

# AWS GLUE FOR HARD ETL (EXTRACT, TRANSFORM, AND LOAD) ACTIVITY

1. FIRST STEP IS TO CREATE THE ROLE FOR OUR AWS GLUE IN ORDER TO GET FULL PERMISSION. IN THE SERVICES, CHOOSE IAM.

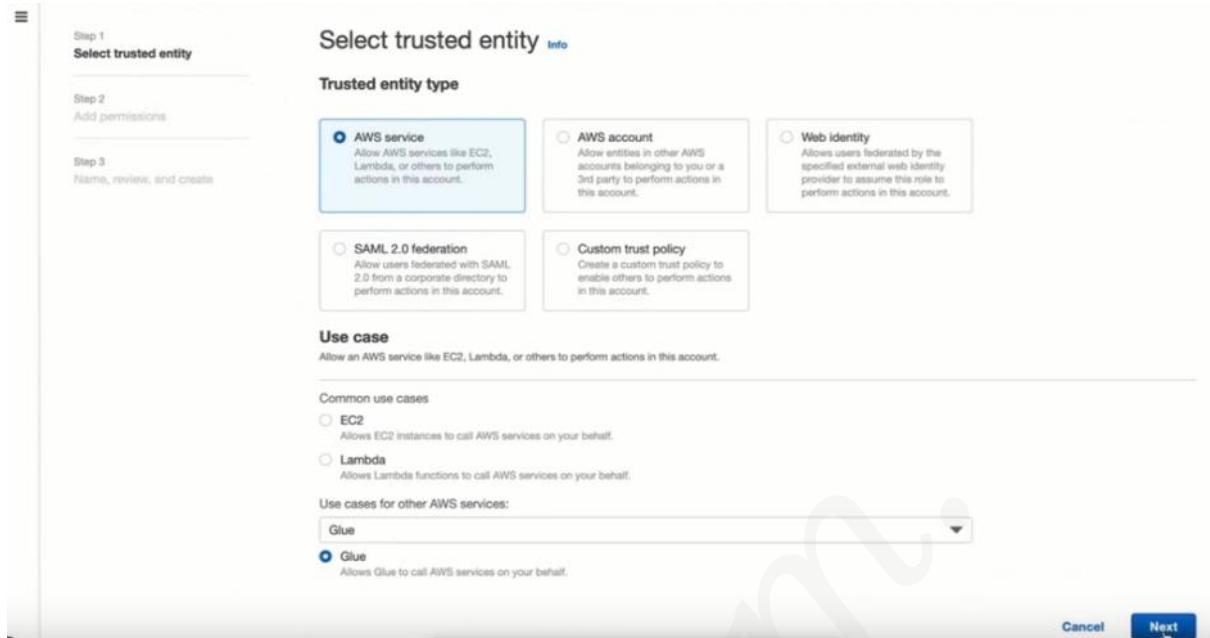
The screenshot shows the AWS search interface with 'IAM' typed into the search bar. The results are filtered under the 'Services' category. The top result is 'IAM' with a brief description: 'Manage access to AWS resources'. Below it are three other services: 'IAM Identity Center (successor to AWS Single Sign-On)', 'Resource Access Manager', and 'Serverless Application Repository'. On the left sidebar, there are links for 'Features (19)', 'Resources New', 'Blogs (1,565)', 'Documentation (46,740)', 'Knowledge Articles (30)', 'Tutorials (2)', 'Events (12)', and 'Marketplace (493)'.

2. CLICK ROLES, THEN, CREATE ROLE.

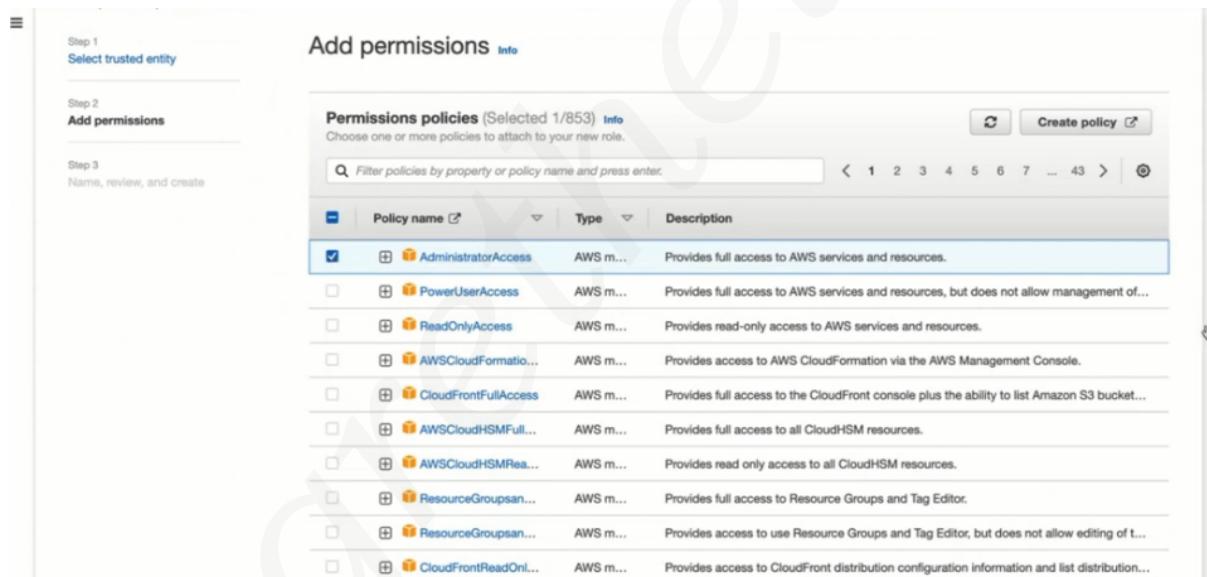
The screenshot shows the 'Roles' page within the IAM service. The left sidebar has navigation links for 'Identity and Access Management (IAM)', 'Dashboard', 'Access management' (with 'User groups', 'Users', 'Roles' selected), 'Policies', 'Identity providers', 'Account settings', 'Access reports' (with 'Access analyzer', 'Archive rules', 'Analyzers', 'Settings'). The main content area shows a table of roles. The first role listed is 'AWSReservedSSO\_AWSAdministratorAccess\_d6359d25c2eb8818'. The table includes columns for 'Role name' and 'Trusted entities'. A 'Create role' button is located at the top right of the table area.

Role name	Trusted entities
AWSReservedSSO_AWSAdministratorAccess_d6359d25c2eb8818	Identity Provider: arn:aws:iam::749601114921:saml-provider/AWSSSC
AWSReservedSSO_AWSPowerUserAccess_72a7e95974f3825f	Identity Provider: arn:aws:iam::749601114921:saml-provider/AWSSSC
AWSReservedSSO_AWSReadOnlyAccess_1ec8cf602ddfb0ec	Identity Provider: arn:aws:iam::749601114921:saml-provider/AWSSSC
AWSReservedSSO_AWSServiceCatalogAdminFullAccess_d1fe8e1088162bc9	Identity Provider: arn:aws:iam::749601114921:saml-provider/AWSSSC
AWSReservedSSO_AWSServiceCatalogEndUserAccess_7c60b01db8043761	Identity Provider: arn:aws:iam::749601114921:saml-provider/AWSSSC
AWSServiceRoleForRedshift	AWS Service: redshift (Service-Linked Role)
AWSServiceRoleForSSO	AWS Service: sso (Service-Linked Role)
AWSServiceRoleForSupport	AWS Service: support (Service-Linked Role)

**3. CHOOSE AWS SERVICE IN THE TRUSTED ENTITY TYPE. AND IN THE USE CASES FOR OTHER AWS SERVICES, CHOOSE AND SELECT GLUE. CLICK, NEXT.**



**4. FOR TRAINING PURPOSES, WE WILL SELECT THE ADMINISTRATORACCESS POLICY WHICH WILL GIVE THE FULL ACCESS. SCROLL DOWN, THEN, CLICK, NEXT.**



5. IN THE ROLE NAME, WE WILL CALL THIS, GLUEFULLACCESSROLE. ON THE DESCRIPTION, IT WILL BE, ALLOWS GLUE TO CALL AWS SERVICESON YOUR BEHALF. SCROLL DOWN, THEN, CLICK CREATE ROLE.

Step 1  
Select trusted entity

Step 2  
Add permissions

Step 3  
Name, review, and create

### Name, review, and create

#### Role details

Role name  
Enter a meaningful name to identify this role.  
**GlueFullAccessRole**

Description  
Add a short explanation for this role.  
**Allows Glue to call AWS services on your behalf.**

Step 1: Select trusted entities Edit

```
1. {
2.     "Version": "2012-10-17",
3.     "Statement": [
4.         {
5.             "Effect": "Allow",
6.             "Principal": {
7.                 "Service": "glue.amazonaws.com"
8.             },
9.             "Action": "sts:AssumeRole"
10.        }
11.    ]
12. }
```

AdministratorAccess      AWS managed - job function      Permissions policy

#### Tags

Add tags - optional Info  
Tags are key-value pairs that you can add to AWS resources to help identify, organize, or search for resources.

No tags associated with the resource.

**Add tag**  
You can add up to 50 more tags.

**Cancel** **Previous** **Create role**

## 6. THIS WILL NOW BE OUR GLUE ACCESS ROLE.

The screenshot shows the AWS Identity and Access Management (IAM) service. On the left, there's a sidebar with navigation links like 'Dashboard', 'User groups', 'Users', 'Roles', 'Policies', 'Identity providers', and 'Account settings'. Under 'Access management', 'Roles' is selected. The main content area is titled 'Roles (10) Info' and contains a table with columns 'Role name' and 'Trusted entities'. A message at the top says 'Role GlueFullAccessRole created.' A 'View role' button is visible in the top right corner.

Role name	Trusted entities
AWSReservedSSO_AWSAdministratorAccess_d6359d25c2eb8818	Identity Provider: arn:aws:iam::749601114921:saml-provider/AWSSC
AWSReservedSSO_AWSPowerUserAccess_72a7e95974f3825f	Identity Provider: arn:aws:iam::749601114921:saml-provider/AWSSC
AWSReservedSSO_AWSReadOnlyAccess_1ec8cf602ddf80ec	Identity Provider: arn:aws:iam::749601114921:saml-provider/AWSSC
AWSReservedSSO_AWSServiceCatalogAdminFullAccess_d1fe8e1088162bc9	Identity Provider: arn:aws:iam::749601114921:saml-provider/AWSSC
AWSReservedSSO_AWSServiceCatalogEndUserAccess_7c60b01db8043761	Identity Provider: arn:aws:iam::749601114921:saml-provider/AWSSC
AWSServiceRoleForRedshift	AWS Service: redshift (Service-Linked Role)
AWSServiceRoleForSSO	AWS Service: sso (Service-Linked Role)
AWSServiceRoleForSupport	AWS Service: support (Service-Linked Role)
AWSServiceRoleForTrustedAdvisor	AWS Service: trustedadvisor (Service-Linked Role)
GlueFullAccessRole	AWS Service: glue

## 7. GO TO SERVICES, CLICK AMAZON S3 TO CRCREATE OUR BUCKET.

The screenshot shows the AWS Services search interface. The search bar at the top has 'S3' typed into it. Below the search bar, there's a sidebar with sections like 'Hard ETL', 'Services (7)', 'Features (19)', 'Resources New', 'Blogs (1,249)', 'Documentation (20,670)', 'Knowledge Articles (30)', 'Tutorials (12)', 'Events (26)', and 'Marketplace (1,165)'. The main content area is titled 'Services' and shows a list of services. A context menu is open over the 'S3' service icon, with options like 'Open Link in New Tab' (which is highlighted), 'Open Link in New Window', 'Open Link in Incognito Window', 'Save Link As...', 'Copy Link Address', 'Copy', 'Copy Link to Highlight', 'Search Google for "S3"', 'Print...', 'Translate Selection to English', 'Inspect', 'Speech', and 'Services'. At the bottom of the screen, there's a footer with links for 'Amazon S3 File Gateway' and 'See all 7 results ▾'.

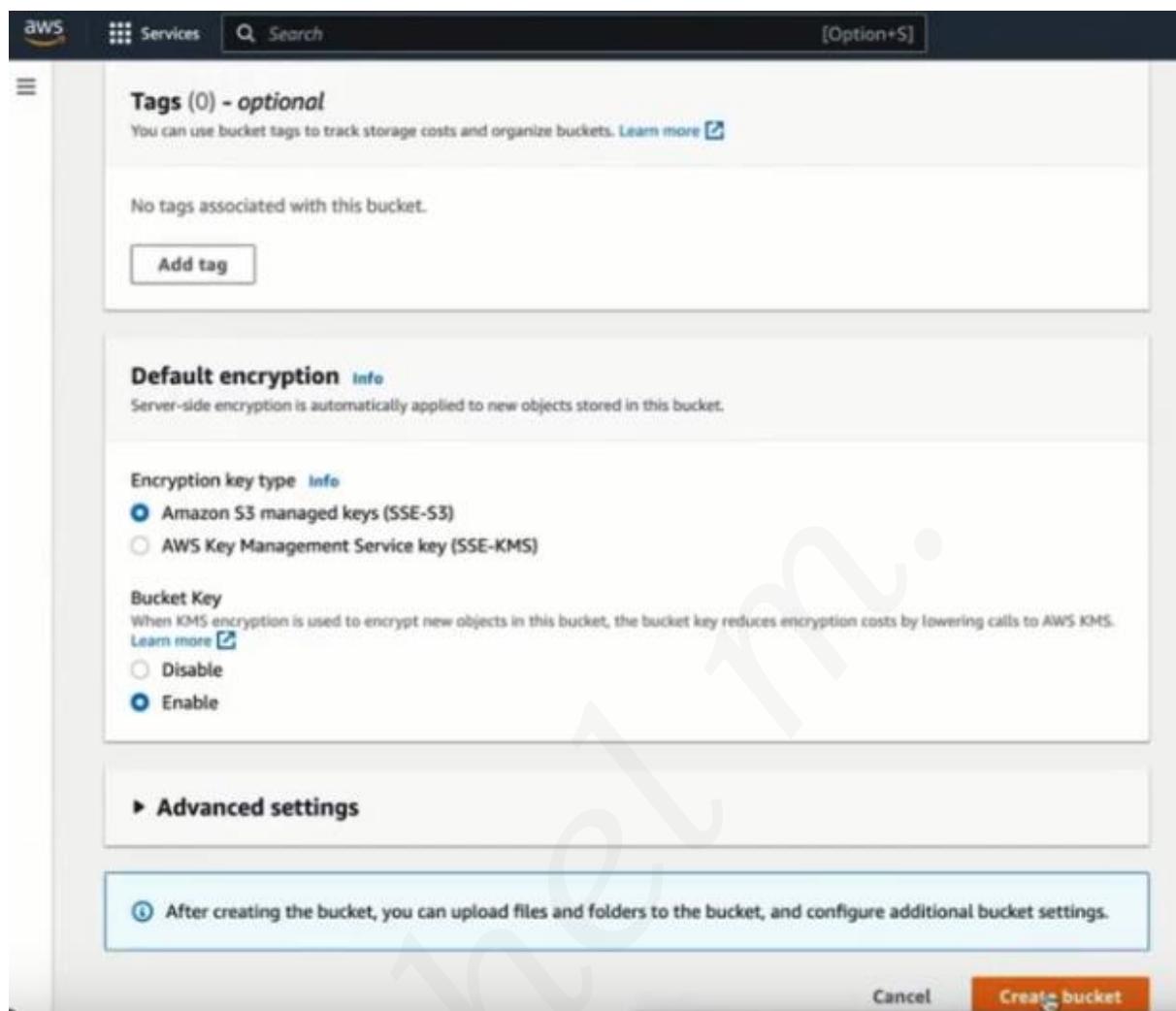
**CLICK CREATE BUCKET.**

The screenshot shows the Amazon S3 landing page. At the top, it says "Storage" and "Amazon S3: Store and retrieve any amount of data from anywhere". Below that, a sub-headline reads "Amazon S3 is an object storage service that offers industry-leading scalability, data availability, security, and performance." To the right, there's a "Create a bucket" button in an orange box. Further down, there's a section titled "How it works" with a video thumbnail labeled "Introduction to Amazon S3" and a "Copy link" button. On the far right, there's a "Pricing" section with a note about no minimum fees and links to monthly calculators and pricing details.

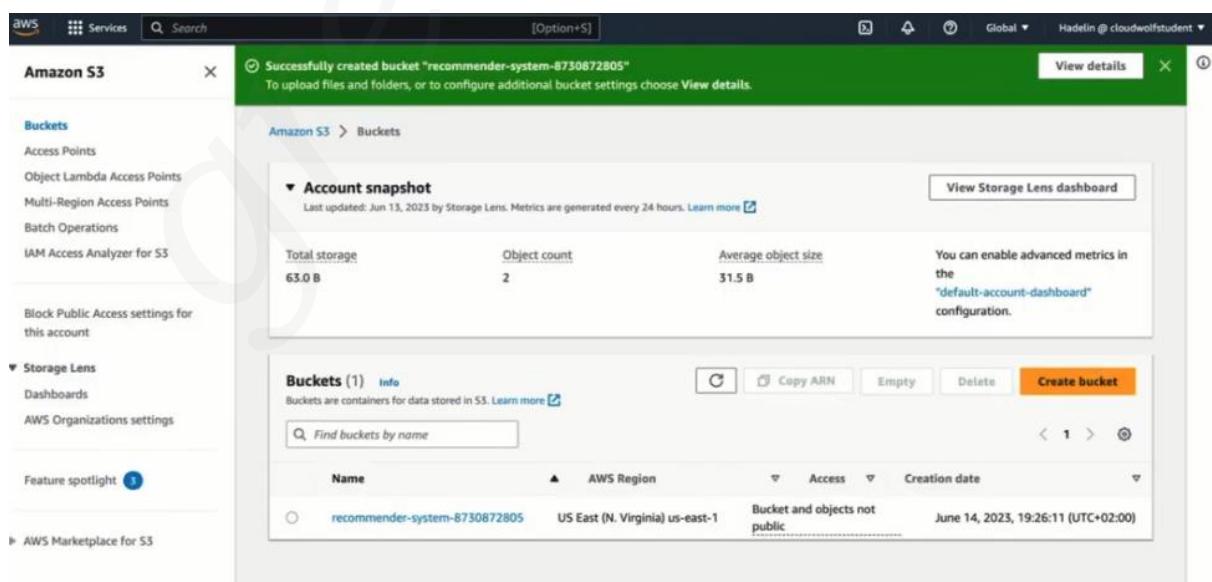
8. SINCE WE ARE BUILDING A RECOMMENDER SYSTEM, WE WILL GIVE A BUCKET NAME `recommender-system-8730872805`. THE NUMBERS GIVEN HERE ARE JUST RANDOM. THIS IS TO MAKE THAT THIS IS GLOBALLY UNIQUE NAME.

The screenshot shows the "Create bucket" wizard. The first step, "General configuration", is active. It asks for a "Bucket name" which is set to "recommender-system-8730872805". It also specifies the "AWS Region" as "US East (N. Virginia) us-east-1". There's an optional "Copy settings from existing bucket" section with a "Choose bucket" button. The second step, "Object Ownership", is shown below. It has two options: "ACLs disabled (recommended)" (selected) and "ACLs enabled". Both options describe how objects are owned by the account or other accounts.

**9. SCROLL DOWN, AND, CLICK CREATE BUCKET.**



**10. THERE YOU GO, WE HAVE THE BUCKET. INSIDE THE BUCKET, WE ARE GOING TO UPLOAD OUR CSV FILE. CLICK THE BUCKET, recommender-system-8730872805.**



**11. INSIDE THE BUCKET, WE ARE GOING TO UPLOAD OUR CSV FILE. CLICK UPLOAD.**

The screenshot shows the Amazon S3 console interface. On the left, there's a sidebar with various options like Buckets, Access Points, Object Lambda Access Points, Multi-Region Access Points, Batch Operations, and IAM Access Analyzer for S3. The main area is titled 'recommender-system-8730872805' and shows the 'Objects' tab selected. At the top of the object list, there are several buttons: Copy S3 URI, Copy URL, Download, Open, Delete, Actions (with a dropdown arrow), and Create folder. Below these buttons is a large red 'Upload' button with a camera icon. Underneath the buttons is a search bar with the placeholder 'Find objects by prefix'. The object list table has columns for Name, Type, Last modified, Size, and Storage class. A message at the bottom states 'No objects' and 'You don't have any objects in this bucket.' There's also a small 'Upload' button at the bottom of the list.

**12. CLICK ADD FILES.**

The screenshot shows the 'Upload' page within the Amazon S3 console. At the top, there's a breadcrumb navigation: 'Amazon S3 > Buckets > recommender-system-8730872805 > Upload'. The main area is titled 'Upload' with a 'Info' link. A large dashed rectangular area is provided for dragging files or choosing them. Below this is a table titled 'Files and folders (0)' with a note: 'All files and folders in this table will be uploaded.' The table has a 'Remove' button, an 'Add files' button (which is highlighted in red), and an 'Add folder' button. A search bar is also present. The table has columns for Name, Folder, Type, and Size. A message at the bottom of the table area says 'No files or folders' and 'You have not chosen any files or folders to upload.' At the bottom of the page is a 'Destination' section with a table showing 'Destination' and the URL 's3://recommender-system-8730872805'.

**13. CHOOSE THE SMALL DATA FILE. CLICK AND UPLOAD THE MOVIE\_RATINGS.CSV FOR OUR S3 SOURCE.**



**14. YOU CAN SEE THAT THE MOVIE\_RATINGS IS NOW UPLOADED.**

Upload succeeded  
View details below.

### Upload: status

The information below will no longer be available after you navigate away from this page.

Summary	Succeeded	Failed
Destination s3://recommender-system-8730872805	1 file, 2.3 MB (100.00%) Succeeded	0 files, 0 B (0%) Failed

**Files and folders** | Configuration

**Files and folders (1 Total, 2.3 MB)**

Name	Folder	Type	Size	Status	Error
movie_ratings.csv	-	text/csv	2.3 MB	Succeeded	-

Amazon S3 > Buckets > recommender-system-8730872805

**recommender-system-8730872805** Info

Objects | Properties | Permissions | Metrics | Management | Access Points

**Objects (1)**

Objects are the fundamental entities stored in Amazon S3. You can use Amazon S3 inventory [Get object](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

Name	Type	Last modified	Size	Storage class
movie_ratings.csv	csv	June 14, 2023, 19:27:06 (UTC+02:00)	2.3 MB	Standard

**15. WE NOW NEED TO CREATE THE DATA SOURCE IN THE DATA CATALOGUE OF GLUE. OPEN THE AWS GLUE.**

The screenshot shows the AWS Glue search results page. The search bar at the top contains the query 'glue'. On the left, there is a sidebar titled 'Hard ETL' with sections for 'Saved jobs' (1), 'Visual' (1), and a '+' button. Below these are links for 'Services (5)', 'Features (10)', 'Resources New', 'Blogs (340)', 'Documentation (8,717)', 'Knowledge Articles (30)', 'Events (7)', and 'Marketplace (258)'. The main content area is titled 'Services' and shows 'See all 5 results'. The first result is 'AWS Glue', which is described as a serverless data integration service. It includes a 'Top features' section with links to 'AWS Glue Studio', 'Data Catalog', 'Crawlers', and 'Workflows'. Other results listed are 'AWS Glue DataBrew', 'AWS Lake Formation', and 'Athena'.

**16. GO TO THE DATA CATALOG, CLICK DATABASES. THEN, CLICK, ADD DATABASE. WE WILL BE CREATING ONE OF GLUE.**

The screenshot shows the AWS Glue Data Catalog Databases page. The left sidebar has sections for 'Getting started', 'ETL jobs', 'Visual ETL', 'Notebooks', 'Job run monitoring', 'Data Catalog tables', 'Data connections', 'Workflows (orchestration)', 'Data Catalog', 'Tables', 'Stream schema registries', 'Schemas', 'Connections', 'Crawlers', 'Classifiers', 'Catalog settings', 'Data Integration and ETL', 'ETL jobs', 'Visual ETL', 'Notebooks', 'Job run monitoring', 'Interactive Sessions', 'Data classification tools', and 'Sensitive data detection'. The main content area is titled 'Databases (0)' and includes a sub-header 'A database is a set of associated table definitions, organized into a logical group.' A search bar 'Filter databases' is present. Below is a table with columns 'Name', 'Description', 'Location URI', and 'Created on (UTC)'. The table displays the message 'No resources' and 'No resources to display.'

**17. SINCE WE WILL BE CREATING A DATABASE FOR THE MOVIE RATINGS, WE WILL NAME IT MOVIE-RATINGS-GLEUE-DATABASE. THEN, CLICK CREATE DATABASE.**

AWS Glue > Databases > Add database

Create a database

Create a database in the AWS Glue Data Catalog.

**Database details**

Name: movie-ratings-glue-database

Database name is required, in lowercase characters, and no longer than 255 characters.

Location - optional

Set the URI location for use by clients of the Data Catalog.

Description - optional

Enter text

Descriptions can be up to 2048 characters long.

Cancel Create database

**18. THIS WILL NOW BE THE FIRST POINT IN THE DATA CATALOG. WE WILL THEN CREATE A TABLE WTIHIN THE MOVIE-RATINGS-GLEUE-DATABASE DATABASE. THE TABLE IS GOING TO BE EXACTLY THE TABLE CONTAINING THE MOVIE RATINGS IN THE CSV FILE. TO DO SO, WE ARE GOING TO ENTER OUR DATABASE.**

**CLICK THE DATABASE movie-ratings-glue-database.**

AWS Glue > Databases

Databases (1)

A database is a set of associated table definitions, organized into a logical group.

Last updated (UTC) June 14, 2023 at 17:28:45

Add database

Name	Description	Location URI	Created on (UTC)
movie-ratings-glue-database	-	-	June 14, 2023 at 17:28:43

**19. CLICK ADD TABLES USING CRAWLER. BASICALLY, THE CRAWLER IS THE TOOL THAT CAN EXTRACT THE DATA IN A CSV FILE. IT EXTRACTS THE METADATA AND CAN ALSO POPULATE THE DATA IN THE CSV FILE.**

AWS Glue > Databases > movie-ratings-glue-database

movie-ratings-glue-database

Database properties

Name	Description	Location	Created on (UTC)
movie-ratings-glue-database	-	-	June 14, 2023 at 17:28:43

Tables (0)

Last updated (UTC) June 14, 2023 at 17:29:02

Add tables using crawler

Name	Database	Location	Classification	Deprecated	View data
No available tables					

**20. ENTER UNIQUE CRAWLER NAME. HERE, WE WILL CALL IT MOVE-RATINGS-CRAWLER. CLICK NEXT.**

AWS Glue > Crawlers > Add crawler

Step 1 Set crawler properties

Step 2 Choose data sources and classifiers

Step 3 Configure security settings

Step 4 Set output and scheduling

Step 5 Review and create

**Set crawler properties**

**Crawler details** Info

Name

Description can be up to 255 characters long. Some character set including control characters are prohibited.

Description - optional

Descriptions can be up to 2048 characters long.

Tags - optional Use tags to organize and identify your resources.

Cancel Next

**21. CHOOSE NOT YET IN THE DATA SOURCE CONFIGURATION. THEN CLICK ADD A DATA SOURCE.**

AWS Glue > Crawlers > Add crawler

Step 1 Set crawler properties

Step 2 Choose data sources and classifiers

Step 3 Configure security settings

Step 4 Set output and scheduling

Step 5 Review and create

**Choose data sources and classifiers**

**Data source configuration**

Is your data already mapped to Glue tables?

Not yet Select one or more data sources to be crawled.

Yes Select existing tables from your Glue Data Catalog.

**Data sources (0) Info**

The list of data sources to be scanned by the crawler.

Type	Data source	Parameters
You don't have any data sources.		

Add a data source

Custom classifiers - optional A classifier checks whether a given file is in a format the crawler can handle. If it is, the classifier creates a schema in the form of a StructType object that matches that data format.

Cancel Previous Next

**22. DATA SOURCE IS S3, LOCATION OF S3 DATA IS IN THIS ACCOUNT. THEN, BROWSE S3.**

**Add data source**

**Data source**  
Choose the source of data to be crawled.

S3

**Network connection - optional**  
Optionally include a Network connection to use with this S3 target. Note that each crawler is limited to one Network connection so any other S3 targets will also use the same connection (or none, if left blank).

**Location of S3 data**

In this account  
 In a different account

**S3 path**  
Browse for or enter an existing S3 path.

s3://bucket/prefix/object

All folders and files contained in the S3 path are crawled. For example, type s3://MyBucket/MyFolder/ to crawl all objects in MyBucket.

**Subsequent crawler runs**  
This field is a global field that affects all S3 data sources.

Crawl all sub-folders  
Crawl all folders again with every subsequent crawl.  
 Crawl new sub-folders only  
Only Amazon S3 folders that were added since the last crawl will be crawled. If the schemas are compatible, new partitions will be added to existing tables.  
 Crawl based on events  
Rely on Amazon S3 events to control what folders to crawl.

Crawl only a subset of files

**23. IN THE BROWSE S3, CHOOSE THE BUCKET LISTED BELOW, THE BUCKET WE CREATED AWHILE AGO.**

**Choose S3 path**

S3 buckets

**Buckets (1/1)**

Name	Creation date
recommender-system-8730872805	2023-06-14T17:26:11.000Z

**24. SCROLL DOWN, IN THE SUBSEQUENT CRAWLER RUNS, CHOOSE CRAWL ALL SUB-FOLDERS. THEN CLICK, ADD DATA SOURCE.**

**Data source**  
Choose the source of data to be crawled.  
S3

**Network connection - optional**  
Optional: include a Network connection to use with this S3 target. Note that each crawler is limited to one Network connection so any other S3 targets will also use the same connection (or none, if left blank).  
Clear selection Add new connection

**Location of S3 data**  
 In this account  
 In a different account

**S3 path**  
Browse for or enter an existing S3 path.  
s3://recommender-system-87308728 View Browse S3  
All folders and files contained in the S3 path are crawled. For example, type s3://MyBucket/MyFolder/ to crawl all objects in MyFolder within MyBucket.

**Subsequent crawler runs**  
This field is a global field that affects all S3 data sources.  
 Crawl all sub-folders  
Crawl all folders again with every subsequent crawl.  
 Crawl new sub-folders only  
Only Amazon S3 folders that were added since the last crawl will be crawled. If the schemas are compatible, new partitions will be added to existing tables.  
 Crawl based on events  
Rely on Amazon S3 events to control what folders to crawl.  
 Sample only a subset of files  
 Exclude files matching pattern

**25. SELECT S3 IN THE DATA SOURCE. THEN, CLICK, NEXT.**

**AWS Glue** AWS Glue > Crawlers > Add crawler

**Step 1** Set crawler properties  
**Step 2** Choose data sources and classifiers  
**Step 3** Configure security settings  
**Step 4** Set output and scheduling  
**Step 5** Review and create

**Choose data sources and classifiers**

**Data source configuration**

Is your data already mapped to Glue tables?  
 Not yet  
Select one or more data sources to be crawled.  
 Yes  
Select existing tables from your Glue Data Catalog.

**Data sources (1) Info**  
The list of data sources to be scanned by the crawler.

Type	Data source	Parameters
S3	s3://recommender-system-8730...	Recrawl all

**Custom classifiers - optional**  
A classifier checks whether a given file is in a format the crawler can handle. If it is, the classifier creates a schema in the form of a StructType object that matches that data format.

Cancel Previous Next

**26. ENTER THE IAM ROLE WE JUST CREATED. THIS IS TO GIVE US FULL PERMISSION FROM GLUE. SELECT THE ROLE THAT WE`VE CREATED, GLUEFULLACCESSROLE. THEN, CLICK NEXT.**

The screenshot shows the AWS Glue 'Add crawler' wizard at Step 3: Configure security settings. On the left sidebar, under 'Step 3' 'Configure security settings', there is a list of steps: Step 1 Set crawler properties, Step 2 Choose data sources and classifiers, Step 4 Set output and scheduling, and Step 5 Review and create. The main panel displays the 'Configure security settings' step. It includes an 'IAM role' section with a dropdown menu titled 'Choose an IAM role'. The dropdown lists several IAM roles, with 'GlueFullAccessRole' highlighted. Other listed roles include 'AWSReservedSSO\_AWSAdministratorAccess\_d6359d25c2eb881', 'AWSReservedSSO\_AWSPowerUserAccess\_72a7e95974f3825f', 'AWSReservedSSO\_AWSReadOnlyAccess\_1ecBcf602ddff80ec', 'AWSReservedSSO\_AWSServiceCatalogAdminFullAccess\_d1fe8e1088162bc9', and 'AWSReservedSSO\_AWSServiceCatalogEndUserAccess\_7c60b01db8043761'. Below the dropdown is a note: 'Allows Glue to call all AWS services on your behalf.' At the bottom right of the panel are 'Cancel', 'Previous', and 'Next' buttons.

The screenshot shows the AWS Glue 'Add crawler' wizard at Step 3: Configure security settings. The left sidebar shows steps 1 through 5. The main panel has an 'IAM role' section with 'GlueFullAccessRole' selected. Below this, there is a 'Lake Formation configuration - optional' section with a checkbox for 'Use Lake Formation credentials for crawling S3 data source'. A note below the checkbox states: 'Checking this box will allow the crawler to use Lake Formation credentials for crawling the data source. If the data source is registered in another account, you must provide the registered account ID. Otherwise, the crawler will crawl only those data sources associated to the account. Only applicable to S3, Glue Catalog and Iceberg data sources.' There is also a 'Security configuration - optional' section with a note: 'Enable at-rest encryption with a security configuration.' At the bottom right are 'Cancel', 'Previous', and 'Next' buttons.

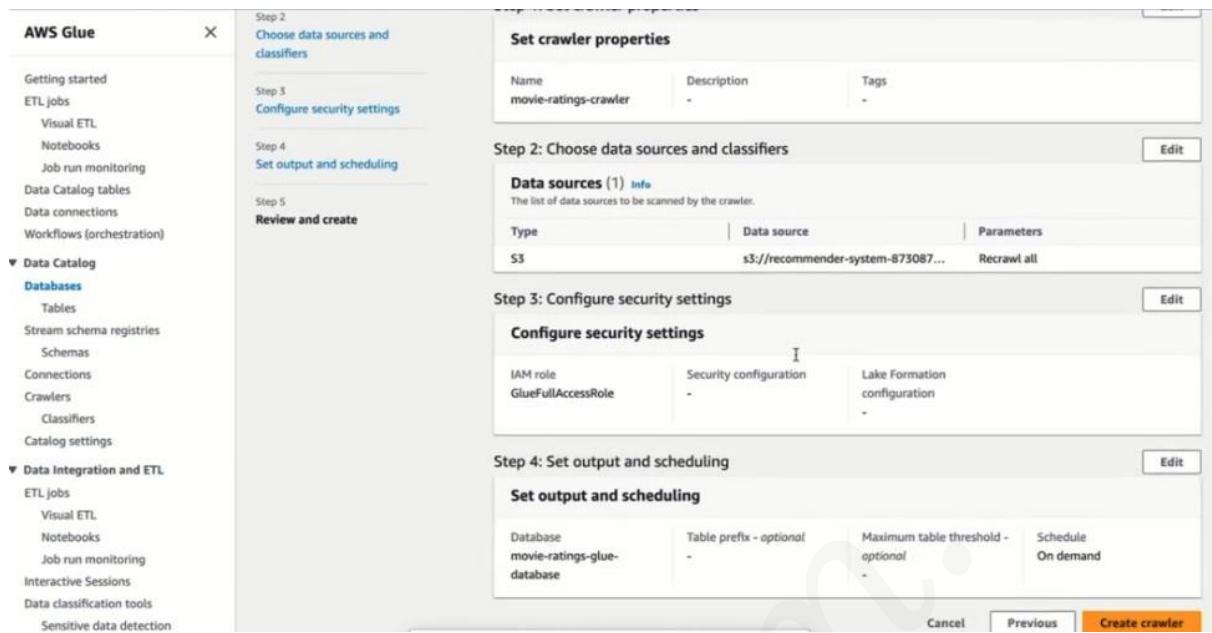
## 27. IN THE TARGET DATABASE, SELECT THE ONE THAT WE CREATED.

The screenshot shows the AWS Glue interface for creating a new crawler. On the left, a sidebar lists various services like AWS Glue, ETL jobs, Data Catalog tables, and Data Integration and ETL. The main panel is titled 'Set output and scheduling'. It includes sections for 'Output configuration' (with a 'Target database' dropdown set to 'movie-ratings-glue-database'), 'Maximum table threshold - optional' (set to 0), and 'Crawler schedule' (set to 'On demand'). A note at the bottom says 'ON THE FREQUENCY, WE KEEP IT ON DEMAND. THEN, CLICK NEXT.'

ON THE FREQUENCY, WE KEEP IT ON DEMAND. THEN, CLICK NEXT.

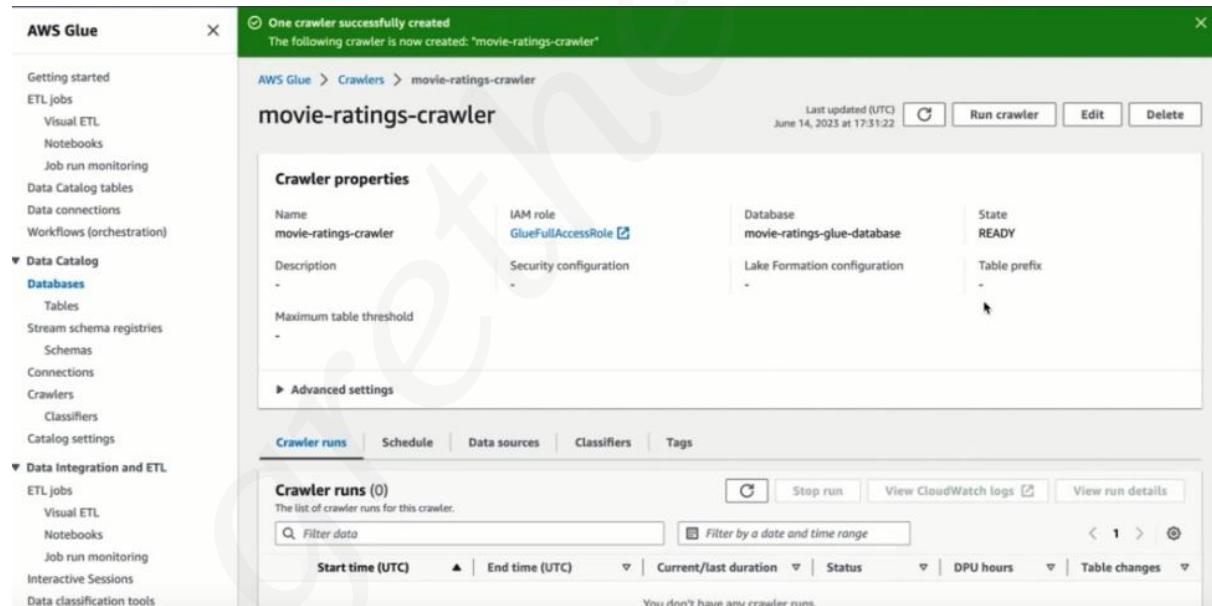
This screenshot shows the same 'Set output and scheduling' step as the previous one, but with a key difference: the 'Frequency' dropdown is now set to 'On demand'. The rest of the configuration remains the same, including the target database and maximum table threshold.

## 28. REVIEW THE CONFIGURATION, THEN, CLICK CREATE CRAWLER.



## 29. WE NOW HAVE JUST CREATED OUR CRAWLER. WE WILL NOW RUN IT IN ORDER TO DO THIS EXTRACTION OF THE DATA AND PUT IT IN THE TABLE OF THIS DATABASE WE JUST CREATED WITHIN GLUE.

**CLICK RUN CRAWLER.**



### 30. SELECT THE CRAWLER, THEN, CLICK, RUN.

AWS Glue

Getting started  
ETL jobs  
Visual ETL  
Notebooks  
Job run monitoring  
Data Catalog tables  
Data connections  
Workflows (orchestration)

Crawlers

Crawlers (1/1) Info

Name	State	Last run	Last run tim...	Log	Table changes ...
movie-ratings-cr...	Ready	-	-	-	-

31. IT IS NOW SUCCESSFULLY RUNNING. IT WILL EXTRACT ALL THE DATA FROM THE CSV FILE TO PUT IT IN THE TABLE OF THE DATABASE WE CREATED. WE CREATED TWO THINGS SO FAR, THE DATABASE FIRST, THEN, WE ARE CREATING THE TABLE OF THIS DATABASE.

Starting crawler

Attempting to start run crawler "movie-ratings-crawler"

AWS Glue > Crawlers

Crawlers

Crawlers (1/1) Info

Name	State	Last run	Last run tim...	Log	Table changes ...
movie-ratings-cr...	Ready	-	-	-	-

IT IS NOW RUNNING AND IT WILL TAKE APPROX. 2 MINUTES. AT THE END, WE WILL SEE THAT WE HAVE THE MOVE RATINGS TABLE CONTAINING THE EXACT SAME DATA AS IN THE CSV FILE. AND THAT IS THE DATA CATALOG ELEMENT AND THE ELEMENT IS A TABLE WHICH WE WILL THEN CONNECT IN OUR ETL PROCESS.

Crawler successfully starting

The following crawler is now starting: "movie-ratings-crawler"

AWS Glue > Crawlers

Crawlers

Crawlers (1/1) Info

Name	State	Last run	Last run tim...	Log	Table changes ...
movie-ratings-cr...	Running	-	-	-	-

ONCE IT IS STOPPING, IT MEANS THAT THE CRAWLING IS DONE AND SOON IT SHOULD SAY SUCCESSFUL.

The screenshot shows the AWS Glue Crawlers page. At the top, a green banner displays a success message: "Crawler successfully starting" and "The following crawler is now starting: 'movie-ratings-crawler'". Below the banner, the page title is "AWS Glue > Crawlers" and the section title is "Crawlers". A brief description states: "A crawler connects to a data store, progresses through a prioritized list of classifiers to determine the schema for your data, and then creates metadata tables in your data catalog." The main content area is titled "Crawlers (1/1) Info" and contains a table with one row. The table columns are: Name, State, Schedule, Last run, Last run time, Log, and Table changes. The single row shows "movie-ratings-cr..." with a state of "Stopping".

The screenshot shows the AWS Glue Crawlers page again. The green banner now displays a success message: "Crawler successfully starting" and "The following crawler is now starting: 'movie-ratings-crawler'". The page title and section title remain the same. The main content area shows the crawler has moved from "Stopping" to "Ready" status, indicated by a green checkmark icon. The last run status is "Succeeded" and the timestamp is "June 14, 2023 a...".

## 32. ONCE READY OR SUCCESSFUL, GO TO THE DATA CATALOG, DATABASE, THEN, TABLES. REFRESH IT TO SEE THE TABLES. CLICK THE TABLE AVAILABLE IN THE LIST.

The screenshot shows the AWS Glue Data Catalog Tables page. The left sidebar includes sections like Getting started, ETL jobs, Visual ETL, Notebooks, Job run monitoring, Data Catalog tables (with sub-options Databases, Tables, Stream schema registries, Schemas, Connections, Crawlers, Classifiers), and Data Flow. The main content area shows a green banner: "Crawler successfully starting" and "The following crawler is now starting: 'movie-ratings-crawler'". The title is "Tables" and a description states: "A table is the metadata definition that represents your data, including its schema. A table can be used as a source or target in a job definition." The main table is titled "Tables (1)" and lists one table: "recommendation\_system\_8730872805". The table details are: Name (recommendation\_system\_8730872805), Database (movie-ratings-glue-data), Location (s3://recommendation-system), Classification (csv), and Deprecated (None). There are buttons for Delete, Data quality, Add tables using crawler, and Add table.

**33. THIS IS NOW OUR TABLE OVERVIEW. THERE, WE CAN`T SEE THE VALUES WHICH GIVES OUR TABLE NOT VERY INTUITIVE. TO HAVE A GREAT LOOK OF OUR TABLE, WE WILL NEED TO USE ANOTHER SERVICE.**

The screenshot shows the AWS Glue Table Overview interface. On the left, there's a sidebar with navigation links for AWS Glue services like Getting started, ETL jobs, Data Catalog tables, and Data Integration and ETL. The main area has tabs for 'Table overview' and 'Data quality New'. Under 'Table overview', the 'Table details' tab is selected, showing information such as Name (recommender\_system\_8730872805), Description (-), Database (movie-ratings-glue-database), Classification (CSV), Location (s3://recommender-system-8730872805/), Connection (-), Deprecated (-), Last updated (June 14, 2023 at 17:33:30), Input format (org.apache.hadoop.mapred.TextInputFormat), Output format (org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat), and Serde serialization lib (org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe). Below this, the 'Schema' tab is selected, displaying a table schema with four columns: #, Column name, Data type, and Partition key. The schema entries are:

#	Column name	Data type	Partition key
1	userid	bigint	-
2	movieid	bigint	-
3	rating	double	-
4	timestamp	bigint	-

**34. IN THE SEARCH BAR, LOOK FOR ATHENA. THIS IS A QUERY SERVICE WHERE WE CAN MANIPULATE THE DATA SO WE CAN HAVE A GOOD LOOK OF OUR DATA.**

The screenshot shows the AWS search interface with a search bar containing 'Athena'. The results are displayed under the 'Services' category. The top result is 'Athena' (Serverless interactive analytics service). Below it, under 'Features', are listed 'Data sources' (Athena feature), 'Workgroups' (Athena feature), 'Notebooks' (Athena feature), and 'SQL queries' (Athena feature). The sidebar on the left is identical to the one in the previous screenshot, showing the AWS Glue navigation menu.

HERE, YOU WILL SEE THE DATA SOURCE, DATABASE, AND THE TABLE THAT WE JUST CREATED.

The screenshot shows the Amazon Athena Query editor interface. On the left, there's a sidebar labeled 'Data' with dropdown menus for 'Data source' (set to 'AwsDataCatalog') and 'Database' (set to 'movie-ratings-glue-database'). Below these are sections for 'Tables and views' (with a 'Create' button) and a search bar ('Filter tables and views'). Under 'Tables and views', there's a 'Tables (1)' section showing a single entry: 'recommender\_system\_8730872805'. To the right of the sidebar is the main area titled 'Query 1' containing a table with one row. At the bottom of the main area are buttons for 'Run', 'Explain', 'Cancel', 'Clear', and 'Create'. Below the table is a 'Results' section with 'Copy' and 'Download results' buttons.

35. TO LOOK FOR THE TABLE, CLICK THE THREE DOTS ON THE RIGHT SIDE OF THE TABLE, THEN, CLICK PREVIEW TABLE.

This screenshot is similar to the previous one, showing the Amazon Athena Query editor. The 'Tables and views' section now has a context menu open over the 'recommender\_system\_8730872805' table. The menu options include 'Run Query', 'Preview Table' (which is highlighted with a blue border), 'Generate table DDL', 'Insert', 'Insert into editor', 'Manage', 'Delete table', 'View properties', and 'View in Glue'. The rest of the interface remains the same, with the main query area and results section visible.

**36. THIS WILL BE THE RESULT OR THE PREVIEW OF THE TABLE. IN THE QUERY SECTION, WE LIMIT THE TABLE TO 10. YOU CAN SET THE LIMIT BASED ON YOUR PREFERENCE.**

The screenshot shows the Amazon Athena Query editor interface. The top navigation bar includes 'Amazon Athena > Query editor' and tabs for 'Editor', 'Recent queries', 'Saved queries', and 'Settings'. A 'Workgroup' dropdown is set to 'primary'. The main area has two tabs: 'Query 1' and 'Query 2'. 'Query 1' contains the SQL command: 'SELECT \* FROM "movie-ratings-glue-database"."recommender\_system\_8730872805" limit 10'. Below the SQL editor is a table browser with sections for 'Tables (1)' and 'Views (0)'. The 'Tables' section shows 'recommender\_system\_8730872805'. The bottom panel displays the 'Query results' tab, which is 'Completed'. It shows the execution time: 'Time in queue: 113 ms', 'Run time: 623 ms', and 'Data scanned: 690.49 KB'. The 'Results (10)' section lists ten rows of data from the 'recommender\_system' table, including columns: #, userid, movieid, rating, and timestamp.

This screenshot shows the detailed results of the query from the previous image. The table has columns: #, userid, movieid, rating, and timestamp. The data consists of 10 rows, each representing a user rating for a specific movie. The rows are numbered 1 through 10. The 'userid' column is consistently '1'. The 'movieid' column contains values such as 110, 147, 858, 1221, 1246, 1968, 2762, 2918, 2959, and 4226. The 'rating' column shows ratings like 1.0, 4.5, 5.0, 5.0, 5.0, 4.0, 4.5, 5.0, 4.0, and 4.0. The 'timestamp' column shows timestamps such as 1425941529, 1425942435, 1425941523, 1425941546, 1425941556, 1425942148, 1425941300, 1425941593, 1425941601, and 1425942228.

#	userid	movieid	rating	timestamp
1	1	110	1.0	1425941529
2	1	147	4.5	1425942435
3	1	858	5.0	1425941523
4	1	1221	5.0	1425941546
5	1	1246	5.0	1425941556
6	1	1968	4.0	1425942148
7	1	2762	4.5	1425941300
8	1	2918	5.0	1425941593
9	1	2959	4.0	1425941601
10	1	4226	4.0	1425942228

**37. BACK INTO THE AWS GLUE, WE ARE DONE WITH THE FIRST DATA SOURCE, WE HAVE CREATED THE TABLE, AND THE TABLE IS POPULATED WITH THE DATA.**

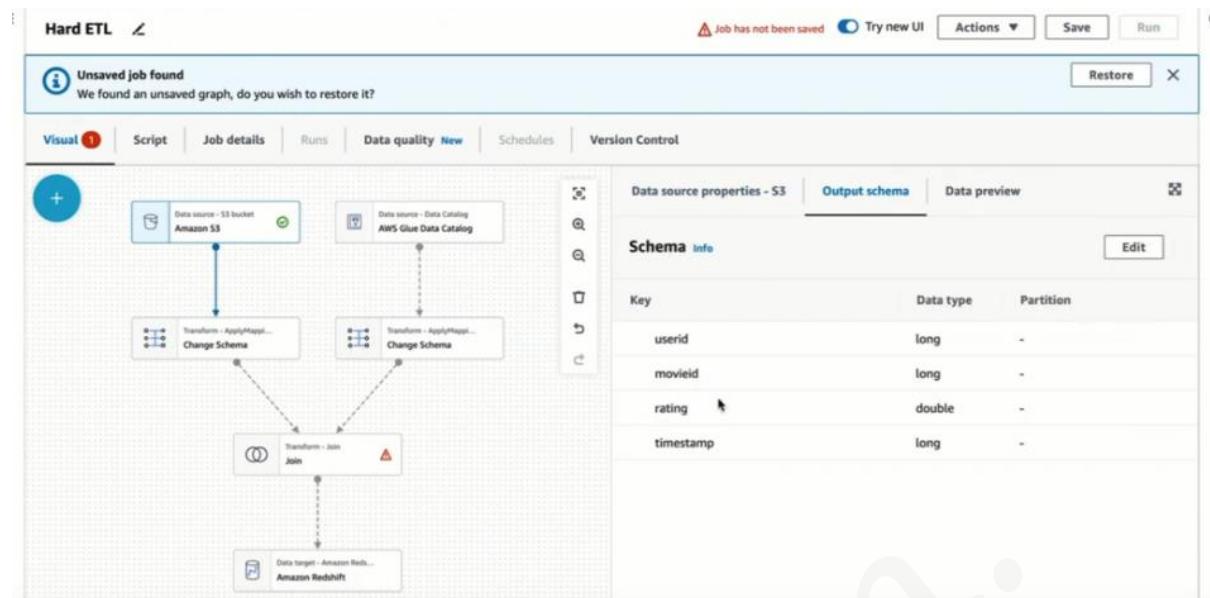
**38. WE NOW GO TO THE VISUAL EDITOR AND WE ARE GOING TO CONNECT THE DATA. WE JUST CREATED THE DATA SOURCE S3, BUT, NOW WE ARE GOING TO CONNECT IT.**

**IN THE S3 SOURCE TYPE, CLICK DATA CATALOG TABLE.**

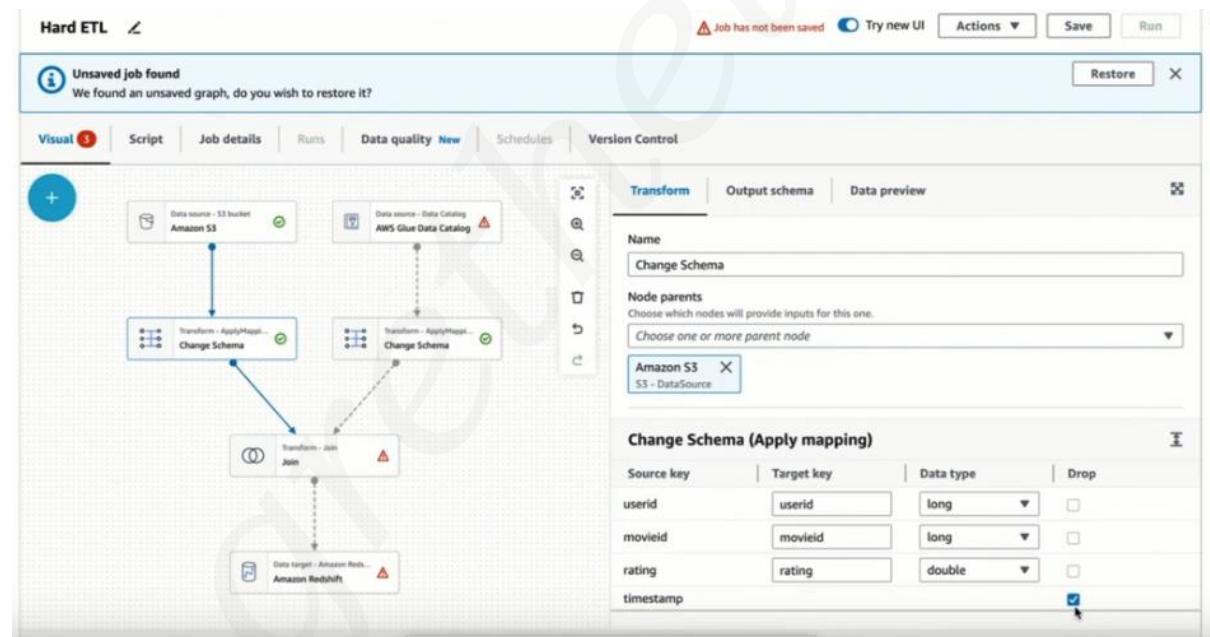
**IN THE DATABASE, CLICK THE DATABASE WE CREATED, MOVIE-RATINGS-GLUE-DATABASE.**

**IN THE TABLE, CLICK THE TABLE THAT WAS CREATED.**

### 39. IN THE OUTPUT SCHEMA, KEEP IT DEFAULT.



### 40. PROCEED TO THE TRANSFORM-APPLYMAPPING PROCEDURE. SINCE WE ARE BUILDING A RECOMMENDER SYSTEM, WE CAN CHECK OR CHANGE THE SCHEMA BASED ON THEIR RELEVANCE. HERE, WE CHOOSE TO DROP THE TIMESTAMP WHICH WE FOUND IRRELEVANT IN BUILDING THE RECOMMENDER SYSTEM.



THIS IS THE KIND OF TRANSFORMATION WE CAN DO HERE WITH THIS APPLY MAPPING TRANSFORM. THE FIRST DATA SOURCE IS DONE. WE HAVE EXTRACTED IT PROPERLY WITH THE DATA CATALOG AND THEN WE HAVE TRANSFORMED IT PROPERLY THROUGH THE APPLY MAPPING TRANSFORM PROCEDURE.

## 41. PROCEED TO THE SECOND DATA SOURCE, THE RDS. IN THE SERVICES, LOOK FOR RDS.

The screenshot shows the AWS Services search interface. The search bar at the top contains the text 'RDS'. On the left, there is a sidebar with several service categories: AWS Glue, Data Catalog tables, Data Catalog, Data Integration and, and ETL jobs. The main content area displays search results for 'Services (14)'. The first result is 'RDS' (Managed Relational Database Service), which is highlighted with a blue border. Below it are other services: 'AWS FIS' (improve resiliency and performance with controlled experiments), 'Database Migration Service' (Managed Database Migration Service), and 'Amazon OpenSearch Service' (Run open-source OpenSearch or Elasticsearch using Managed Clusters or Serverless de...). At the bottom of the results list, there is a link to 'See all 14 results ▶'.

## 42. CLICK CREATE DATABASE.

The screenshot shows the Amazon RDS dashboard. The left sidebar includes links for Dashboard, Databases, Query Editor, Performance insights, Snapshots, Exports in Amazon S3, Automated backups, and Reserved instances. The main area features a callout box with the text: 'Try the new Amazon RDS Multi-AZ deployment option for MySQL and PostgreSQL. For your Amazon RDS for MySQL and PostgreSQL workloads, improve transactional commit latencies by 2x, experience faster failover typically less than 35 seconds and, get read scalability with two readable standby DB instances by deploying the Multi-AZ DB cluster.' It includes a 'Create database' button and a link to 'Learn more'. Below the callout, there are sections for 'Resources' and 'Recommended for you'.

**43. WE WILL CREATE A RDS DATABASE POWERED BY MYSQL. SO, CHOOSE EASY CREATE. THEN, CLICK MYSQL IN THE ENGINE TYPE OF CONFIGURATION.**

RDS > Create database

## Create database

**Choose a database creation method** [Info](#)

Standard create  
You set all of the configuration options, including ones for availability, security, backups, and maintenance.

Easy create  
Use recommended best-practice configurations. Some configuration options can be changed after the database is created.

**Configuration**

Engine type [Info](#)

Aurora (MySQL Compatible)

Aurora (PostgreSQL Compatible)

MySQL

MariaDB

PostgreSQL

Oracle

**44. CHOOSE FREE TIER.**

Edition

MySQL Community

DB instance size

Production  
db.r6g.xlarge  
4 vCPUs  
32 GiB RAM  
500 GiB  
1.017 USD/hour

Dev/Test  
db.r6g.large  
2 vCPUs  
16 GiB RAM  
100 GiB  
0.231 USD/hour

Free tier  
db.t3.micro  
2 vCPUs  
1 GiB RAM  
20 GiB  
0.020 USD/hour

DB instance identifier

Type a name for your DB instance. The name must be unique across all DB instances owned by your AWS account in the current AWS Region.

**45. GIVE DB INSTANCE IDENTIFIER A NAME. WE WILL CALL IT, customer-features-rds-db-instance. KEEP THE MASTER USERNAME AS ADMIN. CHOOSE THE PASSWORD, THEN, CONFIRM, IT. KEEP THE OTHERS AS DEFAULT. THEN CLICK CREATE.**

**DB instance identifier**  
Type a name for your DB instance. The name must be unique across all DB instances owned by your AWS account in the current AWS Region.

The DB instance identifier is case-insensitive, but is stored as all lowercase (as in "mydbinstance"). Constraints: 1 to 60 alphanumeric characters or hyphens. First character must be a letter. Can't contain two consecutive hyphens. Can't end with a hyphen.

**Master username** [Info](#)  
Type a login ID for the master user of your DB instance.  
  
1 to 16 alphanumeric characters. First character must be a letter.

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

**Master password** [Info](#)

Constraints: At least 8 printable ASCII characters. Can't contain any of the following: / (slash), '(single quote)', "(double quote) and @ (at sign).

**Confirm master password** [Info](#)

**► Set up EC2 connection - optional**  
You can also set up a connection to an EC2 instance after creating the database. Go to the database list page or the database details page, choose Actions, and then choose Set up to EC2 connection.

**► View default settings for Easy create**  
Easy create sets the following configurations to their default values, some of which can be changed later. If you want to change

**46. THE CREATING STATUS WILL TAKE UP TO FEW MINUTES.**

**Amazon RDS**

Dashboard  
Databases  
Query Editor  
Performance insights  
Snapshots  
Exports in Amazon S3  
Automated backups  
Reserved instances  
Proxies

Subnet groups  
Parameter groups  
Option groups  
Custom engine versions

**RDS > Databases**

**> Creating database customer-features-rds-db-instance**  
Your database might take a few minutes to launch.  
You can use settings from customer-features-rds-db-instance to simplify configuration of suggested database add-ons while we finish creating your DB for you.  
How was your experience creating an Amazon RDS database? [Provide feedback](#)

**Databases (1)**

Group resources  Modify Actions

Filter by databases

DB identifier	Status	Role	Engine	Region & AZ	Size	Actions
customer-features-rds-db-instance	Creating	Instance	MySQL Community	-	db.t3.micro	-

The screenshot shows the Amazon RDS 'Databases' page. A success message at the top states: 'Successfully created database customer-features-rds-db-instance'. Below it, a note says: 'You can use settings from customer-features-rds-db-instance to simplify configuration of suggested database add-ons while we finish creating your DB for you.' A link 'How was your experience creating an Amazon RDS database? Provide feedback' is present. A modal window titled 'Consider creating a Blue/Green Deployment to minimize downtime during upgrades' provides information about using Blue/Green Deployments to minimize downtime during database upgrades. The main table lists one database entry:

DB identifier	Status	Role	Engine	Region & AZ	Size	Actions
customer-features-rds-db-instance	Backing-up	Instance	MySQL Community	us-east-1c	db.t3.micro	-

## 47. WE WILL NOW CREATE A TABLE CONTAINING THE CUSTOMER FEATURES DATA IN THE CSV FILE INTO THE DATABASE. WE CAN DO IT BY EITHER USING THE CLI OR AN EXTRA TOOL WHICH IS CALLED MYSQL WORKBENCH.

IF YOU DO NOT HAVE THE MYSQL WORKBENCH, YOU CAN GO TO GOOGLE AND SEARCH FOR MYSQL WORKBENCH, THEN, DOWNLOAD, THEN, INSTALL.

<https://dev.mysql.com/downloads/workbench/>

### MySQL Community Downloads

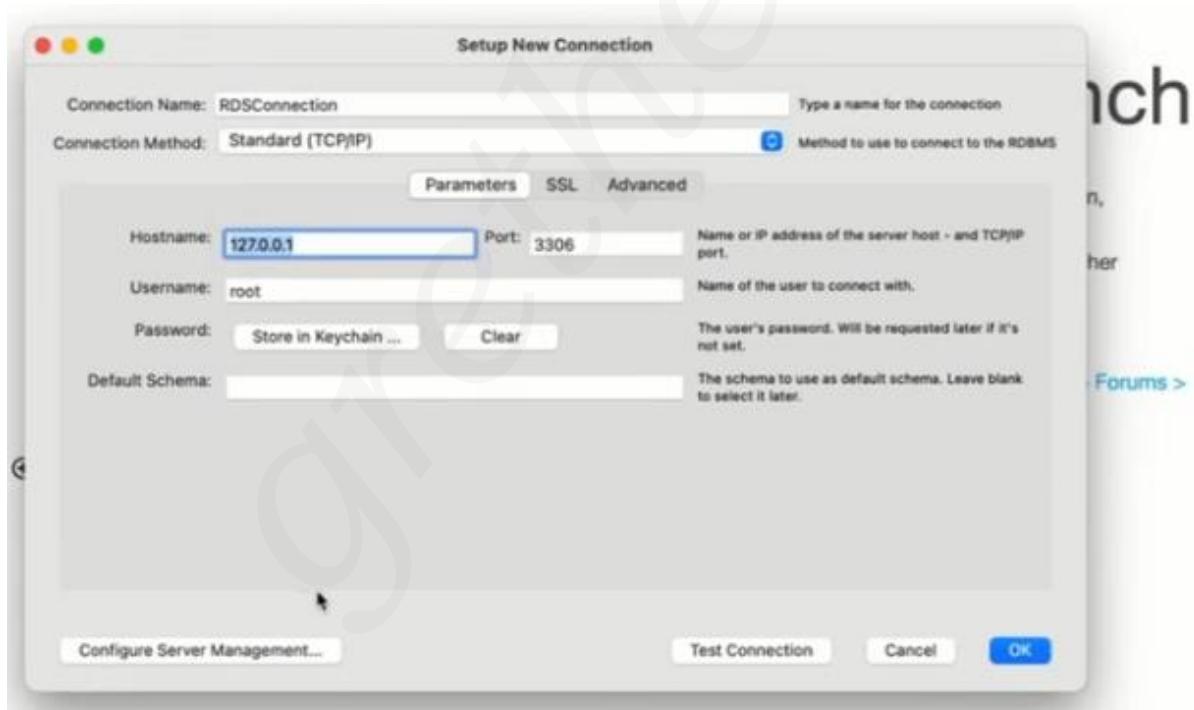
↳ MySQL Workbench

The screenshot shows the MySQL Workbench download page for version 8.0.36. It features tabs for 'General Availability (GA) Releases' (selected), 'Archives', and a help icon. The main section is titled 'MySQL Workbench 8.0.36' and includes a dropdown for 'Select Operating System' set to 'Microsoft Windows'. A 'Recommended Download' section shows the 'MySQL Installer for Windows' package, which is described as containing 'All MySQL Products. For All Windows Platforms. In One Package.' An image of the Windows logo is shown next to the text. Below this, a note states: 'Starting with MySQL 5.6 the MySQL Installer package replaces the standalone MSI packages.' A 'Windows (x86, 32 & 64-bit), MySQL Installer MSI' download button is available, along with a 'Go to Download Page >' link. A 'Other Downloads:' section lists 'Windows (x86, 64-bit), MSI Installer' with version 8.0.36, size 42.0M, and a 'Download' button. A note below it encourages users to verify downloads using MD5 checksums and GnuPG signatures. The note reads: 'We suggest that you use the [MD5 checksums](#) and [GnuPG signatures](#) to verify the integrity of the packages you download.'

**48. ONCE INSTALLED, MAKE A CONNECTION BETWEEN MYSQL WORKBENCH AND RDS DATABASE. OPEN MYSQL WORKBENCH, CLICK THE PLUS (+) BUTTON TO CREATE THE CONECTION.**



**FOR THE CONNECTION NAME, MAKE IT RDSCONNECTION. HOST NAME IS ACCORDING TO YOU WHAT IS GOING TO BE. TO CHECK, GO TO YOUR AMAZON RDS, AND CHECK FOR THE ENDPOINT & PORT.**



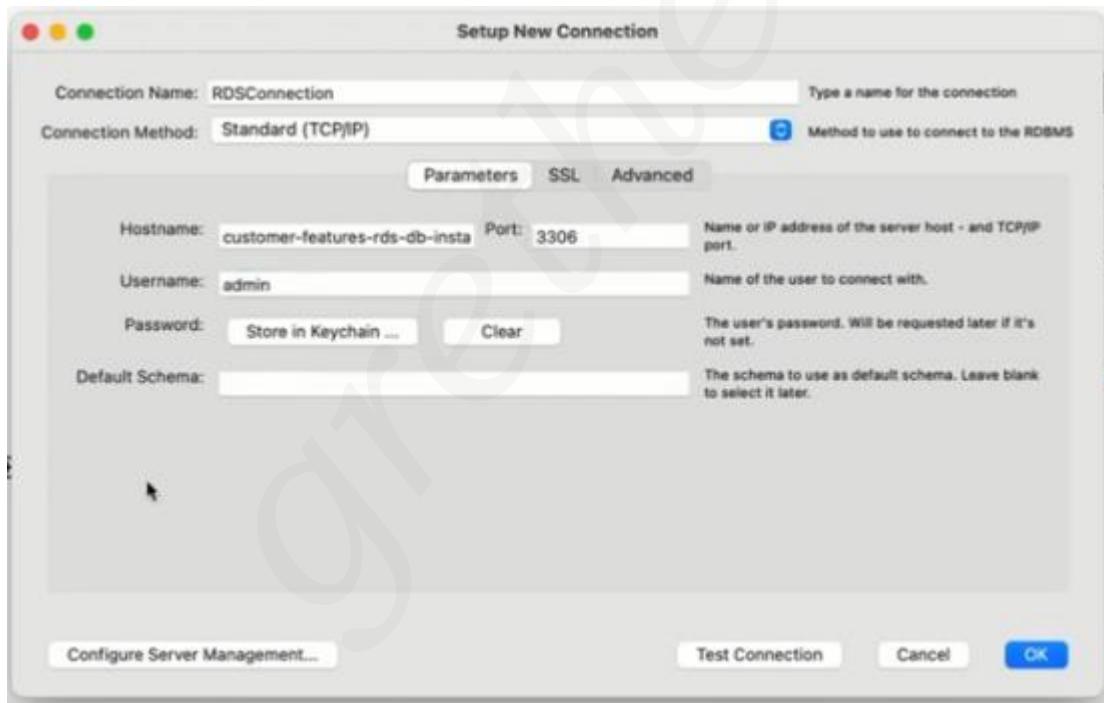
**COPY THE ENDPOINT DETAILS. GO BACK TO THE MYSQL WORKBENCH, THEN, PASTE IT TO THE HOSTNAME. THE PORT 3306 IS THE PORT OF MY MYSQL, SO, IT WILL REMAIN THE SAME.**

The screenshot shows the 'Amazon RDS' interface with the 'Connectivity & security' tab selected. On the left, there's a sidebar with various navigation options like Dashboard, Databases, Query Editor, etc. The main content area displays endpoint details:

Endpoint & port	Networking	Security
Endpoint customer-features-rds-db-instance.crrwucvfcmbc.us-east-1.rds.amazonaws.com	Availability Zone us-east-1c	VPC security groups default (sg-00757adc39bf9625e) Active
Port 3306	VPC vpc-0041e736a2d14cfa2	Publicly accessible No
	Subnet group default-vpc-0041e736a2d14cfa2	Certificate authority <a href="#">Info</a> rds-ca-2019
	Subnets subnet-061069d72790ee8be subnet-0ab745a585106a535 subnet-0536055a78fadcc9a subnet-0c09e23ebb408adf0 subnet-066078a3fb43745dc subnet-0ff66ccebd5b8ad21	Certificate authority date August 22, 2024, 19:08 (UTC+02:00)
	Network type IPv4	DB instance certificate expiration date August 22, 2024, 19:08 (UTC+02:00)

**FOR THE USERNAME, PUT admin, AS IT IS THE NAME WE PUT IN THE DB INSTANCE.**

**THEN, CICK, TEST CONNECTION TO TEST IF IT WORKS PROPERLY.**



## 49. THE RESULT WILL GIVE US FAILED CONNECTION.



## 50. GO BACK TO AMAZON RDS, IF YOU NOTICE IN THE CONNECTIVITY & SECURITY, IN THE SECURITY FIELD - PUBLICLY ACCESSIBLE IS NO.

A screenshot of the Amazon RDS "Connectivity & security" page. The left sidebar shows "Databases" is selected. The main content area has tabs for "Connectivity & security", "Monitoring", "Logs & events", "Configuration", "Maintenance & backups", and "Tags". The "Connectivity & security" tab is active. It displays three sections: "Endpoint & port", "Networking", and "Security".

Endpoint & port	Networking	Security
Endpoint customer-features-rds-db-instance.crrwucvcfmc.us-east-1.rds.amazonaws.com	Availability Zone us-east-1c	VPC security groups default (sg-00757adc39bf9625e) <input checked="" type="checkbox"/> Active
Port 3306	VPC vpc-0041e736a2d14cfa2	Publicly accessible No
	Subnet group default-vpc-0041e736a2d14cfa2	Certificate authority <a href="#">Info</a> rds-ca-2019
	Subnets subnet-061069d72790ee8be subnet-0ab745a585106a535 subnet-0536055a78fadcc5a subnet-0c09e23eb408adf0 subnet-066078a3fb43745dc subnet-0ff66ccebd5b8ad21	Certificate authority date August 22, 2024, 19:08 (UTC+02:00)
	Network type IPv4	DB instance certificate expiration date August 22, 2024, 19:08 (UTC+02:00)

**51. WE HAVE TO CONFIGURE THE SECURITY – PUBLICLY AVAILABLE TO YES. TO DO THIS, CLICK MODIFY.**

The screenshot shows the Amazon RDS console for a database named "customer-features-rds-db-instance". At the top, there is a green banner with a success message: "Successfully created database customer-features-rds-db-instance". Below the banner, the database identifier is listed as "customer-features-rds-db-instance". The "Summary" section displays various metrics: CPU usage at 5.54%, Status as "Available", Class as "db.t3.micro", Role as "Instance", Current activity with 0 connections, Engine as "MySQL Community", and Region & AZ as "us-east-1c". Below the summary, there are tabs for "Connectivity & security", "Monitoring", "Logs & events", "Configuration", "Maintenance & backups", and "Tags". The "Connectivity & security" tab is selected, showing sub-sections for "Endpoint & port", "Networking", and "Security".

**SCROLL DOWN, AND FIND THE CONFIGURATION WHERE WE CAN SET THE PUBLIC ACCESS TO YES.  
IN THE CONNECTIVITY, CLICK ADDITIONAL CONFIGURATION. SELECT, PUBLICLY ACCESSIBLE.**

The screenshot shows the "Connectivity" configuration page for the database instance. Under "Network type", the "IPv4" option is selected, with a note that resources can communicate only over the IPv4 addressing protocol. The "DB subnet group" is set to "default-vpc-0041e736a2d14cfa2". Under "Security group", the "Choose security groups" dropdown contains "default". In the "Certificate authority" section, the value "rds-ca-2019" is listed. At the bottom, there is a "▼ Additional configuration" section which includes a "Public access" field.

Choose security groups ▾

default X

Certificate authority [Info](#)

Using a server certificate provides an extra layer of security by validating that the connection is being made to an Amazon database. It does so by checking the server certificate that is automatically installed on all databases that you provision.

rds-ca-2019 ▾

▼ Additional configuration

Public access

Publicly accessible

RDS assigns a public IP address to the database. Amazon EC2 instances and other resources outside of the VPC can connect to your database. Resources inside the VPC can also connect to the database. Choose one or more VPC security groups that specify which resources can connect to the database.

Not publicly accessible

No IP address is assigned to the DB instance. EC2 instances and devices outside the VPC can't connect.

Database port

Specify the TCP/IP port that the DB instance will use for application connections. The application connection string must specify the port number. The DB security group and your firewall must allow connections to the port. [Learn more](#) ⓘ

3306

**SCROLL DOWN, THEN SELECT, CONTINUE. YOU WILL SEE WHAT IS MODIFIED.**

**CLICK MODIFY DB INSTANCE.**

RDS > Databases > Modify DB instance: customer-features-rds-db-instance

## Modify DB instance: customer-features-rds-db-instance

**Summary of modifications**

You are about to submit the following modifications. Only values that will change are displayed. Carefully verify your changes and click **Modify DB Instance**.

Attribute	Current value	New value
Public accessibility	No	Yes

**Schedule modifications**

When to apply modifications

Apply during the next scheduled maintenance window

Current maintenance window: June 17, 2023 10:19 - 10:49 UTC+2

Apply immediately

The modifications in this request and any pending modifications will be asynchronously applied as soon as possible, regardless of the maintenance window setting for this database instance.

Cancel Back **Modify DB instance**

## Modify DB instance: customer-features-rds-db-instance

### Summary of modifications

You are about to submit the following modifications. Only values that will change are displayed. Carefully verify your changes and click Modify DB Instance.

Attribute	Current value	New value
Public accessibility	No	Yes

### Schedule modifications

#### When to apply modifications

- Apply during the next scheduled maintenance window

Current maintenance window: June 17, 2023 10:19 - 10:49 UTC+2

- Apply immediately

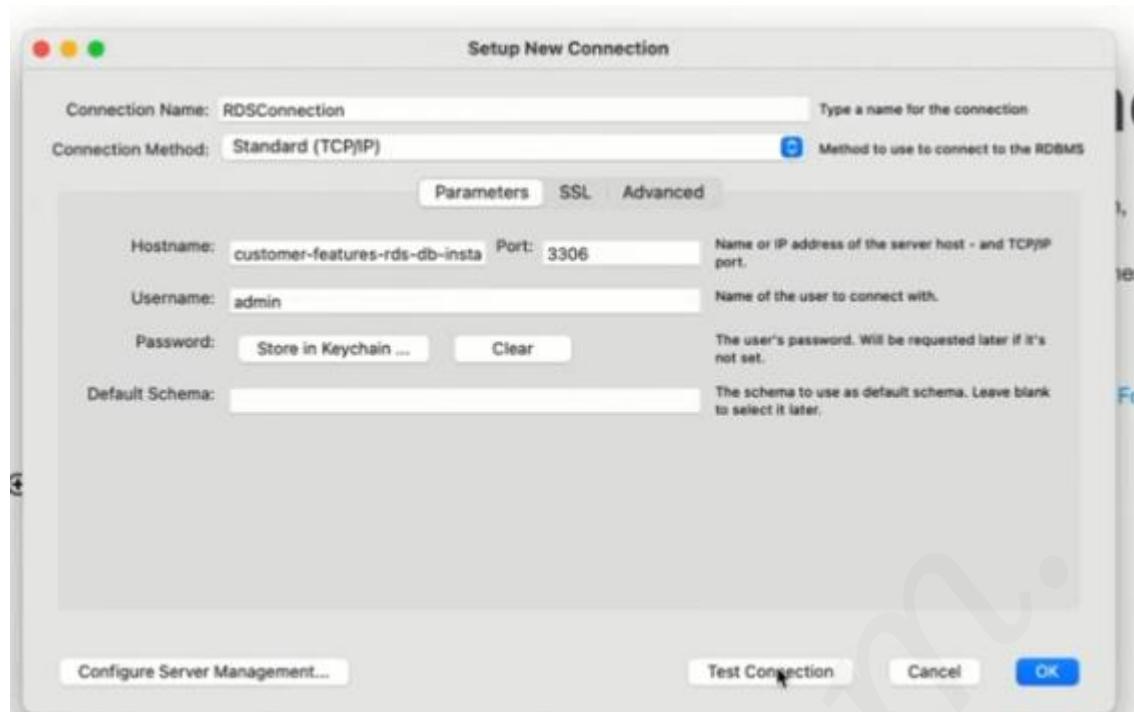
The modifications in this request and any pending modifications will be asynchronously applied as soon as possible, regardless of the maintenance window setting for this database instance.

## SUCCESSFULLY MOODIED INSTANCE.

The screenshot shows the Amazon RDS console interface. On the left, there is a navigation sidebar with various options like Dashboard, Databases, Query Editor, etc. The main area displays a success message: "Successfully modified instance customer-features-rds-db-instance". Below this, another message says "Successfully created database customer-features-rds-db-instance". A callout box provides information about Blue/Green Deployments. The "Databases" section shows a table with one row for "customer-features-rds-db-instance".

DB identifier	Status	Role	Engine	Region & AZ	Size	Actions
customer-features-rds-db-instance	Available	Instance	MySQL Community	us-east-1c	db.t3.micro	1 Action

## 52. GO BACK TO MYSQL WORKBENCH, THEN, TEST THE CONNECTION.



## 53. HERE, WE WILL GET THE FAILED CONNECTION AGAIN.



**54. THIS IS BECAUSE WE FORGOT TO EDIT THE INBOUND RULE. SO, TO ACHIEVE THE CONNECTION, WE NEED TO FIX THE ACCESS ISSUE, SPECIFICALLY THE INBOUND RULES OF THE SECURITY GROUP ATTACHED TO THE RDS DATABASE INSTANCE.**

**SO NOW, WE ARE GOING TO EDIT THE INBOUND RULE IN ORDER TO OPEN THE PORT 3306, THE MYSQL PORT.**

**IN THE SEARCH BAR, GO TO EC2.**

The screenshot shows the AWS search interface. In the top navigation bar, there is a search bar with the text 'EC2'. Below the search bar, the results are displayed under the heading 'Search results for 'EC2'' and 'Try searching with longer queries for more relevant results'. The results are categorized into 'Services (12)', 'Features (53)', and 'Resources (New)'. Under 'Services', the first result is 'EC2' with the subtext 'Virtual Servers in the Cloud'. Other services listed include 'EC2 Image Builder', 'Amazon Inspector', and 'AWS Firewall Manager'. On the left side, there is a sidebar for 'Amazon RDS' with various options like Dashboard, Databases, Query Editor, and Performance insights.

**55. GO TO THE SECURITY GROUP AND CREATE SECURITY GROUP.**

The screenshot shows the AWS Security Groups page. The left sidebar includes links for Launch Templates, Spot Requests, Savings Plans, Reserved Instances, Dedicated Hosts, Scheduled Instances, Capacity Reservations, Images (AMIs, AMI Catalog), Elastic Block Store (Volumes, Snapshots, Lifecycle Manager), and Network & Security (Security Groups, Elastic IPs, Placement Groups, Key Pairs, Network Interfaces). The main content area displays a table titled 'Security Groups (3)'. The table has columns for Name, Security group ID, Security group name, VPC ID, Description, and Owner. The data is as follows:

Name	Security group ID	Security group name	VPC ID	Description	Owner
-	sg-0ce2e3b4e3be04e92	SG-Open-HTTP	vpc-0041e736a2d14cfa2	Allows HTTP Traffic	7496011
-	sg-04c7ce4cb6a0e059f	launch-wizard-1	vpc-0041e736a2d14cfa2	launch-wizard-1 create...	7496011
-	sg-00757adc39bf9625e	default	vpc-0041e736a2d14cfa2	default VPC security gr...	7496011

**56. WE WILL CALL IT, SG-OPEN-MYSQL. DESCRIPTION IS ALLOWS MYSQL ACCESS TO DEVELOPERS. AND MAKE THE VPC DEFAULT.**

A security group acts as a virtual firewall for your instance to control inbound and outbound traffic. To create a new security group, complete the fields below.

**Basic details**

Security group name [Info](#)  
SG-Open-MySQL  
Name cannot be edited after creation.

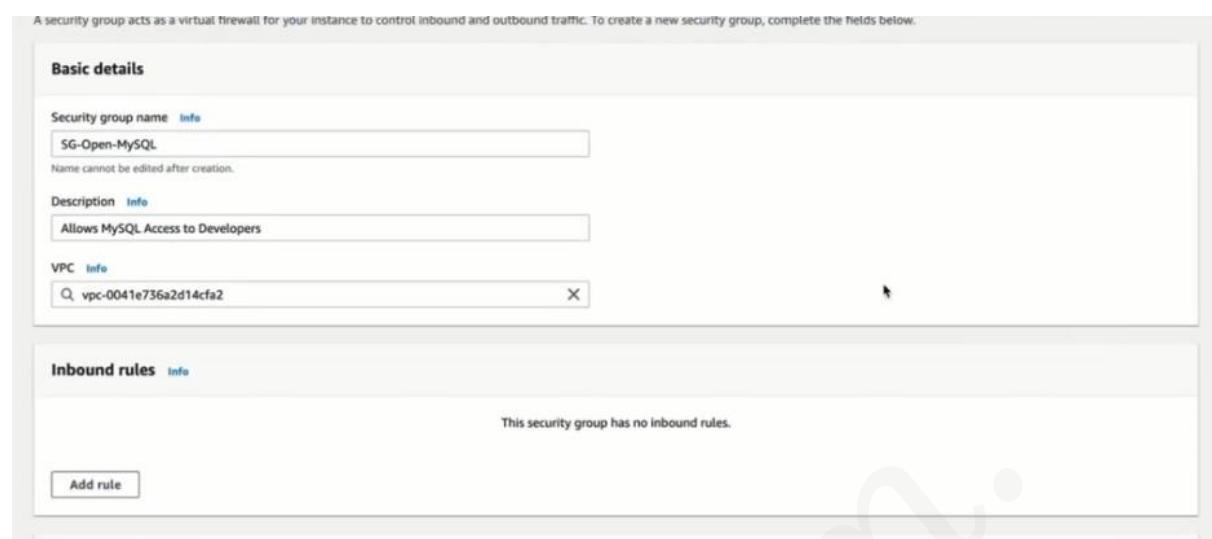
Description [Info](#)  
Allows MySQL Access to Developers

VPC [Info](#)  
Q vpc-0041e736a2d14cfa2 X

**Inbound rules** [Info](#)

This security group has no inbound rules.

Add rule



**57. CLICK ADD RULE IN THE INBOUND RULE. IN THE TYPE, SELECT CUSTOM TCP. PROTOCOL IS TCP. PORT IS 3306. DESTINATION IS ANYWHERE.**

VPC [Info](#)  
Q vpc-0041e736a2d14cfa2 X

**Inbound rules** [Info](#)

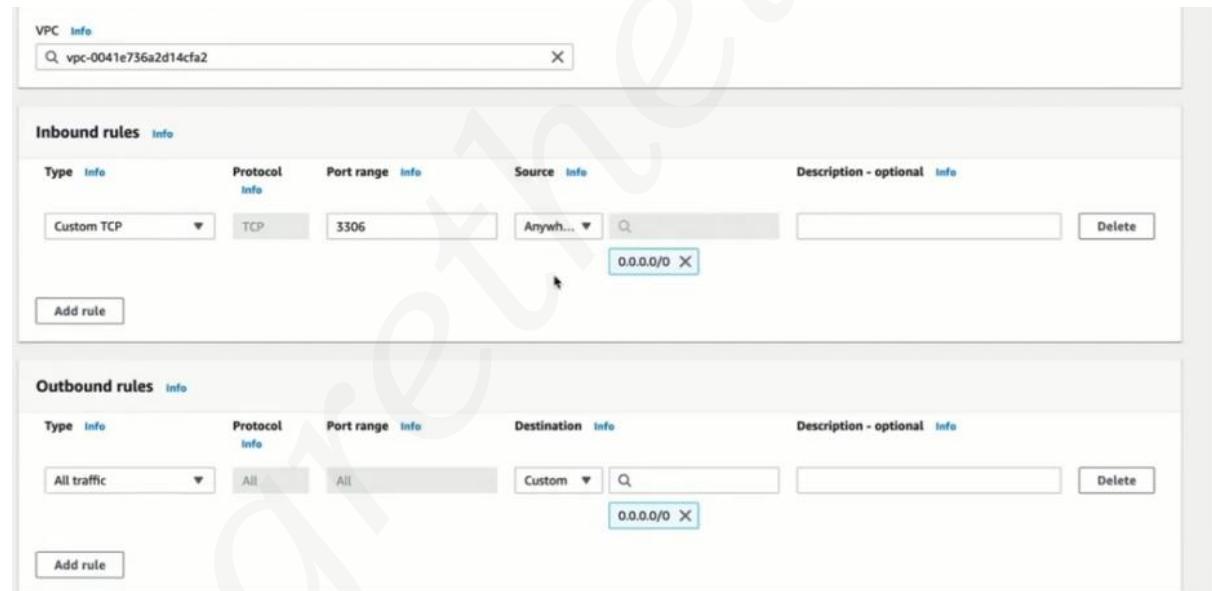
Type <a href="#">Info</a>	Protocol <a href="#">Info</a>	Port range <a href="#">Info</a>	Source <a href="#">Info</a>	Description - optional <a href="#">Info</a>
Custom TCP	TCP	3306	Anywh... ▾	<input type="text"/> 0.0.0.0/0 X

Add rule

**Outbound rules** [Info](#)

Type <a href="#">Info</a>	Protocol <a href="#">Info</a>	Port range <a href="#">Info</a>	Destination <a href="#">Info</a>	Description - optional <a href="#">Info</a>
All traffic	All	All	Custom ▾	<input type="text"/> 0.0.0.0/0 X

Add rule



## 58. MAKE THE OUTBOUND RULE DEFAULT. THEN, CLICK CREATE SECURITY GROUP.

**Outbound rules**

Type	Protocol	Port range	Destination	Description - optional
All traffic	All	All	Custom	0.0.0.0/0

**Tags - optional**

No tags associated with the resource.

**Create security group**

## 59. IT IS NOW IN THE SECURITY GROUP LIST. BUT, WE HAVE TO ATTACH IT TO THE RDS DATABASE INSTANCE.

**Security Groups (4)**

Name	Security group ID	Security group name	VPC ID	Description	Owner
-	sg-04c7ce4cb6a0e059f	launch-wizard-1	vpc-0041e736a2d14cfa2	launch-wizard-1 create...	7496011
-	sg-0ce2e3b4e3be04e92	SG-Open-HTTP	vpc-0041e736a2d14cfa2	Allows HTTP Traffic	7496011
-	sg-00757adc39bf9625e	default	vpc-0041e736a2d14cfa2	default VPC security gr...	7496011
-	sg-05eae22adcf0dce56	SG-Open-MySQL	vpc-0041e736a2d14cfa2	Allows MySQL Access t...	7496011

**sg-05eae22adcf0dce56 - SG-Open-MySQL**

**Details**

Security group name SG-Open-MySQL	Security group ID sg-05eae22adcf0dce56	Description Allows MySQL Access to Developers	VPC ID vpc-0041e736a2d14cfa2
Owner 749601114921	Inbound rules count 1 Permission entry	Outbound rules count 1 Permission entry	

**Inbound rules**

Name	Security group rule...	IP version	Type	Protocol	Port rang...
-	sgr-000a62b4b15cc1b...	IPv4	MySQL/Aurora	TCP	3306

## 60. GO BACK TO AMAZON RDS, THEN, CLICK MODIFY.

The screenshot shows the Amazon RDS console with the database 'customer-features-rds-db-instance' selected. The 'Modify' button is located in the top right corner of the main content area. The left sidebar contains navigation links for various RDS features like Dashboard, Databases, and Performance insights.

## 61. SCROLL DOWN AND GO TO CONNECTIVITY, THEN, SECURITY GROUP. THEN YOU WILL FIND THE ONE THAT WE JUST CREATED. CHOOSE DEAFULT AND SG-OPEN-MYSQL.

The screenshot shows the 'Connectivity & security' tab in the Amazon RDS console. Under the 'Security' section, the 'DB subnet group' dropdown is set to 'default-vpc-0041e736a2d14cfa2'. Below it, the 'Choose security groups' dropdown lists several security groups, with 'default' and 'SG-Open-MySQL' selected. The 'default' option is highlighted with a blue selection bar.

## 62. SCROLL DOWN AND CLICK CONTINUE.

The screenshot shows the 'Modify DB instance' configuration page. It includes sections for CloudWatch Logs (with options for Audit log, Error log, General log, and Slow query log), IAM role (using the RDS service-linked role), Maintenance (enabling auto minor version upgrade), DB instance maintenance window (set to Saturday 08:00 UTC for 0.5 hours), and Deletion protection (disabled). At the bottom right is a large orange 'Continue' button.

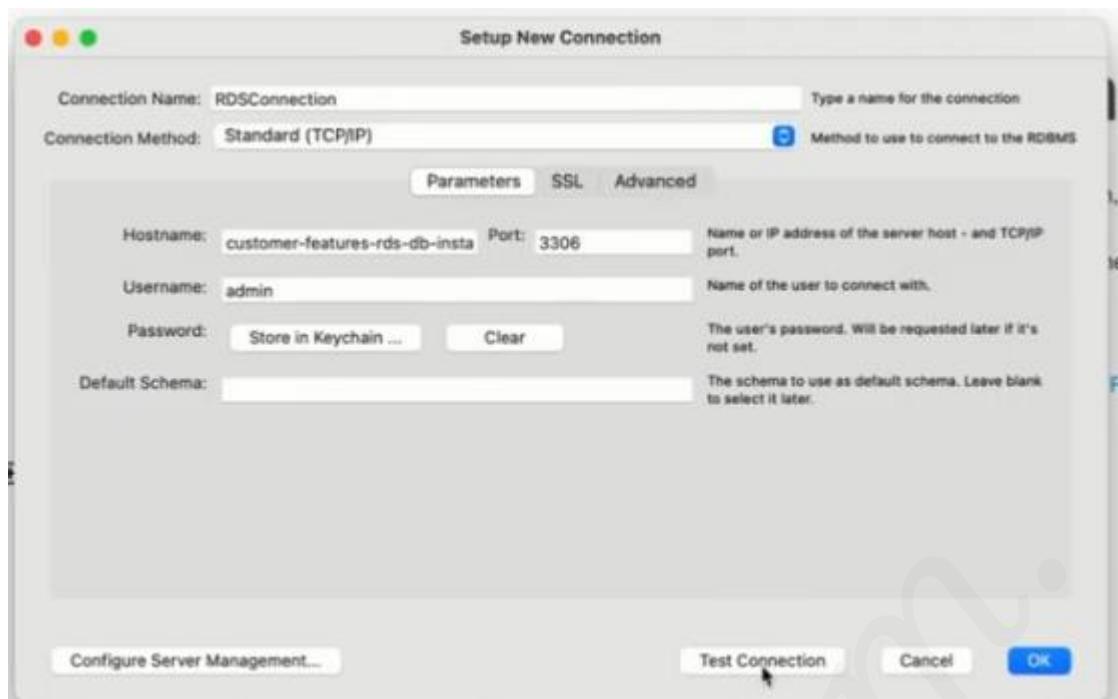
**63. HERE, YOU WILL SEE THE SUMMARY OF MODIFICATIONS. THEN, CLICK MODIFY DB INSTANCE.**

The screenshot shows the 'Modify DB instance' page for the database 'customer-features-rds-db-instance'. On the left, there's a sidebar with various RDS management links like Dashboard, Databases, Query Editor, etc. The main area has two sections: 'Summary of modifications' and 'Schedule modifications'. In the 'Summary of modifications' section, it lists a single change: 'Security group' set to 'default'. In the 'Schedule modifications' section, there are two options: 'Apply during the next scheduled maintenance window' (selected) and 'Apply immediately'. Below these, there are 'Cancel', 'Back', and 'Modify DB instance' buttons.

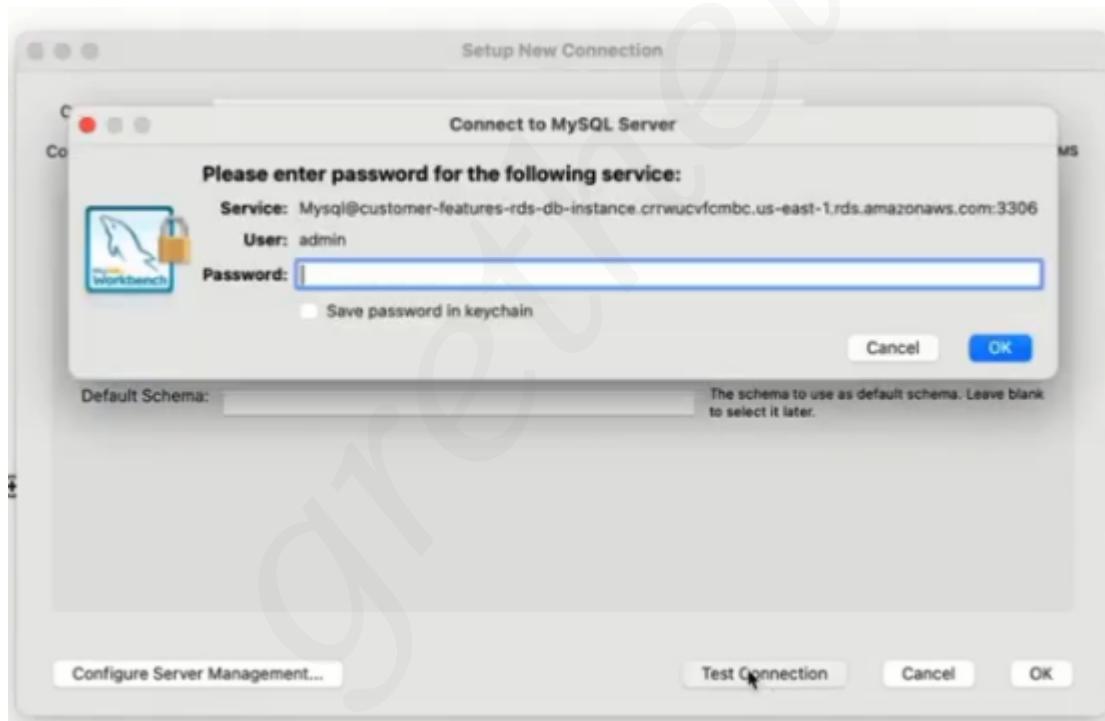
**64. SUCCESSFULLY MODIFIED.**

The screenshot shows the 'Databases' page with one entry: 'customer-features-rds-db-instance'. A success message at the top says 'Successfully modified instance customer-features-rds-db-instance'. A note below it encourages using Blue/Green Deployments. The table lists the database details: Status (Available), Role (Instance), Engine (MySQL Community), Region & AZ (us-east-1c), Size (db.t3.micro), and Actions (1 Action).

## 65. GO BACK TO MYSQL WORKBENCH TO TEST THE CONNECTION.



CONGRATULATIONS! THE CONNECTION IS SUCCESSFUL! ENTER THE PASSWORD.

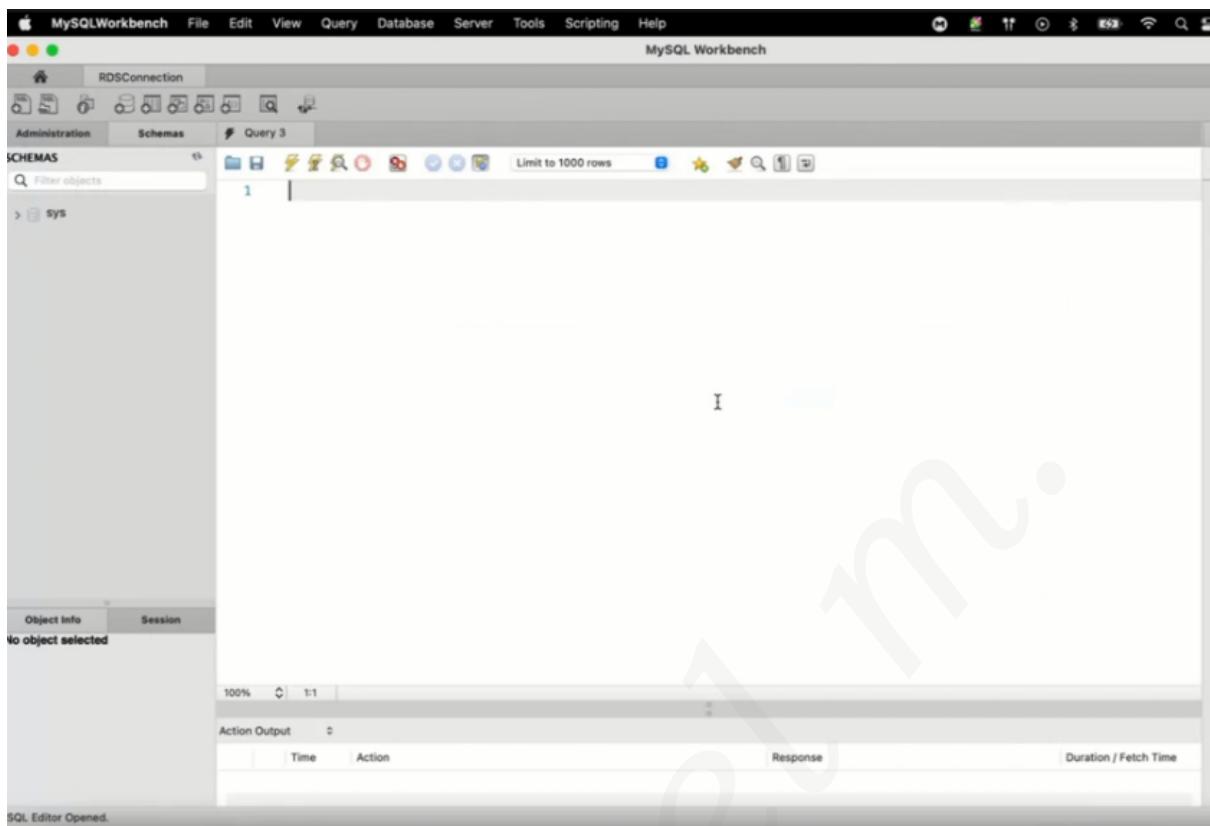




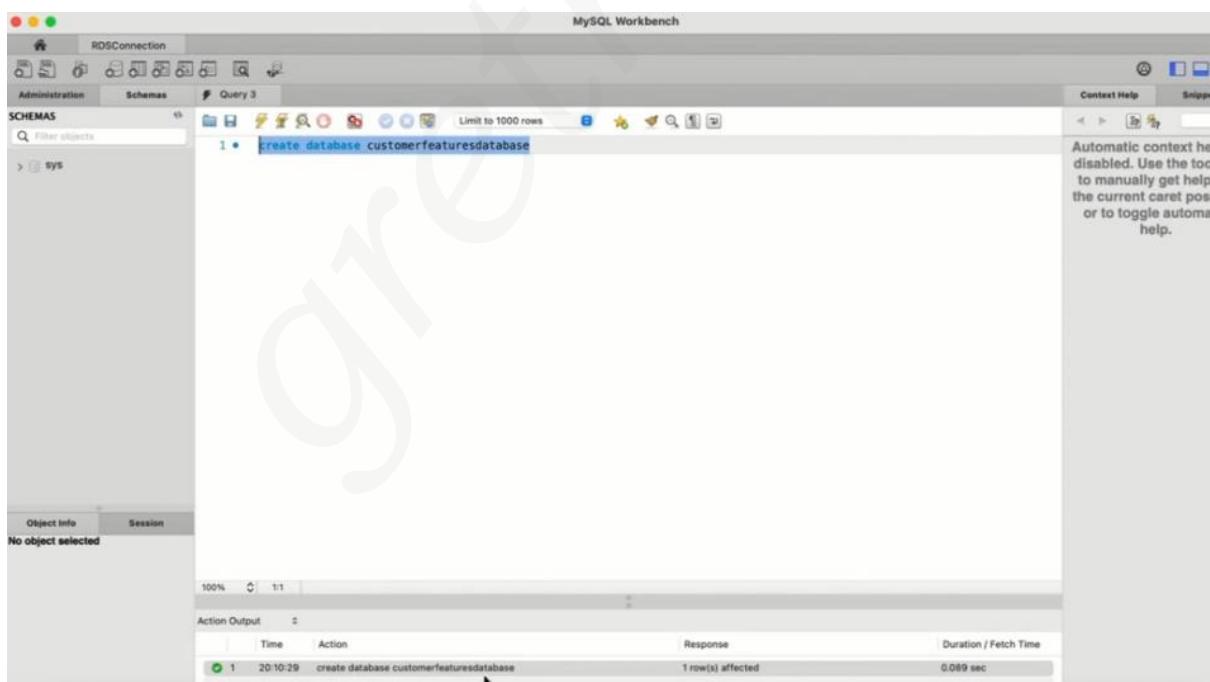
**66. THE CONNECTION IS NOW ESTABLISHED, DOUBLE CLICK THE RDS CONNECTION. IT IS NOW OPENING THE SQL EDITOR.**



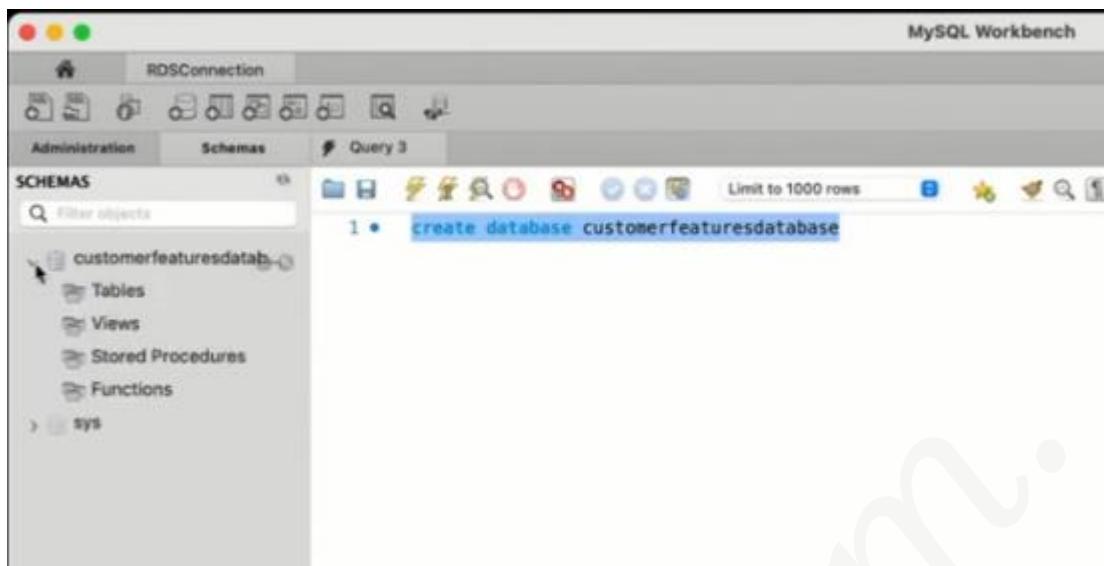
**67. AND THIS IS WHERE WE ARE ABLE TO CREATE OUR TABLE WITHIN OUR RDS DATABASE INSTANCE.**



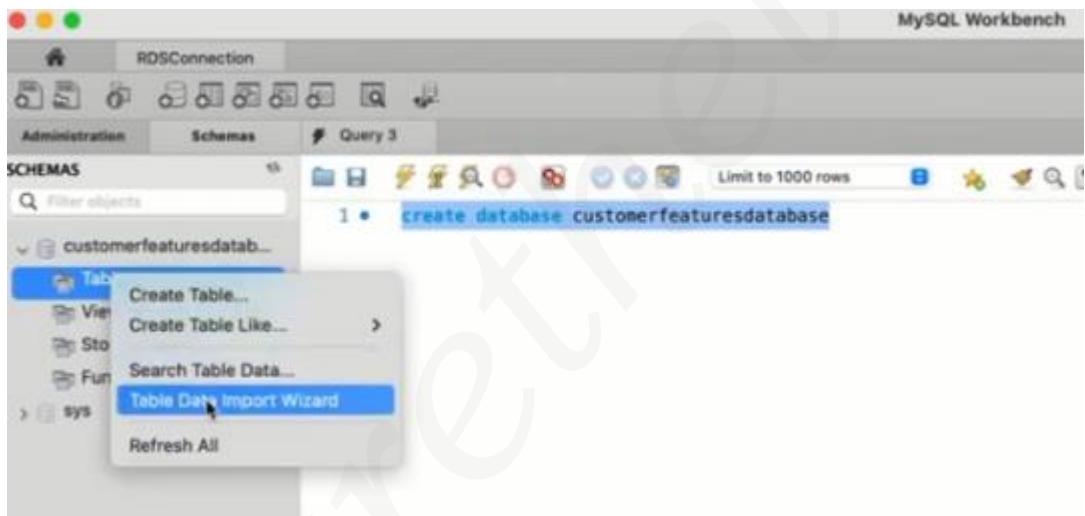
**68. WE WILL NOW RUN A QUERY, CREATE DATABASE...THEN EXECUTE.**



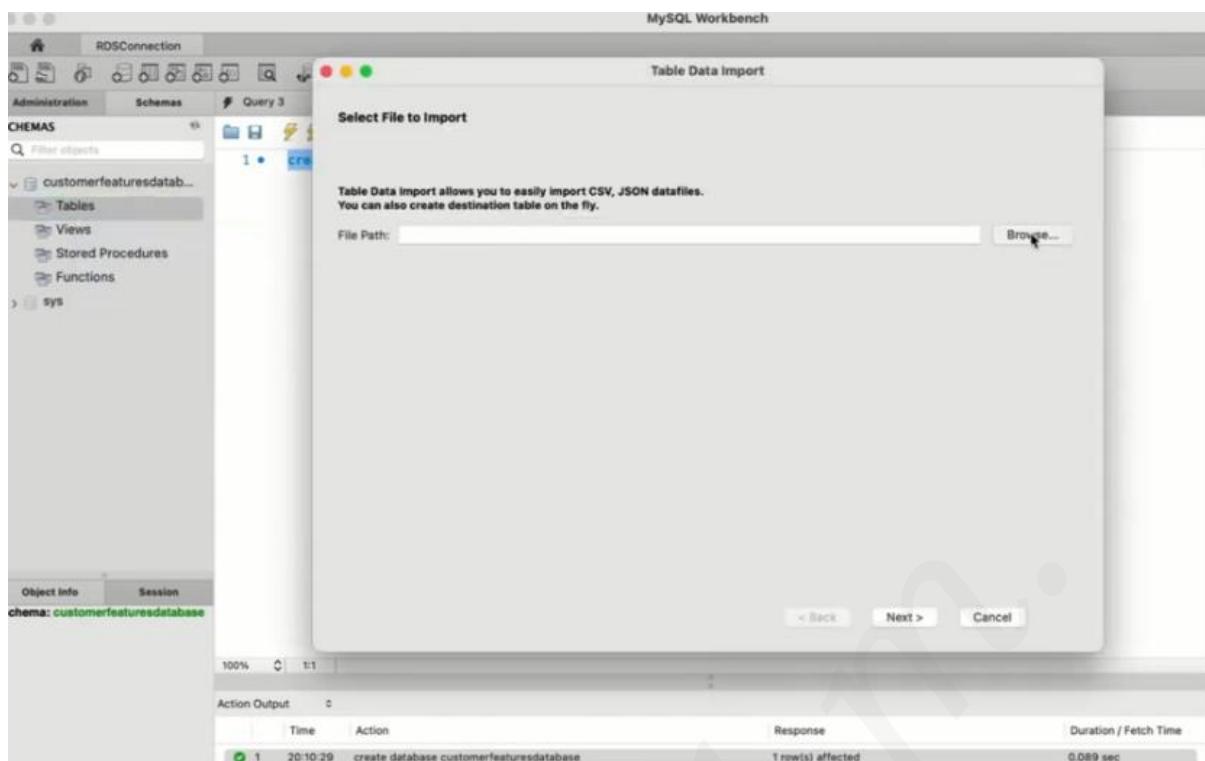
**69. REFRESH THE SCHEMA TO SEE THE DATABASE.**



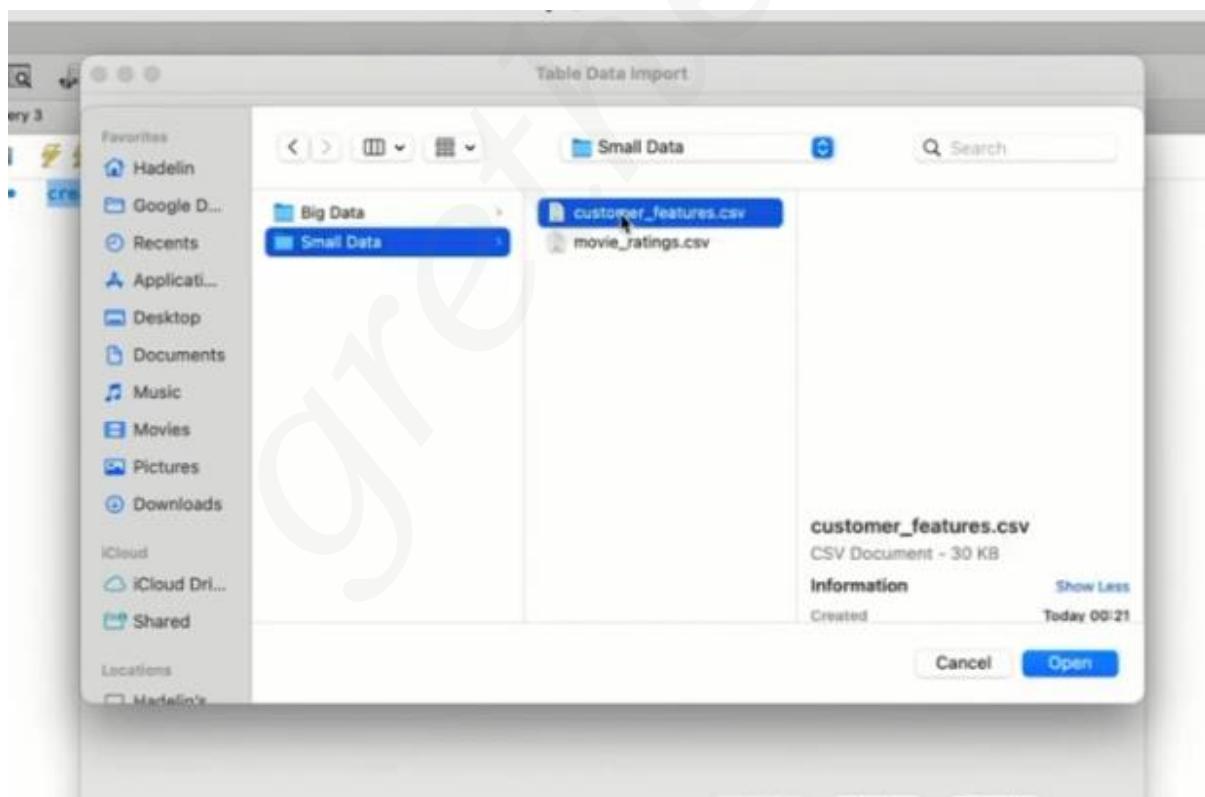
**70. CLICK TABLE, CLICK RIGHT, SELECT TABLE DATA IMPORT WIZARD.**



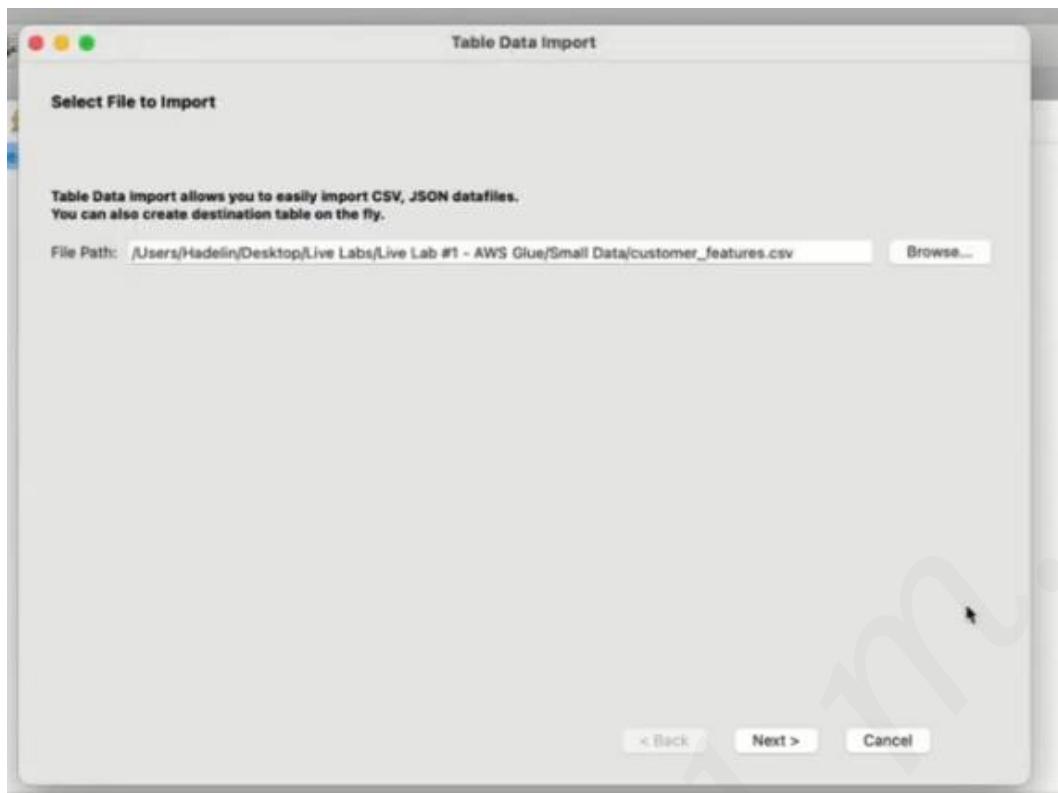
## 71. BROWSE THE MACHINE TO FIND THE CSV FILE.



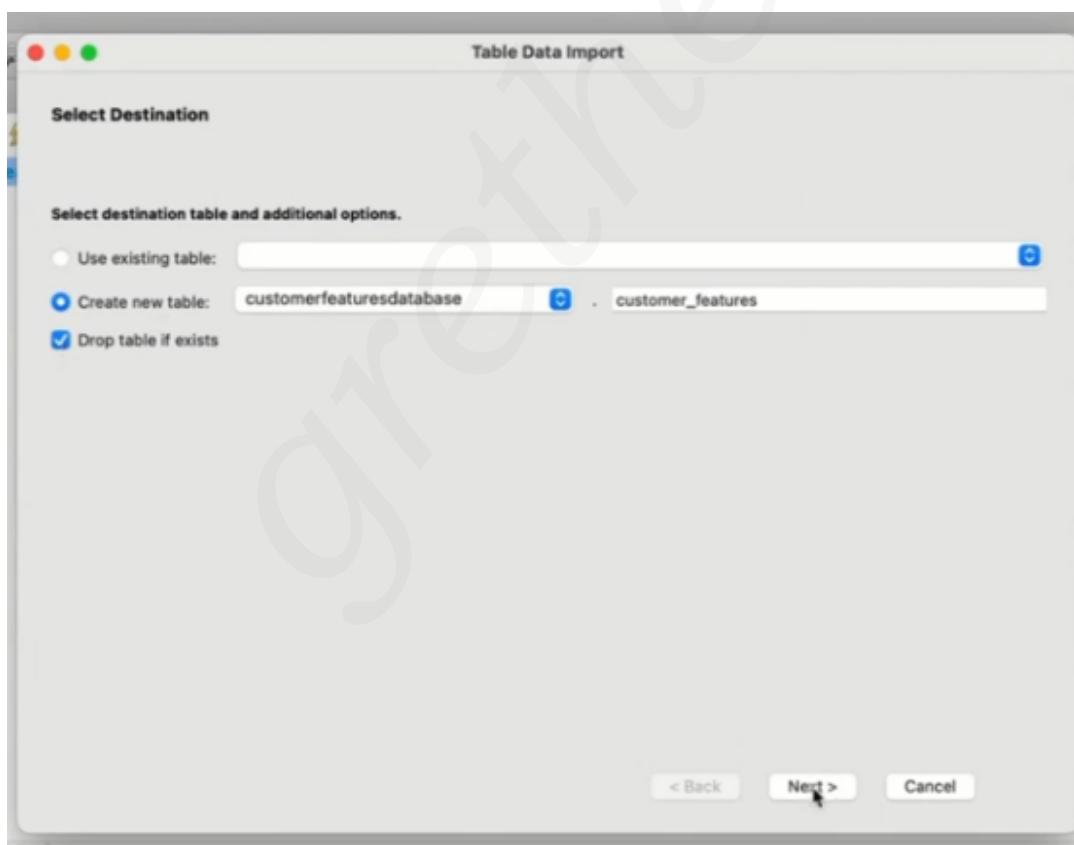
HERE, I CHOSE THE SMALL DATA, CUSTOMER FEATURES FILE.



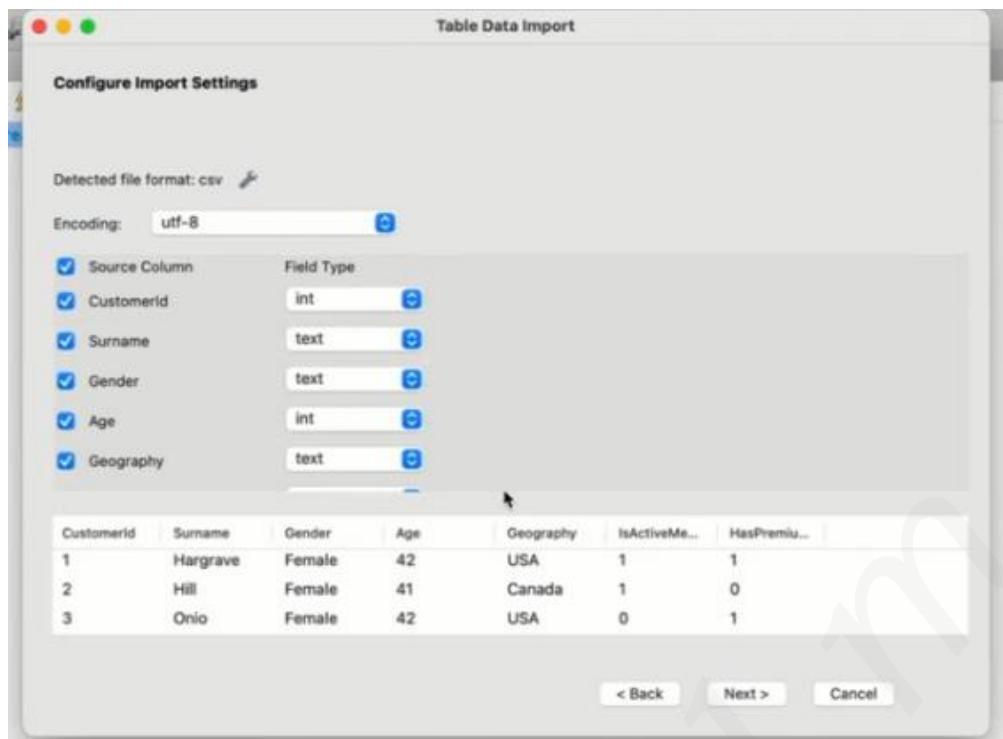
**72. CLICK NEXT.**



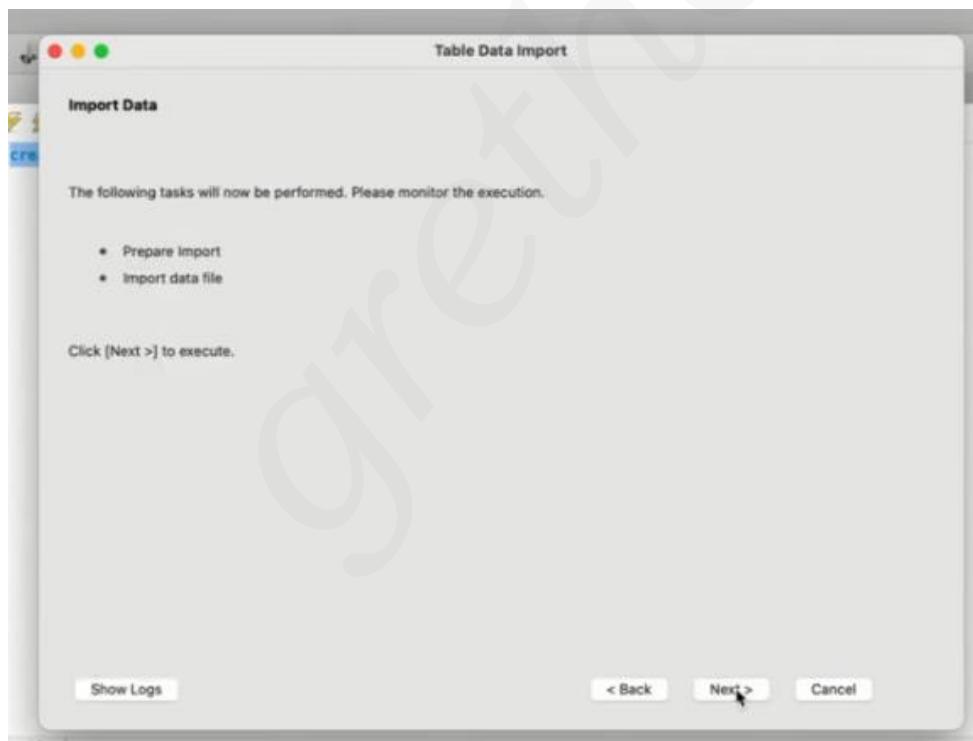
**73. YOU CAN CLICK THE DROP TABLE IF EXISTS IF YOU WANT TO. THEN, CLICK NEXT.**



**74. YOU CAN CHANGE THE TYPE IF YOU WANT. BUT, I JUST LEAVE IT AS IT IS. CLICK, NEXT.**

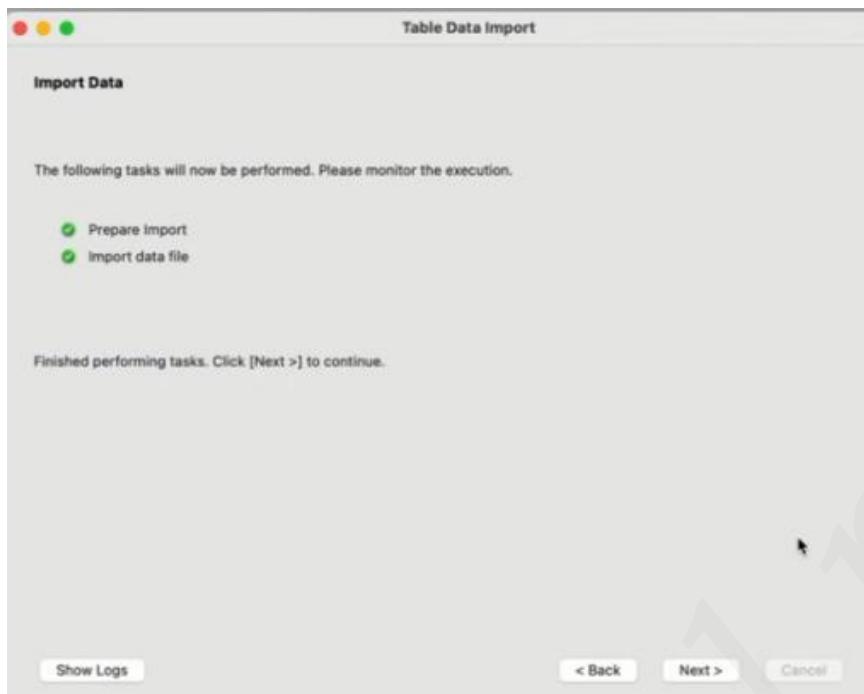


**75. IN THE IMPORT DATA, CLICK NEXT TO EXECUTE.**



**76. IT IS NOW BEGINNING THE IMPORT AND IT IS IMPORTING THE DATA TO POPULATE IT IN THE TABLE OF THE DATABASE IN THE RDS DATABASE INSTANCE THAT WE CREATED AND THAT WE CONNECTED SUCCESSFULLY TO MYSQL WORKBENCH.**

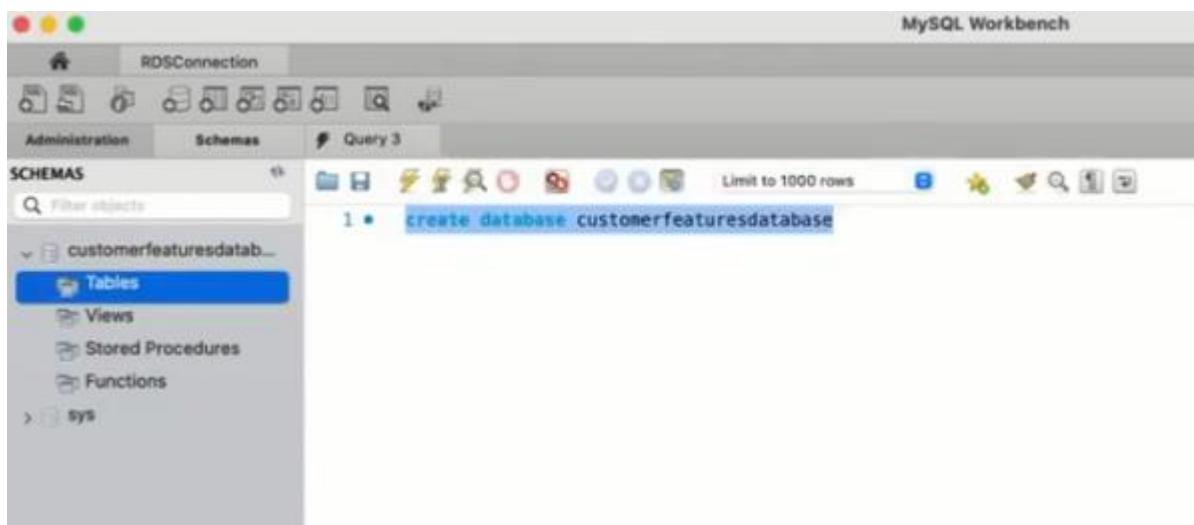
**CLICK NEXT.**



**CLICK FINISH. THE RECORDED IS THEN FINISHED.**



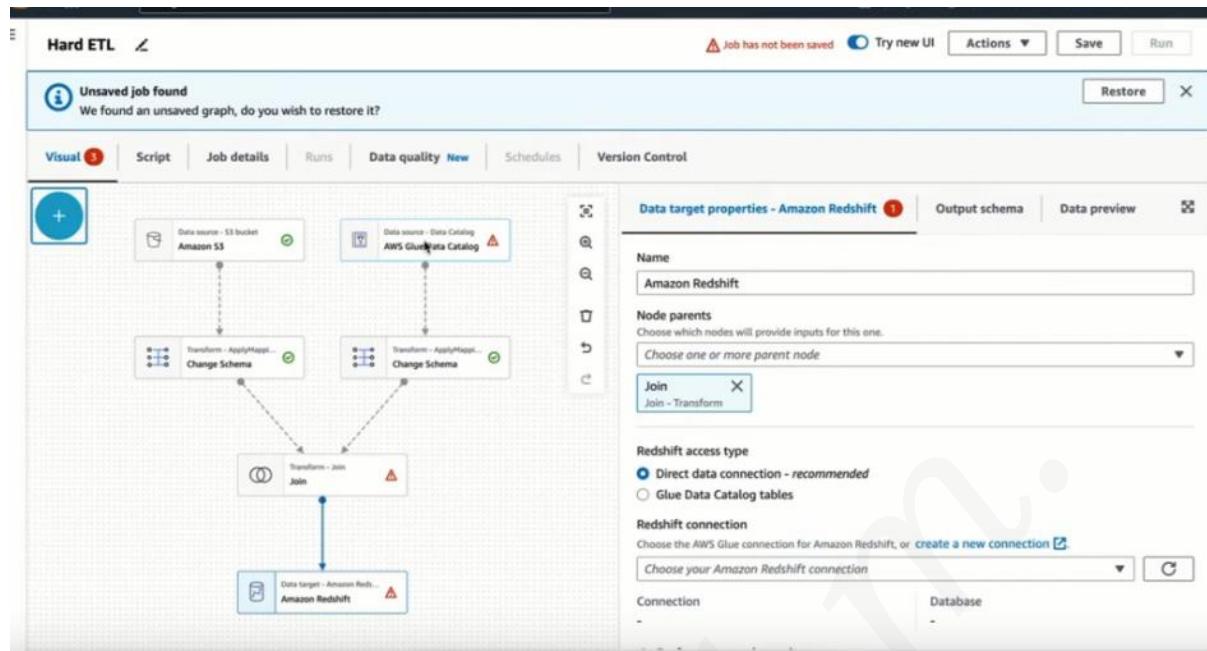
**77. IN THE RDS DATABASE, WE HAVE THE DATA POPULATED.**



**78. GO BACK TO THE AMAZON RDS CONSOLE. OUR DATABASE INSTANCE HAS THE TABLE WELL POPULATED WITH THE DATA OF OUR CSV FILE CONTAINING THE CUSTOMER FEATURES.**

A screenshot of the Amazon RDS console. The left sidebar shows "Databases" selected. The main area shows the "customer-features-rds-db-instance" database. The "Summary" tab is active, displaying details like DB identifier, CPU usage (3.52%), Status (Available), Class (db.t3.micro), Role, Current activity (0 Connections), Engine (MySQL Community), and Region &amp; AZ (us-east-1c). Below the summary are tabs for "Connectivity &amp; security", "Monitoring", "Logs &amp; events", "Configuration", "Maintenance &amp; backups", and "Tags". The "Connectivity &amp; security" tab is selected, showing the Endpoint (customer-features-rds-db-instance.crrwucvfcmbc.us-east-1.rds.amazonaws.com), Networking (Availability Zone us-east-1c, VPC vpc-0041e736a2d14cfa2), and Security (VPC security groups SG-Open-MySQL (sg-05eae22adcf0dce56) Active, default (sg-00757adc39bf9625e) Active).

**79. SO IN THE ETL PROCESS, WE HAVE DONE THE EXTRACT STEP. NOW, WE HAVE TO CONNECT OUR RDS TO THE ETL PROCESS. WE HAVE TO CONNECT THE RDS TO GLUE.**



**80. THE NEXT STEP THAT WE NEED TO DO IS TO CREATE AN ELEMENT FOR OUR RDS DATABASE INSTANCE, THEREFORE, AND NOW TABLE IN THE DATA CATAKOG. SO, WE ARE GOING BACK TO THE ETL, DATABASES, AND THIS TIME WE ARE GOING TO CREATE THE CUSTOMER FEATURES DATABASE.**

**CLICK ADD DATABASE.**

The screenshot shows the AWS Glue Data Catalog interface. The left sidebar shows navigation options like "Getting started", "ETL jobs", "Visual ETL", "Notebooks", "Job run monitoring", "Data Catalog tables", "Data connections", and "Workflows (orchestration)". The main area is titled "Databases (1)" and shows a table with one entry: "movie-ratings-glue-database". The table has columns: Name, Description, Location URI, and Created on (UTC). The entry "movie-ratings-glue-database" has a creation date of "June 14, 2023 at 17:28:43". At the top right of the table, there are "Edit", "Delete", and "Add database" buttons.

## 81. NAME IT AS CUSTOMERR-FEATURES-GLEUE-DATABASE. THEN CLICK CREATE DATABASE.

The screenshot shows the 'Create a database' dialog box within the AWS Glue interface. On the left, there's a sidebar with navigation links for AWS Glue, Data Catalog, and Data Integration and ETL. The main area is titled 'Create a database' with the sub-instruction 'Create a database in the AWS Glue Data Catalog.' Below this is a 'Database details' section. It has a 'Name' field containing 'customer-features-glue-database', a 'Location - optional' field with a URI placeholder, and a 'Description - optional' field with a text input area. At the bottom right are 'Cancel' and 'Create database' buttons.

## 82. THE SECOND DATABASE IS CREATED. BUT, WE NEED TO CONNECT THE RDS DATABASE THAT WE CREATED AND POPULATE IT TO MYSQL WORKBENCH TO THE GLUE DATABASE.

The screenshot shows the 'Databases' list page within the AWS Glue interface. The sidebar includes links for Getting started, ETL jobs, Data Catalog tables, Data connections, and Workflows (orchestration). Under 'Data Catalog', there are 'Tables', 'Stream schema registries', 'Schemas', and 'Connections'. The main area displays a table with two entries:

Name	Description	Location URI	Created on (UTC)
customer-features-glue-database	-	-	June 14, 2023 at 18:29:06
movie-ratings-glue-database	-	-	June 14, 2023 at 17:28:43

## 83. HERE, WE WILL USE THE CONNECTION INSTEAD OF THE CRAWLER. BECAUSE WE HAVE TO ESTABLISH THIS CONNECTION FROM RDS TO AWS GLUE.

GO TO AWS GLUE. DATA CATALOG, CONNECTIONS. CLICK CREAT CONNECTIONS.

The screenshot shows the 'Connectors' list page within the AWS Glue interface. The sidebar includes links for Getting started, ETL jobs, Data Catalog tables, Data connections, and Workflows (orchestration). Under 'Data Catalog', there are 'Tables', 'Stream schema registries', 'Schemas', 'Connections', 'Crawlers', 'Classifiers', and 'Catalog settings'. The main area has sections for 'Marketplace connectors' (with a 'Go to AWS Marketplace' button) and 'Custom connectors' (with a 'Create custom connector' button). Below these are sections for 'Connectors (0)' and 'Connections (0)', each with a 'Filter connections by property' search bar and a table header for Name, Type, and Last modified.

**84. ENTER A NAME FOR THE CONNECTION. WE CALL IT RDS CONNECTION.  
THEN, CONNECTION TYPE IS AMAZON RDS.  
IN DATABASE ENGINE, CHOOSE MYSQL.**

The screenshot shows the 'Create connection' page in the AWS Glue console. On the left, a sidebar lists various AWS Glue services: Getting started, ETL jobs, Visual ETL, Notebooks, Job run monitoring, Data Catalog tables, Data connections (highlighted), Workflows (orchestration), Data Catalog, Databases, Tables, Stream schema registries, Schemas, Connections, Crawlers, Classifiers, Catalog settings, Data Integration and ETL, ETL jobs, Visual ETL, Notebooks, Job run monitoring, Interactive Sessions, and Data classification tools. The main area is titled 'Create connection' and contains two sections: 'Connection properties' and 'Connection access'. In 'Connection properties', the 'Name' field is set to 'RDSConnection', 'Connection type' is set to 'Amazon RDS', and 'Database engine' is set to 'MySQL'. There is also a checkbox for 'Require SSL connection' which is unchecked. In 'Connection access', the 'Database instances' dropdown is set to 'customer-features-rds-db-instance', and the 'Database name' field is set to 'customerfeaturesdatabase'. Under 'Credential type', 'Username and password' is selected. The 'Username' field is filled with 'admin' and the 'Password' field contains a masked value. At the bottom right are 'Cancel' and 'Create connection' buttons.

**85. IN CONNECTION ACCESS, DATABASE INSTANCES, REFRESH, THEN, SELECT YOUR RDS  
DATABASE INSTANCE.  
IN THE DATABASE NAME, COPY THE DATABASE NAME IN THE MYSQL.  
CONFIDENTIAL TYPE, USE THE USERNAME THAT WAS SET EARLIER WHICH IS THE ADMIN.  
INPUT THE PASSWORD. THEN CLICK CREATE CONNECTION.**

This screenshot shows the 'Connection access' configuration page. It includes fields for 'Database instances' (set to 'customer-features-rds-db-instance'), 'Database name' (set to 'customerfeaturesdatabase'), 'Credential type' (set to 'Username and password'), 'Username' (set to 'admin'), and 'Password' (a masked input field). At the bottom right are 'Cancel' and 'Create connection' buttons.

## 86. RDS CONNECTION IS NOW CREATED

The screenshot shows the AWS Glue Connectors page. On the left, there is a sidebar with various navigation options. In the center, under the 'Connectors' section, there are two main categories: 'Marketplace connectors' and 'Custom connectors'. Below these is a table titled 'Connectors (0)' with a 'Actions' dropdown. At the bottom of this section is a table titled 'Connections (1)' with a 'Actions' dropdown, showing one entry: 'RDSConnection' (Type: JDBC, Last modified: Jun 14, 2023). A success message at the top of the page states: "RDSConnection connector successfully added. To begin using your connection you must create a job."

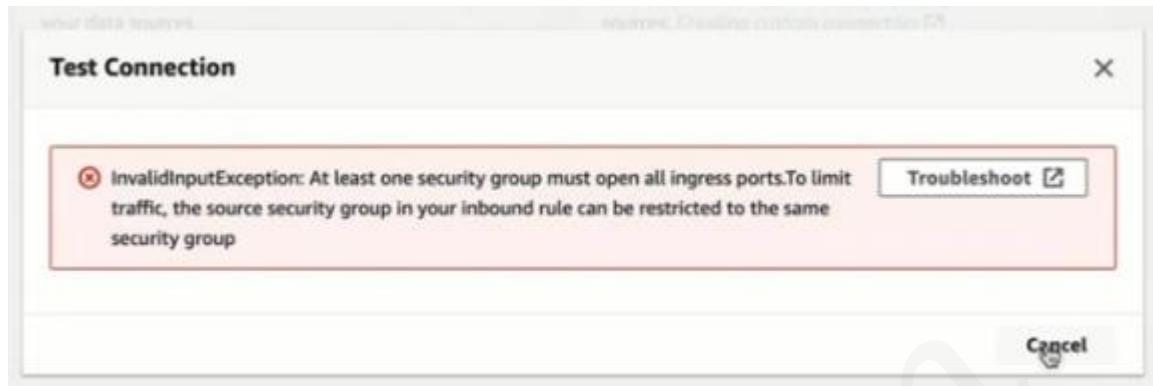
## 87. TEST THE CONNECTION. SELECT THE CONNECTION. CLICK ACTIONS. CLICK TEST CONNECTION.

This screenshot shows the 'Connections (1)' table from the previous step. The table has columns for Name, Type, and Last modified. The single entry is 'RDSConnection' (Type: JDBC, Last modified: Jun 14, 2023). The 'Actions' dropdown is open, and the 'Test connection' option is highlighted with a mouse cursor.

## 88. SELECT THE IAM ROLE – GLUEFULLACCESSROLE. SELECT CONFIRM.

This screenshot shows the 'Test Connection' dialog. It has a search bar for 'Choose an option' where 'GlueFullAccessRole' is selected. Below the search bar, a tooltip explains: 'Allows Glue to call AWS services on your behalf.' At the bottom right of the dialog are 'Cancel' and 'Confirm' buttons.

89. HERE, WE WILL GET THE FAILED TEST CONNECTION. TO SOLVE THE ISSUE, WE HAVE TO FOLLOW WHAT IS SUGGESTED. BASED ON THE TROUBLESHOOT RECOMMENDATION, AT LEAST ONE SECURITY GROUP MUST OPEN ALL INGRESS PORTS AND TO LIMIT TRAFFIC, THE SOURCE SECURITY GROUP IN YOUR INBOUND RULE CAN BE RESTRICTED TO THE SAME SECURITY GROUP. SO, BASICALLY, IT IS REFERRING TO OUR RDS DATABASE INSTANCE,



90. GO BACK TO THE RDS DATABASE INSTANCE. IN THE SECURITY FIELD, YOU WILL SEE THOSE TWO (2) VPC SECURITY GROUPS, THE ONE THAT WE`VE CREATED, SG-OPEN-MYSQL, AND THE DEFAULT ONE. BASED ON THE PREVIOUS TROUBLESHOOT SUGGESTIONS, WE NEED TO OPEN AT LEAST ONE SECURITY GROUP OF ALL INGRESS PORTS. SO, INSTEAD OF OPENING ONLY THE 3306 PORT, WE ARE GOING TO OPEN ALL PORTS IN ONE OF THE TWO SECURITY GROUPS.

A screenshot of the Amazon RDS 'customer-features-rds-db-instance' configuration page. The left sidebar shows navigation options like Dashboard, Databases, and Connectivity &amp; security. The main summary table shows the instance identifier as 'customer-features-rds-db-instance', CPU usage at 3.52%, status as Available, class as db.t3.micro, role as Instance, current activity as 0 connections, engine as MySQL Community, and region as us-east-1c. The 'Connectivity &amp; security' tab is selected, showing the endpoint as 'customer-features-rds-db-instance.crrwucvfcmbc.us-east-1.rds.amazonaws.com' on port 3306, and the networking and security groups. The security group 'SG-Open-MySQL (sg-05eae22adcf0dce56)' is listed as active.

**91. GO BACK TO THE EC2 INSTANCE, SECURITY GROUPS. SELECT SG-OPEN-MYSQL. GO TO INBOUND RULES. THERE, WE ARE GOING TO ADD THE NEW RULE WHICH OPENS ALL PORTS.**

The screenshot shows the AWS EC2 Security Groups interface. On the left, there's a navigation sidebar with options like EC2 Dashboard, EC2 Global View, Events, Limits, Instances, Images, Elastic Block Store, and so on. The main area displays a table of security groups. One row is selected: "sg-05eae22adcf0dce56 - SG-Open-MySQL". This row has a checkmark next to it. The table columns include Name, Security group ID, Security group name, VPC ID, Description, and Owner. Below the table, a tab bar shows "Inbound rules" is selected. A message box says "You can now check network connectivity with Reachability Analyzer" with a "Run Reachability Analyzer" button.

**92. IN THE INBOUND RULES, CLICK EDIT INBOUND RULES. THEN ADD RULE.**

The screenshot shows the "Edit inbound rules" dialog for the SG-Open-MySQL security group. At the top, there's a message box with "You can now check network connectivity with Reachability Analyzer" and a "Run Reachability Analyzer" button. The main area shows the "Inbound rules (1/1)" section with one existing rule: "sgr-000a62b4b15cc1b...". This rule is for MySQL/Aurora, TCP port 3306, from a custom source. Below this, the breadcrumb navigation shows "EC2 > Security Groups > sg-05eae22adcf0dce56 - SG-Open-MySQL > Edit inbound rules". The "Edit inbound rules" form has fields for "Security group rule ID" (set to "sgr-000a62b4b15cc1b15"), "Type" (set to "MySQL/Aurora"), "Protocol" (set to "TCP"), "Port range" (set to "3306"), "Source" (set to "Custom"), and "Description - optional" (empty). There's also a "Delete" button and an "Add rule" button. At the bottom, there are "Cancel", "Preview changes", and "Save rules" buttons.

**93. IN THE ADD RULE, CLICK ALL TCP IN THE TYPE TO OPEN ALL PORTS FROM 0 – 65535. IN THE SOURCE, SELECT ANYWHERE-IPV4. CLICK SAVE RULES.**

The screenshot shows the 'Edit inbound rules' section of the AWS EC2 Security Groups interface. It displays two rules:

- Security group rule ID:** sgr-000a62b4b15cc1b15  
**Type:** MySQL/Aurora  
**Protocol:** TCP  
**Port range:** 3306  
**Source:** Custom (0.0.0.0/0)
- Security group rule ID:** -  
**Type:** All TCP  
**Protocol:** TCP  
**Port range:** 0 - 65535  
**Source:** Anywh... (0.0.0.0/0)

At the bottom right, there are 'Cancel', 'Preview changes', and 'Save rules' buttons. The 'Save rules' button is highlighted in orange.

**94. WE ARE NOW GOING TO SEE IF IT FIXES THE ISSUE. HERE, THE SECURITY GROUP IS MODIFIED.**

The screenshot shows the main EC2 Security Groups page. A success message at the top states: "Inbound security group rules successfully modified on security group (sg-05eae22adcf0dce56 | SG-Open-MySQL)".

The main table lists security groups:

Name	Security group ID	Security group name	VPC ID	Description	Owner
sg-0ce2e3b4e3be04e92	SG-Open-HTTP	vpc-0041e736a2d14cfa2	Allows HTTP Traffic	7496011	
<input checked="" type="checkbox"/> sg-05eae22adcf0dce56	SG-Open-MySQL	vpc-0041e736a2d14cfa2	Allows MySQL Access t...	7496011	
sg-04c7ce4cb6a0e059f	launch-wizard-1	vpc-0041e736a2d14cfa2	launch-wizard-1-create...	7496011	
sg-00757adc39bf9625e	default	vpc-0041e736a2d14cfa2	default VPC security gr...	7496011	

The selected security group is "sg-05eae22adcf0dce56 - SG-Open-MySQL". The "Inbound rules" tab is active, showing two rules:

- Details: You can now check network connectivity with Reachability Analyzer. Run Reachability Analyzer.
- Inbound rules (2):

**95. GO BACK TO CONNECTORS TO TEST THE CONNECTION AGAIN. IN THE CONNECTIONS, CLICK RDSCONNECTION, CLICK ACTIONS, CHOOSE TEST CONNECTION.**

The screenshot shows the AWS Glue Connectors page. On the left, there's a sidebar with various navigation options like Getting started, ETL jobs, Data Catalog tables, Data connections, etc. The main area is titled 'Connectors' and shows two sections: 'Marketplace connectors' and 'Custom connectors'. Under 'Marketplace connectors', there's a button 'Go to AWS Marketplace'. Under 'Custom connectors', there's a button 'Create custom connector'. Below these sections, there's a table for 'Connections' with one entry: 'RDSConnection' (Type: JDBC, Last modified: Jun 14, 2023). There are buttons for 'Actions', 'Create connection', and 'Create job'.

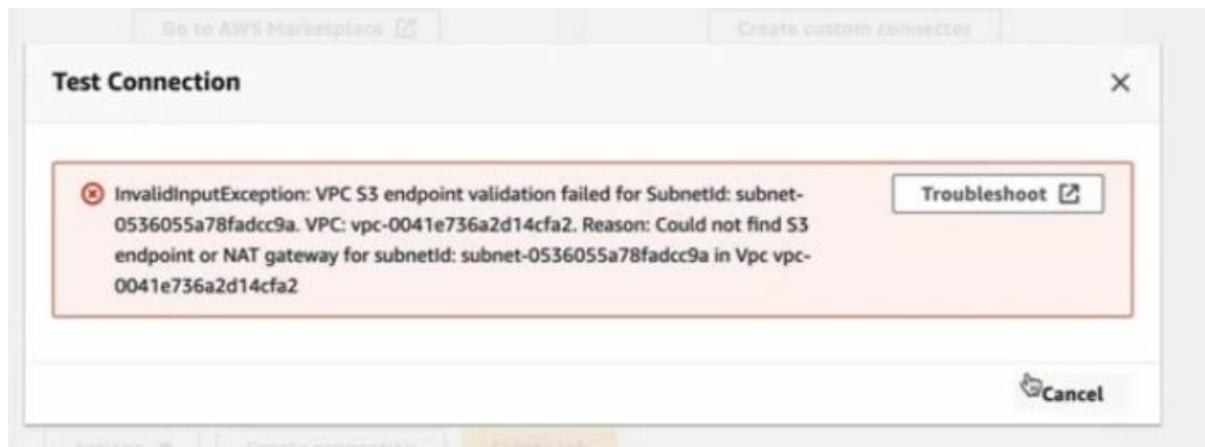
**SELECT TEST CONNECTION.**

This screenshot shows the 'Connections (1)' page. It has a table with columns 'Name', 'Type', and 'Last modified'. The single entry is 'RDSConnection' (Type: JDBC, Last modified: Jun 14, 2023). There are buttons for 'Actions', 'Create connection', and 'Create job' at the top.

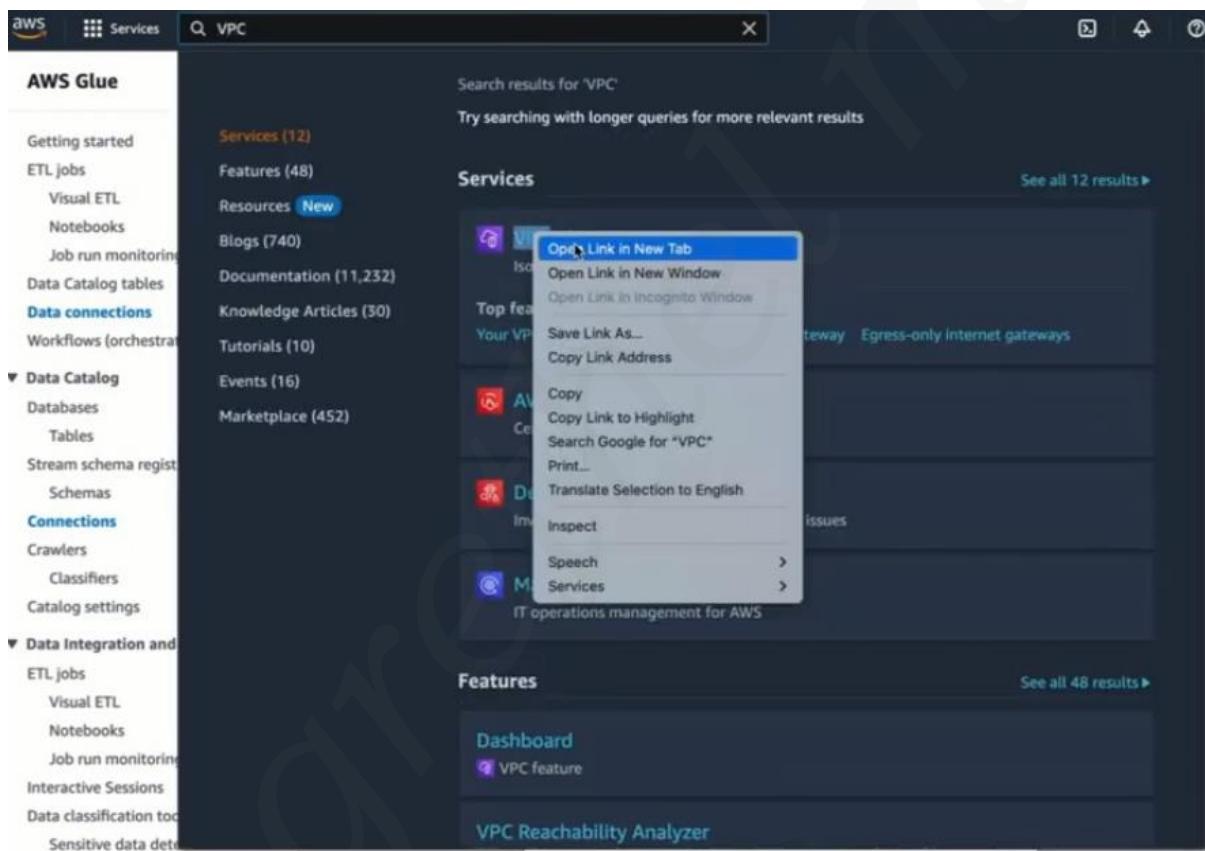
**CHOOSE GLUEFULLACCESSROLE. THEN, CONFIRM.**

This screenshot shows the 'Test Connection' dialog box. It asks to choose an IAM role and lists 'GlueFullAccessRole' as an option. This role is described as 'Allows Glue to call AWS services on your behalf.' At the bottom, there are 'Cancel' and 'Confirm' buttons.

UNFORTUNATELY, WE HAVE ANOTHER MESSAGE ERROR. BASED ON THE TROUBLESHOOT MESSAGE, WE NEED TO CREATE A VPC S3 ENDPOINT.



96. GO BACK TO THE SEARCH BAR, TYPE VPC. THEN, RIGHT CLICK, OPEN LINK IN NEW TAB.



**97. IN THE VCPC, GO TO THE ENDPOINTS. IT IS NORMAL NOT TO SEE ANY ENDPOINTS. BUT AS A SOLUTION TO THE CONNECTION, WE ARE GOING TO CREATE AN ENDPOINT.**

The screenshot shows the AWS VPC Endpoints page. The left sidebar has sections for Virtual private cloud (Your VPCs, Subnets, Route tables, Internet gateways, Egress-only internet gateways, Carrier gateways, DHCP option sets, Elastic IPs, Managed prefix lists), Endpoints (Endpoint services, NAT gateways, Peering connections), and Security (Network ACLs, Security groups). The main content area is titled 'Endpoints' and shows a search bar and a table with columns: Name, VPC endpoint ID, VPC ID, and Service name. A message at the bottom says 'No endpoint found'.

**98. CLICK CREATE ENDPOINTS**

The screenshot shows the same VPC Endpoints page as above, but the 'Create endpoint' button in the top right corner is highlighted in orange.

**99. THE ENDPOINT SHOULD BE THE VPC ENDPOINT TO THE S3. SO, WE ARE GOING TO CALL IT VPC ENDPOINT TO S3.**

The screenshot shows the 'Create endpoint' configuration page. At the top, it says 'VPC > Endpoints > Create endpoint'. The main section is titled 'Create endpoint' with a 'Info' link. It says 'There are three types of VPC endpoints – Interface endpoints, Gateway Load Balancer endpoints, and Gateway endpoints. Interface endpoints and Gateway Load Balancer endpoints are powered by AWS PrivateLink, and use an Elastic Network Interface (ENI) as an entry point for traffic destined to the service. Interface endpoints are typically accessed using the public or private DNS name associated with the service, while Gateway endpoints and Gateway Load Balancer endpoints serve as a target for a route in your route table for traffic destined for the service.' Below this is a 'Endpoint settings' section with a 'Name tag - optional' field containing 'VPC Endpoint to S3'. There are several service category options: 'AWS services' (selected), 'PrivateLink Ready partner services', 'AWS Marketplace services', 'EC2 Instance Connect Endpoint', and 'Other endpoint services'.

**100. IN THE SERVICE CATERGORY, SELECT AWS SERVICES.**

## Endpoint settings

Name tag - *optional*  
Creates a tag with a key of 'Name' and a value that you specify.

VPC Endpoint to S3

**Service category**  
Select the service category

- AWS services**  
Services provided by Amazon
- PrivateLink Ready partner services**  
Services with an AWS Service Ready designation
- AWS Marketplace services**  
Services that you've purchased through AWS Marketplace
- EC2 Instance Connect Endpoint**  
An elastic network interface that allow you to connect to resources in a private subnet
- Other endpoint services**  
Find services shared with you by service name

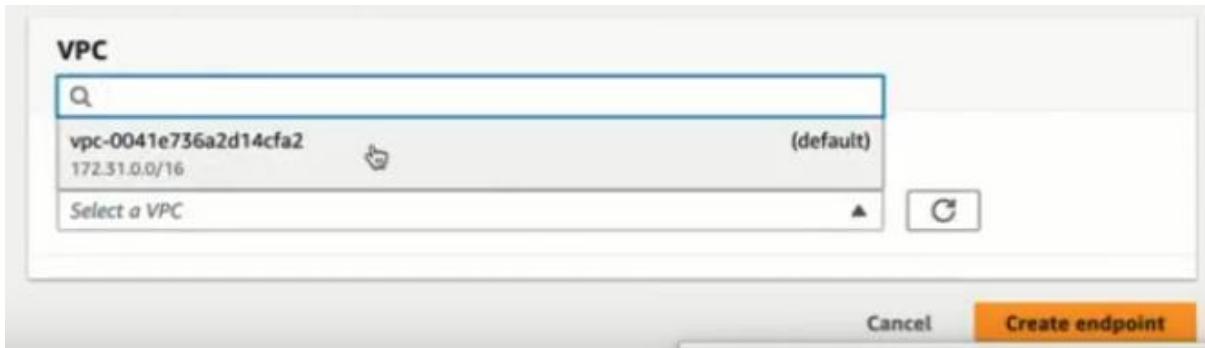
**101. IN THE SERVICES, SELECT SERVICE NAME = COM.AMAZONAWS.US-EAST-1.S3**

Services (225)			
<input type="text" value="S3"/> <span>X</span>		<	1 2 3 4 5 6 7 ... 23 > <span>①</span>
Use: "S3"		Owner	Type
Client filters values		amazon	Interface
Service Name = com.amazonaws.s3-global.accesspoint		amazon	Interface
Service Name = com.amazonaws.us-east-1.s3	...	amazon	Interface
Service Name = com.amazonaws.us-east-1.s3-outposts	...	amazon	Interface
com.amazonaws.us-east-1.acm-pca		amazon	Interface
com.amazonaws.us-east-1.airflow.api		amazon	Interface

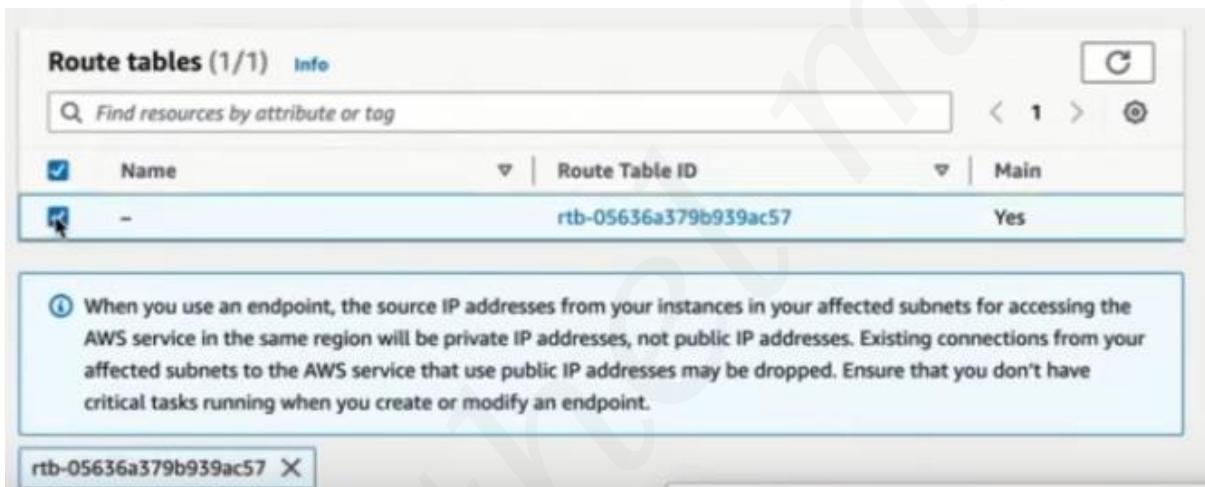
**102. SELECT GATEWAY**

Services (1/2)			
<input type="text"/> Find resources by attribute or tag			
Service Name = com.amazonaws.us-east-1.s3 <span style="border: 1px solid #ccc; padding: 2px;">X</span>		<a href="#">Clear filters</a>	
Service Name	Owner		Type
 com.amazonaws.us-east-1.s3	amazon		Gateway
 com.amazonaws.us-east-1.s3	amazon		Interface

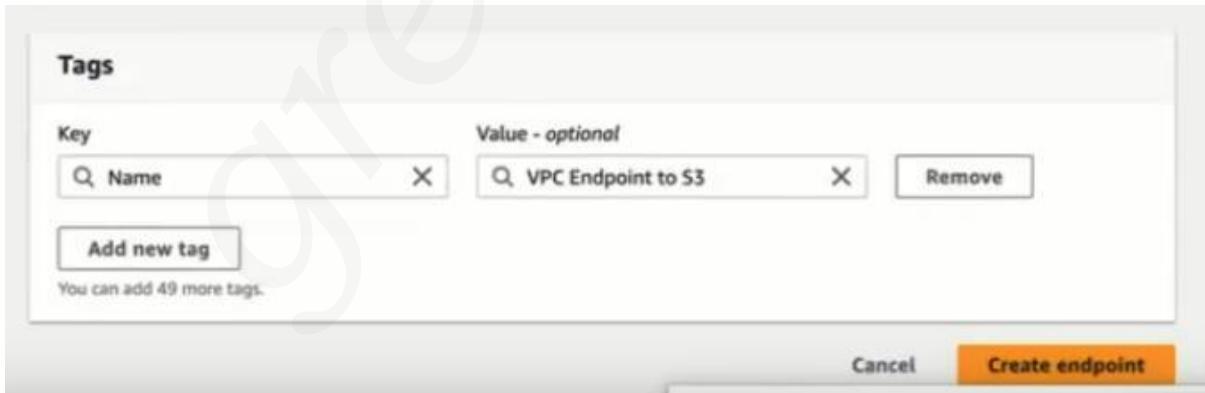
103. SELECT THE VPC. REMEMBER TO CONNECT IT TO THE SUBNET.



104. WE WILL CONNECT THE SUBNET THROUGH THE ROUTE TABLES BECAUSE THAT IS WHAT CONTAINS THE CONNECTION TO THE SUBNETS.



105. CLICK CREATE ENDPOINT.



**106. GO CHECK THE ENDPOINTS. HERE, WE SHOULD SEE THE ONE THAT WE HAVE CREATED.**

The screenshot shows the AWS VPC dashboard. A green success message at the top says "Successfully created VPC endpoint vpce-0dfd95341ef6181af". Below it, a table titled "Endpoints (1/1)" lists one entry: "VPC Endpoint to S3" with "vpce-0dfd95341ef6181af" as the VPC endpoint ID, "vpc-0041e736a2d14cfa2" as the VPC ID, and "com.amazonaws.us-east-1.s3" as the Service name. The table has columns for Name, VPC endpoint ID, VPC ID, and Service name. At the bottom, there are tabs for Details, Route tables, Policy, and Tags.

**107. CHECK IF IT FIXES THE ISSUE. GO BACK TO THE CONNECTORS**

The screenshot shows the AWS Glue Connectors page. On the left, a sidebar lists various AWS Glue services like Getting started, ETL jobs, Data connections, Data Catalog, Connections, and Data Integration and ETL. The "Connections" section is currently selected. The main area shows two sections: "Marketplace connectors" (with a "Go to AWS Marketplace" button) and "Custom connectors" (with a "Create custom connector" button). Below these is a table titled "Connectors (0)" with an info link. It has a search bar and columns for Name, Type, and Last modified. At the bottom, there's a table titled "Connections (1)" with an info link. It also has a search bar and columns for Name, Type, and Last modified. The single connection listed is "RDSConnection" of type "JDBC" last modified on "Jun 14, 2023".

**108. CLICK RDSCONNECTION, ACTION THE TEST CONNECTION.**

The screenshot shows a table with three columns: Name, Type, and Last modified. The first row contains 'RDSConnection', 'JDBC', and 'Jun 14, 2023'. The 'Actions' button is highlighted with a blue border.

Name	Type	Last modified
RDSConnection	JDBC	Jun 14, 2023

The screenshot shows a table with three columns: View details, Type, and Last modified. The first row contains 'by property', 'JDBC', and 'Jun 14, 2023'. The 'Actions' button is highlighted with a blue border. A sidebar on the left lists 'View details', 'Delete', 'Edit', and 'Test connection', with 'Test connection' being the active item.

View details	Type	Last modified
by property	JDBC	Jun 14, 2023

**109. SELECT GLUEFULLACCESS, THEN, CONFIRM.**

The screenshot shows a 'Test Connection' dialog box. It has a dropdown menu for 'IAM role' containing 'GlueFullAccessRole'. Below the dropdown is a note: 'Ensure that this role has permission to access your data store.' and a link 'Create IAM role.' At the bottom are 'Cancel' and 'Confirm' buttons.

**110. IT SHOULD BE SUCCESSFULLY CONNECTED BY NOW. CONGRATULATIONS YOU DID THE HARD PART! WE CAN NOW PROCEED TO THE NEXT PROCEDURE WHICH IS TO HANDLE THE REDSHIFT.**

The screenshot shows a 'Test Connection' dialog box with a green success message: 'Successfully connected to the data store with connection RDSConnection.' and a 'View log' button. At the bottom are 'Cancel' and 'Confirm' buttons.

**111. IN THE AWS GLUE, SELECT DATA CATALOG, THEN, DATABASES. IN THE DATABASES, SELECT CUSTOMER-FEATURES-GLUE-DATABASE.**

The screenshot shows the AWS Glue interface with the 'Data Catalog' section selected. Under 'Databases', there are two entries:

Name	Description	Location URI	Created on (UTC)
customer-features-glue-database	-	-	June 14, 2023 at 18:29:06
movie-ratings-glue-database	-	-	June 14, 2023 at 17:28:43

**112. IN THE CUSTOMER-FEATURES-GLUE-DATABASE, CLICK, ADD TABLES USING CRAWLER.**

The screenshot shows the 'customer-features-glue-database' properties page. Under 'Tables (0)', it says 'No available tables'. There is a button labeled 'Add tables using crawler'.

**113. ENTER UNIQUE CRAWLER NAME. CLICK, NEXT.**

The screenshot shows the 'Set crawler properties' step of the 'Add crawler' wizard. In the 'Crawler details' section, the 'Name' field is filled with 'customer-features-crawler'. The 'Description - optional' field is empty. At the bottom, there is a 'Tags - optional' section with a note: 'Use tags to organize and identify your resources.' A 'Next' button is visible at the bottom right.

**114. CLICK NOT YET IN THE “IS YOUR DATA ALREADY MAPPED TO GLUE TABLES?”. THEN, ADD A DATA SOURCE.**

The screenshot shows the AWS Glue 'Add crawler' wizard. The current step is 'Step 2: Choose data sources and classifiers'. On the left, there's a sidebar with steps: Step 1 (Set crawler properties), Step 2 (Choose data sources and classifiers, which is active), Step 3 (Configure security settings), Step 4 (Set output and scheduling), and Step 5 (Review and create). The main area is titled 'Choose data sources and classifiers' and contains a 'Data source configuration' section. It asks 'Is your data already mapped to Glue tables?' with two options: 'Not yet' (selected) and 'Yes'. Below this is a 'Data sources (0)' section with a table header 'Type | Data source | Parameters'. A message says 'You don't have any data sources.' with a 'Add a data source' button. At the bottom right are 'Cancel', 'Previous', and 'Next' buttons.

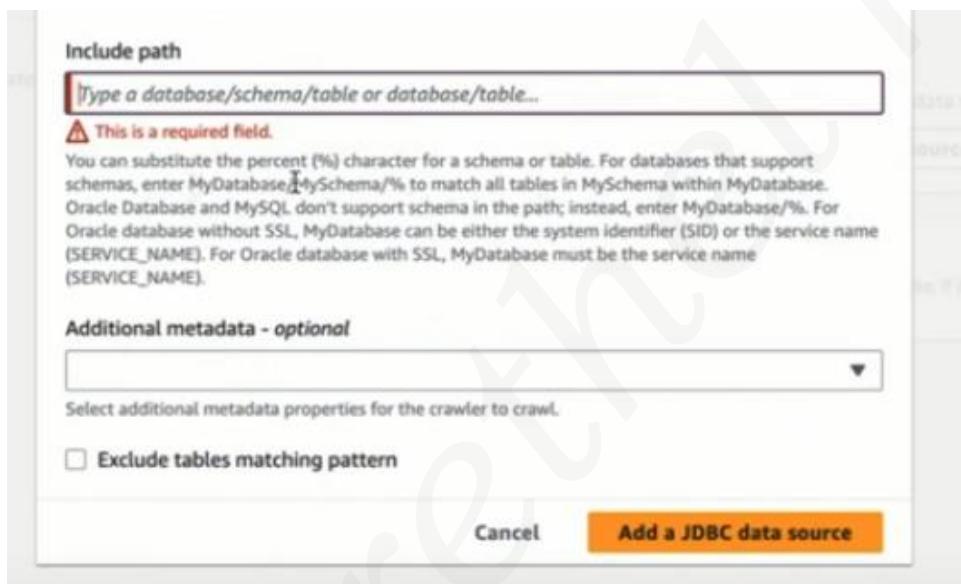
**115. IN THE DATA SOURCE, SELECT JDBC WHICH MEANS JAVA DATABASE CONNECTION.**

The screenshot shows the 'Add data source' dialog box. The 'Data source' section asks 'Choose the source of data to be crawled.' It lists several options: 'S3' (selected), 'JDBC' (with a note 'JDBC stream as the data source'), 'DynamoDB' (with a note 'Dynamo DB as the data source'), 'DocumentDB/MongoDB' (with a note 'Mongo DB as the data source'), 'Delta Lake' (with a note 'Delta Lake as the data source'), and 'In a different account' (unchecked). Below this is an 'S3 path' section with a text input field containing 's3://bucket/prefix/object', a 'View' button, and a 'Browse S3' button. A note at the bottom says 'All folders and files contained in the S3 path are crawled. For example, type s3://MyBucket/MyFolder/ to crawl all objects in MyFolder within MyBucket.'

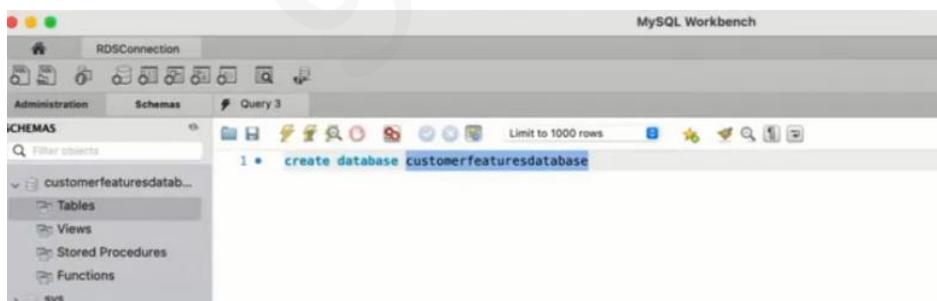
**116. IN THE CONNECTION, SELECT RDSCONNECTION THAT WE HAVE JUST CREATED AND TESTED SUCCESSFULLY.**



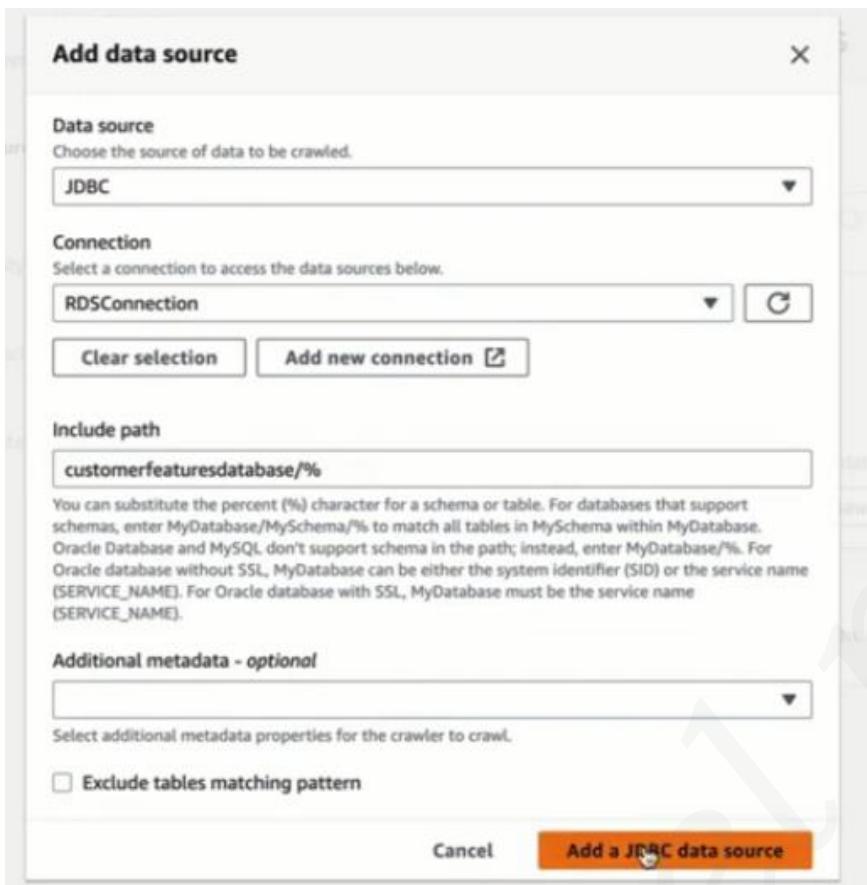
**117. IN THE INCLUDE PATH, WE HAVE TO MAKE SURE THAT WE INCLUDE THE RIGHT PATH. IT IS GOOD TO READ AND TAKE INTO CONSIDERATION WHAT IS WRITTEN BELOW.**



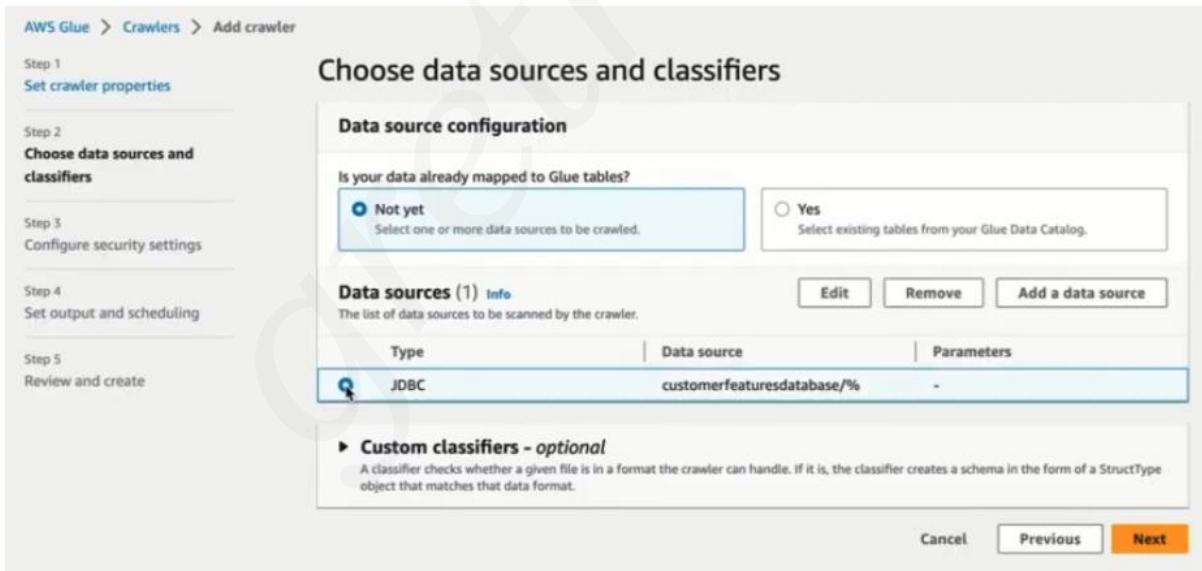
**118. GO BACK TO THE MYSQL WORKBENCH TO DOUBLE CHECK THE DATABASE NAME THAT WE HAVE JUST CREATED PREVIOUSLY.**



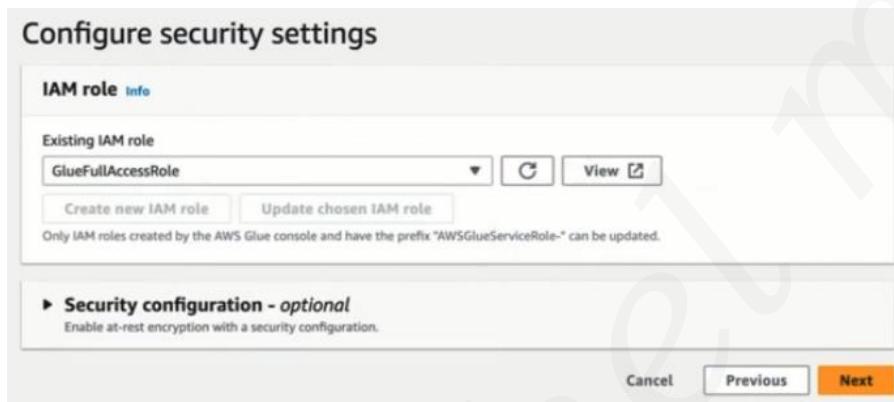
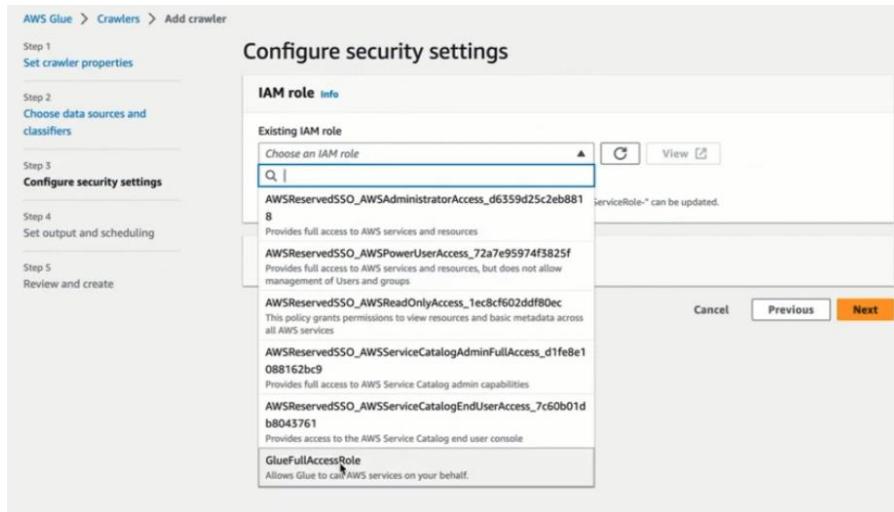
**119. FOLLOW THE DATABASE NAME IN THE MYSQL AND THE INSTRUCTION GIVEN IN PUTTING THE NAME IN THE ‘INCLUDE PATH’. FINALLY, CLICK ADD A JDBC DATA SOURCE.**



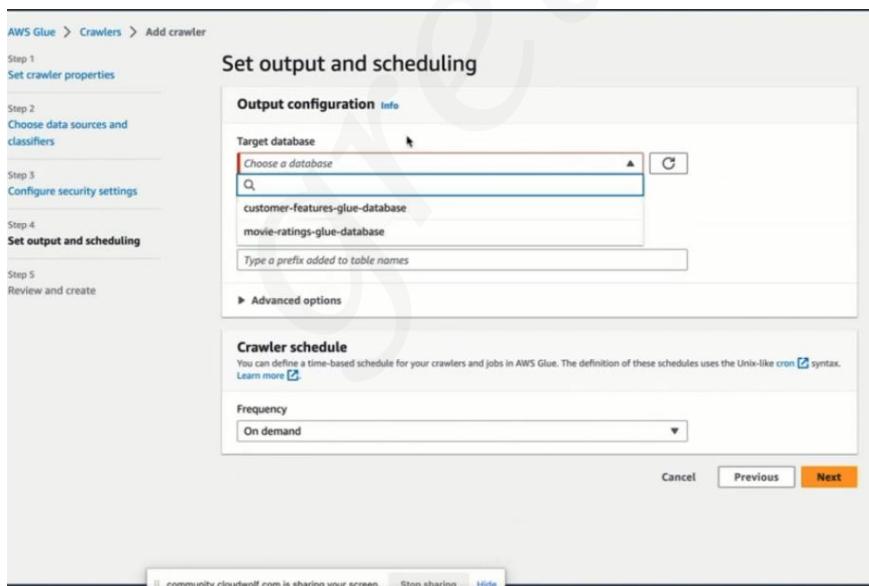
**120. SELECT JDBC. THEN, CLICK NEXT.**



**121. IN THE CONFIGURE SECURITY SETTINGS, EXISTING IAM ROLE, SELECT GLUEFULLACCESSROLE. THEN, CLICK NEXT.**



**122. IN THE TARGET DATABASE, SELECT CUSTOMER-FEATURES-GLUE-DATABASE. KEEP THE ON-DEMAND FREQUENCY OF THE CRAWLER SCHEDULE. IT MEANS THAT IT IS GOING TO CRAWL WHEN WE CLICK THE RUN BUTTON. THEN, CLICK NEXT.**



**Target database**

▼
C

**Clear selection**
**Add database**

**Table name prefix - optional**

*Type a prefix added to table names*

### 123. REVIEW, THEN, CLICK CREATE CRAWLER.

AWS Glue > Crawlers > Add crawler

Step 1 Set crawler properties

Step 2 Choose data sources and classifiers

Step 3 Configure security settings

Step 4 Set output and scheduling

Step 5 Review and create

**Review and create**

**Step 1: Set crawler properties**

<b>Set crawler properties</b>		
Name	Description	Tags
customer-features-crawler	-	-

**Step 2: Choose data sources and classifiers**

<b>Data sources (1) <small>Info</small></b>		
The list of data sources to be scanned by the crawler.		
Type	Data source	Parameters
JDBC	customerfeaturesdatabase/%	-

**Step 3: Configure security settings**

<b>Configure security settings</b>		
IAM role	Security configuration	Lake Formation configuration
GlueFullAccessRole	-	-

**Step 4: Set output and scheduling**

<b>Set output and scheduling</b>		
Database	Table prefix - optional	Schedule
customer-features-glue-database	-	On demand

Cancel
Previous
Create crawler

**124. CRAWLER SUCCESSFULLY CREATED, BUT, WE NEED TO RUN IT. ON THE UPPER RIGHT SIDE, CLICK RUN CRAWLER.**

The screenshot shows the AWS Glue interface. On the left, there's a sidebar with various options like 'Getting started', 'ETL jobs', 'Data Catalog tables', 'Data connections', 'Workflows (orchestration)', 'Data Catalog', 'Databases', 'Tables', 'Stream schema registries', 'Schemas', 'Connections', 'Crawlers', 'Classifiers', 'Catalog settings', 'Data Integration and ETL', 'ETL jobs', 'Visual ETL', 'Notebooks', 'Job run monitoring', and 'Interactive Sessions'. The main area is titled 'customer-features-crawler' and contains sections for 'Crawler properties' (Name: customer-features-crawler, IAM role: GlueFullAccessRole, Database: customer-features-glue-database, State: READY) and 'Crawler runs' (0). A prominent 'Run crawler' button is located at the top right of the crawler runs section.

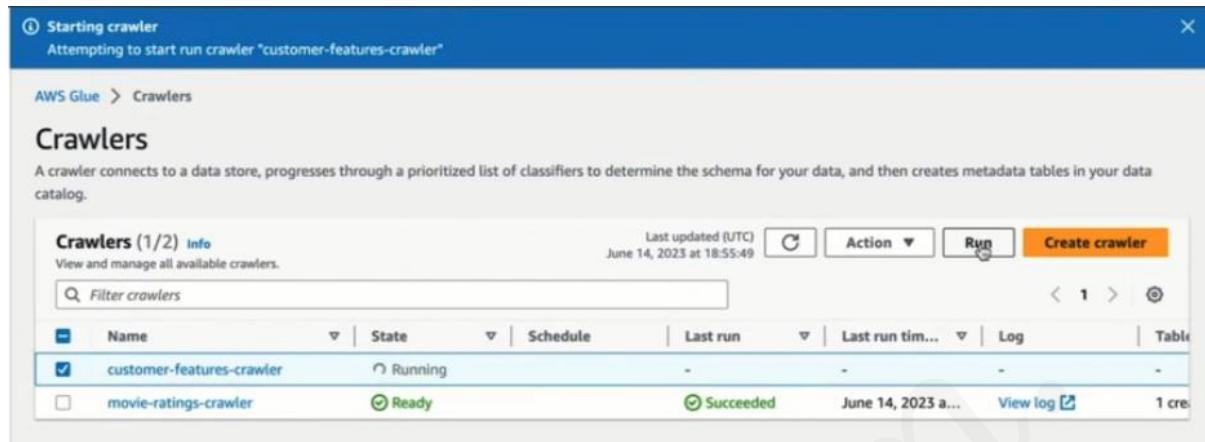
**125. OR YOU CAN CLICK THE CRAWLERS, REFRESH TO SEE THE CRAWLER.**

This screenshot shows the 'Crawlers' page in AWS Glue. The sidebar on the left is identical to the previous screenshot. The main area is titled 'Crawlers' and includes a brief description: 'A crawler connects to a data store, progresses through a prioritized list of classifiers to determine the schema for your data, and then creates metadata tables in your data catalog.' Below this, there's a table titled 'Crawlers (1) Info' with one entry: 'movie-ratings-cr...' (State: Ready, Last run: Succeeded, June 14, 2023). A 'Create crawler' button is located at the top right of the table.

**126. SELECT THE NEW CRAWLER, CUSTOMER-FEATURES-CRAWLER. THEN,CLICK RUN.**

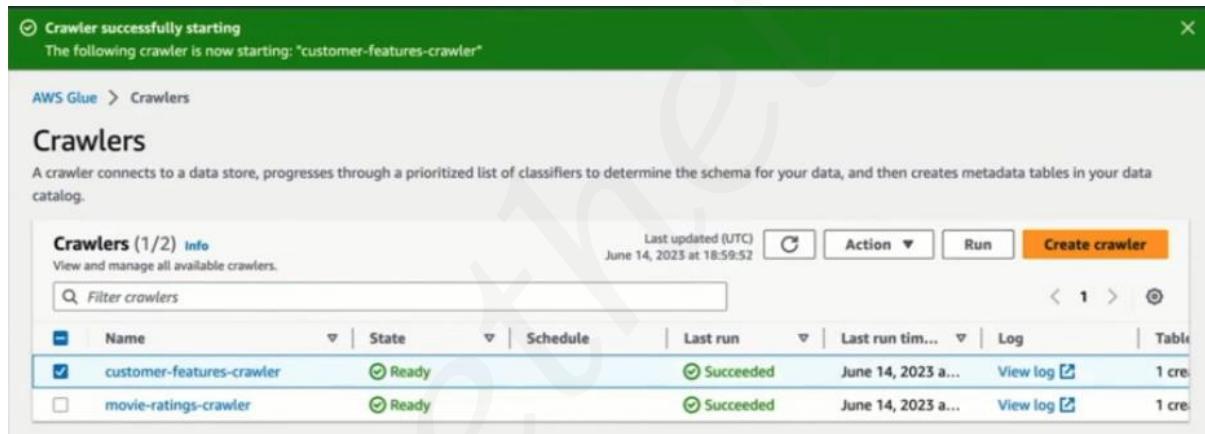
This screenshot shows the 'Crawlers' page again, but with a selection made. The 'customer-features-crawler' row is highlighted with a blue border. The 'Run' button is highlighted in the top right corner of the table header.

**127.** AT THE TOP, YOU WILL SEE THE BLUE FLAG WHICH IS STATING THAT THE CRAWLER IS STARTING. IT IS GOING TO CRAWL THE DATA FROM THE RDS DATABASE TABLE THAT IS IN THE MYSQL WORKBENCH. IT IS NOW RUNNING AND EXTRACTING THE DATA AS IN THE FIRST STEP OF THE ETL PROCESS WHICH MEANS THAT IT IS IN THE EXTRACT STEP.



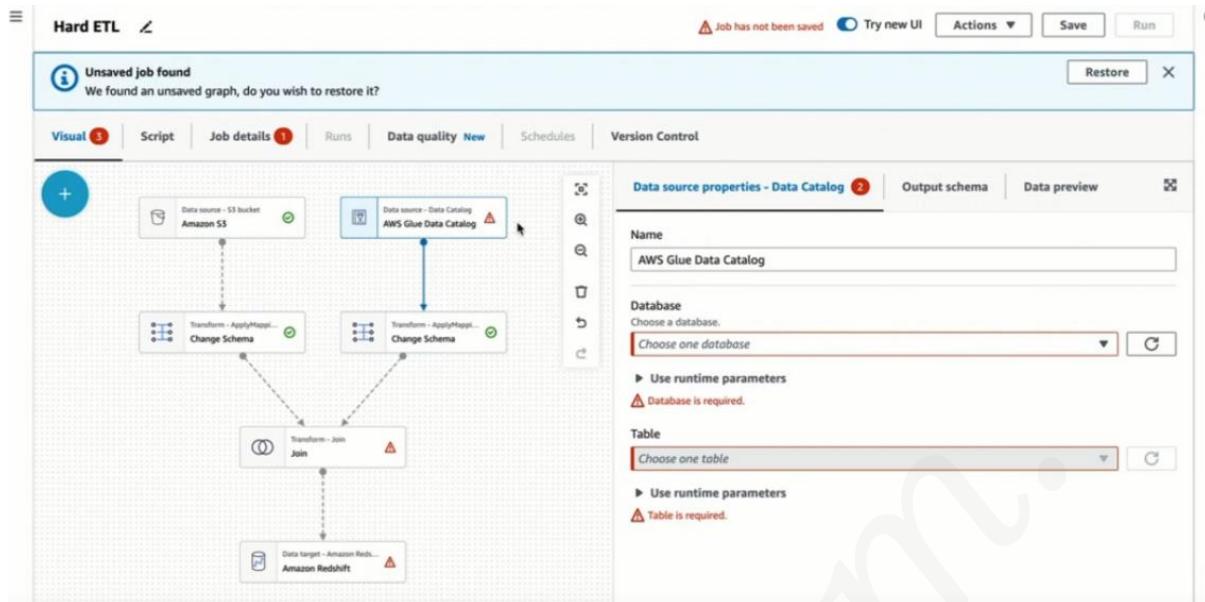
The screenshot shows the AWS Glue Crawlers interface. At the top, a blue header bar displays a status message: "Starting crawler Attempting to start run crawler 'customer-features-crawler'". Below this, the main title "Crawlers" is shown, followed by a brief description: "A crawler connects to a data store, progresses through a prioritized list of classifiers to determine the schema for your data, and then creates metadata tables in your data catalog." A table titled "Crawlers (1/2) Info" lists two crawlers: "customer-features-crawler" (Running) and "movie-ratings-crawler" (Ready). The "customer-features-crawler" row has a green checkmark icon next to it. Action buttons include "Action", "Run", and "Create crawler".

**128.** CRAWLER SUCCESSFULLY STARTING. THE NEXT STEP IS TO GO BACK TO THE VISUAL EDITOR.



The screenshot shows the AWS Glue Crawlers interface. A green header bar at the top displays a success message: "Crawler successfully starting The following crawler is now starting: 'customer-features-crawler'". Below this, the main title "Crawlers" is shown, followed by a brief description: "A crawler connects to a data store, progresses through a prioritized list of classifiers to determine the schema for your data, and then creates metadata tables in your data catalog." A table titled "Crawlers (1/2) Info" lists two crawlers: "customer-features-crawler" (Ready) and "movie-ratings-crawler" (Ready). Both rows have green checkmark icons next to them. Action buttons include "Action", "Run", and "Create crawler".

**129. HERE, WE ARE GOING TO CONNECT THE DATA CATALOG ELEMENT RELATED TO OUR DS DATABASE IN THE ETL PROCESS.**

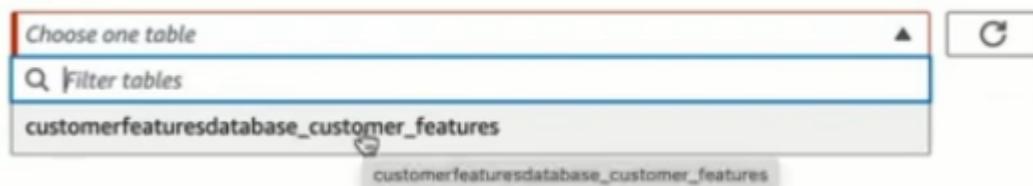


**130. IN THE NAME PORTION, MAKE SURE THAT THE AWS GLUE DATA CATALOG IS SELECTED. IN THE DATABASE, SELECT CUSTOMER-FEATURE-GLUE-DATABASE THAT WE CREATED.**

This screenshot shows the "Data source properties - Data Catalog" configuration screen. The "Name" field is filled with "AWS Glue Data Catalog". In the "Database" section, a dropdown menu is open, showing "Choose one database". Below the dropdown is a search bar with the placeholder "Filter databases". Two databases are listed: "customer-features-glue-database" and "movie-ratings-glue-database". The "customer-features-glue-database" entry is highlighted with a blue background. At the bottom of the screen, there are two status messages: "▶ Use runtime parameters" and "⚠ Table is required.". The left side of the interface features a "Version Control" sidebar.

**131. IN THE TABLE, CHOOSE CUSTOMERFEATURESDATABASE\_CUSTOMER\_FEATURES.**

Table



**132. IT SHOULD THEN LOOK LIKE THIS. THIS IS THE STEP OF THE EXTRACTION OF OUR SECOND DATA SOURCE RELATED TO RDS.**

Version Control

Data source properties - Data Catalog      Output schema      Data preview

Name  
AWS Glue Data Catalog

Database  
Choose a database.  
customer-features-glue-database

▶ Use runtime parameters

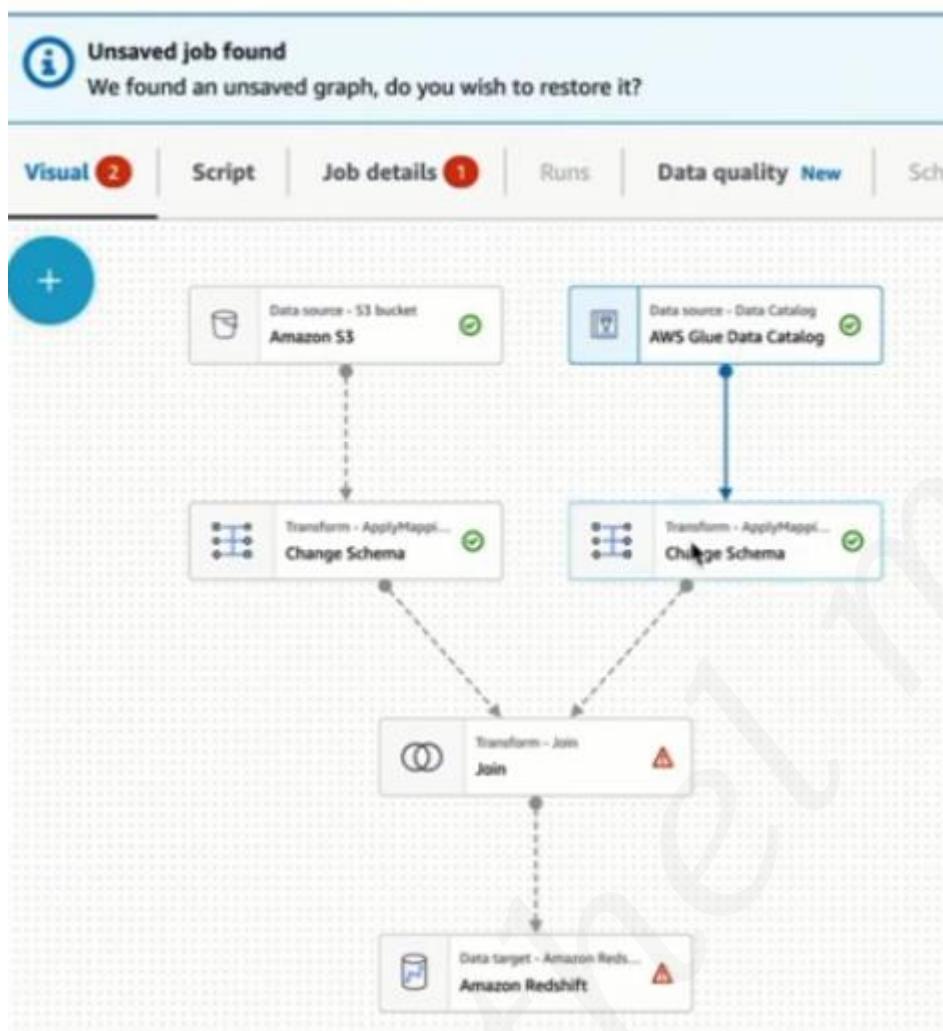
Table  
customerfeaturesdatabase\_customer\_features

▶ Use runtime parameters

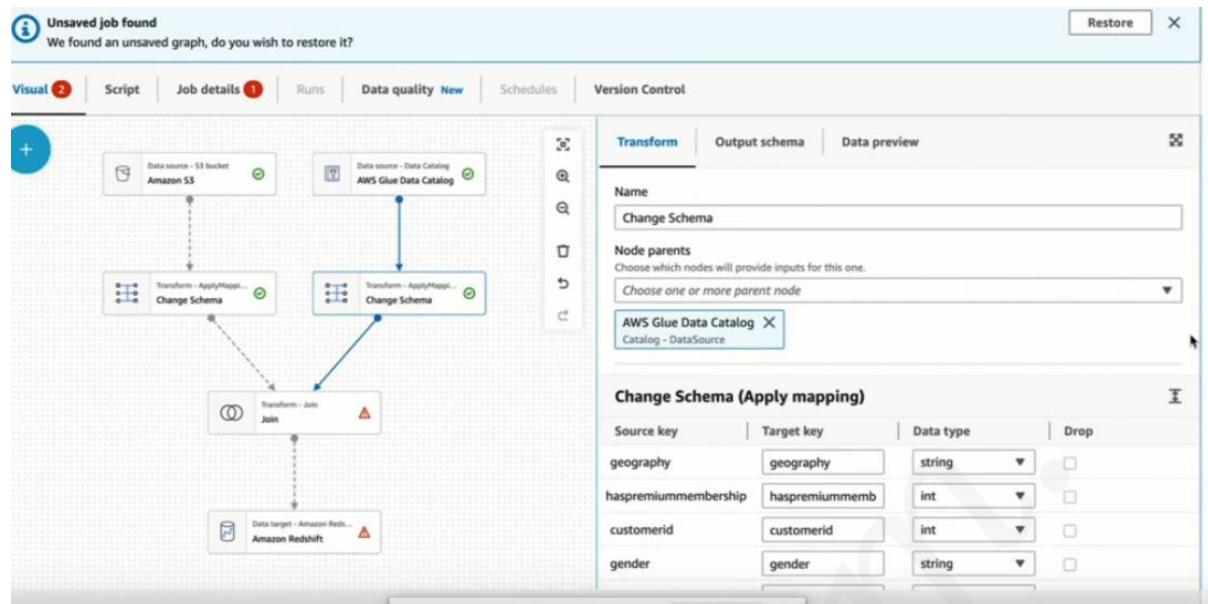
This screenshot shows the "Data source properties - Data Catalog" tab of a AWS Glue Data Catalog configuration interface. It includes sections for Name (set to "AWS Glue Data Catalog"), Database (set to "customer-features-glue-database"), and Table (set to "customerfeaturesdatabase\_customer\_features"). Each section has a "Use runtime parameters" link below it. The "Output schema" and "Data preview" tabs are also visible at the top.

133. WE, THEN, MOVE ON TO THE TRANSFORM CHANGE SCHEMA PHASE.

Hard ETL ↴



LET'S CHECK IF WE ARE GOING TO TRANSFORM A BIT OUR DATA RELATED TO THE CUSTOMER FEATURES.



HERE, WE CAN CHANGE THE TARGET KEY AND DATA TYPES BASED ON THE PREFERENCE. BUT FOR THIS TRAINING PURPOSES, WE WILL PUT IT AS IT IS.

The screenshot shows the "Version Control" section of the AWS Glue Data Catalog interface. It has tabs for "Transform", "Output schema", and "Data preview". The "Transform" tab shows a dropdown for "Choose which nodes will provide inputs for this one" set to "Choose one or more parent node". Under "AWS Glue Data Catalog X Catalog - DataSource", a table titled "Change Schema (Apply mapping)" lists eight columns: Source key, Target key, Data type, and Drop. The data is as follows:

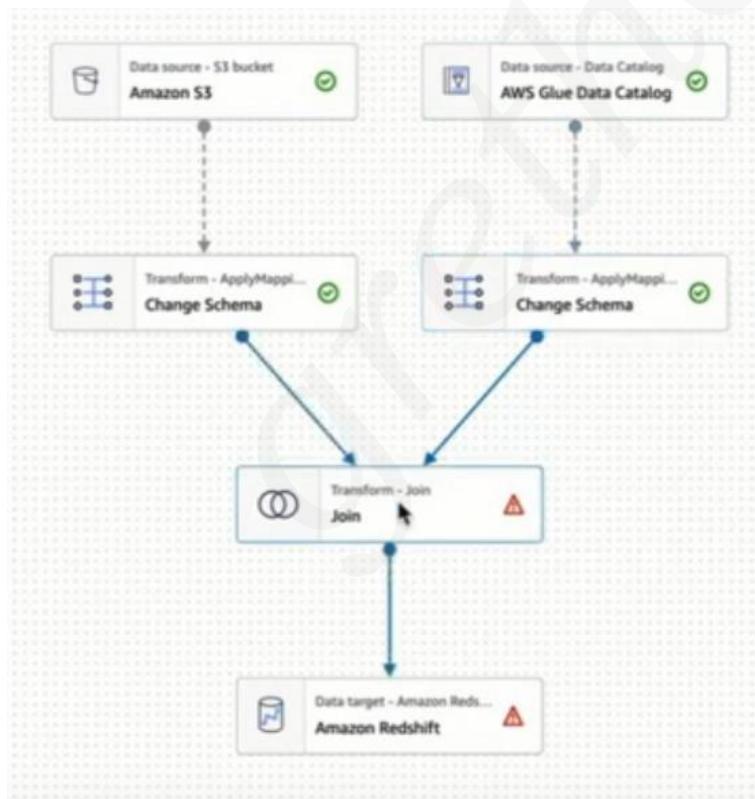
Source key	Target key	Data type	Drop
geography	geography	string	<input type="checkbox"/>
haspremiummembership	haspremiummemb	int	<input type="checkbox"/>
customerid	customerid	int	<input type="checkbox"/>
gender	gender	string	<input type="checkbox"/>
isactivemember	isactivemember	int	<input type="checkbox"/>
surname	surname	string	<input type="checkbox"/>
age	age	int	<input type="checkbox"/>

**134. BUT TO TAKE INTO CONSIDERATION IN ORDER TO BUILD THE RECOMMENDER SYSTEM, WE CAN CHECK IF THERE IS ANY FEATURE OR VARIABLES THAT WE CAN DROP BECAUSE THEY ARE NOT RELEVANT TO MAKING A RECOMMENDER SYSTEM FOR MOVIES. FOR EXAMPLE, THE SURNAME MIGHT NOT BE RELEVANT, SO, WE ARE GOING TO DROP THIS VARIABLE.**

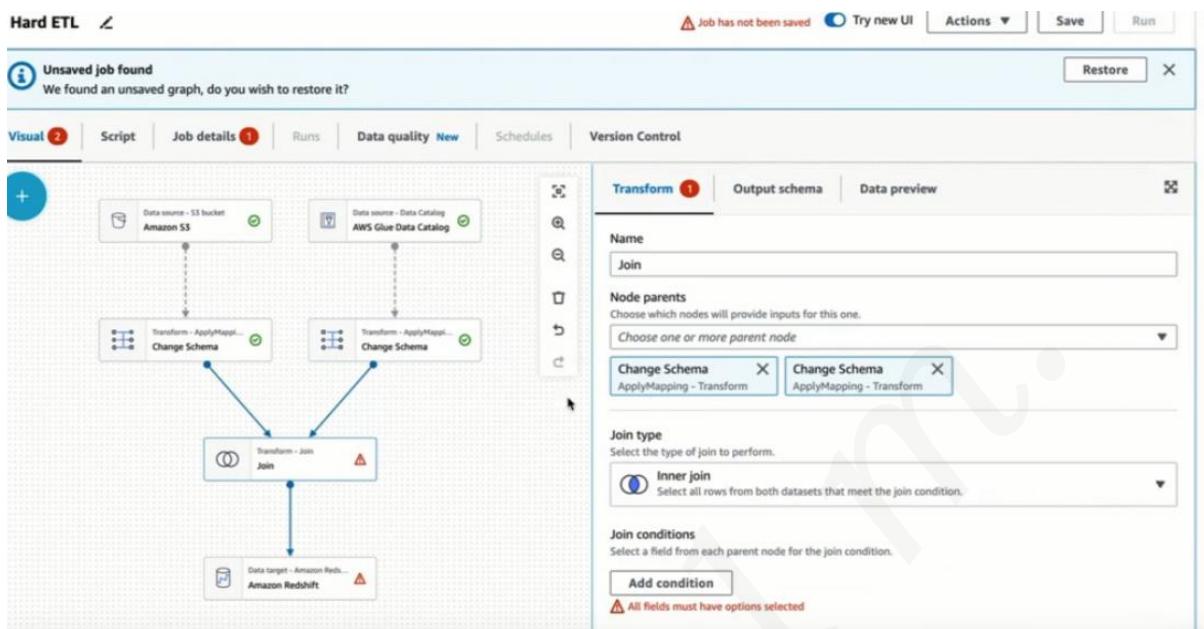
Screenshot of the AWS Glue Data Catalog interface showing the 'Transform' tab selected. Under 'Node parents', it says 'Choose which nodes will provide inputs for this one.' and 'Choose one or more parent node'. A dropdown menu shows 'AWS Glue Data Catalog X Catalog - DataSource'. Below this is a table titled 'Change Schema (Apply mapping)'. The table has columns: Source key, Target key, Data type, and Drop. The rows are:

Source key	Target key	Data type	Drop
geography	geography	string	<input type="checkbox"/>
haspremiummembership	haspremiummemb	int	<input type="checkbox"/>
customerid	customerid	int	<input type="checkbox"/>
gender	gender	string	<input type="checkbox"/>
isactivermember	isactivermember	int	<input type="checkbox"/>
surname			<input checked="" type="checkbox"/>
age	age	int	<input type="checkbox"/>

**135. WE WILL THEN PROCEED TO THE NEXT STEP, JOIN, WHICH IS STILL PART OF THE TRANSFORM STEP IN THE ETL PROCESS.**



**136. HERE, WE ARE GOING TO DO THE INNER JOIN TO JOIN THE RESULTED TABLES OF THE TRANSFORMATION COMING FROM THE S3, THE MOVIE RATINGS, AND THE ONE COMING FROM THE RDS DATABASE TABLE OF THE CUSTOMER FEATURE. WE ARE JOINING THESE TWO TABLE AND WE HAVE TO JOIN THEM THROUGH A COMMON DENOMINATOR OR A COMMON VARIABLE.**



**137. TO CHOOSE THE COMMON VARIABLE, WE NEED TO CLICK THE ADD CONDITION IN THE JOIN CONDITIONS FIELD.**

**Join type**  
Select the type of join to perform.

**Inner join**  
Select all rows from both datasets that meet the join condition.

**Join conditions**  
Select a field from each parent node for the join condition.

**Add condition**

**⚠ All fields must have options selected**

**138. IN THE CHANGE SCHEMA, CHOOSE CUSTOMERID AND USERID. WE ARE GOING TO JOIN THEM THROUGH THIS COMMON DENOMINATOR OR VARIABLE WHICH ARE THE IDS OF THE CUSTOMERS OR USERS OF THIS MOVIE STREAMING PLATFORM.**

**Join conditions**

Select a field from each parent node for the join condition.

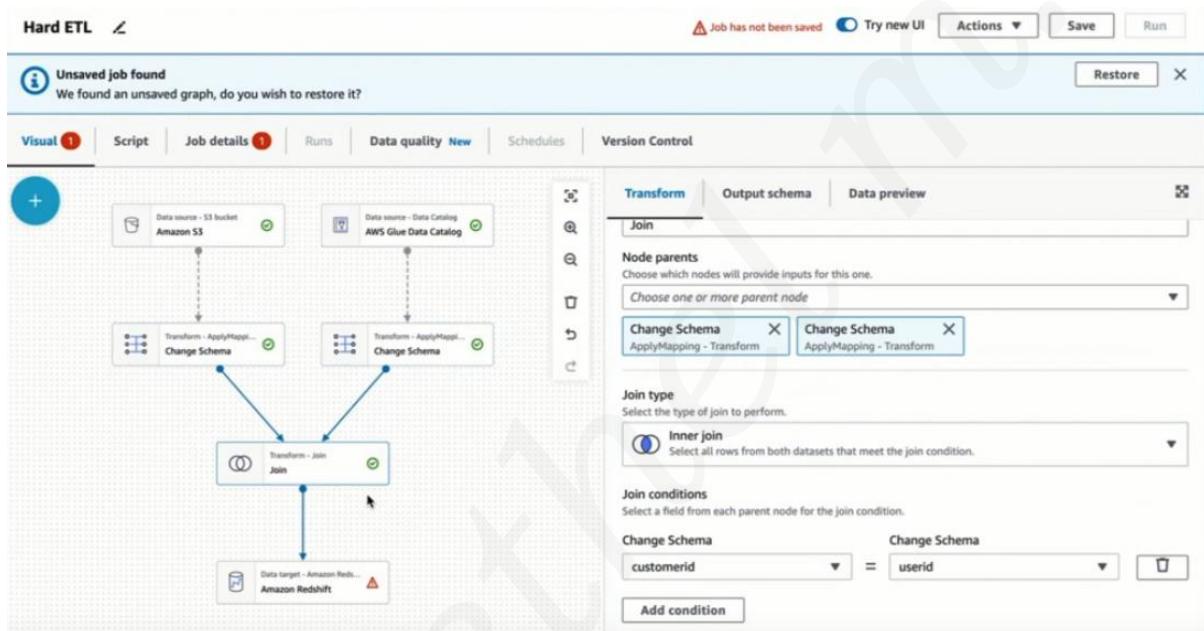
Change Schema      Change Schema

customerid      =      userid

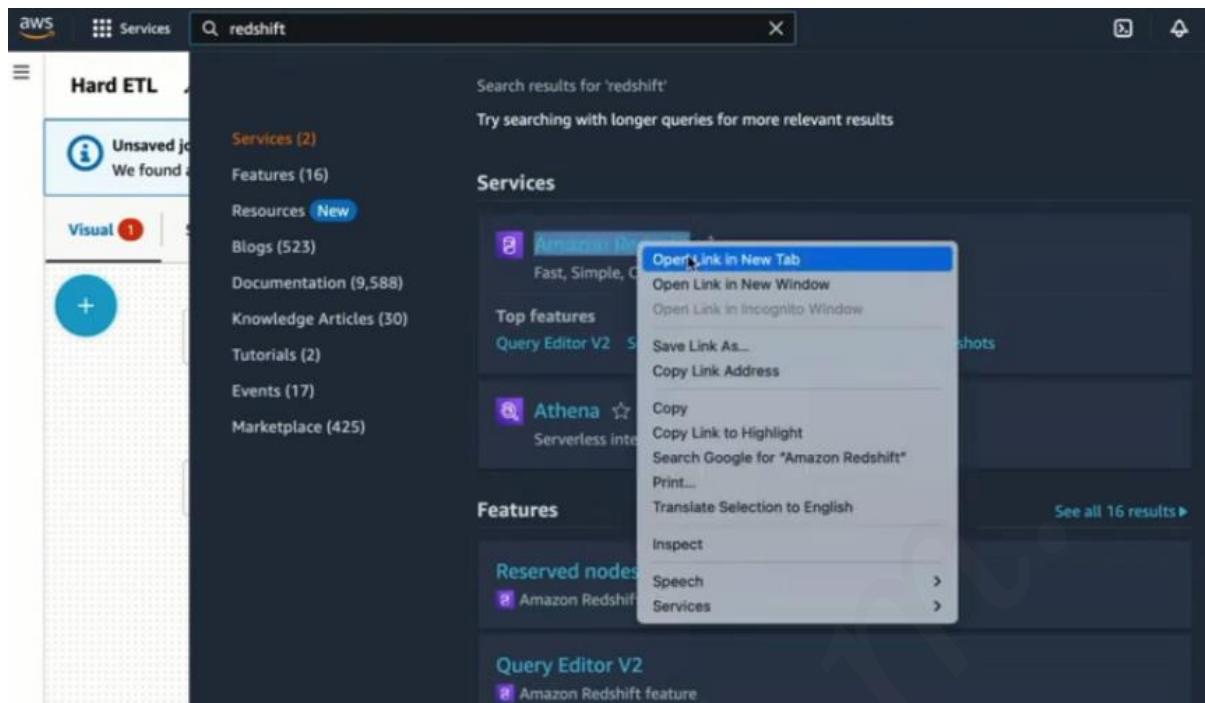
Add condition



**139. THIS WILL THEN VALIDATES THE TRANSFORM CELL. YOU ARE NOW DONE TO THE TRANSFORM PART OF THE ETL PROCESS. KEEP THE TAB OPEN.**



**140. MOVING ON TO THE FINAL STEP IS THE TARGET AMAZON REDSHIFT. IN THE SERVICES, TYPE REDSHIFT. RIGHT CLICK, THEN, OPEN IT IN THE NEW TAB.**



**141. CLICK THE TRY REDSHIFT SERVERLESS FREE TRIAL.**



**142. CLICK DEFAULT SETTINGS.**

Amazon Redshift Serverless > Get started with Amazon Redshift Serverless

## Get started with Amazon Redshift Serverless Info

To start using Amazon Redshift Serverless, set up your serverless data warehouse and create a database. You will receive \$293.60 credit towards your Redshift Serverless usage in this account.

### Configuration

**Use default settings**  
Default settings have been defined to help you get started. You can change them at any time later.

**Customize settings**  
Customize your settings for your specific needs.

▼ How it works

Next steps

**143. IN THE PERMISSIONS, WE HAVE TO CREATE ANOTHER ROLE**

### Permissions

① Associate an IAM role so that your serverless endpoint can LOAD and UNLOAD data. You can create an IAM role as the default for this configuration that has the [AmazonRedshiftAllCommandsFullAccess](#) policy attached. This policy includes permissions to run SQL commands to COPY, UNLOAD, and query data with Amazon Redshift Serverless. This policy also grants permissions to run SELECT statements for related services, such as Amazon S3, Amazon CloudWatch logs, Amazon SageMaker, and AWS Glue. You won't be able to run these SQL commands without an IAM role attached to your namespace.

#### Associated IAM roles (0)

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.

Set default  Manage IAM roles

Q. Search for associated IAM role by name, status, or role type

IAM roles	Status	Role type
No resources	No associated IAM roles	

Associate IAM role

#### 144. GO BACK TO IAM. CLICK CREATE ROLE.

The screenshot shows the AWS IAM Roles list page. At the top, a green banner indicates "Role GlueFullAccessRole created." Below the banner, the page title is "Roles (10) Info". A sub-header says, "An IAM role is an identity you can create that has specific permissions with credentials that are valid for short durations. Roles can be assumed by entities that you trust." There is a search bar and a navigation bar with buttons for "Create role", "Delete", and "View role". The main table lists ten roles, each with a checkbox, a role name, and its identity provider or service linked role information. One role, "AWSReservedSSO\_AWSReservedSSO\_AWSPowerUserAccess\_72a7e95974f3825f", is highlighted with a red circle and a "Deletion failed." message.

#### 145. CHOOSE AWS SERVICE IN THE TRUSTED ENTITY TYPE.

The screenshot shows the "Step 1 Select trusted entity" screen. It includes a "Trusted entity type" section with four options: "AWS service" (selected), "AWS account", "Web identity", and "SAML 2.0 federation". Below this is a "Use case" section with the instruction "Allow an AWS service like EC2, Lambda, or others to perform actions in this account." A note states, "Redshift is selected." The sidebar shows steps 2 and 3: "Add permissions" and "Name, review, and create".

#### 146. IN THE 'USE CASES FOR OTHER AWS SERVICES', SELECT REDSHIFT.

The screenshot shows a dropdown menu titled "Choose a service to view use case". The menu lists various AWS services: Network Firewall, Nimble Studio, OpenSearch Ingestion, OpsWorks, Panorama, Personalize, Purchase Orders, QLDB, RAM, RDS, Reachability Analyzer, Redshift (selected), Rekognition, Resource Groups, RoboMaker, Roles Anywhere, S3, SMS, and SNS. The "Redshift" option is highlighted with a red circle.

## 147. SELECT REDSHIFT – CUSTOMIZABLE. THEN, CLICK NEXT.

### Use case

Allow an AWS service like EC2, Lambda, or others to perform actions in this account.

#### Common use cases

- EC2**  
Allows EC2 instances to call AWS services on your behalf.
- Lambda**  
Allows Lambda functions to call AWS services on your behalf.

#### Use cases for other AWS services:

Redshift

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

[Cancel](#) [Next](#)

## 148. KEEP THE ADMINISTRATOR ACCESS TO GIVE US FULL ACCESS. THEN, CLICK NEXT.

Step 2  
**Add permissions**

Step 3  
Name, review, and create

**Permissions policies (Selected 1/853) [Info](#)**  
Choose one or more policies to attach to your new role.

Filter policies by property or policy name and press enter.

Policy name	Type	Description
<input checked="" type="checkbox"/> AdministratorAccess	AWS m...	Provides full access to AWS services and resources.
<input type="checkbox"/> PowerUserAccess	AWS m...	Provides full access to AWS services and resources, but does not allow management of...
<input type="checkbox"/> ReadOnlyAccess	AWS m...	Provides read-only access to AWS services and resources.
<input type="checkbox"/> AWSCloudFormatio...	AWS m...	Provides access to AWS CloudFormation via the AWS Management Console.
<input type="checkbox"/> CloudFrontFullAcc...	AWS m...	Provides full access to the CloudFront console plus the ability to list Amazon S3 bucket...
<input type="checkbox"/> AWSCloudHSMFull...	AWS m...	Provides full access to all CloudHSM resources.
<input type="checkbox"/> AWSCloudHSMRea...	AWS m...	Provides read only access to all CloudHSM resources.
<input type="checkbox"/> ResourceGroupsan...	AWS m...	Provides full access to Resource Groups and Tag Editor.
<input type="checkbox"/> ResourceGroupsan...	AWS m...	Provides access to use Resource Groups and Tag Editor, but does not allow editing of t...
<input type="checkbox"/> CloudFrontReadOnl...	AWS m...	Provides access to CloudFront distribution configuration information and list distribution...
<input type="checkbox"/> CloudSearchFullAc...	AWS m...	Provides full access to the Amazon CloudSearch configuration service.
<input type="checkbox"/> CloudSearchReadO...	AWS m...	Provides read only access to the Amazon CloudSearch configuration service.
<input type="checkbox"/> CloudWatchFullAcc...	AWS m...	Provides full access to CloudWatch.

<input type="checkbox"/>	ResourceGroupsFullAccess	AWS managed policy	Provides full access to Resource Groups and Tag Editor.
<input type="checkbox"/>	ResourceGroupsReadOnlyAccess	AWS managed policy	Provides access to use Resource Groups and Tag Editor, but does not allow editing of them.
<input type="checkbox"/>	CloudFrontReadOnlyAccess	AWS managed policy	Provides access to CloudFront distribution configuration information and list distribution.
<input type="checkbox"/>	CloudSearchFullAccess	AWS managed policy	Provides full access to the Amazon CloudSearch configuration service.
<input type="checkbox"/>	CloudSearchReadOnlyAccess	AWS managed policy	Provides read only access to the Amazon CloudSearch configuration service.
<input type="checkbox"/>	CloudWatchFullAccess	AWS managed policy	Provides full access to CloudWatch.
<input type="checkbox"/>	CloudWatchReadOnlyAccess	AWS managed policy	Provides read only access to CloudWatch.
<input type="checkbox"/>	CloudWatchLogsFullAccess	AWS managed policy	Provides full access to CloudWatch Logs.
<input type="checkbox"/>	CloudWatchLogsReadOnlyAccess	AWS managed policy	Provides read only access to CloudWatch Logs.
<input type="checkbox"/>	AWSDirectConnectFullAccess	AWS managed policy	Provides full access to AWS Direct Connect via the AWS Management Console.
<input type="checkbox"/>	AWSDirectConnectReadOnlyAccess	AWS managed policy	Provides read only access to AWS Direct Connect via the AWS Management Console.
<input type="checkbox"/>	AmazonAppStreamFullAccess	AWS managed policy	Provides full access to Amazon AppStream via the AWS Management Console.
<input type="checkbox"/>	AmazonAppStreamReadOnlyAccess	AWS managed policy	Provides read only access to Amazon AppStream via the AWS Management Console.

#### ► Set permissions boundary - optional Info

Set a permissions boundary to control the maximum permissions this role can have. This is not a common setting, but you can use it to delegate permission management to others.

[Cancel](#) [Previous](#) [Next](#)

## 149. GIVE THE NAME, REDSHIFTFULLACCESSROLE.

Step 1  
Select trusted entity

Name, review, and create

Step 2  
Add permissions

Role details

Step 3  
Name, review, and create

Role name  
Enter a meaningful name to identify this role.

RedshiftFullAccessRole

Maximum 64 characters. Use alphanumeric and '-' characters.

Description  
Add a short explanation for this role.

Allows Redshift clusters to call AWS services on your behalf.

Maximum 1000 characters. Use alphanumeric and '-' characters.

Step 1: Select trusted entities

```

1. [
2.     "Version": "2012-10-17",
3.     "Statement": [
4.         {
5.             "Effect": "Allow",
6.             "Action": [
7.                 "sts:AssumeRole"
8.             ],
9.             "Principal": {
10.                 "Service": [
11.                     "redshift.amazonaws.com"
12.                 ]
13.             }
14.         }
15.     ]
16. 
```

**150. SCROLL DOWN, THEN, CLICK CREATE ROLE.**

Step 2: Add permissions

Permissions policy summary

Policy name	Type	Attached as
AdministratorAccess	AWS managed - job function	Permissions policy

Tags

Add tags - optional [Info](#)  
Tags are key-value pairs that you can add to AWS resources to help identify, organize, or search for resources.

No tags associated with the resource.

Add tag

You can add up to 50 more tags.

Cancel Previous Create role

**151. THE ROLE IS NOW SUCCESSFULLY CREATED.**

Identity and Access Management (IAM)

View role X

Role RedshiftFullAccessRole created.

IAM > Roles

Roles (13) [Info](#)

An IAM role is an identity you can create that has specific permissions with credentials that are valid for short durations. Roles can be assumed by entities that you trust.

Delete Create role

Dashboard

**152. GO BACK TO REDSHIFT. CLICK ASSOCIATE IAM ROLE.**

Services Search (Option+5)

run SQL commands to COPY, UNLOAD, and query data with Amazon Redshift Serverless. This policy also grants permissions to run SELECT statements for related services, such as Amazon S3, Amazon CloudWatch logs, Amazon SageMaker, and AWS Glue. You won't be able to run these SQL commands without an IAM role attached to your namespace.

Associated IAM roles (0)

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.

Set default Manage IAM roles

Search for associated IAM role by name, status, or role type

< 1 >

IAM roles	Status	Role type
No resources No associated IAM roles		

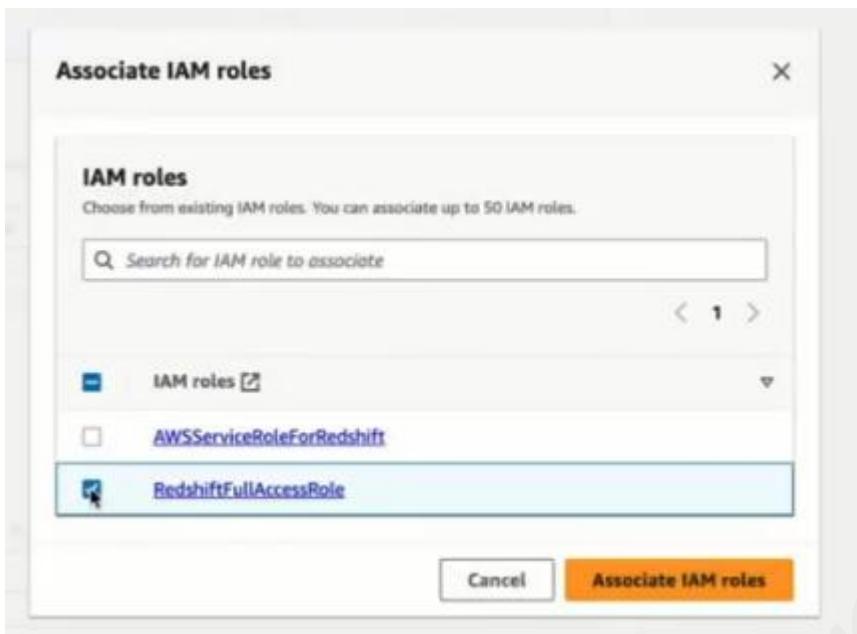
Associate IAM role

Encryption and security

AWS KMS encryption  
AWS owned KMS key

Audit logging  
Off

153. SELECT REDSHIFTFULLACCESSROLE. THEN, CLICK ASSOCIATE IAM ROLES.



154. SELECT REDSHIFTFULLACCESSROLE.

④ Associate an IAM role so that your serverless endpoint can LOAD and UNLOAD data. You can create an IAM role as the default for this configuration that has the **AmazonRedshiftAllCommandsFullAccess** policy attached. This policy includes permissions to run SQL commands to COPY, UNLOAD, and query data with Amazon Redshift Serverless. This policy also grants permissions to run SELECT statements for related services, such as Amazon S3, Amazon CloudWatch logs, Amazon SageMaker, and AWS Glue. You won't be able to run these SQL commands without an IAM role attached to your namespace.

**Associated IAM roles (1/1)**  
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.

**Set default** ▾ **Manage IAM roles** ▾

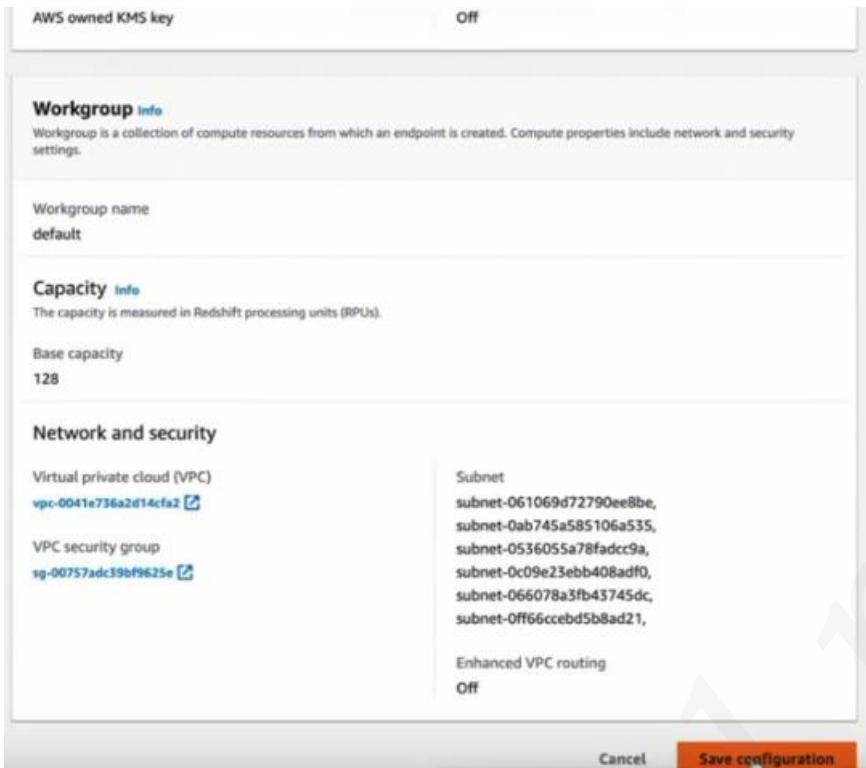
**Search for associated IAM role by name, status, or role type**

IAM roles	Status	Role type
RedshiftFullAccessRole	Not applied	--

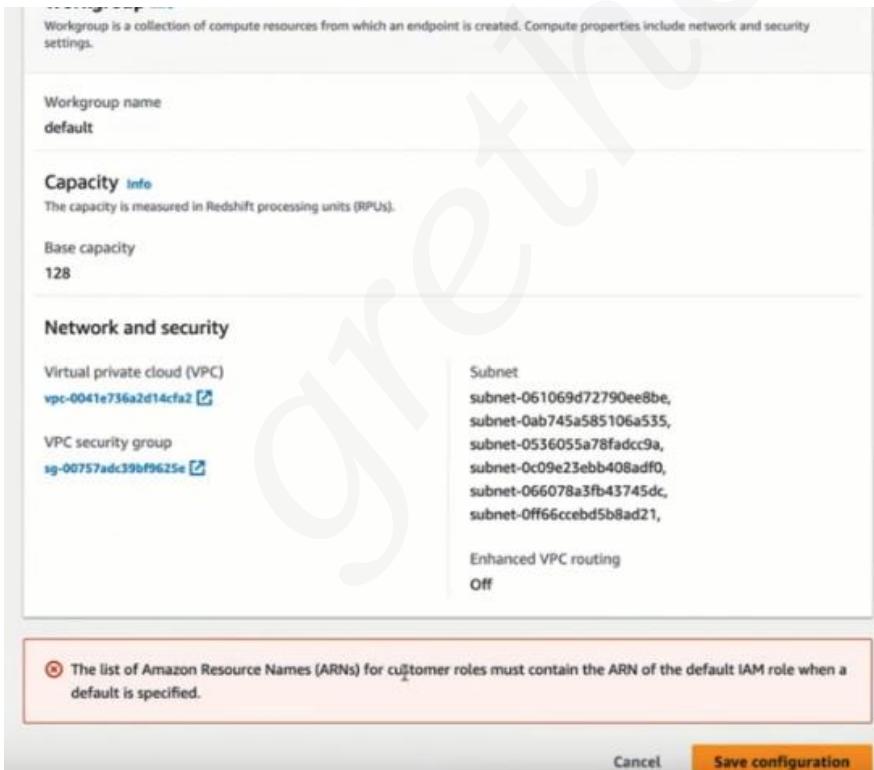
**Encryption and security**

AWS KMS encryption	Audit logging
AWS owned KMS key	Off

**155. SCROLL DOWN, THEN, SAVE CONFIGURATION.**



**156. HERE, WE WILL GET A MESSAGE THAT SAYS SOMETHING ABOUT ARN.**



## 157. TO FIX THE ISSUE, SCROLL UP, CLICK CUSTOMIZE SETTINGS.

To start using Amazon Redshift Serverless, set up your serverless data warehouse and create a database. You will receive \$293.60 credit towards your Redshift Serverless usage in this account.

**Configuration**

Use default settings  
Default settings have been defined to help you get started. You can change them at any time later.

Customize settings  
Customize your settings for your specific needs.

**Namespace Info**  
Namespace is a collection of database objects and users. Data properties include database name and password, permissions, and encryption and security.

**Namespace name**  
This is a unique name that defines the namespace.  
  
The name must be from 3-64 characters. Valid characters are a-z (lowercase only), 0-9 (numbers), and - (hyphen).

**▼ Database name and password**

**Database name**  
The name of the first database in the Amazon Redshift Serverless environment.  
  
The name must be 1-64 alphanumeric characters (lowercase only), and it can't be a reserved word.

**Admin user credentials**  
IAM credentials provided as your default admin user credentials. To add a new admin username and password, customize admin user credentials.  
 Customize admin user credentials

## 158. IN THE NAMESPACE, LET'S CALL IT MYNAMESPACE.

**Namespace Info**  
Namespace is a collection of database objects and users. Data properties include database name and password, permissions, and encryption and security.

**Namespace name**  
This is a unique name that defines the namespace.  
  
The name must be from 3-64 characters. Valid characters are a-z (lowercase only), 0-9 (numbers), and - (hyphen).

**▼ Database name and password**

**Database name**  
The name of the first database in the Amazon Redshift Serverless environment.  
  
The name must be 1-64 alphanumeric characters (lowercase only), and it can't be a reserved word.

**Admin user credentials**  
IAM credentials provided as your default admin user credentials. To add a new admin username and password, customize admin user credentials.  
 Customize admin user credentials  
To use the default IAM credentials, clear this option.

**▼ Permissions**

Associate an IAM role so that your serverless endpoint can LOAD and UNLOAD data. You can create an IAM role as the default for this configuration that has the **AmazonRedshiftAllCommandsFullAccess**  policy attached. This policy includes permissions to run SQL commands to COPY, UNLOAD, and query data with Amazon Redshift Serverless. This policy also

**159. SCROLL DOWN, CLICK ASSOCIATE IAM ROLE.**

The screenshot shows the AWS Lambda console with the 'Associated IAM roles' section. A callout box highlights a note about associating an IAM role for serverless endpoints. Below the note, there's a search bar and two buttons: 'Set default' and 'Manage IAM roles'. A large button labeled 'Associate IAM role' is prominently displayed.

Associated IAM roles (0)

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.

Set default Manage IAM roles

Q Search for associated IAM role by name, status, or role type

< 1 >

IAM roles	Status	Role type
No resources	No associated IAM roles	

Associate IAM role

▼ Security and encryption

**160. SELECT REDSHIFTFULLACCESSROLE, THEN, ASSOCIATE IAM ROLES.**

The screenshot shows the 'Associate IAM roles' dialog box. It lists available IAM roles: 'AWSLambdaVPCAccessRole' (selected), 'AWSServiceRoleForRedshift', and 'RedshiftFullAccessRole' (highlighted). At the bottom are 'Cancel' and 'Associate IAM roles' buttons.

Associate IAM roles

IAM roles

Choose from existing IAM roles. You can associate up to 50 IAM roles.

Q Search for IAM role to associate

< 1 >

IAM roles
AWSLambdaVPCAccessRole
AWSServiceRoleForRedshift
RedshiftFullAccessRole

Cancel Associate IAM roles

## 161. SELECT REDSHIFTFULLACCESSROLE.

The screenshot shows the AWS IAM Roles page. At the top, there are buttons for "Set default" and "Manage IAM roles". Below that is a search bar with placeholder text "Search for associated IAM role by name, status, or role type". A navigation bar indicates "1" item. The main table has columns for "Status" and "Role type". A checkbox next to "IAM roles" is checked. In the table, there is one row for "RedshiftFullAccessRole" which is marked as "Not applied".

**Security and encryption**

**⚠️** Your data is encrypted by default with an AWS owned key. To choose a different key, customize your encryption settings.

Customize encryption settings (advanced)

**Audit logging** Info  
Collects logging information for the database.

**Export these logs:**

User log  
 Connection log  
 User activity log

## 162. IN THE WORKGROUP, WE WILL CALL IT, MYWORKGROUP.

**Workgroup** Info  
Workgroup is a collection of compute resources from which an endpoint is created. Compute properties include network and security settings.

**Workgroup name**  
This is a unique name that defines the workgroup.  
  
The name must be from 3-64 characters. Valid characters are a-z (lowercase only), 0-9 (numbers), and - (hyphen).

**Capacity**  
Set the base capacity used to process your data warehouse workloads. The capacity is measured in Redshift processing units (RPUs). To improve query performance, increase the RPU value.

**Base capacity**  
Base RPU capacity is set to 128 RPUs by default. To change the base RPU capacity, choose another value from the list.  
 ▾  
Range must be 8-512 in increments of 8.

**Network and security**

**163. SCROLL DOWN, CLICK SAVE CONFIGURATION.**

Virtual private cloud (VPC)  
This VPC defines the virtual networking environment for this database.

vpc-0041e736a2d14cfa2

VPC security groups  
This VPC security group defines which subnets and IP ranges can be used in the VPC.

Choose one or more security groups

sg-00757adc39bf9625e X

Subnet  
The subnet in the chosen VPC that is associated with the specified database.

Choose three or more subnet IDs

subnet-061069d72790ee8be X subnet-0ab745a585106a535 X

subnet-0536055a78fadcc9a X subnet-0c09e23ebb408adf0 X

subnet-066078a3fb43745dc X subnet-0ff66ccebd5b8ad21 X

Enhanced VPC routing  
Turning on this option routes network traffic between your serverless database and data repositories through a VPC instead of the internet.

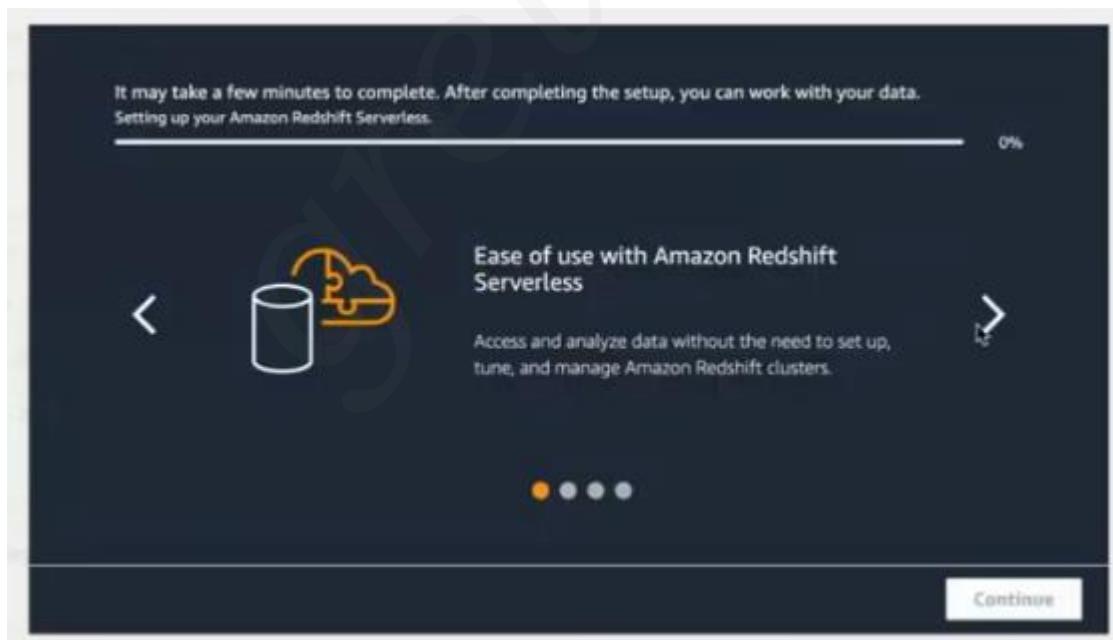
Turn on enhanced VPC routing

**② The list of Amazon Resource Names (ARNs) for customer roles must contain the ARN of the default IAM role when a default is specified.**

Cancel **Save configuration**

**164. AND NOW, IT IS SUPPOSED TO WORK. CONGRATULATIONS, YOU DID THE FINAL TOUCH OF ETL PROCESS! WE NOW HAVE A VERY WELL-ORGANIZED STUDIO OR PLATFORM WHERE WE CAN HAVE ALL OUR DATA COLLECTED AND GATHERED AND JOINED. THANKS TO THE ETL PROCESS THAT WE IMPLEMENTED IN GLUE!**

**NOW, CLICK CONTINUE.**



**165. YOU WILL NOW THEN SEE THE SERVERLESS DASHBOARD. CLICK THE MYNAMESPACE. THEN, CLICK QUERY DATA.**

Successfully setup Amazon Redshift Serverless

Review your configuration settings. To query data, go to query editor.

Query data

Amazon Redshift Serverless

Serverless dashboard

Namespace overview

Total snapshots: 0

Datashares in my account: 0

Datashares requiring authorization: 0

Datashares from other accounts: 0

Datashares requiring association: 0

Namespaces / Workgroups

Namespace	Status	Workgroup	Status
mynamespace	Available	myworkgroup	Available

Queries metrics

Free trial

Free trial credits remaining: \$293.60 out of \$300.00

Free trial expiration: September 12, 2023

Alarms (0)

View in CloudWatch

Successfully setup Amazon Redshift Serverless

Review your configuration settings. To query data, go to query editor.

Query data

Amazon Redshift Serverless > Namespace configuration > mynamespace

mynamespace

General information

Namespace	Status	Admin user name
mynamespace	Available	admin
Namespace ID	Date created	Database name
4fe86c69-31d5-4aff-98cd-e52c1c2d1be6	June 14, 2023, 21:10 (UTC+02:00)	dev
Namespace ARN	Storage used	Total table count
arn:aws:redshift-serverless:us-east-1:749601114921:n...	0 bytes	-

Workgroup

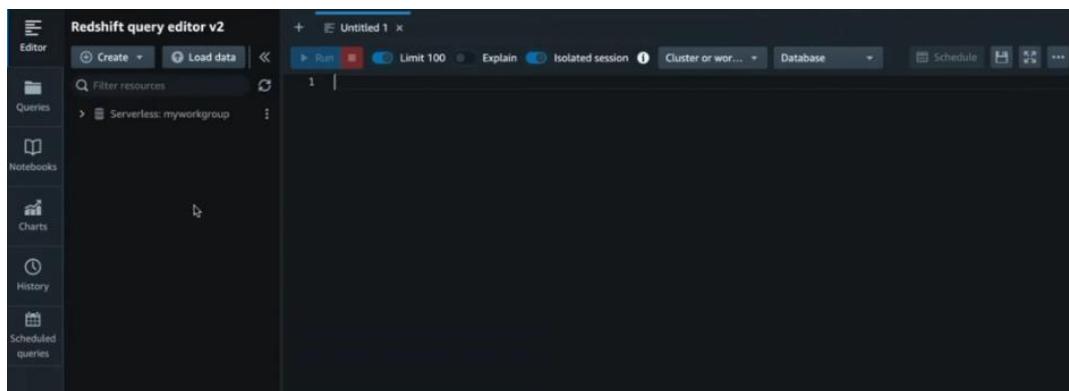
Workgroup name

Set up compute resources for your workgroup.

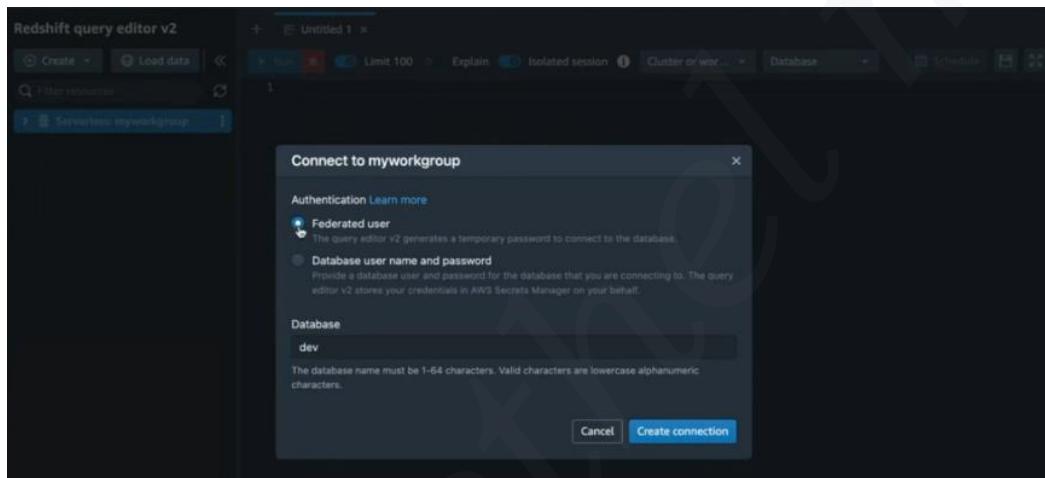
Workgroup	Status
myworkgroup	Available

Actions

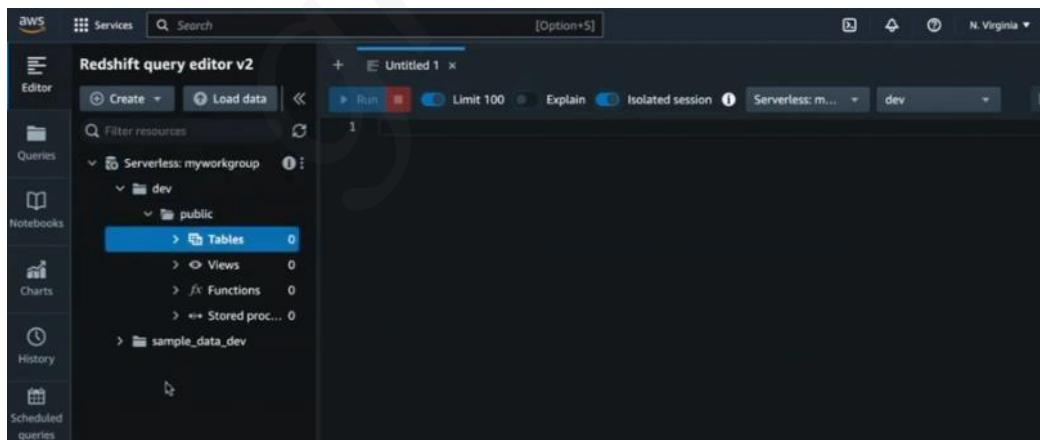
**166. INSIDE THE QUERY DATA, WE ARE GOING TO CREATE OUR FINAL TABLE THAT IS GOING TO CONTAIN THE FINAL RESULT OF THIS WHOLE ETL PROCESS.**  
**FIRST, WE ARE GOING TO CREATE THE CONNECTION. CLICK SERVERLESS MYWORKGROUP.**



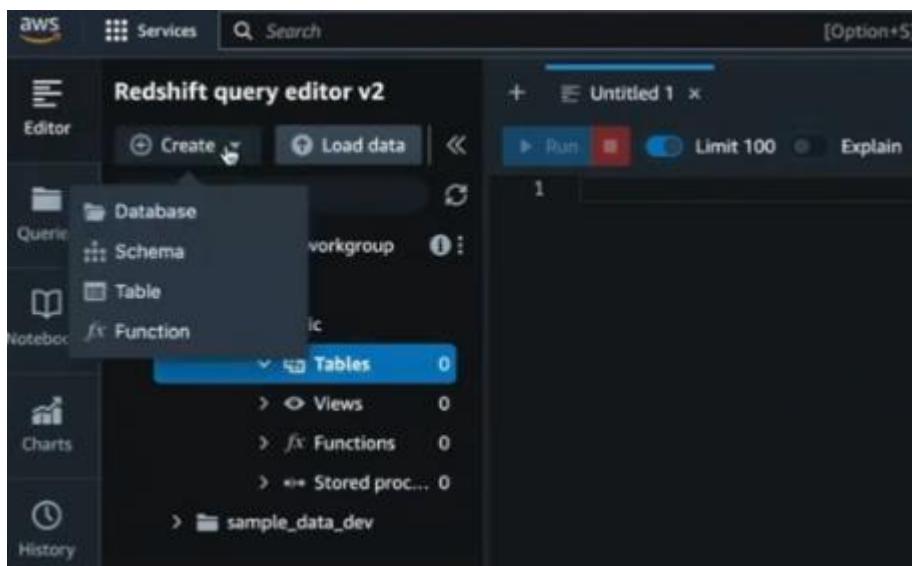
**167. KEEP FEDERATED USER AND THE DEV IN THE DATABASE. THEN. CLICK CREATE CONNECTION.**



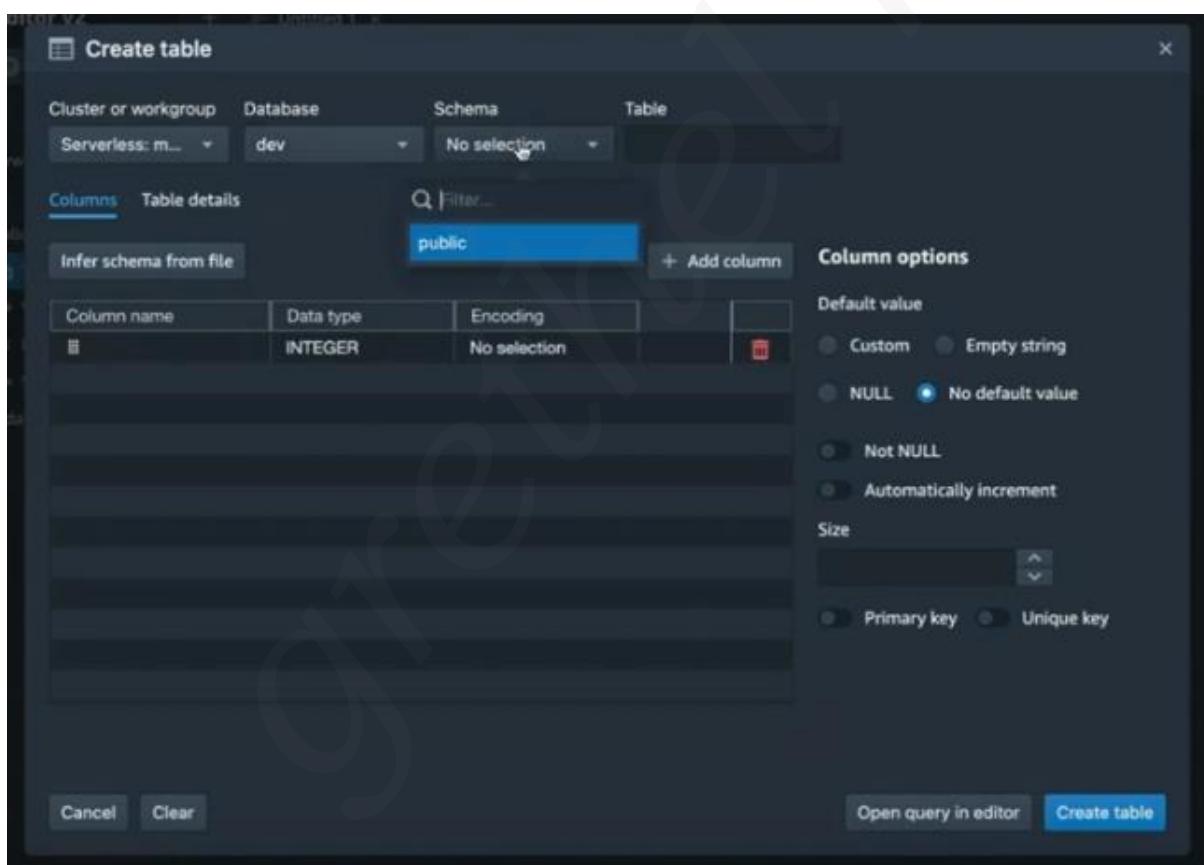
**168. CONNECTION SUCCESFUL. EXPAND THE DEV, PUBLIC, AND TABLES FOLDERS. THEN, WE ARE GOING TO CREATE THE TABLE.**



**169. CLICK CREATE, THEN CHOOSE TABLE.**



**170. IN CLUSTER OR WORKGROUP, SELECT SERVERLESS MY WORKGROUP; DATABASE IS DEV; SCHEMA IS PUBLIC.**



**171. NAME THE TABLE AS FINALTABLE. THEN ADD THE COLUMNS.**

The screenshot shows the 'Create table' dialog in the AWS Glue console. The 'Table' tab is selected, showing the table name 'finaltable'. The 'Columns' tab is active, displaying a single column configuration. The column is named 'id', has a data type of 'INTEGER', and no encoding is selected. To the right, there are 'Column options' for setting a default value (No default value is selected), size (size is set to 10), and key types (Primary key is selected). At the bottom right, the 'Create table' button is highlighted in blue.

**172. BASED ON OUR VISUAL EDITOR, IN THE OUTPUT SCHEMA, WE CAN CHECK OUR COLUMN NAMES AND ADD IT IN OUR QUERY TABLE.**

The screenshot shows the 'Hard ETL' visual editor in the AWS Glue console. It displays a data pipeline graph with two data sources: 'Data source - S3 bucket Amazon S3' and 'Data source - Data Catalog AWS Glue Data Catalog'. Both sources feed into separate 'Transform - ApplyMapping Change Schema' nodes. The outputs of these nodes merge into a single 'Transform - Join' node, which then points to the 'Data target - Amazon Redshift' node. On the right, the 'Output schema' tab is selected, showing the target table's schema with the following columns:

Key	Data type
geography	string
haspremiummembership	int
customerid	int
gender	string
isactivemember	int
age	int
userid	long
movieid	long
rating	double

173. FOLLOWING THE OUTPUT SCHEMA, THIS SHOULD BE THE EXPECTED OUTPUT. THEN, CLICK CREATE TABLE.

The screenshot shows the 'Create table' wizard in the AWS Redshift console. The top navigation bar includes 'Cluster or workgroup' (Serverless: m...), 'Database' (dev), 'Schema' (public), and 'Table' (finaltable). The 'Columns' tab is selected, showing the following schema:

Column name	Data type	Encoding	Actions
geography	CHAR	No selection	trash
haspremiummem...	INTEGER	No selection	trash
customerid	BIGINT	No selection	trash
gender	CHAR	No selection	trash
isactivemember	INTEGER	No selection	trash
age	INTEGER	No selection	trash
userid	BIGINT	No selection	trash
movied	BIGINT	No selection	trash
rating	DECIMAL	No selection	trash

On the right side, there are 'Column options' for each column, including 'Default value' (set to 'No default value'), 'Size' (set to 1000), and key selection buttons ('Primary key' and 'Unique key').

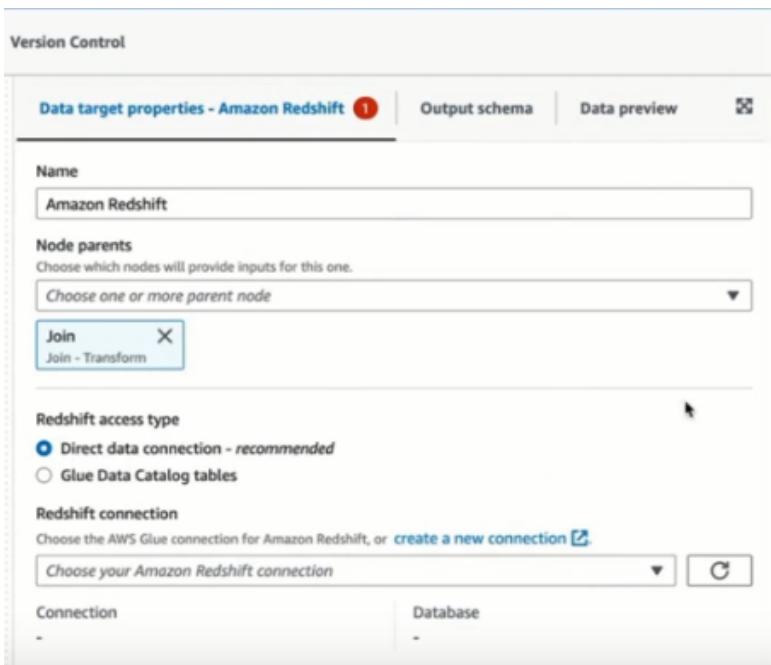
At the bottom left are 'Cancel' and 'Clear' buttons. At the bottom right are 'Open query in editor' and 'Create table' buttons.

174. FINALTABLE IS NOW CREATED SUCCESSFULLY.

The screenshot shows the Redshift query editor v2. The sidebar on the left includes 'Editor', 'Queries', 'Notebooks', 'Charts', and 'History'. The main area displays the 'Redshift query editor v2' interface with a tree view of database objects. A green notification bar at the top right states 'finaltable table is created successfully.' The tree view shows:

- Serverless: myworkgroup
  - dev
    - public
      - Tables (1)
        - finaltable
      - Views 0
      - Functions 0
      - Stored proc... 0
  - sample\_data\_dev

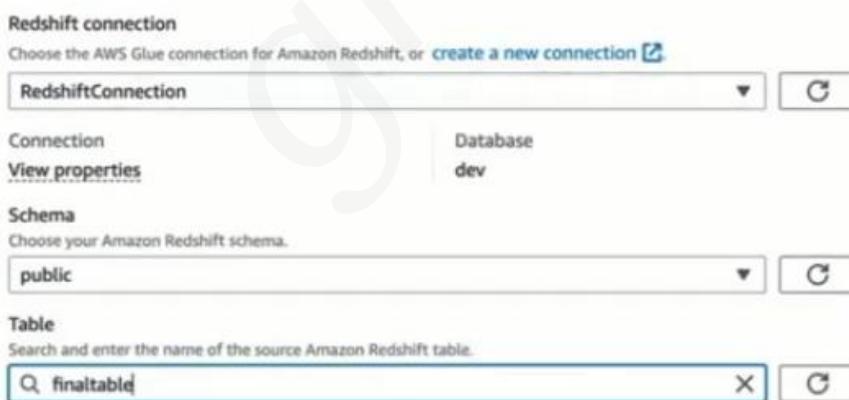
**175. MAKE THE FINAL CONNECTION USING THE VISUAL EDITOR. CLICK THE DATA TARGET PROPERTIES. SELECT DIRECT DATA CONNECTION IN THE REDSHIFT ACCESS TYPE.**



**176. IN THE REDSHIFT CONNECTION, CLICK THE REFRESH BUTTON TO SEE THE CONNECTION THAT WE`VE CREATED IN THE QUERY EDITOR. THEN, SELECT REDSHIFT CONNECTION.**



**177. IN THE SCHEMA, CHOOSE PUBLIC. IN THE TABLE, CHOOSE FINALTABLE.**



## 178. KEEP THE APPEND (INSERT) TO TARGET TABLE.

Schema  
Choose your Amazon Redshift schema.

Table  
Search and enter the name of the source Amazon Redshift table.

Handling of data and target table

APPEND (insert) to target table  
AWS Glue will append data to existing columns of the table and discard any extra columns.

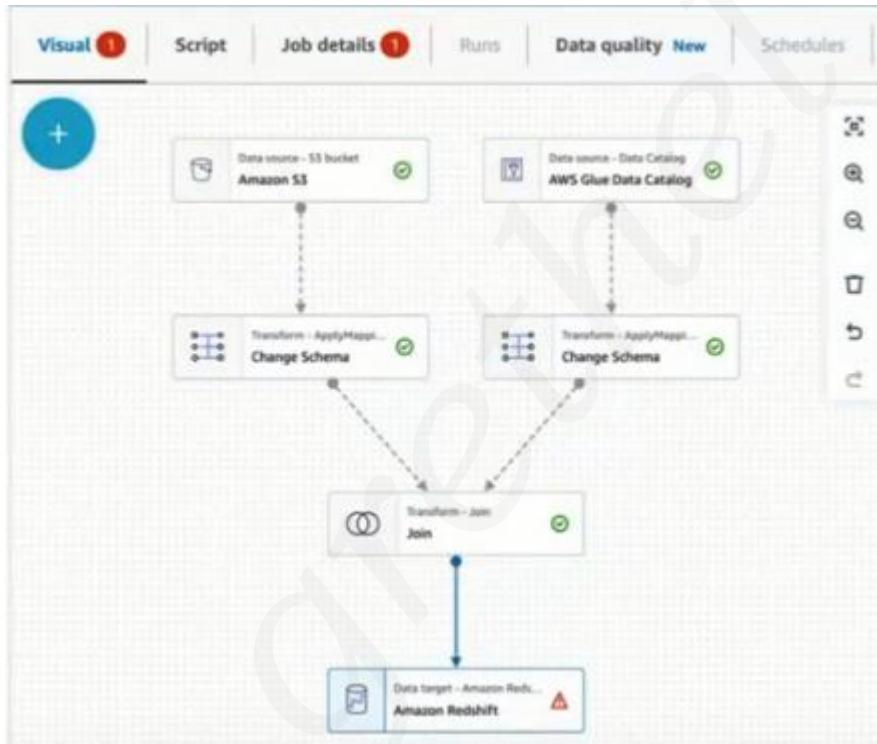
MERGE data into target table  
AWS Glue will either update or append data to the table based on a set of conditions.

TRUNCATE target table  
Same as Append, except AWS Glue will first clear the contents of the table.

DROP and recreate target table  
AWS Glue will delete and recreate the table with the schema from the source data.

Also update existing records in target table  
Update records already in "finaltable" in addition to adding new records.

## 179. TO SAVE THE JOB, CLICK JOB DETAILS.



**180. IN THE IAM ROLE, CHOOSE GLUEFULLACCESSROLE.**

The screenshot shows the 'Job details' tab selected in the top navigation bar. Under 'Basic properties', the 'Name' field is set to 'Hard ETL'. The 'Description - optional' field is empty. In the 'IAM Role' section, a dropdown menu titled 'Choose one IAM role' is open, showing a list of roles. The 'GlueFullAccessRole' is listed and highlighted with a blue border, indicating it is selected. A tooltip for 'GlueFullAccessRole' states: 'Allows Glue to call AWS services on your behalf.' Other roles listed include 'Spark'.

**181. TO LOOK FOR THE SCRIPT OF THE WHOLE ETL PROCESS, CLICK SCRIPT.**

The screenshot shows the 'Script' tab selected in the top navigation bar. A message at the top says 'Unsaved job found' with a note: 'We found an unsaved graph, do you wish to restore it?'. Below the tabs, there's a 'Script (Locked)' section with an 'Info' link. To the right are buttons for 'Generate classic script', 'Download script', and 'Edit script'. The main area contains the generated Python code:

```
1 import sys
2 from awsglue.transforms import *
3 from awsglue.utils import getResolvedOptions
4 from pyspark.context import SparkContext
5 from awsglue.context import GlueContext
6 from awsglue.job import Job
7 from awsglue import DynamicFrame
8
9 args = getResolvedOptions(sys.argv, ["JOB_NAME"])
10 sc = SparkContext()
11 glueContext = GlueContext(sc)
12 spark = glueContext.spark_session
13 job = Job(glueContext)
14 job.init(args["JOB_NAME"], args)
15
16 # Script generated for node Amazon S3
17 AmazonS3_node1686763165660 = glueContext.create_dynamic_frame.from_catalog(
18     database="movie-ratings-glue-database",
19     table_name="recommender_system_8730872805",
20     transformation_ctx="AmazonS3_node1686763165660",
21 )
22
23 # Script generated for node AWS Glue Data Catalog
24 AWSGlueDataCatalog_node1686763177727 = glueContext.create_dynamic_frame.from_catalog(
25     database="customer-features-glue-database",
```

**182. GOING BACK TO THE VISUAL, AT THE TOP, CLICK SAVE.**

The screenshot shows the AWS Glue visual editor interface. At the top, there is a warning message "Job has not been saved" and a "Save" button. Below the toolbar, a context menu is open over a "Job" node, with options like "Run with parameters", "Restore", "Delete", "Reset bookmark", "Download", "Upload", "Version control", "Push to repository", and "Pull from repository".

The main workspace displays a "Data target properties - Amazon Redshift" configuration. It includes fields for "Name" (set to "Amazon Redshift"), "Node parents" (with a dropdown menu "Choose one or more parent node" containing "Join" and "Join - Transform"), and "Redshift access type" (set to "Direct data connection - recommended").

On the right side, there is a "Redshift connection" section with a dropdown "RedshiftConnection" and a "Database" field set to "dev".

The screenshot shows the AWS Glue visual editor interface with the title "Hard ETL". At the top, there is a warning message "Job has not been saved" and a "Save" button. Below the toolbar, a context menu is open over a "Job" node, with options like "Run with parameters", "Restore", "Delete", "Reset bookmark", "Download", "Upload", "Version control", "Push to repository", and "Pull from repository".

The main workspace displays a complex ETL graph. It starts with two "Data source" nodes: "Data source - S3 bucket Amazon S3" and "Data source - Data Catalog AWS Glue Data Catalog". These feed into two "Transform - ApplyMapping..." nodes, each followed by a "Change Schema" node. The outputs of these four nodes converge at a "Transform - Join" node. Finally, the output of the join node feeds into a "Data target - Amazon Redshift" node.

On the right side, there is a "Data target properties - Amazon Redshift" configuration panel with fields for "Name" (set to "Amazon Redshift"), "Node parents" (with a dropdown menu "Choose one or more parent node" containing "Join" and "Join - Transform"), and "Redshift access type" (set to "Direct data connection - recommended").

On the far right, there are sections for "Output schema" and "Data preview".

**183. GOING BACK TO THE AWS GLUE, IN THE ETL JOBS, REFRESH THE ‘YOUR JOBS’ TO CHECK THE HARD ETL THAT WE DID.**

The screenshot shows the AWS Glue Studio interface. On the left, there's a sidebar with navigation links for AWS Glue, ETL jobs, Data Catalog, and Data Integration and ETL. The main area is titled "AWS Glue Studio" and shows the "Create job" section with five options: "Visual with a source and target" (selected), "Visual with a blank canvas", "Spark script editor", "Python Shell script editor", and "Jupyter Notebook". Below this, the "Source" and "Target" fields are set to "Amazon S3". Under "Your jobs (1)", there's a table with one entry: "Hard ETL" (Glue ETL), last modified on 6/14/2023, 9:29:03 PM, and AWS Glue version 3.0. There are "Actions" and "Run Job" buttons at the top of the "Your jobs" table.

**184. ON THE VISUAL EDITOR, CLICK RUN. BUT FOR THE SAKE OF THIS TRAINING, WE WILL NOT CLICK THE RUN BUTTON AS IT WILL INCUR SOME CHARGES. THE PURPOSE HERE IS TO JUST SHOW THE WHOLE PROCESS.**

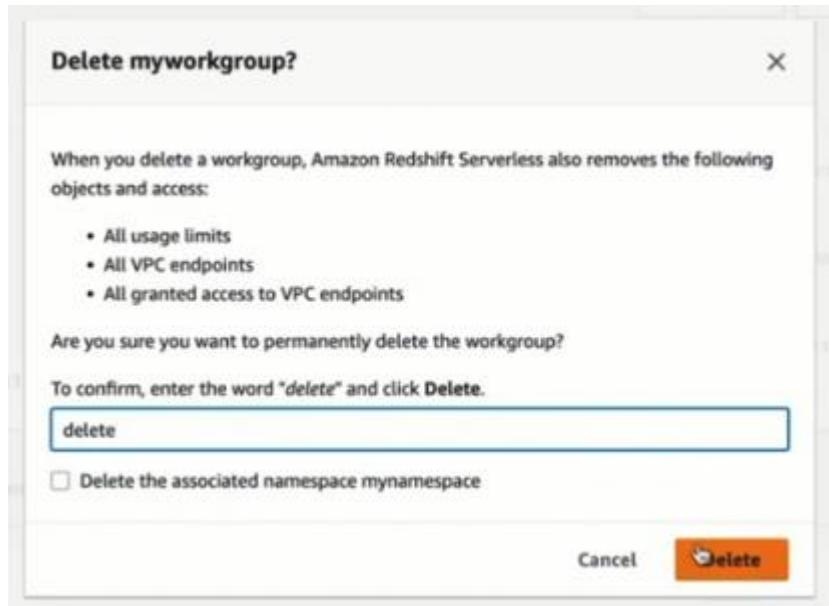
**CONGRATULATIONS, YOU DID THE WHOLE ETL PROCESS SUCCESSFULLY!**

The screenshot shows the "Hard ETL" job details page. A green header bar indicates "Successfully updated job" and provides a "Run" button. Below the header, tabs include "Visual" (selected), "Script", "Job details", "Runs", "Data quality", "Schedules", and "Version Control". The "Visual" tab shows two data sources: "Data source - S3 bucket Amazon S3" and "Data source - Data Catalog AWS Glue Data Catalog". On the right, there's a "Data target properties - Amazon Redshift" section with tabs for "Output schema" and "Data preview".

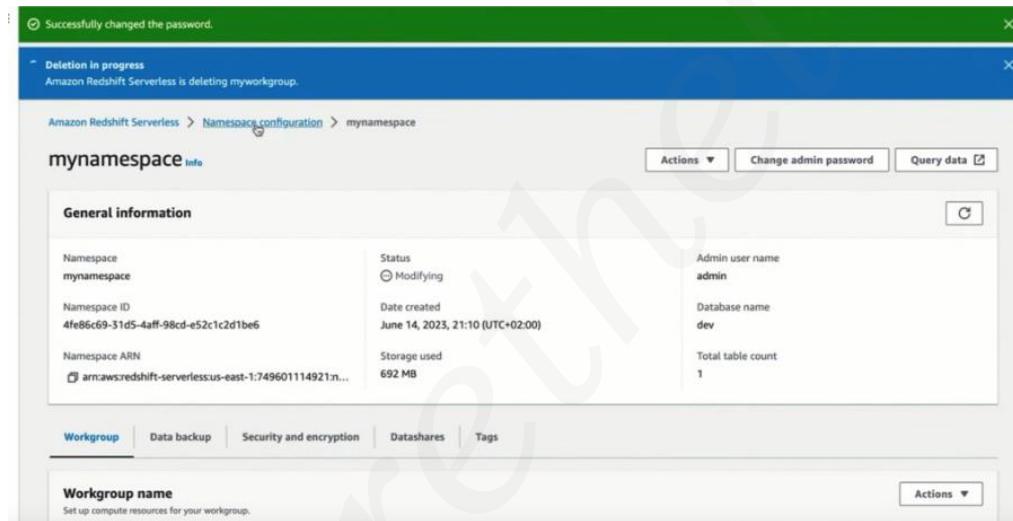
**185. TO AVOID CHARGES, WE NEED TO DELETE EVERYTHING. IN DOING THE CLEAN UP, FIRST, GO TO REDSHIFT, CLICK WORKGROUP NAME. CLICK ACTIONS, THEN, DELETE.**

The screenshot shows the "Workgroup" management page. It has tabs for "Workgroup" (selected), "Data backup", "Security and encryption", "Datashares", and "Tags". Under "Workgroup name", it shows "myworkgroup" and "Status Available". On the right, there are "Actions" and "Delete" buttons.

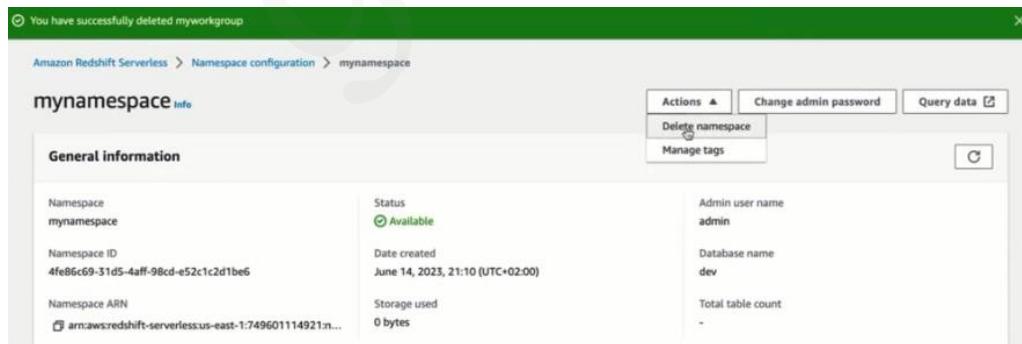
**ENTER DELETE. THEN, CLICK DELETE.**



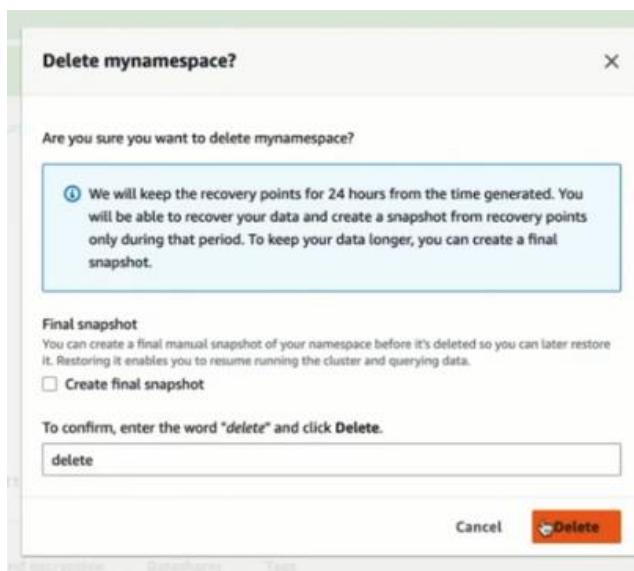
**186. THE NEXT TO DELETE IS THE NAME SPACE. GO TO THE NAMESPACE CONFIGURATION.**



**CLICK THE MYNAMESPACE, ACTIONS, THEN DELETE.**



**TYPE DELETE. CLICK DELETE.**

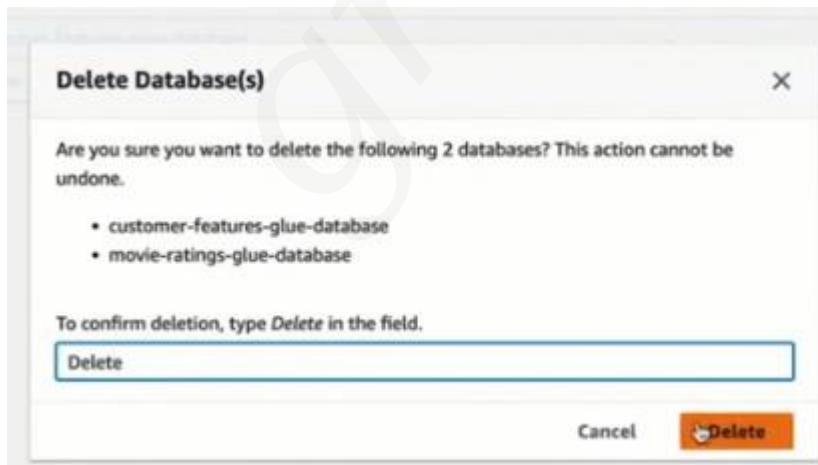


**187. NEXT IS TO CLEAN UP THE GLUE. CLICK THE DATABASES IN THE DATA CATALOG FIELD. SELECT BOTH DATABASES. CLICK DELETE.**

The screenshot shows the AWS Glue Data Catalog interface. On the left, a sidebar lists various options like "Getting started", "ETL jobs", and "Data Catalog". Under "Data Catalog", "Databases" is selected, showing a list of two databases: "customer-features-glue-database" and "movie-ratings-glue-database". Both databases have checkboxes next to their names, which are checked. The main area displays the database details with columns for Name, Description, Location URI, and Created on (UTC).

Name	Description	Location URI	Created on (UTC)
customer-features-glue-database	-	-	June 14, 2023 at 18:29:06
movie-ratings-glue-database	-	-	June 14, 2023 at 17:28:43

**ENTER DELETE, THEN, CLICK DELETE.**



## DATABASES SUCCESSFULLY DELETED.

The screenshot shows the AWS Glue Tables interface. At the top, a green header bar indicates "Some databases successfully deleted" with the message: "The following databases are now deleted: 'customer-features-glue-database', 'movie-ratings-glue-database'". Below this, the "Tables" section is displayed. A sub-header states: "A table is the metadata definition that represents your data, including its schema. A table can be used as a source or target in a job definition." A table header row includes columns for Name, Database, Location, Classification, Deprecated, and View data. A message below the table says "No available tables".

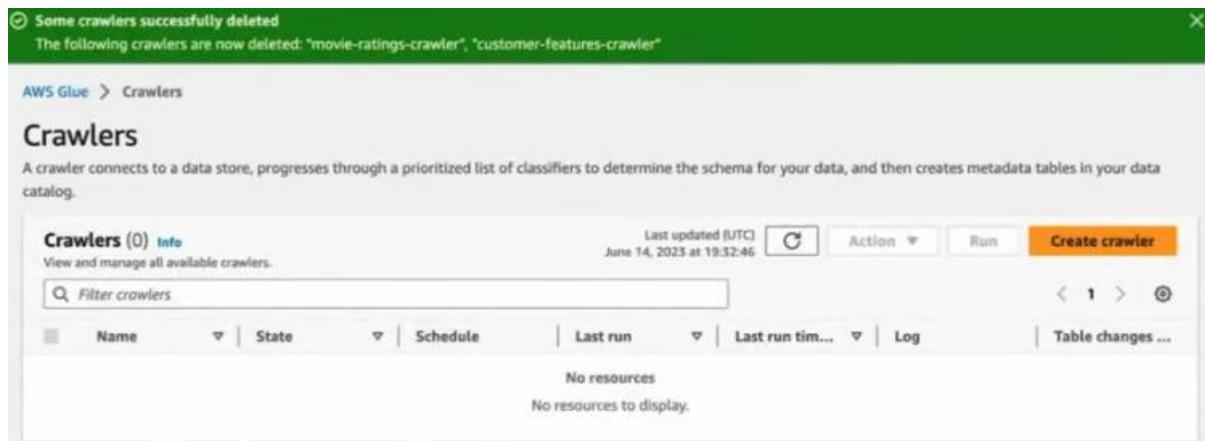
## 188. NEXT DELETE THE CRAWLERS. SELECT BOTH CRAWLERS. SELECT ACTIONS, THEN, DELETE CRAWLERS.

The screenshot shows the AWS Glue Crawlers interface. A green header bar indicates "Some databases successfully deleted" with the message: "The following databases are now deleted: 'customer-features-glue-database', 'movie-ratings-glue-database'". Below this, the "Crawlers" section is displayed. A sub-header states: "A crawler connects to a data store, progresses through a prioritized list of classifiers to determine the schema for your data, and then creates metadata tables in your data catalog." A table header row includes columns for Name, State, Schedule, Last run, and Last run time. Two crawlers are listed: "customer-features-crawler" (Ready, Succeeded, June 14, 2022) and "movie-ratings-crawler" (Ready, Succeeded, June 14, 2022). On the right side of the table, there is a context menu with options: Edit crawler, Duplicate crawler, Delete crawlers, View details, Resume schedule, Pause schedule, and Stop run. A sidebar on the left lists various AWS Glue services and features.

ENTER DELETE, CLICK DELETE.

The screenshot shows a confirmation dialog titled "Delete crawlers". The message asks: "Are you sure you want to delete the following 2 crawlers? This action cannot be undone." Below this is a list of the selected crawlers: "customer-features-crawler" and "movie-ratings-crawler". A text input field at the bottom prompts: "To confirm deletion, type Delete in the field." The word "Delete" is typed into this field. At the bottom right are two buttons: "Cancel" and a large orange "Delete" button with a cursor icon pointing to it.

## SUCCESSFULLY DELETED CRAWLERS.



189. THEN, DELETE THE DATA CONNECTIONS. IN THE CONNECTIONS, CLICK ACTIONS, THEN, DELETE.

The screenshot shows the AWS Glue Connectors list page. The left sidebar includes links for Getting started, ETL jobs, Data Catalog tables, Data connections, Data Catalog, Connections, and Data Integration and ETL. The main content area is titled "Connectors Info" and contains two sections: "Marketplace connectors" (with a "Go to AWS Marketplace" button) and "Custom connectors" (with a "Create custom connector" button). Below these is a table titled "Connectors (0) Info" with columns: Name, Type, and Last modified. A "Actions" dropdown menu is open over the first row, showing options: Actions ▲, Create connection, and Delete. The "Delete" option is highlighted with a cursor. The table shows one entry: "by property" (Type: JDBC, Last modified: Jun 14, 2023). At the bottom of the page, a modal dialog titled "Delete RDSConnection" is displayed, containing the message: "Deleting RDSConnection will permanently remove it as a source in Glue Studio. You will need to update your jobs accordingly." It also shows a status message: "Delete in progress." At the bottom of the dialog are "Cancel" and "Delete" buttons, with the "Delete" button being the primary target of the cursor.

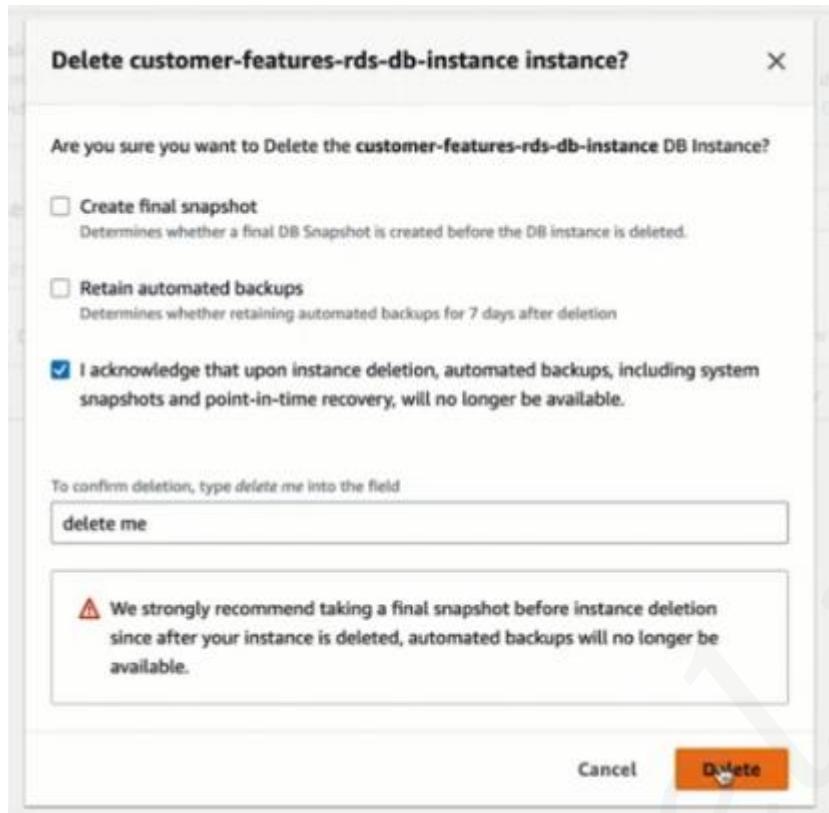
**190. THEN, DELETE THE ETL JOBS. CHOOSE HARD ETL. CLICK ACTIONS, THEN, DELETE JOB.**

The screenshot shows the AWS Glue Studio interface. On the left, the navigation menu includes sections for Getting started, ETL jobs (selected), Notebooks, Job run monitoring, Data Catalog tables, Data connections, Workflows (orchestration), Data Catalog (selected), Data Integration and ETL (selected), and ETL jobs. The main area displays the 'Create job' screen with five options: Visual with a source and target (selected), Visual with a blank canvas, Spark script editor, Python Shell script editor, and Jupyter Notebook. Below these, the 'Source' is set to 'Amazon S3' and the 'Target' is set to 'Amazon S3'. A table titled 'Your jobs (1)' lists a single job named 'Hard ETL' (Glue ETL). The 'Actions' column for this job has a dropdown menu open, showing options like Edit job, Clone job, Schedule job, Delete job(s) (which is highlighted with a blue border), and Reset job bookmark. A separate modal window titled 'Delete job(s)' asks if the user is sure they want to delete the selected job. The 'Delete' button in this modal is also highlighted with a blue border.

**191. IN THE AMAZON RDS, WE NEED TO DELETE THE DATABASES. SELECT THE DATABASE INSTANCE OF THE RDS. CLICK ACTIONS, THEN, DELETE.**

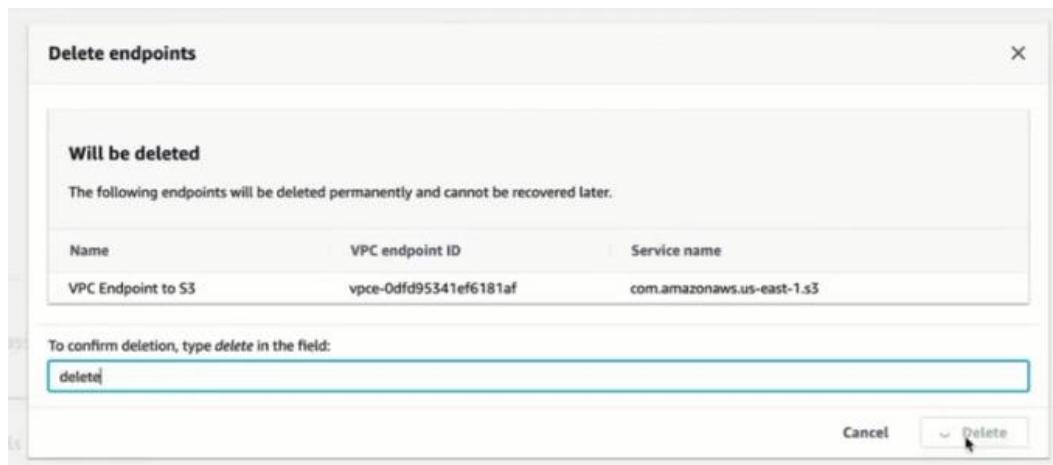
The screenshot shows the Amazon RDS Databases page. The left sidebar lists various RDS management features: Dashboard, Databases (selected), Query Editor, Performance insights, Snapshots, Exports in Amazon S3, Automated backups, Reserved instances, Proxies, Subnet groups, Parameter groups, Option groups, Custom engine versions, Events, Event subscriptions, Recommendations (with a notification count of 1), and Certificate update. The main content area shows a table titled 'Databases (1)'. The table has columns for DB identifier, Status, Role, and Engine. One row is selected, showing 'customer-features-rds-db-instance' with 'Available' status, 'Instance', and 'MySQL' engine. To the right of the table is a 'Actions' column with a dropdown menu. The 'Delete' option in this menu is highlighted with a blue border. Other actions listed in the menu include: Quick Actions - New, Convert to Multi-AZ deployment, Stop temporarily, Reboot, Set up EC2 connection, Create read replica, Create Aurora read replica, Create Blue/Green Deployment - new, Promote, Take snapshot, Restore to point in time, Migrate snapshot, Create RDS Proxy, and Create ElastiCache cluster - new.

UNSELECT CREATE FINAL SNAPSHOT AND RETAIN AUTOMATED BACKUPS. CLICK THE ACKNOWLEDGMENT. TYPE DELETE ME IN THE FIELD. CLICK DELETE.

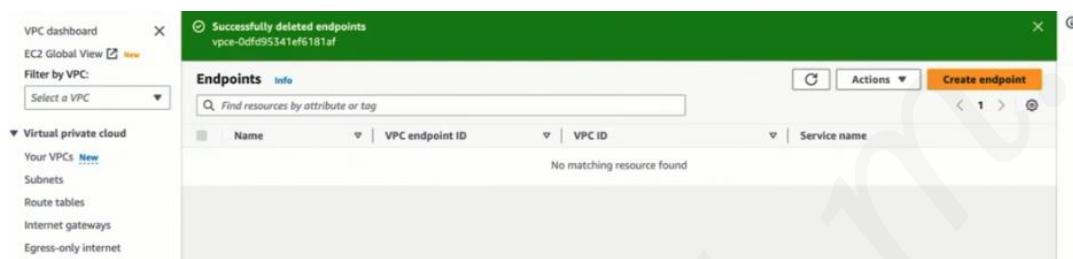


192. GO TO THE VPC DASHBOARD TO DELETE THE ENDPOINT VPC. SELECT THE AVAILABLE ENDPOINTS, ACTIONS, AND THEN DELETE VPC ENDPOINTS.

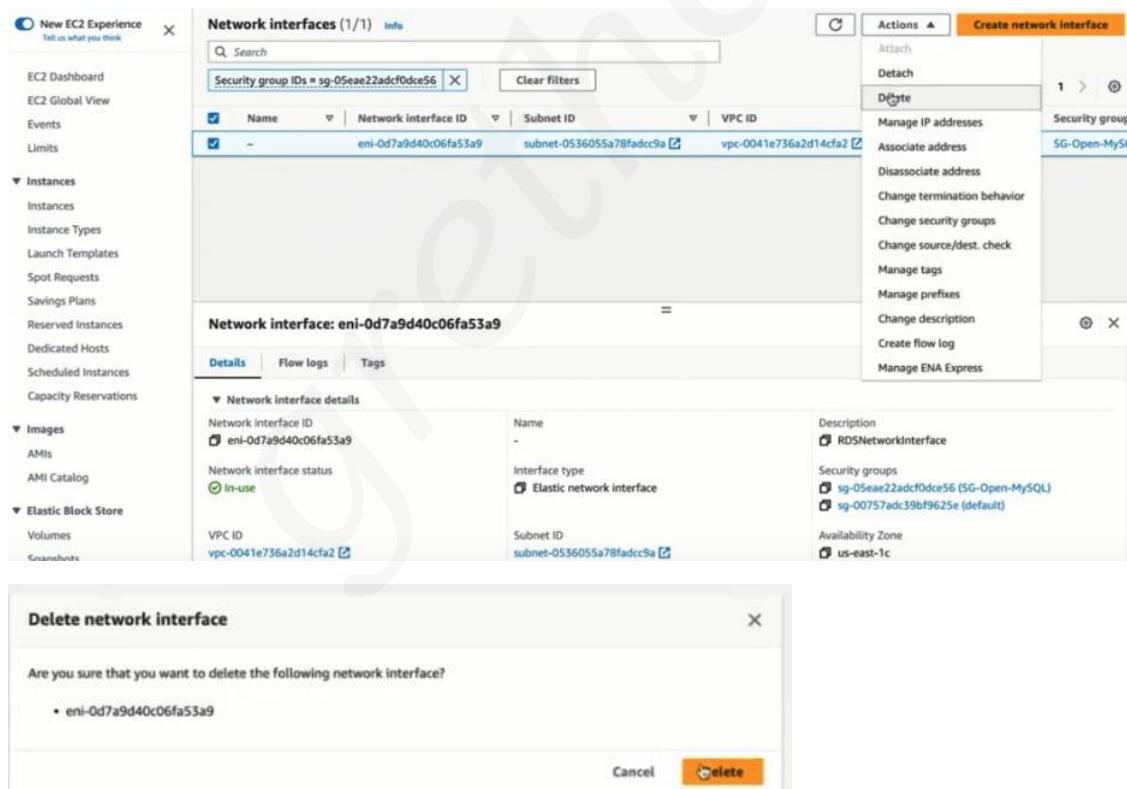
The screenshot shows the AWS VPC dashboard. On the left, there's a sidebar with options like "Virtual private cloud", "Endpoints", and "Actions". The main area displays a table of endpoints. One endpoint is selected: "VPC Endpoint to S3" with ID "vpce-0dfd95341ef6181af", VPC ID "vpc-0041e736a2d14cfa2", and service "com.amazonaws.s3". The "Actions" column for this endpoint includes "View details", "Manage subnets", "Manage security groups", "Manage route tables", "Manage policy", "Modify private DNS name", "Modify endpoint settings", "Manage tags", and "Delete VPC endpoints". Below the table, a detailed view for "vpce-0dfd95341ef6181af / VPC Endpoint to S3" is shown with tabs for "Details", "Route tables", "Policy", and "Tags".



**SUCCESSFULLY DELETED ENDPOINTS.**



**193. GO TO THE EC2 TO DELETE THE SECURITY GROUPS. BUT FIRST DELETE THE NETWORK INTERFACE. SELECT THE AVAILABLE NETWORK INTERFACE; ACTIONS; THEN, DELETE.**



**194. SELECT THE SECURITY GROUP, CLICK ACTIONS, THEN, DELETE SECURITY GROUPS.**

The first screenshot shows a list of security groups:

Name	Security group ID	Security group name	VPC ID	Description	Owner
-	sg-0ce2e3b4e3be04e92	SG-Open-HTTP	vpc-0041e736a2d14cfa2	Allows HTTP Traffic	749601
<input checked="" type="checkbox"/>	sg-05eae22adcf0dce56	SG-Open-MySQL	vpc-0041e736a2d14cfa2	Allows MySQL Access t...	749601
-	sg-04c7ce4cb6a0e059f	launch-wizard-1	vpc-0041e736a2d14cfa2	launch-wizard-1 create...	749601
-	sg-00757adc39bf9625e	default	vpc-0041e736a2d14cfa2	default VPC security gr...	749601

The second screenshot shows the 'Inbound rules' tab for the selected security group 'SG-Open-MySQL':

Name	Security group ID	Security group name	VPC ID	Description	Owner
-	sg-0ce2e3b4e3be04e92	SG-Open-HTTP	vpc-0041e736a2d14cfa2	Allows HTTP Traffic	749601
<input checked="" type="checkbox"/>	sg-05eae22adcf0dce56	SG-Open-MySQL	vpc-0041e736a2d14cfa2	Allows MySQL Access t...	749601
-	sg-04c7ce4cb6a0e059f	launch-wizard-1	vpc-0041e736a2d14cfa2	launch-wizard-1 create...	749601
-	sg-00757adc39bf9625e	default	vpc-0041e736a2d14cfa2	default VPC security gr...	749601

The third screenshot shows a confirmation dialog for deleting the selected security group:

Delete security groups

Are you sure that you want to delete this security group?

- sg-05eae22adcf0dce56 - SG-Open-MySQL

Cancel **Delete**

**195. GO TO S3 TO CLEAN UP THE BUCKET. SELECT THE BUCKET, CLICK EMPTY.**

The screenshot shows the AWS S3 buckets interface:

**Account snapshot:**

- Total storage: 63.0 B
- Object count: 2
- Average object size: 31.5 B
- You can enable advanced metrics in the "default-account-dashboard" configuration.

**Buckets (1) Info:**

Buckets are containers for data stored in S3. Learn more [\[link\]](#)

Name	AWS Region	Access	Creation date
recommender-system-8730872805	US East (N. Virginia) us-east-1	Bucket and objects not public	June 14, 2023, 19:26:11 (UTC+02:00)

**TYPE PERMANENTLY DELETE. CLICK EMPTY.**

The screenshot shows the 'Empty bucket' page for the 'recommender-system-8730872805' bucket. It includes a warning about the permanence of deletion, a note about lifecycle rules, and a text input field where 'permanently delete' is typed. The 'Empty' button is highlighted at the bottom right.

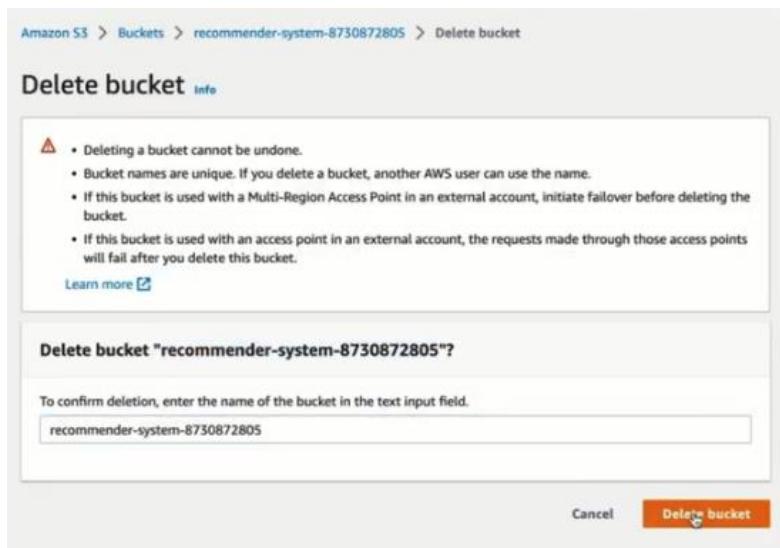
**CLICK EXIT.**

The screenshot shows the 'Empty bucket: status' page after deletion. It displays a success message, a summary table with one successfully deleted object, and a 'Delete' button.

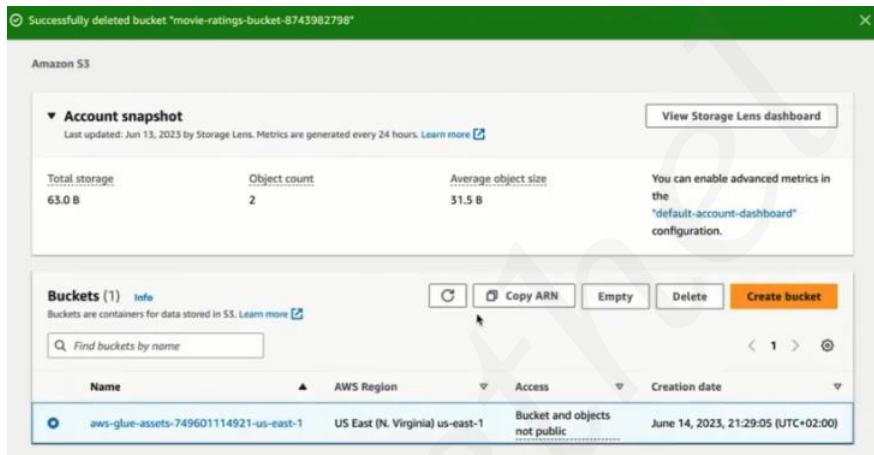
**196. SELECT THE BUCKET. CLICK DELETE.**

The screenshot shows the 'Buckets' list page with one bucket named 'recommender-system-8730872805'. The 'Delete' button for this bucket is highlighted.

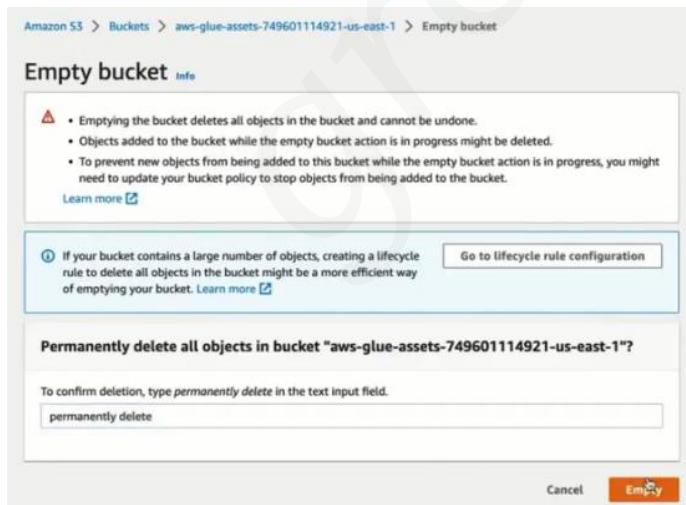
ENTER THE NAME OF THE BUCKET IN THE TEXT INPUT FIELD. CLICK DELETE BUCKET.



197. BUCKET DELETED. CLEAN THE ONE THAT WAS POPULATED AUTOMATICALLY WITH THE GLUE STUFF THAT WE DID. CLICK IT, THEN CHOOSE EMPTY.



TYPE PERMANENTLY DELETE. CLICK EMPTY.



**198. SELECT THE BUCKET AGAIN, CLICK DELETE.**

The screenshot shows the AWS S3 Buckets list. There is one bucket entry:

Name	AWS Region	Access	Creation date
aws-glue-assets-749601114921-us-east-1	US East (N. Virginia) us-east-1	Bucket and objects not public	June 14, 2023, 21:29:05 (UTC+02:00)

**ENTER THE NAME, THEN, DELETE BUCKET.**

The screenshot shows the 'Delete bucket' confirmation dialog. It includes a warning message, a text input field for the bucket name, and a 'Delete Bucket' button.

**Delete bucket** Info

**⚠** • Deleting a bucket cannot be undone.  
• Bucket names are unique. If you delete a bucket, another AWS user can use the name.  
• If this bucket is used with a Multi-Region Access Point in an external account, initiate failover before deleting the bucket.  
• If this bucket is used with an access point in an external account, the requests made through those access points will fail after you delete this bucket.

[Learn more](#)

**Delete bucket "aws-glue-assets-749601114921-us-east-1"?**

To confirm deletion, enter the name of the bucket in the text input field.

[Cancel](#) [Delete Bucket](#)

**SUCCESSFULLY DELETED BUCKET.**

The screenshot shows the Amazon S3 home page. A green banner at the top indicates the bucket was successfully deleted. The main content area features the S3 logo and a call-to-action to 'Create a bucket'.

**Storage**

**Amazon S3**

Store and retrieve any amount of data from anywhere

Amazon S3 is an object storage service that offers industry-leading scalability, data availability, security, and performance.

**Create a bucket**

Every object in S3 is stored in a bucket. To upload files and folders to S3, you'll need to create a bucket where the objects will be stored.

[Create bucket](#)

**199. TO DELETE THE ROLES, GO TO IAM. SELECT THE REDSHIFTFULLACCESSROLE AND GLUEFULLACCESSROLE. SELECT DELETE AT THE TOP.**

The screenshot shows the AWS IAM Roles page. On the left, there's a sidebar with options like Dashboard, Access management, Policies, Identity providers, and Account settings. The main area lists several roles:

Role Name	Description
AWSReservedSSO_AWSServiceCatalogAdminFullAccess_d1fe8e1088162bc9	Identity Provider: arn:aws:iam::749601114921:saml-provider/AWSSSC
AWSReservedSSO_AWSServiceCatalogEndUserAccess_7c60b01db8043761	Identity Provider: arn:aws:iam::749601114921:saml-provider/AWSSSC
AWSServiceRoleForRDS	AWS Service: rds (Service-Linked Role)
AWSServiceRoleForRedshift	AWS Service: redshift (Service-Linked Role) ✖ Deletion failed.
AWSServiceRoleForSSO	AWS Service: sso (Service-Linked Role) ✖ Deletion failed.
AWSServiceRoleForSupport	AWS Service: support (Service-Linked Role)
AWSServiceRoleForTrustedAdvisor	AWS Service: trustedadvisor (Service-Linked Role)
<b>GlueFullAccessRole</b>	AWS Service: glue Selected for deletion
rds-monitoring-role	AWS Service: monitoring.rds
<b>RedshiftFullAccessRole</b>	AWS Service: redshift Selected for deletion

At the bottom of the list, there's a "Roles Anywhere" section with an "Info" link and a "Manage" button.

**ENTER DELETE. THEN, DELETE.**

The screenshot shows a modal dialog titled "Delete 2 roles?". It asks if the user wants to delete two roles permanently, noting that this will also delete all their inline policies and any attached instance profiles. The roles listed are "GlueFullAccessRole" and "RedshiftFullAccessRole". Below the list, it says "Note: Recent activity usually appears within 4 hours. Data is stored for a maximum of 365 days, depending when your region began supporting this feature." and "This action cannot be undone." A text input field at the bottom contains the word "delete". At the bottom right are "Cancel" and "Delete" buttons.

**ROLES SUCCESSFULLY DELETED. CLEAN UP ALL DONE. CONGRATULATIONS!**

The screenshot shows the AWS IAM Roles page again. At the top, there are two green success messages: "Role RedshiftFullAccessRole created" and "Roles deleted.". Below this, the main "Roles (11) Info" section is visible, and at the bottom, there's a "Create role" button.