CS 79B Final Project - Part 1

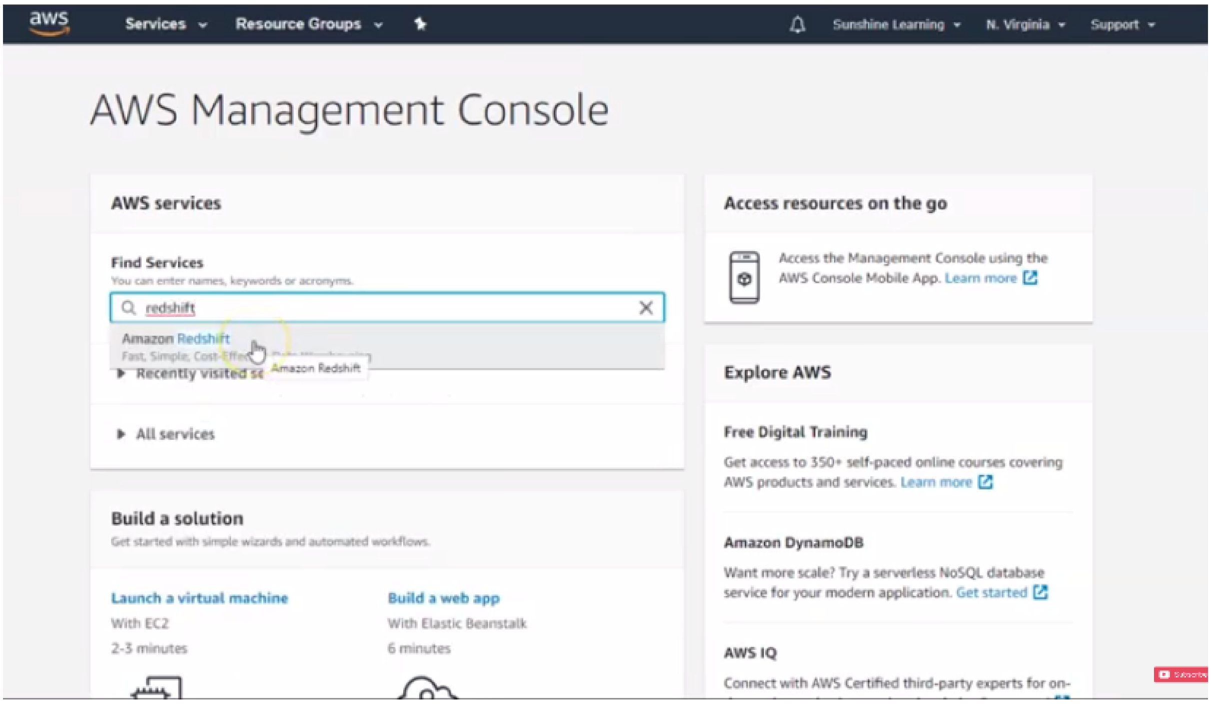
Identify an AWS service from the Database, Analytics or Migration category group not discussed in class from the AWS Console. Make sure that the underlined items are covered in your answer. Create a word or text document and upload it.

The AWS Service selected for the Final Project is AWS Redshift

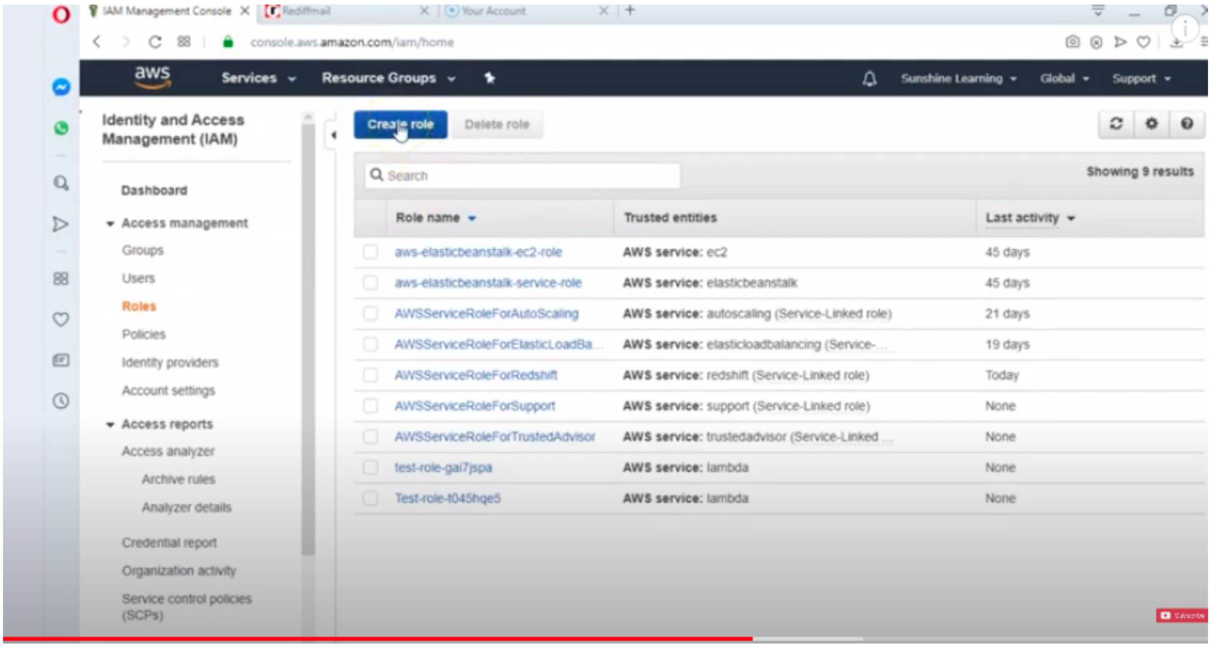
(1) For this service, describe the typical way to interact with the service (that is, a list of actions or steps users' takes to implement this service. Similar to instructions provided to create a MySQL DB, Image Rekognition etc.)

Steps to interact with Amazon Redshift:

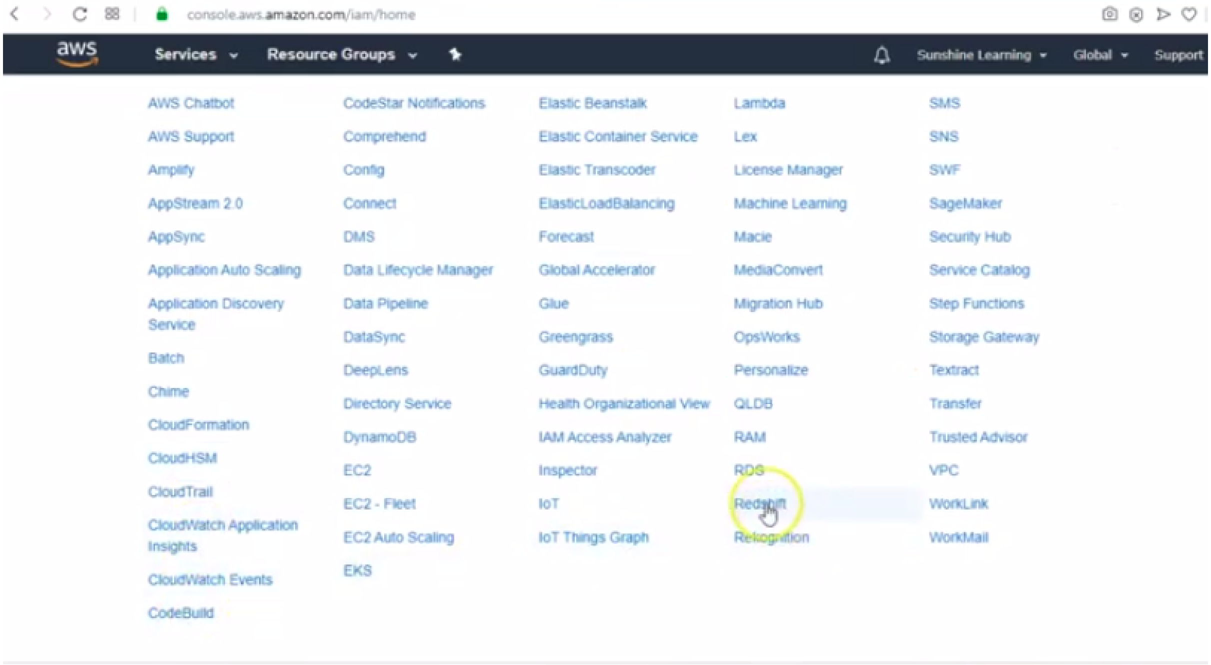
1. *Search for Amazon Redshift in Amazon Management Console*


aws
Services
Resource Groups
Sunshine Learning
N. Virginia
Support
AWS Management Console
AWS services
Find Services
You can enter names, keywords or acronyms.
redshift
Amazon Redshift
Fast, Simple, Cost-Effec
Amazon Redshift
Recently visited se
All services
Build a solution
Get started with simple wizards and automated workflows.
Launch a virtual machine
With EC2
2-3 minutes
Build a web app
With Elastic Beanstalk
6 minutes
Access resources on the go
Access the Management Console using the AWS Console Mobile App. Learn more
Explore AWS
Free Digital Training
Get access to 350+ self-paced online courses covering AWS products and services. Learn more
Amazon DynamoDB
Want more scale? Try a serverless NoSQL database service for your modern application. Get started
AWS IQ
Connect with AWS Certified third-party experts for on-

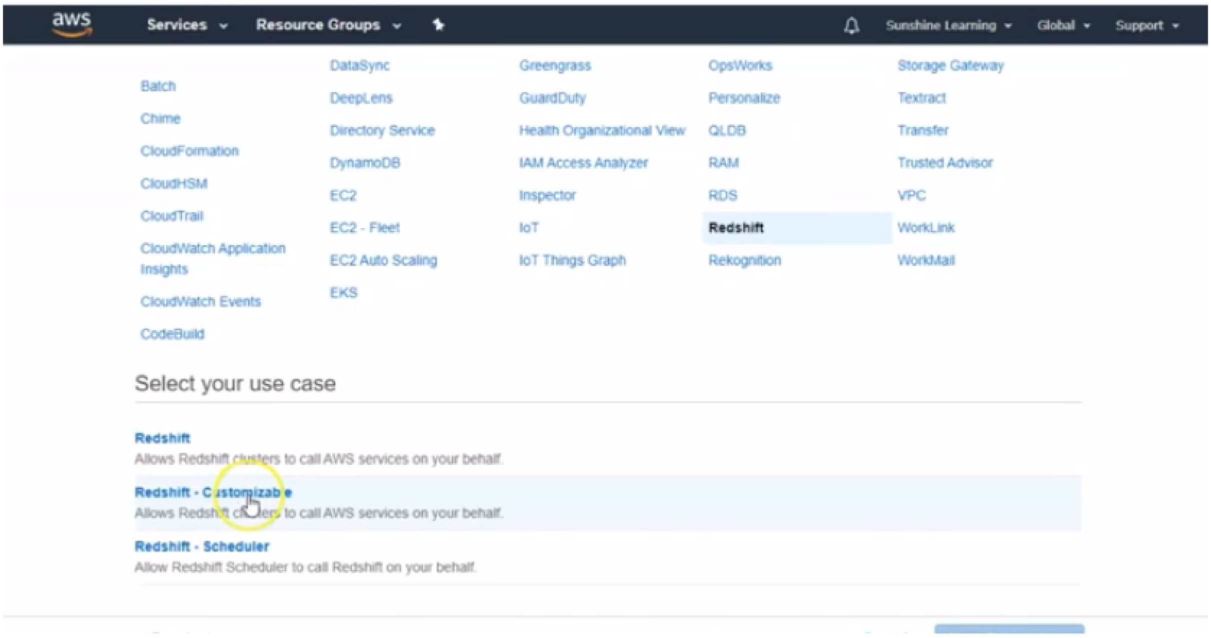
2. *Before creating Cluster, we need to* ***create an IAM role****. Select* ***Redshift*** *under IAM Role because Redshift will be calling S3 service. We will putting up data in S3 and then it will be uploaded to Redshift*



Role name	Trusted entities	Last activity
aws-elasticbeanstalk-ec2-role	AWS service: ec2	45 days
aws-elasticbeanstalk-service-role	AWS service: elasticbeanstalk	45 days
AWSServiceRoleForAutoScaling	AWS service: autoscaling (Service-Linked role)	21 days
AWSServiceRoleForElasticLoadBa...	AWS service: elasticloadbalancing (Service-...	19 days
AWSServiceRoleForRedshift	AWS service: redshift (Service-Linked role)	Today
AWSServiceRoleForSupport	AWS service: support (Service-Linked role)	None
AWSServiceRoleForTrustedAdvisor	AWS service: trustedadvisor (Service-Linked ...	None
test-role-gai7jspa	AWS service: lambda	None
Test-role-t045hqe5	AWS service: lambda	None



3. *Click on **Redshift Customizable** in the Use Case*



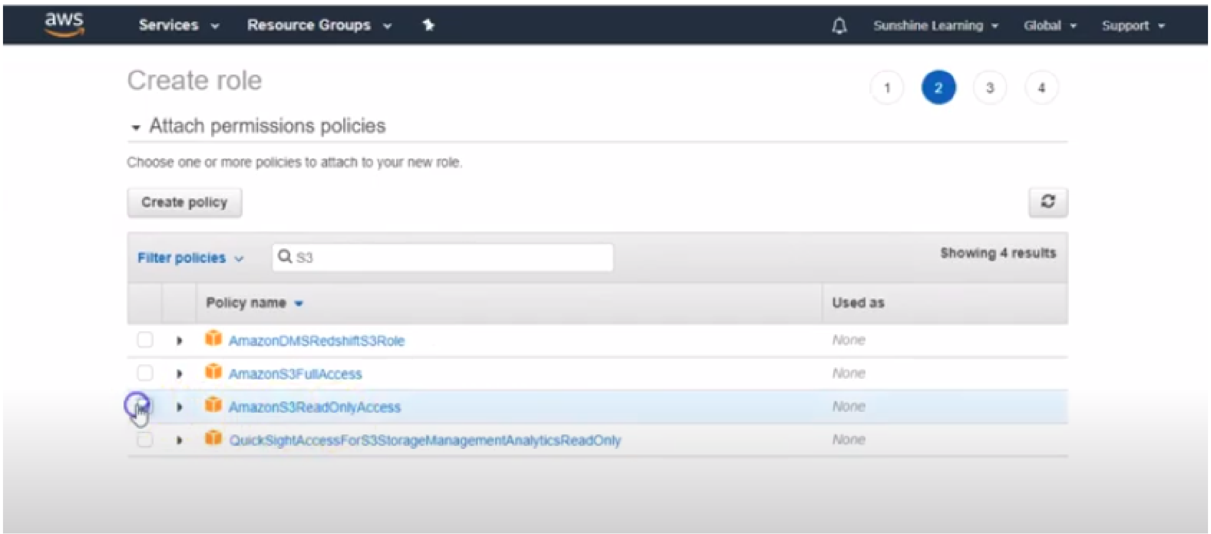
Select your use case

Redshift
Allows Redshift clusters to call AWS services on your behalf.

Redshift - Customizable
Allows Redshift clusters to call AWS services on your behalf.

Redshift - Scheduler
Allow Redshift Scheduler to call Redshift on your behalf.

4. *Under Policy Permissions, select **AMAZONS3READONLYACCESS** and assign permissions to this Role*



Create role

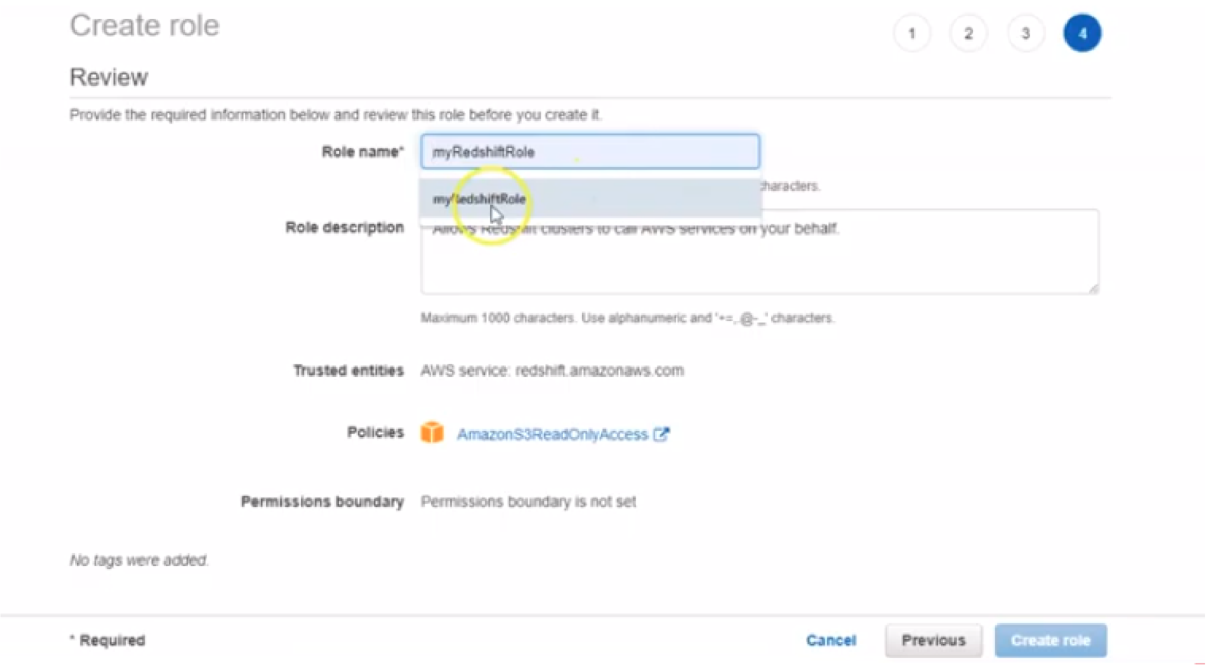
Attach permissions policies

Choose one or more policies to attach to your new role.

Showing 4 results

	Policy name	Used as
□	AmazonDMSRedshiftS3Role	None
□	AmazonS3FullAccess	None
☑	AmazonS3ReadOnlyAccess	None
□	QuickSightAccessForS3StorageManagementAnalyticsReadOnly	None

5. *Enter **Role** name and click on **Create Role***



Create role

Review

Provide the required information below and review this role before you create it.

Role name* myRedshiftRole

myRedshiftRole

characters.

Role description Allows Redshift clusters to call AWS services on your behalf.

Maximum 1000 characters. Use alphanumeric and '+=,.@-_' characters.

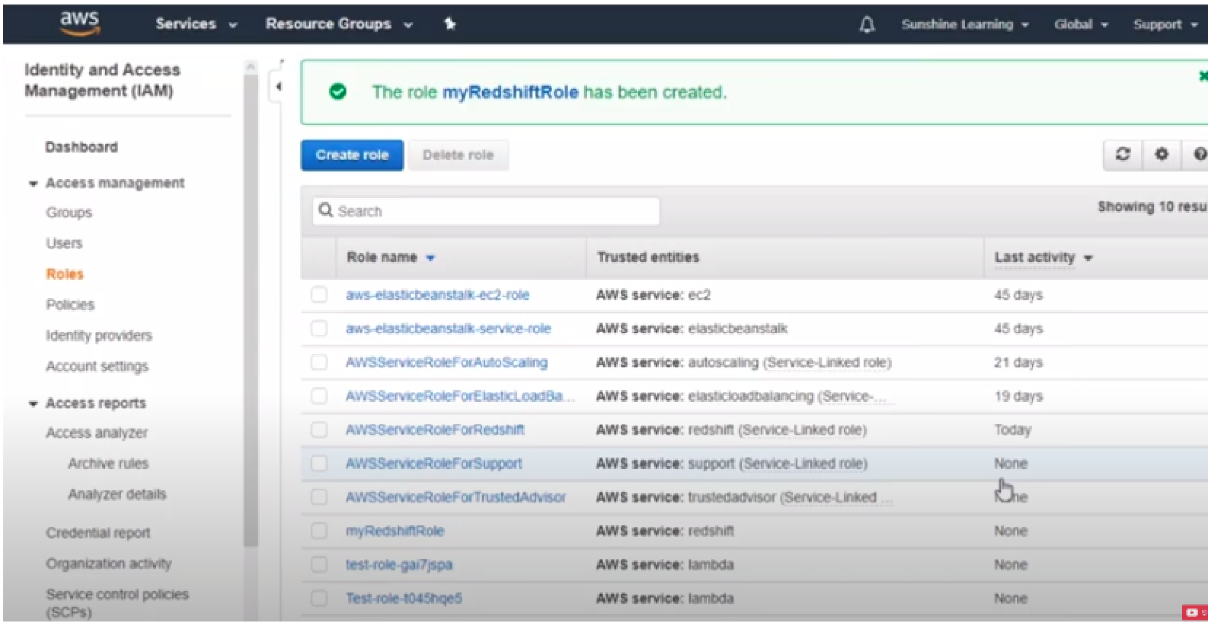
Trusted entities AWS service: redshift.amazonaws.com

Policies AmazonS3ReadOnlyAccess

Permissions boundary Permissions boundary is not set

No tags were added.

* Required

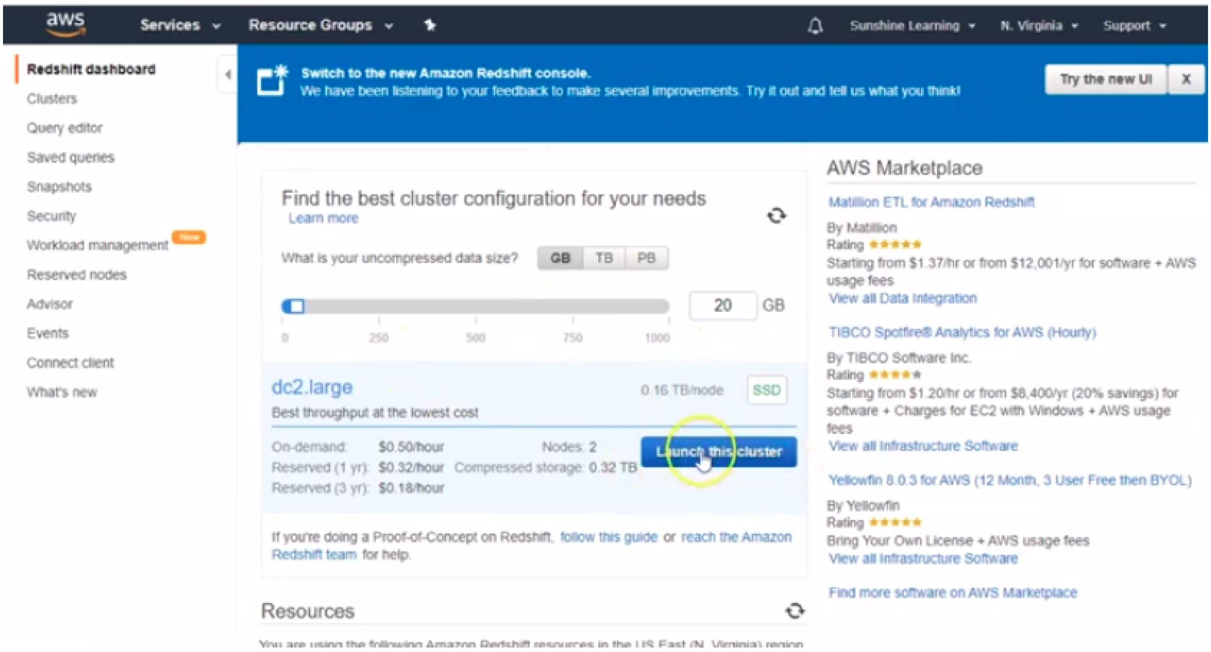


The role **myRedshiftRole** has been created.

Showing 10 resu

Role name	Trusted entities	Last activity
aws-elasticbeanstalk-ec2-role	**AWS service:** ec2	45 days
aws-elasticbeanstalk-service-role	**AWS service:** elasticbeanstalk	45 days
AWSServiceRoleForAutoScaling	**AWS service:** autoscaling (Service-Linked role)	21 days
AWSServiceRoleForElasticLoadBa...	**AWS service:** elasticloadbalancing (Service-...	19 days
AWSServiceRoleForRedshift	**AWS service:** redshift (Service-Linked role)	Today
AWSServiceRoleForSupport	**AWS service:** support (Service-Linked role)	None
AWSServiceRoleForTrustedAdvisor	**AWS service:** trustedadvisor (Service-Linked ...	None
myRedshiftRole	**AWS service:** redshift	None
test-role-gai7jspa	**AWS service:** lambda	None
Test-role-t045hqe5	**AWS service:** lambda	None

6. *Next, go to **Redshift** and create a **Cluster.** Also select the **uncompressed data size**: GB/TB/PB*



Switch to the new Amazon Redshift console.
We have been listening to your feedback to make several improvements. Try it out and tell us what you think!

Find the best cluster configuration for your needs
Learn more

What is your uncompressed data size? GB

20 GB

dc2.large 0.16 TB/node
Best throughput at the lowest cost

On-demand: $0.50/hour Nodes: 2
Reserved (1 yr): $0.32/hour Compressed storage: 0.32 TB
Reserved (3 yr): $0.18/hour

If you're doing a Proof-of-Concept on Redshift, follow this guide or reach the Amazon Redshift team for help.

Resources
You are using the following Amazon Redshift resources in the US East (N. Virginia) region

AWS Marketplace

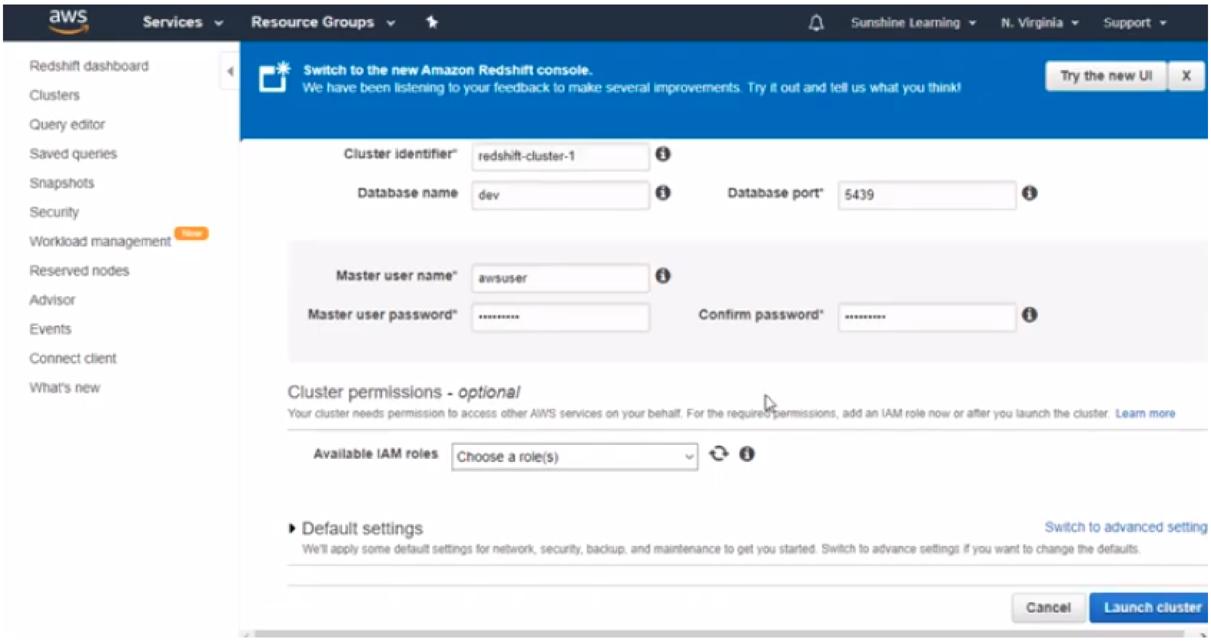
Matillion ETL for Amazon Redshift
By Matillion
Rating ★★★★★
Starting from $1.37/hr or from $12,001/yr for software + AWS usage fees
View all Data Integration

TIBCO Spotfire® Analytics for AWS (Hourly)
By TIBCO Software Inc.
Rating ★★★★★
Starting from $1.20/hr or from $8,400/yr (20% savings) for software + Charges for EC2 with Windows + AWS usage fees
View all Infrastructure Software

Yellowfin 8.0.3 for AWS (12 Month, 3 User Free then BYOL)
By Yellowfin
Rating ★★★★★
Bring Your Own License + AWS usage fees
View all Infrastructure Software

Find more software on AWS Marketplace

7. *Give **Cluster name, No: of Nodes, Master Username and Password** for Redshift Cluster*



Switch to the new Amazon Redshift console.
We have been listening to your feedback to make several improvements. Try it out and tell us what you think!

Cluster identifier* redshift-cluster-1

Database name dev

Database port* 5439

Master user name* awsuser

Master user password* ••••••••••

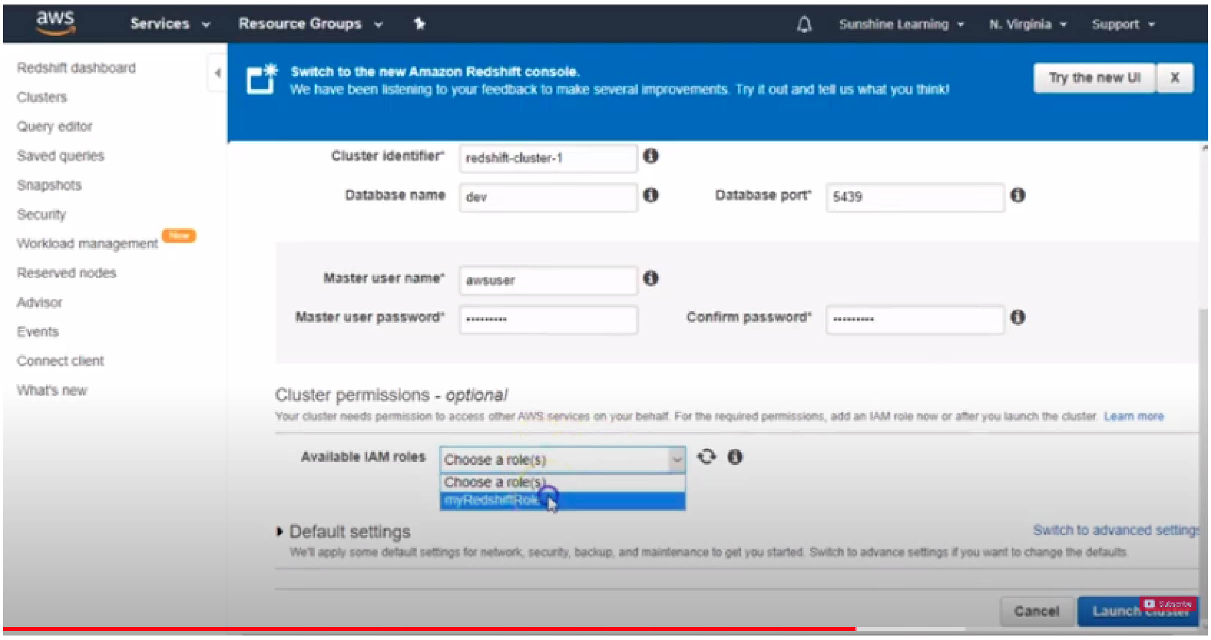
Confirm password* ••••••••••

Cluster permissions - *optional*
Your cluster needs permission to access other AWS services on your behalf. For the required permissions, add an IAM role now or after you launch the cluster. Learn more

Available IAM roles Choose a role(s)

Default settings
We'll apply some default settings for network, security, backup, and maintenance to get you started. Switch to advance settings if you want to change the defaults.

Switch to advanced setting



Switch to the new Amazon Redshift console.
We have been listening to your feedback to make several improvements. Try it out and tell us what you think!

Cluster identifier* redshift-cluster-1

Database name dev

Database port* 5439

Master user name* awsuser

Master user password* ●●●●●●●●●

Confirm password* ●●●●●●●●●

Cluster permissions - *optional*

Your cluster needs permission to access other AWS services on your behalf. For the required permissions, add an IAM role now or after you launch the cluster. Learn more

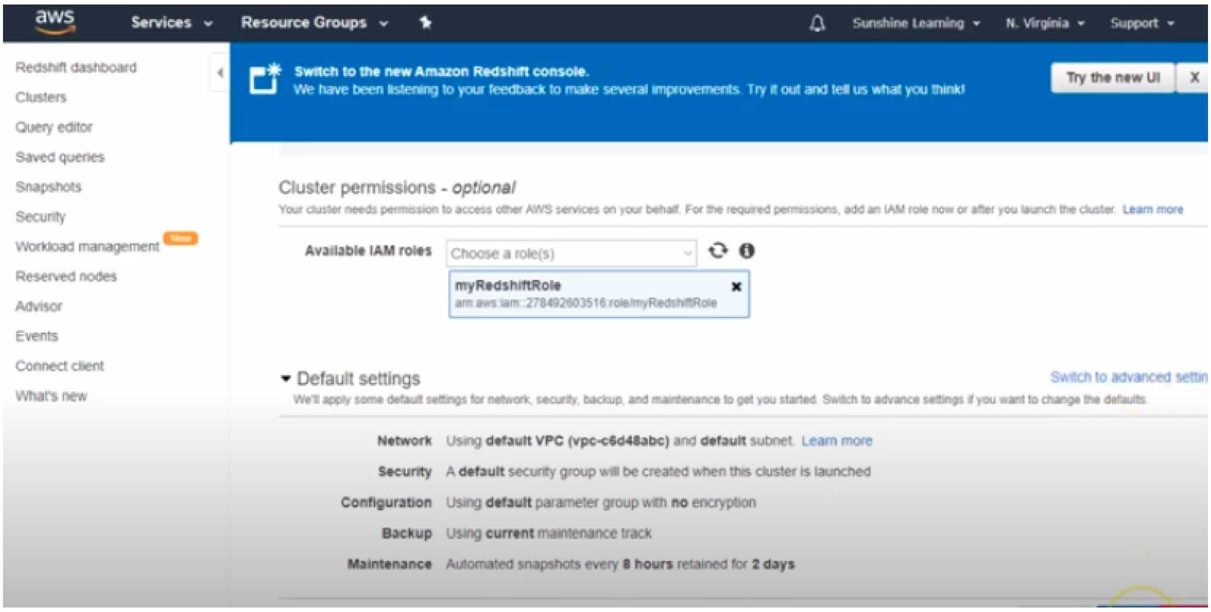
Available IAM roles Choose a role(s)
Choose a role(s)
myRedshiftRole

Default settings

We'll apply some default settings for network, security, backup, and maintenance to get you started. Switch to advance settings if you want to change the defaults.

Switch to advanced settings

8. *You can change **Default VPC** from Default Settings*



Switch to the new Amazon Redshift console.
We have been listening to your feedback to make several improvements. Try it out and tell us what you think!

Cluster permissions - *optional*

Your cluster needs permission to access other AWS services on your behalf. For the required permissions, add an IAM role now or after you launch the cluster. Learn more

Available IAM roles Choose a role(s)

myRedshiftRole
arn:aws:iam::278492603516:role/myRedshiftRole

Default settings

We'll apply some default settings for network, security, backup, and maintenance to get you started. Switch to advance settings if you want to change the defaults.

Switch to advanced settin

Network Using **default VPC (vpc-c6d48abc)** and **default** subnet. Learn more

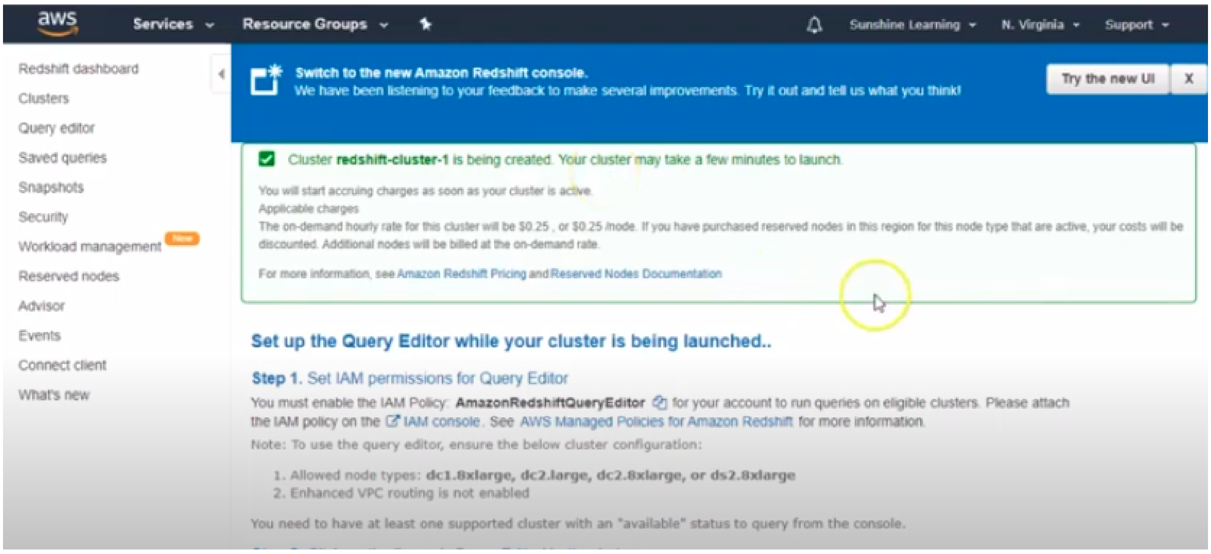
Security A **default** security group will be created when this cluster is launched

Configuration Using **default** parameter group with **no** encryption

Backup Using **current** maintenance track

Maintenance Automated snapshots every **8 hours** retained for **2 days**

9. ***Redshift Cluster** is being created*



Switch to the new Amazon Redshift console.
We have been listening to your feedback to make several improvements. Try it out and tell us what you think!

☑ Cluster **redshift-cluster-1** is being created. Your cluster may take a few minutes to launch.

You will start accruing charges as soon as your cluster is active.
Applicable charges
The on-demand hourly rate for this cluster will be $0.25 , or $0.25 /node. If you have purchased reserved nodes in this region for this node type that are active, your costs will be discounted. Additional nodes will be billed at the on-demand rate.

For more information, see Amazon Redshift Pricing and Reserved Nodes Documentation

Set up the Query Editor while your cluster is being launched..

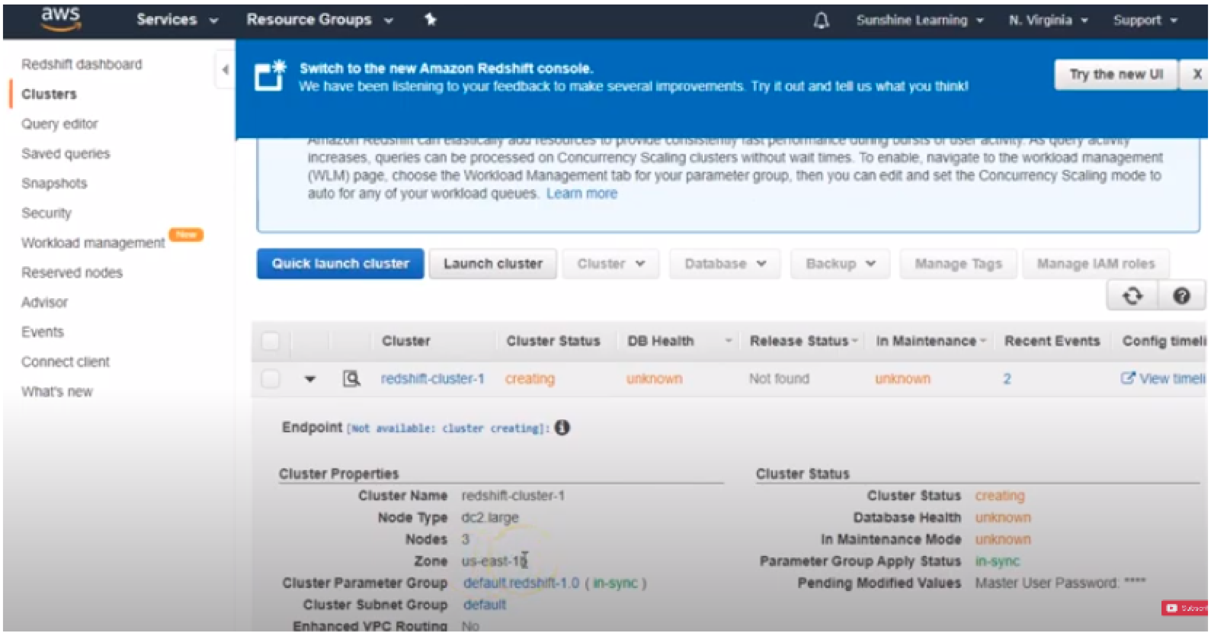
Step 1. Set IAM permissions for Query Editor

You must enable the IAM Policy: **AmazonRedshiftQueryEditor** for your account to run queries on eligible clusters. Please attach the IAM policy on the IAM console. See AWS Managed Policies for Amazon Redshift for more information.

Note: To use the query editor, ensure the below cluster configuration:

1. Allowed node types: **dc1.8xlarge, dc2.large, dc2.8xlarge, or ds2.8xlarge**
2. Enhanced VPC routing is not enabled

You need to have at least one supported cluster with an "available" status to query from the console.



Switch to the new Amazon Redshift console.
We have been listening to your feedback to make several improvements. Try it out and tell us what you think!

Amazon Redshift can elastically add resources to provide consistently fast performance during bursts of user activity. As query activity increases, queries can be processed on Concurrency Scaling clusters without wait times. To enable, navigate to the workload management (WLM) page, choose the Workload Management tab for your parameter group, then you can edit and set the Concurrency Scaling mode to auto for any of your workload queues. Learn more

Cluster	Cluster Status	DB Health	Release Status	In Maintenance	Recent Events	Config timeli
redshift-cluster-1	creating	unknown	Not found	unknown	2	View timeli

Endpoint [Not available: cluster creating]:

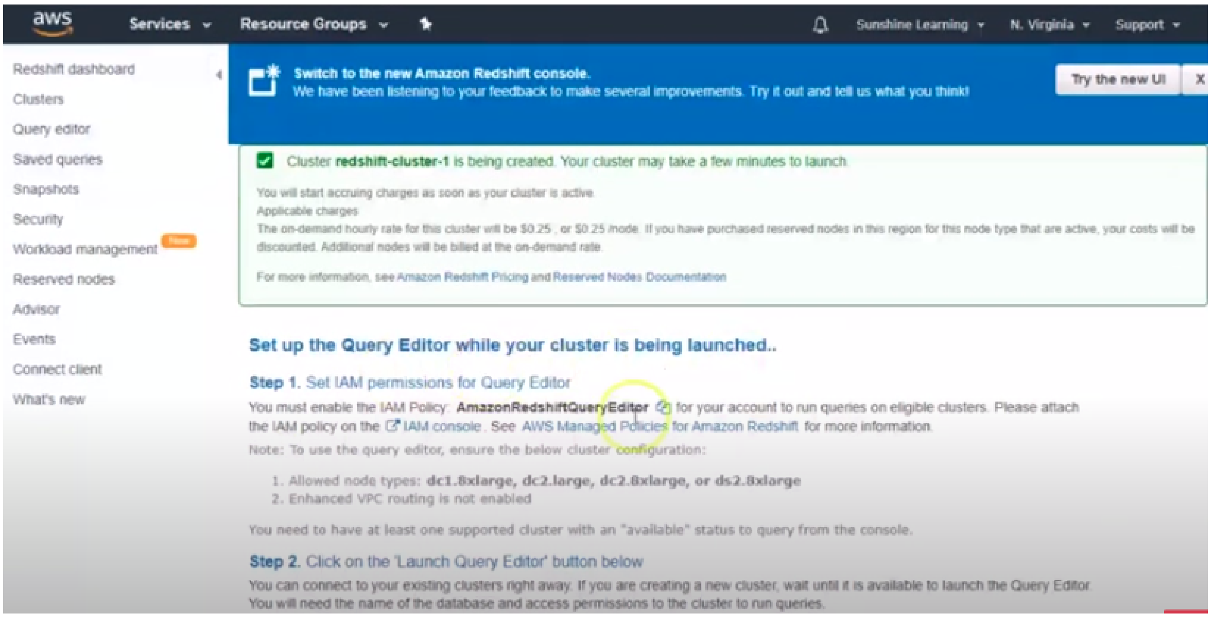
Cluster Properties

Cluster Name	redshift-cluster-1
Node Type	dc2.large
Nodes	3
Zone	us-east-1b
Cluster Parameter Group	default.redshift-1.0 (in-sync)
Cluster Subnet Group	default
Enhanced VPC Routing	No

Cluster Status

Cluster Status	creating
Database Health	unknown
In Maintenance Mode	unknown
Parameter Group Apply Status	in-sync
Pending Modified Values	Master User Password: ****

*10. To run SQL queries, you can use **Query Editor***



Switch to the new Amazon Redshift console.
We have been listening to your feedback to make several improvements. Try it out and tell us what you think!

☑ Cluster **redshift-cluster-1** is being created. Your cluster may take a few minutes to launch.

You will start accruing charges as soon as your cluster is active.
Applicable charges
The on-demand hourly rate for this cluster will be $0.25 , or $0.25 /node. If you have purchased reserved nodes in this region for this node type that are active, your costs will be discounted. Additional nodes will be billed at the on-demand rate.

For more information, see Amazon Redshift Pricing and Reserved Nodes Documentation

Set up the Query Editor while your cluster is being launched..

Step 1. Set IAM permissions for Query Editor

You must enable the IAM Policy: **AmazonRedshiftQueryEditor** for your account to run queries on eligible clusters. Please attach the IAM policy on the IAM console. See AWS Managed Policies for Amazon Redshift for more information.

Note: To use the query editor, ensure the below cluster configuration:

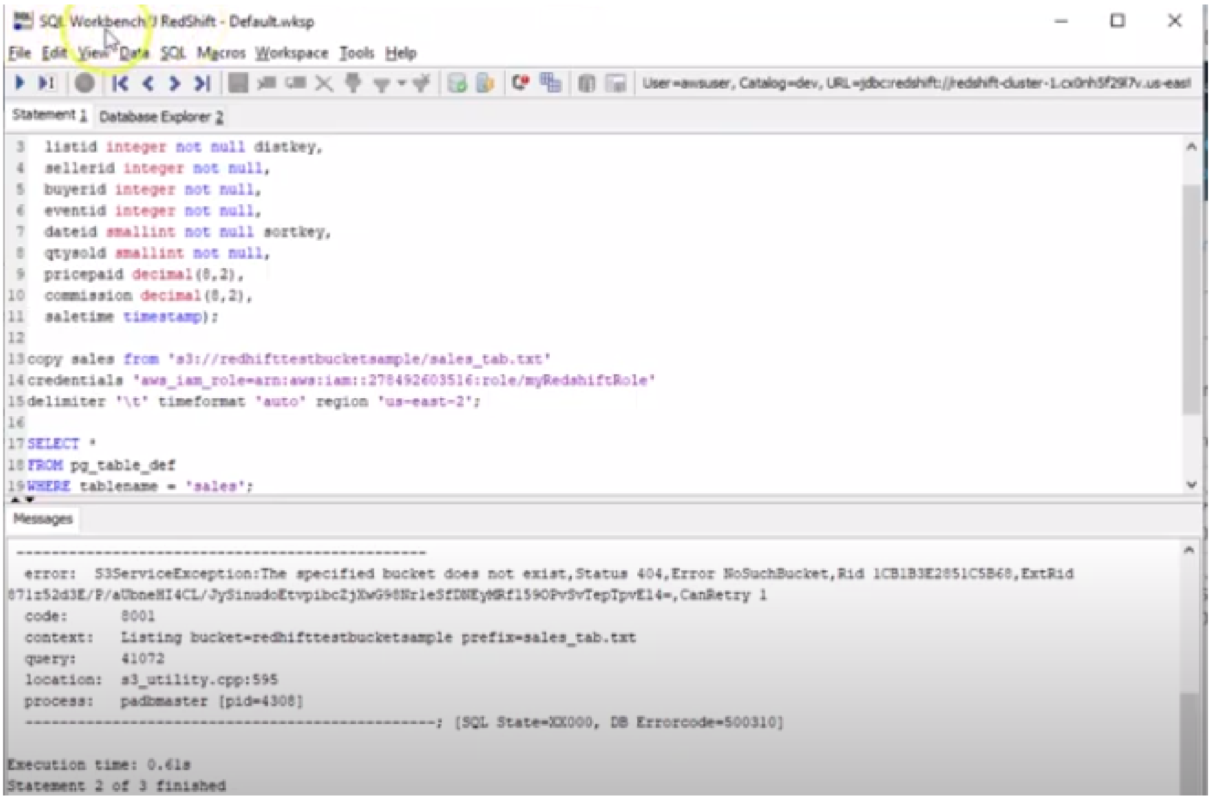
1. Allowed node types: **dc1.8xlarge, dc2.large, dc2.8xlarge, or ds2.8xlarge**
2. Enhanced VPC routing is not enabled

You need to have at least one supported cluster with an "available" status to query from the console.

Step 2. Click on the 'Launch Query Editor' button below

You can connect to your existing clusters right away. If you are creating a new cluster, wait until it is available to launch the Query Editor. You will need the name of the database and access permissions to the cluster to run queries.

11. *To create Tables in Redshift, use **SQL Workbench/J** client to connect to **Redshift DB***

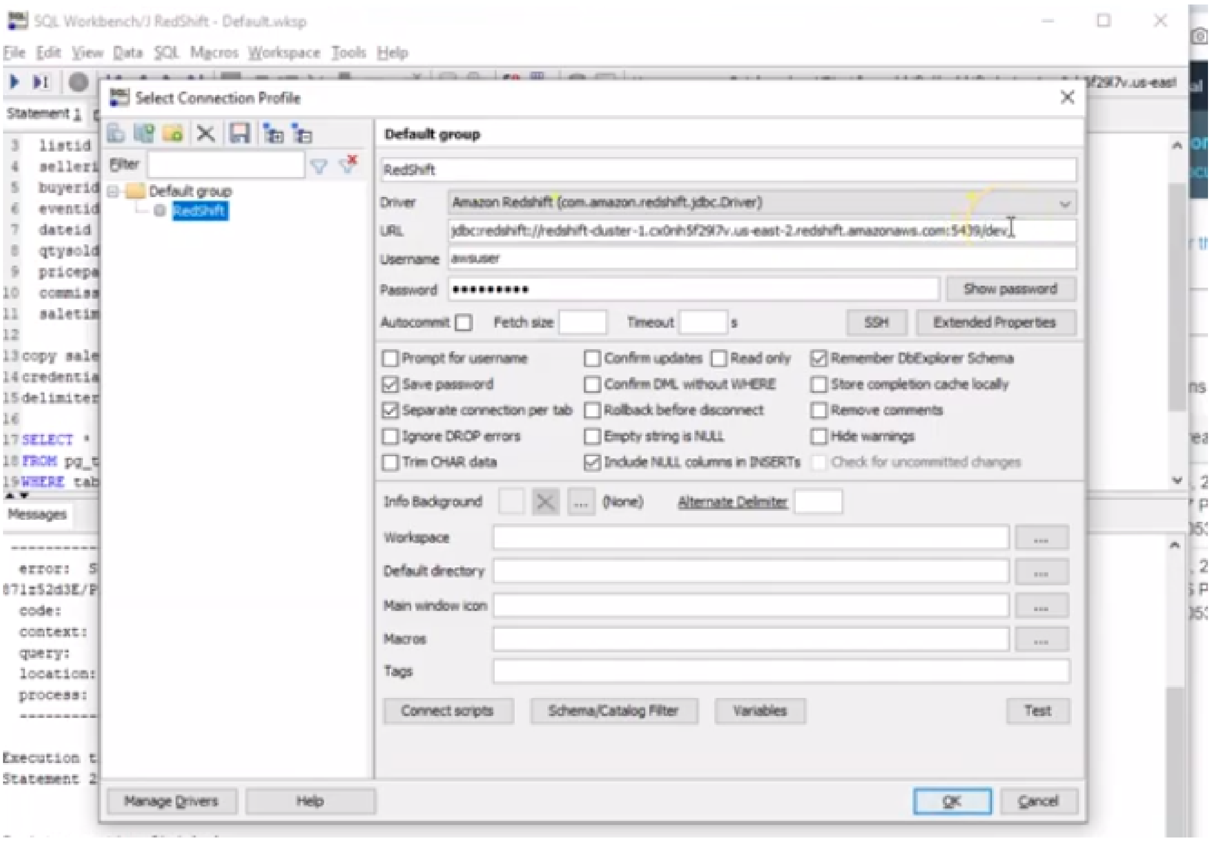


User=awsuser, Catalog=dev, URL=jdbc:redshift://redshift-cluster-1.cx0nh5f297v.us-east

```
 listid integer not null distkey,
 sellerid integer not null,
 buyerid integer not null,
 eventid integer not null,
 dateid smallint not null sortkey,
 qtysold smallint not null,
 pricepaid decimal(8,2),
 commission decimal(8,2),
 saletime timestamp);

copy sales from 's3://redhifttestbucketsample/sales_tab.txt'
credentials 'aws_iam_role=arn:aws:iam::278492603516:role/myRedshiftRole'
delimiter '\t' timeformat 'auto' region 'us-east-2';

SELECT *
FROM pg_table_def
WHERE tablename = 'sales';
```

```
 -----------------------------------------------
  error:  S3ServiceException:The specified bucket does not exist,Status 404,Error NoSuchBucket,Rid 1CB1B3E2851C5B68,ExtRid
871z52d3E/P/aUbneHI4CL/JySinudoEtvpibcZjXwG98Nr1eSfDNEyMRf159OPvSvTepTpvE14=,CanRetry 1
  code:      8001
  context:   Listing bucket=redhifttestbucketsample prefix=sales_tab.txt
  query:     41072
  location:  s3_utility.cpp:595
  process:   padbmaster [pid=4308]
  -----------------------------------------------; [SQL State=XX000, DB Errorcode=500310]

Execution time: 0.61s
Statement 2 of 3 finished
```

12. *Connect to Redshift DB using **JDBC URL** with **PORT 5439.** Select the **Driver, Username and Password***


SQL Workbench/J RedShift - Default.wksp

Select Connection Profile

Default group

RedShift

Driver: Amazon Redshift (com.amazon.redshift.jdbc.Driver)

URL: jdbc:redshift://redshift-cluster-1.cx0nh5f2917v.us-east-2.redshift.amazonaws.com:5439/dev

Username: awsuser

Password: ••••••••••

Autocommit □ Fetch size Timeout s

□ Prompt for username	□ Confirm updates □ Read only	☑ Remember DbExplorer Schema
☑ Save password	□ Confirm DML without WHERE	□ Store completion cache locally
☑ Separate connection per tab	□ Rollback before disconnect	□ Remove comments
□ Ignore DROP errors	□ Empty string is NULL	□ Hide warnings
□ Trim CHAR data	☑ Include NULL columns in INSERTs	□ Check for uncommitted changes

Info Background (None) Alternate Delimiter

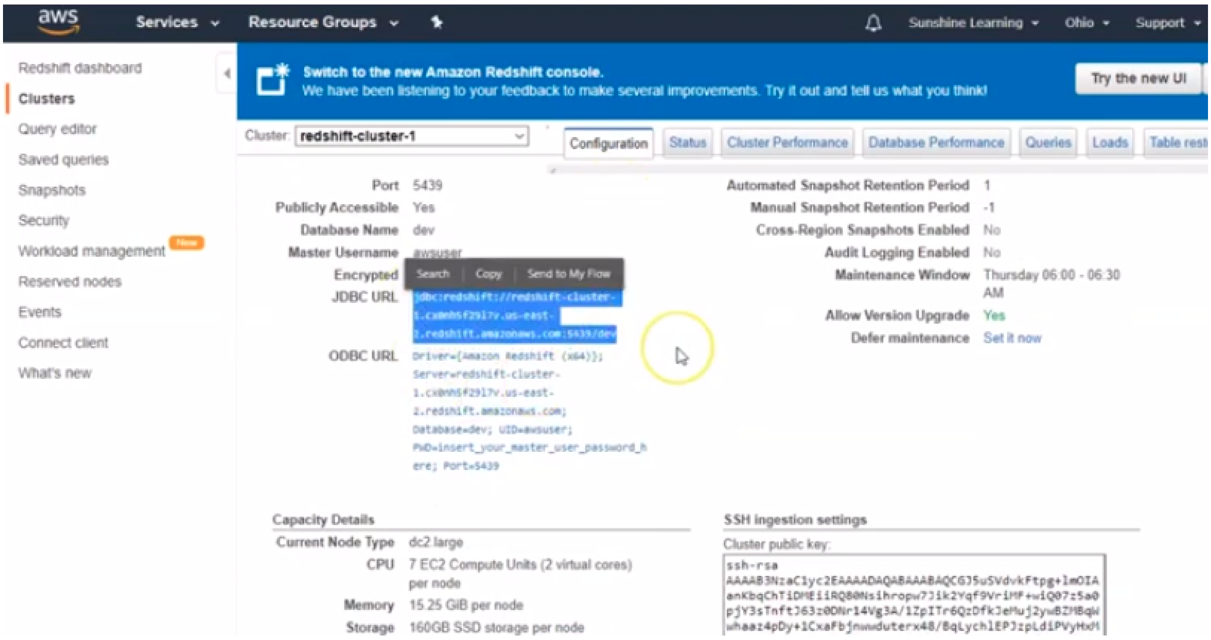
Workspace

Default directory

Main window icon

Macros

Tags



Switch to the new Amazon Redshift console.
We have been listening to your feedback to make several improvements. Try it out and tell us what you think!

Cluster: redshift-cluster-1

Port	5439
Publicly Accessible	Yes
Database Name	dev
Master Username	awsuser
Encrypted	
JDBC URL	jdbc:redshift://redshift-cluster-1.cxmnhsf2917v.us-east-2.redshift.amazonaws.com:5439/dev
ODBC URL	Driver={Amazon Redshift (x64)}; Server=redshift-cluster-1.cxmnhsf2917v.us-east-2.redshift.amazonaws.com; Database=dev; UID=awsuser; PWD=insert_your_master_user_password_here; Port=5439

Automated Snapshot Retention Period	1
Manual Snapshot Retention Period	-1
Cross-Region Snapshots Enabled	No
Audit Logging Enabled	No
Maintenance Window	Thursday 06:00 - 06:30 AM
Allow Version Upgrade	Yes
Defer maintenance	Set it now

Capacity Details

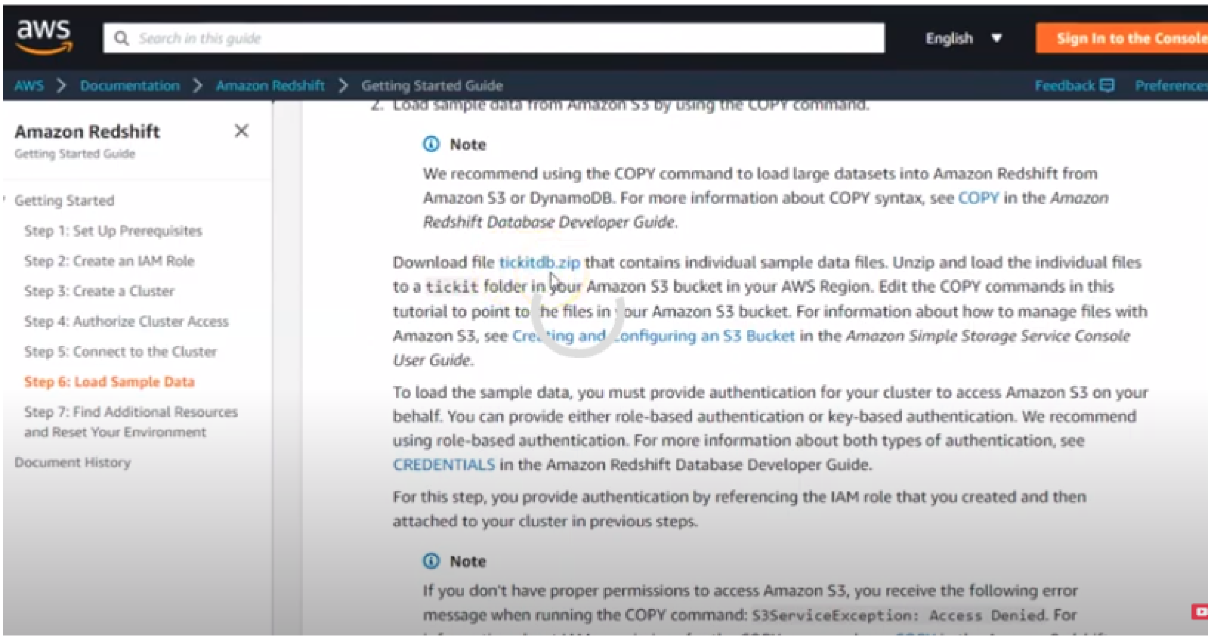
Current Node Type	dc2.large
CPU	7 EC2 Compute Units (2 virtual cores) per node
Memory	15.25 GiB per node
Storage	160GB SSD storage per node

SSH ingestion settings

Cluster public key:

```
ssh-rsa
AAAAB3NzaC1yc2EAAAADAQABAAABAQCGJ5uSVdvkFtpg+lmOIA
anKbqChTiDMEiiRQ80Nsihropw7Jik2Yqf9VriMF+wiQ07z5a0
pjY3sTnftJ63z0DNr14Vg3A/1ZpITr6QzDfkJeMuj2ywBZMBqW
whaaz4pDy+1CxaFbjnwwduterx48/BqLychlEPJzpLdiPVyHxM
```

13. *Download sample data from below location and load it to* ***S3***


aws
Search in this guide
English
Sign In to the Console
AWS > Documentation > Amazon Redshift > Getting Started Guide
Feedback
Preferences
Amazon Redshift
Getting Started Guide
Getting Started
Step 1: Set Up Prerequisites
Step 2: Create an IAM Role
Step 3: Create a Cluster
Step 4: Authorize Cluster Access
Step 5: Connect to the Cluster
Step 6: Load Sample Data
Step 7: Find Additional Resources and Reset Your Environment
Document History

2. Load sample data from Amazon S3 by using the COPY command.

Note

We recommend using the COPY command to load large datasets into Amazon Redshift from Amazon S3 or DynamoDB. For more information about COPY syntax, see COPY in the *Amazon Redshift Database Developer Guide*.

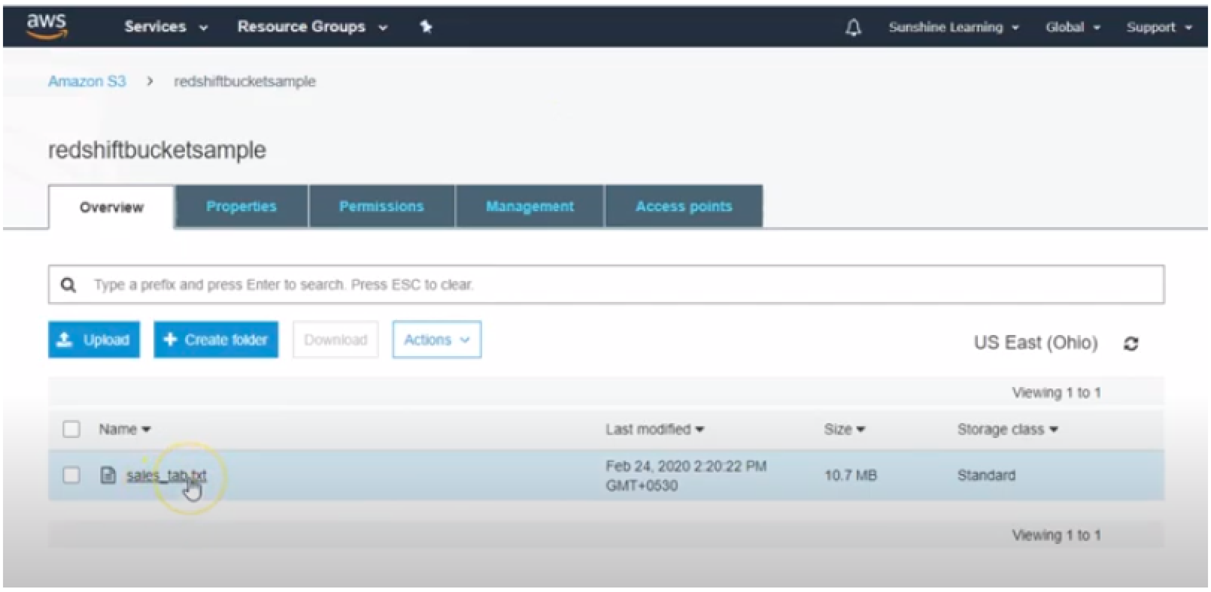
Download file tickitdb.zip that contains individual sample data files. Unzip and load the individual files to a `tickit` folder in your Amazon S3 bucket in your AWS Region. Edit the COPY commands in this tutorial to point to the files in your Amazon S3 bucket. For information about how to manage files with Amazon S3, see Creating and configuring an S3 Bucket in the *Amazon Simple Storage Service Console User Guide*.

To load the sample data, you must provide authentication for your cluster to access Amazon S3 on your behalf. You can provide either role-based authentication or key-based authentication. We recommend using role-based authentication. For more information about both types of authentication, see CREDENTIALS in the Amazon Redshift Database Developer Guide.

For this step, you provide authentication by referencing the IAM role that you created and then attached to your cluster in previous steps.

Note

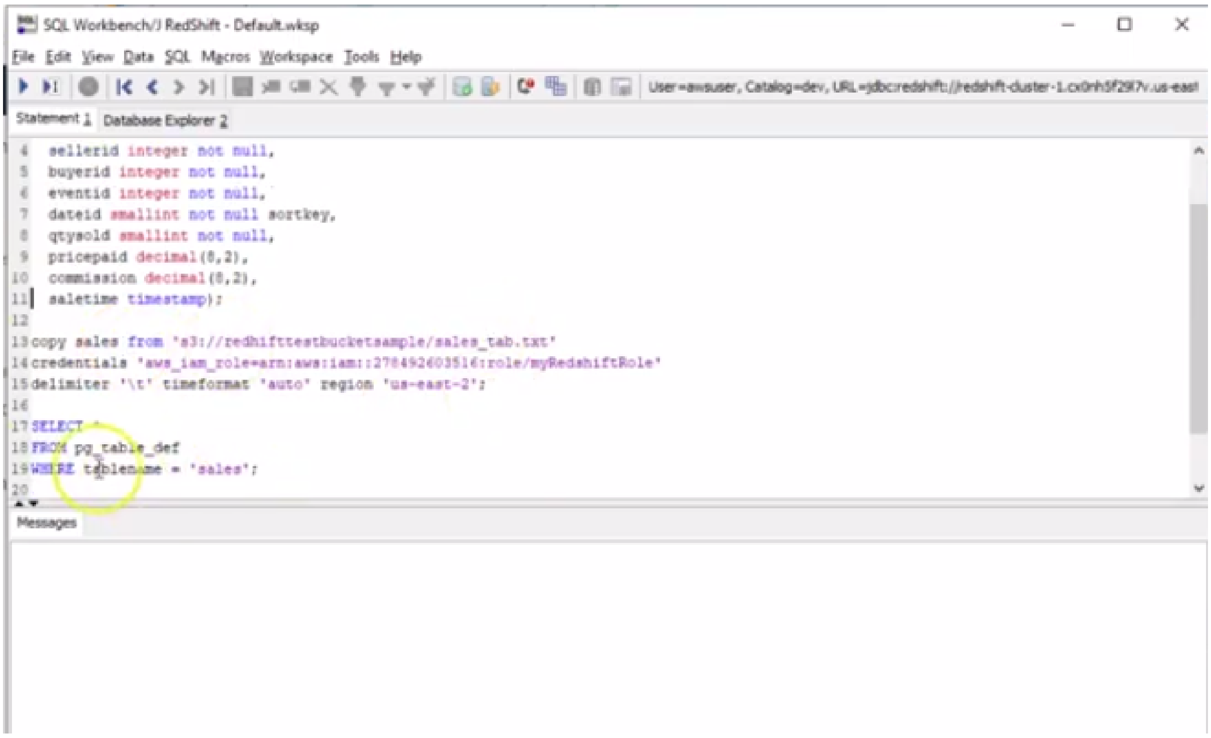
If you don't have proper permissions to access Amazon S3, you receive the following error message when running the COPY command: `S3ServiceException: Access Denied`. For



redshiftbucketsample

Name	Last modified	Size	Storage class
sales_tab.txt	Feb 24, 2020 2:20:22 PM GMT+0530	10.7 MB	Standard

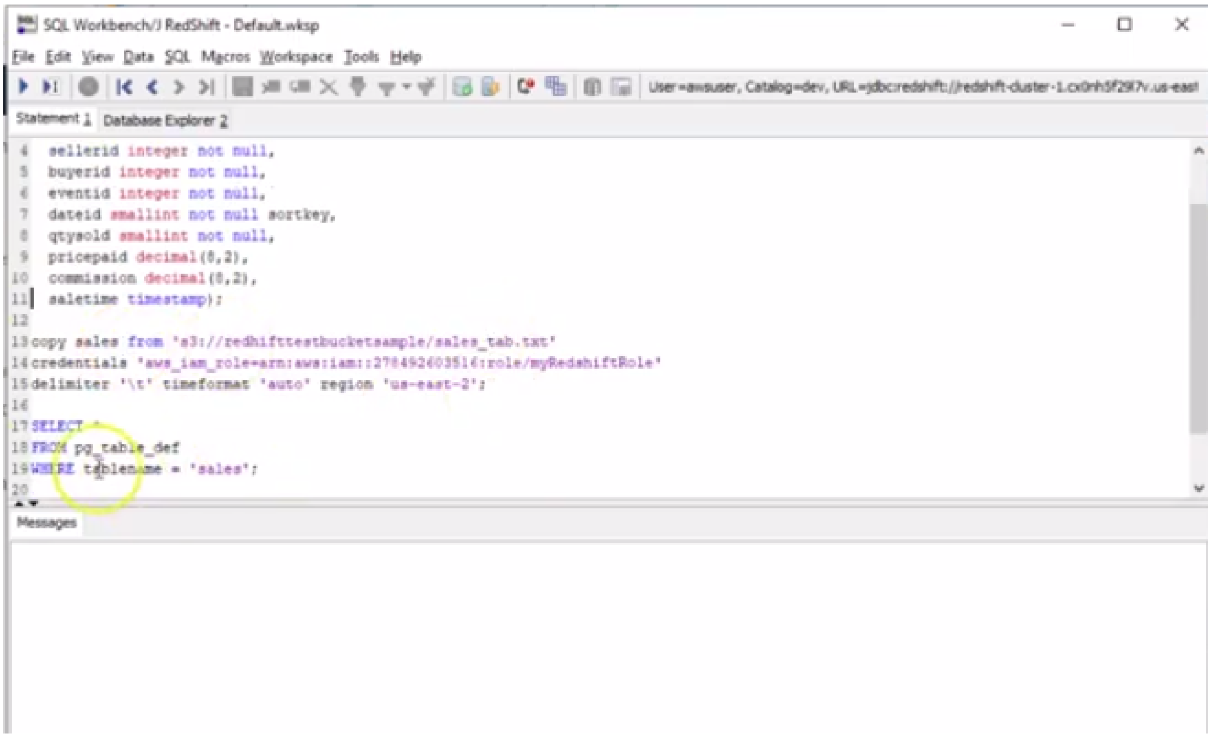
14. *Write queries to **Create** table, Use **copy** statement to copy file from S3 to Redshift DB, Use **Select** statement to query the results from Sales table*



SQL Workbench/J RedShift - Default.wksp

```
 sellerid integer not null,
 buyerid integer not null,
 eventid integer not null,
 dateid smallint not null sortkey,
 qtysold smallint not null,
 pricepaid decimal(8,2),
 commission decimal(8,2),
 saletime timestamp);

copy sales from 's3://redhifttestbucketsample/sales_tab.txt'
credentials 'aws_iam_role=arn:aws:iam::278492603516:role/myRedshiftRole'
delimiter '\t' timeformat 'auto' region 'us-east-2';

SELECT *
FROM pg_table_def
WHERE tablename = 'sales';
```

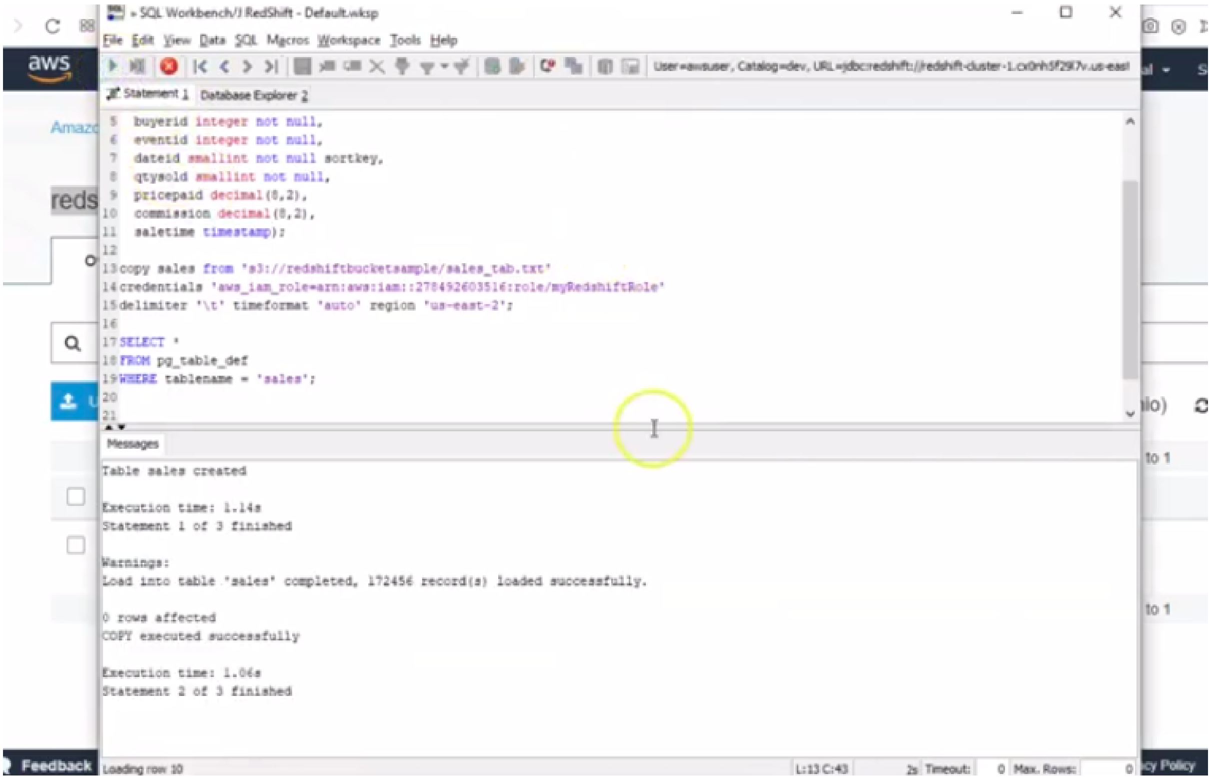
15. *In **copy** statement **credentials**, define the **Role ARM***



SQL Workbench/J RedShift - Default.wksp

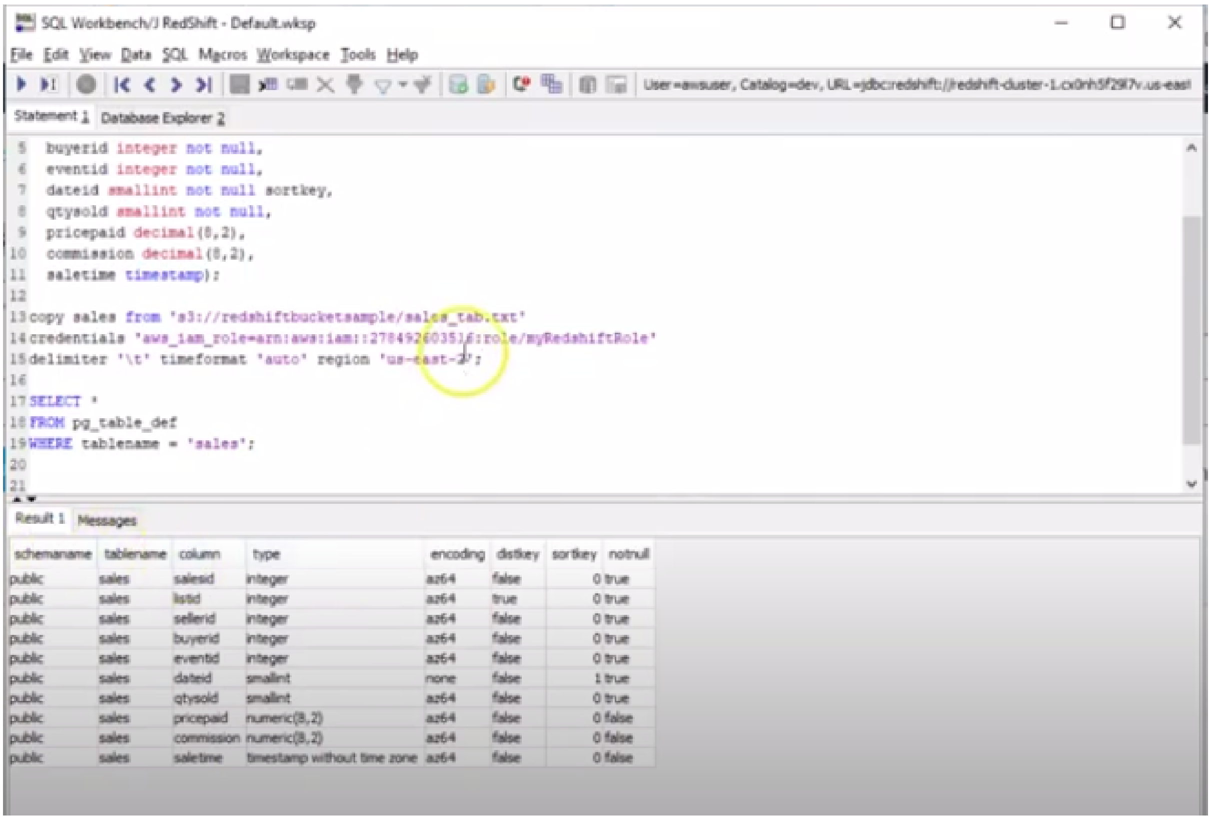
```
 sellerid integer not null,
 buyerid integer not null,
 eventid integer not null,
 dateid smallint not null sortkey,
 qtysold smallint not null,
 pricepaid decimal(8,2),
 commission decimal(8,2),
 saletime timestamp);

copy sales from 's3://redhifttestbucketsample/sales_tab.txt'
credentials 'aws_iam_role=arn:aws:iam::278492603516:role/myRedshiftRole'
delimiter '\t' timeformat 'auto' region 'us-east-2';

SELECT *
FROM pg_table_def
WHERE tablename = 'sales';
```

16. *Execute all these 3 queries. You should be able to Connect to Database and display the tables in Amazon Redshift*



```
  buyerid integer not null,
  eventid integer not null,
  dateid smallint not null sortkey,
  qtysold smallint not null,
  pricepaid decimal(8,2),
  commission decimal(8,2),
  saletime timestamp);

copy sales from 's3://redshiftbucketsample/sales_tab.txt'
credentials 'aws_iam_role=arn:aws:iam::278492603516:role/myRedshiftRole'
delimiter '\t' timeformat 'auto' region 'us-east-2';

SELECT *
FROM pg_table_def
WHERE tablename = 'sales';
```

```
Table sales created

Execution time: 1.14s
Statement 1 of 3 finished

Warnings:
Load into table 'sales' completed, 172456 record(s) loaded successfully.

0 rows affected
COPY executed successfully

Execution time: 1.06s
Statement 2 of 3 finished
```



```
  buyerid integer not null,
  eventid integer not null,
  dateid smallint not null sortkey,
  qtysold smallint not null,
  pricepaid decimal(8,2),
  commission decimal(8,2),
  saletime timestamp);

copy sales from 's3://redshiftbucketsample/sales_tab.txt'
credentials 'aws_iam_role=arn:aws:iam::278492603516:role/myRedshiftRole'
delimiter '\t' timeformat 'auto' region 'us-east-2';

SELECT *
FROM pg_table_def
WHERE tablename = 'sales';
```

schemaname	tablename	column	type	encoding	distkey	sortkey	notnull
public	sales	salesid	integer	az64	false	0	true
public	sales	listid	integer	az64	true	0	true
public	sales	sellerid	integer	az64	false	0	true
public	sales	buyerid	integer	az64	false	0	true
public	sales	eventid	integer	az64	false	0	true
public	sales	dateid	smallint	none	false	1	true
public	sales	qtysold	smallint	az64	false	0	true
public	sales	pricepaid	numeric(8,2)	az64	false	0	false
public	sales	commission	numeric(8,2)	az64	false	0	false
public	sales	saletime	timestamp without time zone	az64	false	0	false

(2) How does AWS charge for this service?

Amazon Redshift costs less to operate than any other data warehouse. It starts small at $0.25 per hour and scale up to petabytes of data and thousands of concurrent users.

On-Demand Pricing:

Example:

In US West (Northern California), for Dense Compute EC2 dc2.large - $0.33 per Hour

Redshift Pricing Spectrum:

$5.00 per terabyte of data scanned

Concurrency Scaling pricing:

A 10 DC2.8XL node Redshift cluster in the US-East costs $48 per hour.

Redshift managed storage pricing:

In US West (Northern California), for Storage/month - $0.0271 per GB

(3) Is it a part of the free tier?

The Amazon Redshift free trial program is not part of the AWS Free Tier

(4) Provide URLs to documentation for this service.

https://aws.amazon.com/redshift/

https://docs.aws.amazon.com/redshift/latest/dg/c_high_level_system_architecture.html

https://aws.amazon.com/redshift/pricing/

https://docs.aws.amazon.com/redshift/latest/dg/welcome.html

(5) URLs to AWS videos about this service.

https://www.youtube.com/watch?v=7bfOIlAyxlg

(6) Watch the videos you located.

https://www.youtube.com/watch?v=7bfOIlAyxlg

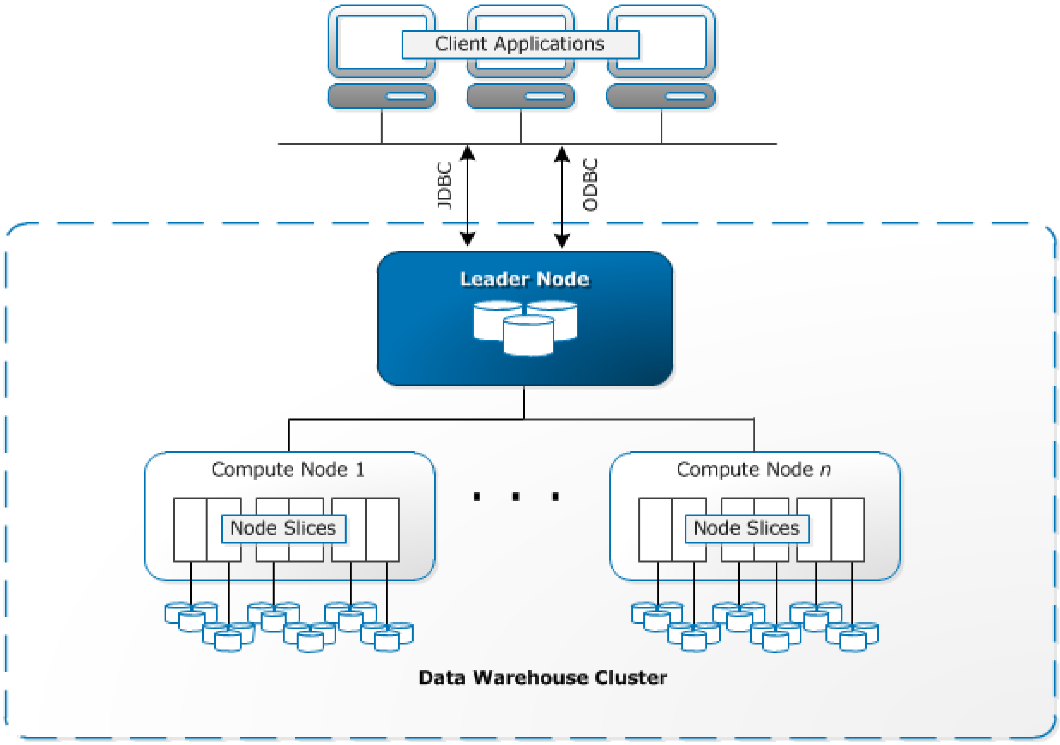
(7) For one of the videos you listed, answer the following questions:

- **Summarize a few key points made in the video.**

Definition of Amazon Redshift: It's a Cloud-based Data warehouse service that is used for collecting and storing data. It's also used to get and analyze the data using BI (Business Intelligence) tools and simplifies the process of handling large sets of data.

Data warehouse Definition: Data warehouse is a repository where the data is stored

1. Before Redshift, people used to fetch the data from the Data warehouse.
2. Fetching data from Data warehouse was a complicated task because the developer and Data warehouse might be located in different geographic locations and there might be network connectivity issues, internet connectivity challenges, security challenges, and a lot of maintenance is required to manage Data warehouse.



3. **Cons of Traditional Data warehouse service:**
    a. It's time consuming process to download or get data from Data warehouse
    b. Maintenance cost was high
    c. There is possibility of losing information in between downloading of data
    d. Data rigidity was an issue
4. All these problems can be solved with Amazon Redshift
5. **Pros:**
    a. **Costs** less when compared to other Cloud data warehouse products in the market. You can have a large Data warehouse or combine databases in a Data warehouse at a very low cost.
    b. It's the fastest Data warehouse (in performance) in market
    c. It has more than 15k customers
    d. Whatever you use, you pay for it ONLY

e. **Scalability:** If you want to increase the nodes in your database, then you can increase it on the fly. You need not switch off to scale.
f. **Availability:** It’s highly available in multiple Availability Zones.
g. You can have multiple clusters in Redshift, you can define own VPC and Security Group for each Cluster and increase the security of Redshift
h. **Flexibility:** You can remove cluster, create new cluster, take a snapshot of the deleting cluster, move the cluster to different region
i. You can have a simple migration from a traditional database to a Cloud Redshift data migration by in-built tools in Cloud being connected to a traditional database.

- **Identify two interesting quotes that were made.**

  Architecture of Amazon Redshift:
  - We have Compute Node which does data processing
  - Leader Node which gives instructions to Compute Node and it also manages Client applications that require data from Redshift.
  - The client applications connect with JDBC (Java Database Connectivity) and ODBC (Open Database Connectivity).
  - Amazon Redshift can monitor connections from Client applications using JDBC.
  - ODBC allows Client applications to have live or direct data interaction with Amazon Redshift. The Leader Node can get information from Compute nodes.
  - Compute nodes (Nodes) processes the data, they are a set of computing resources and when combined together are called Clusters.
  - A Cluster is a set of Computing nodes called Nodes.
  - Nodes which are combined together are called Data warehouse Cluster.
  - We can have 1...100 Compute nodes which is a scalable solution.
  - Each Cluster has a Database in the form of Node.
  - The Leader Node manages the interaction between Client application and Compute Node i.e acts as bridge between both. It also analyzes and develops designs in order to carry out database operations.
  - At high-level, Leader Node sends out instructions to be performed or executed by Compute Node and sends the response to Client applications.
    - The Leander Node runs the programs and assigns the code to individual compute nodes.
    - The Compute Node executes the program and shares the results to Leader Node for final aggregation and then it’s delivered to Client application
  - Each Computer Node is categorized into slices and each slice is allocated with specific memory space, where it processes it’s workload.
  - These Node slices work in parallel and hence the reason why Redshift is the fastest Data warehouse when compared to traditional and currently existing Data warehouses

- **What new facts did you learn from watching this video.**

1. We have additional concepts called Column storage and Compression.
2. **Column storage:** Data is stored in columns which helps in optimizing query performance and quicker output. Below is the example. It makes the data more structured and easy to extract.
3. Compression: To save Column storage we can use Compression as an attribute. It’s a Column-level operation which decreases storage requirements and improves query performance.

- **What was the best part of the video? Why?**

  Use cases: DNA, a Telecommunications company is facing issues with handling Website data and Amazon S3 data which leads to slow process of the application. They overcame this issue by using Amazon Redshift and they noticed a 52% increase in application performance.

- **What questions remain in your mind after watching the video? Why?**

None