

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.

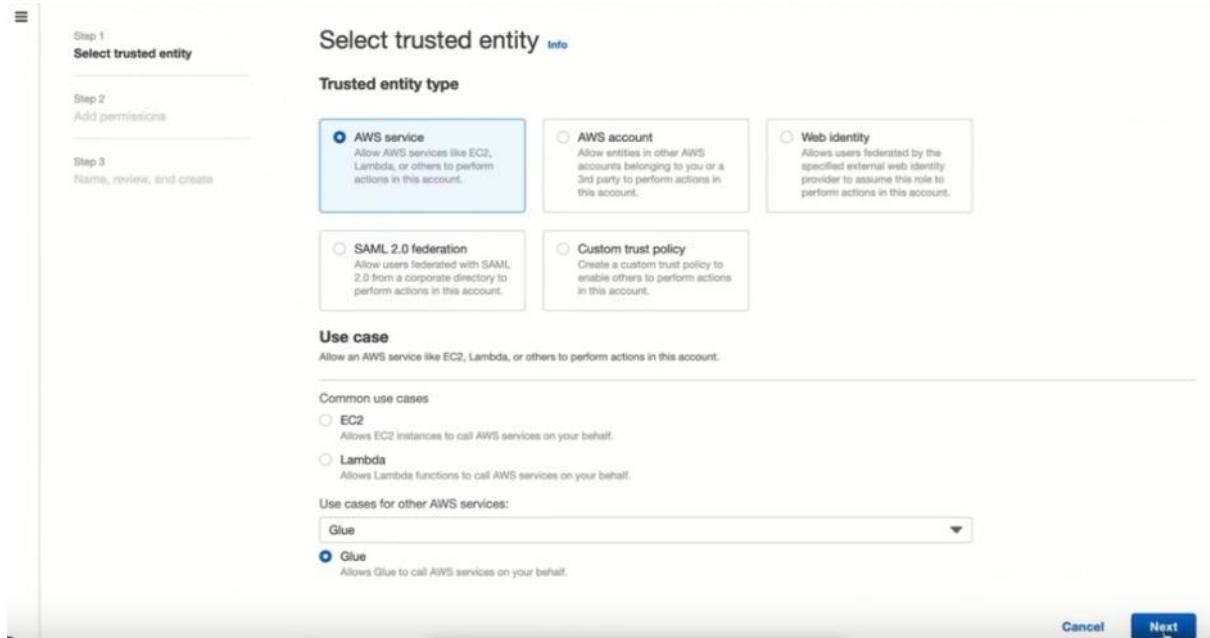
The screenshot shows the AWS IAM service page. On the left, there's a sidebar with categories like Services (9), Features (19), Resources (New), Blogs (1,565), Documentation (46,740), Knowledge Articles (30), Tutorials (2), Events (12), and Marketplace (493). The main area is titled 'Services' and shows the 'IAM' service card. The card has a star icon, the text 'Manage access to AWS resources', and a 'Top features' section with links to Groups, Users, Roles, Policies, and Access Analyzer. Below the card, there are three more cards: 'IAM Identity Center (successor to AWS Single Sign-On)', 'Resource Access Manager', and 'Serverless Application Repository'. A link 'See all 9 results ▶' is at the top right of the services list.

2. CLICK ROLES, THEN, CREATE ROLE.

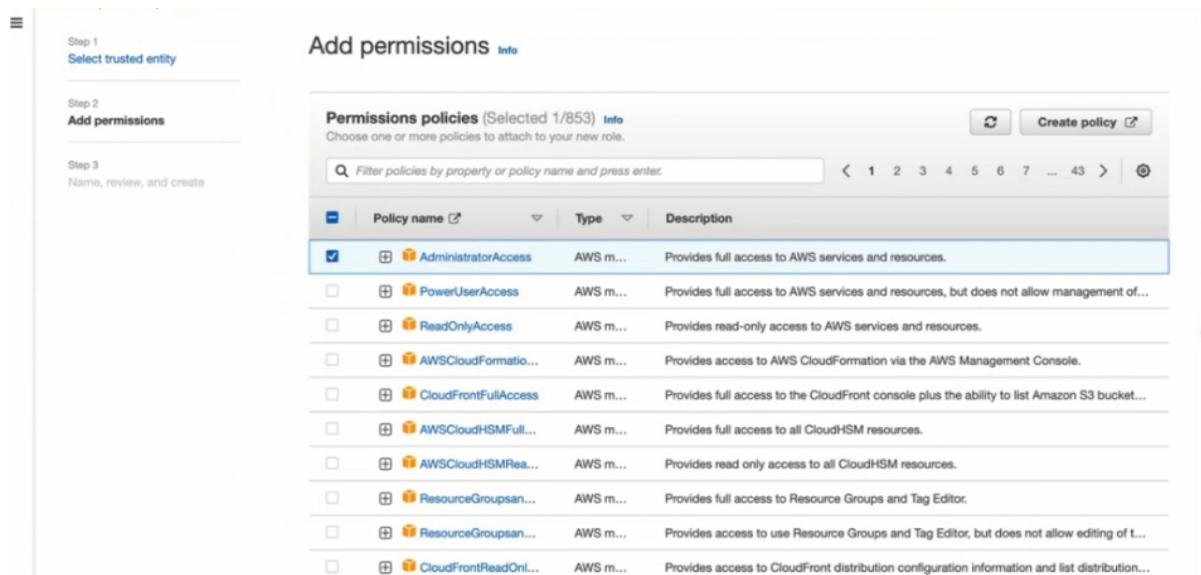
The screenshot shows the 'Roles' page under the IAM service. The left sidebar includes 'Identity and Access Management (IAM)', 'Dashboard', and sections for 'Access management' (User groups, Users, Roles, Policies, Identity providers, Account settings), 'Access reports' (Access analyzer, Archive rules, Analyzers, Settings), and a search bar. The main area shows a table of roles with columns for 'Role name' and 'Trusted entities'. The table lists several roles, each with an 'Edit' button, a 'Delete' button, and a note about deletion status (e.g., 'Deletion failed'). The first role listed is 'AWSReservedSSO_AWSAdministratorAccess_d6359d25c2eb8818'.

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) Deletion failed.
AWSServiceRoleForSSO	AWS Service: sso (Service-Linked Role) Deletion failed.
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.

The screenshot shows the final step of creating a new AWS IAM role. The left sidebar lists three steps: Step 1 (Select trusted entity), Step 2 (Add permissions), and Step 3 (Name, review, and create). The main area is titled "Name, review, and create".

Role details

Role name: GlueFullAccessRole

Description: Allows Glue to call AWS services on your behalf.

Step 1: Select trusted entities

```

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  ]

```

Permissions

- 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.

Buttons: 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', 'Access management', 'Roles', 'Access reports', and 'Service control policies (SCPs)'. The main 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.' Below the table, there are buttons for 'View role', 'Delete', and 'Create role'.

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_d1fe8a1088162bc9	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 page. The search bar at the top has 'S3' typed into it. On the left, there's a sidebar with sections like 'Hard ETL', 'Services (7)', 'Features (19)', 'Resources New', 'Visual (1)', and a '+' button. The main area lists services: 'Services (7)' (with 'Bucket' highlighted), 'Features (19)', 'Amazon S3 File Gateway'. A context menu is open over the 'Open Link in New Tab' option for the 'Bucket' item. The menu includes options like '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'.

CLICK CREATE BUCKET.

The screenshot shows the Amazon S3 landing page. At the top right, there is a large orange button labeled "Create a bucket". To the left of the button, there is a brief description: "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." Below this, there is a smaller orange button labeled "Create bucket".

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.

How it works

aws Introduction to Amazon S3 Copy link

Pricing

With S3, there are no minimum fees. You only pay for what you use. Prices are based on the location of your S3 bucket.

Estimate your monthly bill using the AWS Simple Monthly Calculator

View 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 current step is "General configuration".

Bucket name: recommender-system-8730872805

Bucket name must be unique within the global namespace and follow the bucket naming rules. See rules for bucket naming.

AWS Region: US East (N. Virginia) us-east-1

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

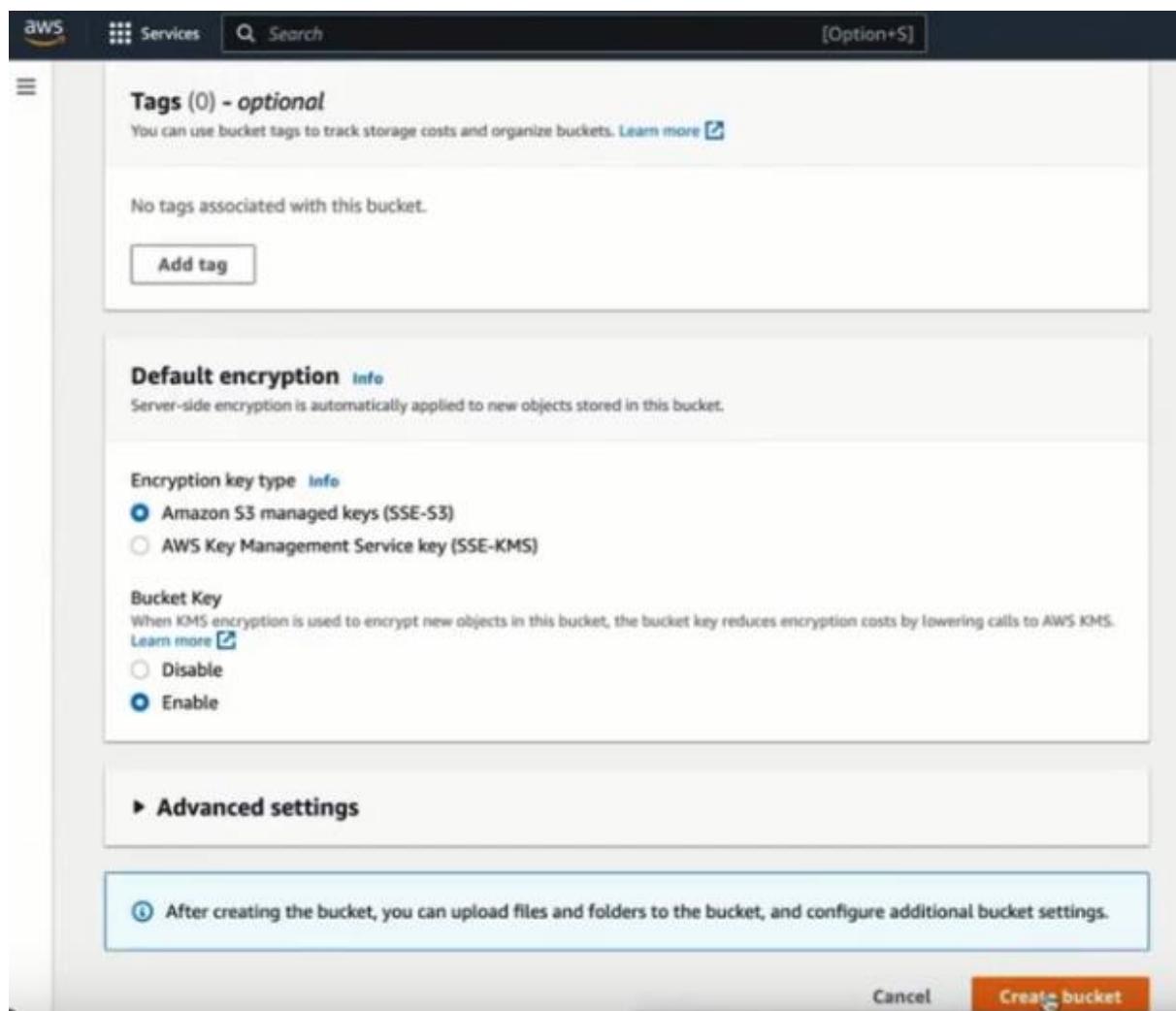
Object Ownership: Info

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

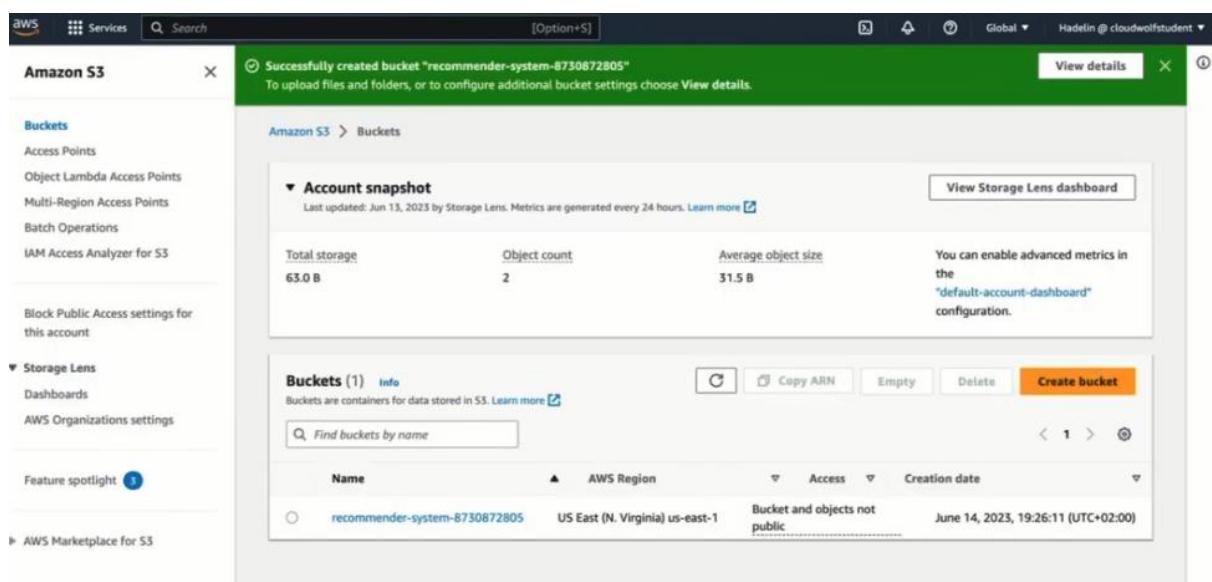
ACLs disabled (recommended)
All objects in this bucket are owned by this account. Access to this bucket and its objects is specified using only policies.

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

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 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. Below the tabs are buttons for 'Actions' (with 'Upload' highlighted), 'Copy S3 URI', 'Copy URL', 'Download', 'Open', 'Delete', and 'Create folder'. A search bar says 'Find objects by prefix'. A table header includes '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.' An 'Upload' button is located at the bottom right of the table area.

12. CLICK ADD FILES.

The screenshot shows the 'Upload' page within the Amazon S3 console. At the top, it displays the path 'Amazon S3 > Buckets > recommender-system-8730872805 > Upload'. The main area is titled 'Upload' and contains instructions: 'Add the files and folders you want to upload to S3. To upload a file larger than 160GB, use the AWS CLI, AWS SDK or Amazon S3 REST API. Learn more'. Below this is a large dashed rectangular area with the text 'Drag and drop files and folders you want to upload here, or choose Add files or Add folder.' A 'Files and folders (0)' section shows a table with columns 'Name', 'Folder', 'Type', and 'Size'. A 'Remove' button is available for each row. Buttons for 'Add files' and 'Add folder' are also present. A search bar 'Find by name' is at the top of the table. 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.' The final section is 'Destination', which shows '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%)	0 files, 0 B (0%)

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 (1)

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) 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 with sections for 'Hard ETL' (including 'Unsaved jobs' with 1 item), 'Visual' (with 1 item), and other links like 'Documentation', 'Knowledge Articles', 'Events', and 'Marketplace'. The main content area is titled 'Services' and shows a list of services. The first item in the list is 'AWS Glue', which is highlighted with a blue border. Below it are 'AWS Glue DataBrew', 'AWS Lake Formation', and 'Athena'. Each service entry includes a small icon, a name, a star rating, and a brief description. To the right of the list, there is a 'Data preview' section and a 'View' button. At the bottom right of the main content area, there is a 'Actions' dropdown menu.

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 contains a message: 'A database is a set of associated table definitions, organized into a logical group.' Below this is a table with columns: Name, Description, Location URI, and Created on (UTC). The table currently displays 'No resources' and 'No resources to display.' At the top right of the main content area, there are buttons for 'Edit', 'Delete', and 'Add database' (which is highlighted in orange).

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.

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

Add database

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 Add table

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).

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://bucket/prefix/object

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 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)

Find bucket

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

Cancel Choose

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
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).

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 X 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

Getting started Step 1 Set crawler properties

ETL jobs Step 2 Choose data sources and classifiers

Notebooks Step 3 Configure security settings

Job run monitoring Step 4 Set output and scheduling

Data Catalog tables Step 5 Review and create

Data connections

Workflows (orchestration)

Databases Tables

Stream schema registries Schemas

Connections

Crawlers Classifiers

Catalog settings

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.

AWS Glue > Crawlers > Add crawler

Configure security settings

IAM role [Info](#)

Existing IAM role

[View](#)

GlueFullAccessRole
Allows Glue to call AWS services on your behalf.

Cancel Previous Next

AWS Glue > Crawlers > Add crawler

Configure security settings

IAM role [Info](#)

Existing IAM role

[View](#)

Only IAM roles created by the AWS Glue console and have the prefix "AWSGlueServiceRole-" can be updated.

Lake Formation configuration - optional

Allow the crawler to use Lake Formation credentials for crawling the data source. [Learn more](#)

Use Lake Formation credentials for crawling S3 data source
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.

► Security configuration - optional

Enable at-rest encryption with a security configuration.

Cancel Previous Next

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

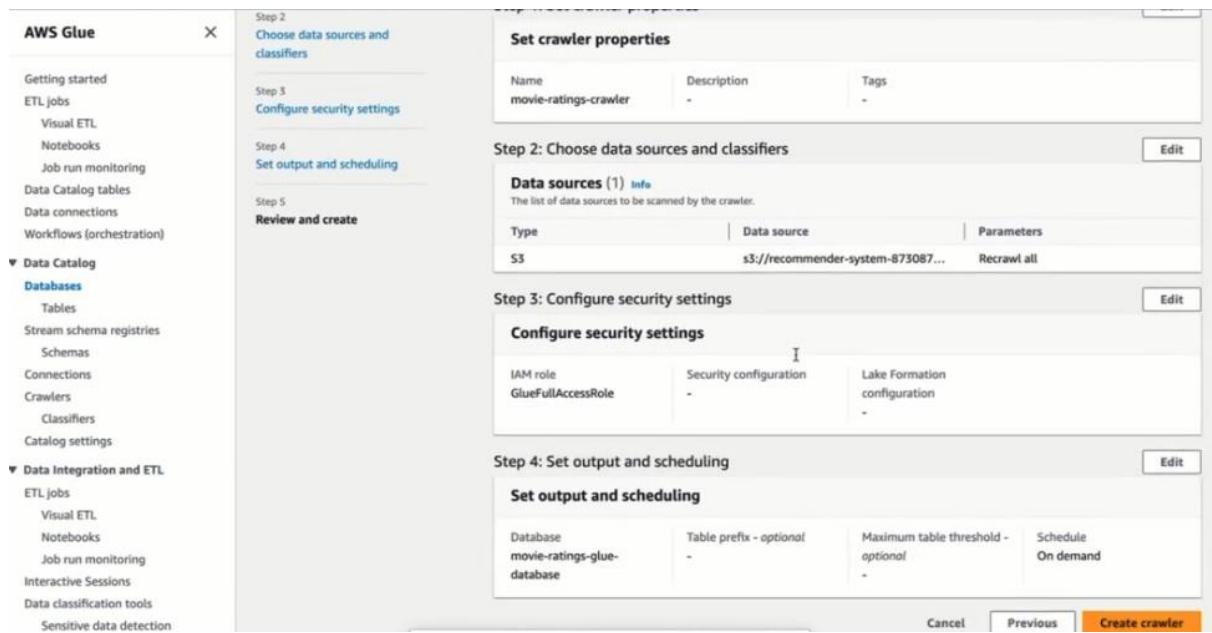
The screenshot shows the AWS Glue interface for creating a new crawler. On the left, a sidebar lists various options like 'Getting started', 'ETL jobs', 'Data Catalog tables', 'Data connections', and 'Workflows (orchestration)'. The main panel is titled 'Set output and scheduling' and contains the following sections:

- Output configuration**:
 - Target database**: A dropdown menu set to 'movie-ratings-glue-database'.
 - Table name prefix - optional**: An input field containing 'movie-name-prefix->openuid'.
 - Maximum table threshold - optional**: A text input field with the placeholder 'Type a number greater than 0'.
- Crawler schedule**:
 - A note: 'You can define a time-based schedule for your crawlers and jobs in AWS Glue. The definition of these schedules uses the Unix-like cron syntax.' with a 'Learn more' link.
 - Frequency**: A dropdown menu set to 'On demand'.

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

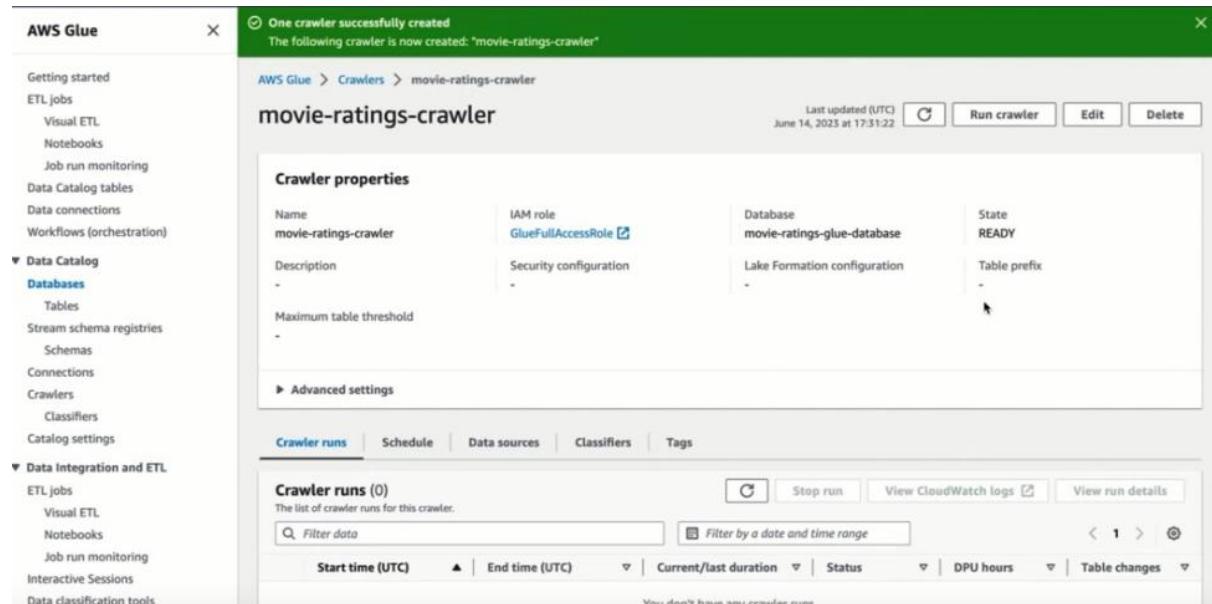
This screenshot shows the same 'Set output and scheduling' step as the previous one, but with the 'Next' button at the bottom highlighted in orange. The interface elements are identical to the first screenshot, including the sidebar navigation and the configuration fields for target database, table prefix, and crawler frequency.

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.

The screenshot shows the AWS Glue interface with the 'Crawlers' section selected. A success message at the top states: 'One crawler successfully created' and 'The following crawler is now created: "movie-ratings-crawler"'. The crawler list table has one item: 'movie-ratings-cr...' with a status of 'Ready'. The 'Run' button is highlighted in orange.

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.

The screenshot shows the AWS Glue interface with the 'Crawlers' section selected. A progress message at the top says: 'Starting crawler' and 'Attempting to start run crawler "movie-ratings-crawler"'. The crawler list table shows 'movie-ratings-cr...' with a status of 'Ready'. The 'Run' button is highlighted in orange.

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.

The screenshot shows the AWS Glue interface with the 'Crawlers' section selected. A progress message at the top says: 'Crawler successfully starting' and 'The following crawler is now starting: "movie-ratings-crawler"'. The crawler list table shows 'movie-ratings-cr...' with a status of 'Running'. The 'Run' button is highlighted in orange.

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 the 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 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 titled "Crawlers (1/1) Info" shows one crawler: "movie-ratings-cr..." with a state of "Stopping". The table includes columns for Name, State, Schedule, Last run, Last run time, Log, and Table changes. The last row shows "1 created".

The screenshot shows the AWS Glue Crawlers page again. The green banner now displays "Crawler successfully starting" and "The following crawler is now starting: 'movie-ratings-crawler'". The crawler status has changed to "Ready". The table below shows the crawler in a "Succeeded" state with the timestamp "June 14, 2023 a...". The table columns are identical to the previous screenshot.

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 for Getting started, ETL jobs, Visual ETL, Notebooks, Job run monitoring, Data Catalog tables (selected), Data connections, Workflows (orchestration), Data Catalog (expanded), Databases, Tables (selected), Stream schema registries, Schemas, Connections, Crawlers, and Classifiers. The main content area shows the "Tables" section with the sub-header: "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 titled "Tables (1)" shows one table: "recommendation_system_8730872805". The table includes columns for Name, Database, Location, Classification, Deprecated, and View. The last column shows "Table" with a link.

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 page. On the left, there's a sidebar with navigation links for AWS Glue, Data Catalog tables, and Data Integration and ETL. The main area has tabs for 'Table overview' (selected), 'Data quality', and 'New'. Under 'Table details', the table name is 'recommender_system_8730872805', located is 's3://recommender-system-8730872805/', and the input format is 'org.apache.hadoop.mapred.TextInputFormat'. Under 'Advanced properties', the database is 'movie-ratings-glue-database', classification is 'CSV', and last updated is 'June 14, 2023 at 17:33:30'. The 'Schema' tab is selected, showing a table with four columns: #, Column name, Data type, Partition key, and Comment. The columns are: 1 (userid, bigint, -), 2 (movieid, bigint, -), 3 (rating, double, -), and 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 categorized under 'Services' and 'Features'. Under 'Services', 'Athena' is listed as a 'Serverless interactive analytics service'. Under 'Features', there are sections for '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. In the top left, it says "Amazon Athena > Query editor". Below that is a navigation bar with tabs: "Editor" (which is selected), "Recent queries", "Saved queries", and "Settings". To the right of the tabs is a "Workgroup" dropdown set to "primary". On the left side, there's a sidebar titled "Data" with sections for "Data source" (set to "AwsDataCatalog"), "Database" (set to "movie-ratings-glue-database"), and "Tables and views" (with a "Create" button). Under "Tables and views", there's a search bar and a table section showing "Tables (1)" and "Views (0)". The main area is titled "Query 1" and contains a SQL editor with the placeholder "SQL Ln 1, Col 1" and buttons for "Run", "Explain", "Cancel", "Clear", and "Create". Below the SQL editor is a "Query results" tab and a "Results" tab. The "Results" tab has buttons for "Copy" and "Download results".

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 shows "Tables (1)" with the entry "recommender_system_8730872805". A context menu is open over this table entry, listing options: "Run Query", "Preview Table" (which is highlighted in blue), "Generate table DDL", "Insert", "Insert into editor", "Manage", "Delete table", "View properties", and "View in Glue". The rest of the interface is identical to the first screenshot, including the "Query 1" SQL editor and the "Results" tab below.

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 'Tables and views' section with 'Tables (1)' and 'Views (0)'. The 'Tables' section shows 'recommender_system_8730872805'. The bottom section 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 are presented in a table titled 'Results (10)' with columns: '#', 'userid', 'movieid', 'rating', and 'timestamp'. The results are as follows:

#	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

This is a detailed view of the 'Results (10)' table from the previous screenshot. The table has columns: '#', 'userid', 'movieid', 'rating', and 'timestamp'. The data is identical to the one shown in the previous screenshot, listing 10 rows where user 1 has rated various movies with their respective ratings and timestamps.

#	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.

Name	Description	Database	Classification
recommender_system_8730872805	-	movie-ratings-glue-database	csv
Location	Connection	Deprecated	Last updated
s3://recommender-system-8730872805/	-	-	June 14, 2023 at 17:33:30
Input format	Output format	Serde serialization lib	
org.apache.hadoop.mapred.TextInputFormat	org.apache.hadoop.hive.serde2.lazy.LazyKeyTextOutputFormat	org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe	

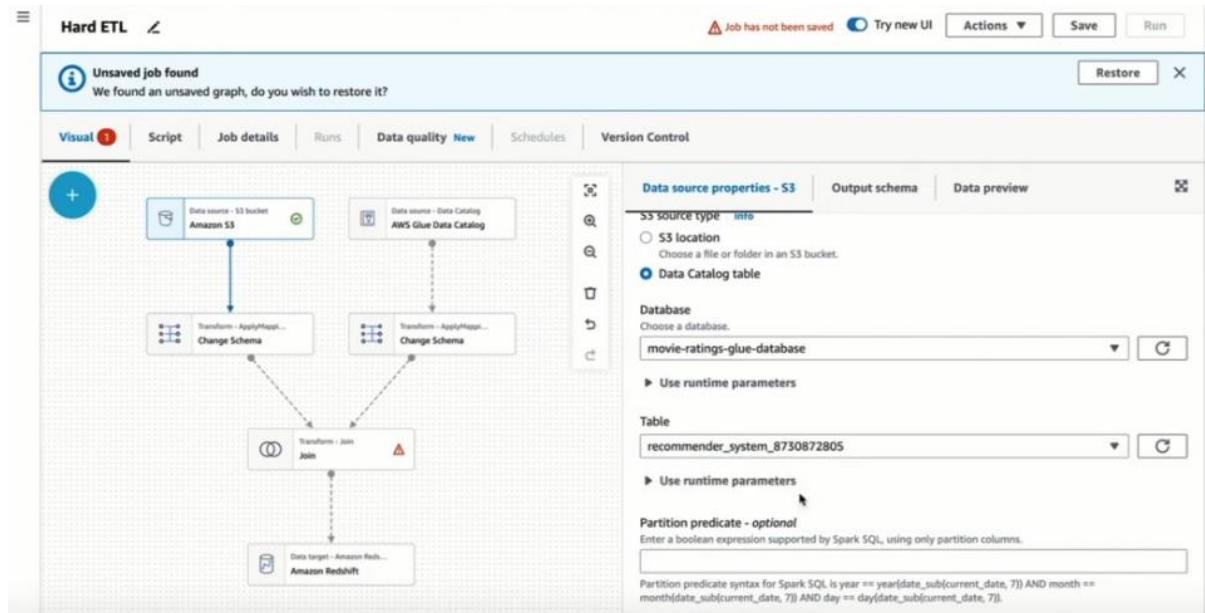
Schema (4)
View and manage the table schema.
Edit schema as JSON | Edit schema

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.

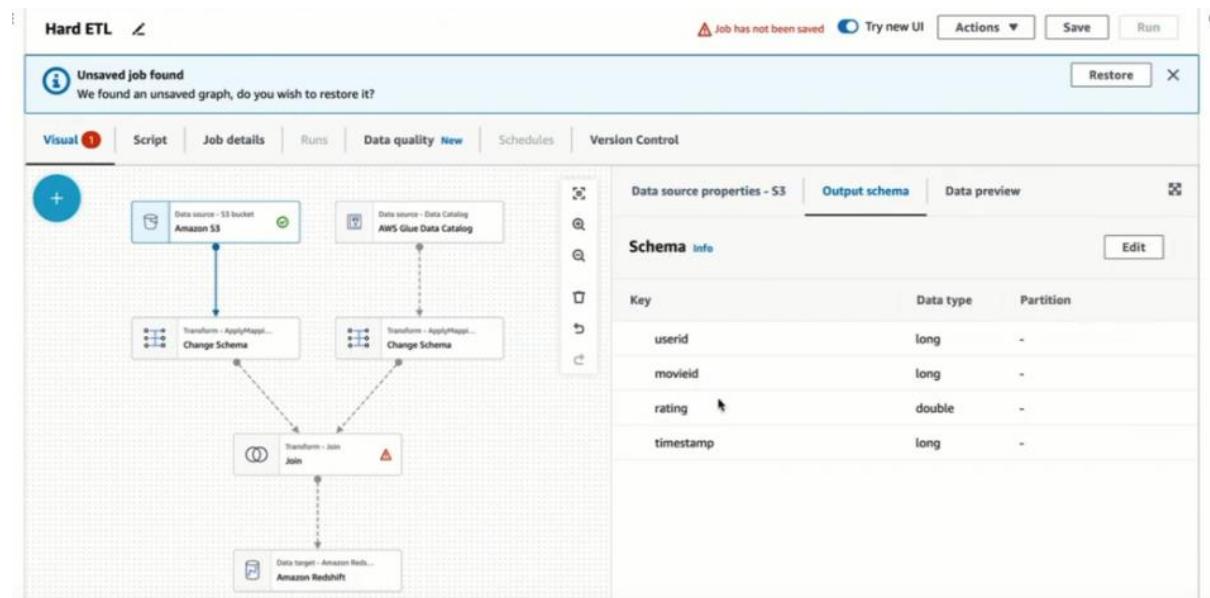
IN THE S3 SOURCE TYPE, CLICK DATA CATALOG TABLE.

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

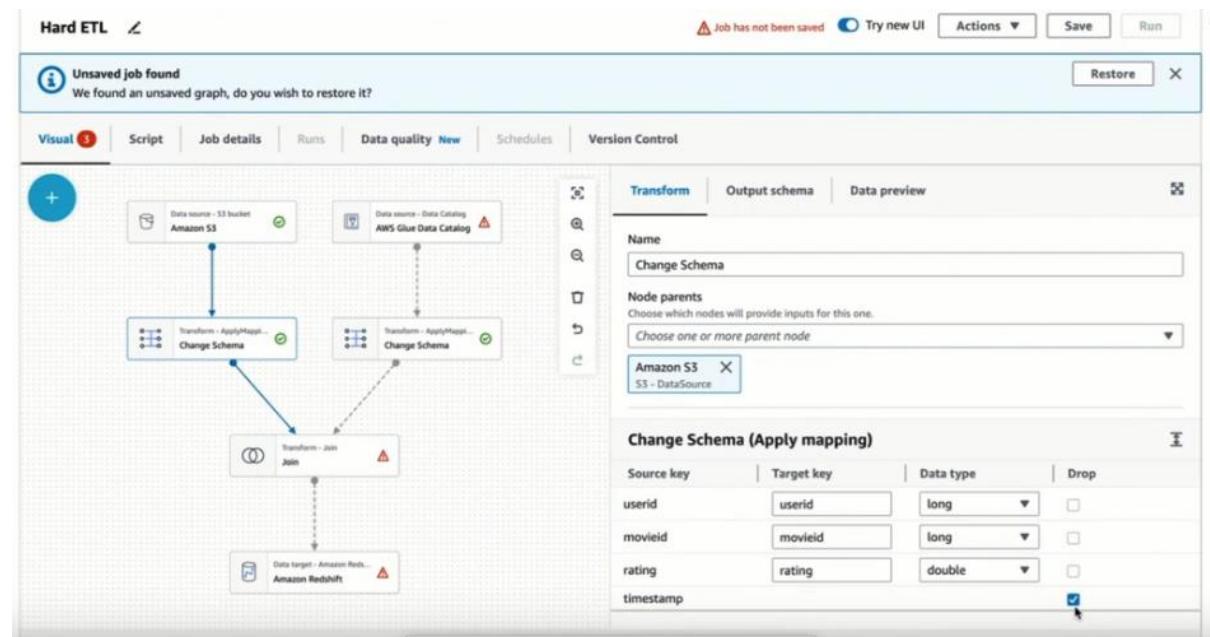
IN THE TABLE, CLICK THE TABLE THAT WAS CREATED.



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 is titled 'Services (14)' and shows a list of services. The first item in the list is 'RDS' (Managed Relational Database Service), which is highlighted with a blue border. Below it are other services: AWS FIS, Database Migration Service, and Amazon OpenSearch Service. At the bottom of the list, there is a section titled 'Features' with a link to 'See all 29 results'.

42. CLICK CREATE DATABASE.

The screenshot shows the Amazon RDS 'Create database' dialog. The dialog box is centered over the main RDS dashboard. It contains a message encouraging the use of the Multi-AZ deployment option for MySQL and PostgreSQL, mentioning improved commit latencies and faster failover. Below the message is a prominent orange 'Create database' button. To the right of the button, there is a link to 'Or, Restore Multi-AZ DB Cluster from Snapshot'. The background of the dialog is white, while the message and button are set against a light gray background.

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

> 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](#)

View credential details X

RDS > Databases

Consider creating a Blue/Green Deployment to minimize downtime during upgrades
You may want to consider using Amazon RDS Blue/Green Deployments and minimize your downtime during upgrades. A Blue/Green Deployment provides a staging environment for changes to production databases. [RDS User Guide](#) [Aurora User Guide](#)

Databases (1)

Group resources Modify Actions ▾ Restore from S3 Create database

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. On the left, a sidebar lists various RDS-related options like Dashboard, Databases, Query Editor, etc. The main area displays a success message: 'Successfully created database customer-features-rds-db-instance'. It also includes a tooltip about Blue/Green Deployments and a link to the RDS User Guide. The 'Databases' table lists one 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

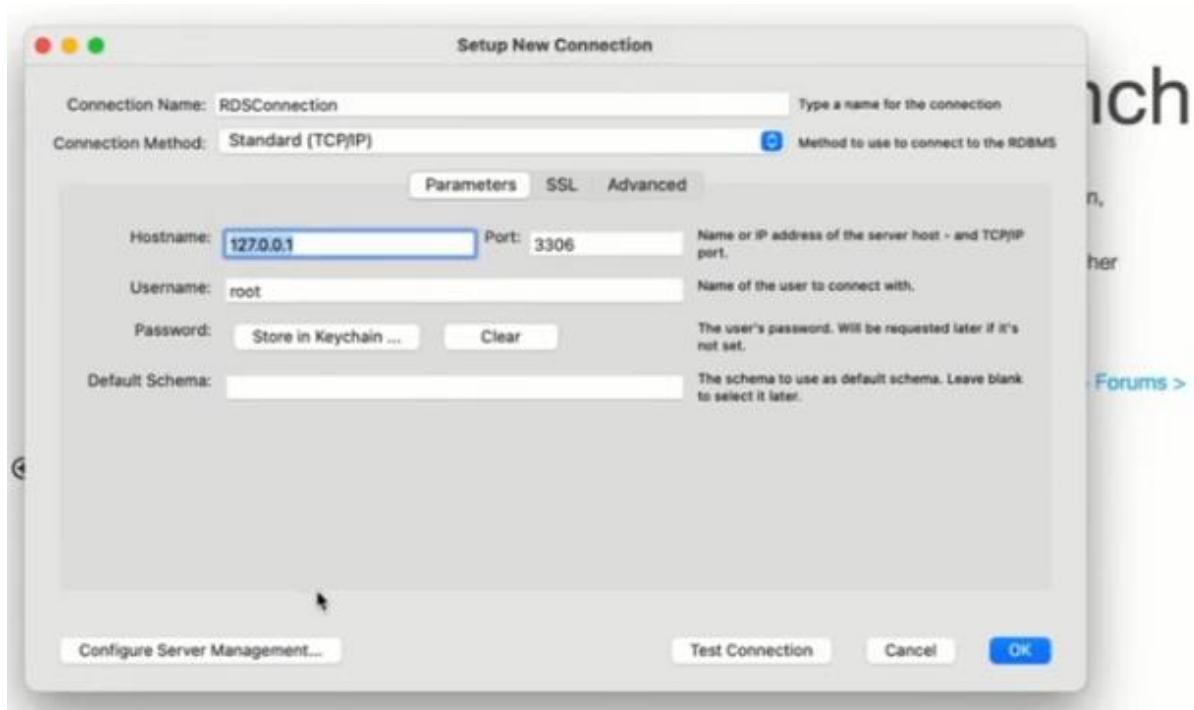
◀ MySQL Workbench

The screenshot shows the MySQL Workbench download page for version 8.0.36. At the top, there are tabs for 'General Availability (GA) Releases' (selected), 'Archives', and a help icon. Below this, it says 'MySQL Workbench 8.0.36'. A dropdown menu 'Select Operating System:' is set to 'Microsoft Windows'. The 'Recommended Download:' section features a large button for the 'MySQL Installer for Windows' (Windows x86, 32 & 64-bit, MySQL Installer MSI). To the right is an image of the Windows logo. Below the button, it says 'All MySQL Products. For All Windows Platforms. In One Package.' and 'Starting with MySQL 5.6 the MySQL Installer package replaces the standalone MSI packages.' Further down, there's a 'Go to Download Page >' button. The 'Other Downloads:' section lists the 'Windows (x86, 64-bit), MSI Installer' file (version 8.0.36, 42.0M, MD5: 2156fe0cb6f5ed83908e4636ba86390a | Signature). At the bottom, a note encourages users to verify downloads using MD5 checksums and GnuPG signatures.

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.



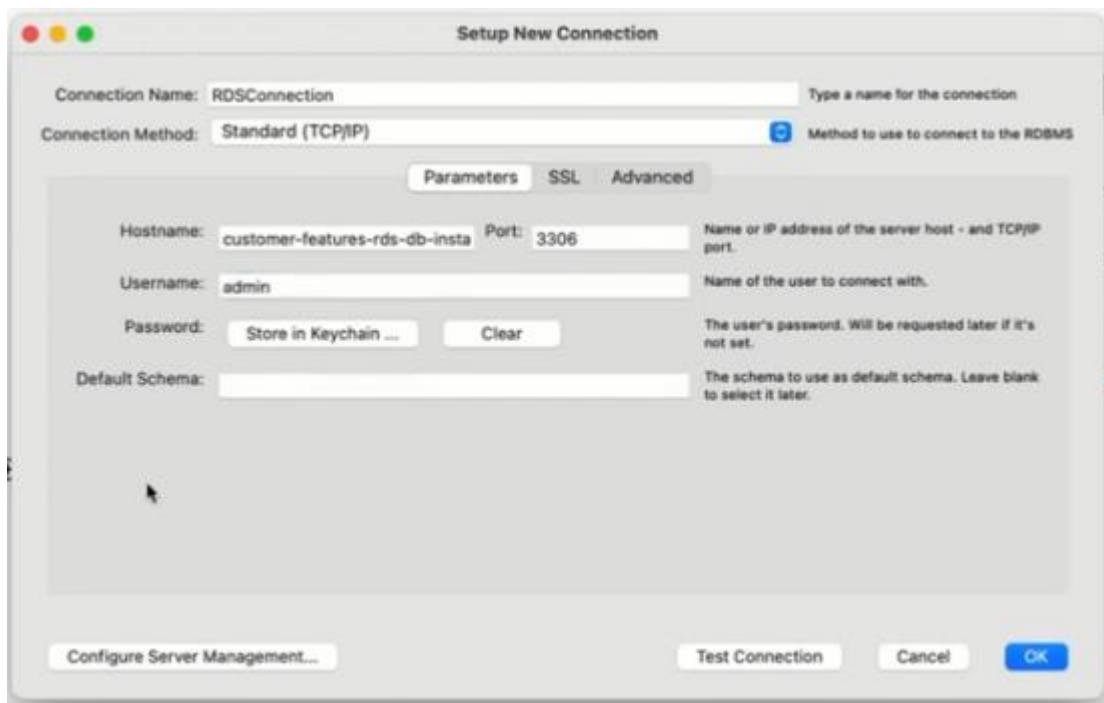
COPY THE ENDPOINT DETAILS. GO BACK TO THE MYSQL WORKBENCH, THEN, PASTE IT TO THE HOSTNAME. THE PORT 3306 IS THE PORT OF MY MYSQL, SO, IT WILL REMAIN THE SAME.

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:

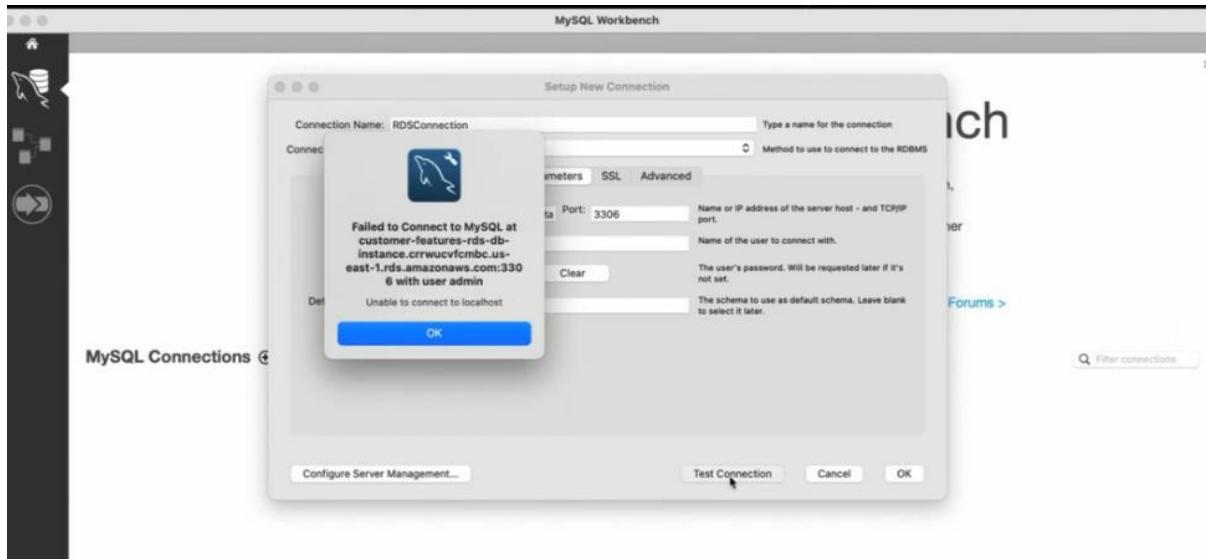
Endpoint & port	Networking	Security
Endpoint customer-features-rds-db-instance.crrwucvfcmbc.us-east-1.rds.amazonaws.com	Availability Zone us-east-1c	VPC security groups default (sg-00757adc39bf9625e) Active
Port 3306	VPC vpc-0041e736a2d14cfa2	Publicly accessible No
	Subnet group default-vpc-0041e736a2d14cfa2	Certificate authority Info rds-ca-2019
	Subnets subnet-061069d72790ee8be subnet-0ab745a585106a535 subnet-053605a78fadcc9a 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 navigation options like Dashboard, Databases, Query Editor, etc. The main content area has tabs for Connectivity & security, Monitoring, Logs & events, Configuration, Maintenance & backups, and Tags. The Connectivity & security tab is selected. It displays configuration details in three columns: Endpoint & port, Networking, and Security. In the Security section, the "Publicly accessible" field is set to "No". Other visible details include the VPC security group "default (sg-00757adc39bf9625e)" and active status. The Networking section lists the VPC "vpc-0041e736a2d14cfa2" and Subnet group "default-vpc-0041e736a2d14cfa2". The Subnets section lists several subnet IDs. The Network type is IPv4. The DB instance certificate expiration date is listed as 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-tabs 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 same database instance. On the left, there is a sidebar with links like Dashboard, Databases (which is 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, and Recommendations (with a notification count of 1). The main area is titled "Connectivity" and contains sections for "Network type", "DB subnet group", "Security group", and "Certificate authority". At the bottom, there is a section titled "Additional configuration" which includes a "Public access" field. The "Public access" field is currently set to "No".

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

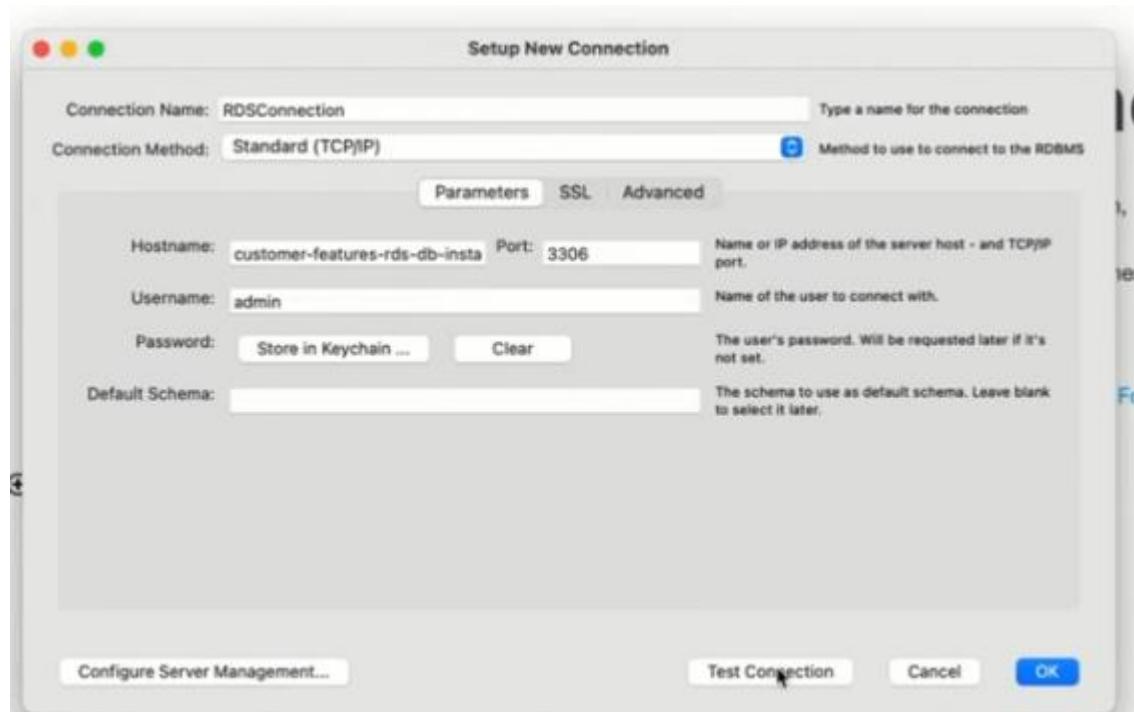
The modifications in this request and any pending modifications will be asynchronously applied as soon as possible, regardless of the maintenance window setting for this database instance.

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 main content area shows a table of databases, with one row selected 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 sidebar menu for 'Amazon RDS' is visible, showing options like Dashboard, Databases, Query Editor, and Performance insights. The main search results are displayed under the heading 'Search results for "EC2"' with the sub-heading 'Services (12)'. The results include:

- EC2** ★ Virtual Servers in the Cloud
- EC2 Image Builder ★ A managed service to automate build, customize and deploy OS images
- Amazon Inspector ★ Continual vulnerability management at scale
- AWS Firewall Manager ★ Central management of firewall rules

Below the search results, there is a message: 'Try searching with longer queries for more relevant results'.

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

The screenshot shows the AWS EC2 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: 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.

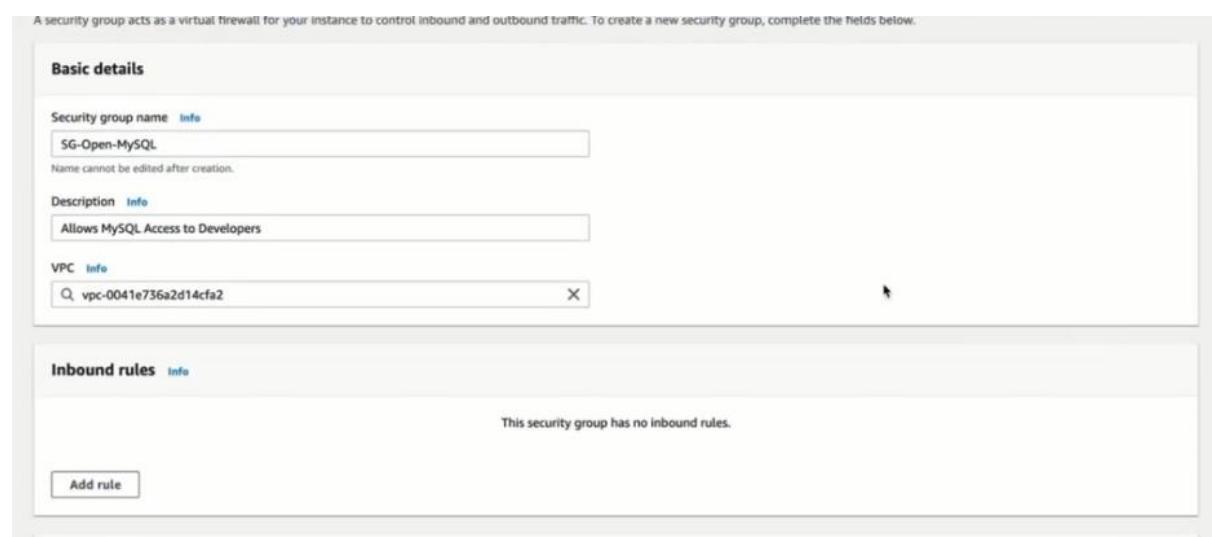
Description [Info](#)
Allows MySQL Access to Developers

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](#)

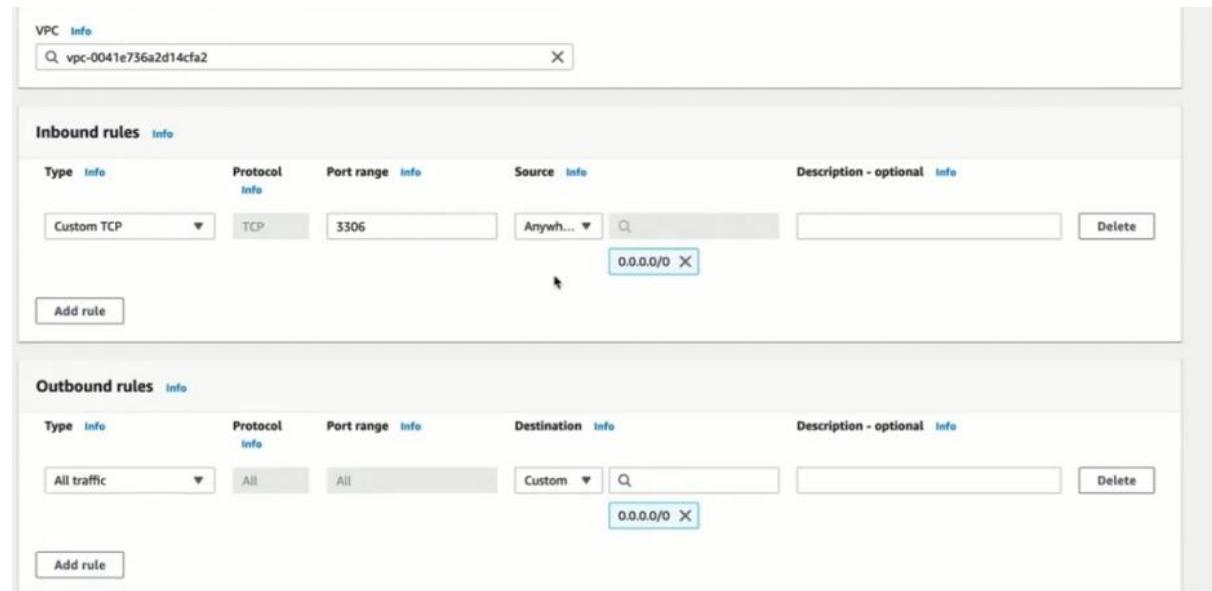
Type Info	Protocol Info	Port range Info	Source Info	Description - optional Info
Custom TCP	TCP	3306	Anywh... ▾	<input type="text"/> 0.0.0.0/0 X

Add rule

Outbound rules [Info](#)

Type Info	Protocol Info	Port range Info	Destination Info	Description - optional Info
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

Add rule

Tags - optional

No tags associated with the resource.

Add new tag

You can add up to 50 more tags

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

Inbound rules

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

Outbound rules

Name	Security group rule...	IP version	Type	Protocol	Port rang...
-	0.0.0.0/0	MySQL/Aurora	TCP	3306	

You can now check network connectivity with Reachability Analyzer

Inbound rules (1/1)

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 'customer-features-rds-db-instance' summary page. The 'Modify' button is located in the top right corner of the main content area. The summary table includes columns for DB identifier, CPU usage, Status, Class, Role, Current activity, Engine, and Region & AZ. Below the summary is a navigation bar with tabs: Connectivity & security, Monitoring, Logs & events, Configuration, Maintenance & backups, and Tags. The 'Connectivity & security' tab is selected.

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' configuration page. Under 'Network type', 'IPv4' is selected. Under 'DB subnet group', 'default-vpc-0041e736a2d14cfa2' is chosen. Under 'Security group', 'Choose security groups' is selected, showing 'default' and 'SG-Open-MySQL' as options. A note at the bottom right says 'Changes made to an Amazon database will take effect after provision.' A 'Continue' button is visible at the bottom right of the configuration panel.

62. SCROLL DOWN AND CLICK CONTINUE.

The screenshot shows the 'Modify DB instance' configuration page. It includes sections for CloudWatch Logs (Audit log, Error log, General log, Slow query log), IAM role (RDS service-linked role), Maintenance (Auto minor version upgrade, Enable auto minor version upgrade checked), DB instance maintenance window (Start day: Saturday, Start time: 08:19 UTC, Duration: 0.5 hours), and Deletion protection (Enable deletion protection unchecked). The 'Continue' button is located at the bottom right of the configuration panel.

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 a 'Summary of modifications' section where it says 'You are about to submit the following modifications. Only values that will change are displayed. Carefully verify your changes and click Modify DB instance.' It lists a single modification: 'Security group' set to 'default' changing to 'default, SG-Open-MySQL'. Below this is a 'Schedule modifications' section with two options: 'Apply during the next scheduled maintenance window' (selected) and 'Apply immediately'. A note under 'Apply immediately' states that modifications will be applied as soon as possible, regardless of the maintenance window setting. At the bottom right are 'Cancel', 'Back', and a prominent orange 'Modify DB instance' button.

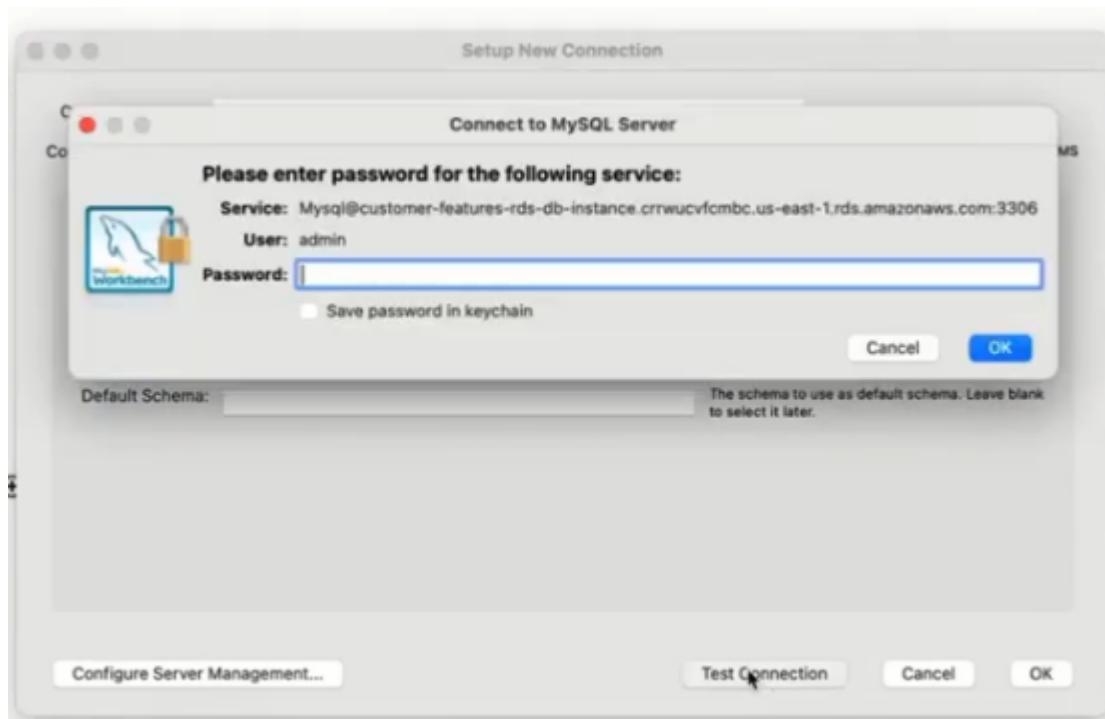
64. SUCCESSFULLY MODIFIED.

The screenshot shows the 'Databases' page with one entry: 'customer-features-rds-db-instance'. A green success message at the top says 'Successfully modified instance customer-features-rds-db-instance'. Below the message is a informational box about Blue/Green Deployments. The main table has columns for DB identifier, Status, Role, Engine, Region & AZ, Size, and Actions. The instance listed is 'customer-features-rds-db-instance' with status 'Available', engine 'MySQL Community', region '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.

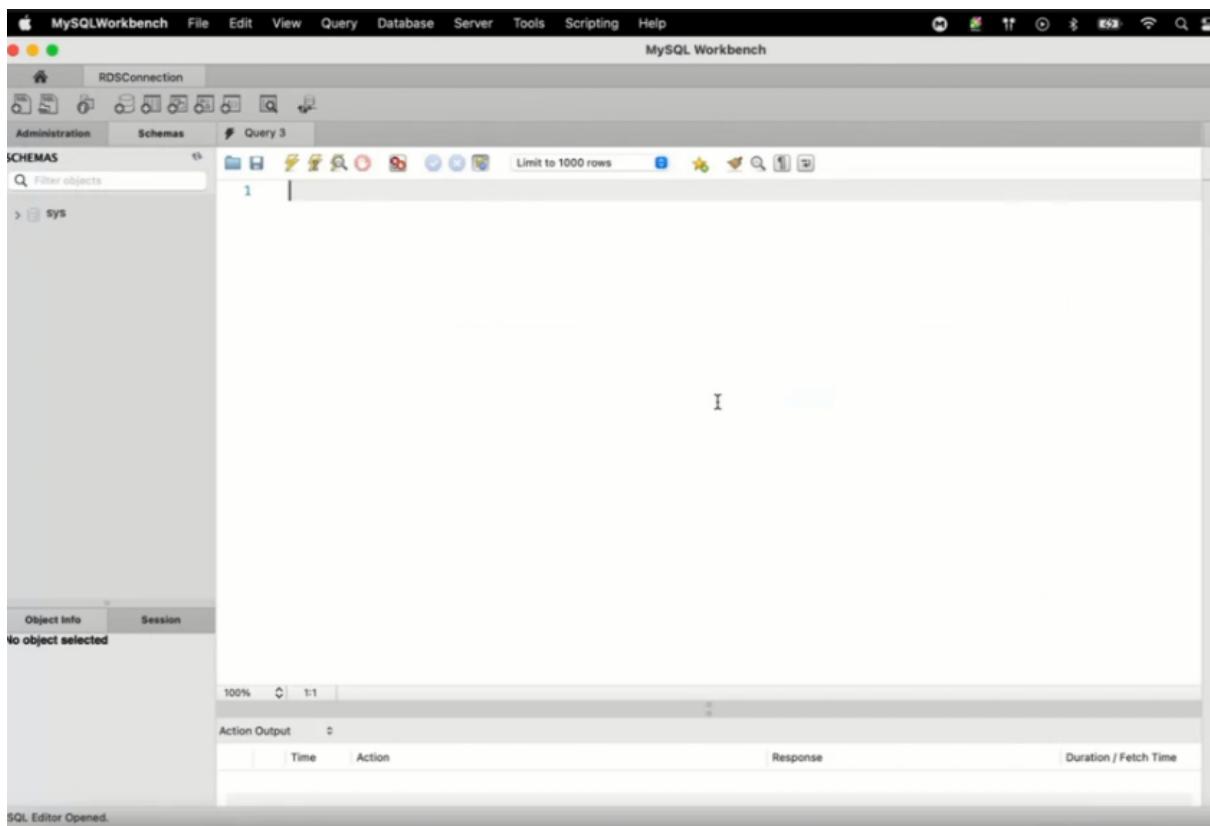




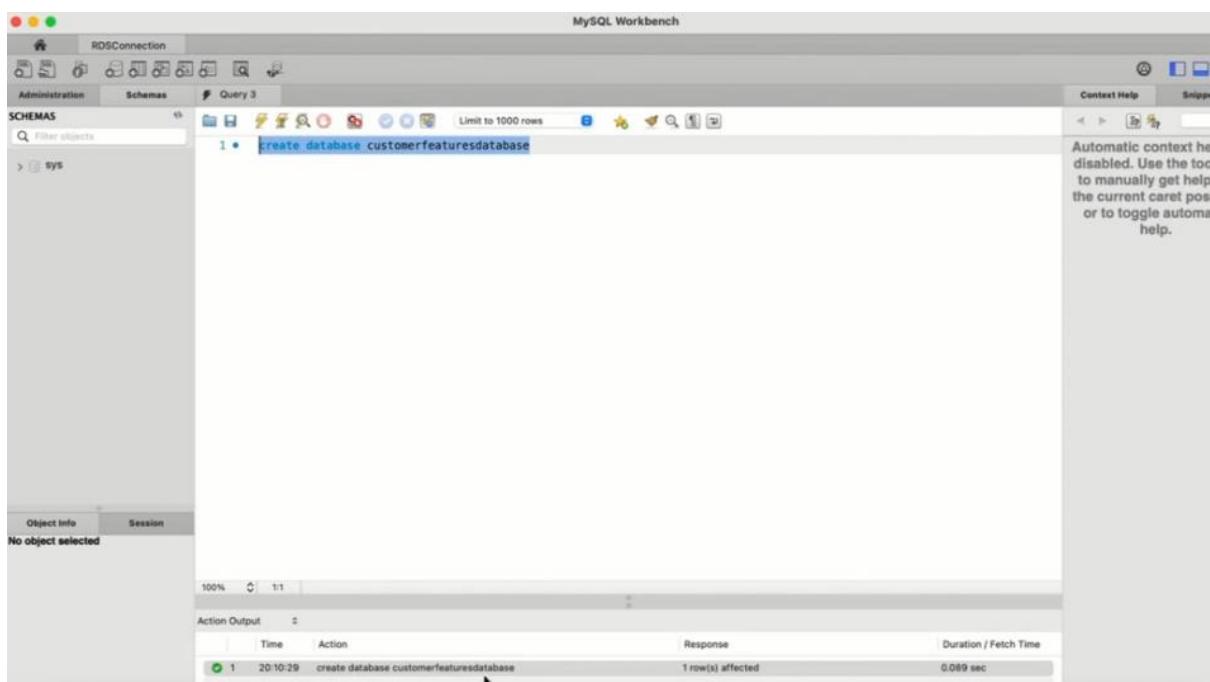
66. THE CONNECTION IS NOW ESTABLISHED, DOUBLE CLICK THE RDS CONNECTION. IT IS NOW OPENING THE SQL EDITOR.

A screenshot of the MySQL Workbench interface. On the left, there's a sidebar with 'MySQL Connections' containing 'RDSConnection'. In the center, a modal window titled 'Opening SQL Editor' is displayed, showing a progress bar and the text 'An SQL editor instance for "RDSConnection" is opening and should be available in a moment. Please stand by...'. There are buttons for 'Cancel' and 'Filter connections'. The background shows the main MySQL Workbench workspace with various tabs and sections like 'Browse Document' and 'Discuss on the Forums'.

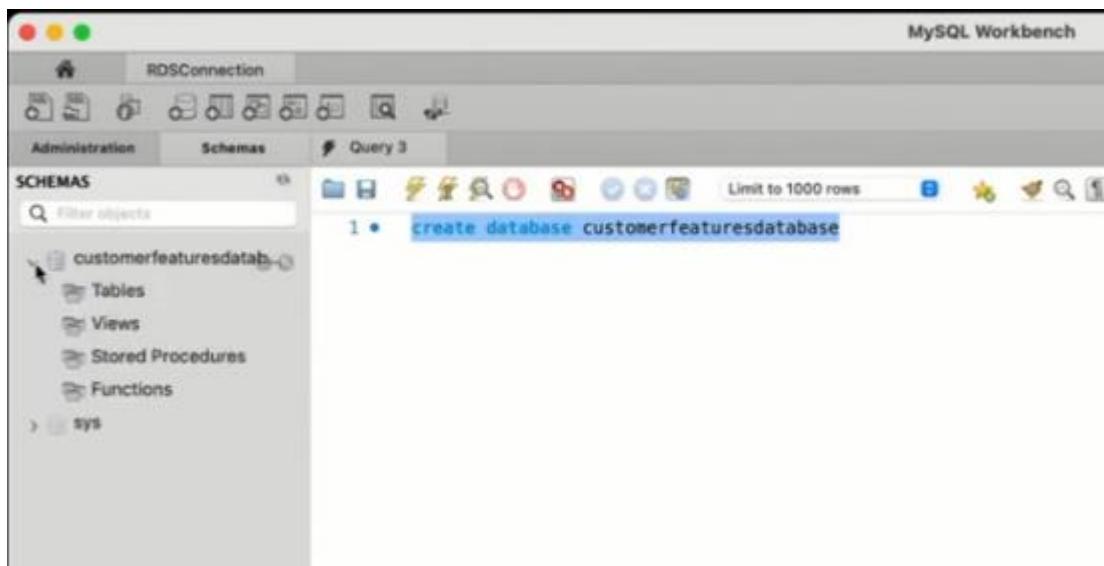
67. AND THIS IS WHERE WE ARE ABLE TO CREATE OUR TABLE WITHIN OUR RDS DATABASE INSTANCE.



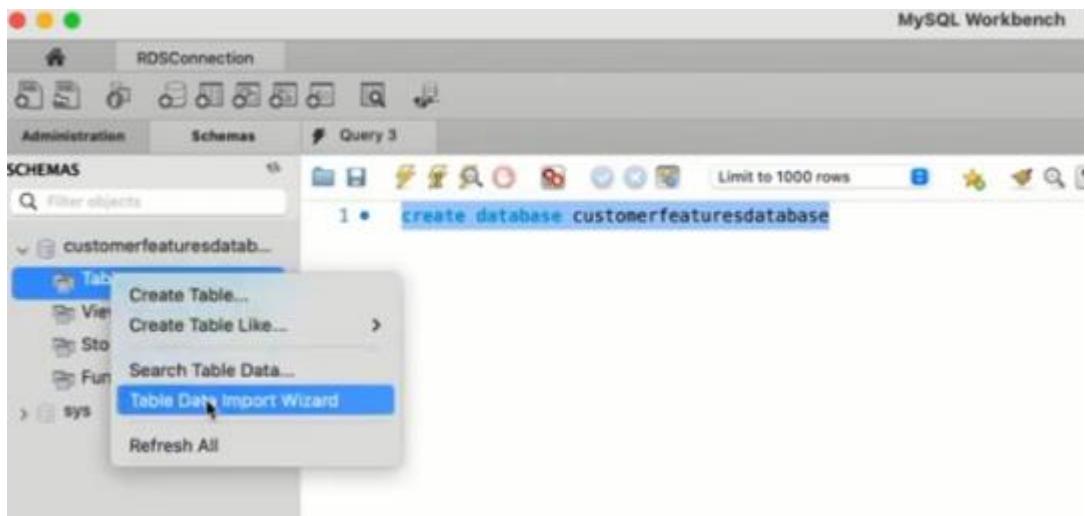
68. WE WILL NOW RUN A QUERY, CREATE DATABASE...THEN EXECUTE.



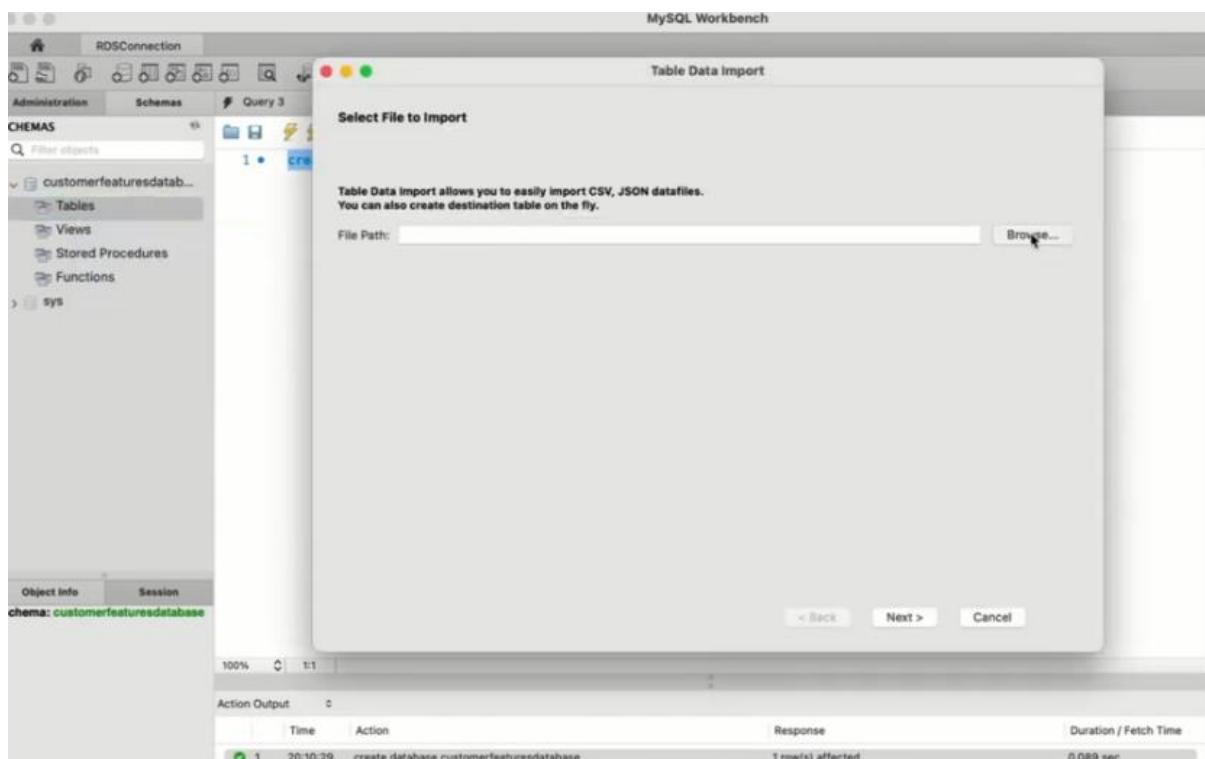
69. REFRESH THE SCHEMA TO SEE THE DATABASE.



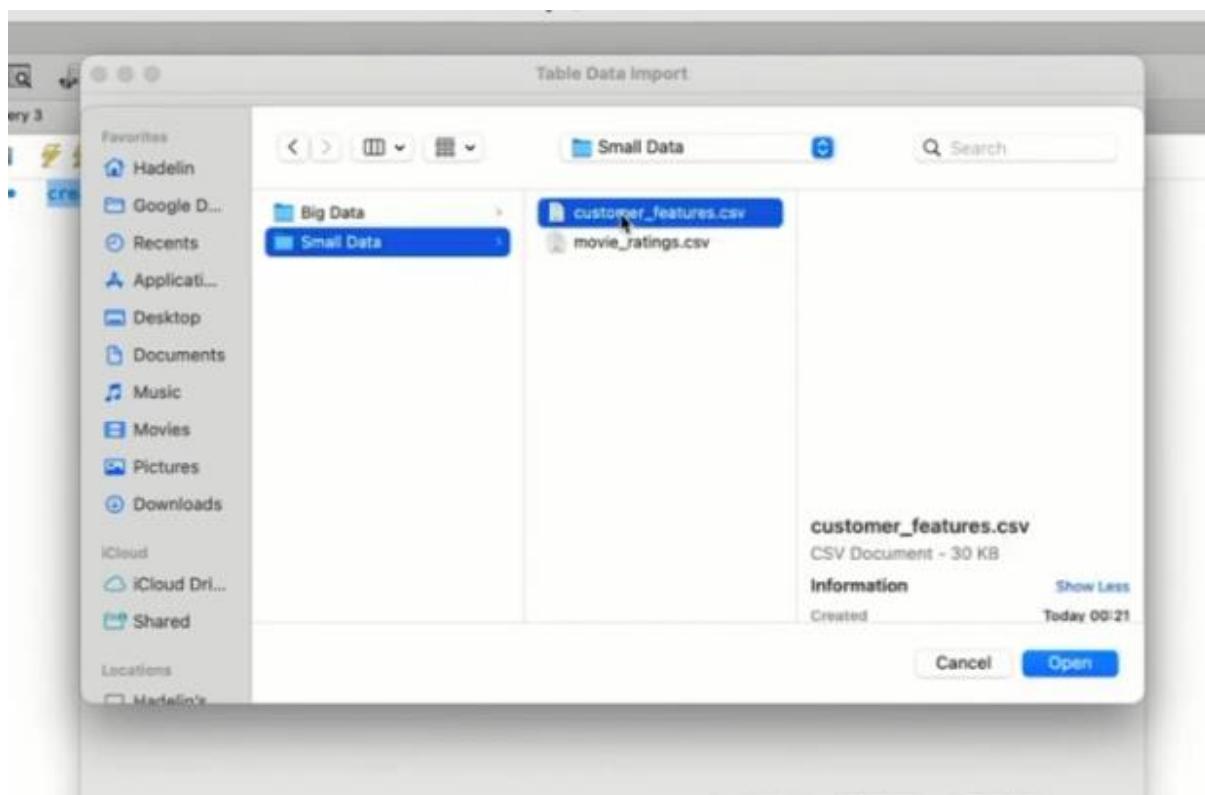
70. CLICK TABLE, CLICK RIGHT, SELECT TABLE DATA IMPORT WIZARD.



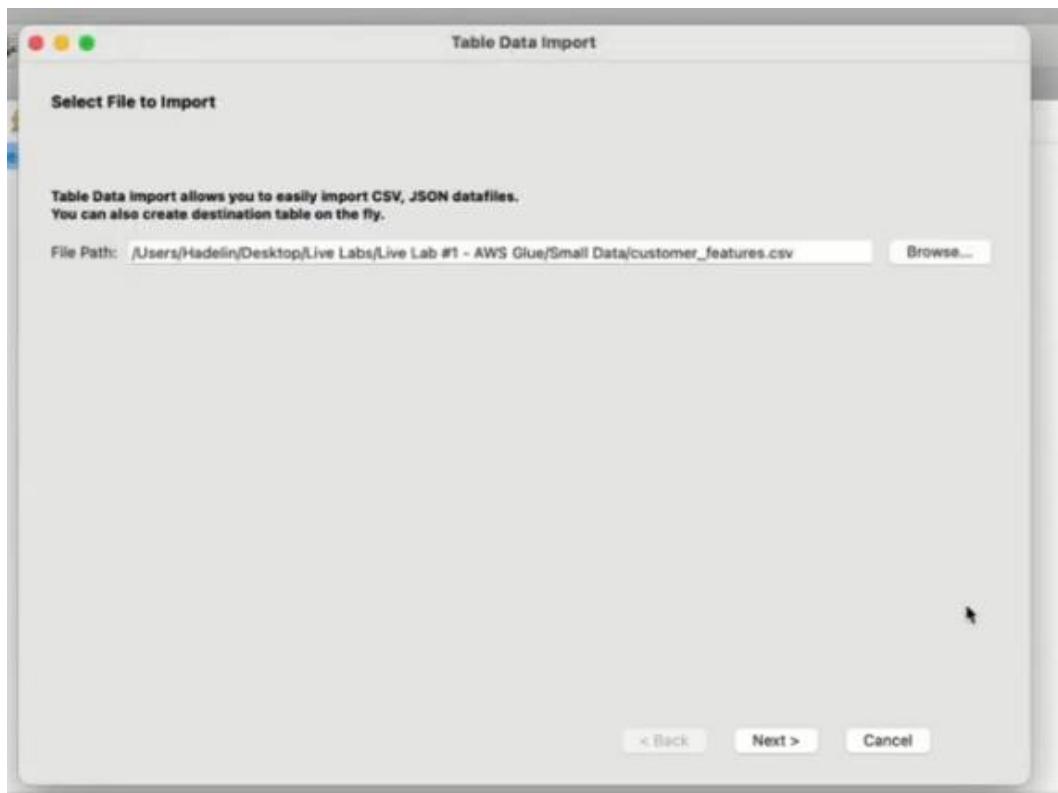
71. BROWSE THE MACHINE TO FIND THE CSV FILE.



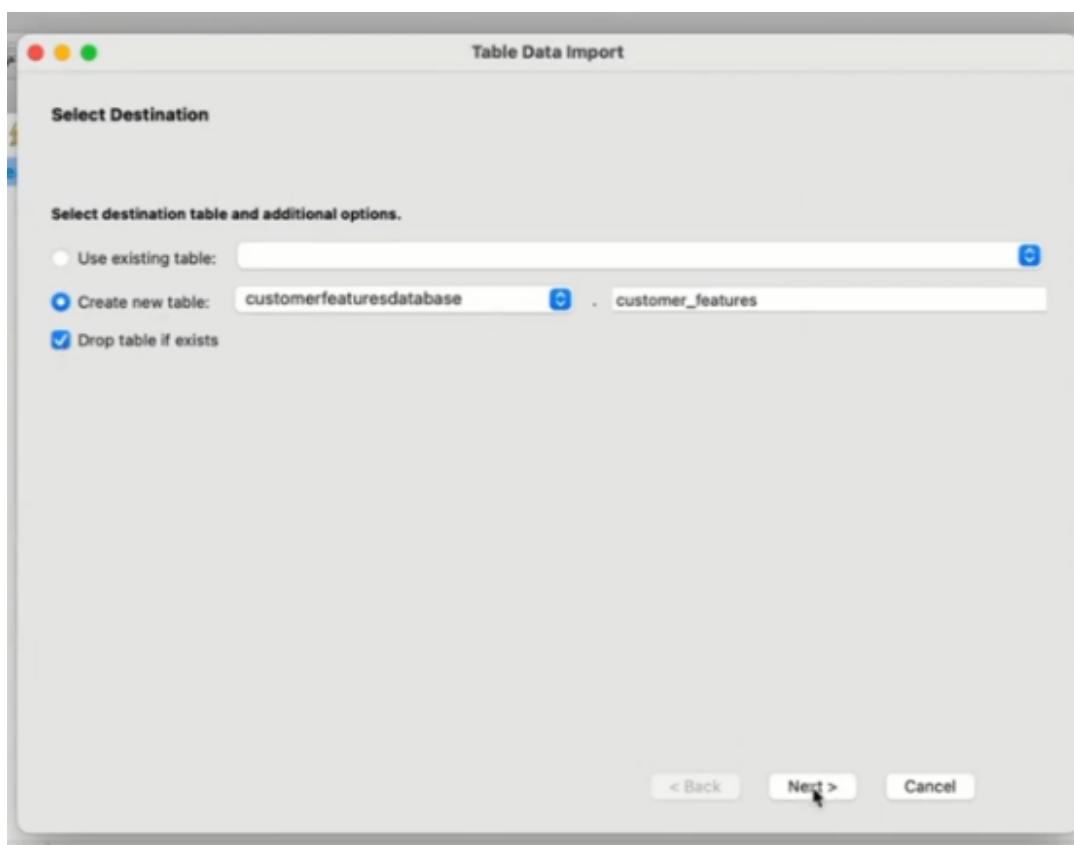
HERE, I CHOSE THE SMALL DATA, CUSTOMER FEATURES FILE.



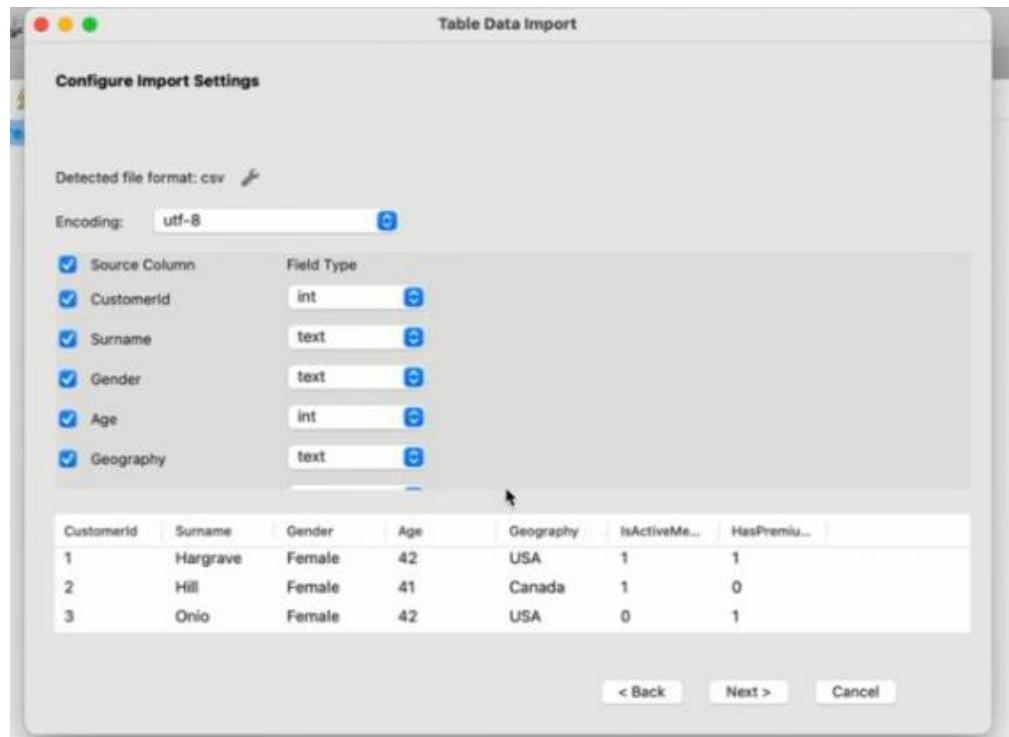
72. CLICK NEXT.



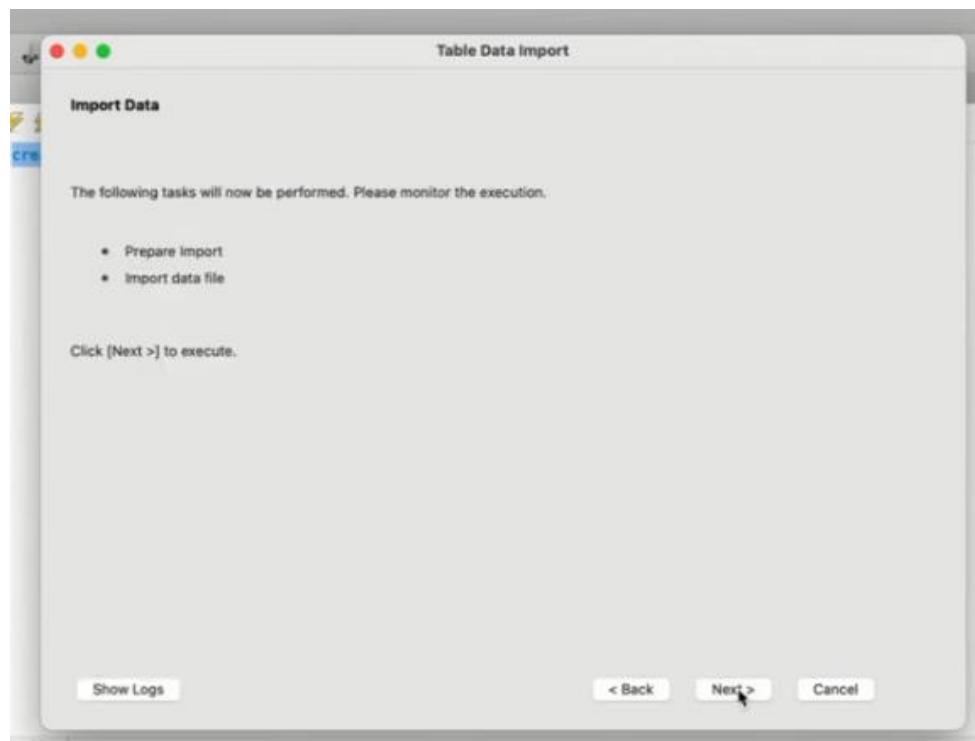
73. YOU CAN CLICK THE DROP TABLE IF EXISTS IF YOU WANT TO. THEN, CLICK NEXT.



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

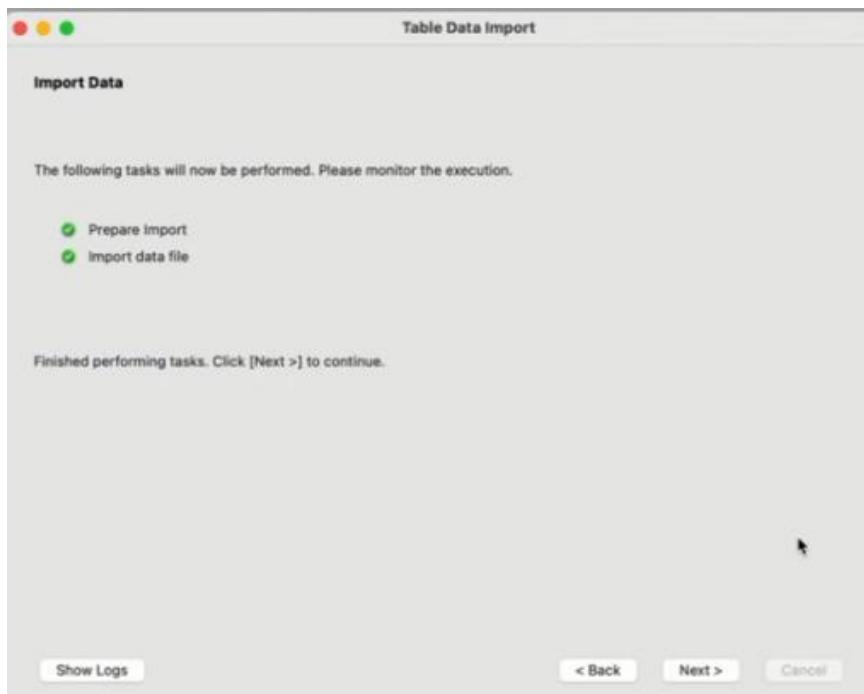


75. IN THE IMPORT DATA, CLICK NEXT TO EXECUTE.



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.

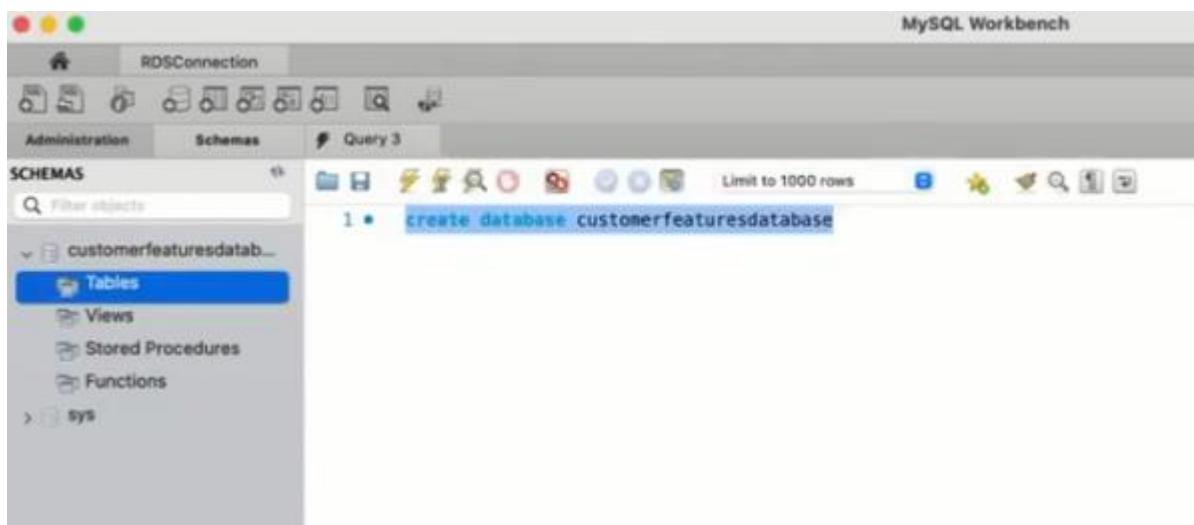
CLICK NEXT.



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 has sections like "Dashboard", "Databases" (which is selected), "Query Editor", etc. The main area shows a database instance named "customer-features-rds-db-instance".

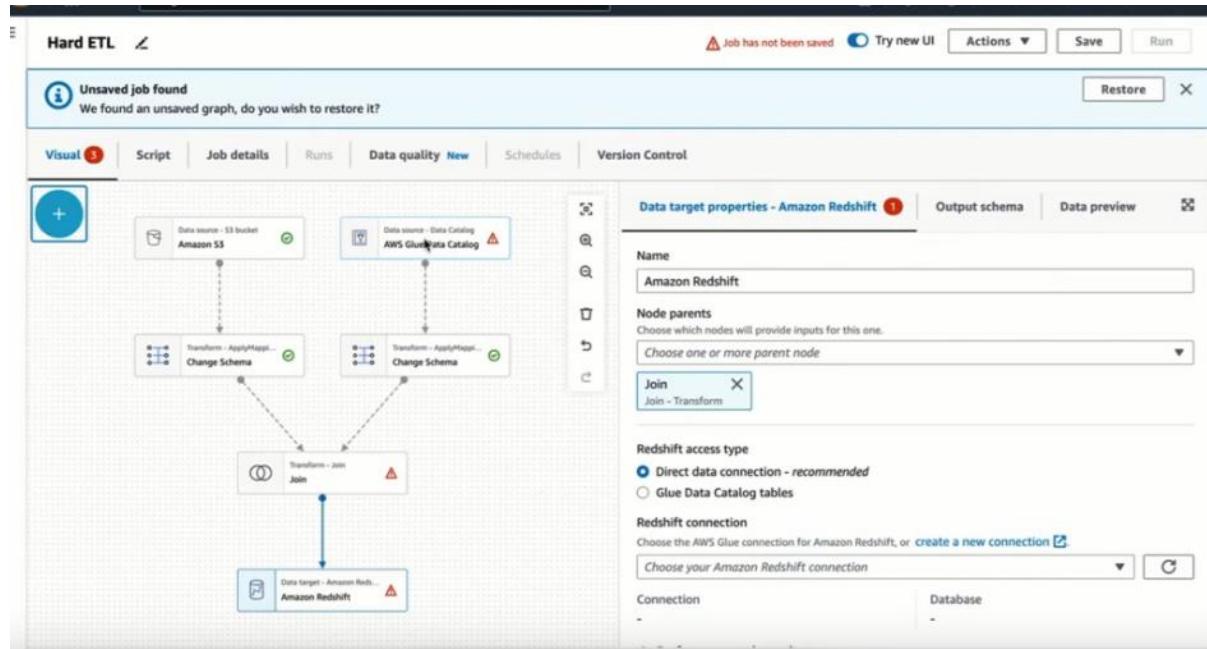
Summary

DB identifier	CPU	Status	Class
customer-features-rds-db-instance	3.52%	Available	db.t3.micro

Connectivity & security

Endpoint & port	Networking	Security
Endpoint customer-features-rds-db-instance.crrwucvfcmbc.us-east-1.rds.amazonaws.com	Availability Zone us-east-1c VPC vpc-0041e736a2d14cfa2	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 has sections for "Getting started", "ETL jobs", "Visual ETL", "Notebooks", "Job run monitoring", "Data Catalog tables", "Data connections", and "Workflows (orchestration)". Under "Data Catalog", there are sections for "Databases", "Tables", "Stream schema registries", "Schemas", "Connections", "Crawlers", "Classifiers", and "Catalog settings". The main content area is titled "Databases (1)" and contains a table with the following data:

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

At the top of the main area, there are buttons for "Edit", "Delete", and "Add database". Above the table, it says "Last updated (UTC) June 14, 2023 at 18:28:32".

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

The screenshot shows the 'Create a database' dialog box in the AWS Glue Data Catalog. The 'Name' field is filled with 'customer-features-glue-database'. The 'Description - optional' field contains 'Enter text'. The 'Location - optional' field is empty. 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 in the AWS Glue Data Catalog. It lists two databases: 'customer-features-glue-database' and 'movie-ratings-glue-database', both created on June 14, 2023.

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 in the AWS Glue Data Catalog. It includes 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 'Create connection' button.

**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, there's a sidebar with various navigation links. The main area is titled 'Create connection' and contains two tabs: 'Connection properties' (selected) and 'Connection access'. Under 'Connection properties', the 'Name' field is set to 'RDSConnection', 'Connection type' is set to 'Amazon RDS', and 'Database engine' is set to 'MySQL'. There's also a checkbox for 'Require SSL connection' which is unchecked. Under 'Connection access', there's a dropdown menu for 'Database instances' containing 'customer-features-rds-db-instance', a 'Database name' input field containing 'customerfeaturesdatabase', and a 'Credential type' section where 'Username and password' is selected. 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 has 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' (represented by a redacted line). At the bottom right are 'Cancel' and 'Create connection' buttons.

86. RDS CONNECTION IS NOW CREATED

The screenshot shows the AWS Glue Connectors page. At the top, a green banner indicates that "RDSConnection" has been successfully added. The left sidebar contains navigation links for various AWS Glue services. The main content area is titled "Connectors" and includes sections for "Marketplace connectors" (with a link to "Go to AWS Marketplace") and "Custom connectors" (with a link to "Create custom connector"). Below these is a table titled "Connectors (0)" with a "Actions" dropdown. Further down is a table titled "Connections (1)" with a "Actions" dropdown, showing one entry: "RDSConnection" (Type: JDBC, Last modified: Jun 14, 2023).

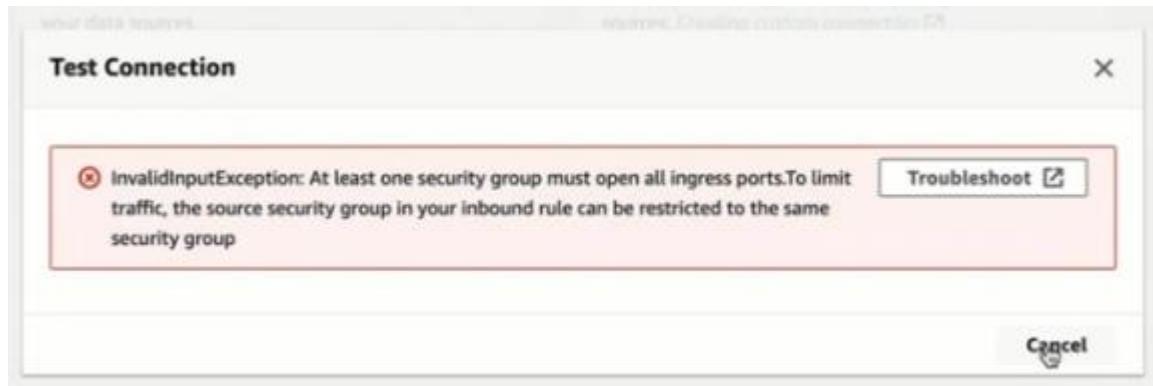
87. TEST THE CONNECTION. SELECT THE CONNECTION. CLICK ACTIONS. CLICK TEST CONNECTION.

The screenshot shows the AWS Glue Connections page. It features a table titled "Connections (1)" with a "Actions" dropdown. The table lists one connection: "RDSConnection" (Type: JDBC, Last modified: Jun 14, 2023). The "Actions" dropdown menu is open, and the "Test connection" option is highlighted with a mouse cursor.

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

The 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 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 & security. The main content area shows the 'Summary' tab with basic instance details: DB identifier (customer-features-rds-db-instance), CPU usage (3.52%), Status (Available), Class (db.t3.micro), Role (Instance), Current activity (0 Connections), Engine (MySQL Community), Region & AZ (us-east-1c). Below the summary is the 'Connectivity & security' tab, which displays the endpoint and port (customer-features-rds-db-instance.crrwucvfcmb.us-east-1.rds.amazonaws.com, Port 3306), networking (Availability Zone us-east-1c, VPC vpc-0041e736a2d14cf2a), and security (VPC security groups: SG-Open-MySQL (sg-05ae22adcf0dce56) and default (sg-00757adc39bf9625e), both marked as Active). The 'Publicly accessible' checkbox is also present.

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, and Elastic Block Store. The main area displays a table of security groups. One row is selected, showing details for 'sg-05eae22adcf0dce56 - SG-Open-MySQL'. Below this, a sub-menu for 'Inbound rules' is open, showing a single rule for MySQL/Aurora traffic on port 3306.

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
<input checked="" type="checkbox"/>	sg-05eae22adcf0dce56	SG-Open-MySQL	vpc-0041e736a2d14cfa2	Allows MySQL Access t...	7496011

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. It lists one existing rule (MySQL/Aurora on port 3306) and provides a form to add a new rule. The 'Add rule' button is visible at the bottom left, and 'Save rules' is at the bottom right.

Name	Type	Protocol	Port range	Source	Description - optional
sgr-000a62b4b15cc1b15	MySQL/Aurora	TCP	3306	Custom	0.0.0.0/0

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 is displayed: "Inbound security group rules successfully modified on security group (sg-05eae22adcf0dce56 | SG-Open-MySQL)".

The "Inbound rules" tab is selected for the security group "sg-05eae22adcf0dce56 - SG-Open-MySQL". The table shows the following rules:

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
<input type="checkbox"/>	sg-04c7ce4cb6a0e059f	launch-wizard-1	vpc-0041e736a2d14cfa2	launch-wizard-1-create...	7496011
<input type="checkbox"/>	sg-00757adc39bf9625e	default	vpc-0041e736a2d14cfa2	default VPC security gr...	7496011

A message at the bottom says: "You can now check network connectivity with Reachability Analyzer". There is a "Run Reachability Analyzer" button.

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

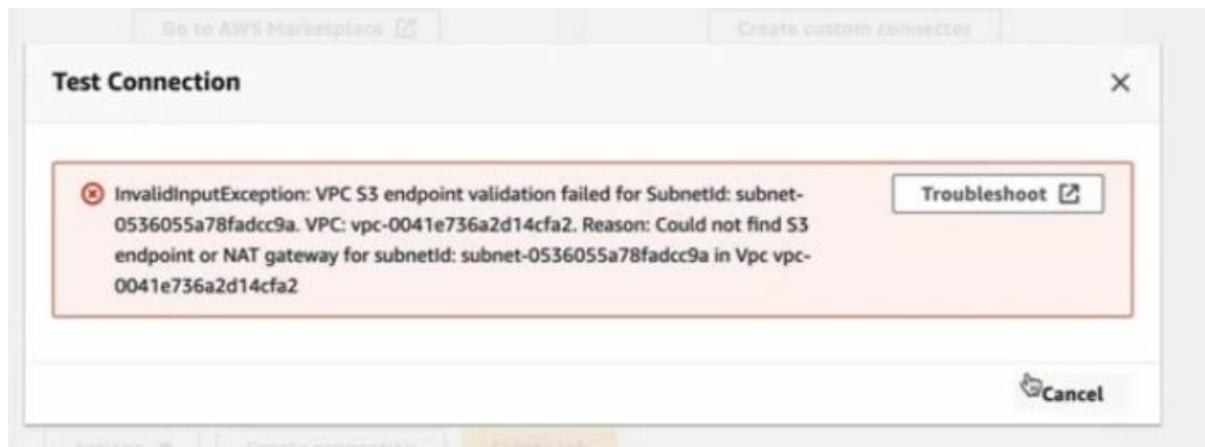
SELECT TEST CONNECTION.

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). Above the table are buttons for 'Actions', 'Create connection', and 'Create job'. A 'Test connection' button is visible in the table row. The table also includes a search bar and sorting/filtering options.

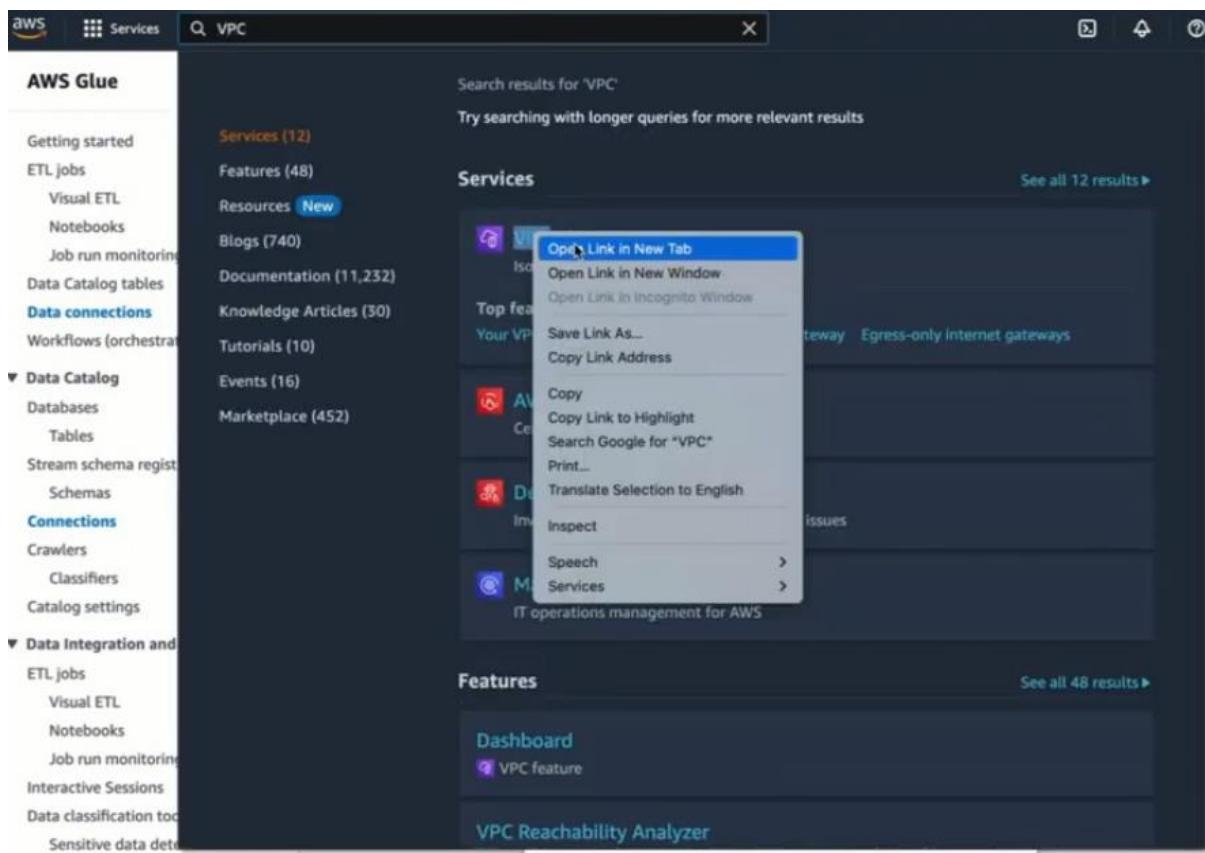
CHOOSE GLUEFULLACCESSROLE. THEN, CONFIRM.

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 is open, showing options like Virtual private cloud, Security, and Endpoints. The Endpoints section is selected. The main content area has a heading 'Endpoints' with a 'Info' link. Below it is a search bar and a table header with columns: Name, VPC endpoint ID, VPC ID, and Service name. The table body displays the message 'No endpoint found'. At the bottom, there's a button labeled 'Select an endpoint'.

98. CLICK CREATE ENDPOINTS

The screenshot shows the same AWS VPC Endpoints page as above, but with a red box highlighting the 'Create endpoint' button in the top right corner of the main content area.

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' wizard. The first step, 'Endpoint settings', is displayed. It includes a 'Name tag - optional' field containing 'VPC Endpoint to S3', and a 'Service category' section with several options: 'AWS services' (selected), 'PrivateLink Ready partner services', 'AWS Marketplace services', 'EC2 Instance Connect Endpoint', and 'Other endpoint services'. The URL for this step is 'VPC > Endpoints > Create endpoint'.

100. IN THE SERVICE CATERGORY, SELECT AWS SERVICES.

The screenshot shows the 'Endpoint settings' configuration screen. At the top, there's a section for a 'Name tag - optional' which contains the text 'VPC Endpoint to S3'. Below this is a 'Service category' section with the sub-instruction 'Select the service category'. There are five options listed in boxes:

- 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

The screenshot shows a list of services titled 'Services (225)'. A search bar at the top contains 'S3'. The results table has columns for 'Owner' and 'Type'. Several rows are visible, including:

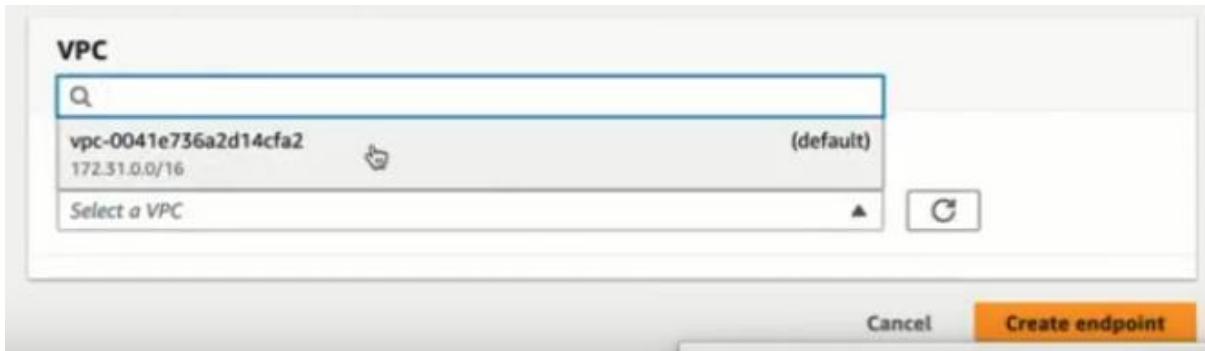
- Use: "S3"
- Client filters values
- Service Name = com.amazonaws.s3-global.accesspoint
- Service Name = com.amazonaws.us-east-1.s3** (highlighted with a mouse cursor)
- Service Name = com.amazonaws.us-east-1.s3-outposts
- com.amazonaws.us-east-1.acm-pca
- com.amazonaws.us-east-1.airflow.api

102. SELECT GATEWAY

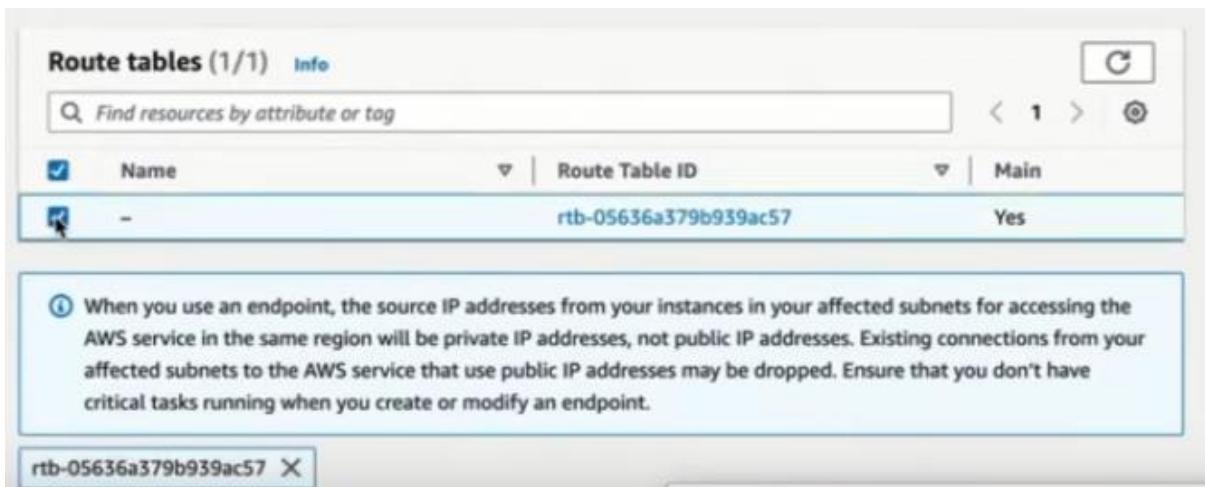
The screenshot shows a list of services titled 'Services (1/2)'. A search bar at the top contains 'Find resources by attribute or tag'. A filter bar below it shows 'Service Name = com.amazonaws.us-east-1.s3' with a clear filters button. The results table has columns for 'Owner' and 'Type'. Two rows are visible:

- Service Name = com.amazonaws.us-east-1.s3 (highlighted with a mouse cursor)
- Service Name = com.amazonaws.us-east-1.s3

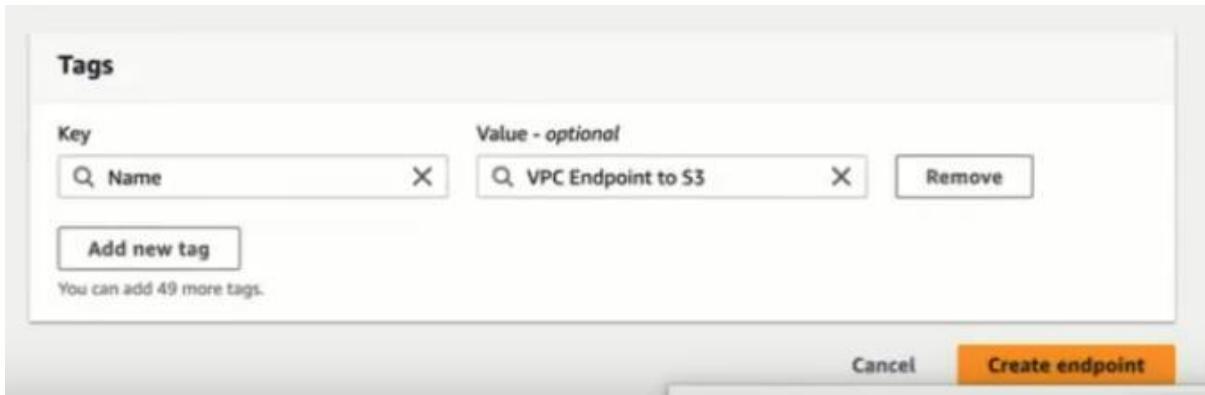
103. SELECT THE VPC. REMEMBER TO CONNECT IT TO THE SUBNET.



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. On the left, there's a sidebar with options like 'Virtual private cloud', 'Your VPCs', 'Subnets', 'Route tables', etc. The main area has a green header bar with the message 'Successfully created VPC endpoint vpce-0dfd95341ef6181af'. Below this is a table titled 'Endpoints (1/1)'. It contains one row with the following data:

Name	VPC endpoint ID	VPC ID	Service name
vpce Endpoint to S3	vpce-0dfd95341ef6181af	vpc-0041e736a2d14cfa2	com.amazonaws.us-east-1.s3

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 interface. The left sidebar includes sections for 'Data connections', 'Connections', and 'Data Integration and ETL'. 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 is a table titled 'Connectors (0)'. At the bottom, there's a table titled 'Connections (1)'.

Name	Type	Last modified
RDSConnection	JDBC	Jun 14, 2023

108. CLICK RDS CONNECTION, 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'. A cursor is hovering over the 'Actions' button for the first row.

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'. A cursor is hovering over the 'Test connection' link in the 'Actions' column for the first row.

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

109. SELECT GLUEFULLACCESS, THEN, CONFIRM.

The screenshot shows a 'Test Connection' dialog box. Under the 'IAM role' section, 'GlueFullAccessRole' is selected in a dropdown menu. There is also a 'Create IAM role.' link. At the bottom right 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.' At the bottom right are 'Cancel' and 'View log' 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)', 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.

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 at Step 2: 'Choose data sources and classifiers'. On the left, a sidebar lists steps: Step 1 (Set crawler properties), Step 2 (Choose data sources and classifiers, currently selected), Step 3 (Configure security settings), Step 4 (Set output and scheduling), and Step 5 (Review and create). The main panel title is 'Choose data sources and classifiers'. It contains a 'Data source configuration' section with a question 'Is your data already mapped to Glue tables?' and 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.

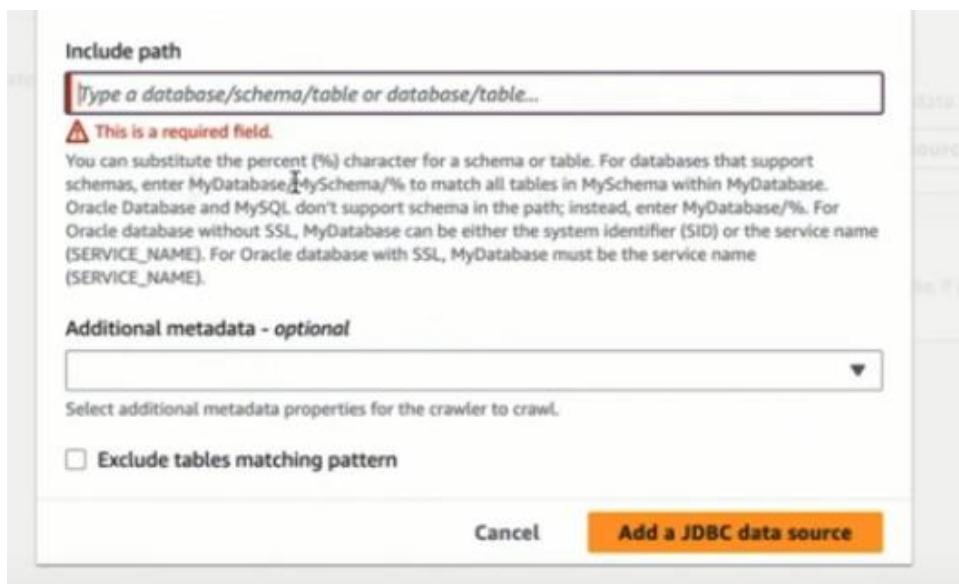
115. IN THE DATA SOURCE, SELECT JDBC WHICH MEANS JAVA DATABASE CONNECTION.

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' (highlighted with a mouse cursor), 'DynamoDB', 'DocumentDB/MongoDB', and 'Delta Lake'. Below these is an option 'In a different account' with an unchecked radio button. The 'S3 path' section asks 'Browse for or enter an existing S3 path.' It includes a text input field with 's3://bucket/prefix/object', a 'View' button, and a 'Browse S3' button. A note at the bottom states: '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.'

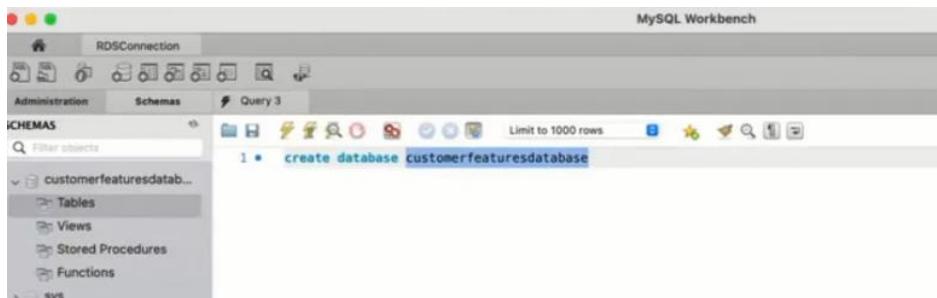
116. IN THE CONNECTION, SELECT RDSCONNECTION THAT WE HAVE JUST CREATED AND TESTED SUCCESSFULLY.



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.

Add data source

Data source

Choose the source of data to be crawled.

JDBC

Connection

Select a connection to access the data sources below.

RDSConnection

Clear selection Add new connection

Include path

customerfeaturesdatabase/%

You can substitute the percent (%) character for a schema or table. For databases that support schemas, enter MyDatabase/MySchema/% to match all tables in MySchema within MyDatabase. Oracle Database and MySQL don't support schema in the path; instead, enter MyDatabase/%. For Oracle database without SSL, MyDatabase can be either the system identifier (SID) or the service name (SERVICE_NAME). For Oracle database with SSL, MyDatabase must be the service name (SERVICE_NAME).

Additional metadata - optional

Select additional metadata properties for the crawler to crawl.

Exclude tables matching pattern

Cancel Add a JDBC data source

120. SELECT JDBC. THEN, 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

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.

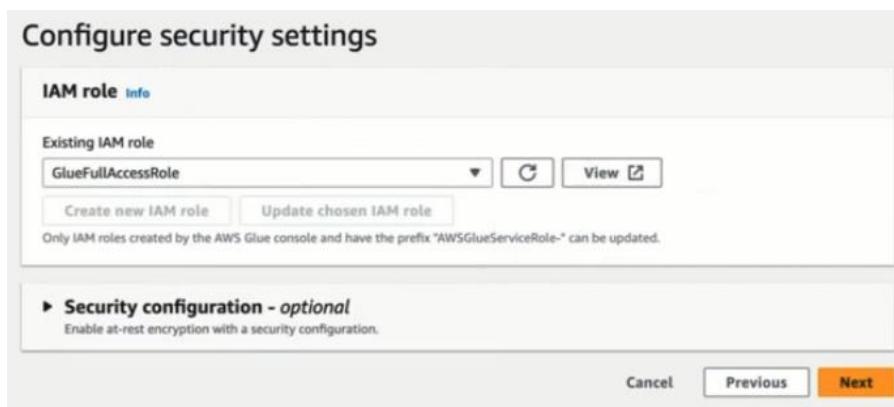
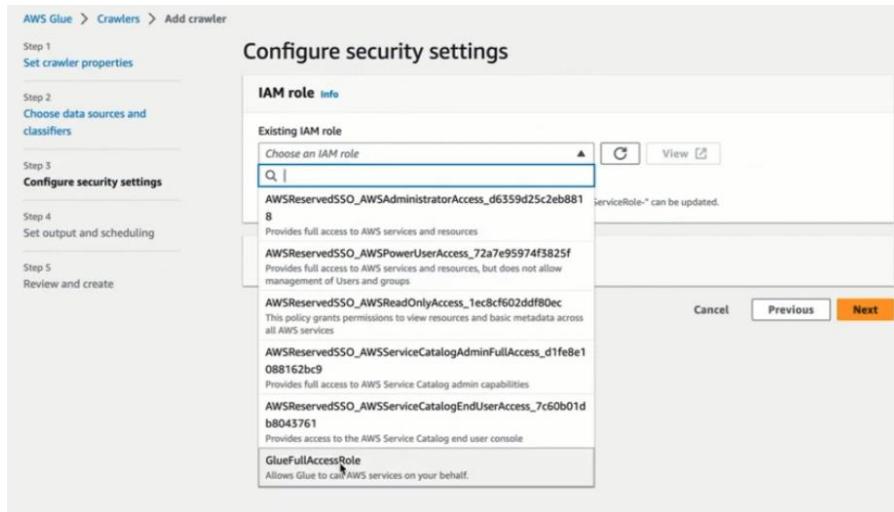
Data sources (1) Info
The list of data sources to be scanned by the crawler.

Type	Data source	Parameters
JDBC	customerfeaturesdatabase/%	-

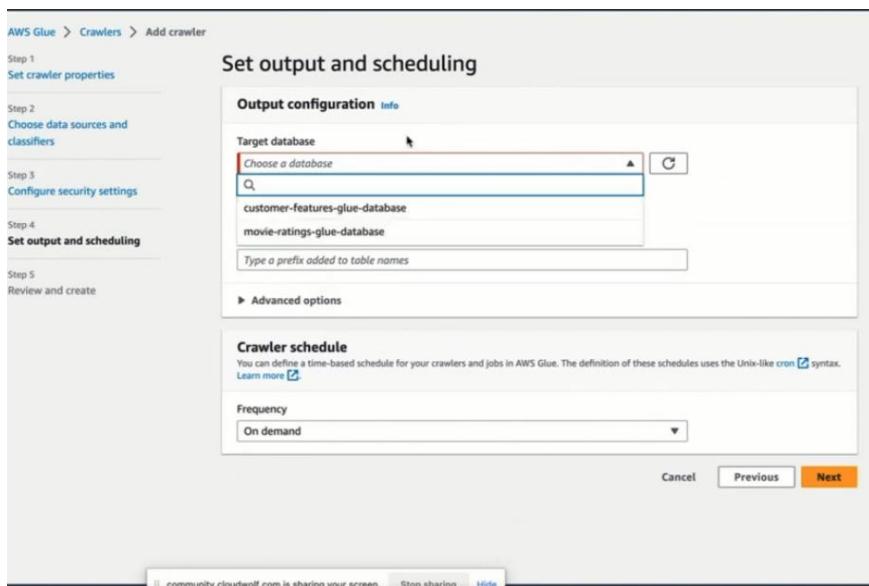
Custom classifiers - optional
A classifier checks whether a given file is in a format the crawler can handle. If it is, the classifier creates a schema in the form of a StructType object that matches that data format.

Cancel Previous Next

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

customer-features-glue-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

Set crawler properties

Name	Description	Tags
customer-features-crawler	-	-

Step 2: Choose data sources and classifiers

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

Type	Data source	Parameters
JDBC	customerfeaturesdatabase/%	-

Step 3: Configure security settings

Configure security settings

IAM role	Security configuration	Lake Formation configuration
GlueFullAccessRole	-	-

Step 4: Set output and scheduling

Set output and scheduling

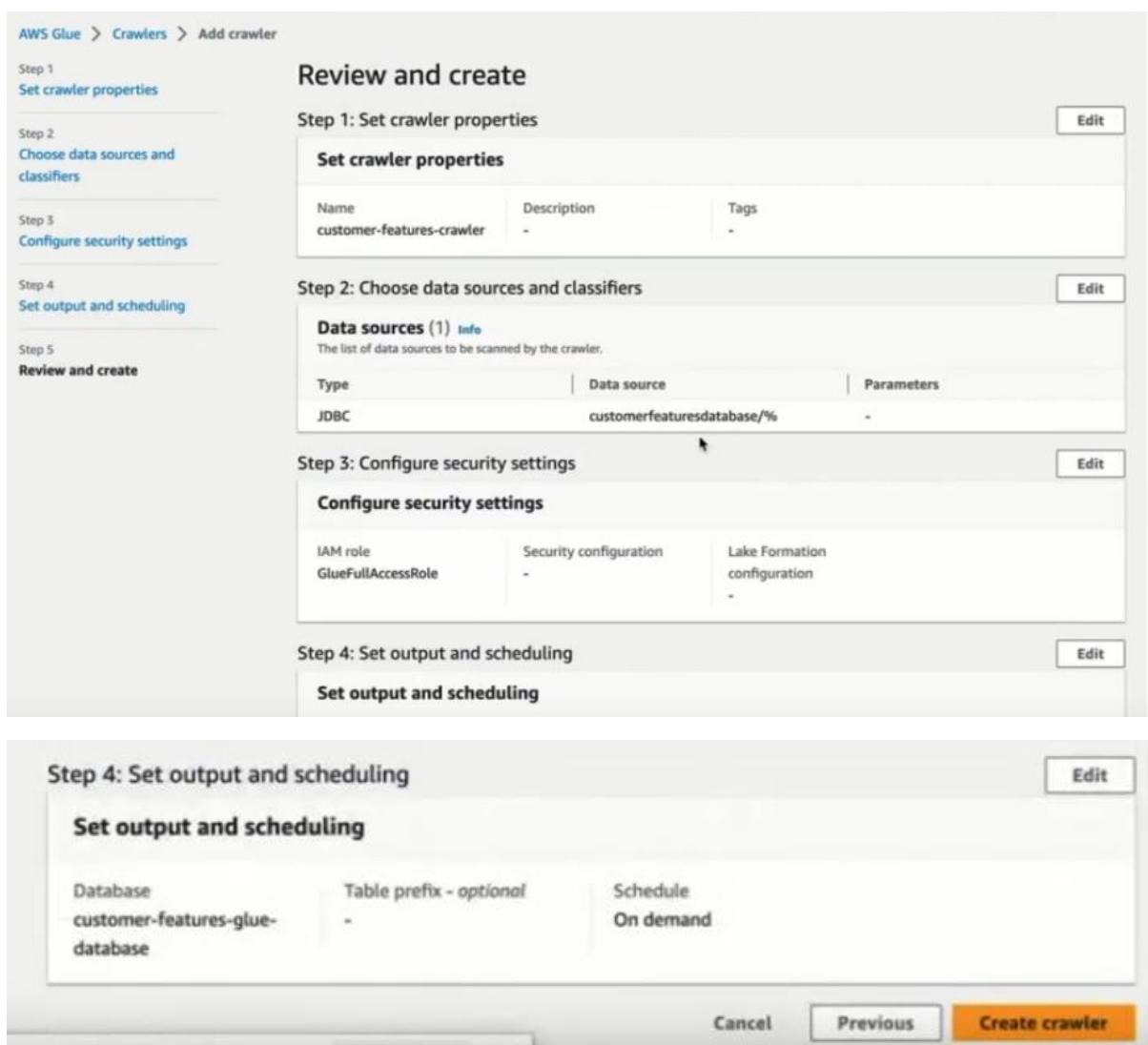
Database	Table prefix - optional	Schedule
customer-features-glue-database	-	On demand

Step 4: Set output and scheduling

Set output and scheduling

Database	Table prefix - optional	Schedule
customer-features-glue-database	-	On demand

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', etc. The main area is titled 'customer-features-crawler' and contains a 'Crawler properties' section with fields for Name, IAM role, Database, and State. Below this is a 'Crawler runs' section which is currently empty. At the top right of the main area, there are buttons for 'Run crawler', 'Edit', and 'Delete'.

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

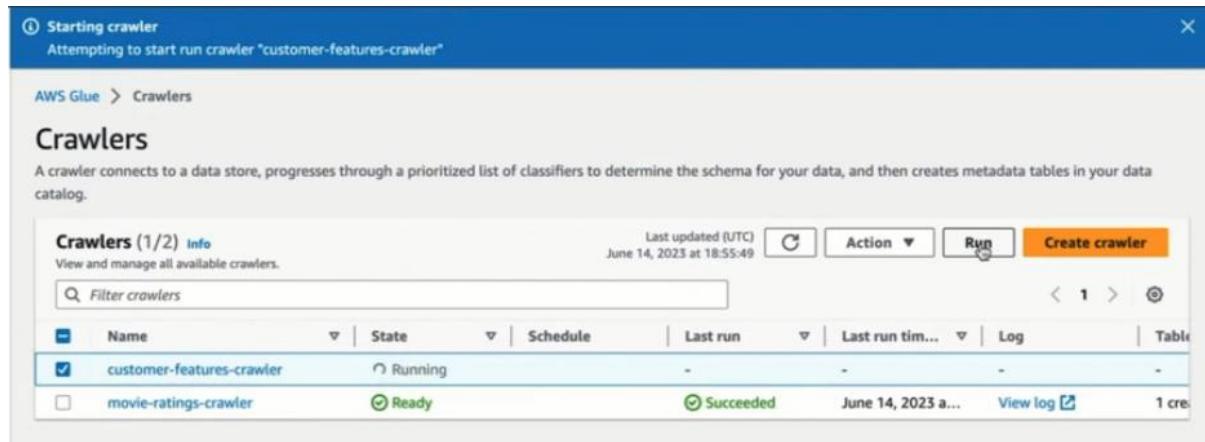
This screenshot shows the 'Crawlers' page in AWS Glue. It lists one crawler named 'customer-features-crawler' which is currently 'Ready'. There are buttons for 'Action', 'Run', and 'Create crawler' at the top right. The table below shows columns for Name, State, Last run, Log, and Table changes.

Name	State	Last run	Log	Table changes
customer-features-cr...	Ready	Succeeded	June 14, 2023 a...	View log

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

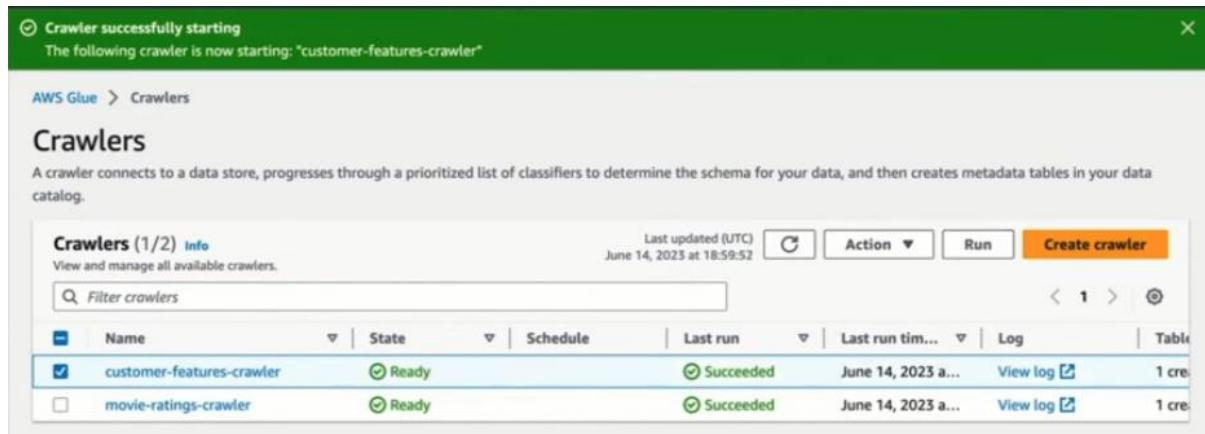
This screenshot shows the 'Crawlers' page again, but this time the 'customer-features-crawler' is selected (indicated by a checked checkbox). The 'Run' button is highlighted in orange at the top right of the table header.

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.



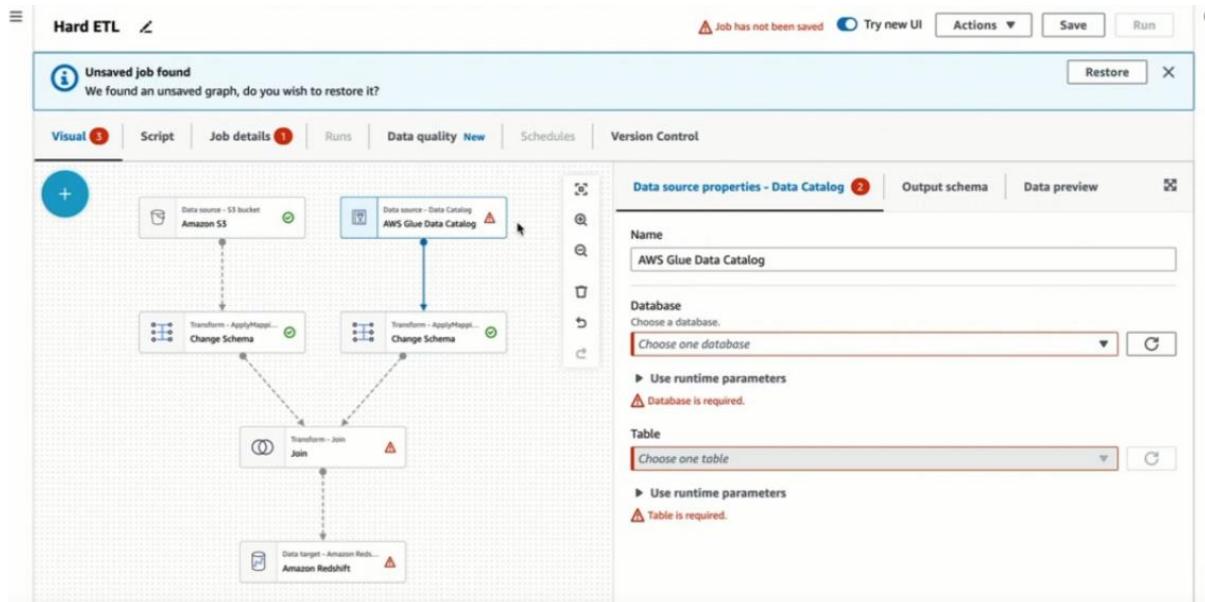
The screenshot shows the AWS Glue Crawlers interface. At the top, a blue header bar displays a status message: "Starting crawler Attempting to start run crawler 'customer-features-crawler'". Below this, the main title "Crawlers" is shown, followed by a sub-instruction: "A crawler connects to a data store, 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 "Run" button highlighted in orange, indicating it is currently running.

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



The screenshot shows the AWS Glue Crawlers interface after the crawler has successfully started. A green header bar 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 sub-instruction: "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, Succeeded) and "movie-ratings-crawler" (Ready, Succeeded). Both rows have "View log" links next to them.

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 provides a detailed view of the "Database" configuration in the "Data source properties - Data Catalog" panel. The "Name" field is filled with "AWS Glue Data Catalog". In the "Database" section, a dropdown menu is open, showing the option "Choose one database". Below the dropdown, a search bar contains the text "Filter databases". Two database entries are listed: "customer-features-glue-database" and "movie-ratings-glue-database". The "customer-features-glue-database" entry is highlighted with a blue selection bar. A tooltip for the database dropdown states: "Choose a database." and "Choose one database". At the bottom of the panel, there are two status messages: "▶ Use runtime parameters" and "⚠ Table is required.".

131. IN THE TABLE, CHOOSE CUSTOMERFEATURESDATABASE_CUSTOMER_FEATURES.

Table

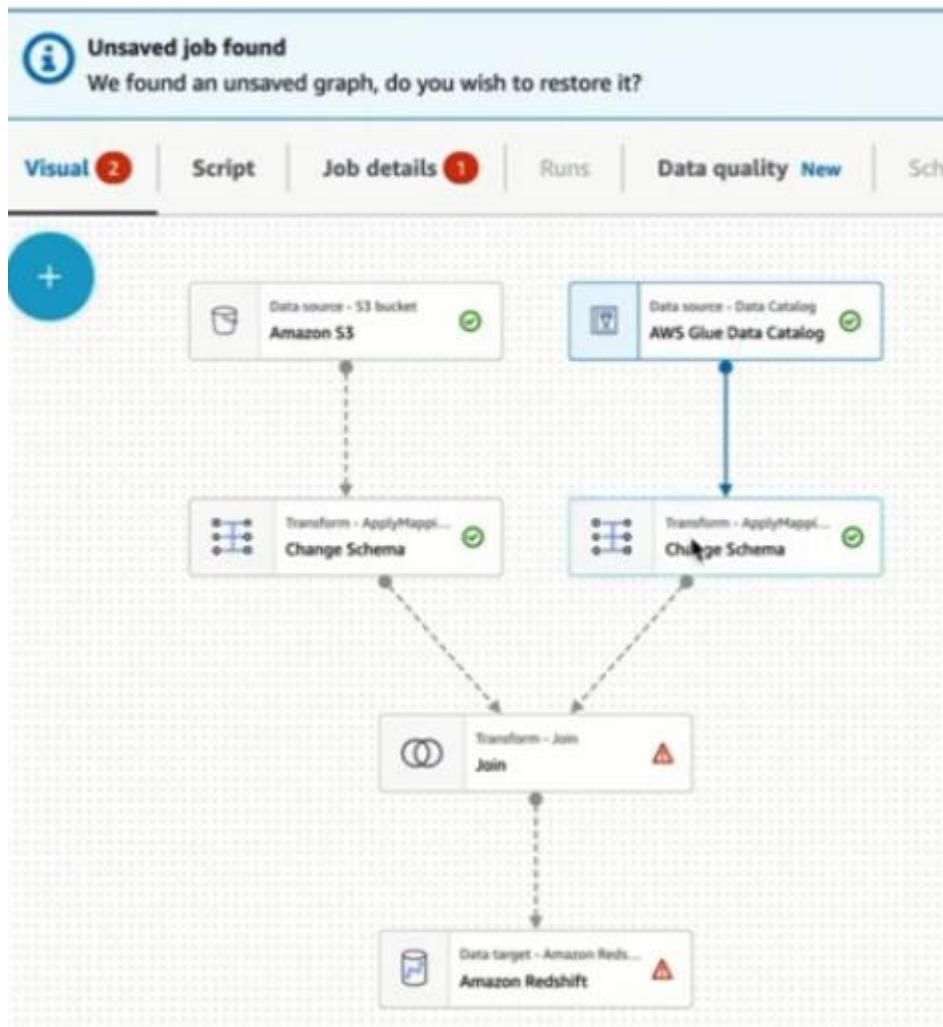
A screenshot of a dropdown menu titled "Choose one table". The menu contains a search bar labeled "Filter tables" and a list of table names. The table "customerfeaturesdatabase_customer_features" is highlighted with a blue border and has a small circular icon with a downward arrow pointing to it at the bottom right. Other visible table names include "Choose one table" and "customerfeaturesdatabase_customer_features".

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

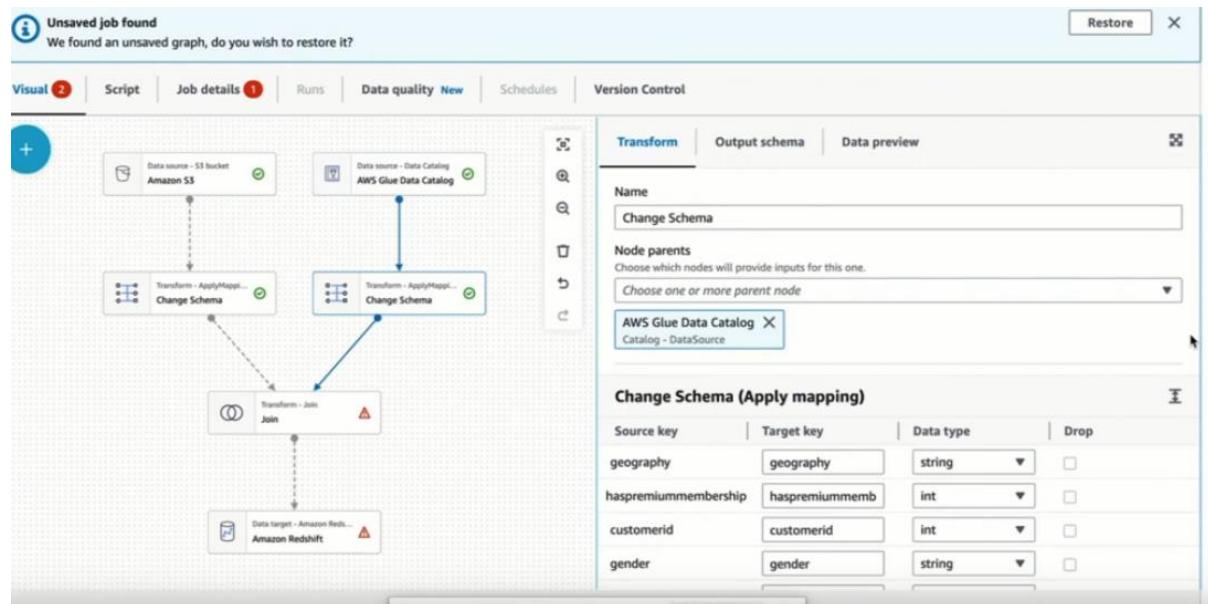
A screenshot of the AWS Glue Data Catalog configuration interface. The top navigation bar shows tabs: "Data source properties - Data Catalog" (selected), "Output schema", and "Data preview".
The "Name" field is set to "AWS Glue Data Catalog".
The "Database" section shows a dropdown menu with "customer-features-glue-database" selected.
The "Table" section shows a dropdown menu with "customerfeaturesdatabase_customer_features" selected.
Both the "Database" and "Table" sections have a "▶ Use runtime parameters" link below them.

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 features tabs for "Transform", "Output schema", and "Data preview". The "Transform" tab displays a dropdown for "Choose one or more parent node" set to "Choose one or more parent node" and a section for "AWS Glue Data Catalog X Catalog - DataSource".

Below this, a table titled "Change Schema (Apply mapping)" lists eight columns: Source key, Target key, Data type, and Drop. The data is as follows:

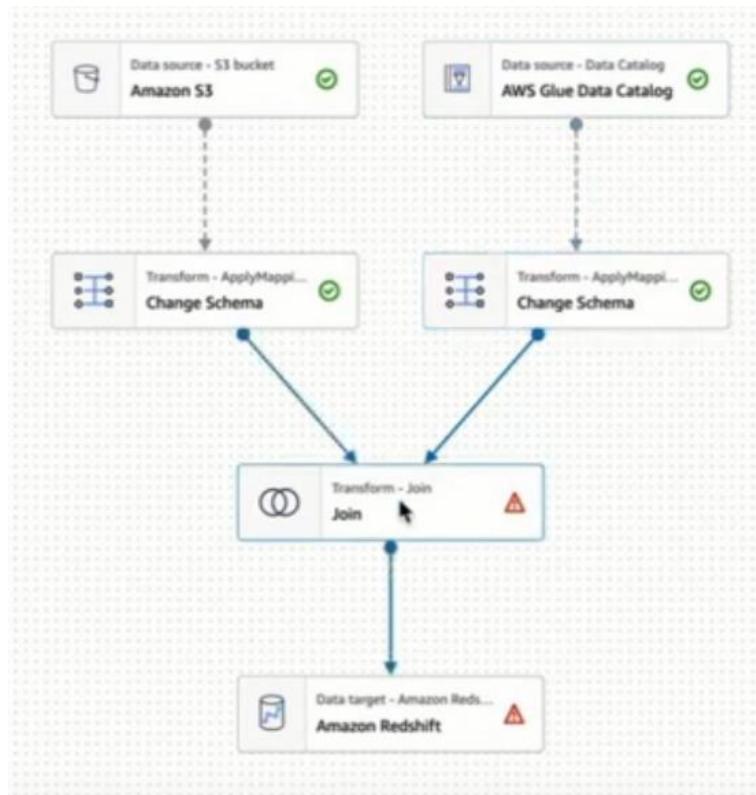
Source key	Target key	Data type	Drop
geography	geography	string	<input type="checkbox"/>
haspremiummembership	haspremiummemb	int	<input type="checkbox"/>
customerid	customerid	int	<input type="checkbox"/>
gender	gender	string	<input type="checkbox"/>
isactivemember	isactivemember	int	<input type="checkbox"/>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

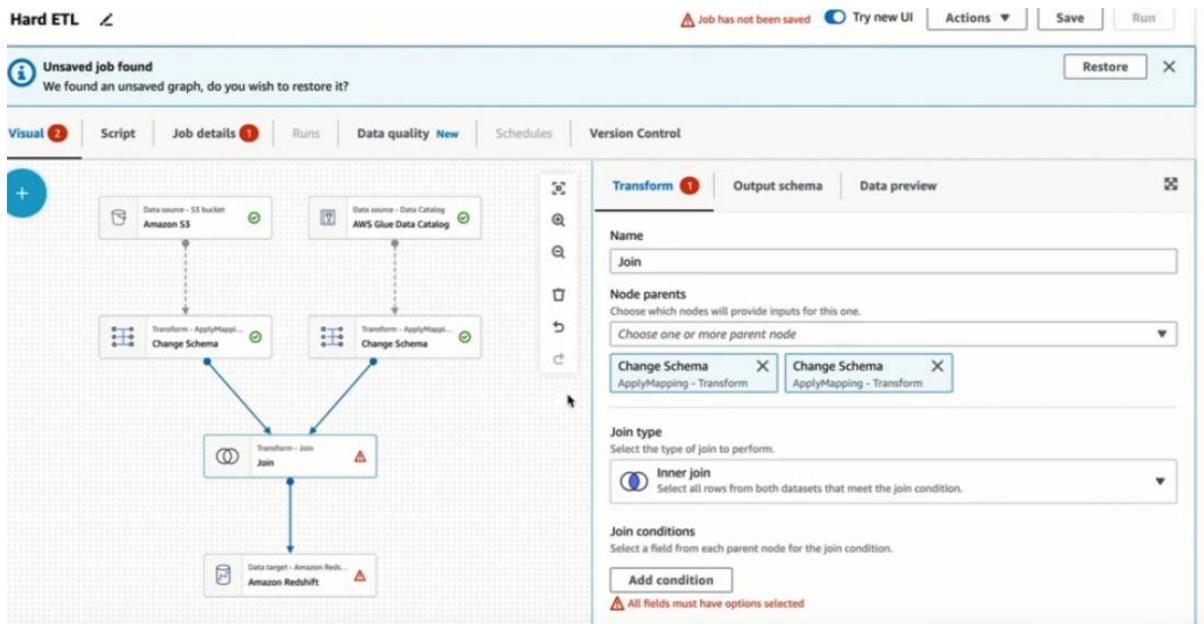
The screenshot shows the 'Change Schema (Apply mapping)' section of the AWS Glue Data Catalog. It displays a table with columns: Source key, Target key, Data type, and Drop. The table contains the following data:

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

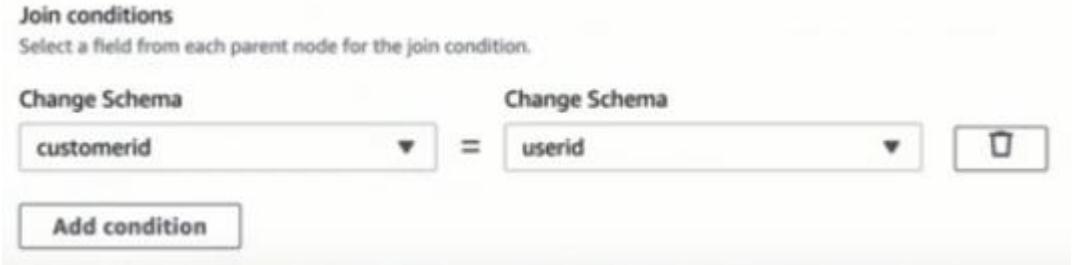
138. IN THE CHANGE SCHEMA, CHOOSE CUSTOMERID AND USERID. WE ARE GOING TO JOIN THEM THROUGH THIS COMMON DENOMINATOR OR VARIABLE WHICH ARE THE IDS OF THE CUSTOMERS OR USERS OF THIS MOVIE STREAMING PLATFORM.

Join conditions

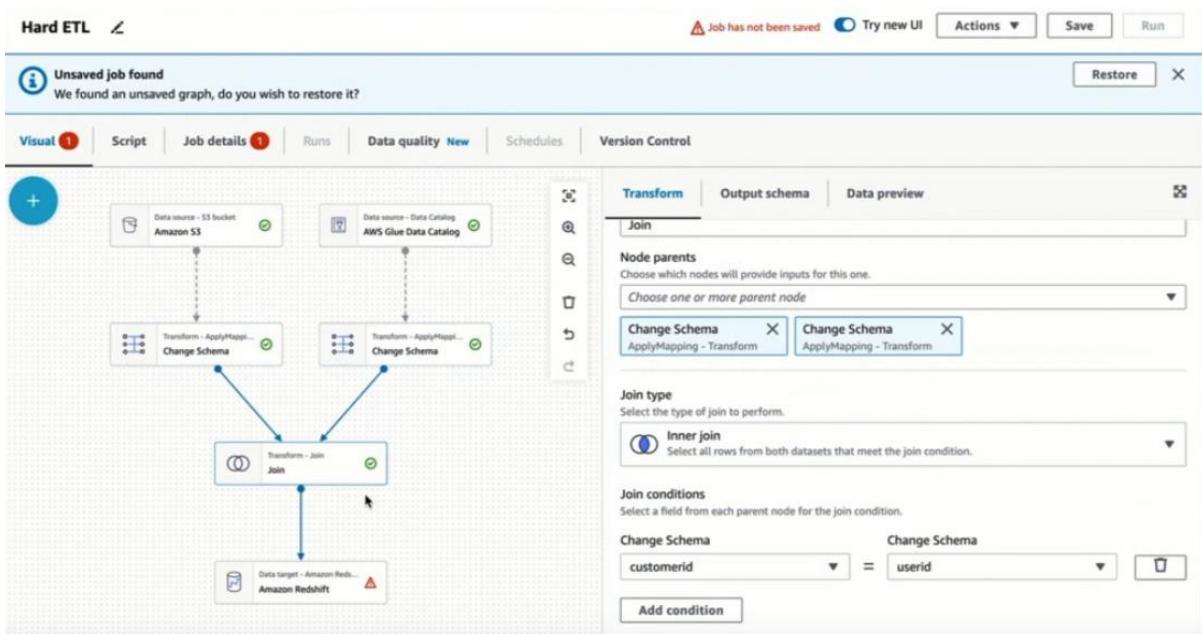
Select a field from each parent node for the join condition.

Change Schema Change Schema

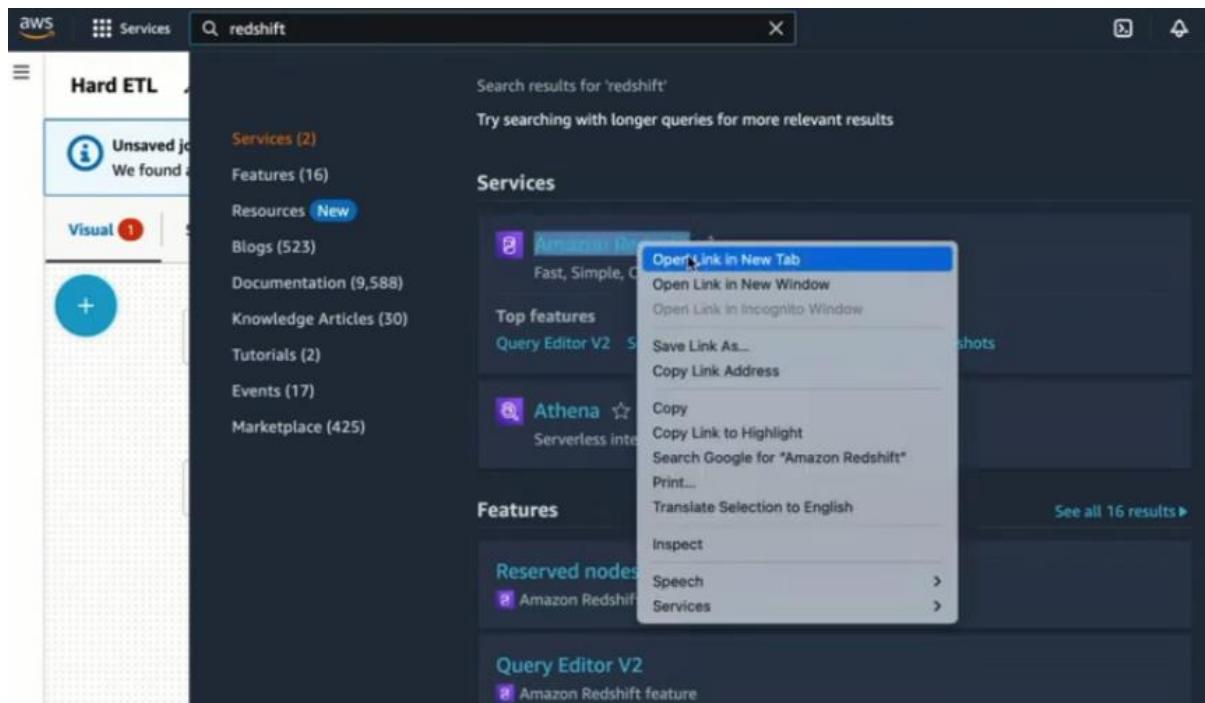
customerid = userid Add condition



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.

The screenshot shows the 'Get started with Amazon Redshift Serverless' page. At the top, there's a navigation bar with 'Amazon Redshift Serverless' and 'Get started with Amazon Redshift Serverless'. Below the navigation, the title 'Get started with Amazon Redshift Serverless' is followed by a link 'Info'. A note says '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.' Under the title, there are two options: 'Use default settings' (selected) and 'Customize settings'. The 'Use default settings' option includes a note: 'Default settings have been defined to help you get started. You can change them at any time later.' Below this, there's a section titled '▼ How it works'.

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

The screenshot shows the 'Permissions' section of the IAM role creation page. It starts with a note: '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.' Below this, there's a section titled 'Associated IAM roles (0)' with a note: '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.' It includes a 'Set default' dropdown and a 'Manage IAM roles' button. A search bar says 'Search for associated IAM role by name, status, or role type'. Below the search bar is a table header with columns 'IAM roles', 'Status', and 'Role type'. The table body shows 'No resources' and 'No associated IAM roles'. There's a 'Associate IAM role' button at the bottom.

144. GO BACK TO IAM. CLICK CREATE ROLE.

The screenshot shows the AWS IAM Roles page. At the top, a success message says "Role GlueFullAccessRole created." Below it, the "Roles (10) Info" section provides a brief description of what an IAM role is. A search bar and navigation buttons are at the top right. The main table lists ten roles, each with a checkbox, a role name, and a "Trusted entities" column showing the identity provider or service. One role, "AWSReservedSSO_AWSRoleForRedshift", has a red error icon next to its name, indicating a failed deletion attempt.

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/AWSSC
AWSReservedSSO_AWSReadOnlyAccess_1ec8cf602ddfb0ec	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
AWSReservedSSO_AWSRoleForRedshift	AWS Service: redshift (Service-Linked Role)
AWSReservedSSO_AWSRoleForSSO	AWS Service: sso (Service-Linked Role)
AWSReservedSSO_AWSRoleForSupport	AWS Service: support (Service-Linked Role)

145. CHOOSE AWS SERVICE IN THE TRUSTED ENTITY TYPE.

This screenshot shows the "Step 1: Select trusted entity" screen of the IAM Create Role wizard. It includes a "Trusted entity type" section with four options: "AWS service" (selected), "AWS account", "SAML 2.0 federation", and "Custom trust policy". Below this is a "Use case" section with a note about allowing actions in the account. The sidebar on the left shows steps 2 and 3 of the wizard.

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

This 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 blue selection bar.

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.



149. GIVE THE NAME, REDSHIFTFULLACCESSROLE.

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.

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

Description
Add a short explanation for this role.

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		

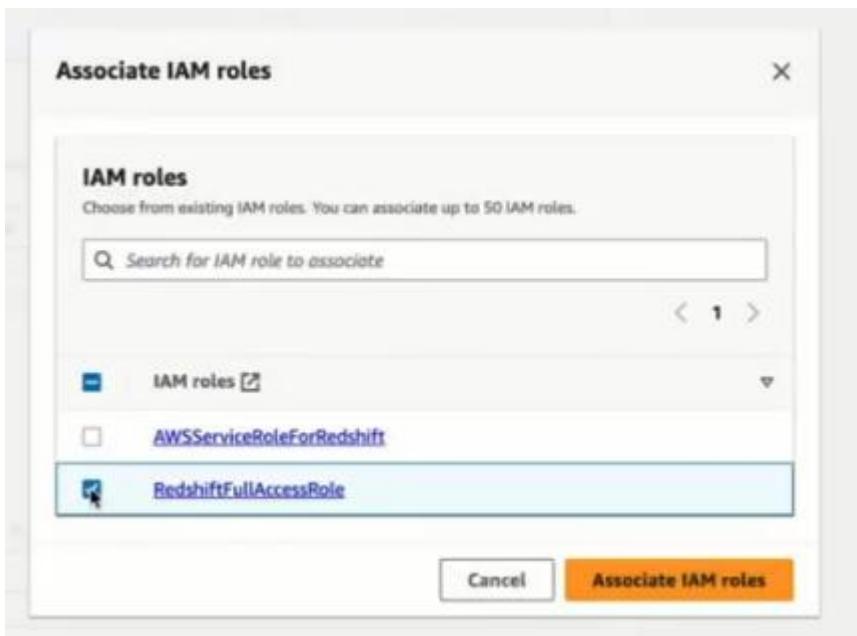
Associate IAM role

Encryption and security

AWS KMS encryption
AWS owned KMS key

Audit logging
Off

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



154. SELECT REDSHIFTFULLACCESSROLE.

The screenshot shows the 'Associated IAM roles' section. A callout box highlights the 'AmazonRedshiftAllCommandsFullAccess' policy, stating it includes permissions for COPY, UNLOAD, and SELECT statements. Below this, the 'Associated IAM roles (1/1)' section lists 'RedshiftFullAccessRole' with a status of 'Not applied'. The section also includes tabs for 'Set default' and 'Manage IAM roles', and a search bar. At the bottom, there's an 'Encryption and security' section with options for AWS KMS encryption and Audit logging.

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 (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** ▾

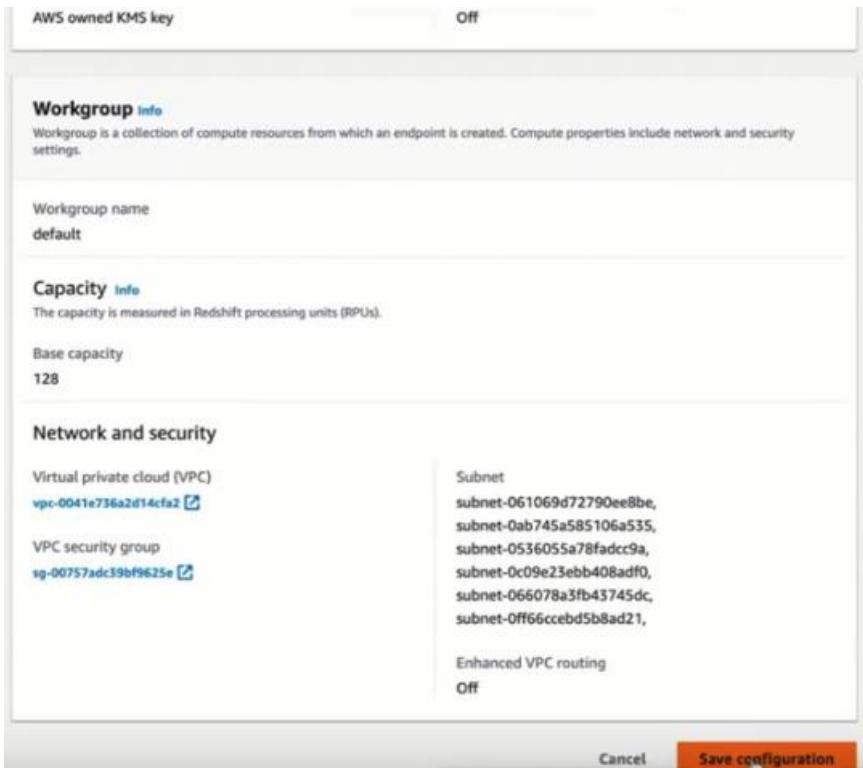
Q Search for associated IAM role by name, status, or role type

IAM roles	Status	Role type
RedshiftFullAccessRole	Not applied	--

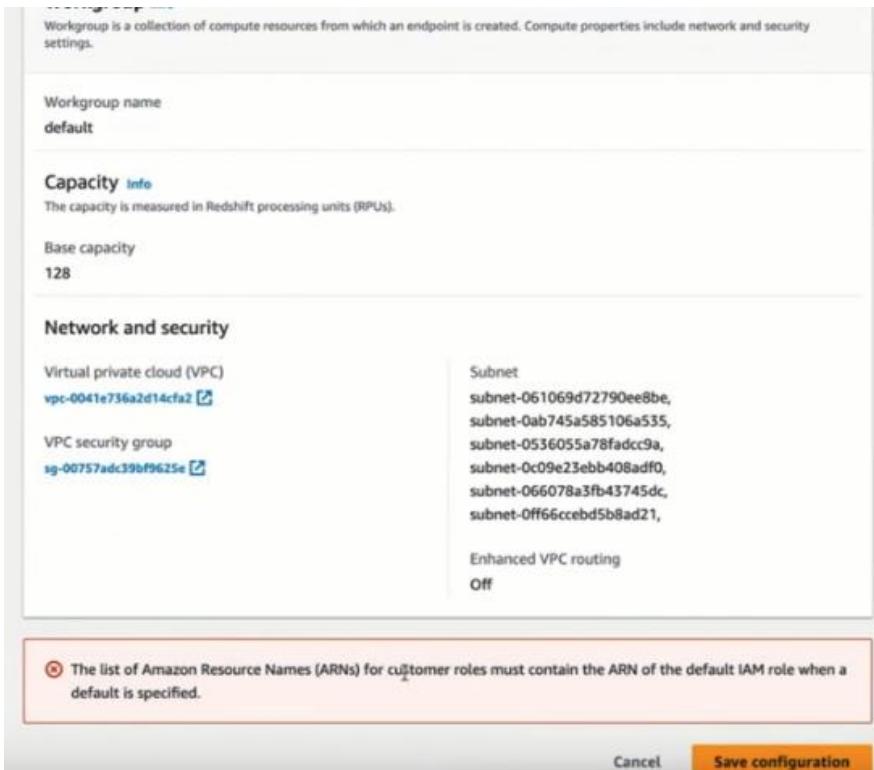
Encryption and security

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 a table header with columns for 'IAM roles', 'Status', and 'Role type'. The table body displays a message: 'No resources' and 'No associated IAM roles'. A button labeled 'Associate IAM role' is visible. At the bottom left, there's a collapsed section titled 'Security and encryption'.

160. SELECT REDSHIFTFULLACCESSROLE, THEN, ASSOCIATE IAM ROLES.

The screenshot shows the 'Associate IAM roles' dialog box. It has a heading 'IAM roles' and a note: 'Choose from existing IAM roles. You can associate up to 50 IAM roles.' Below is a search bar and a table with a single row. The row for 'RedshiftFullAccessRole' is selected, indicated by a blue background. At the bottom are 'Cancel' and 'Associate IAM roles' buttons.

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' result. The main table has columns for 'IAM roles', 'Status', and 'Role type'. One row is selected, showing 'RedshiftFullAccessRole' with 'Not applied' status and a dropdown arrow. Below the table, a section titled 'Security and encryption' contains a warning message: 'Your data is encrypted by default with an AWS owned key. To choose a different key, customize your encryption settings.' There is also an unchecked checkbox for 'Customize encryption settings (advanced)'. Another section, 'Audit logging', includes a link to 'Info' and a note about collecting logging information for the database. Under 'Export these logs:', there are three unchecked checkboxes for 'User log', 'Connection log', and 'User activity log'.

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

The screenshot shows the 'Workgroup' creation page. The title is 'Workgroup [Info](#)'. A description states: 'Workgroup is a collection of compute resources from which an endpoint is created. Compute properties include network and security settings.' The 'Workgroup name' section shows the value 'myworkgroup' in a text input field. A note below says: 'The name must be from 3-64 characters. Valid characters are a-z (lowercase only), 0-9 (numbers), and - (hyphen).'. The 'Capacity' section is expanded, showing the base capacity is set to 128 RPUs. A note says: '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.' Below this is a dropdown menu for 'Base capacity' with the value '128' selected. A note at the bottom of the dropdown says: 'Range must be 8-512 in increments of 8.'. The 'Network and security' section is partially visible at the bottom.

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.

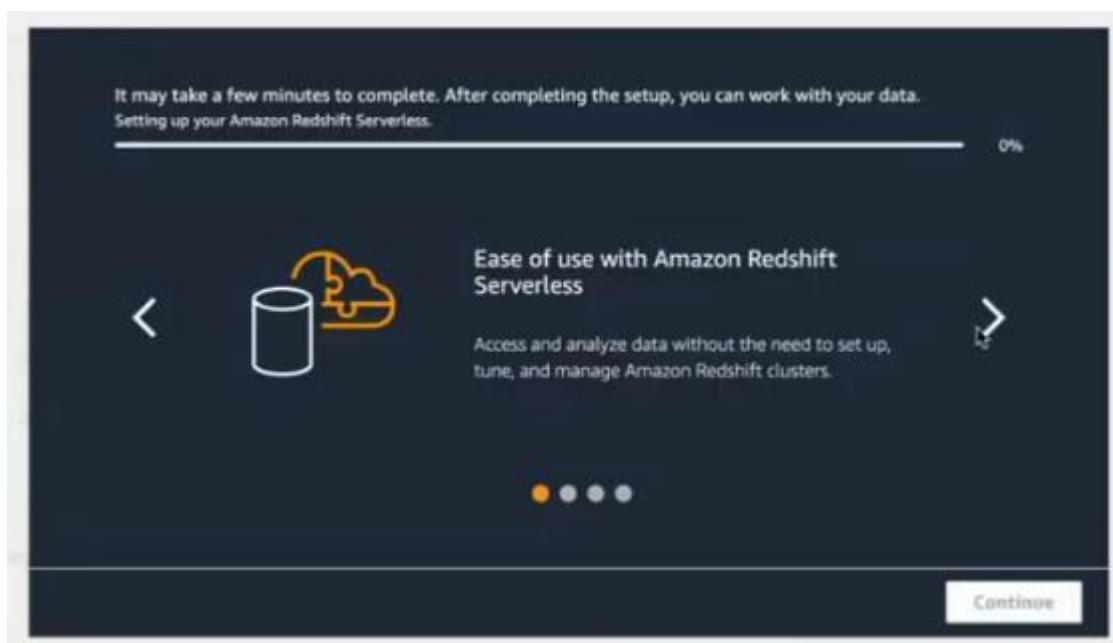
Turn on enhanced VPC routing

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.

The screenshot shows the Amazon Redshift Serverless dashboard. At the top, there is a green header bar with the message "Successfully setup Amazon Redshift Serverless" and a "Query data" button. Below the header, the page title is "Serverless dashboard". The main content area is divided into several sections:

- Namespace overview**: Shows statistics for Total snapshots (0), Datashares in my account (0), Datashares requiring authorization (0), Datashares from other accounts (0), and Datashares requiring association (0).
- Namespaces / Workgroups**: A table showing one namespace: mynamespace with Status Available, Workgroup myworkgroup, and Status Available.
- Queries metrics**: Displays workgroup metrics for myworkgroup over the last hour.
- Free trial**: Shows \$293.60 out of \$300.00 remaining, expiring on September 12, 2023.
- Alarms**: Shows 0 alarms.

The screenshot shows the configuration page for the mynamespace namespace. The top navigation bar includes "Amazon Redshift Serverless > Namespace configuration > mynamespace" and buttons for "Actions", "Change admin password", and "Query data".

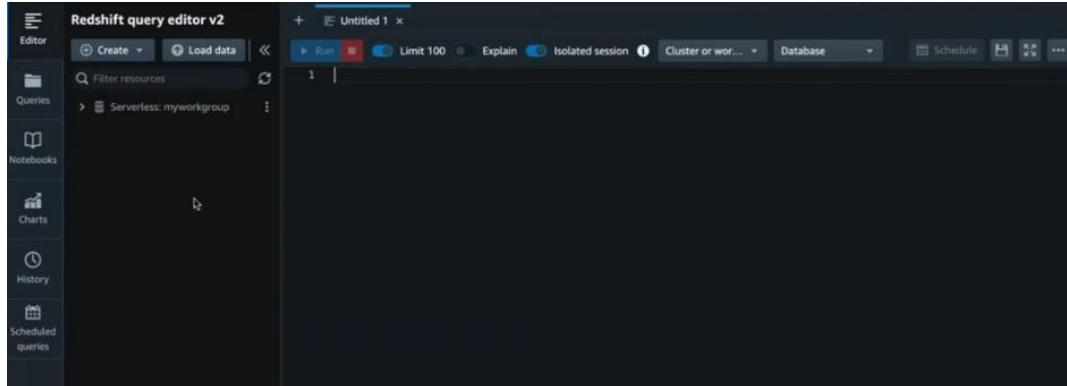
General information section:

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

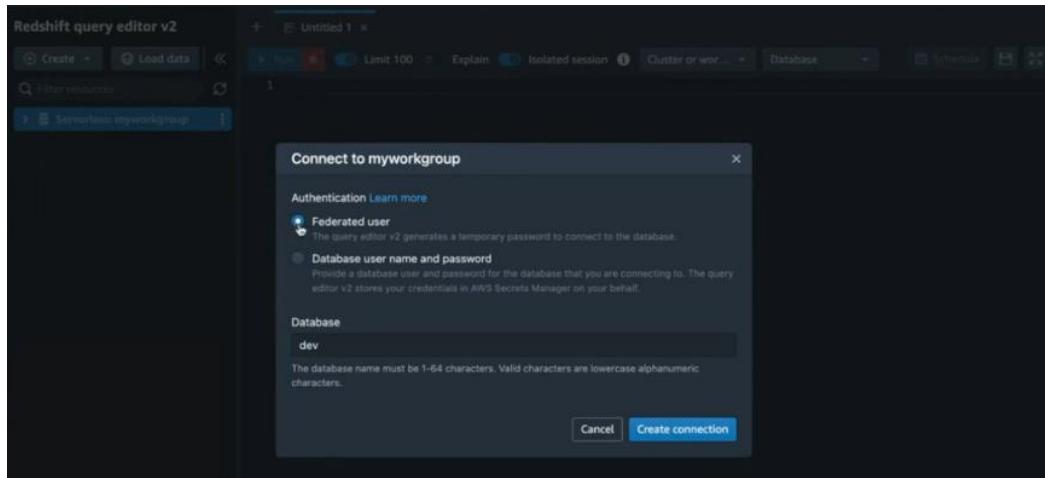
Workgroup tab is selected. The "Workgroup name" section shows a workgroup named "myworkgroup" with Status Available. There is also an "Actions" button.

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.

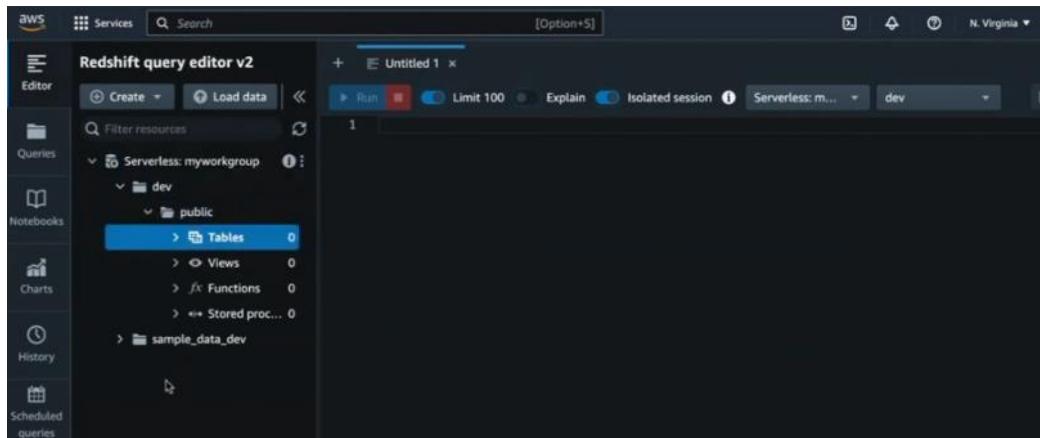
FIRST, WE ARE GOING TO CREATE THE CONNECTION. CLICK SERVERLESS MYWORKGROUP.



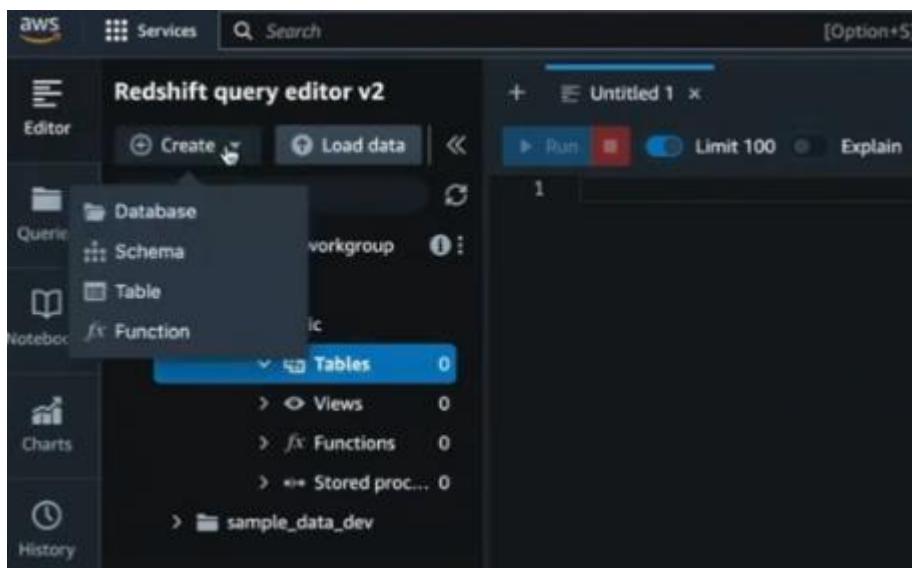
167. KEEP FEDERATED USER AND THE DEV IN THE DATABASE. THEN. CLICK CREATE CONNECTION.



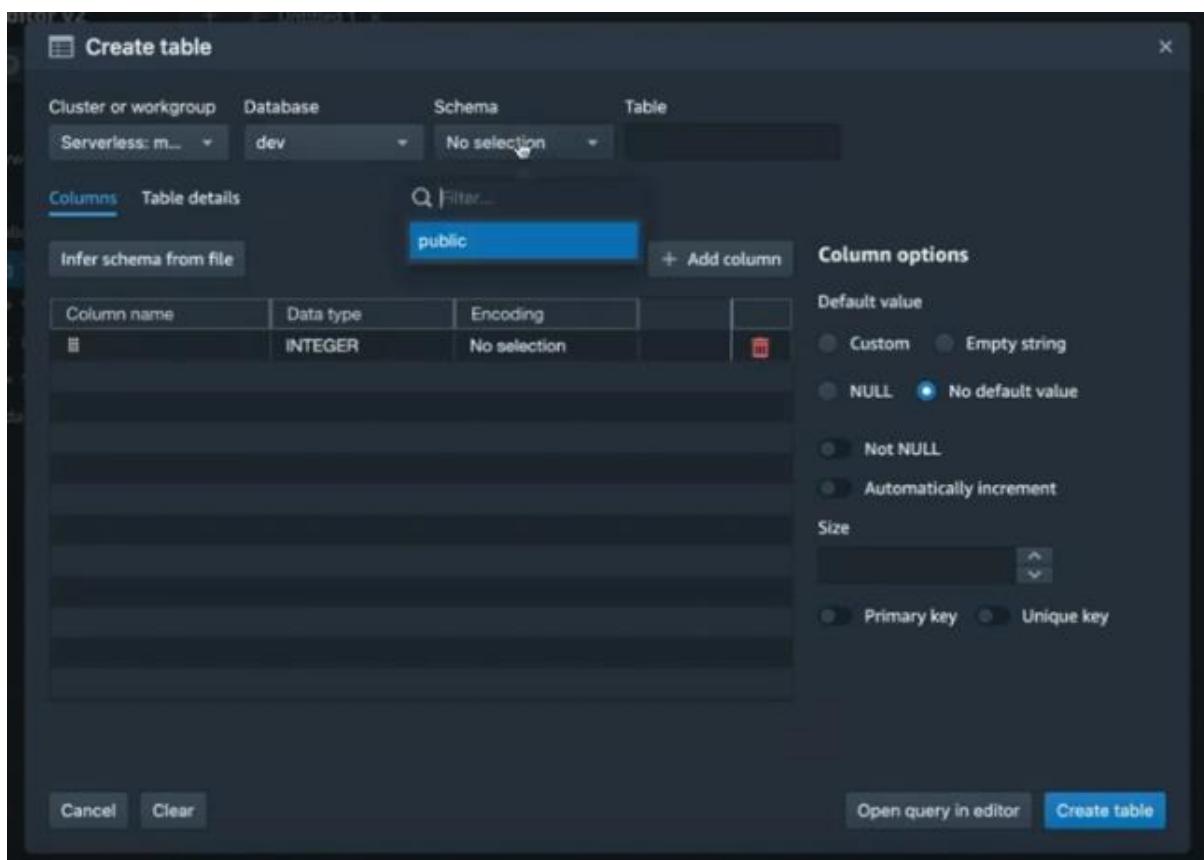
168. CONNECTION SUCCESFUL. EXPAND THE DEV, PUBLIC, AND TABLES FOLDERS. THEN, WE ARE GOING TO CREATE THE TABLE.



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 top navigation bar includes 'Cluster or workgroup' (Serverless: m...), 'Database' (dev), 'Schema' (public), and 'Table' (finaltable). The 'Columns' tab is selected. A single column is defined with the following details:

Column name	Data type	Encoding
id	INTEGER	No selection

On the right, 'Column options' are set to 'No default value' (selected), 'Not NULL' (unchecked), and 'Automatically increment' (unchecked). Below these are 'Size' and 'Primary key' (unchecked) settings. At the bottom are 'Cancel', 'Clear', 'Open query in editor...', and a blue 'Create table' button.

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. The top navigation bar includes 'Visual' (1), 'Script', 'Job details' (1), 'Runs', 'Data quality' (New), 'Schedules', and 'Version Control'. A message 'Unsaved job found' with a note 'We found an unsaved graph, do you wish to restore it?' is displayed. The main area shows a data flow graph with two sources: 'Data source - S3 bucket Amazon S3' and 'Data source - Data Catalog AWS Glue Data Catalog'. These feed into two 'Transform - ApplyMapping' nodes, which then join at a 'Transform - Join' node. The final target is 'Data target - Amazon Redshift Amazon Redshift'. To the right, the 'Output schema' tab of the 'Data target properties - Amazon Redshift' panel is selected, displaying the following schema:

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' dialog box in the AWS Redshift console. The 'Table' tab is selected, showing the following configuration:

- Cluster or workgroup:** Serverless: m...
- Database:** dev
- Schema:** public
- Table:** finaltable

The 'Columns' tab is active, displaying the schema:

Column name	Data type	Encoding	Action
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:** Radio buttons for 'Custom', 'Empty string', 'NULL' (selected), and 'No default value'.
- Size:** A dropdown menu.
- Primary key:** A radio button.
- Unique key:** A radio button.

At the bottom of the dialog box are 'Cancel', 'Clear', 'Open query in editor', and 'Create table' buttons.

174. FINALTABLE IS NOW CREATED SUCCESSFULLY.

The screenshot shows the Redshift query editor interface. The sidebar shows the database structure:

- Editor
- Queries
- Notebooks
- Charts
- History

The main area displays the 'Redshift query editor v2' interface with the following details:

- Services:** AWS
- Search:** Search bar
- Actions:** Create, Load data, Run, Explain, Isolated session, Serverless: m..., dev
- Message:** A green success message: "finaltable table is created successfully."
- Table Structure:** Shows a tree view of tables under 'public' schema:
 - Tables: finaltable (1)
 - Views: 0
 - Functions: 0
 - Stored proc...: 0
- Sample Data:** 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.

The screenshot shows the 'Data target properties - Amazon Redshift' configuration screen. It includes fields for 'Name' (Amazon Redshift), 'Node parents' (Choose one or more parent node), 'Join' (Join - Transform), 'Redshift access type' (selected: Direct data connection - recommended), 'Redshift connection' (Choose your Amazon Redshift connection dropdown), and 'Connection' and 'Database' dropdowns.

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

The screenshot shows the 'Choose your Amazon Redshift connection' dropdown menu. It lists 'RedshiftConnection' (selected) and 'Choose your Amazon Redshift connection' (refresh button).

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

The screenshot shows the 'Schema' and 'Table' configuration sections. Under 'Schema', it shows 'Choose your Amazon Redshift schema.' with 'public' selected. Under 'Table', it shows 'Search and enter the name of the source Amazon Redshift table.' with 'finaltable' entered.

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

Schema
Choose your Amazon Redshift schema.
 C

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

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. We found an unsaved graph, do you wish to restore it?' with 'Restore' and 'X' buttons. Below the message, the script content is displayed. The script is a Python script for an AWS Glue job, starting with imports from awsglue.transforms, awsglue.utils, pyspark.context, awsglue.context, awsglue.job, and awsglue. It then defines arguments, initializes a SparkContext, creates a GlueContext, and starts a Job. The script includes comments for Amazon S3 and AWS Glue Data Catalog connections. Buttons for 'Generate classic script', 'Download script', and 'Edit script' are visible at the top right of the code editor.

```
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",
```

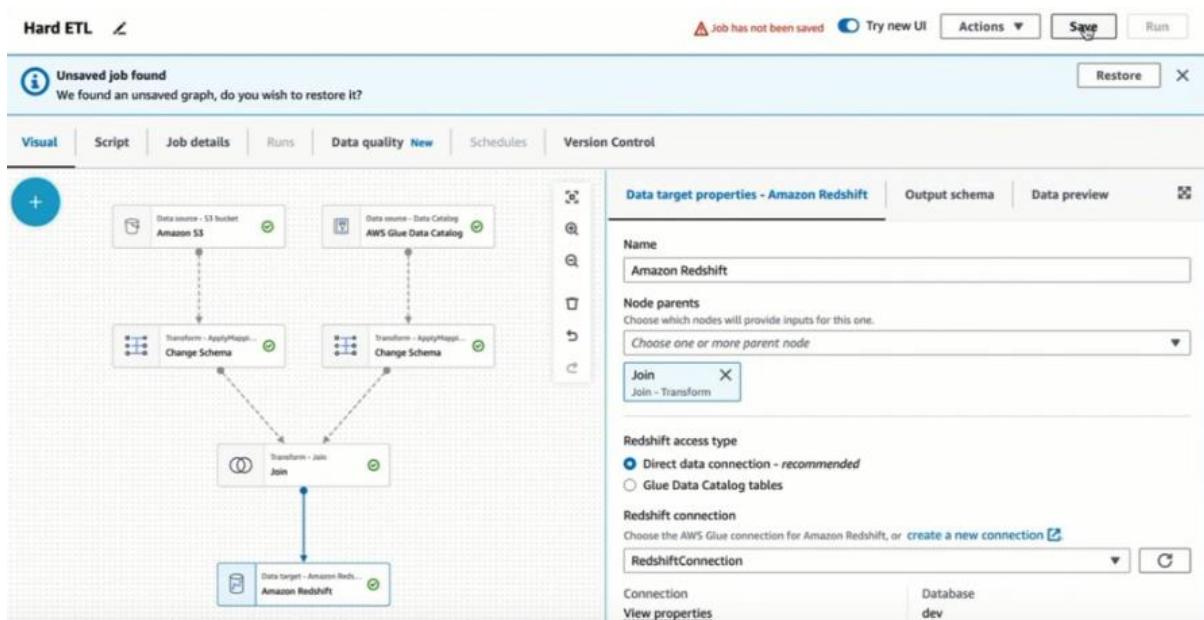
182. GOING BACK TO THE VISUAL, AT THE TOP, CLICK SAVE.

The screenshot shows the AWS Glue visual editor interface. At the top, there is a warning message "Job has not been saved" and a "Try new UI" button. Below this is a toolbar with "Actions" (containing "Run", "Save", and "Run" buttons), "Restore" (with a "X" button), and other options like "Run with parameters".

The main area is titled "Data target properties - Amazon Redshift". It includes fields for "Name" (set to "Amazon Redshift") and "Node parents" (a dropdown menu labeled "Choose one or more parent node").

On the left, there is a "Join" node with a sub-node "Join - Transform".

Under "Redshift access type", the "Direct data connection - recommended" option is selected. Under "Redshift connection", a dropdown menu shows "RedshiftConnection" and a "Database" dropdown set to "dev".



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', '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 is 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. A 'Run job' button is visible at the top right of the 'Your jobs' section.

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.

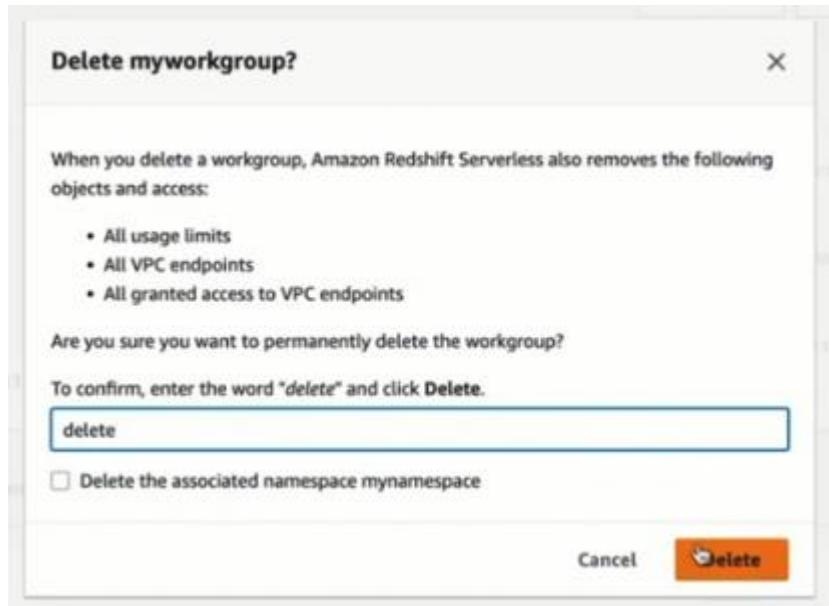
CONGRATULATIONS, YOU DID THE WHOLE ETL PROCESS SUCCESSFULLY!

The screenshot shows the 'Hard ETL' job details page. At the top, a green banner says 'Successfully updated job' and 'Successfully updated job Hard ETL. To run the job choose the Run Job button.' Below the banner are tabs for Visual, Script, Job details, Runs, Data quality, Schedules, and Version Control. The 'Visual' tab is selected. On the left, there are two data source configurations: 'Data source - S3 bucket Amazon S3' and 'Data source - Data Catalog AWS Glue Data Catalog'. On the right, there are tabs for 'Data target properties - Amazon Redshift', 'Output schema', and 'Data preview'. The 'Data target properties' tab is selected. A 'Run' button is located at the top right of the page.

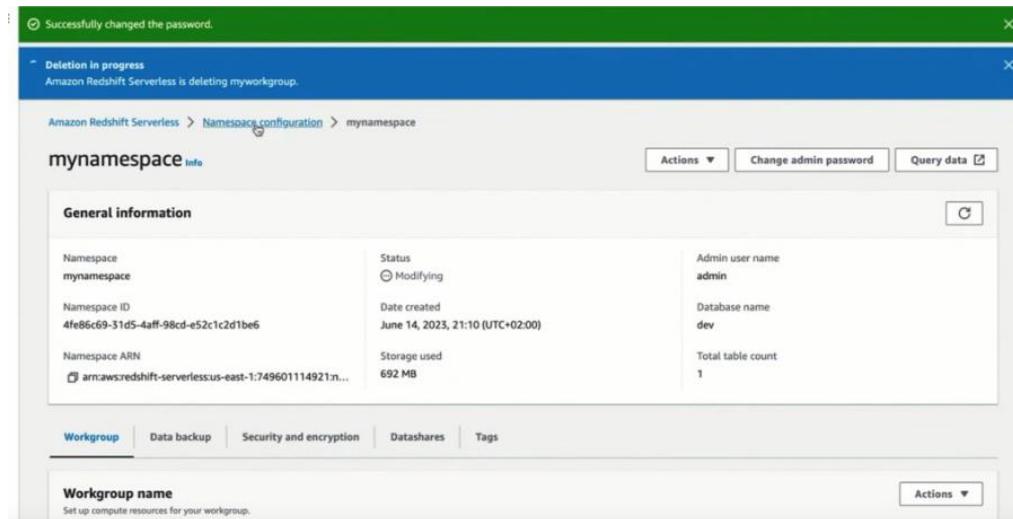
185. TO AVOID CHARGES, WE NEED TO DELETE EVERYTHING. IN DOING THE CLEAN UP, FIRST, GO TO REDSHIFT, CLICK WORKGROUP NAME. CLICK ACTIONS, THEN, DELETE.

The screenshot shows the 'Workgroup' management page. It has tabs for Workgroup, Data backup, Security and encryption, Datashares, and Tags. The 'Workgroup' tab is selected. It displays a table with one row: 'Workgroup myworkgroup' and 'Status Available'. At the top right, there are 'Actions' and 'Delete' buttons.

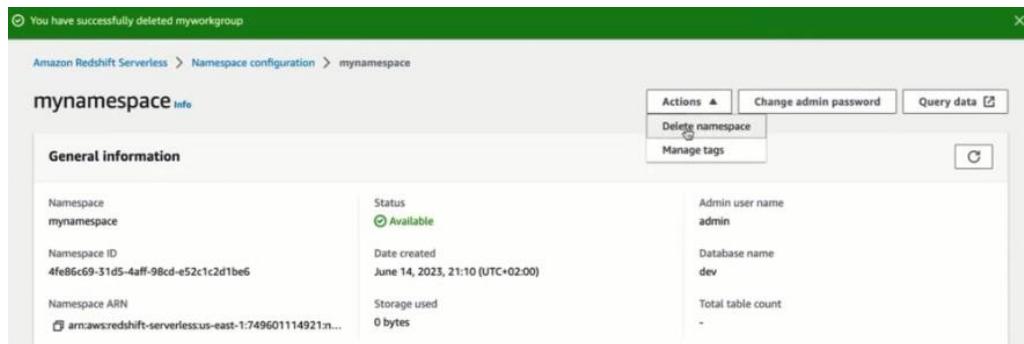
ENTER DELETE. THEN, CLICK DELETE.



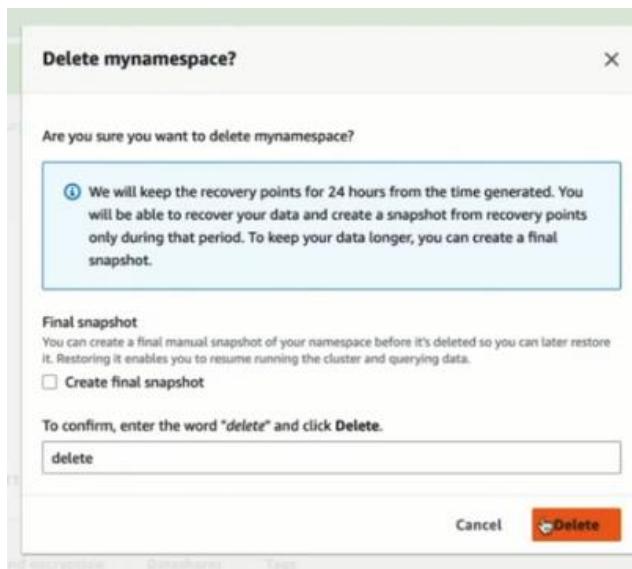
186. THE NEXT TO DELETE IS THE NAME SPACE. GO TO THE NAMESPACE CONFIGURATION.



CLICK THE MYNAMESPACE, ACTIONS, THEN DELETE.



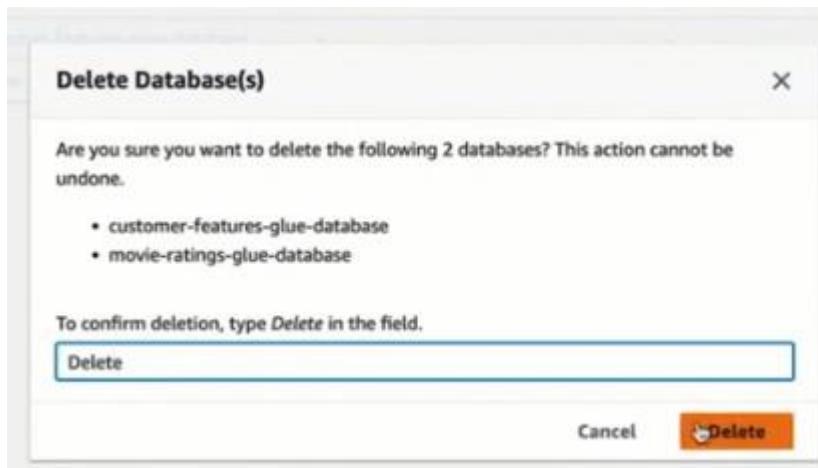
TYPE DELETE. CLICK DELETE.



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 tables". Under "Data Catalog", the "Databases" section is selected, showing a list of two databases: "customer-features-glue-database" and "movie-ratings-glue-database". Both have a creation date of June 14, 2023.

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. 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 "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 (highlighted), Table changes ..., View details, Resume schedule, Pause schedule, and Stop run.

ENTER DELETE, CLICK DELETE.

The screenshot shows a modal dialog titled "Delete crawlers". The message inside reads: "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 asks: "To confirm deletion, type Delete in the field." The word "Delete" is typed into this field. At the bottom right of the dialog are two buttons: "Cancel" and a highlighted "Delete" button.

SUCCESSFULLY DELETED CRAWLERS.

The following crawlers are now deleted: "movie-ratings-crawler", "customer-features-crawler"

AWS Glue > Crawlers

Crawlers

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.

Name	State	Schedule	Last run	Last run time	Log	Table changes
No resources No resources to display.						

Create crawler

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

AWS Glue > Connectors

Connectors

Marketplace connectors

Custom connectors

Actions

Connections (0) Info

Actions ▲ Create connection Create job

View details by property

Delete Edit Test connection

Name Type Last modified

Jun 14, 2023

Delete RDSConnection

Deleting RDSConnection will permanently remove it as a source in Glue Studio. You will need to update your jobs accordingly.

⌚ Delete in progress.

Cancel Delete

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), Visual ETL, Notebooks, Job run monitoring, Data Catalog tables, Data connections, Workflows (orchestration), Data Catalog (selected), DataBases, Tables, Stream schema registries, Schemas, Connections, Crawlers, Classifiers, Catalog settings, and Data Integration and ETL (selected). Under ETL jobs, there are options for Visual ETL, Notebooks, and Job run monitoring.

The main area displays the "Create job" screen. It offers 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" (JSON, CSV, or Parquet files stored in S3) and the "Target" is set to "Amazon S3" (S3 bucket by specifying a bucket path as the data target).

The "Your jobs (1)" section lists a single job named "Hard ETL" (Glue ETL). The Actions menu for this job includes "Edit job", "Clone job", "Schedule job", "Delete job(s)" (which is highlighted with a blue border), and "Reset job bookmark".

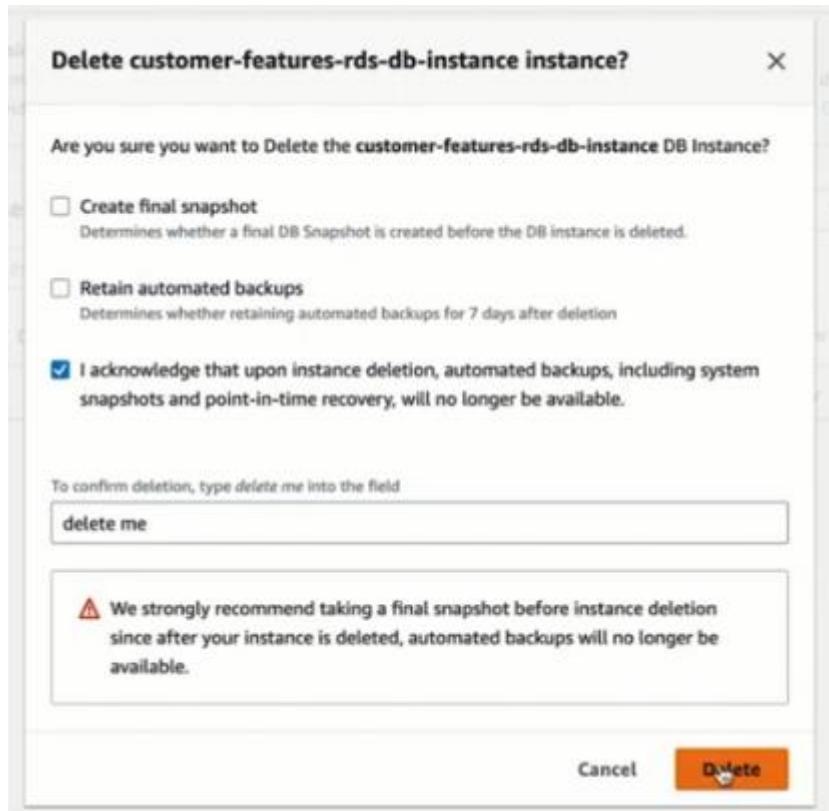
A modal window titled "Delete job(s)" asks, "Are you sure you want to delete the following 1 job(s)? This action cannot be undone." It lists the job "Hard ETL" and provides "Cancel" and "Delete" buttons.

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 includes links for 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 (1), and Certificate update.

The main content area shows a table for "Databases (1)". The table has columns for DB identifier, Status, Role, and Engine. A single row is selected for "customer-features-rds-db-instance" (Status: Available, Instance: MySQL). To the right of the table is a "Actions" menu with options: "Group resources", "Modify", "Restore from S3", "Create database", "Quick Actions - New", "Convert to Multi-AZ deployment", "Stop temporarily", "Reboot", "Delete" (which is highlighted with a blue border), "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.

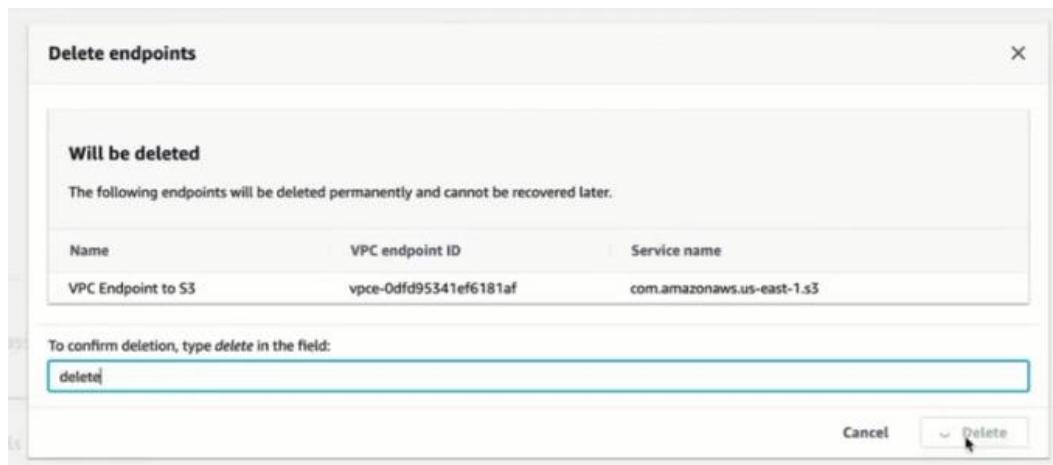
The screenshot shows the AWS VPC dashboard under the "Endpoints" section. It displays a table with one endpoint entry:

Name	VPC endpoint ID	VPC ID	Service
VPC Endpoint to S3	vpce-0dfd95341ef6181af	vpc-0041e736a2d14cfa2	com.amazonaws.s3

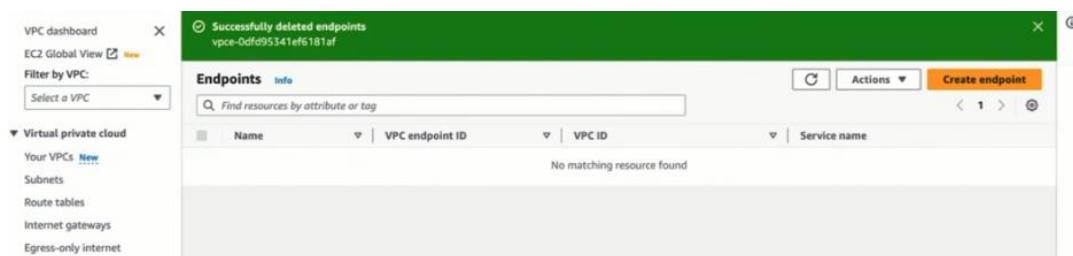
On the right side, there is a "Actions" menu with the following options:

- View details
- Manage subnets
- Manage security groups
- Manage route tables
- Manage policy
- Modify private DNS name
- Modify endpoint settings
- Manage tags
- Delete VPC endpoints

At the bottom, there is a "Details" section for the selected endpoint.



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.

New EC2 Experience

Network interfaces (1/1)

Security group IDs = sg-05eae22adcf0dce56

Name	Network interface ID	Subnet ID	VPC ID
-	eni-0d7a9d40c06fa53a9	subnet-0536055a78fadcc9a	vpc-0041e736a2d14cfa2

Network interface: eni-0d7a9d40c06fa53a9

Details Flow logs Tags

Network interface details

Network interface ID eni-0d7a9d40c06fa53a9	Name -	Description RDSNetworkInterface
Network interface status In-use	Interface type Elastic network interface	Security groups sg-05eae22adcf0dce56 (SG-Open-MySQL) sg-00757adc39bf9625e (default)
VPC ID vpc-0041e736a2d14cfa2	Subnet ID subnet-0536055a78fadcc9a	Availability Zone us-east-1c

Delete network interface

Are you sure that you want to delete the following network interface?

- eni-0d7a9d40c06fa53a9

Cancel **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, with a tooltip message: "Inbound security group rules successfully modified on security group (sg-05eae22adcf0dce56 | SG-Open-MySQL)".

The third screenshot shows a confirmation dialog titled "Delete security groups" asking if you are sure you want to delete the selected security group (sg-05eae22adcf0dce56 - SG-Open-MySQL). It includes a "Cancel" button and a prominent orange "Delete" button.

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

The screenshot shows the AWS S3 buckets interface. On the left, there's a sidebar with navigation links like "Buckets", "Storage Lens", and "AWS Marketplace for S3".

The main area displays an "Account snapshot" with metrics like Total storage (63.0 B), Object count (2), and Average object size (31.5 B). It also includes a note about enabling advanced metrics in the "default-account-dashboard" configuration.

A table lists the buckets:

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)

The "recommender-system-8730872805" bucket is selected, and the "Empty" button in the top right of the table row is highlighted.

TYPE PERMANENTLY DELETE. CLICK EMPTY.

The screenshot shows the 'Empty bucket' dialog box. At the top, the path is shown as 'Amazon S3 > Buckets > recommender-system-8730872805 > Empty bucket'. Below this, the title 'Empty bucket' has an 'Info' link. A warning box contains the following text:

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

A 'Learn more' link is also present. Another box below suggests creating a lifecycle rule:

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

[Go to lifecycle rule configuration](#)

At the bottom, a confirmation message reads: 'Permanently delete all objects in bucket "recommender-system-8730872805"?'. Below it, a note says 'To confirm deletion, type *permanently delete* in the text input field.' A text input field contains the text 'permanently delete'. Two buttons at the bottom are 'Cancel' and a large orange 'Empty' button.

CLICK EXIT.

The screenshot shows the 'Empty bucket: status' dialog box. At the top, a green header bar indicates 'Successfully emptied bucket "recommender-system-8730872805"'. Below this, a note says 'View details below. If you want to delete this bucket, use the delete bucket configuration.' The main area is titled 'Empty bucket: status' with a 'Cancel' button and a red 'Exit' button. A note states 'The details below are no longer available after you navigate away from this page.' A 'Summary' section shows the following data:

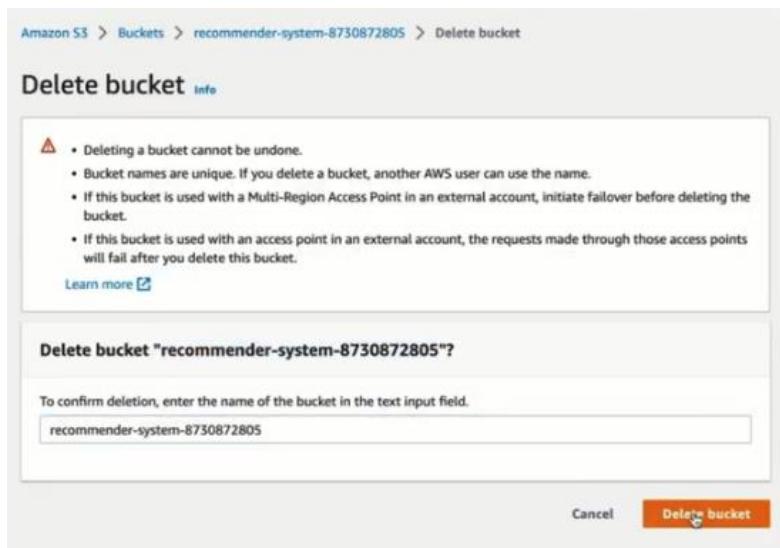
Source	Successfully deleted	Failed to delete
s3://recommender-system-8730872805	Successfully deleted 1 object, 2.3 MB	Failed to delete 0 objects

196. SELECT THE BUCKET. CLICK DELETE.

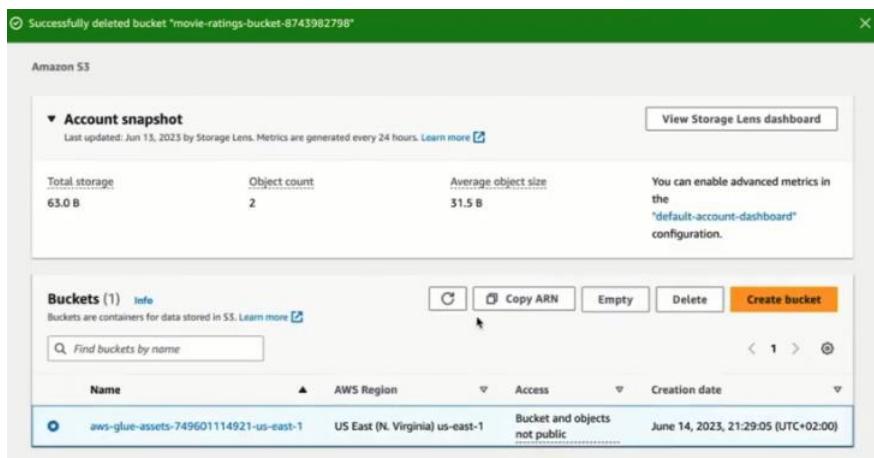
The screenshot shows the 'Buckets (1)' list page. At the top, there are buttons for 'Copy ARN', 'Empty', 'Delete', and 'Create bucket'. A search bar 'Find buckets by name' is also present. The main table lists one bucket:

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)

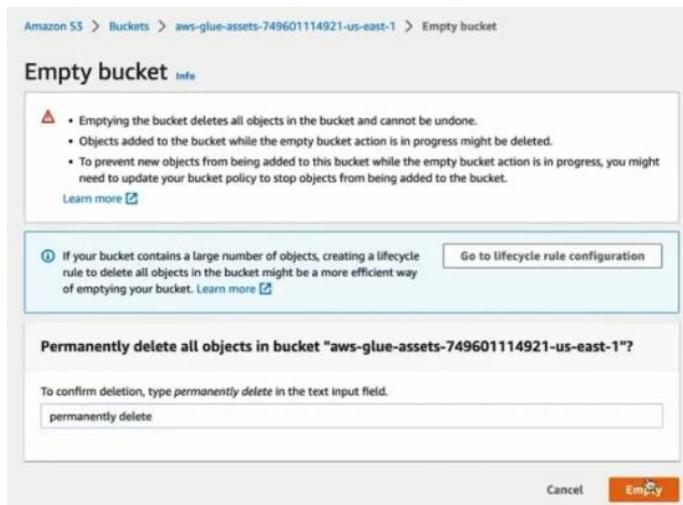
ENTER THE NAME OF THE BUCKET IN THE TEXT INPUT FIELD. CLICK DELETE BUCKET.



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 confirming 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 tagline, and a 'Create a bucket' button.

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, and Roles. Under Roles, 'GlueFullAccessRole' and 'RedshiftFullAccessRole' are selected. The main pane lists various roles with their details and service providers. At the bottom right of the main pane, there's a 'Manage' button.

Role Name	AWS Service
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)
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
rds-monitoring-role	AWS Service: monitoring.rds
RedshiftFullAccessRole	AWS Service: redshift

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 inline policies and attached instance profiles. It lists the two selected roles: 'GlueFullAccessRole' and 'RedshiftFullAccessRole'. A note at the bottom says recent activity is usually within 4 hours. The user is prompted to enter 'delete' in a text input field to confirm. The 'Delete' button is highlighted in blue.

ROLES SUCCESSFULLY DELETED. CLEAN UP ALL DONE. CONGRATULATIONS!

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