smallint	none	false
public	sales	qtsold	smallint	az64	false
public	sales	pricepaid	numeric(8,2)	az64	false

Elapsed time: 45 ms Total rows: 10

The screenshot shows the AWS Redshift Query Editor v2 interface. On the left, the sidebar displays the database structure under 'redshift-cluster-2' / 'dev' / 'public'. A table named 'sales' is selected, showing 8 columns: salesid, listid, sellerid, buyerid, eventid, dateid, qtsold, and pricepaid. The main pane shows the query:

```
1 -- Find total sales on a given calendar date.
2 SELECT sum(qtsold)
3 FROM sales, date
4 WHERE sales.dateid = date.dateid
5 AND caldate = '2008-01-05';
```

The results pane displays the total sales for the specified date:

sum
840

Elapsed time: 284 ms Total rows: 1

dataartistbucket - S3 bucket redshift-cluster-2 - Redshift query editor v2

CV for freshers (4) ... sjsu.zoom.us

Search for services, features, blogs, docs, and more [Alt+S]

You are screen sharing Stop Share

N. Virginia Sanjaya27

Database Services

Redshift query editor v2

+ Create Load data

Filter resources

redshift-cluster-2

dev

public

Tables

category date event listing sales users venue user1

Views Functions Stored procedures sample_data_dev

Run Explain Isolated session redshift-cluster-2 dev

Limit 100

1 -- Find top 10 buyers by quantity
2 SELECT firstname, lastname, total_quantity
3 FROM (SELECT buyerid, sum(qtysold) total_quantity
4 FROM sales
5 GROUP BY buyerid
6 ORDER BY total_quantity desc limit 10) Q, users
7 WHERE Q.buyerid = user_id
8 ORDER BY Q.total_quantity desc;

Result 1 (10)

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

This screenshot shows the AWS Redshift Query Editor interface. The left sidebar lists databases, queries, notebooks, and charts. The main area shows a query editor with a SQL statement to find the top 10 buyers by quantity. The results are displayed in a table titled 'Result 1 (10)' with columns 'firstname', 'lastname', and 'total_quantity'. The elapsed time was 325 ms and 10 rows were returned.

dataartistbucket - S3 bucket redshift-cluster-2 - Redshift query editor v2

CV for freshers (4) ... sjsu.zoom.us

Search for services, features, blogs, docs, and more [Alt+S]

You are screen sharing Stop Share

N. Virginia Sanjaya27

Database Services

Redshift query editor v2

+ Create Load data

Filter resources

redshift-cluster-2

dev

public

Tables

category date event listing sales users venue user1

Views Functions Stored procedures sample_data_dev

Run Explain Isolated session redshift-cluster-2 dev

Limit 100

1 -- Find events in the 99.9 percentile in terms of all time gross sales.
2 SELECT evenename, total_price
3 FROM (SELECT eventid, total_price, ntile(1000) over(order by total_price desc) as
4 percentile
5 FROM (SELECT eventid, sum(pricetpaid) total_price
6 FROM sales
7 GROUP BY eventid)) Q, event E
8 WHERE Q.eventid = E.eventid
9 AND percentile = 1
10 ORDER BY total_price desc;

Result 1 (18)

evenename	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

This screenshot shows the AWS Redshift Query Editor interface. The left sidebar lists databases, queries, notebooks, and charts. The main area shows a query editor with a SQL statement to find events in the 99.9 percentile in terms of all time gross sales. The results are displayed in a table titled 'Result 1 (18)' with columns 'evenename' and 'total_price'. The elapsed time was 326 ms and 18 rows were returned.

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

The screenshot shows the AWS S3 service page. On the left, there's a sidebar with 'Buckets' selected. The main area displays an 'Account snapshot' with a link to 'View Storage Lens dashboard'. Below it is a table titled 'Buckets (1) Info' showing one entry: 'dataartistbucket' from 'US East (N. Virginia) us-east-1' with a creation date of 'September 13, 2022, 13:59:12 (UTC-07:00)'. There are buttons for 'Copy ARN', 'Empty', 'Delete', and 'Create bucket'.

Buckets (1) Info

Name	AWS Region	Access	Creation date
dataartistbucket	US East (N. Virginia) us-east-1	Bucket and objects not public	September 13, 2022, 13:59:12 (UTC-07:00)

This screenshot shows the 'Empty bucket' confirmation dialog for the 'dataartistbucket'. It includes a warning message about deleting all objects, a note about lifecycle rules, and a text input field where 'permanently delete' is typed. A large 'Empty' button is at the bottom.

Empty bucket Info

⚠ • Emptying the bucket deletes all objects in the bucket and cannot be undone.
• Objects added to the bucket while the empty bucket action is in progress might be deleted.
• To prevent new objects from being added to this bucket while the empty bucket action is in progress, you might need to update your bucket policy to stop objects from being added to the bucket.

Learn more [\[Link\]](#)

If your bucket contains a large number of objects, creating a lifecycle rule to delete all objects in the bucket might be a more efficient way of emptying your bucket. [Learn more \[Link\]](#)

Permanently delete all objects in bucket "dataartistbucket"?

To confirm deletion, type *permanently delete* in the text input field.

permanently delete

Cancel **Empty**

The screenshot shows the 'Delete bucket' confirmation dialog in the AWS S3 console. At the top, there's a warning message: '⚠ Deleting a bucket cannot be undone.' and 'Bucket names are unique. If you delete a bucket, another AWS user can use the name.' Below this is a text input field containing 'dataartistbucket'. At the bottom are 'Cancel' and 'Delete bucket' buttons.

Feedback Looking for language selection? Find it in the new Unified Settings [?](#)

© 2022, Amazon Web Services, Inc. or its affiliates. [Privacy](#) [Terms](#) [Cookie preferences](#)

Empty bucket: status [Cancel](#) [Exit](#)

The details below are no longer available after you navigate away from this page.

Summary

Source	Successfully deleted	Failed to delete
s3://dataartistbucket	7 objects, 27.9 MB	0 objects

Step 19: Deletion of all created clusters

The screenshot shows the 'Amazon Redshift' service dashboard. The left sidebar includes options like 'Provisioned clusters dashboard', 'Clusters' (selected), 'Query editor', 'Datashares', 'Configurations', 'Advisor', 'AWS Marketplace', and 'Alarms'. The main area displays connection information for Redshift clusters, including a 'Cluster identifier' dropdown, 'Copy JDBC URL' and 'Copy ODBC URL' buttons, and a 'Driver' section for 'JDBC 4.2 without AWS SDK (jar)'. Below this is a 'Clusters (0)' table with columns: Cluster, Cluster namespace, Status, Storage capacity us..., CPU utilization, and Snapshots. A 'Create cluster' button is located at the bottom of the table.