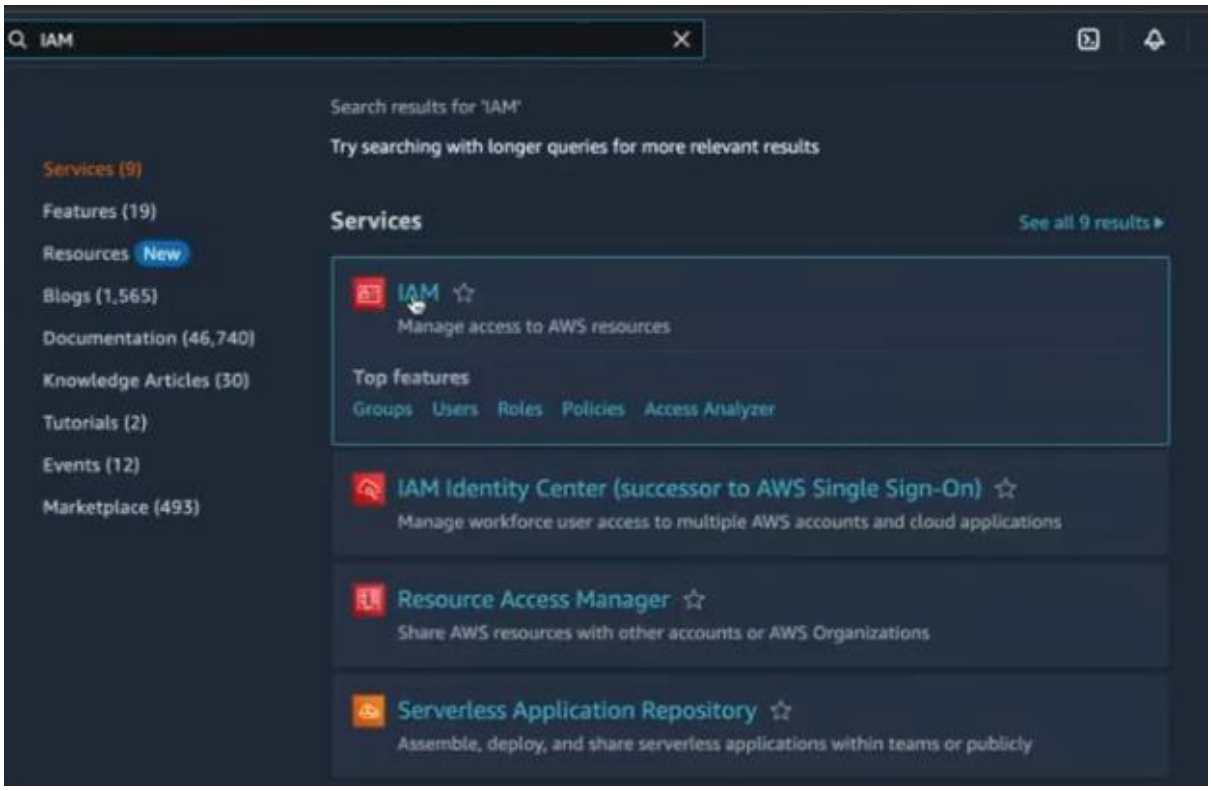
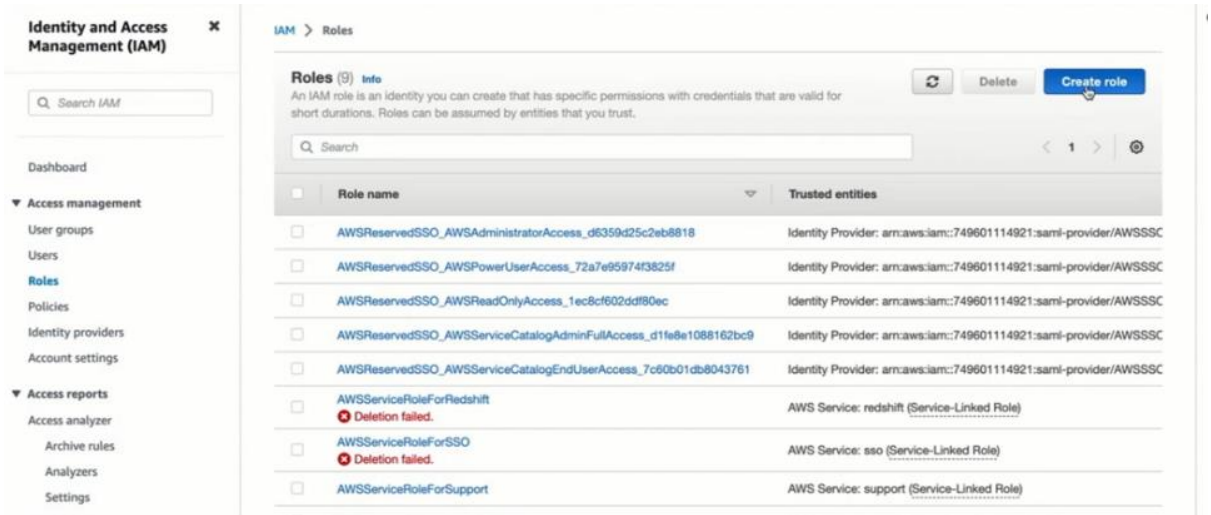
# AWS GLUE FOR HARD ETL (EXTRACT, TRANSFORM, AND LOAD) ACTIVITY

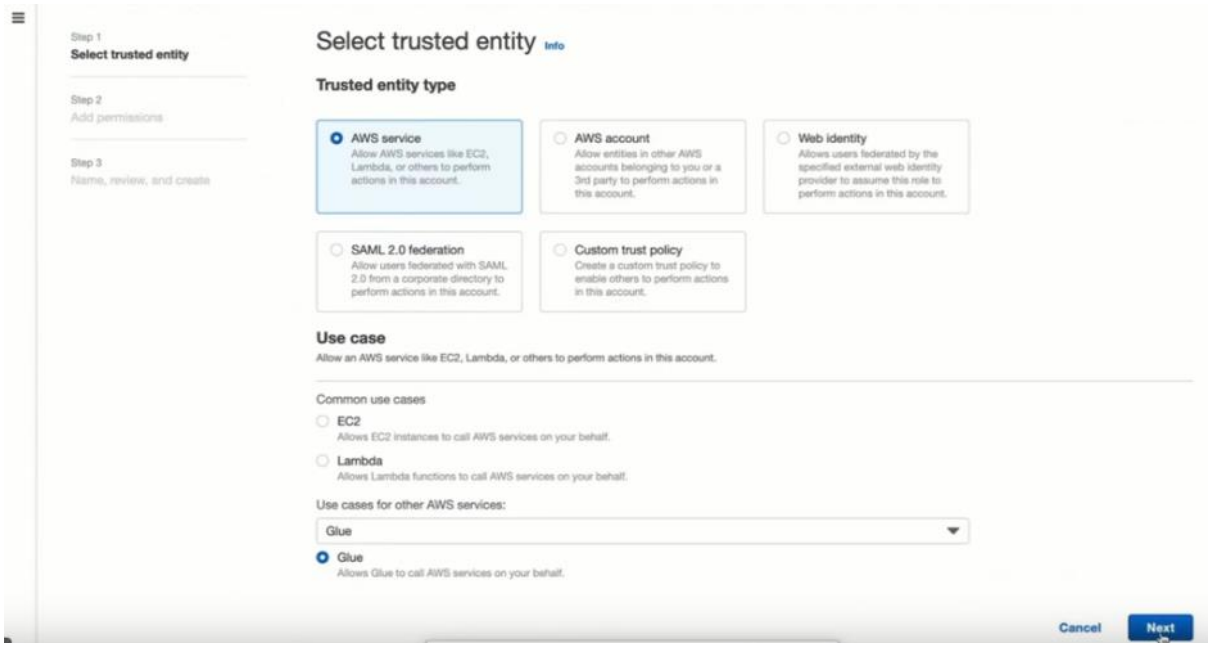
1. **FIRST STEP IS TO CREATE THE ROLE FOR OUR AWS GLUE IN ORDER TO GET FULL PERMISSION. IN THE SERVICES, CHOOSE IAM.**



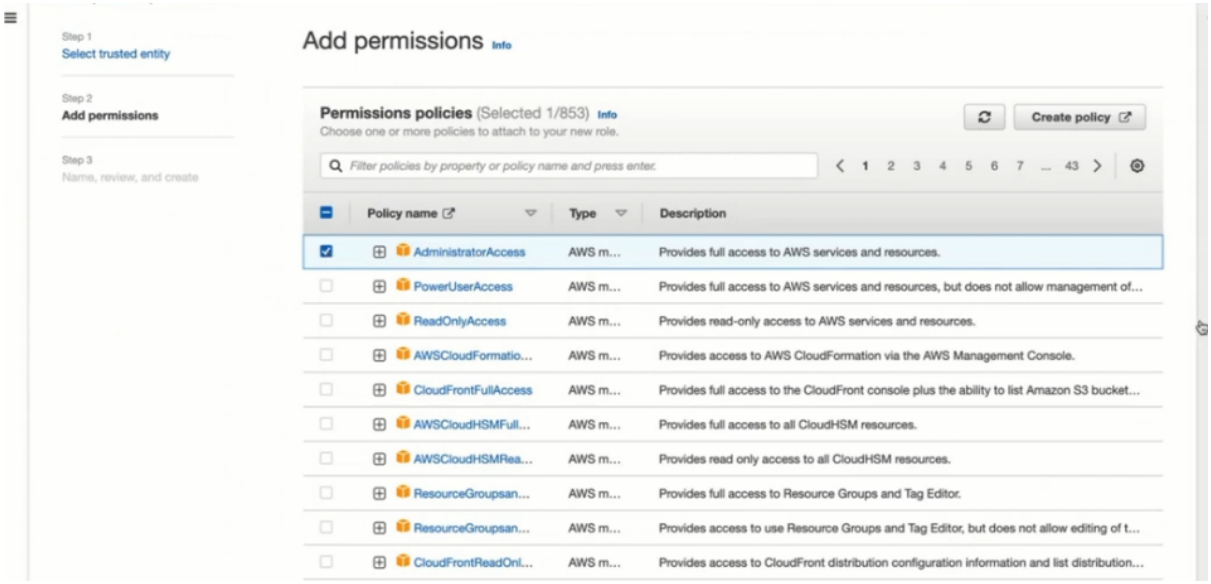

2. **CLICK ROLES, THEN., CREATE ROLE.**



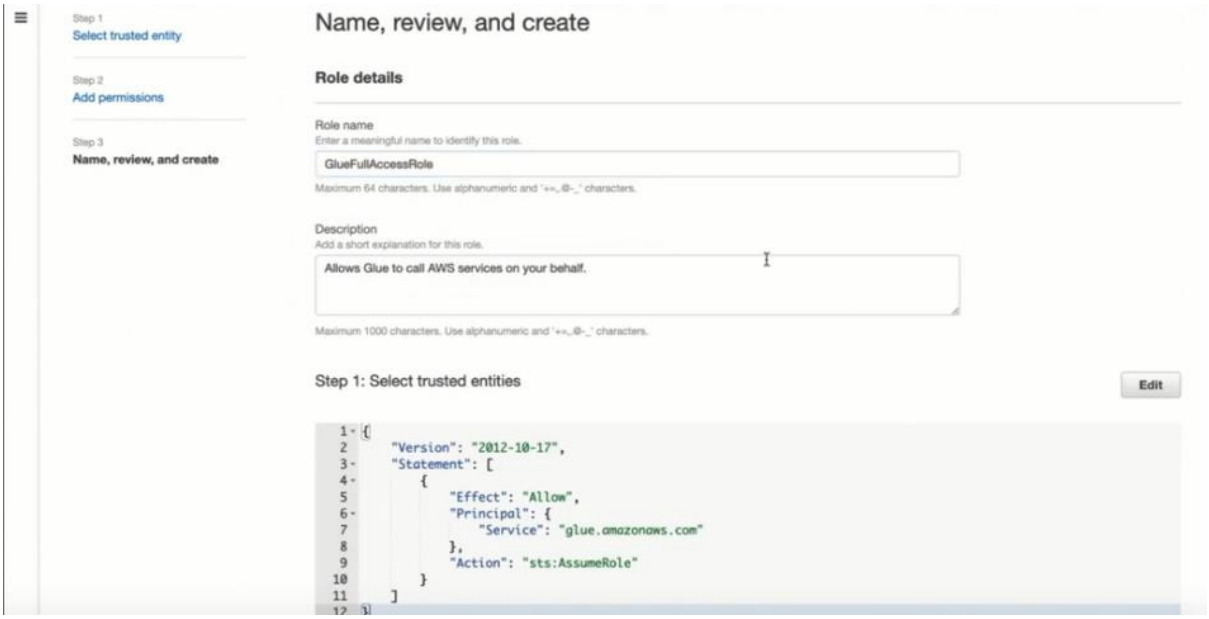

3. **CHOOSE AWS SERVICE IN THE TRUSTED ENTITY TYPE. AND IN THE USE CASES FOR OTHER AWS SERVICES, CHOOSE AND SELECT GLUE. CLICK, NEXT.**




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

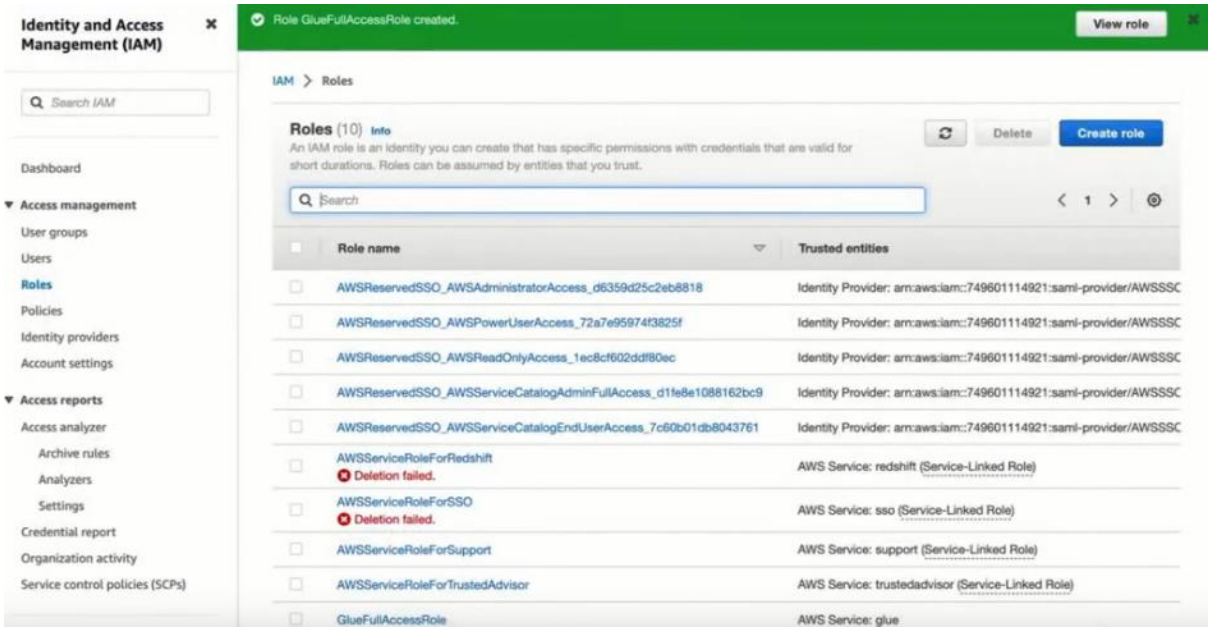



5. **IN THE ROLE NAME, WE WILL CALL THIS, GLUEFULLACCESSROLE. ON THE DESCRIPTION, IT WILL BE, ALLOWS GLUE TO CALL AWS SERVICESON YOUR BEHALF. SCROOL DOWN, THEN, CLICK CREATE ROLE.**

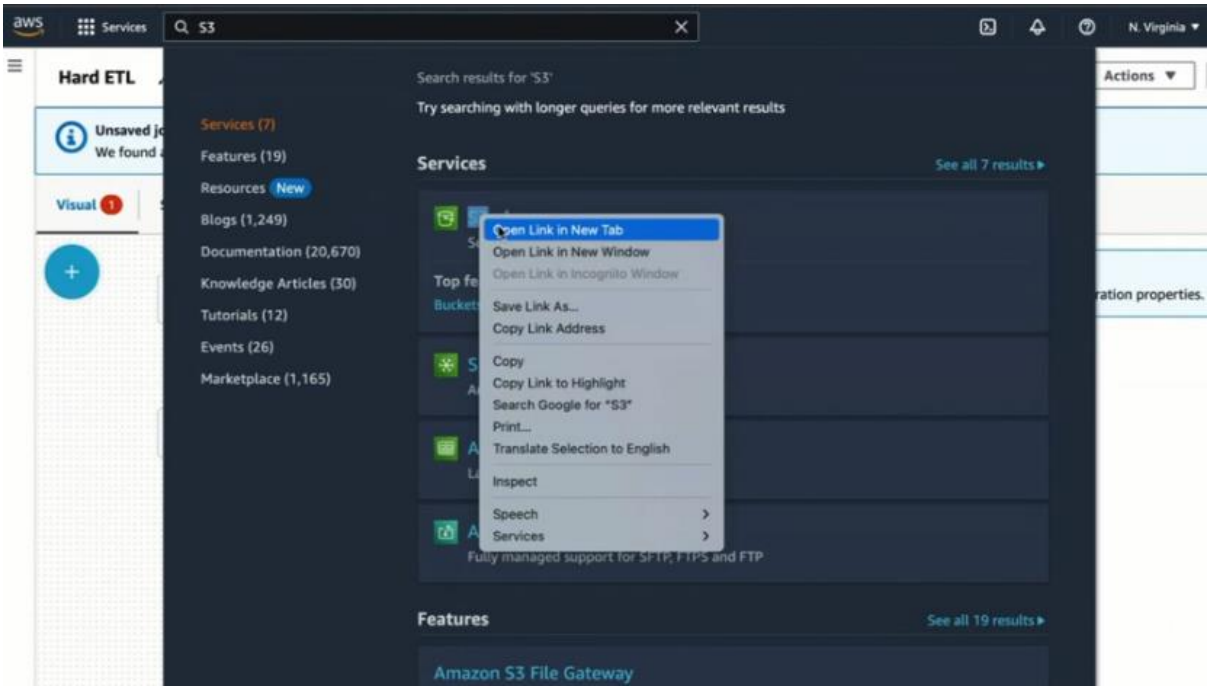




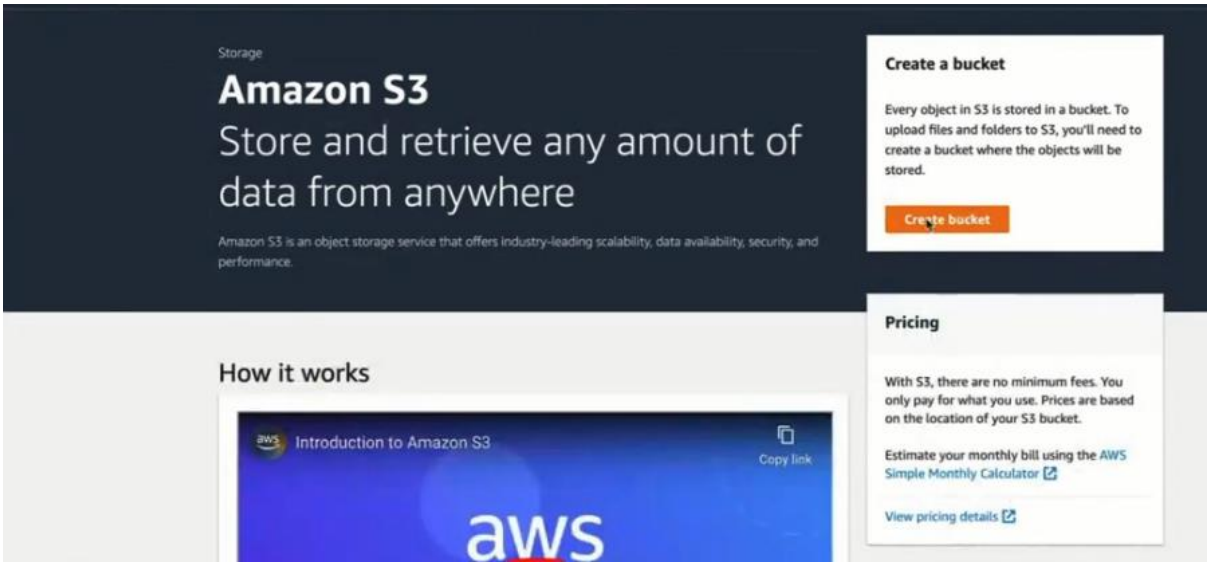
6. **THIS WILL NOW BE OUR GLUE ACCESS ROLE.**



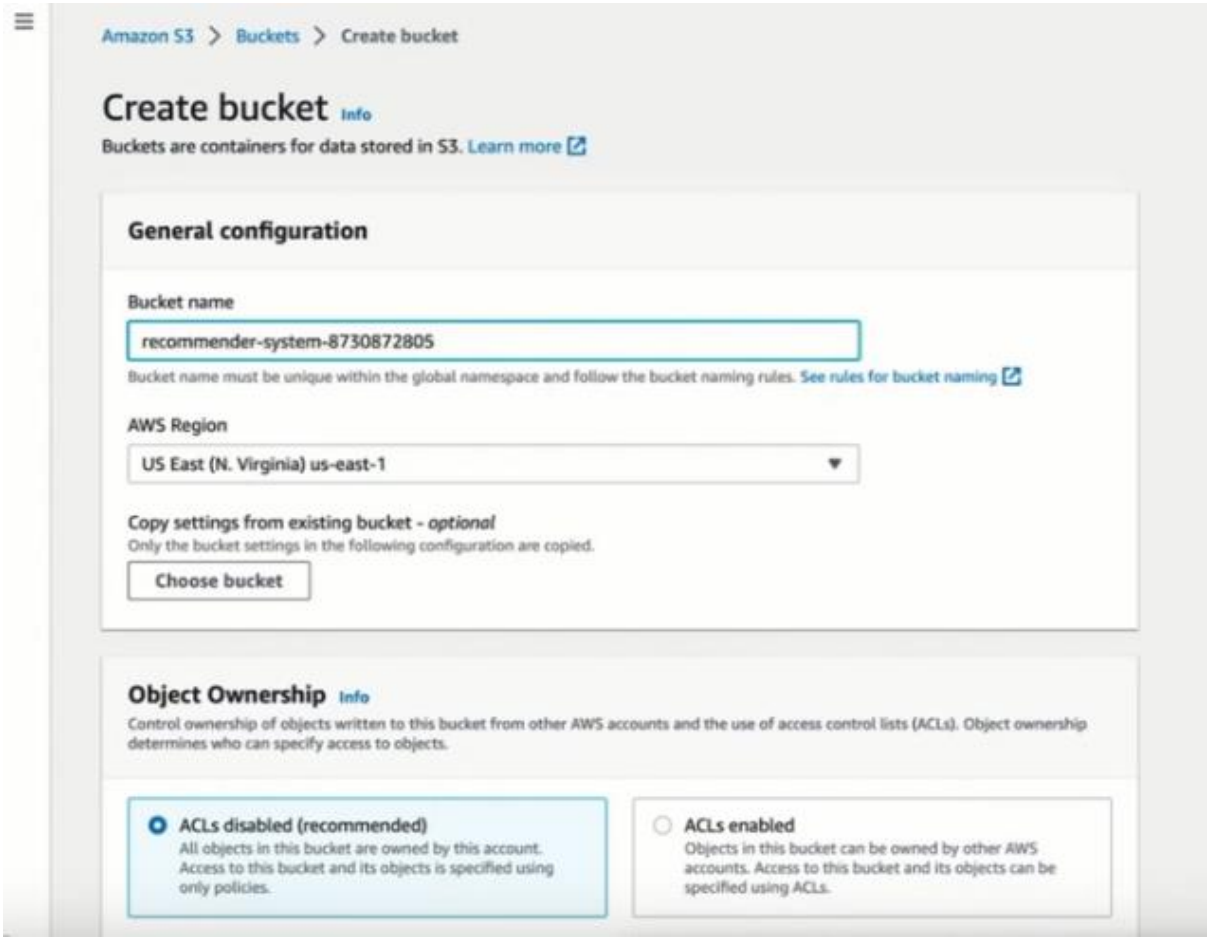

**7. GO TO SERVICES, CLICK AMAZON S3 TO CRREATE OUR BUCKET.**



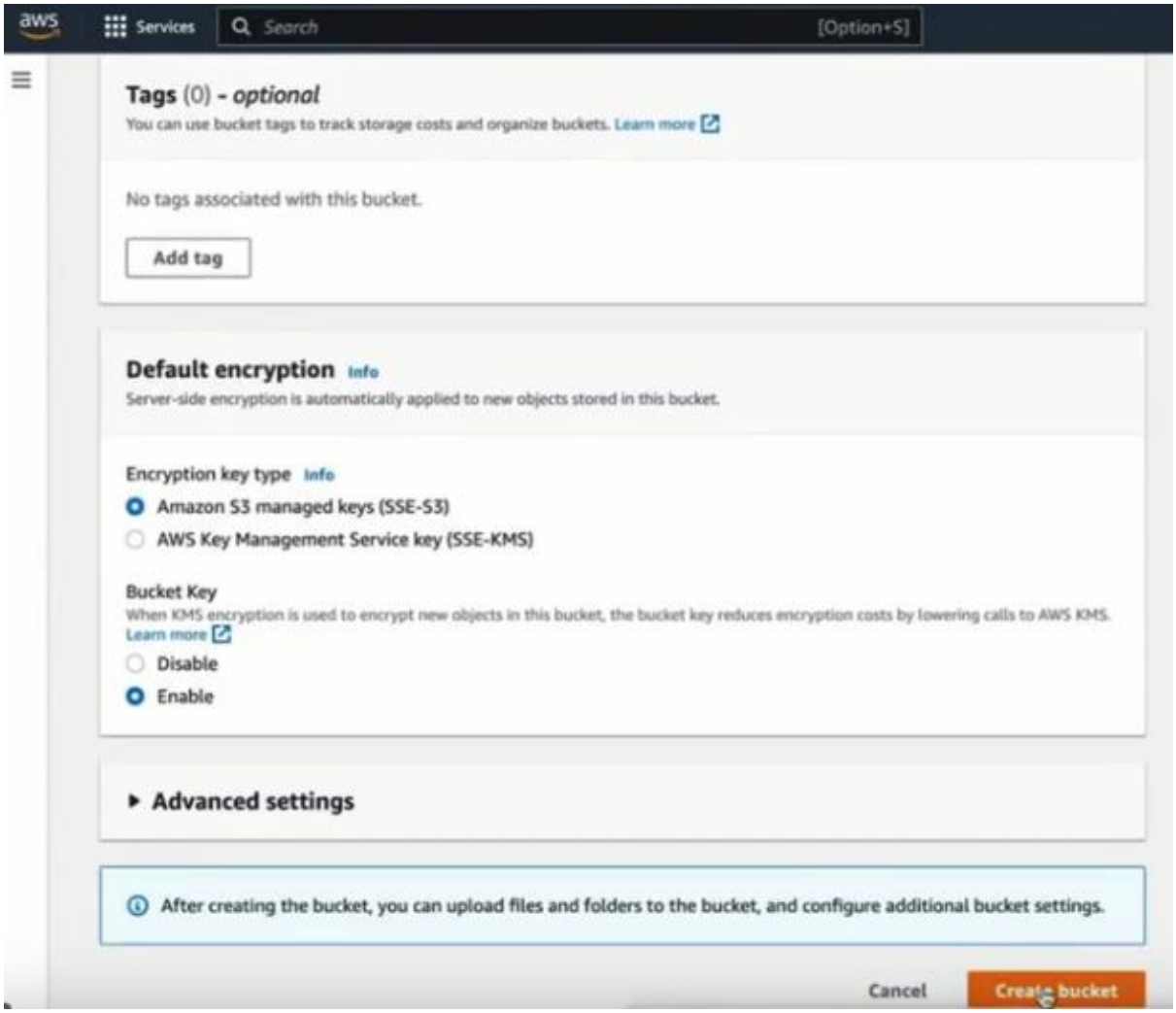

**CLICK CREATE BUCKET.**



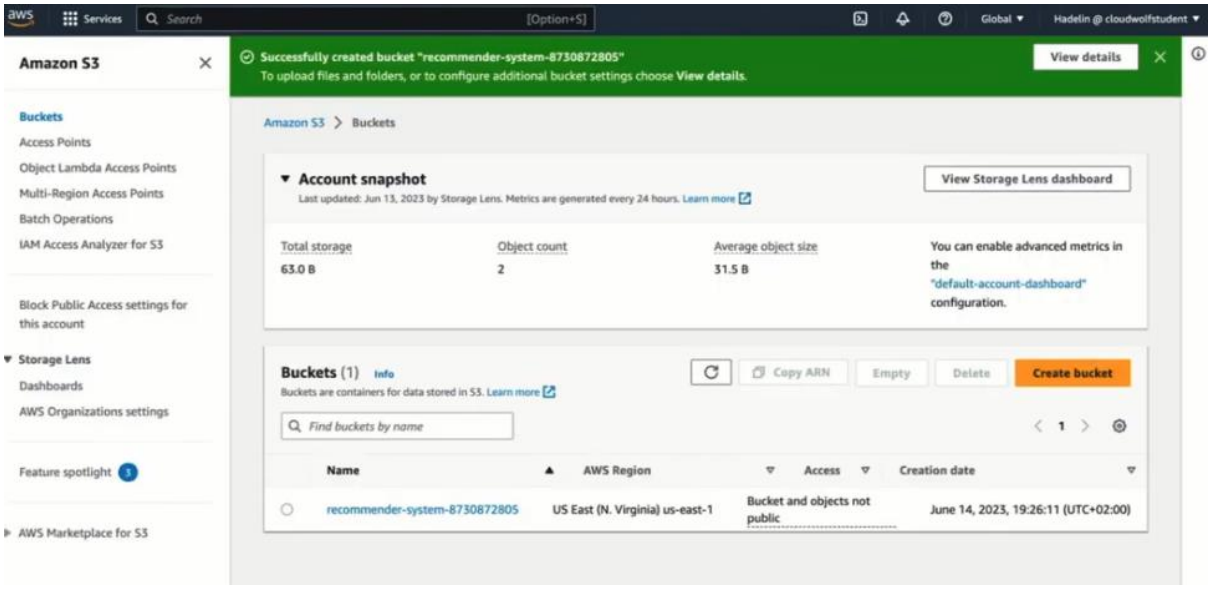

8. **SINCE WE ARE BUILDING A RECOMMENDER SYSTEM, WE WILL GIVE A BUCKET NAME recommender-system-8730872805. THE NUMBERS GIVEN HERE ARE JUST RANDOM. THIS IS TO MAKE THAT THIS IS GLOBALLY UNQUE NAME.**



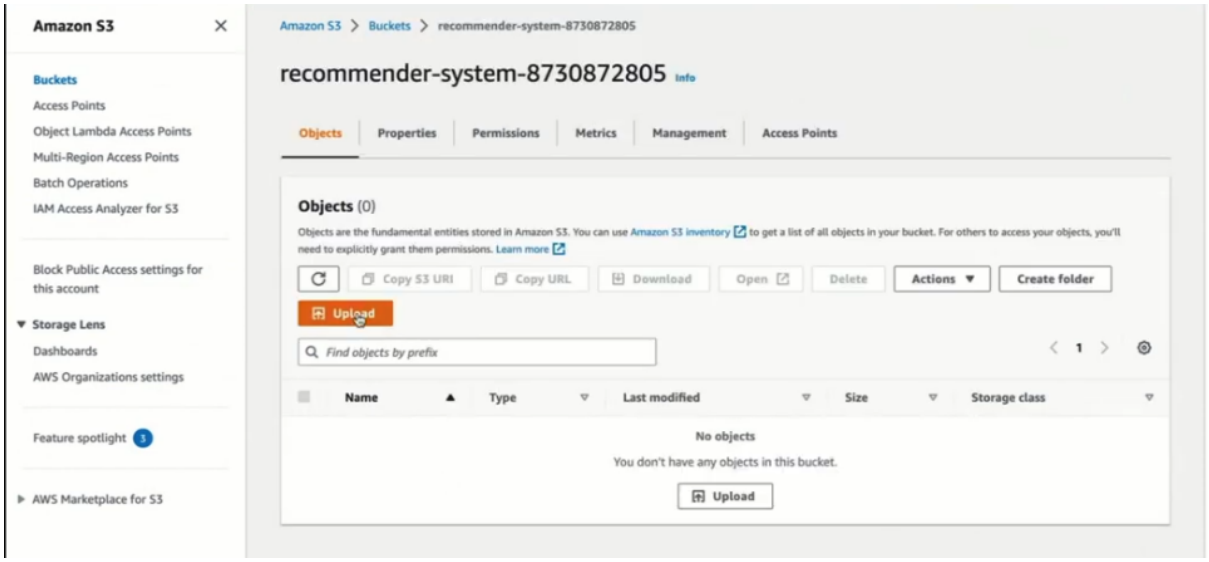

9. **SCROLL DOWN, AND, CLICK CREATE BUCKET.**



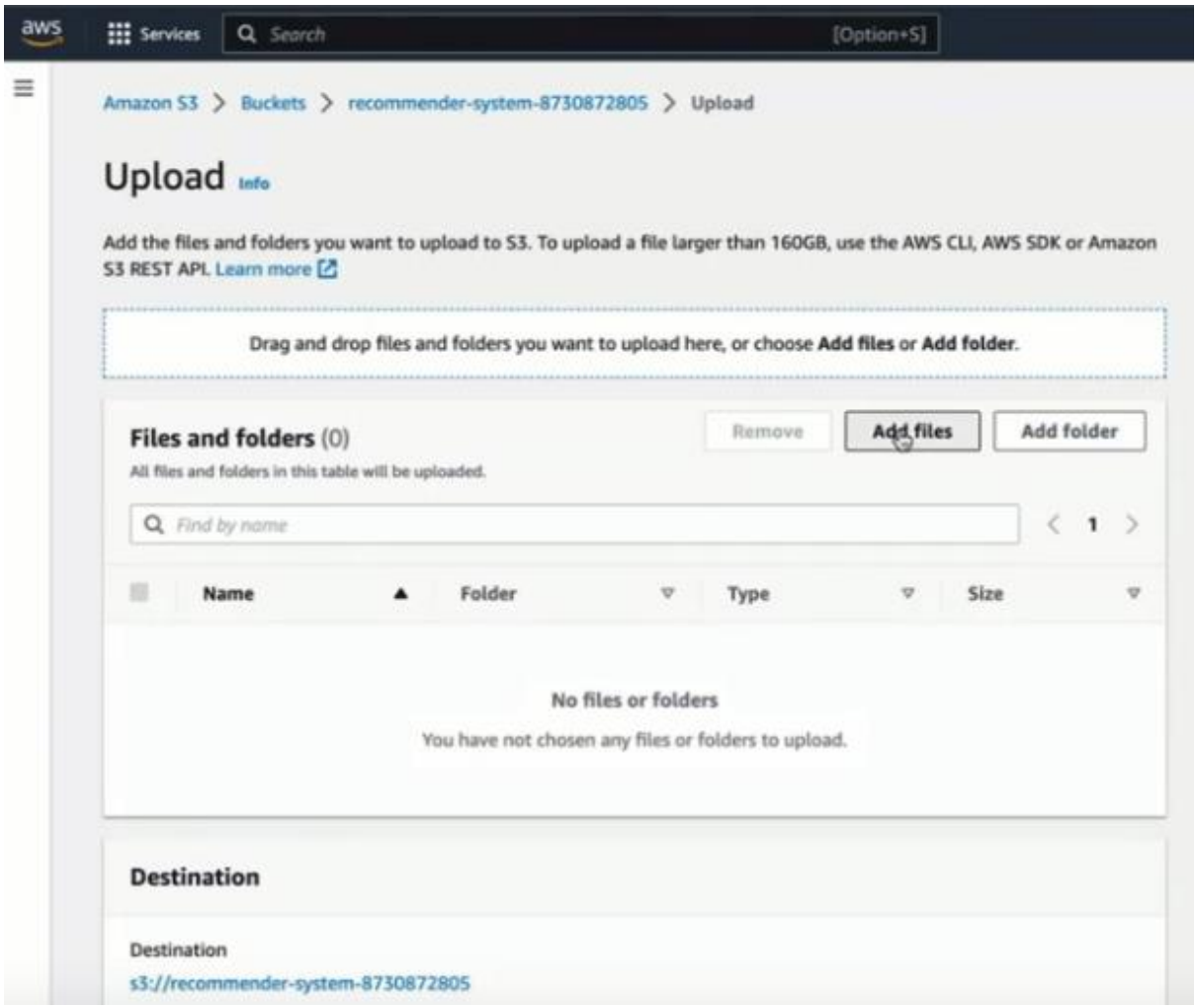

10. **THERE YOU GO, WE HAVE THE BUCKET. INSIDE THE BUCKET, WE ARE GOING TO UPLOAD OUR CSV FILE. CLICK THE BUCKET, recommender-system-8730872805.**




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

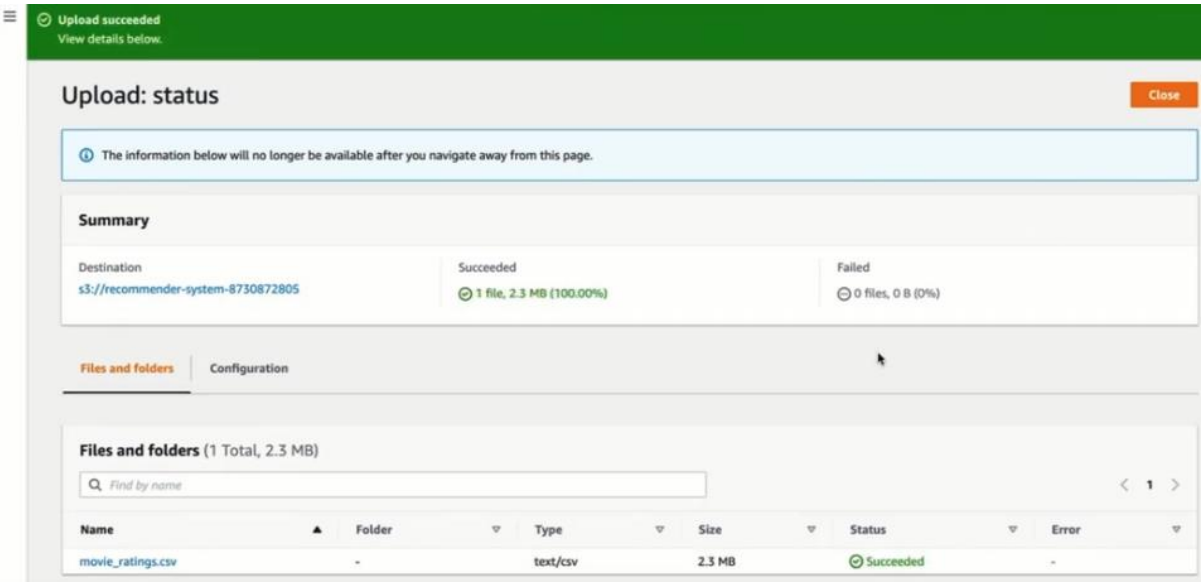


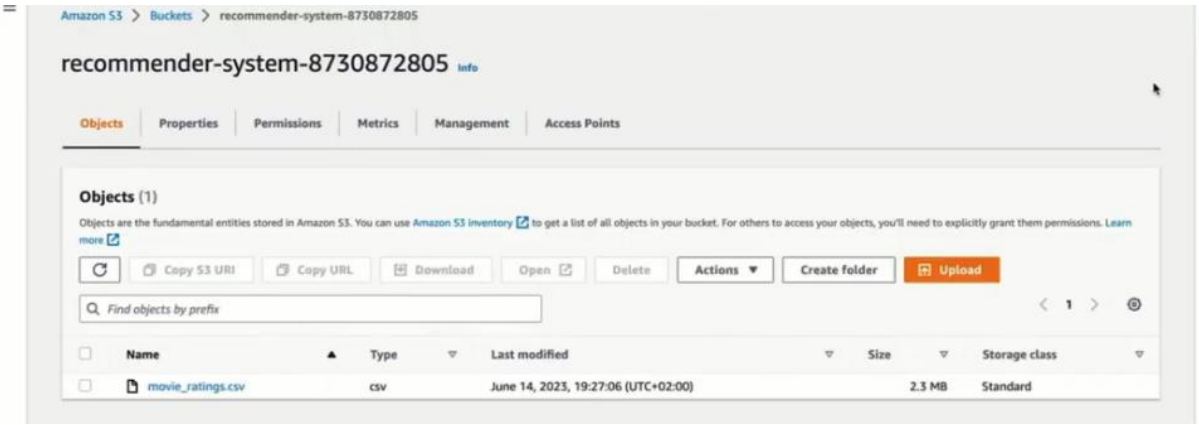

**12. CLICK ADD FILES.**



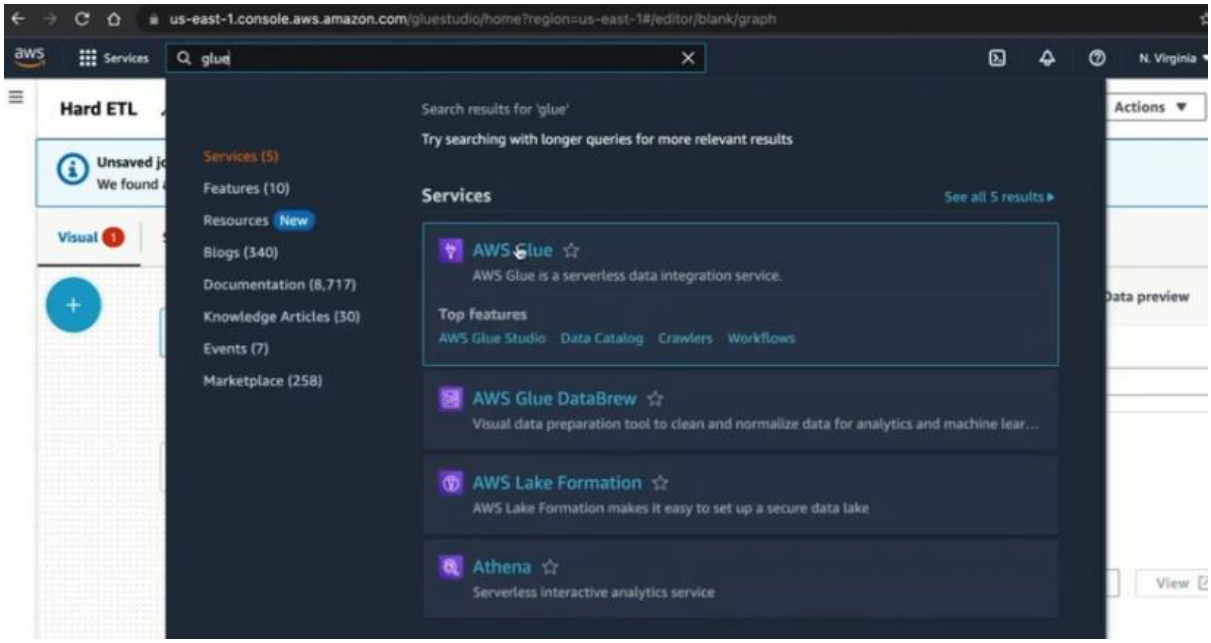
13. **CHOOSE THE SMALL DATA FILE. CLICK AND UPLOAD THE MOVIE_RATINGS.CSV FOR OUR S3 SOURCE.**


14. **YOU CAN SEE THAT THE MOVIE_RATINGS IS NOW UPLOADED.**

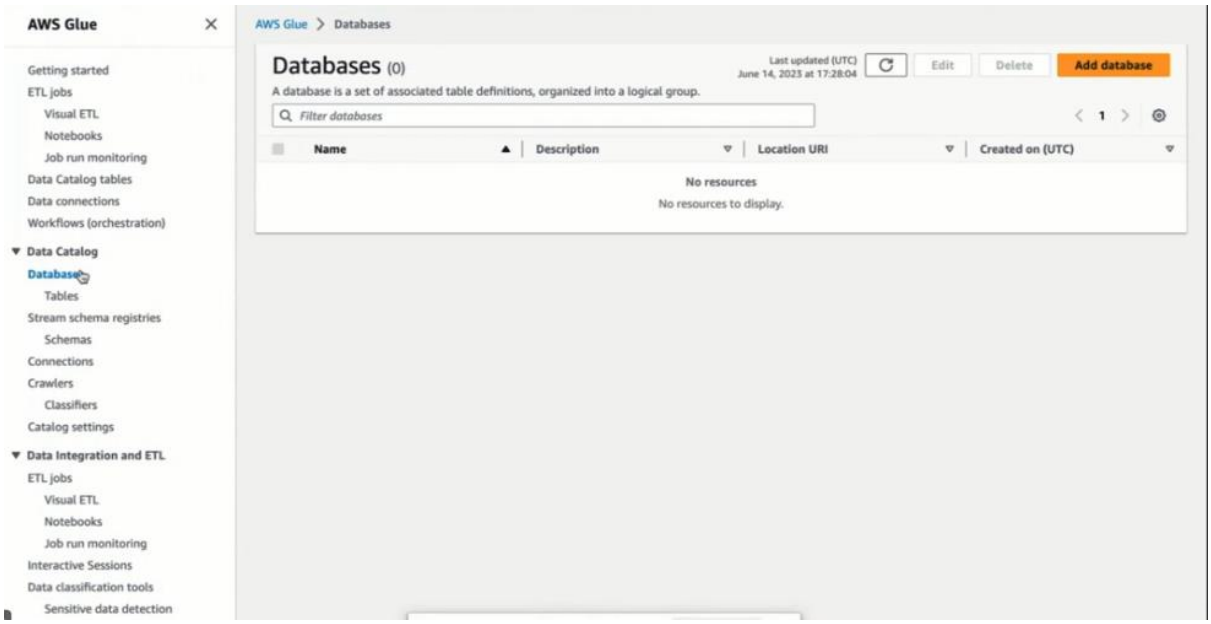


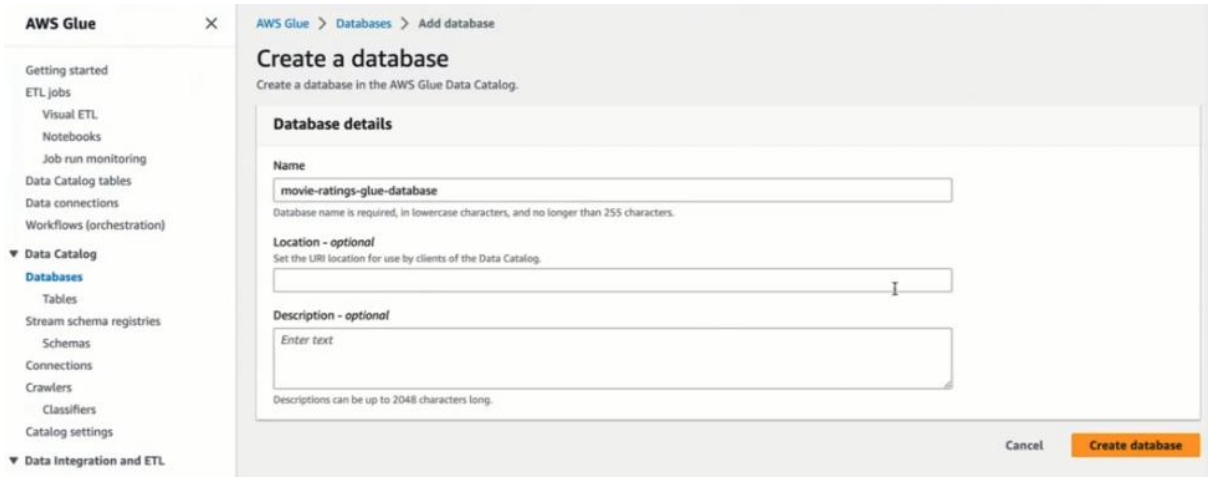

15. **WE NOW NEED TO CREATE THE DATA SOURCE IN THE DATA CATALOGUE OF GLUE. OPEN THE AWS GLUE.**




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

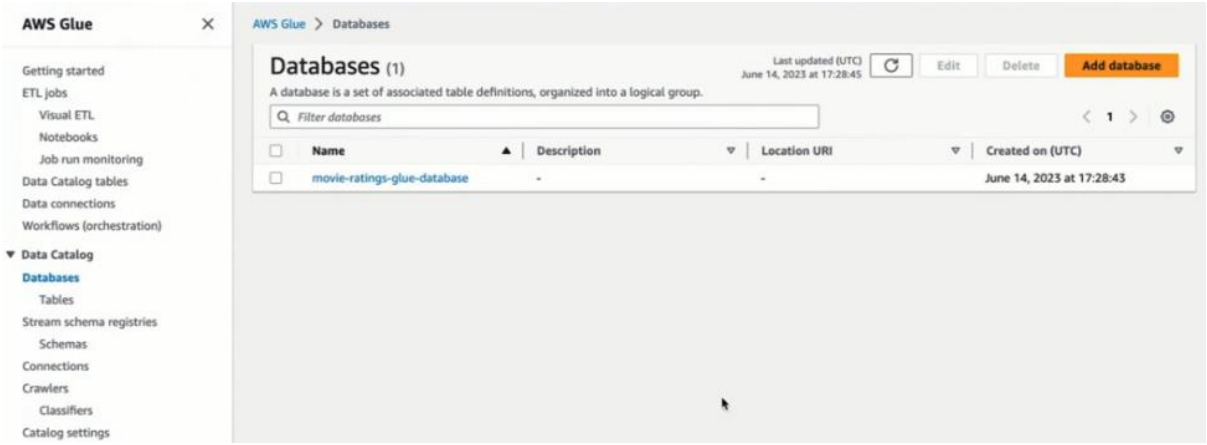



17. **SINCE WE WILL BE CREATING A DATABASE FOR THE MOVIE RATINGS, WE WILL NAME IT MOVIE-RATINGS-GLUE-DATABASE. THEN, CLICK CREATE DATABASE.**

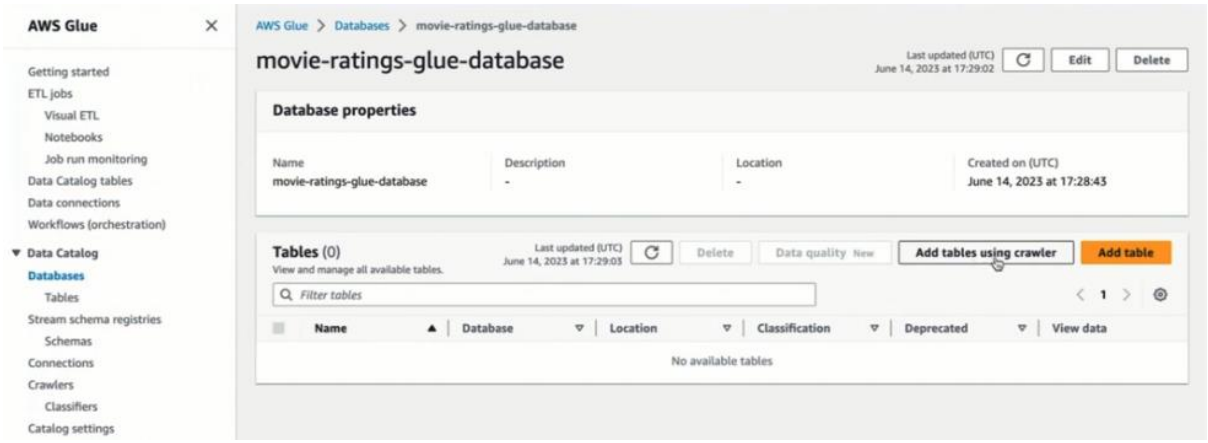



18. **THIS WILL NOW BE THE FIRST POINT IN THE DATA CATALOG. WE WILL THEN CREATE A TABLE WTIHIN THE MOVIE-RATINGS-GLUE-DATABASE DATABASE. THE TABLE IS GOING TO BE EXACTLY THE TABLE CONTAINING THE MOVIE RATINGS IN THE CSV FILE. TO DO SO, WE ARE GOING TO ENTER OUR DATABASE.**

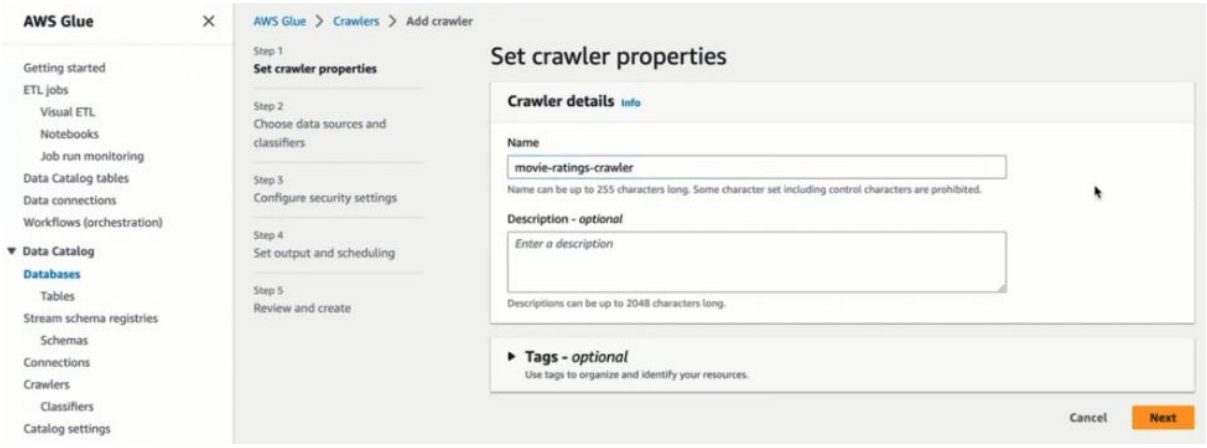
    **CLICK THE DATABASE movie-ratings-glue-database.**



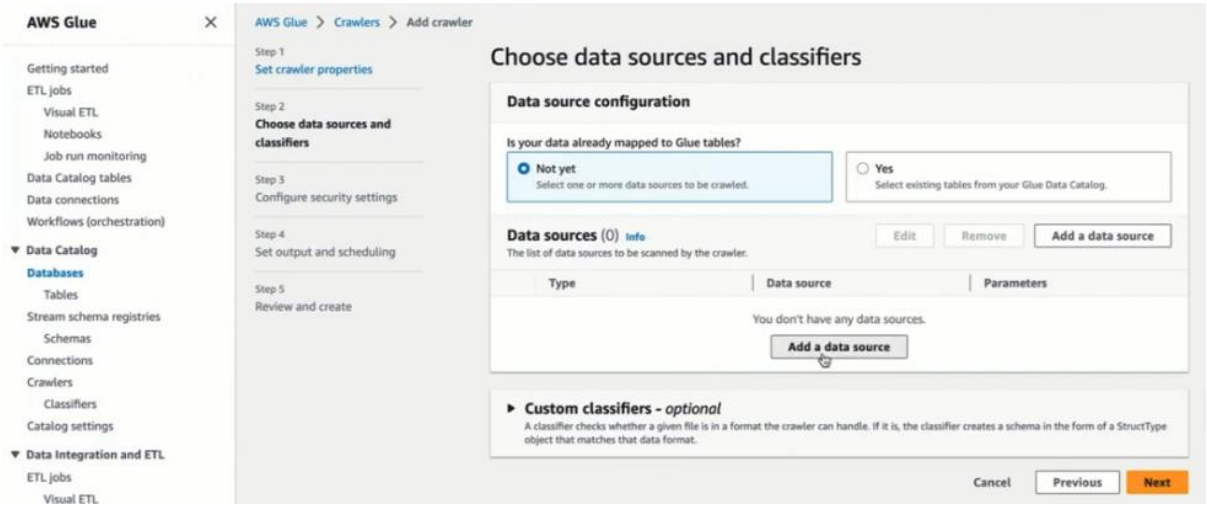

19. **CLICK ADD TABLES USING CRAWLER. BASICALLY, THE CRAWLER IS THE TOOL THAT CAN EXTRACT THE DATA IN A CSV FILE. IT EXTRACTS THE METADATA AND CAN ALSO POPULATE THE DATA IN THE CSV FILE.**



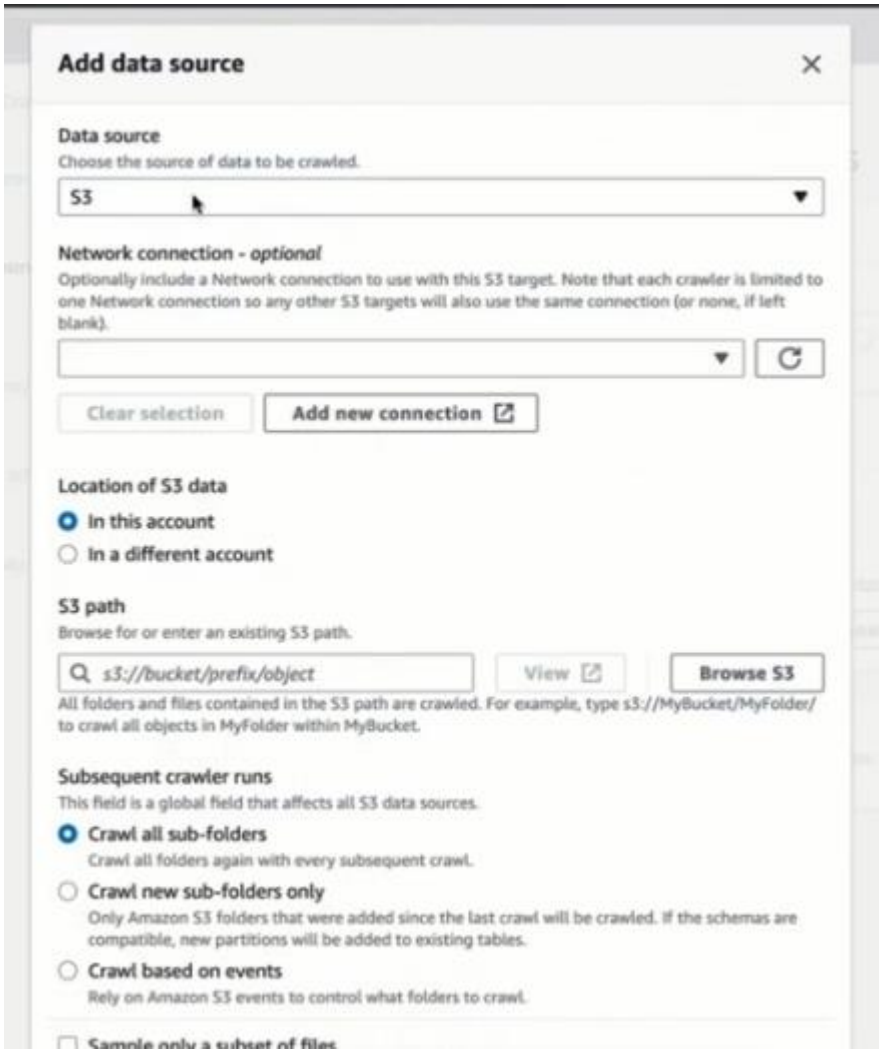

**20. ENTER UNIQUE CRAWLER NAME. HERE, WE WILL CALL IT MOVE-RATINGS-CRAWLER. CLICK NEXT.**



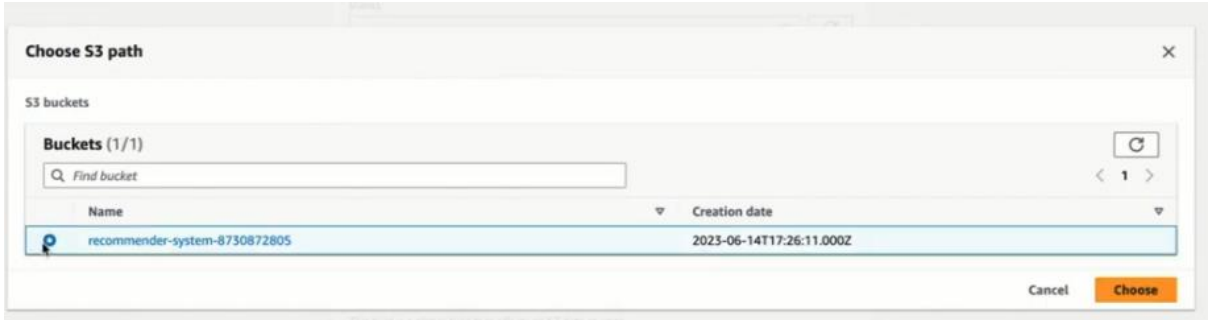

**21. CHOOSE NOT YET IN THE DATA SOURCE CONFIGURATION. THEN CLICK ADD A DATA SOURCE.**



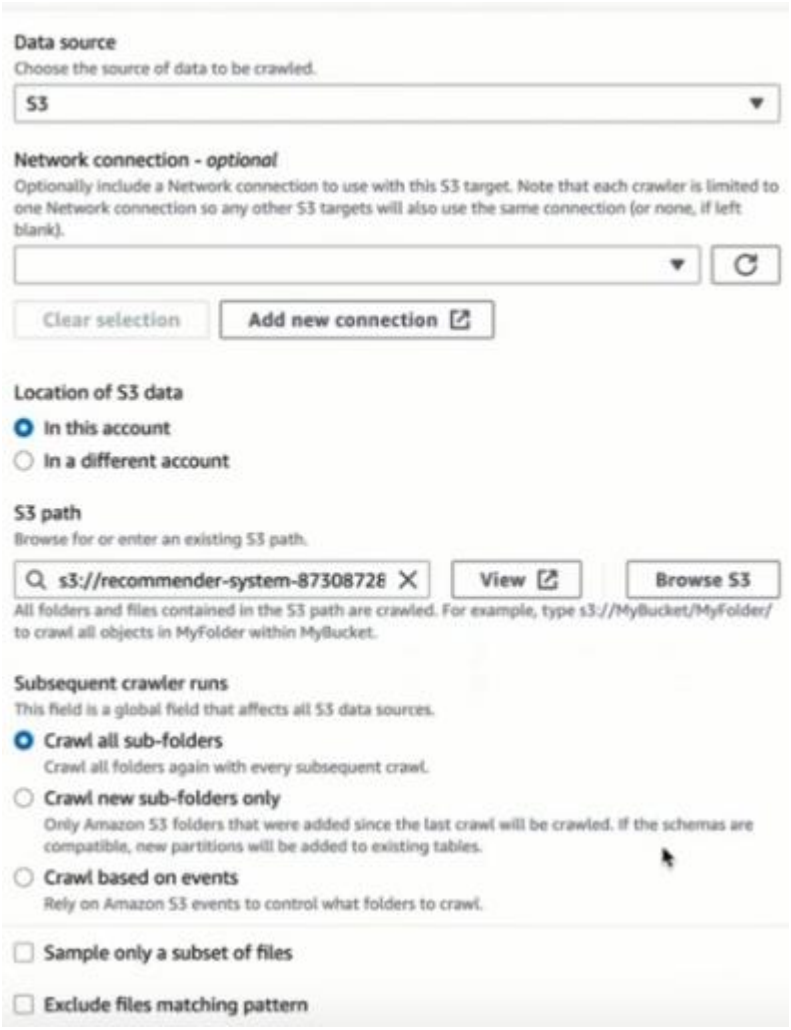
22. **DATA SOURCE IS S3, LOCATION OF S3 DATA IS IN THIS ACCOUNT. THEN, BROWSE S3.**



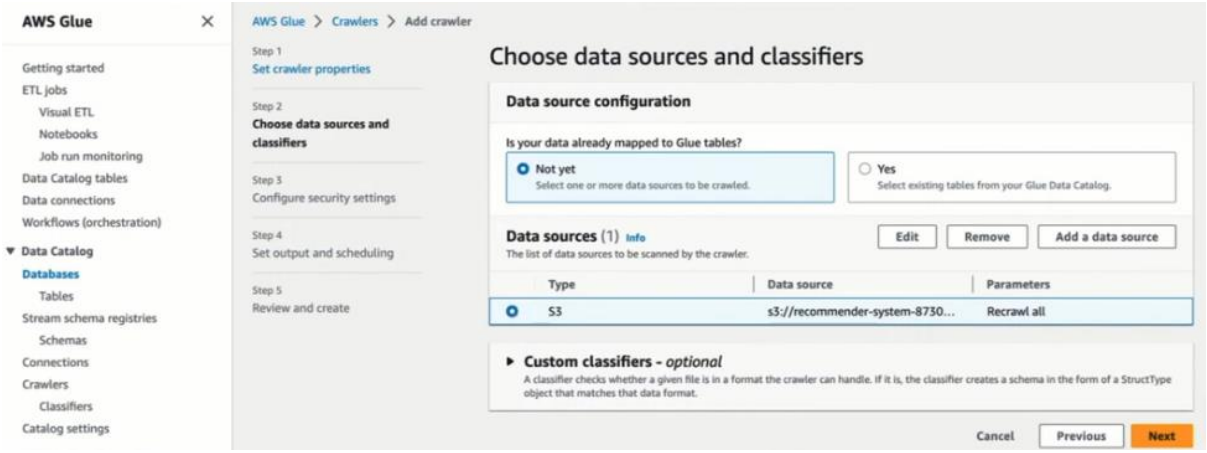

23. **IN THE BROWSE S3, CHOOSE THE BUCKET LISTED BELOW, THE BUCKET WE CREATED AWHILE AGO.**



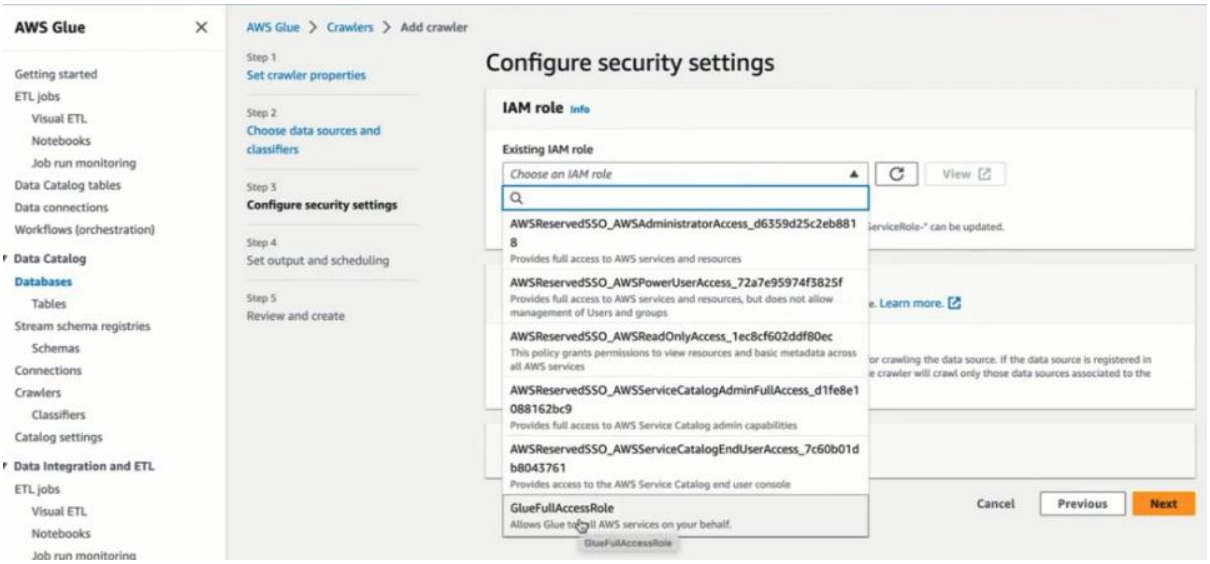

24. **SCROLL DOWN, IN THE SUBSEQUENT CRAWLER RUNS, CHOOSE CRAWL ALL SUB-FOLDERS. THEN CLICK, ADD DATA SOURCE.**




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

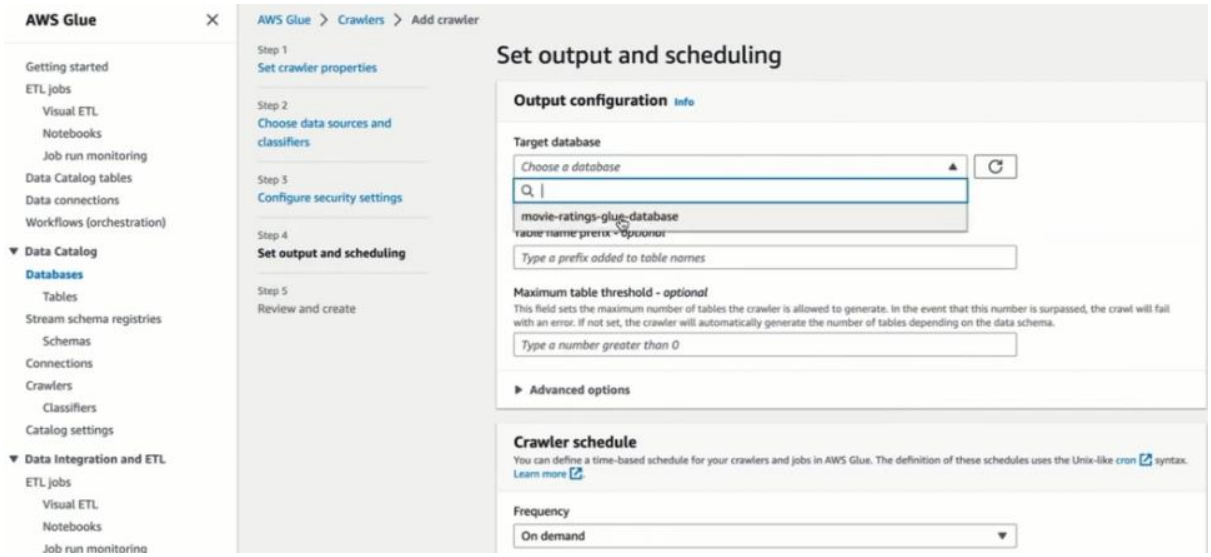



**26. ENTER THE IAM ROLE WE JUST CREATED. THIS IS TO GIVE US FULL PERMISSION FROM GLUE. SELECT THE ROLE THAT WE`VE CREATED, GLUEFULLACCESSROLE. THEN, CLICK NEXT.**

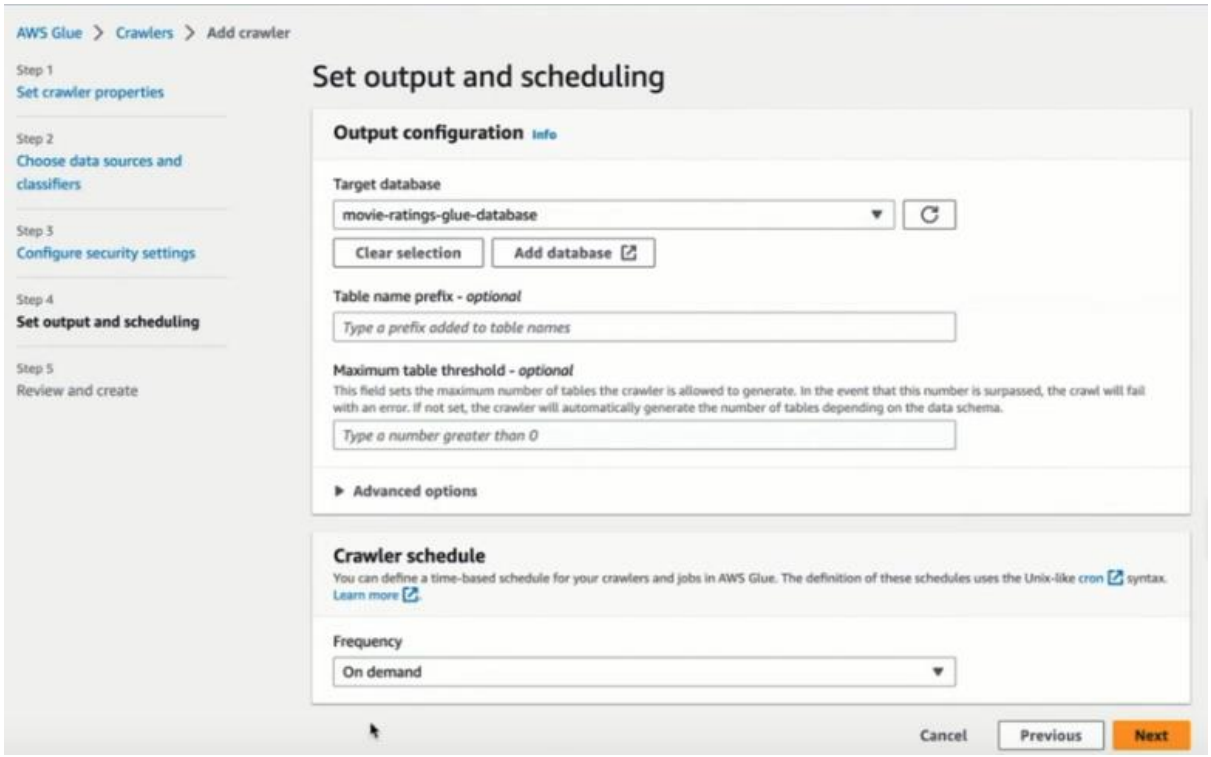




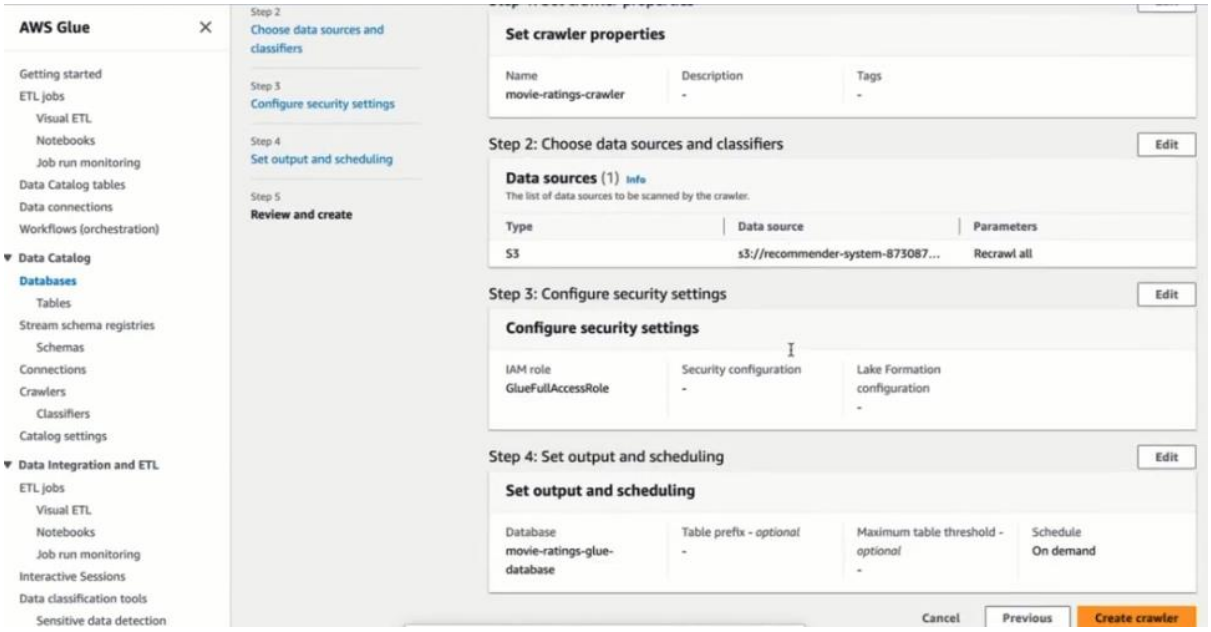
**27. IN THE TARGET DATABASE, SELECT THE ONE THAT WE CREATED.**




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

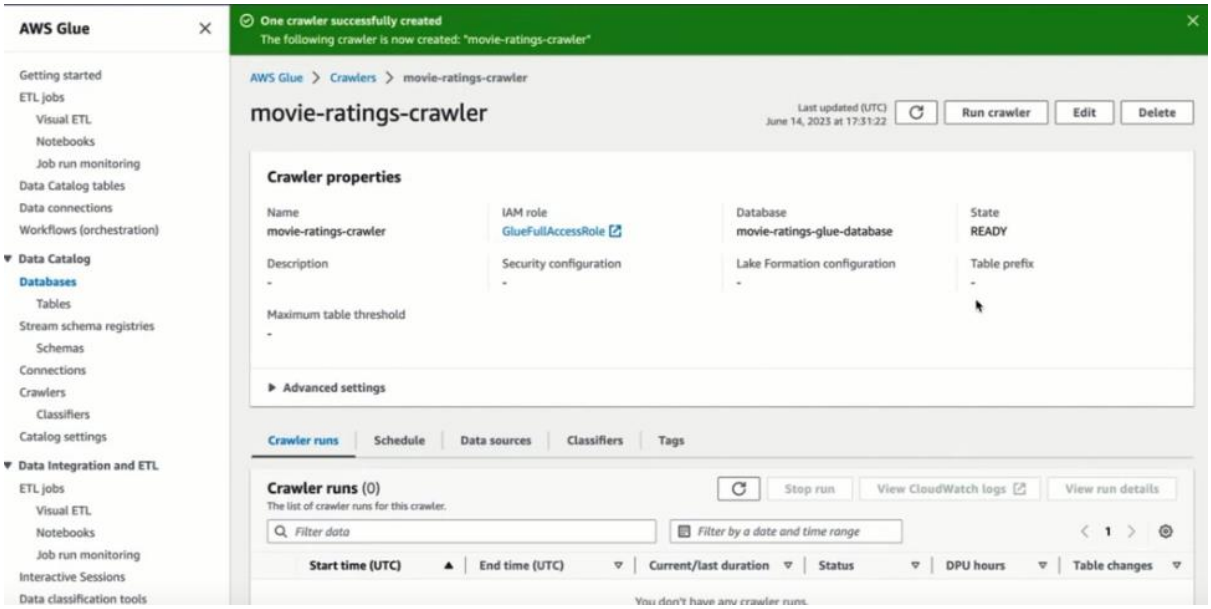



**28. REVIEW THE CONFIGURATION, THEN, CLICK CREATE CRAWLER.**

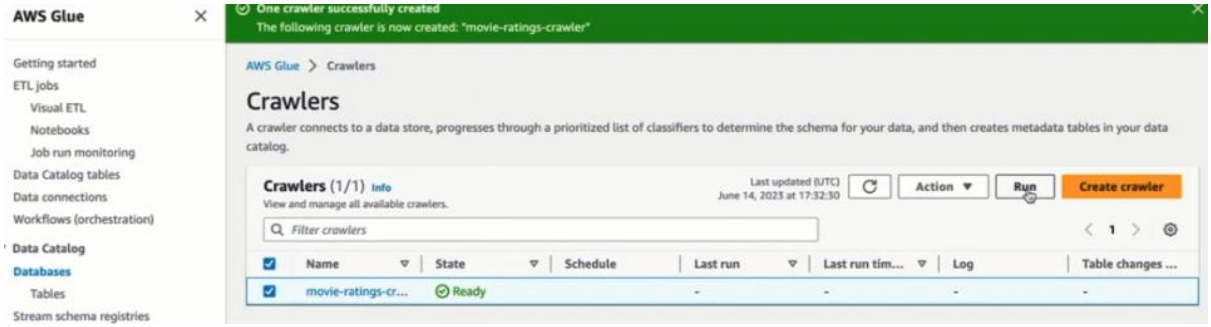



**29. WE NOW HAVE JUST CREATED OUR CRAWLER. WE WILL NOW RUN IT IN ORDER TO DO THIS EXTRACTION OF THE DATA AND PUT IT IN THE TABLE OF THIS DATABASE WE JUST CREATED WITHIN GLUE.**

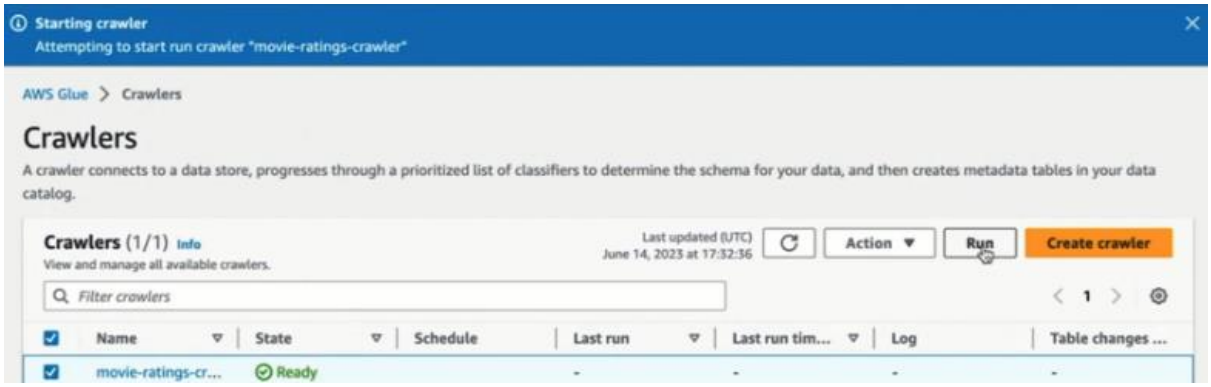
**CLICK RUN CRAWLER.**



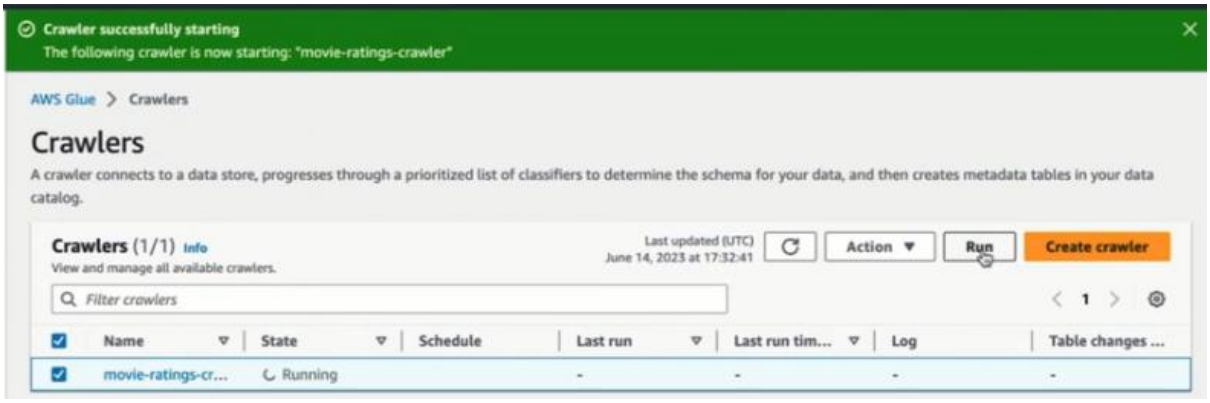

**30. SELECT THE CRAWLER, THEN, CLICK, RUN.**



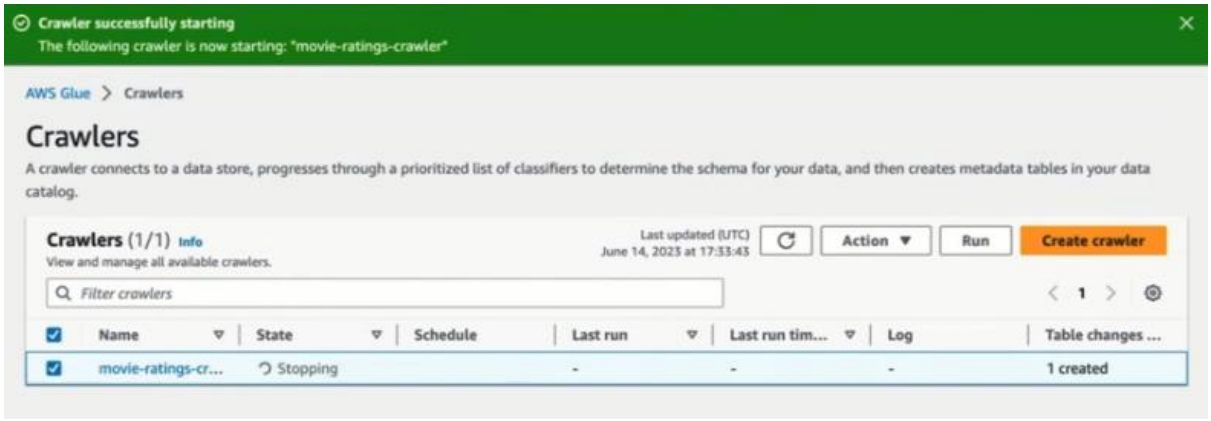

**31. IT IS NOW SUCCESSFULLY RUNNING. IT WILL EXTRACT ALL THE DATA FROM THE CSV FILE TO PUT IT IN THE TABLE OF THE DATABASE WE CREATED. WE CREATED TWO THINGS SO FAR, THE DATABASE FIRST, THEN, WE ARE CREATING THE TABLE OF THIS DATABASE.**

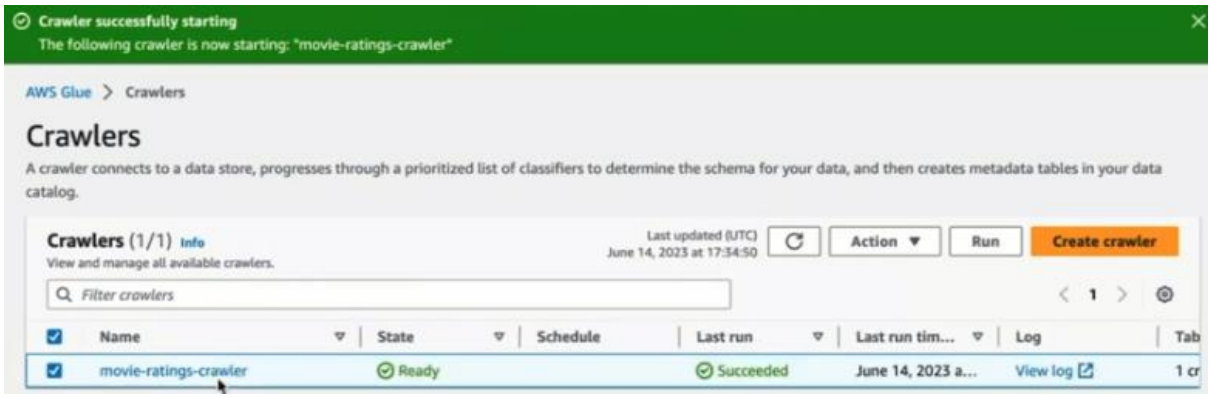



**IT IS NOW RUNNING AND IT WILL TAKE APPROX. 2 MINUTES. AT THE END, WE WILL SEE THAT WE HAVE THE MOVE RATINGS TABLE CONTAINING THE EXACT SAME DATA AS IN THE CSV FILE. AND THAT IS THE DATA CATALOG ELEMENT AND THE ELEMENT IS A TABLE WHICH WE WILL THEN CONNECT IN OUR ETL PROCESS.**

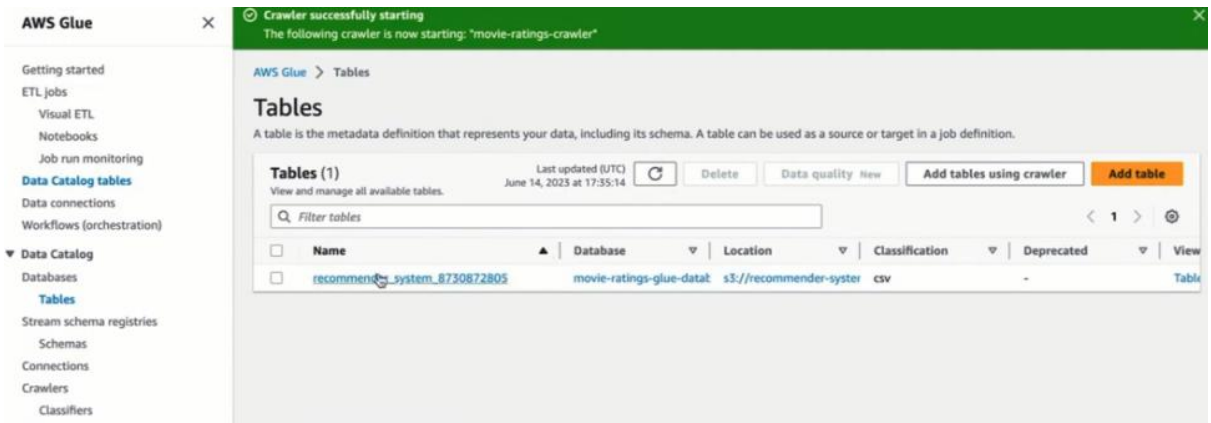



**ONCE IT IS STOPPING, IT MEANS THAT THE CRAWLING IS DONE AND SOON IT SHOULD SAY SUCCESSFUL.**

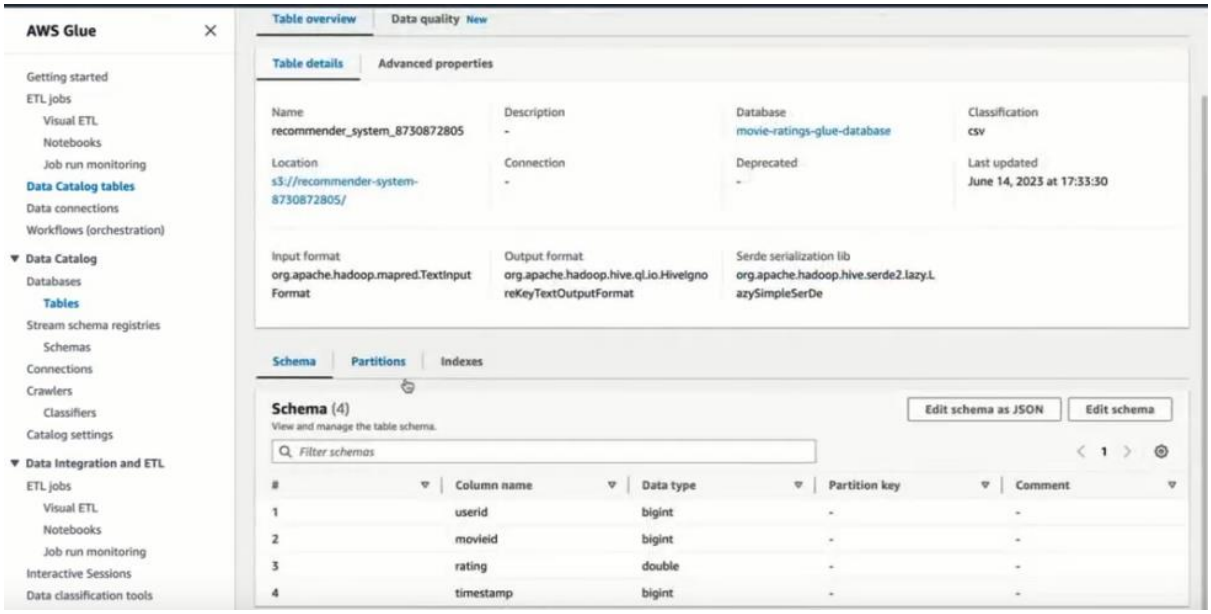


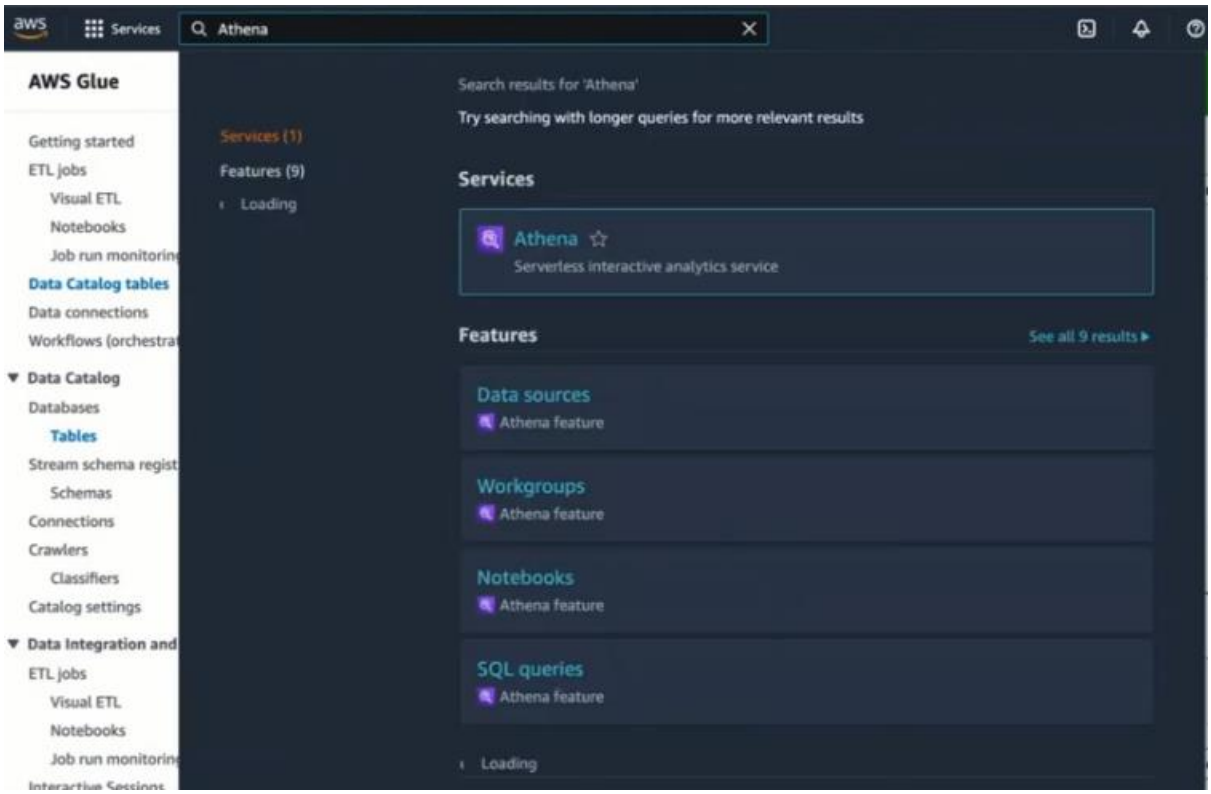

32. **ONCE READY OR SUCCESSFUL, GO TO THE DATA CATALOG, DATABASE, THEN, TABLES. REFRESH IT TO SEE THE TABLES. CLICK THE TABLE AVAILABLE IN THE LIST.**



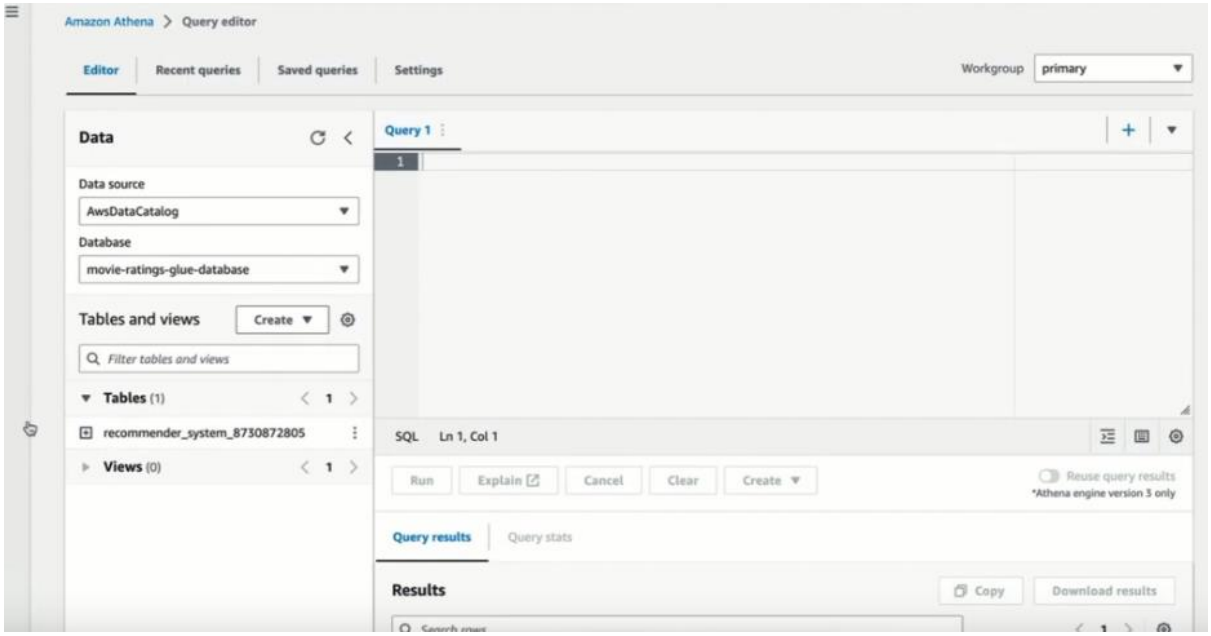

33. **THIS IS NOW OUR TABLE OVERVIEW. THERE, WE CAN`T SEEM TO SEE THE VALUES WHICH GIVES OUR TABLE NOT VERY INTUITIVE. TO HAVE A GREAT LOOK OF OUR TABLE, WE WILL NEED TO USE ANOTHER SERVICE.**



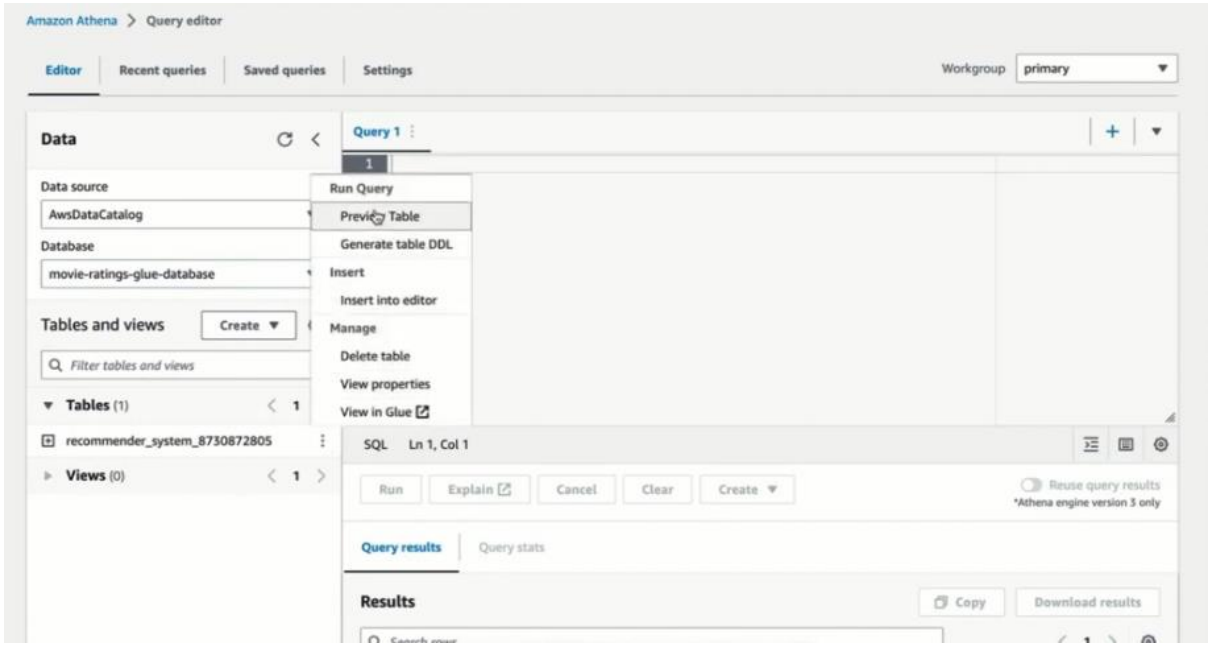

**34. IN THE SEARCH BAR, LOOK FOR ATHENA.THIS IS A QUERY SERVICE WHERE WE CAN MANIPULATE THE DATA SO WE CAN HAVE A GOOD LOOK OF OUR DATA.**



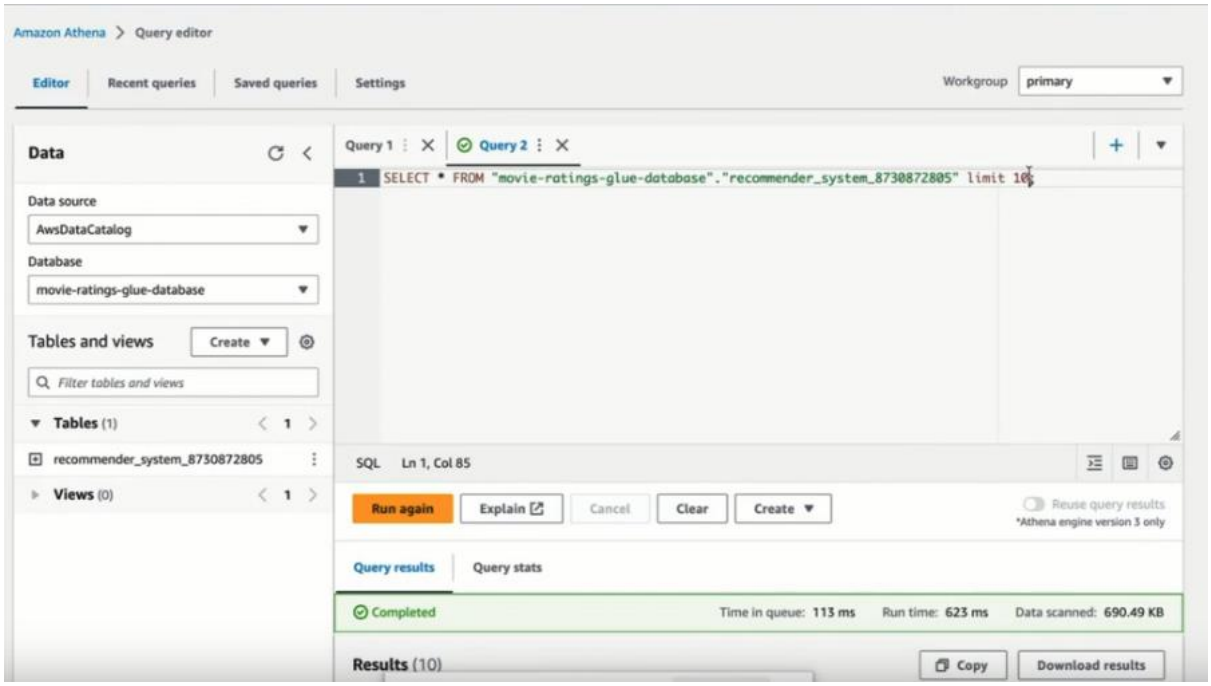
**HERE, YOU WILL SEE THE DATA SOURCE, DATABASE, AND THE TABLE THAT WE JUST CREATED.**

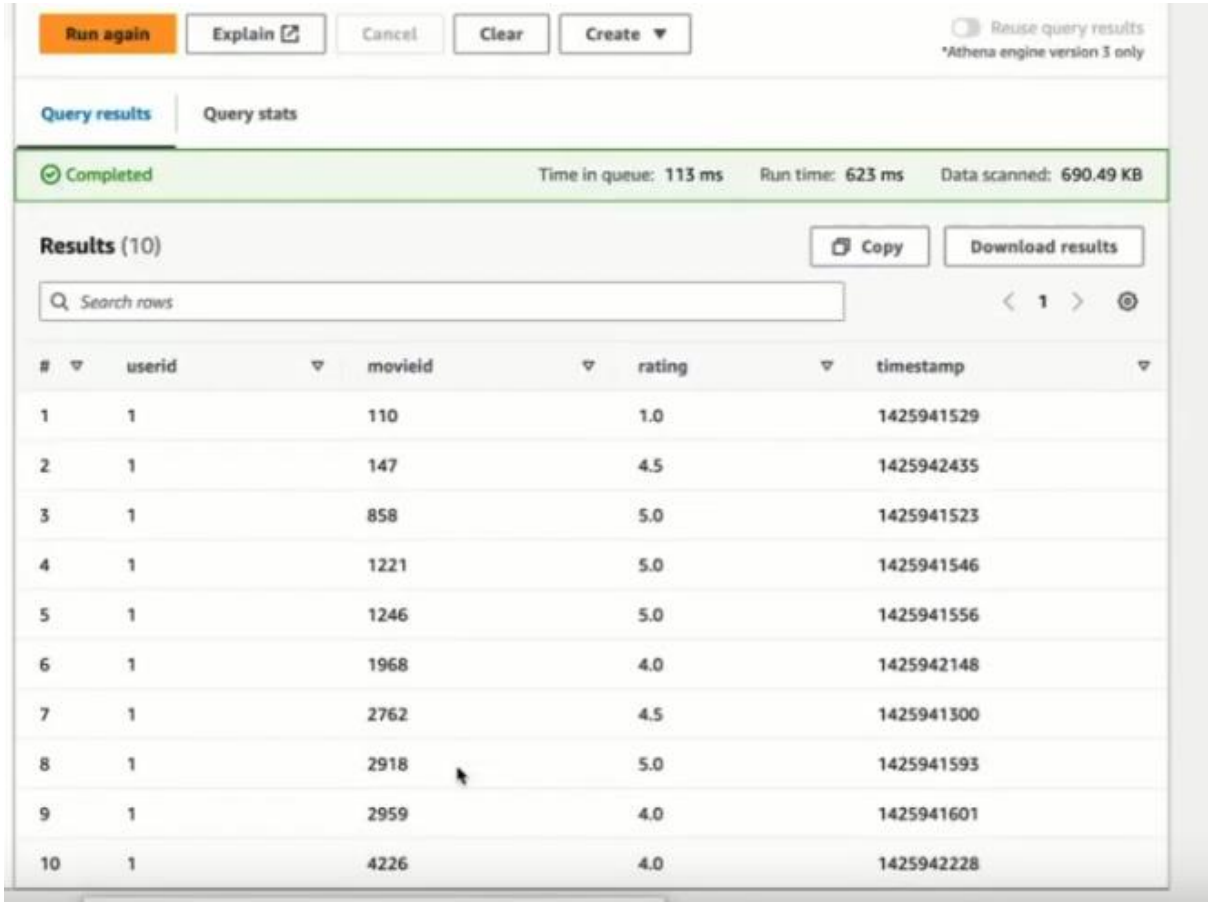



35. **TO LOOK FOR THE TABLE, CLICK THE THREE DOTS ON THE RIGHT SIDE OF THE TABLE, THEN, CLICK PREVIEW TABLE.**

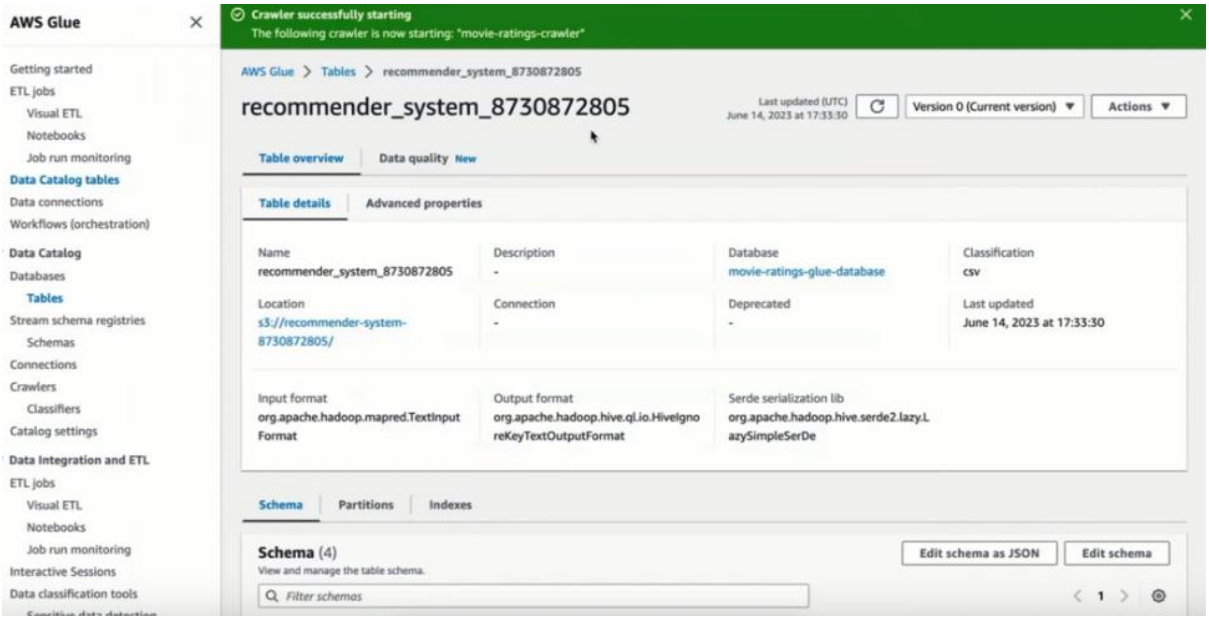



36. **THIS WILL BE THE RESULT OR THE PREVIEW OF THE TABLE. IN THE QUERY SECTION, WE LIMIT THE TABLE TO 10. YOU CAN SET THE LIMIT BASED ON YOUR PREFERENCE.**




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

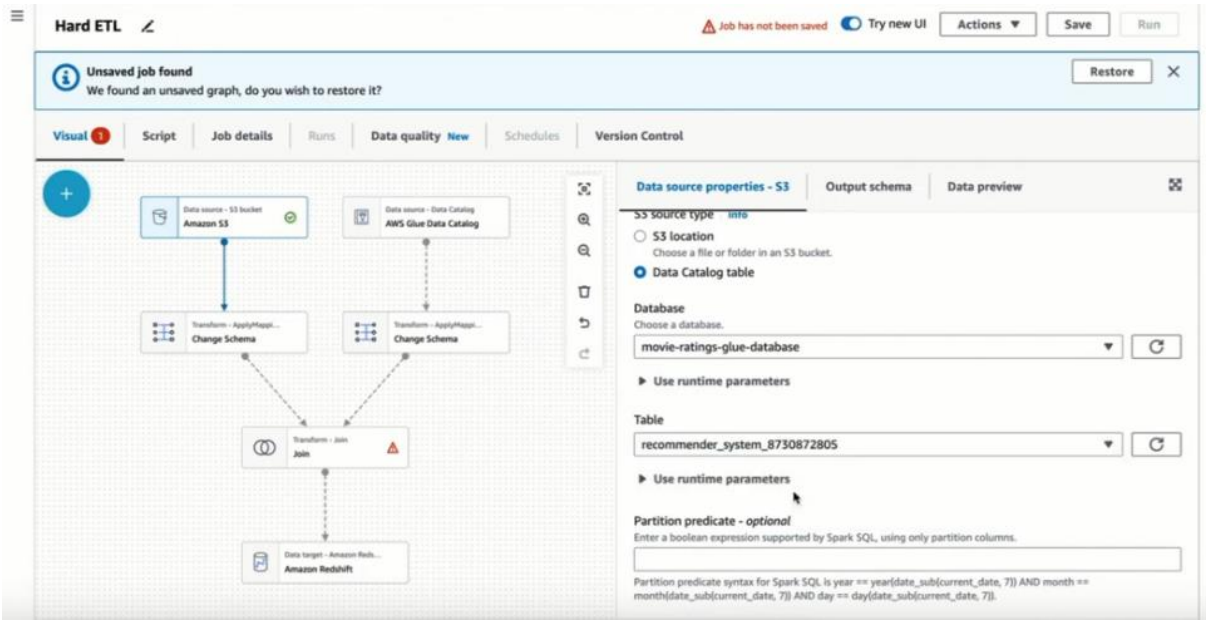



38. **WE NOW GO TO THE VISUAL EDITOR AND WE ARE GOING TO CONNECT THE DATA. WE JUST CREATED THE DATA SOURCE S3, BUT, NOW WE ARE GOING TO CONNECT IT.**

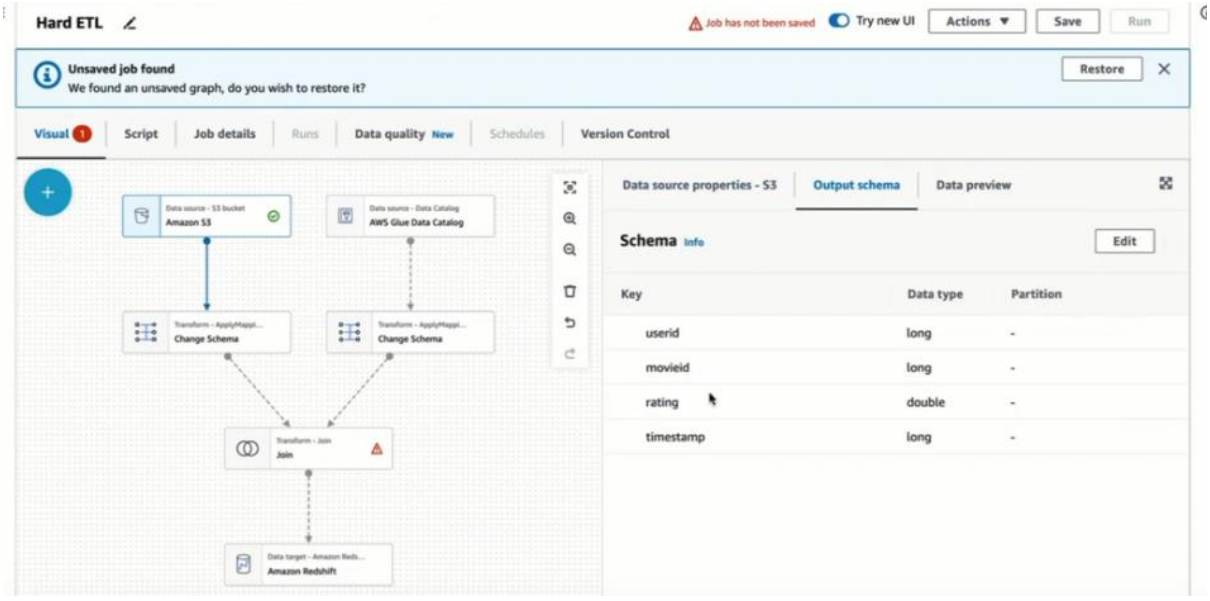
    **IN THE S3 SOURCE TYPE, CLICK DATA CATALOG TABLE.**

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

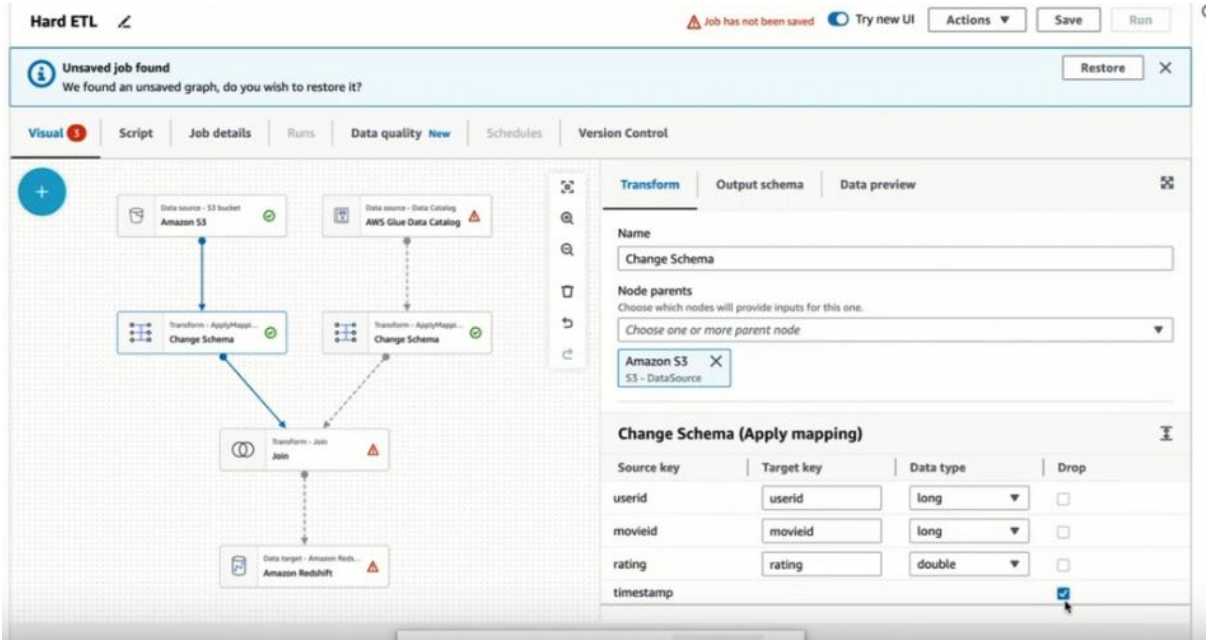
    **IN THE TABLE, CLICK THE TABLE THAT WAS CREATED.**




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

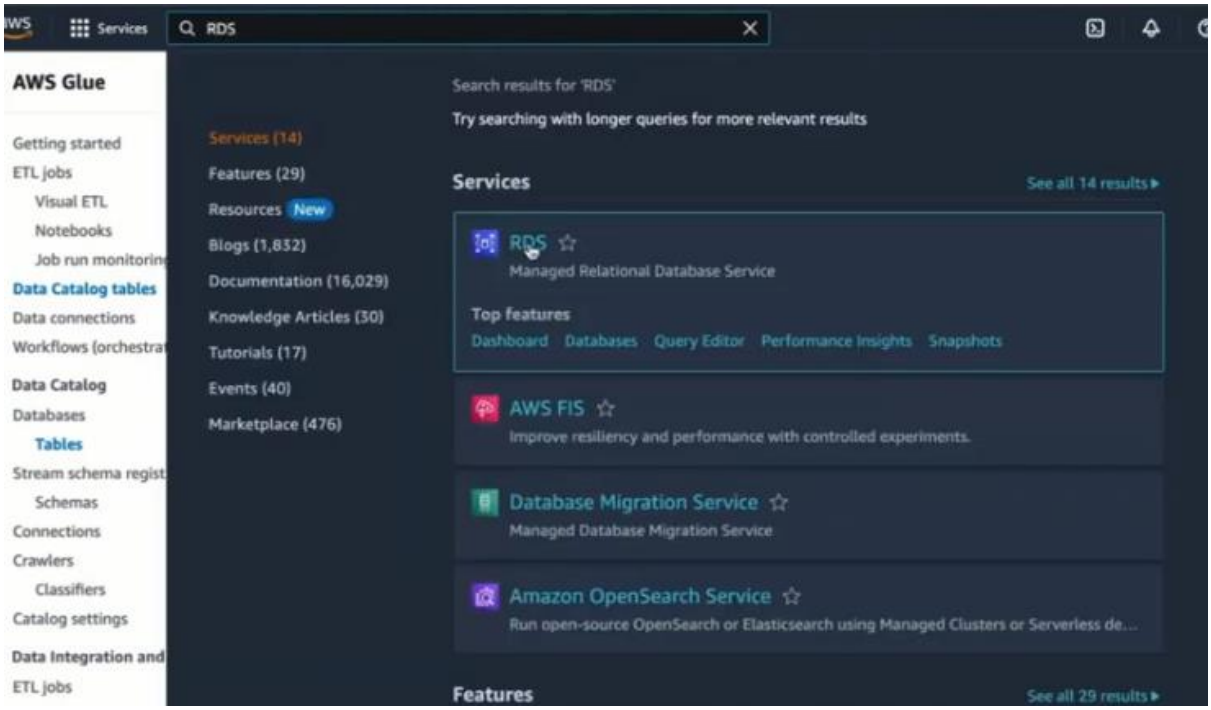



40. **PROCEED TO THE TRANSFORM-APPLYMAPPING PROCEDURE. SINCE WE ARE BUILDING A RECOMMENDER SYSTEM, WE CAN CHECK OR CHANGE THE SCHEMA BASED ON THEIR RELEVANCE. HERE, WE CHOOSE TO DROP THE TIMESTAMP WHICH WE FOUND IRRELEVANT IN BUILDING THE RECOMMENDER SYSTEM.**

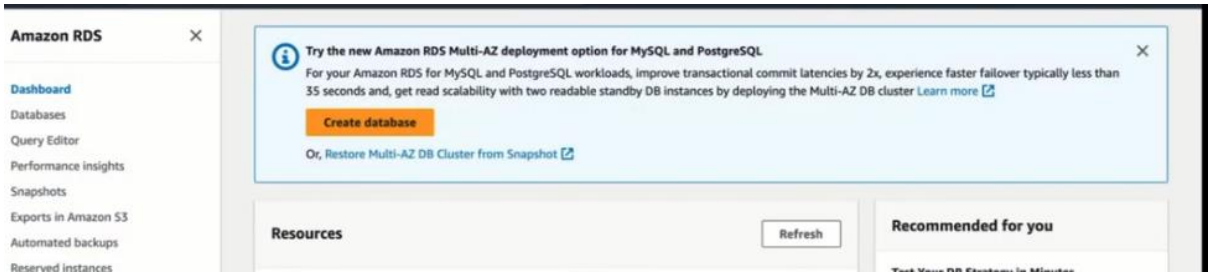



**THIS IS THE KIND OF TRANSFORMATION WE CAN DO HERE WITH THIS APPLY MAPPING TRANSFORM. THE FIRST DATA SOURCE IS DONE. WE HAVE EXTRACTED IT PROPERLY WITH THE DATA CATALOG AND THEN WE HAVE TRANSFORMED IT PROPERLY THROUGH THE APPLY MAPPING TRANSFORM PROCEDURE.**

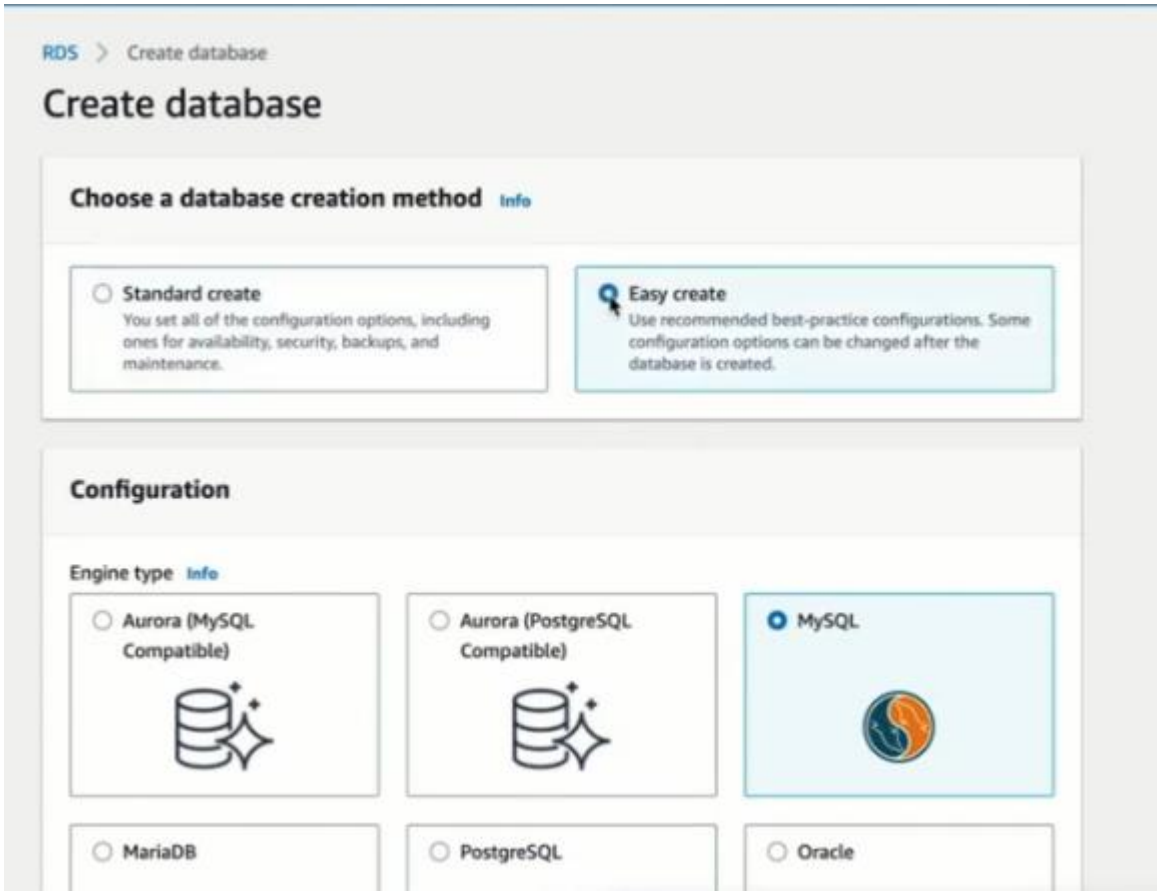
41. **PROCEED TO THE SECOND DATA SOURCE, THE RDS. IN THE SERVICES, LOOK FOR RDS.**




**42. CLICK CREATE DATABASE.**

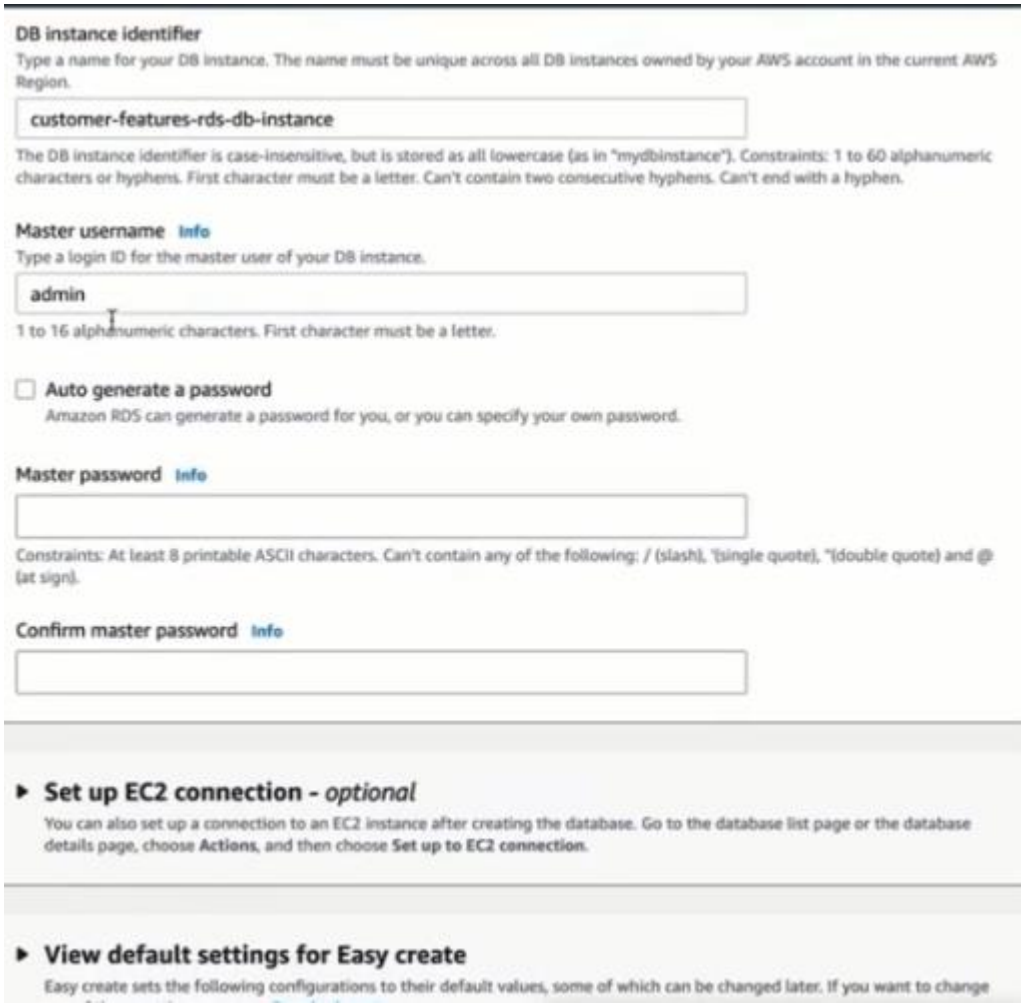



**43. WE WILL CREATE A RDS DATABASE POWERED BY MYSQL. SO, CHOOSE EASY CREATE. THEN, CLICK MYSQL IN THE ENGINE TYPE OF CONFIGURATION.**

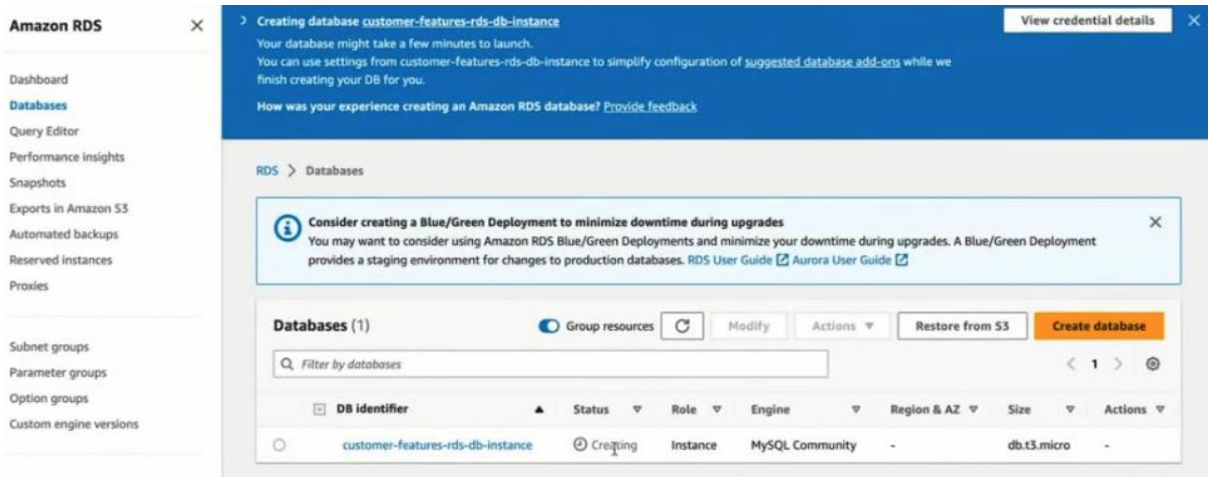


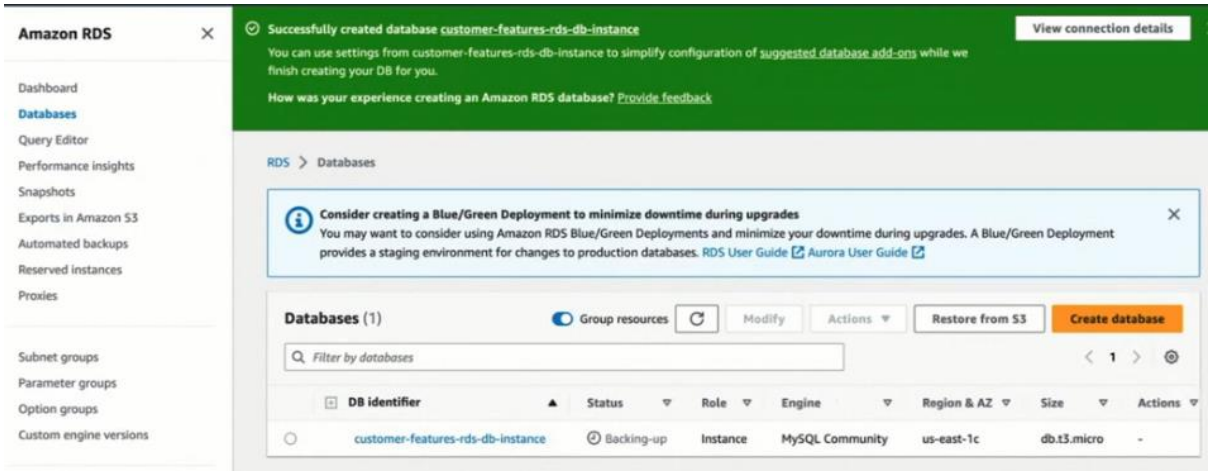

**44. CHOOSE FREE TIER.**


**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.**




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

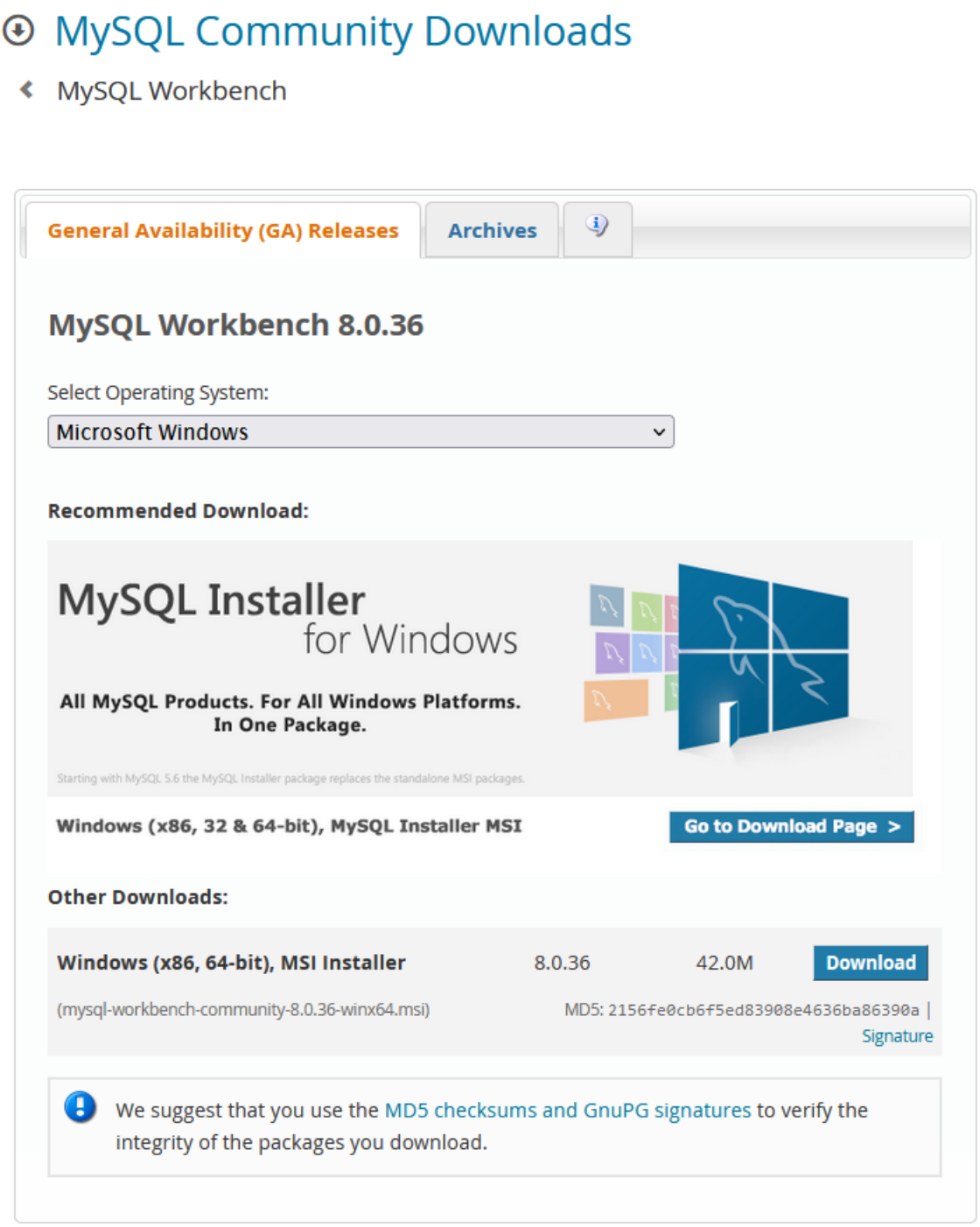




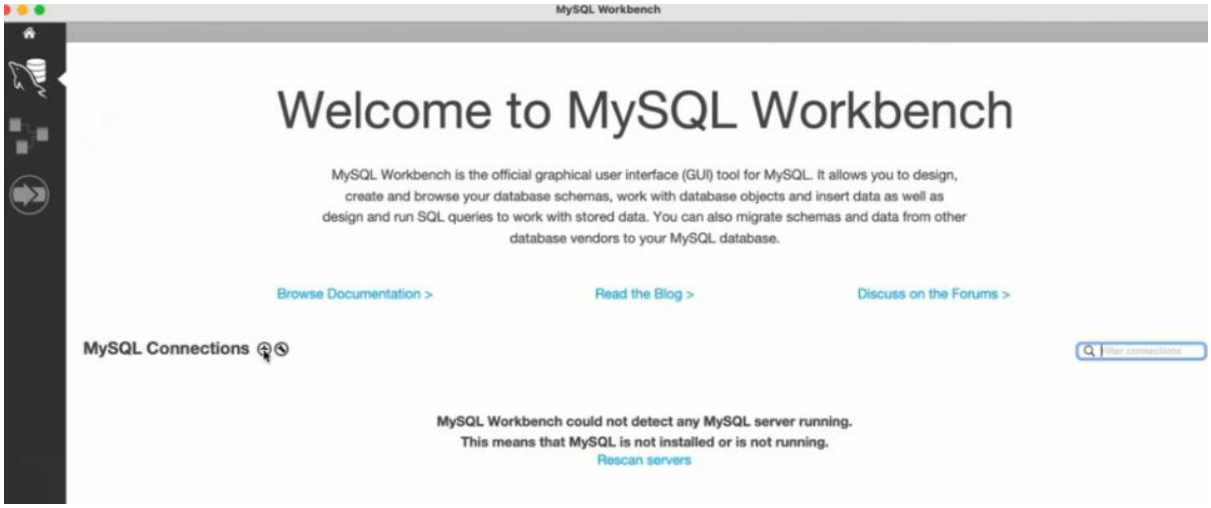

47. **WE WILL NOW CREATE A TABLE CONTAINING THE CUSTOMER FEATURES DATA IN THE CSV FILE INTO THE DATABASE. WE CAN DO IT BY EITHER USING THE CLI OR AN EXTRA TOOL WHICH IS CALLED MYSQL WORKBENCH.**

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

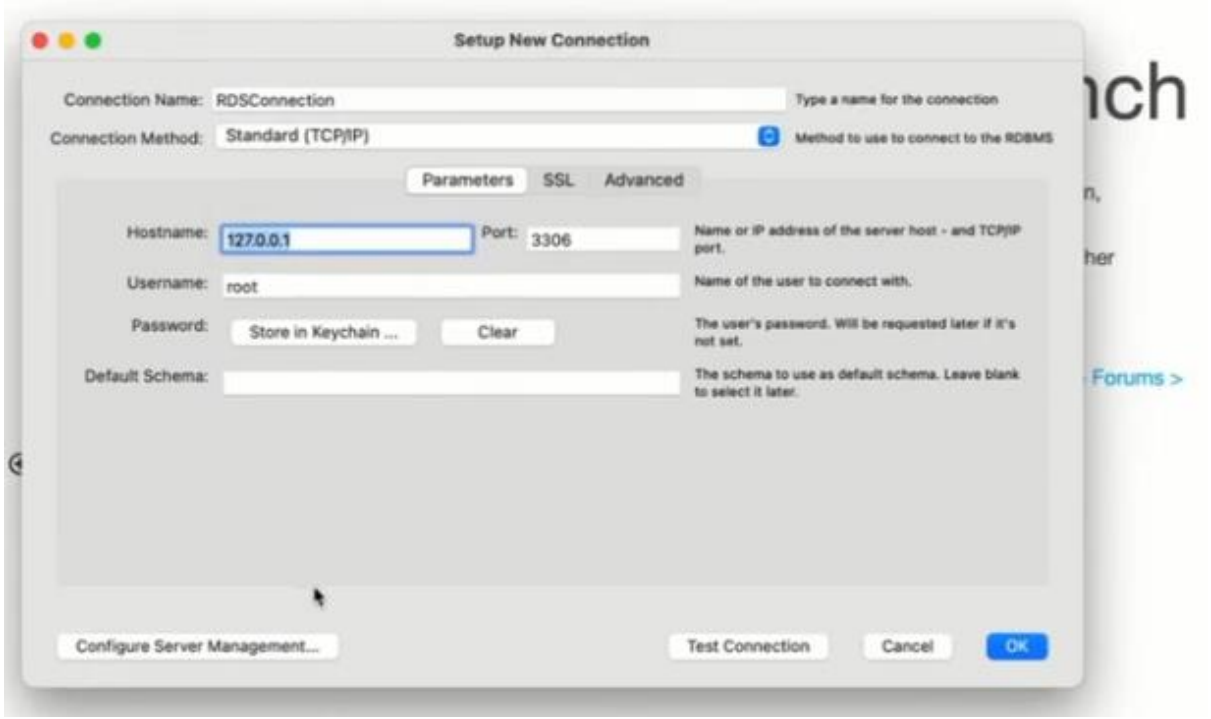
**https://dev.mysql.com/downloads/workbench/**



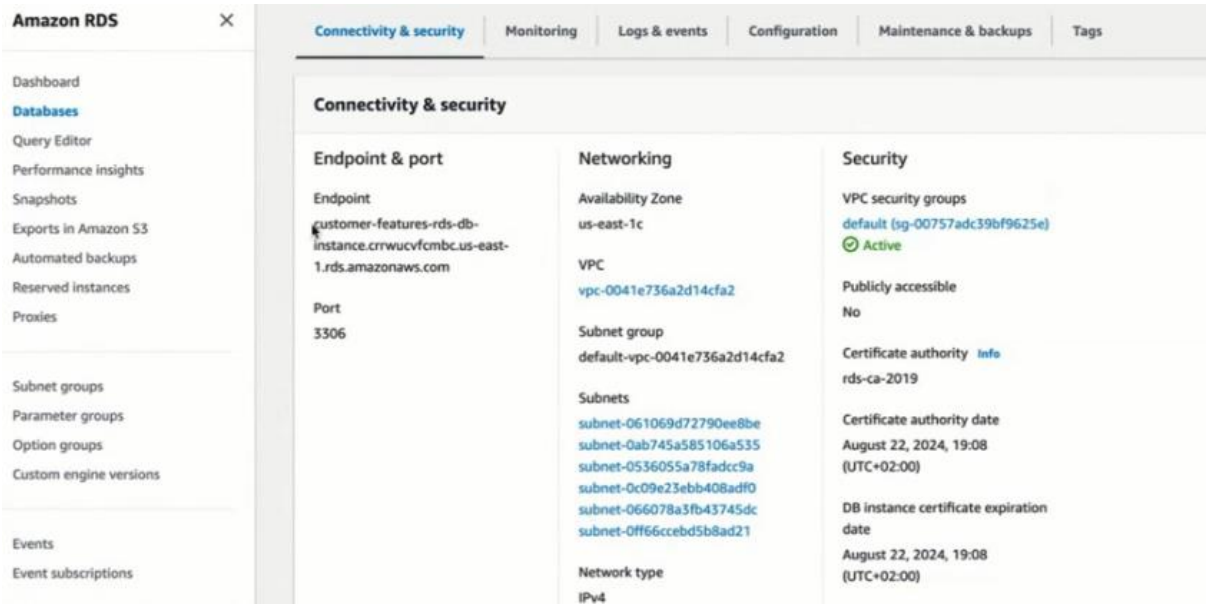

**48. ONCE INSTALLED, MAKE A CONNECTION BETWEEN MYSQL WORKBENCH AND RDS DATABASE. OPEN MYSQL WORKBENCH, CLICK THE PLUS (+) BUTTON TO CREATE THE CONECTION.**




**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.**

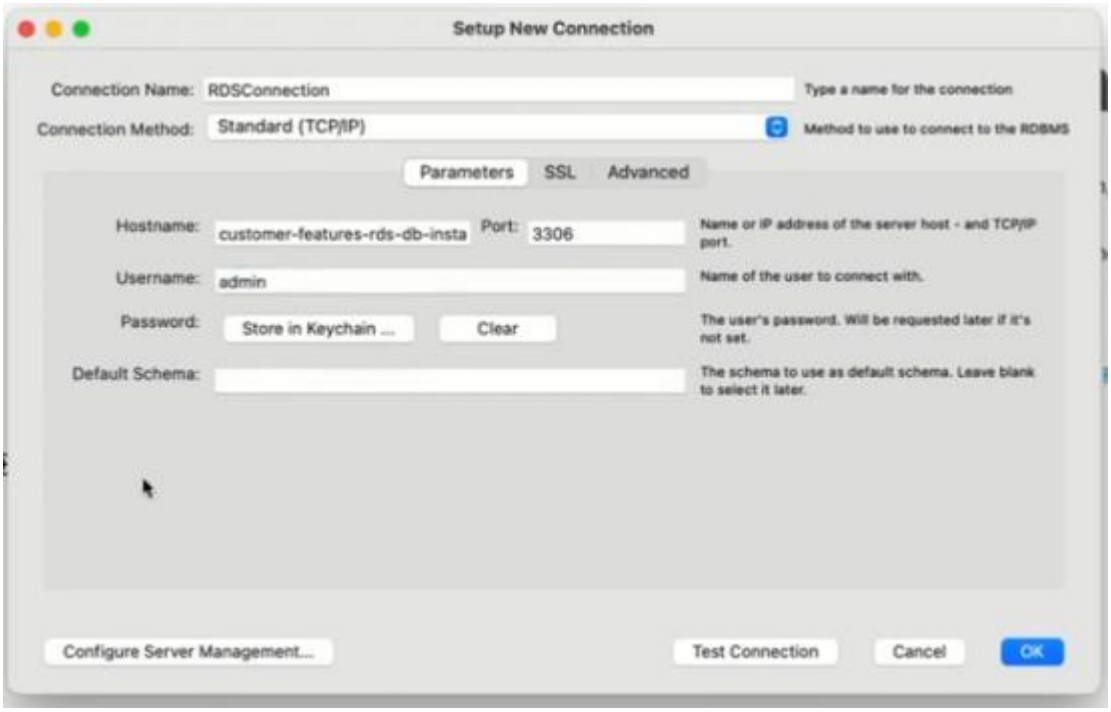



**COPY THE ENDPOINT DETAILS. GO BACK TO THE MYSQL WORKBENCH, THEN, PASTE IT TO THE HOSTNAME. THE PORT 3306 IS THE PORT OF MY MYSQL, SO, IT WILL REMAIN THE SAME.**

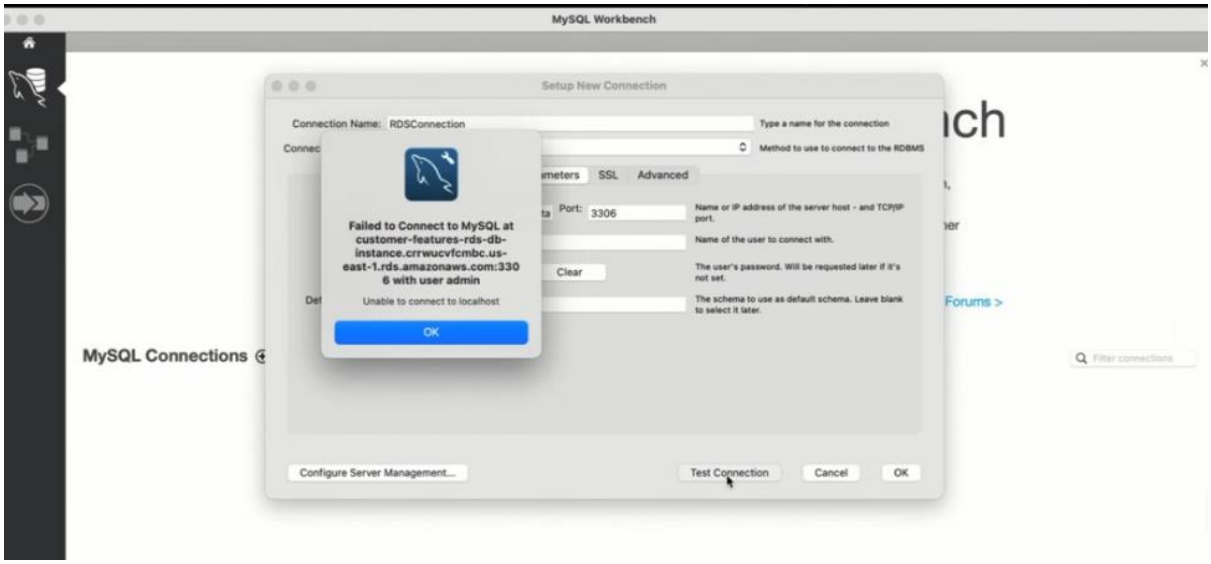



**FOR THE USERNAME, PUT admin, AS IT IS THE NAME WE PUT IN THE DB INSTANCE.**

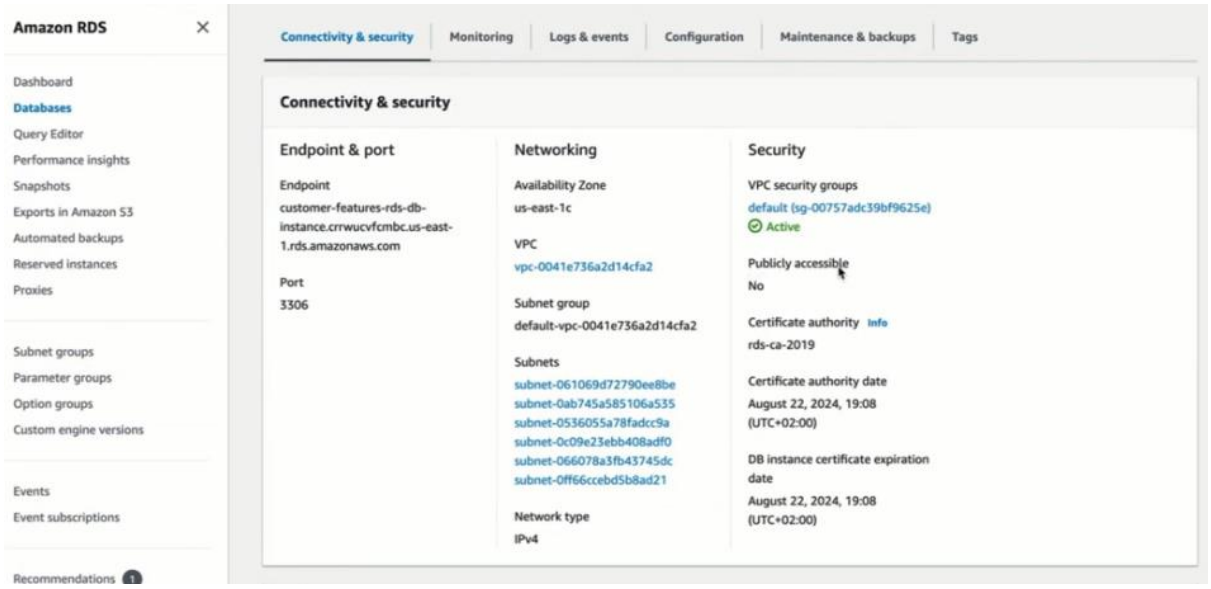
**THEN, CICK, TEST CONNECTION TO TEST IF IT WORKS PROPERLY.**



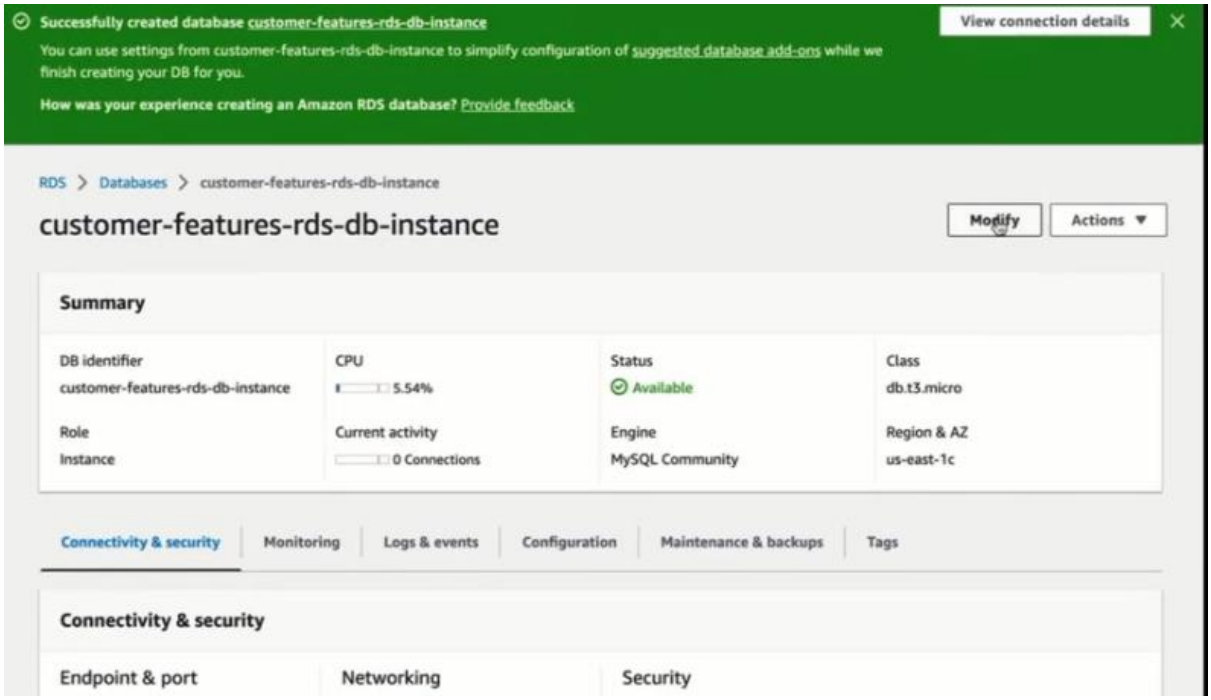

49. **THE RESULT WILL GIVE US FAILED CONNECTION.**




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

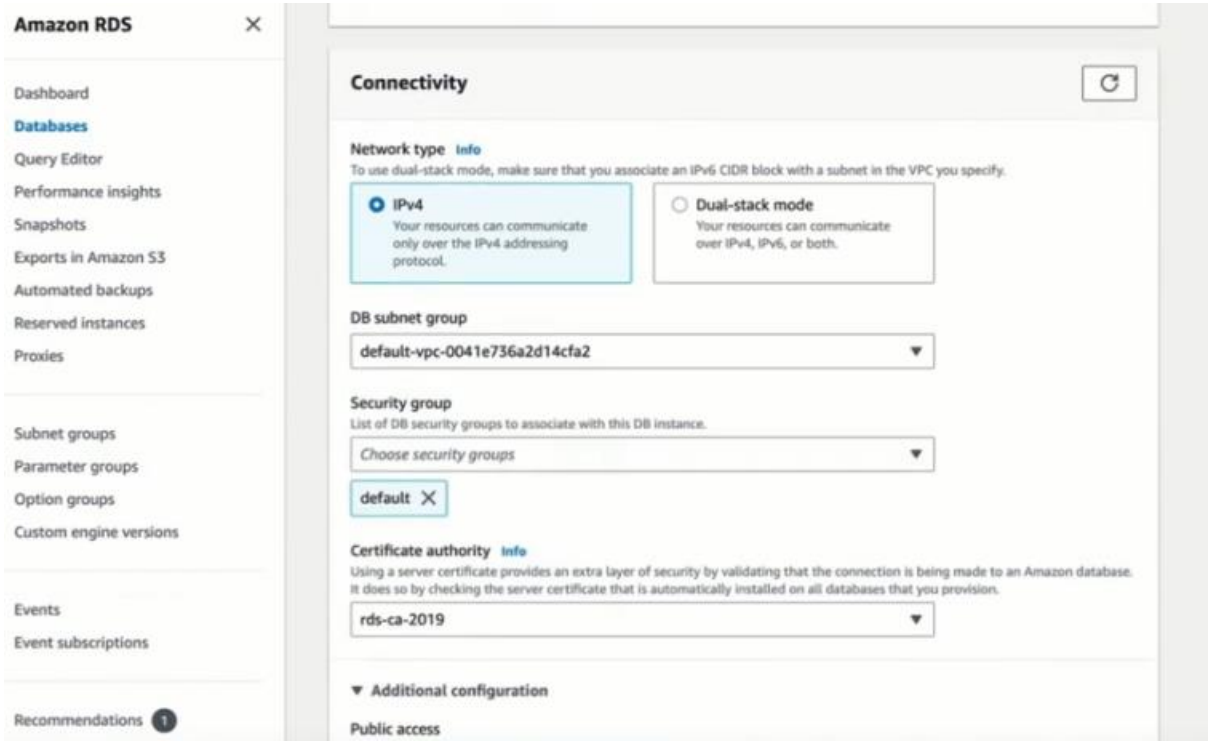


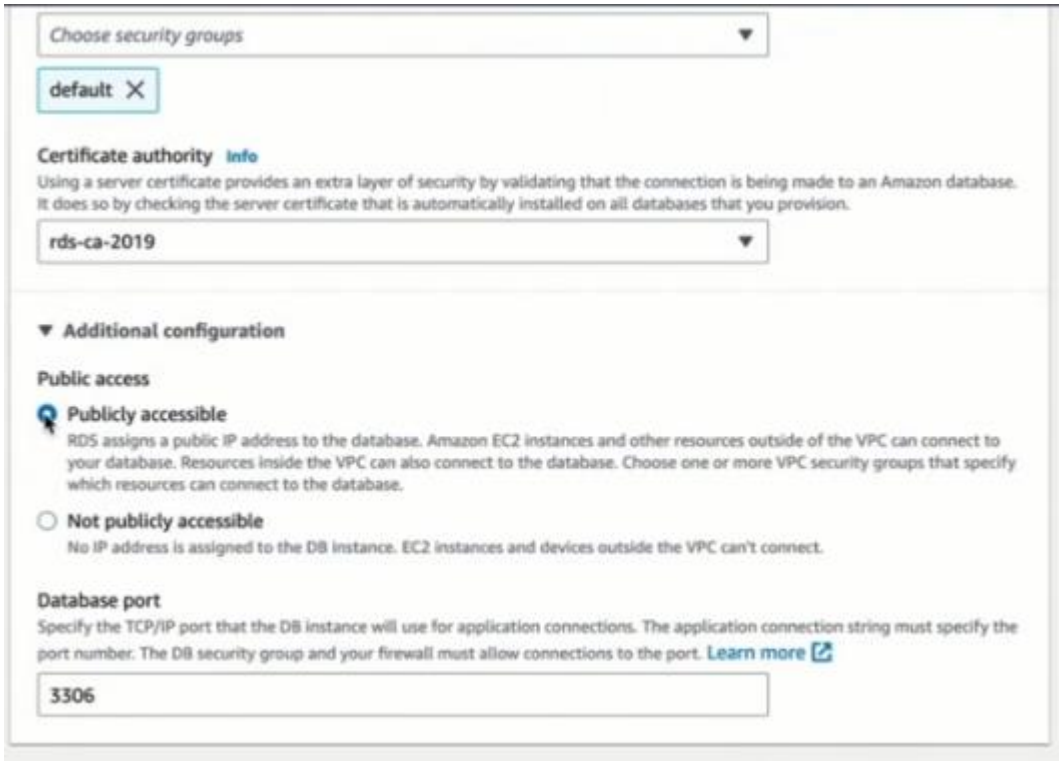

**51. WE HAVE TO CONFIGURE THE SECURITY – PUBLICLY AVAILABLE TO YES. TO DO THIS, CLICK MODIFY.**




**SCROLL DOWN, AND FIND THE CONFIGURATION WHERE WE CAN SET THE PUBLIC ACCESS TO YES.**

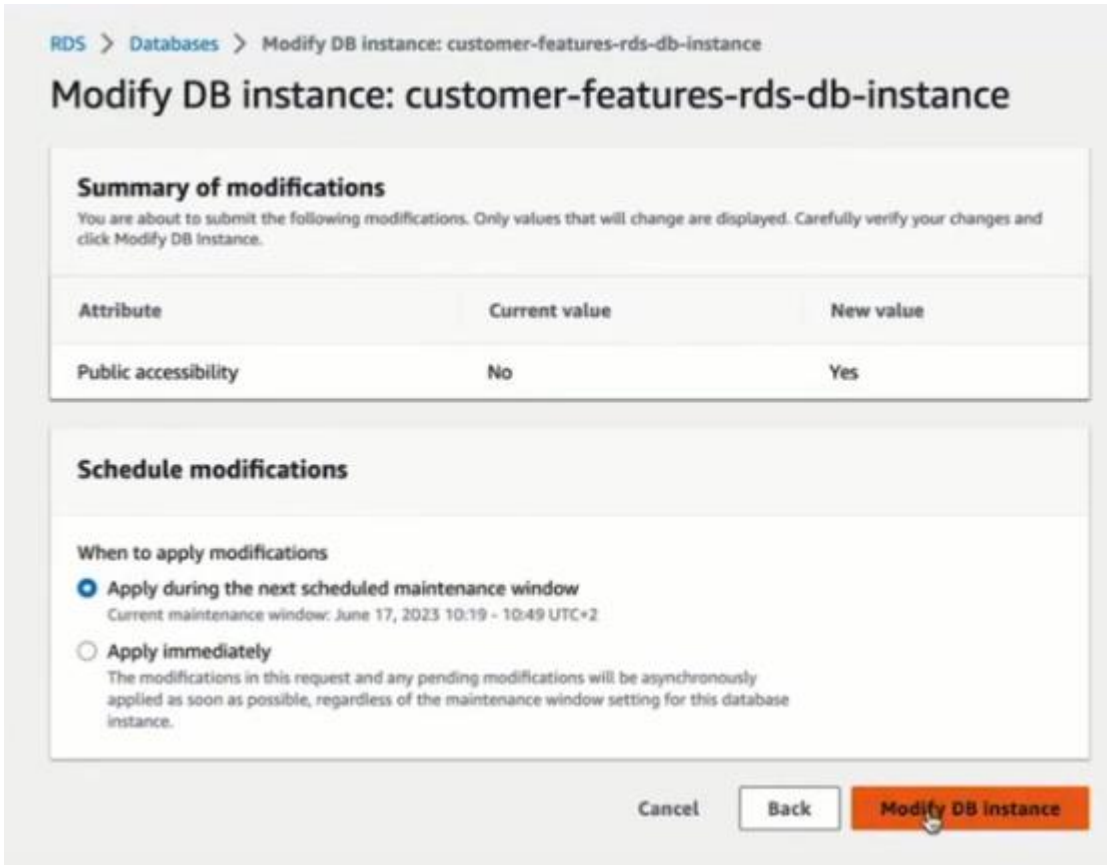
**IN THE CONNECTIVITY, CLICK ADDITIONAL CONFIGURATION. SELECT, PUBLICLY ACCESSIBLE.**

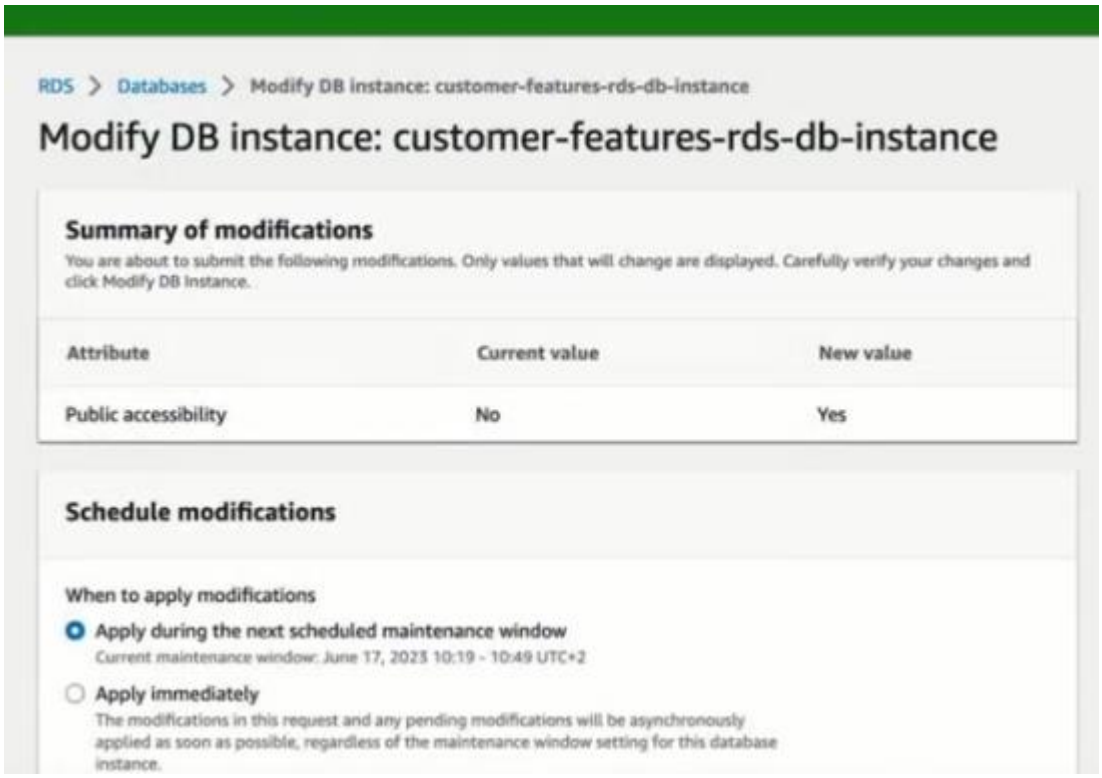




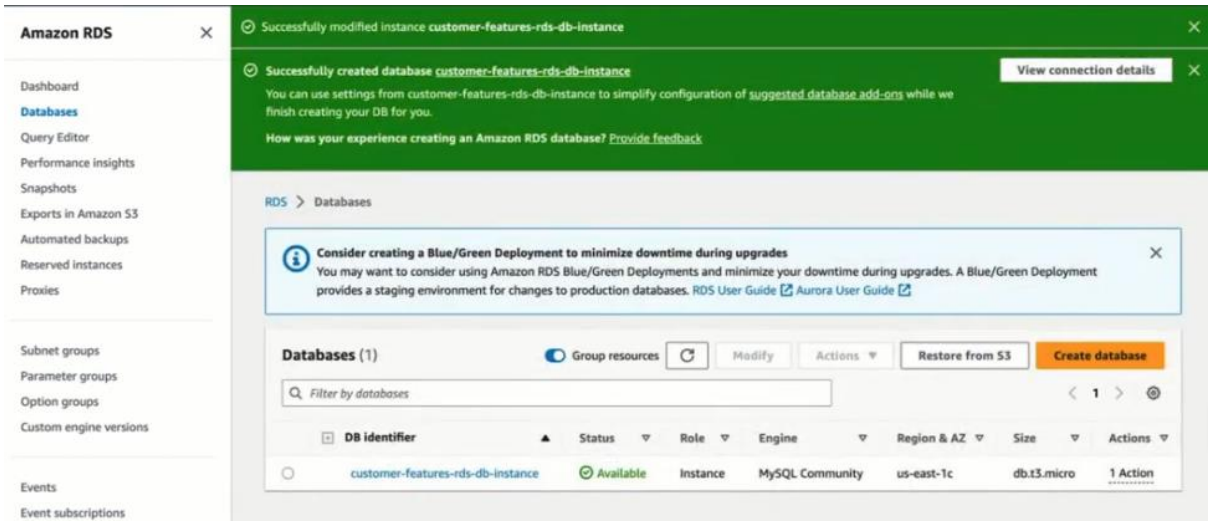

**SCROLL DOWN, THEN SELECT, CONTINUE. YOU WILL SEE WHAT IS MODIFIED.**

**CLICK MODIFY DB INSTANCE.**

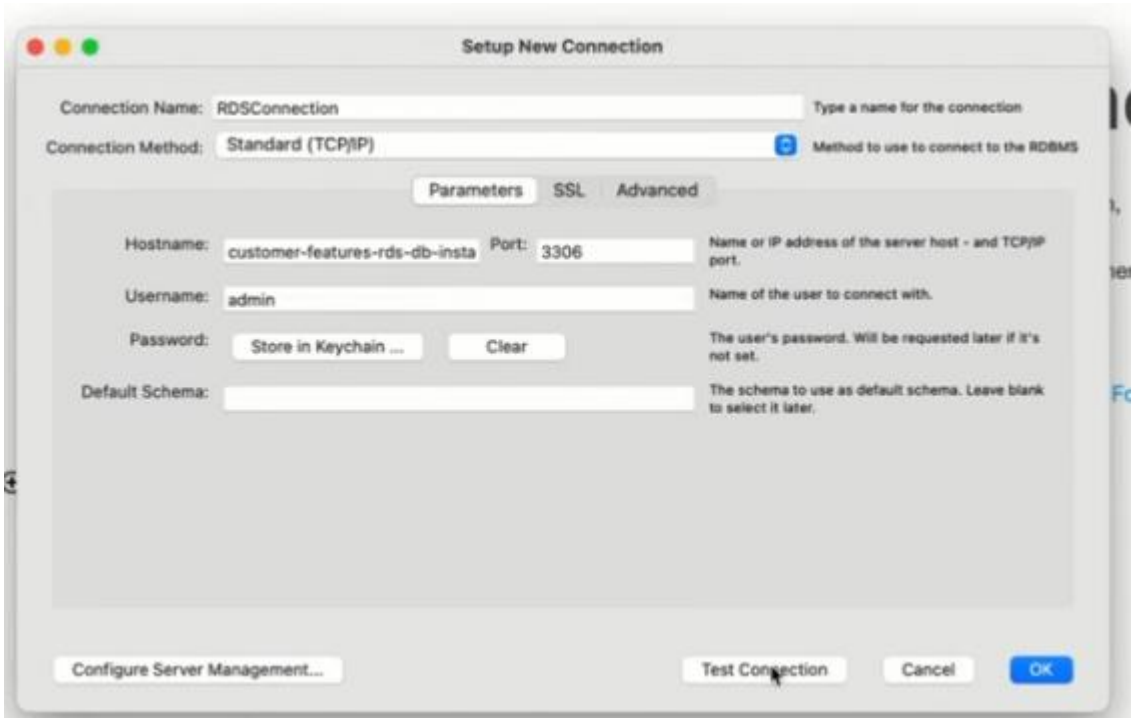

RDS > Databases > Modify DB instance: customer-features-rds-db-instance
Modify DB instance: customer-features-rds-db-instance
Summary of modifications
You are about to submit the following modifications. Only values that will change are displayed. Carefully verify your changes and click Modify DB Instance.
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


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.

**SUCCESSFULLY MOODIED INSTANCE.**




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

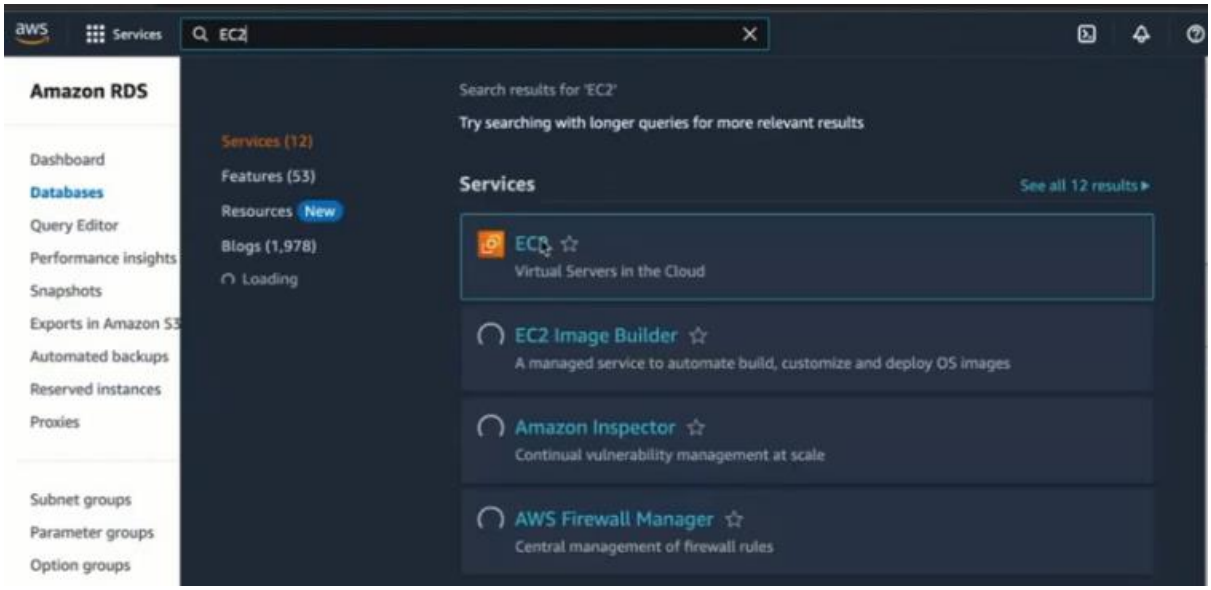



**53. HERE, WE WILL GET THE FAILED CONNECTION AGAIN.**

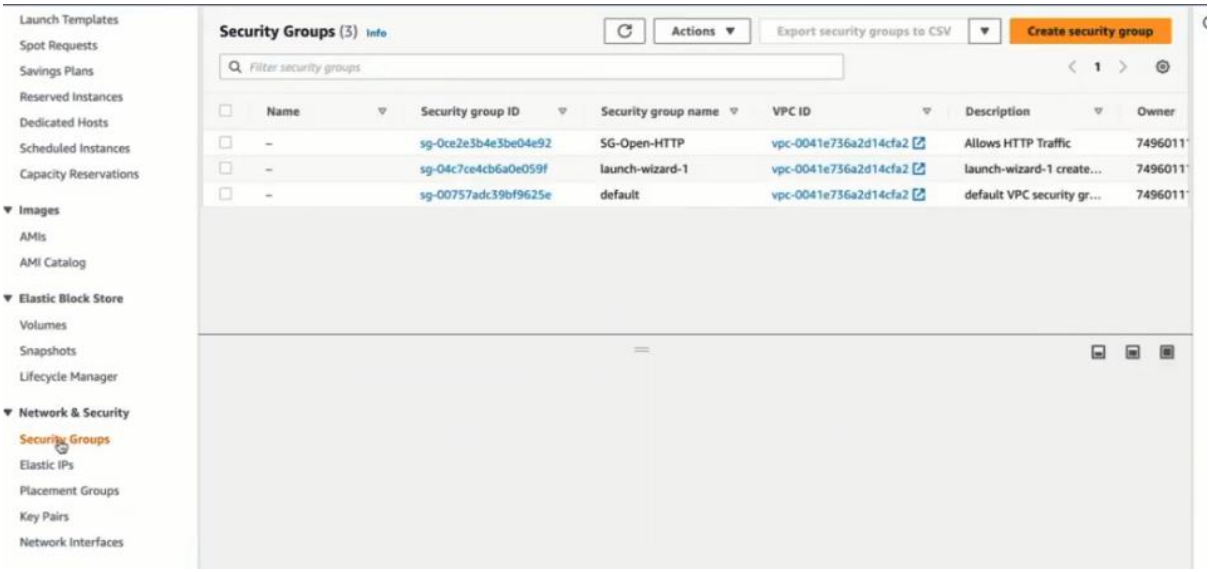

**54. THIS IS BECAUSE WE FORGOT TO EDIT THE INBOUND RULE. SO, TO ACHIEVE THE CONNECTION, WE NEED TO FIX THE ACCESS ISSUE, SPECIFICALLY THE INBOUND RULES OF THE SECURITY GROUP ATTACHED TO THE RDS DATABASE INSTANCE.**

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

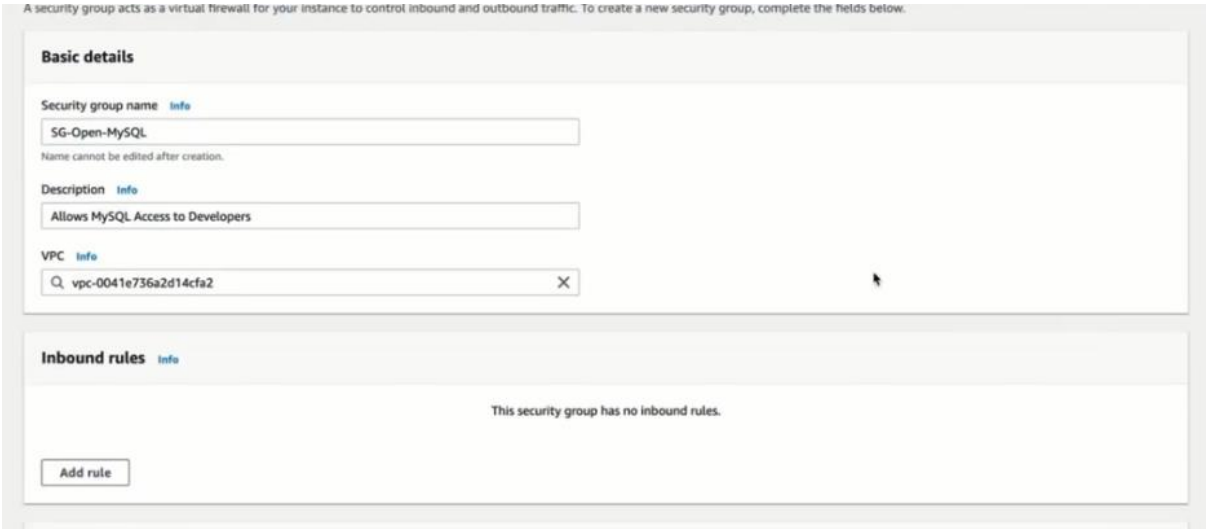
**IN THE SEARCH BAR, GO TO EC2.**



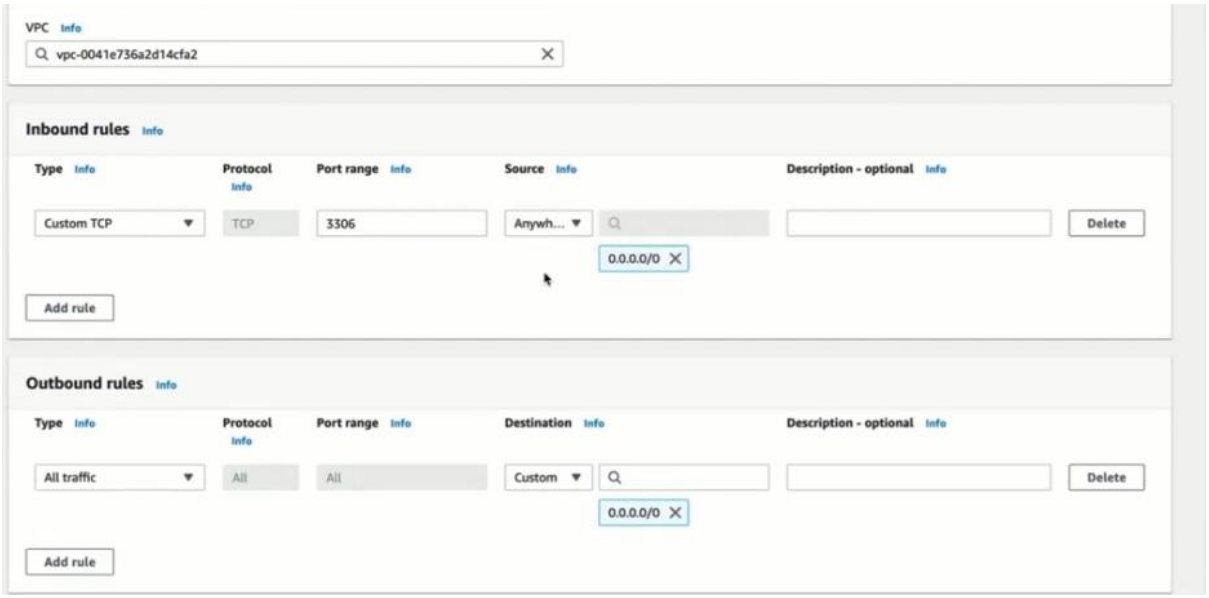
**55. GO TO THE SECURITY GROUP AND CREATE SECURITY GROUP.**



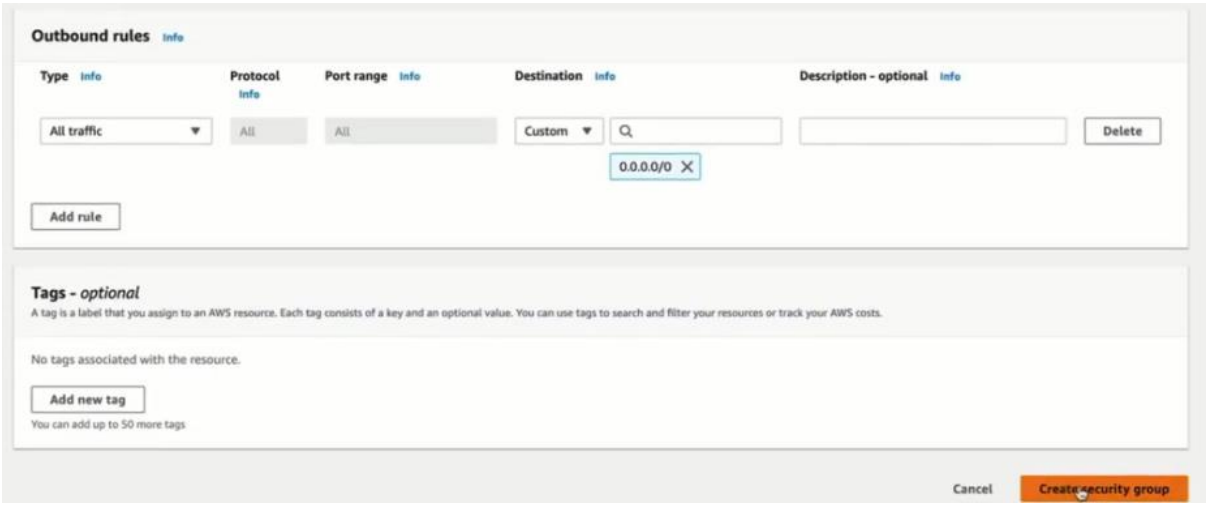

**56. WE WILL CALL IT, SG-OPEN-MYSQL. DESCRIPTION IS ALLOWS MYSQL ACCESS TO DEVELOPERS. AND MAKE THE VPC DEFAULT.**



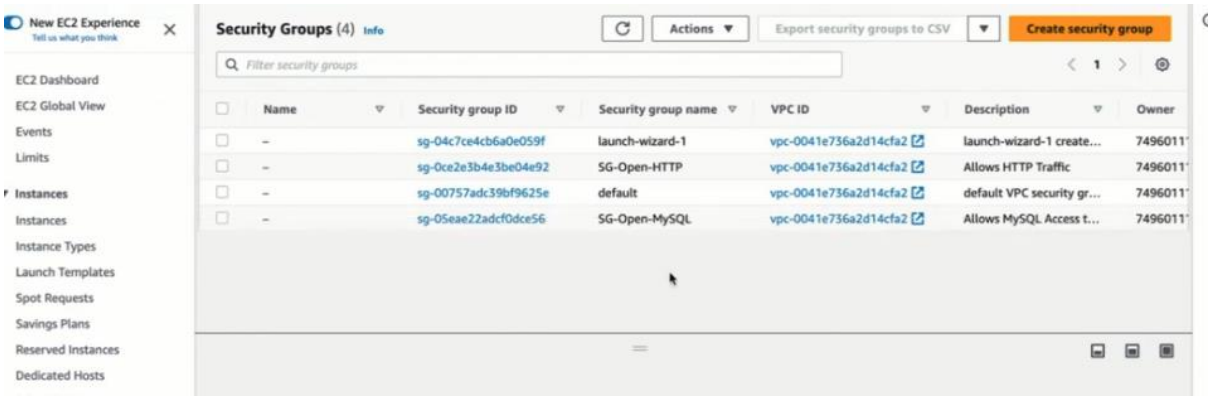

**57. CLICK ADD RULE IN THE INBOUND RULE. IN THE TYPE, SELECT CUSTOME TCP. PROTOCOL IS TCP. PORT IS 3306. DESTINATION IS ANYWHERE.**

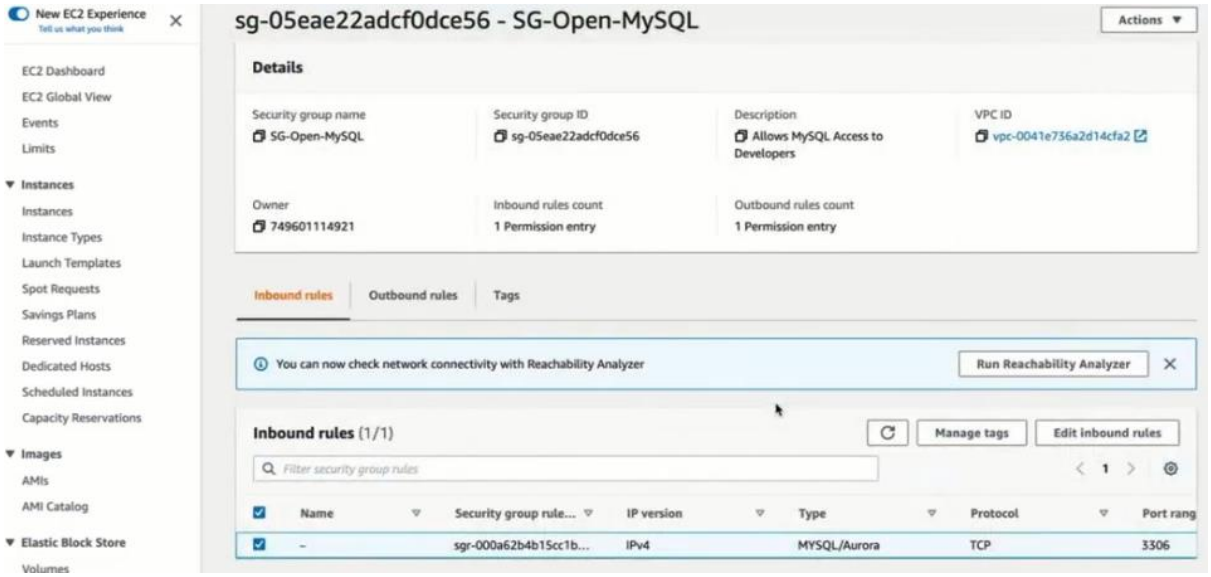



**58. MAKE THE OUTBOUND RULE DEFAULT. THEN, CLICK CREATE SECURITY GROUP.**

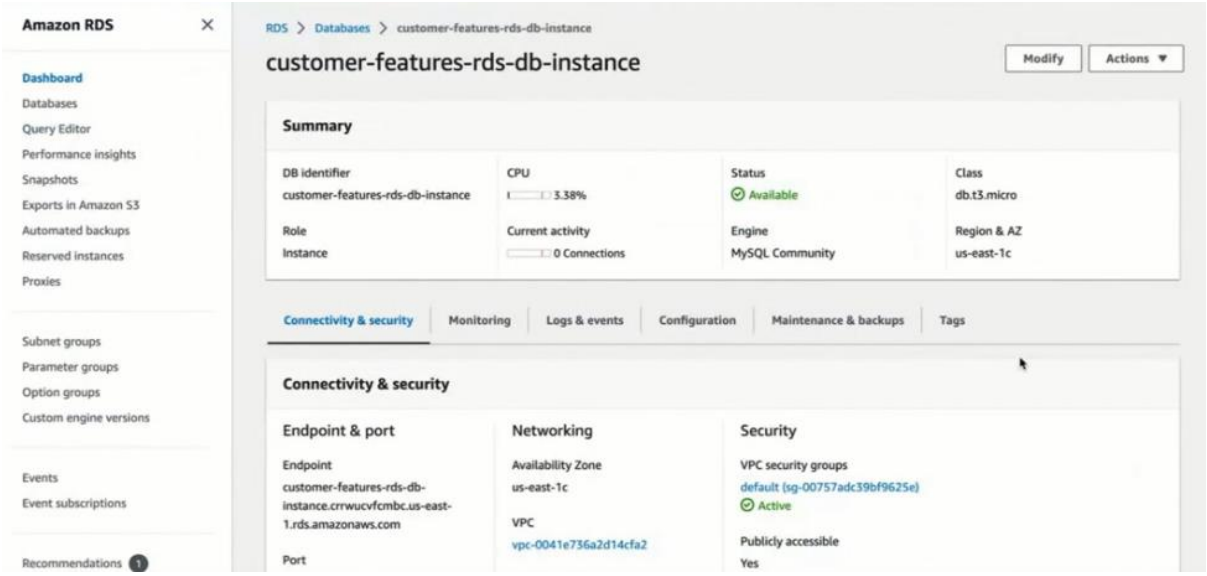



**59. IT IS NOW IN THE SECURITY GROUP LIST. BUT, WE HAVE TO ATTACH IT TO THE RDS DATABASE INSTANCE.**

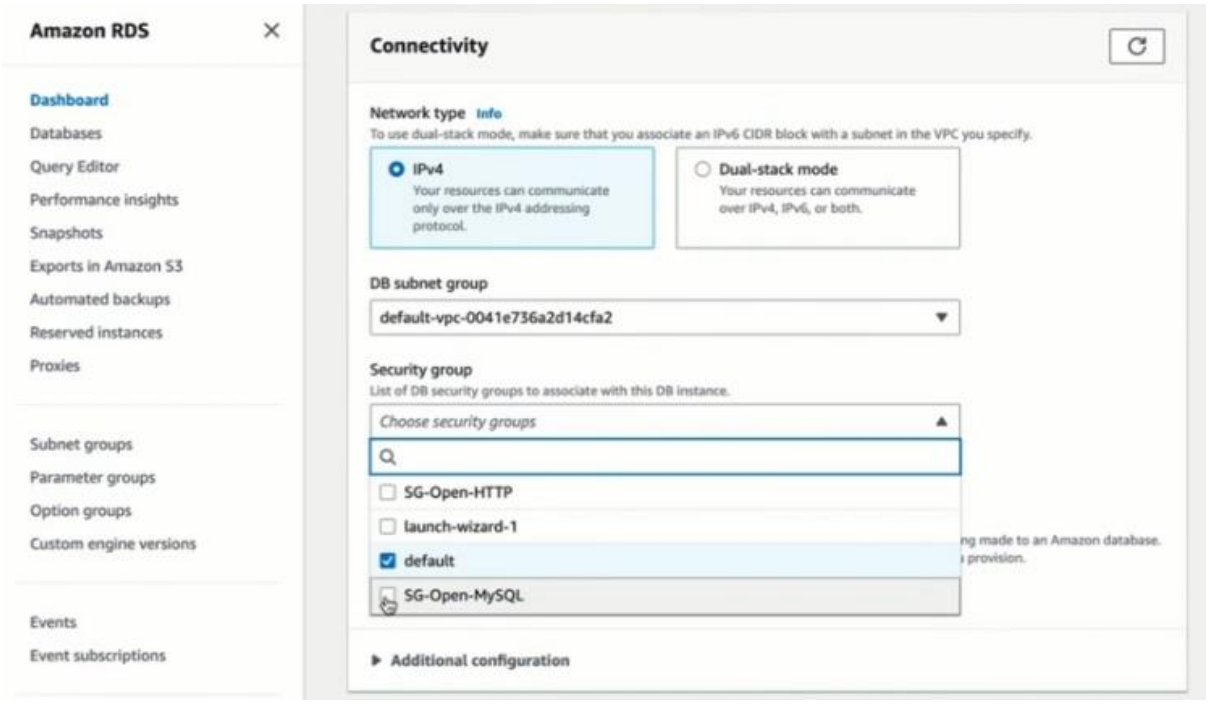


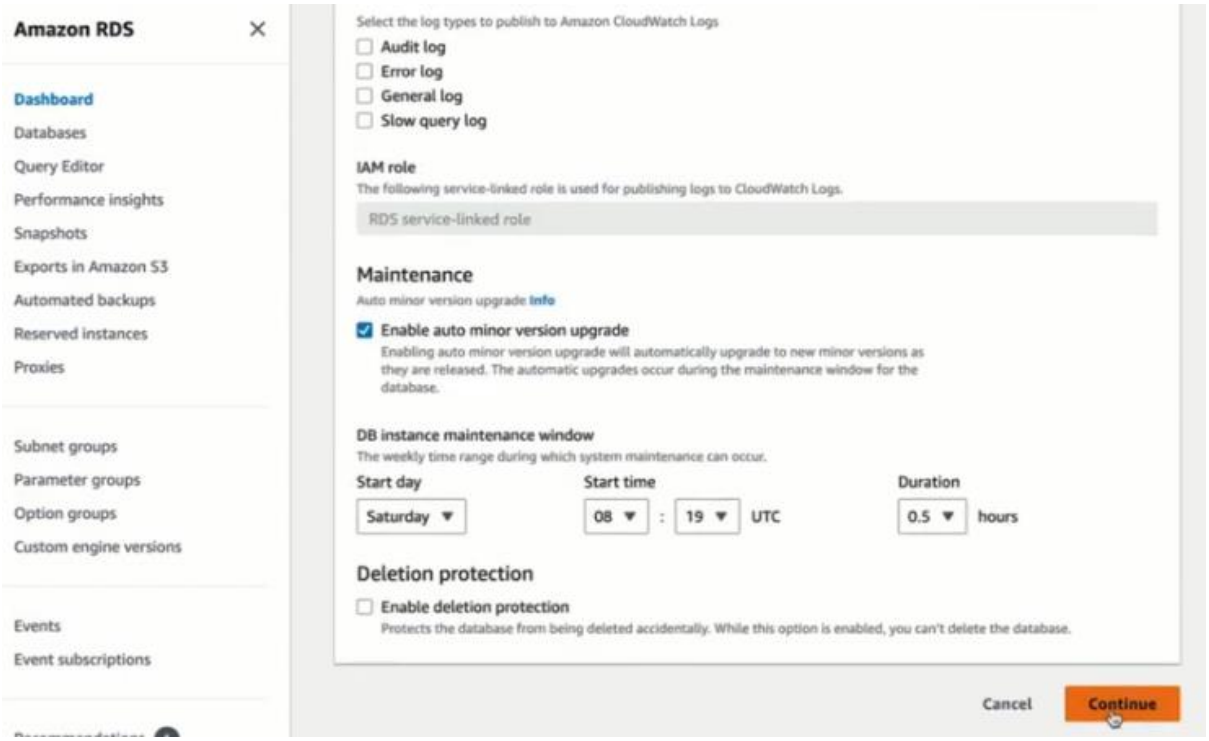

**60. GO BACK TO AMAZON RDS, THEN, CLICK MODIFY.**



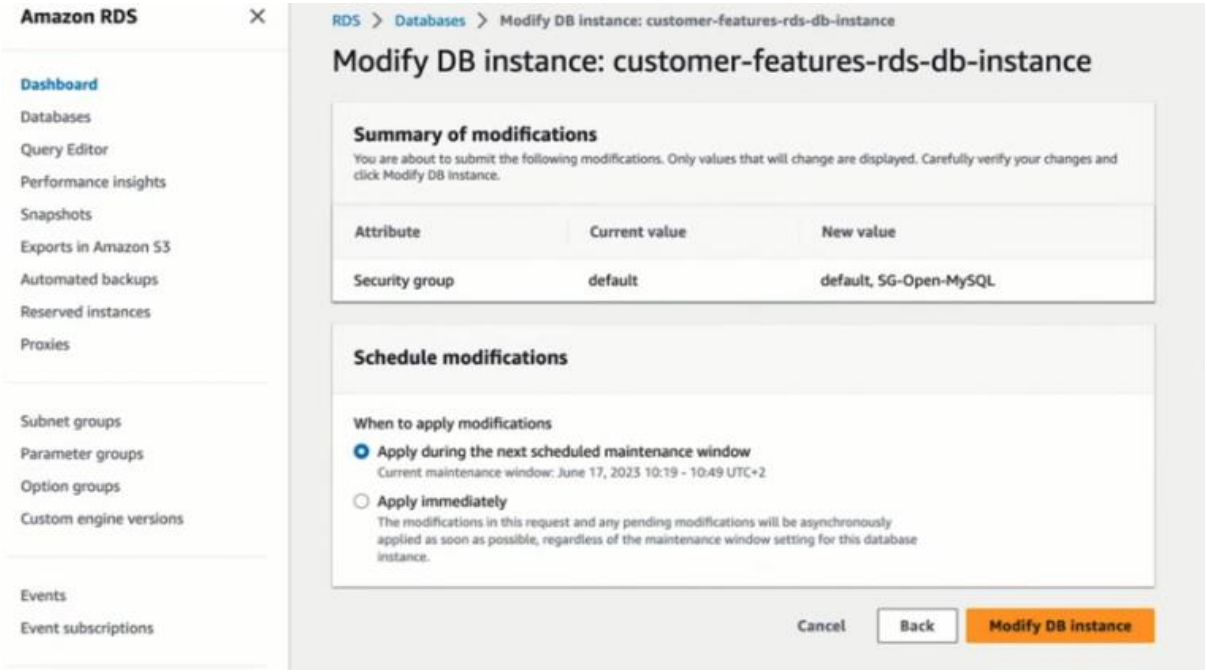
61. **SCROLL DOWN AND GO TO CONNECTIVITY, THEN, SECURITY GROUP. THEN YOU WILL FIND THE ONE THAT WE JUST CREATED. CHOOSE DEAFULT AND SG-OPEN-MYSQL.**



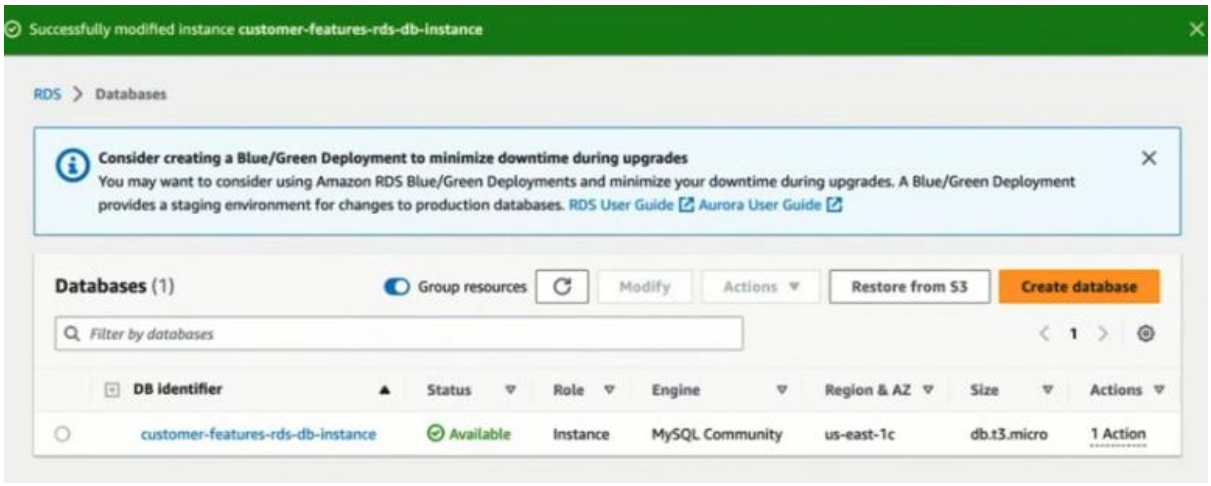

62. **SCROLL DOWN AND CLICK CONTINUE.**



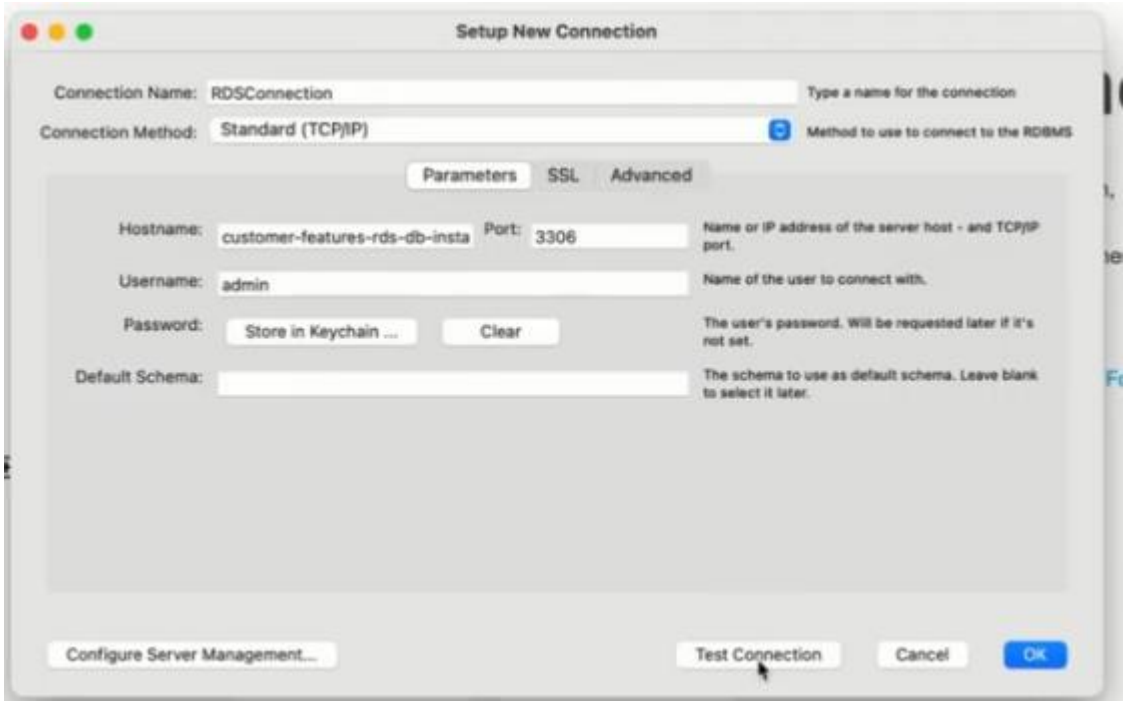
63. **HERE, YOU WILL SEE THE SUMMARY OF MODIFICATIONS. THEN, CLICK MODIFY DB INSTANCE.**



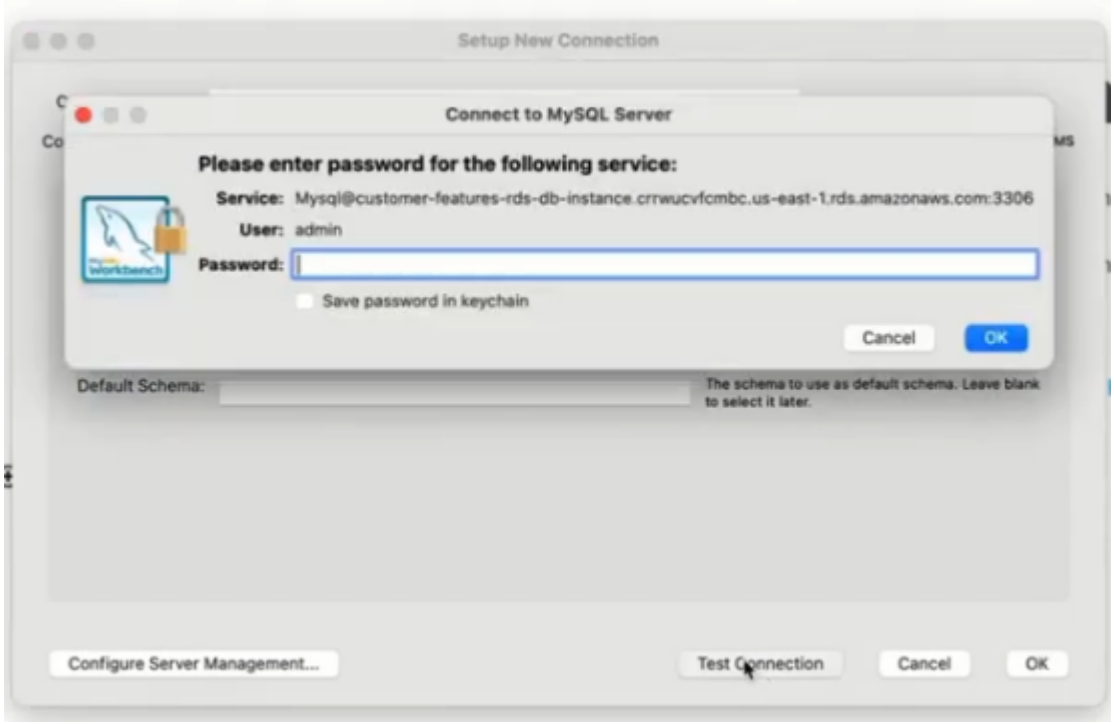

64. **SUCCESSFULLY MODIFIED.**




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

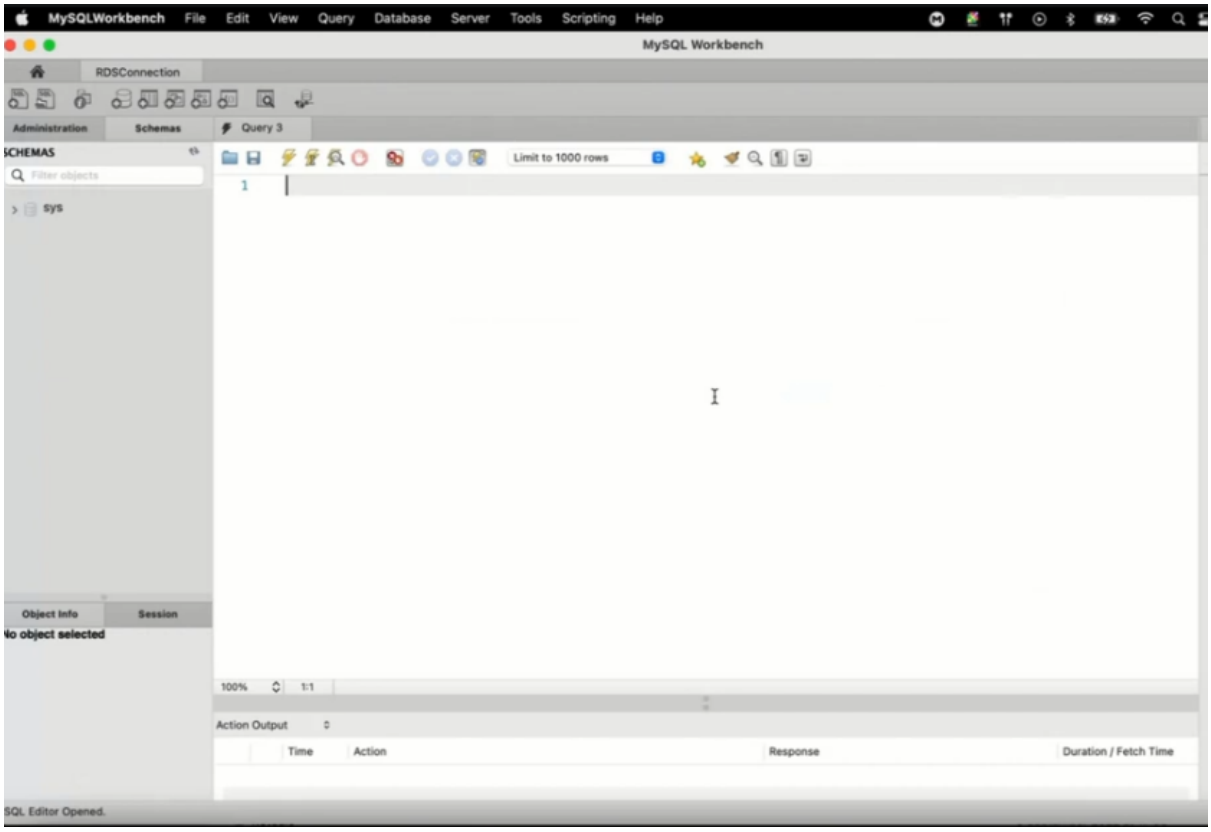



**CONGRATULATIONS! THE CONNECTION IS SUCCESSFUL! ENTER THE PASSWORD.**

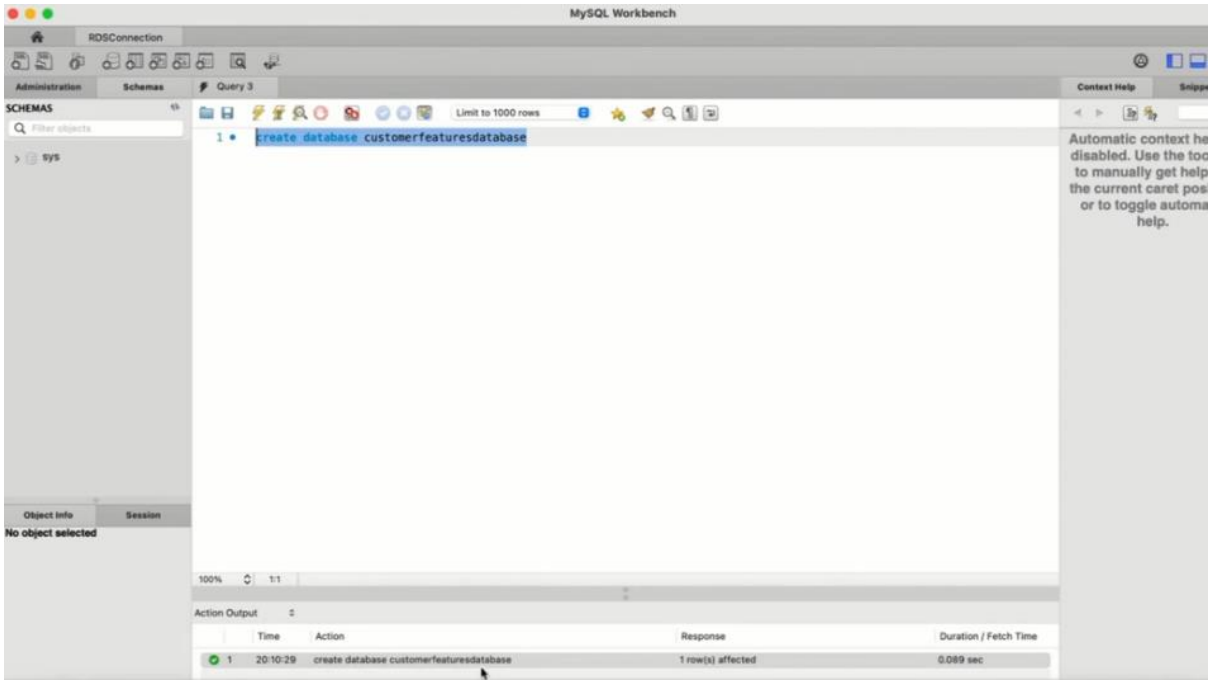



66. **THE CONNECTION IS NOW ESTABLISHED, DOUBLE CLICK THE RDS CONNECTION. IT IS NOW OPENING THE SQL EDITOR.**

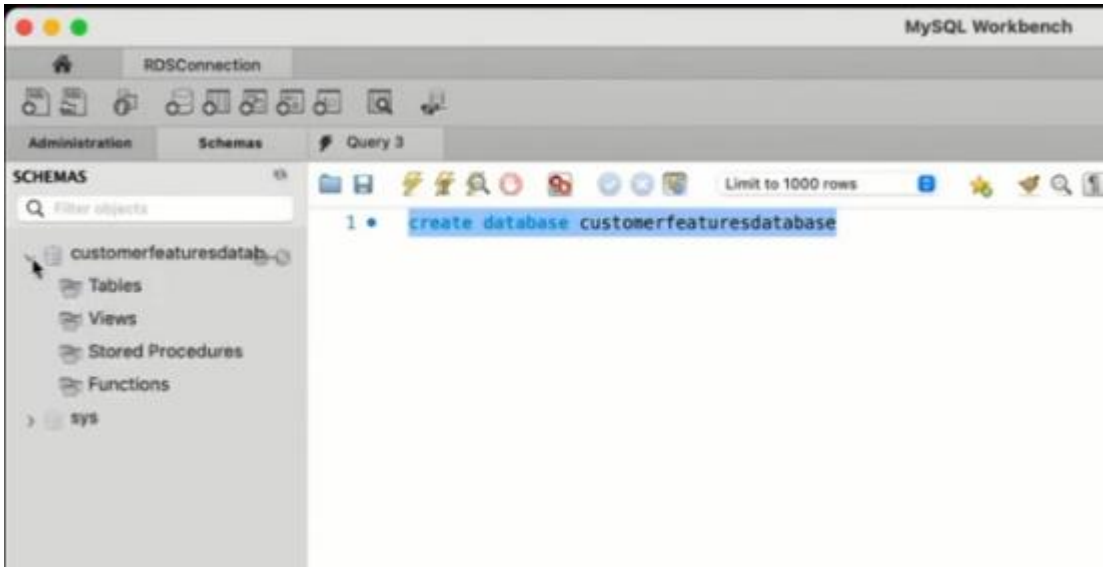

67. **AND THIS IS WHERE WE ARE ABLE TO CREATE OUR TABLE WITHIN OUR RDS DATABASE INSTANCE.**



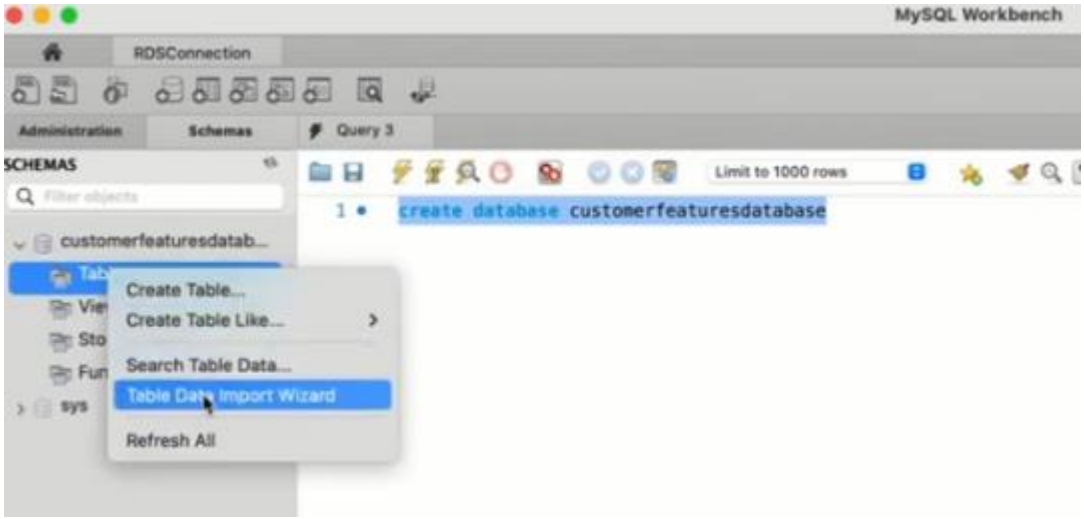

**68. WE WILL NOW RUN A QUERY, CREATE DATABASE…THEN EXECUTE.**



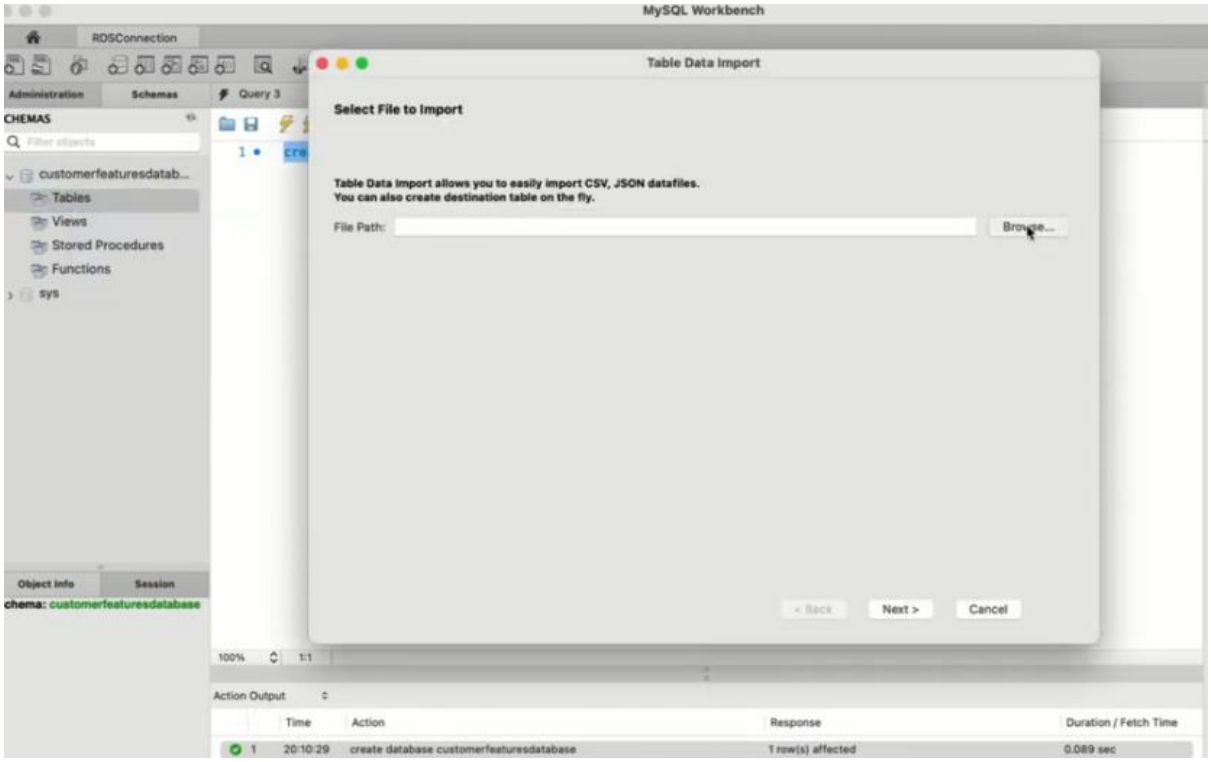

**69. REFRESH THE SCHEMA TO SEE THE DATABASE.**



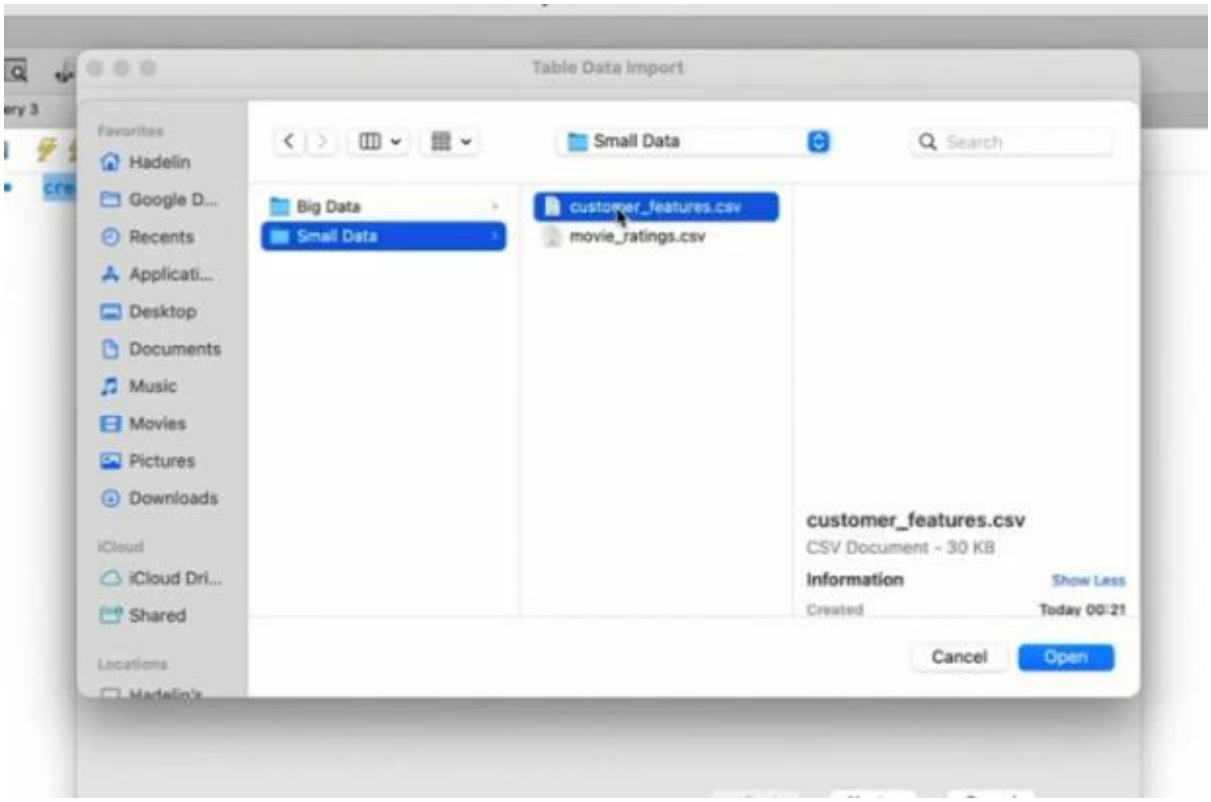

**70. CLICK TABLE, CLICK RIGHT, SELECT TABLE DATA IMPORT WIZARD.**



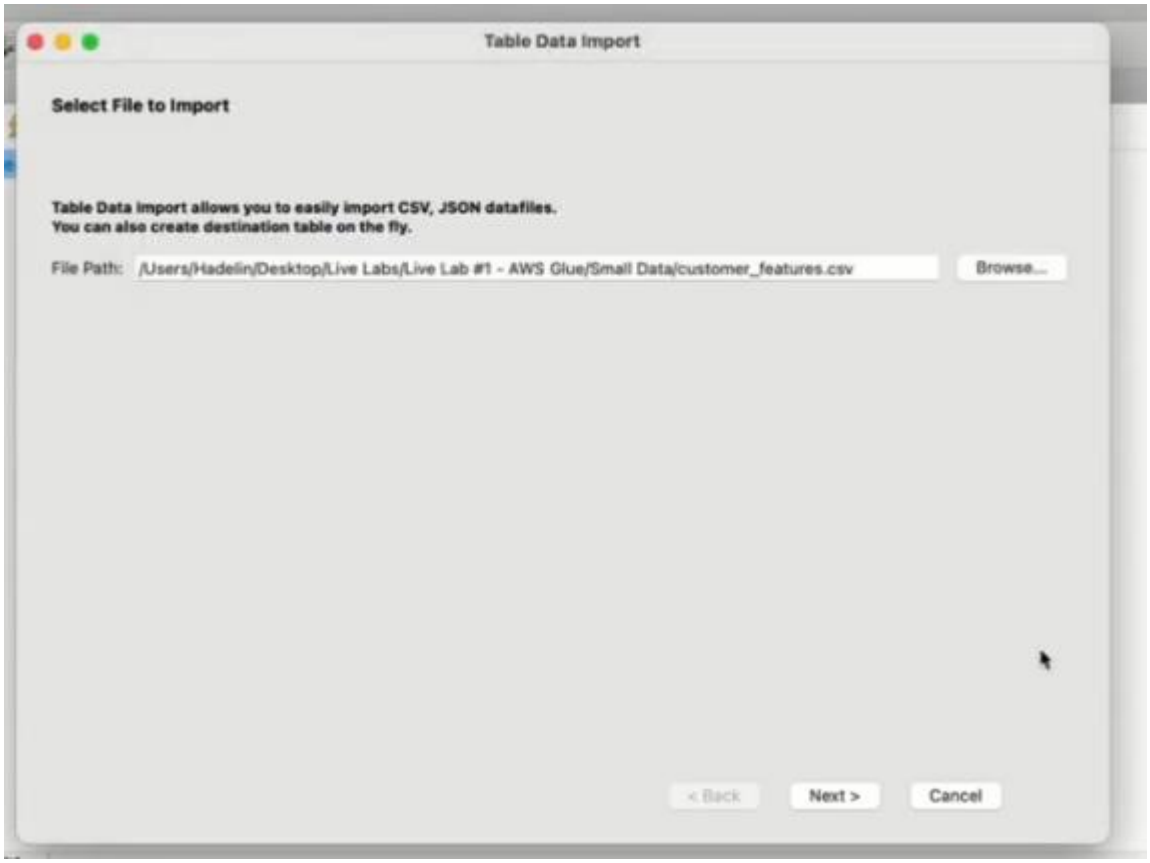

**71. BROWSE THE MACHINE TO FIND THE CSV FILE.**



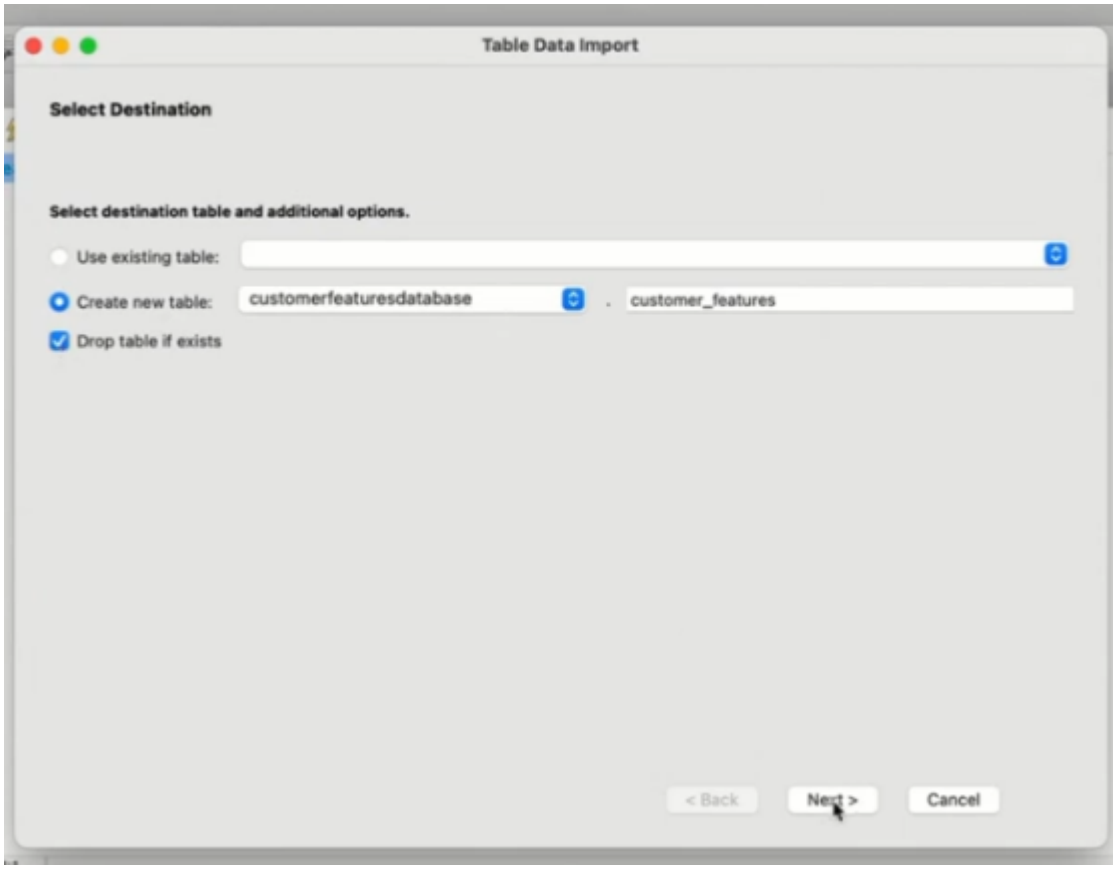

**HERE, I CHOSE THE SMALL DATA, CUSTOMER FEATURES FILE.**



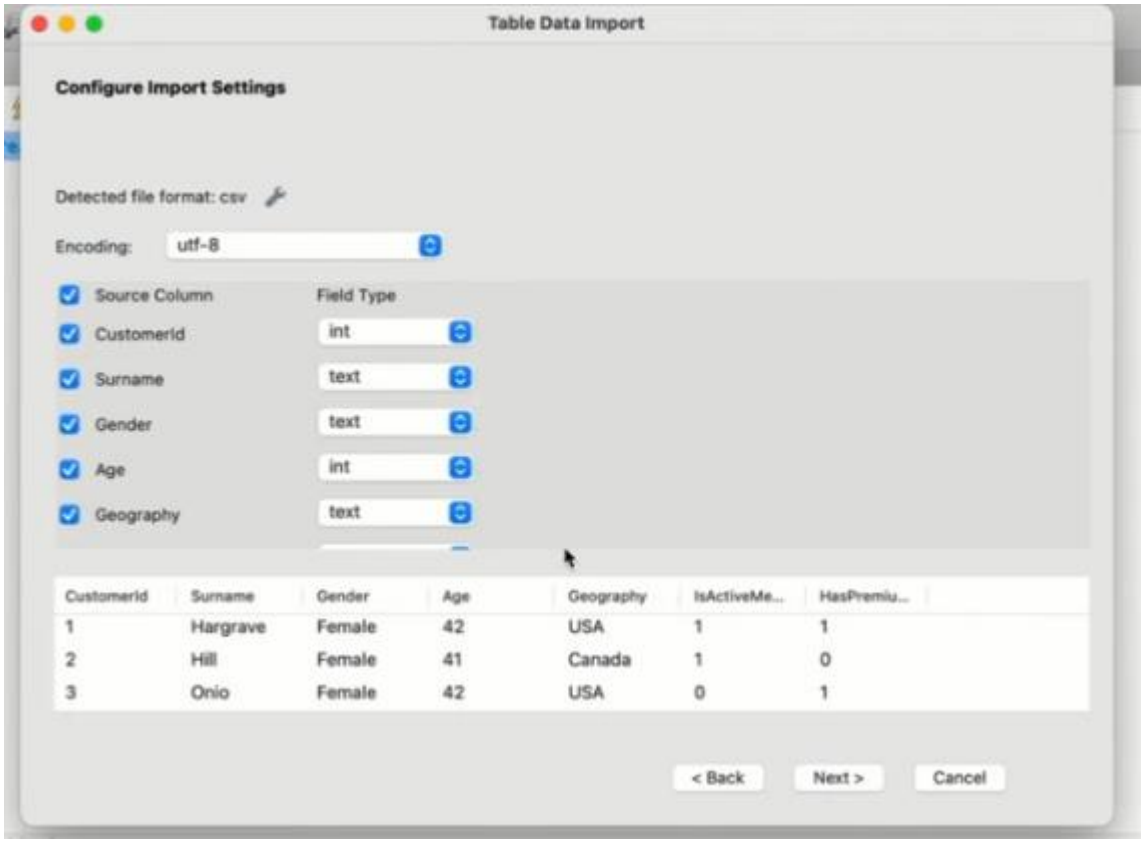
**72. CLICK NEXT.**




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



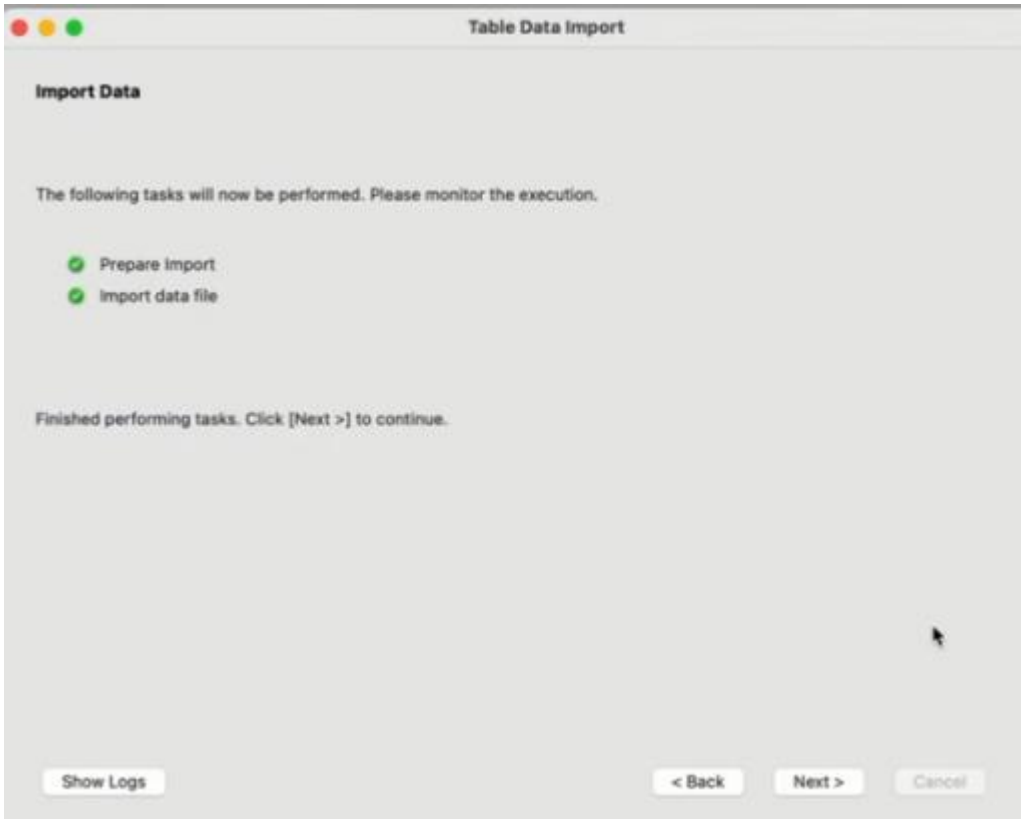

**74. YOU CAN CHANGE THE TYPE IF YOU WANT. BUT, I JUST LEAVE IT AS IT IS. CLICK, NEXT.**



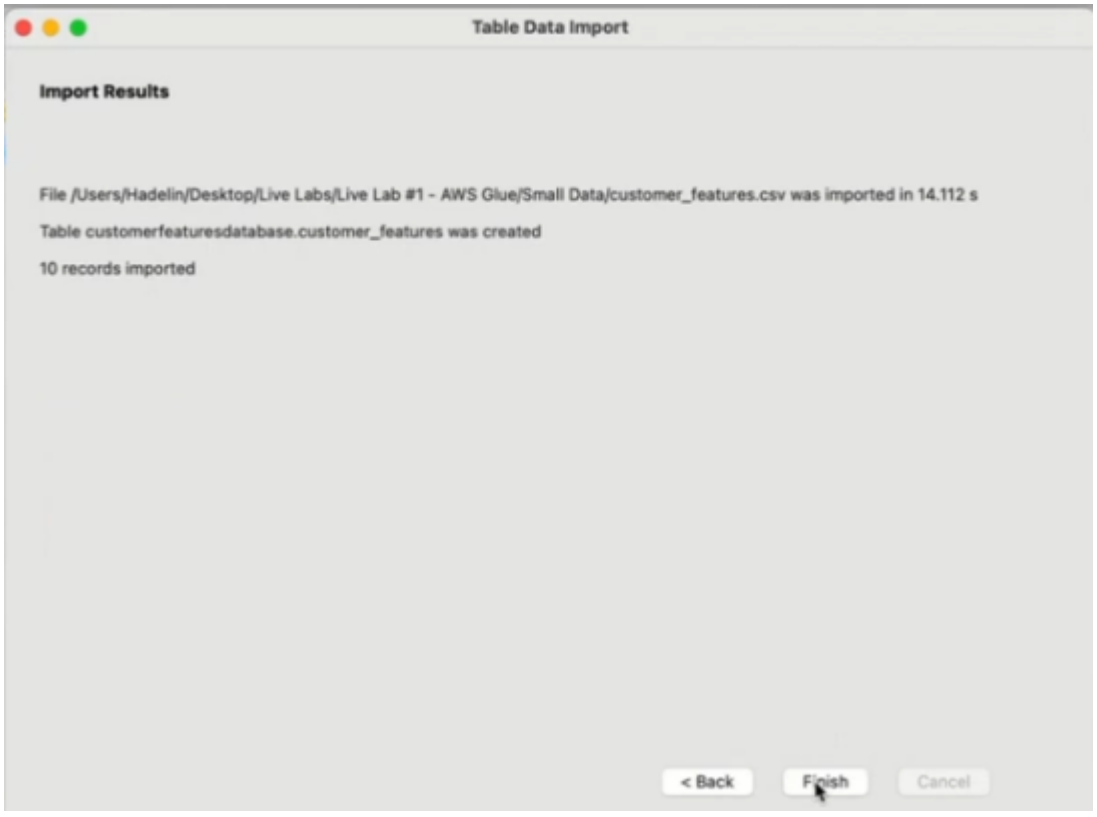
75. **IN THE IMPORT DATA, CLICK NEXT TO EXECUTE.**


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

    **CLICK NEXT.**

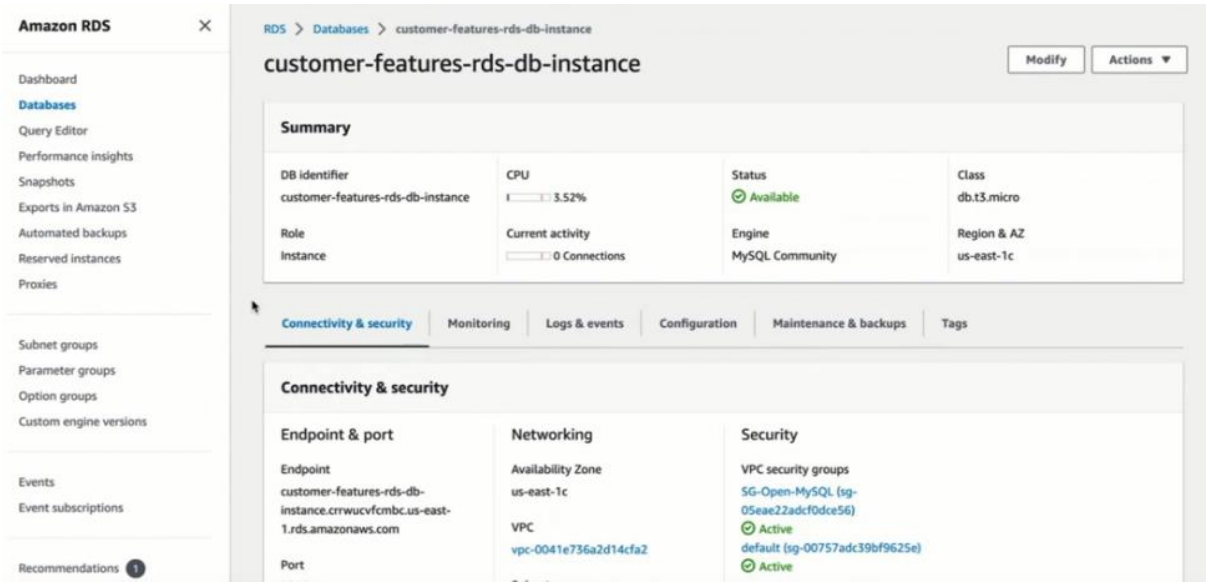



**CLICK FINISH. THE RECORED IS THEN FINISHED.**

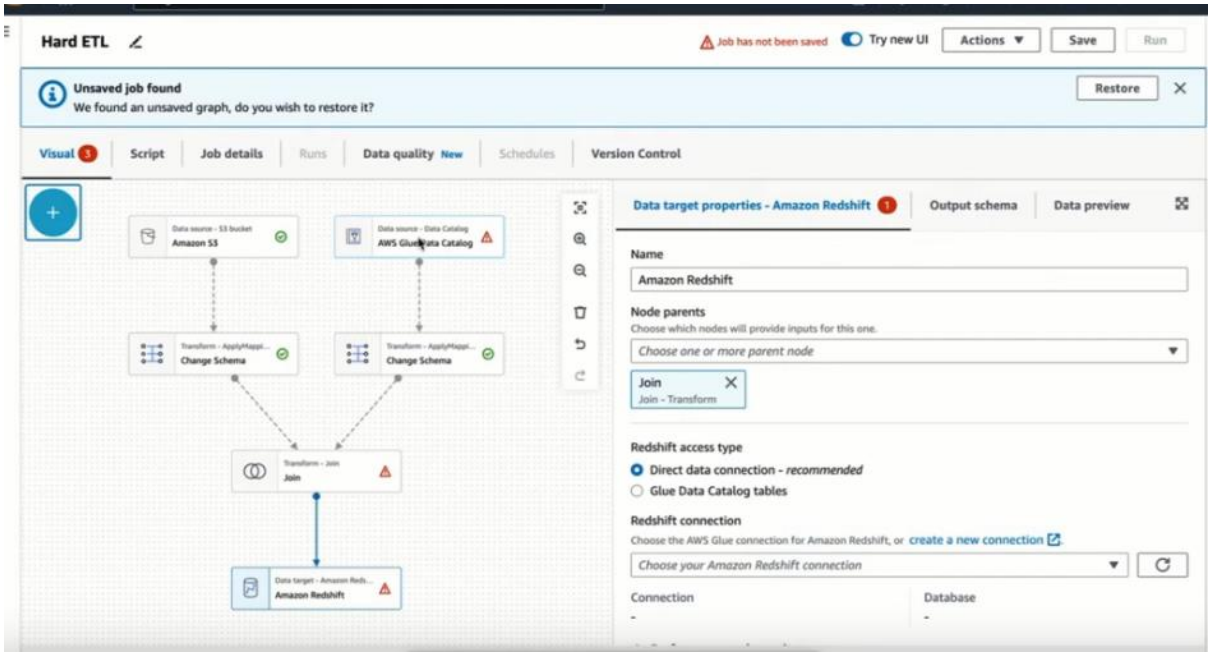


77. **IN THE RDS DATABASE, WE HAVE THE DATA POPULATED.**


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.**

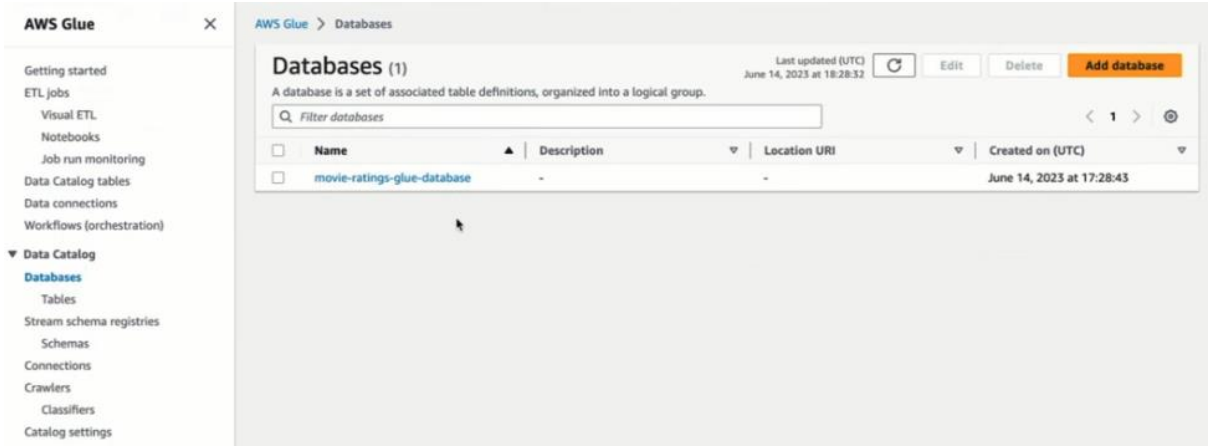



79. **SO IN THE ETL PROCESS, WE HAVE DONE THE EXTRACT STEP. NOW, WE HAVE TO CONNECT OUR RDS TO THE ETL PROCESS. WE HAVE TO CONNECT THE RDS TO GLUE.**

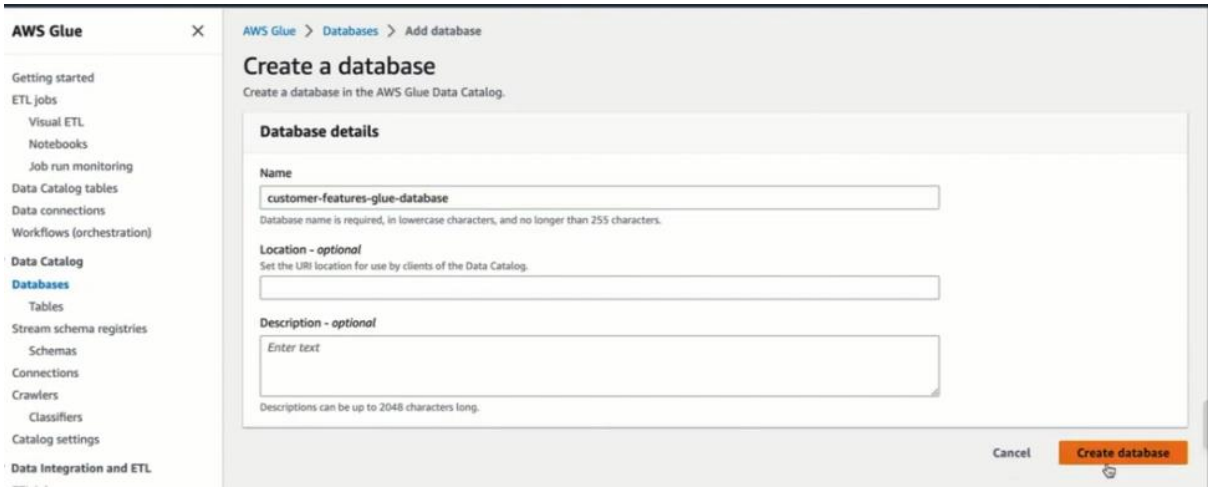



80. **THE NEXT STEP THAT WE NEED TO DO IS TO CREATE AN ELEMENT FOR OUR RDS DATABASE INSTANCE, THEREFORE, AND NOW TABLE IN THE DATA CATAKOG. SO, WE ARE GOING BACK TO THE ETL, DATABASES, AND THIS TIME WE ARE GOING TO CREATE THE CUSTOMER FEATURES DATABASE.**

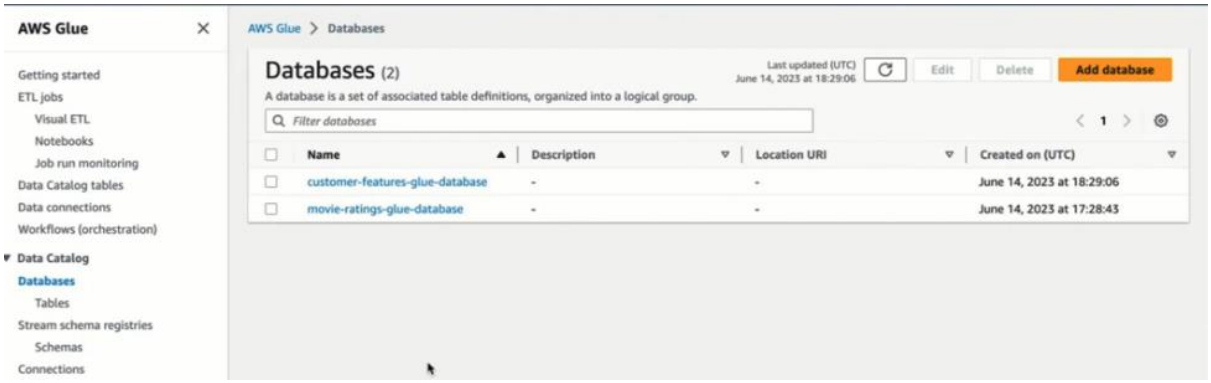
    **CLICK ADD DATABASE.**




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

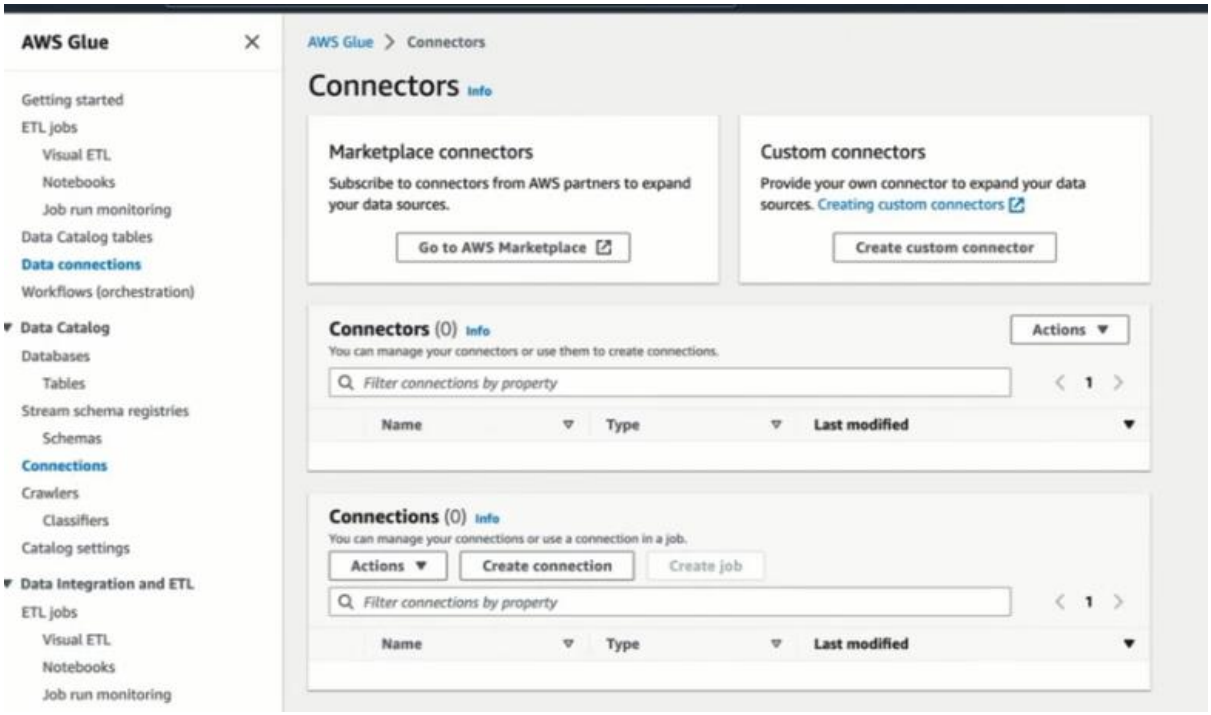



**82. THE SECOND DATABASE IS CREATED. BUT, WE NEED TO CONNECT THE RDS DATABASE THAT WE CREATED AND POPULATE IT TO MYSQL WORKBENCH TO THE GLUE DATABASE.**

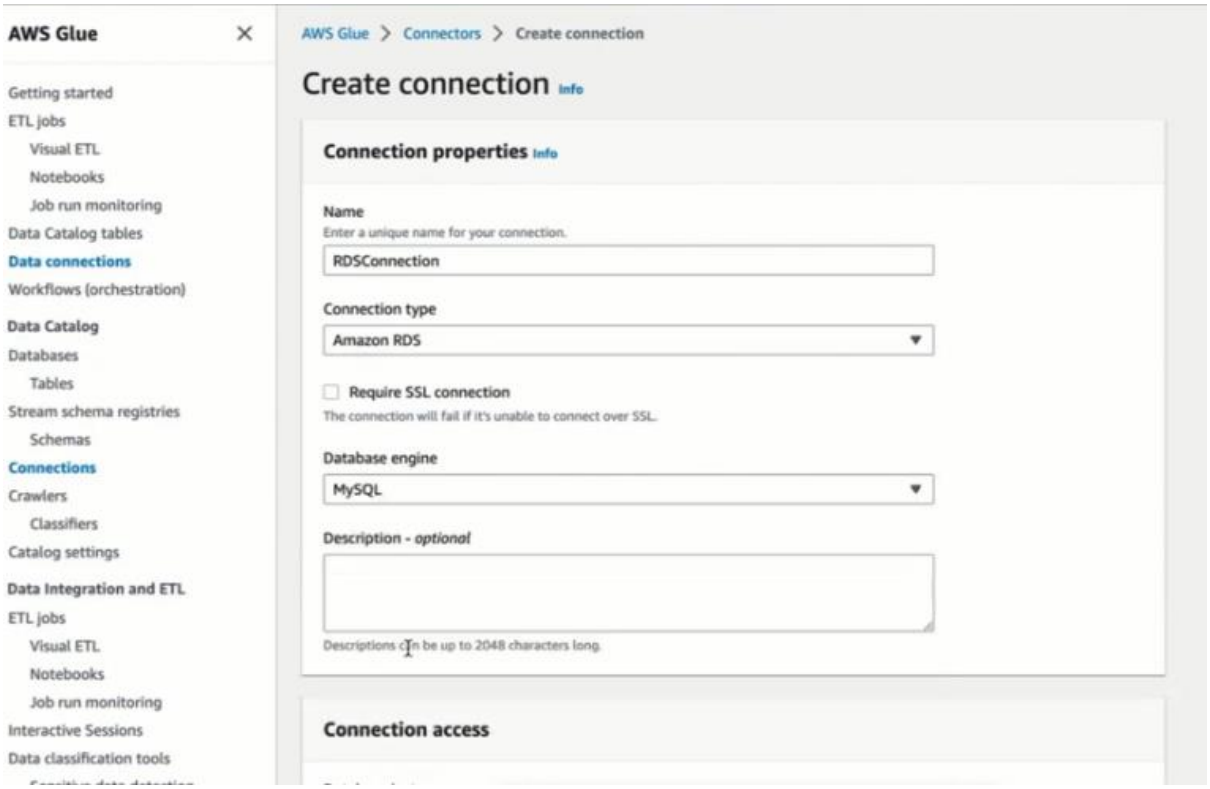



**83. HERE, WE WILL USE THE CONNECTION INSTEAD OF THE CRAWLER. BECAUSE WE HAVE TO ESTABLISH THIS CONNECTION FROM RDS TO AWS GLUE.**

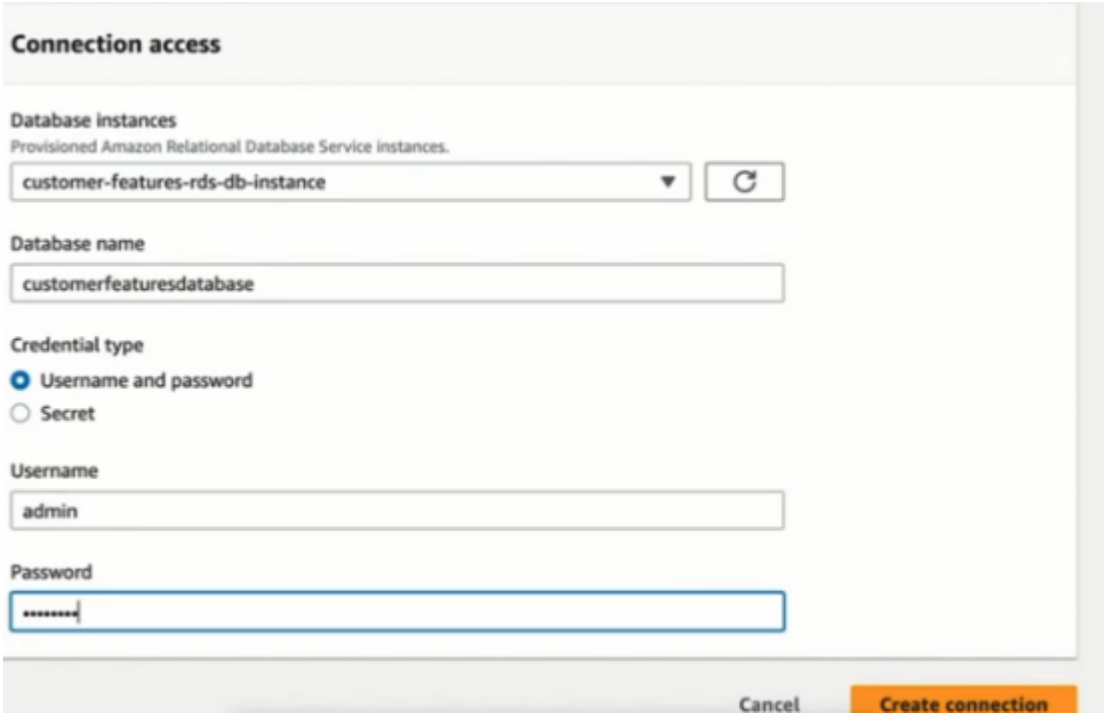
**GO TO AWS GLUE. DATA CATALOG, CONNECTIONS. CLICK CREAT CONNECTIONS.**



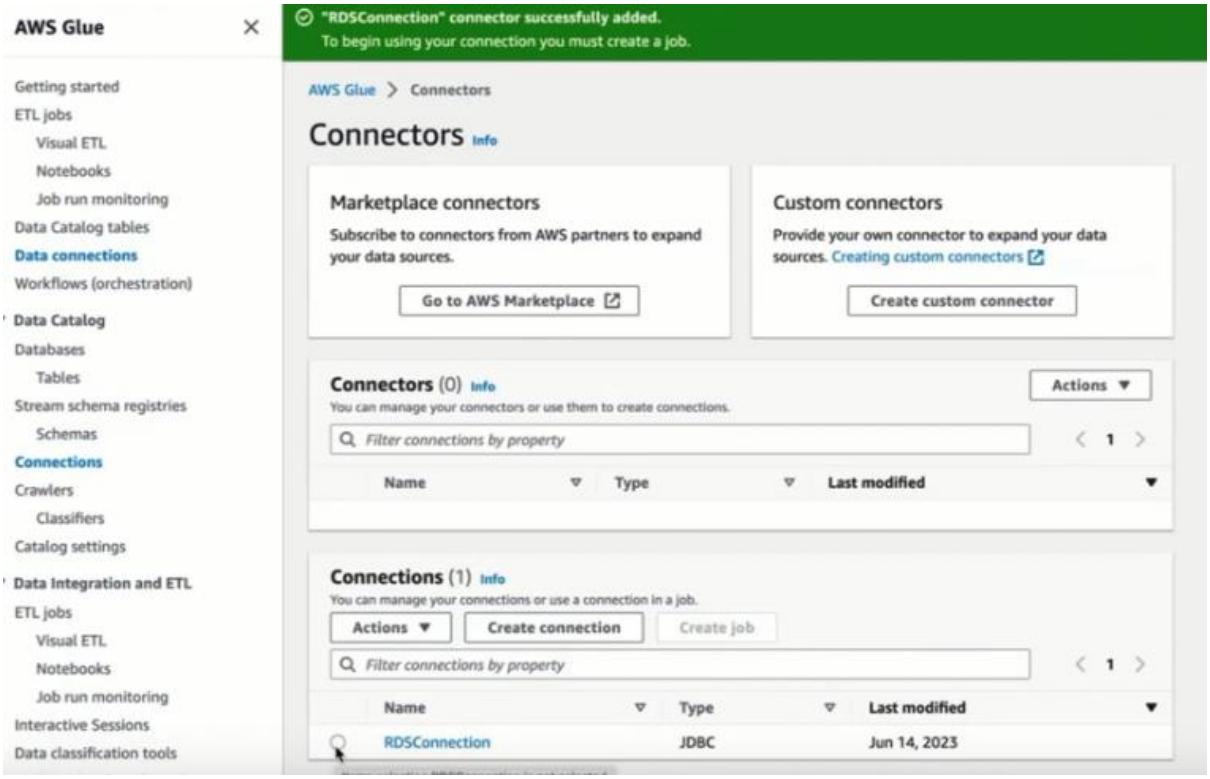

**84. ENTER A NAME FOR THE CONNECTION. WE CALL IT RDS CONNECTION.**
**THEN, CONNECTION TYPE IS AMAZON RDS.**
**IN DATABASE ENGINE, CLOOSE MYSQL.**



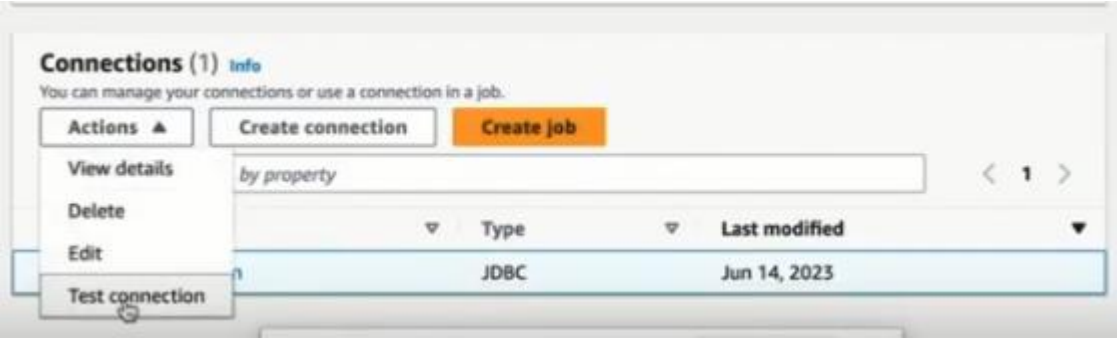
85. **IN CONNECTION ACCESS, DATABASE INSTANCES, REFRESH, THEN, SELECT YOUR RDS DATABASE INSTANCE.**
**IN THE DATABASE NAME, COPY THE DATABASE NAME IN THE MYSQL.**
**CONFIDENTIAL TYPE, USE THE USERNAME THAT WAS SET EARLIER WHICH IS THE ADMIN.**
**INPUT THE PASSWORD. THEN CLICK CREATE CONNECTION.**



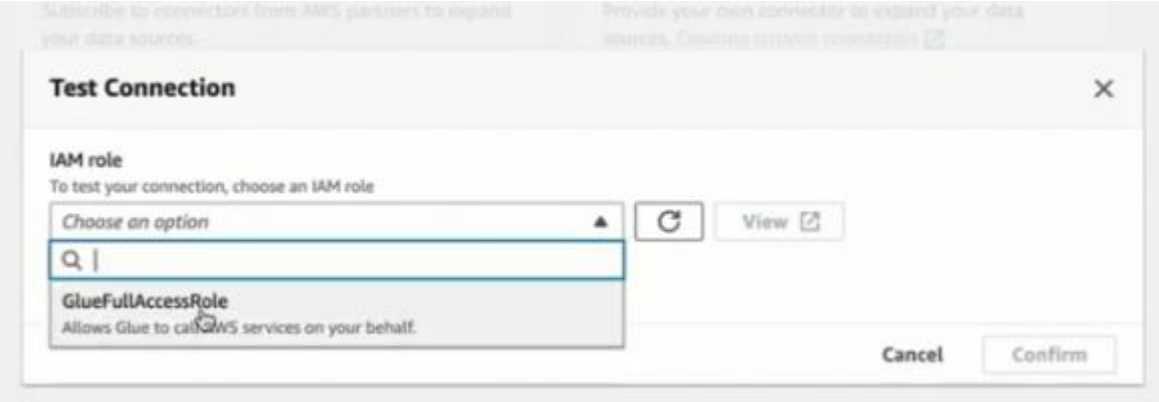

86. **RDS CONNECTION IS NOW CREATED**



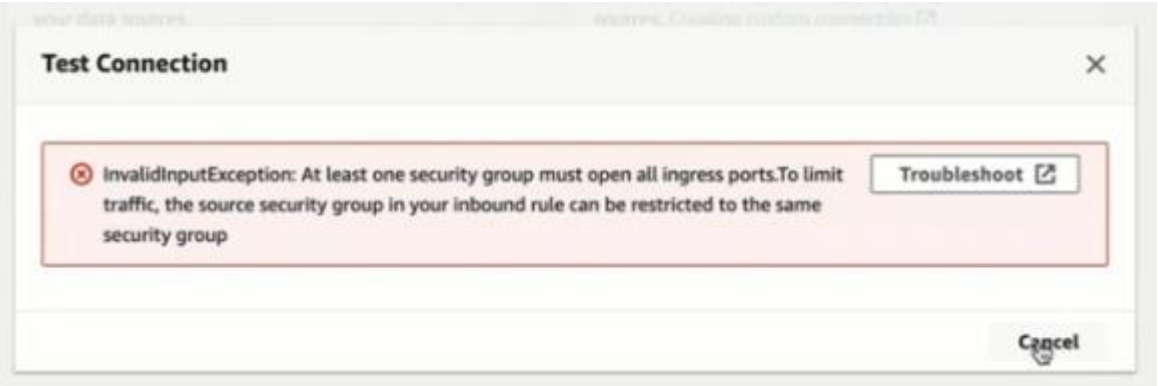
87. **TEST THE CONNECTION. SELECT THE CONNECTION. CLICK ACTIONS. CLICK TEST CONNECTION.**



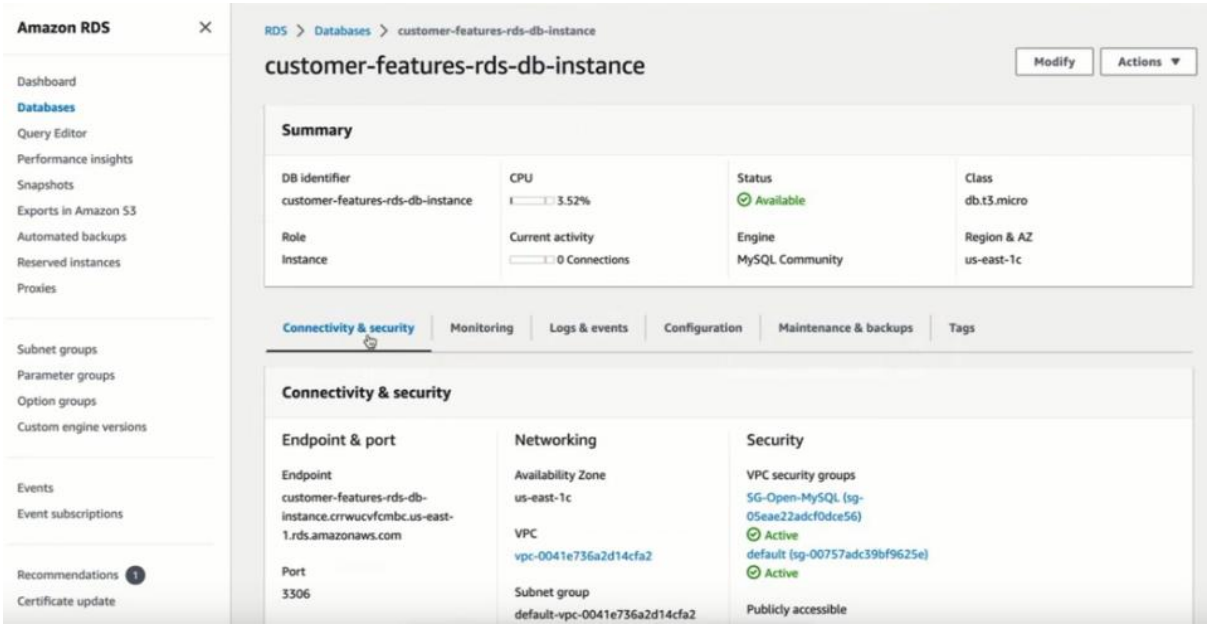

88. **SELECT THE IAM ROLE – GLUEFULLACCESSROLE. SELECT CONFIRM.**



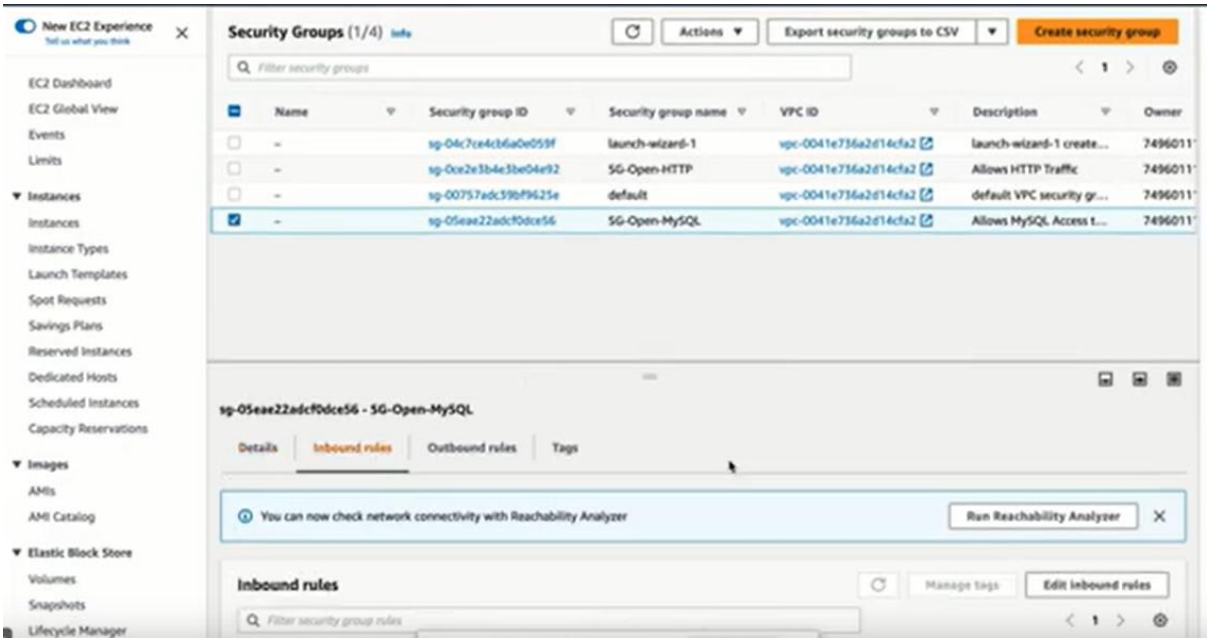

89. **HERE, WE WILL GET THE FAILED TEST CONNECTION. TO SOLVE THE ISSUE, WE HAVE TO FOLLOW WHAT IS SUGGESTED. BASED ON THE TROUBLESHOOT RECOMMENDATION, AT LEAST ONE SECURITY GROUP MUST OPEN ALL INGRESS PORTS AND TO LIMIT TRAFFIC, THE SOURCE SECURITY GROUP IN YOUR INBOUND RULE CAN BE RESTRICTED TO THE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



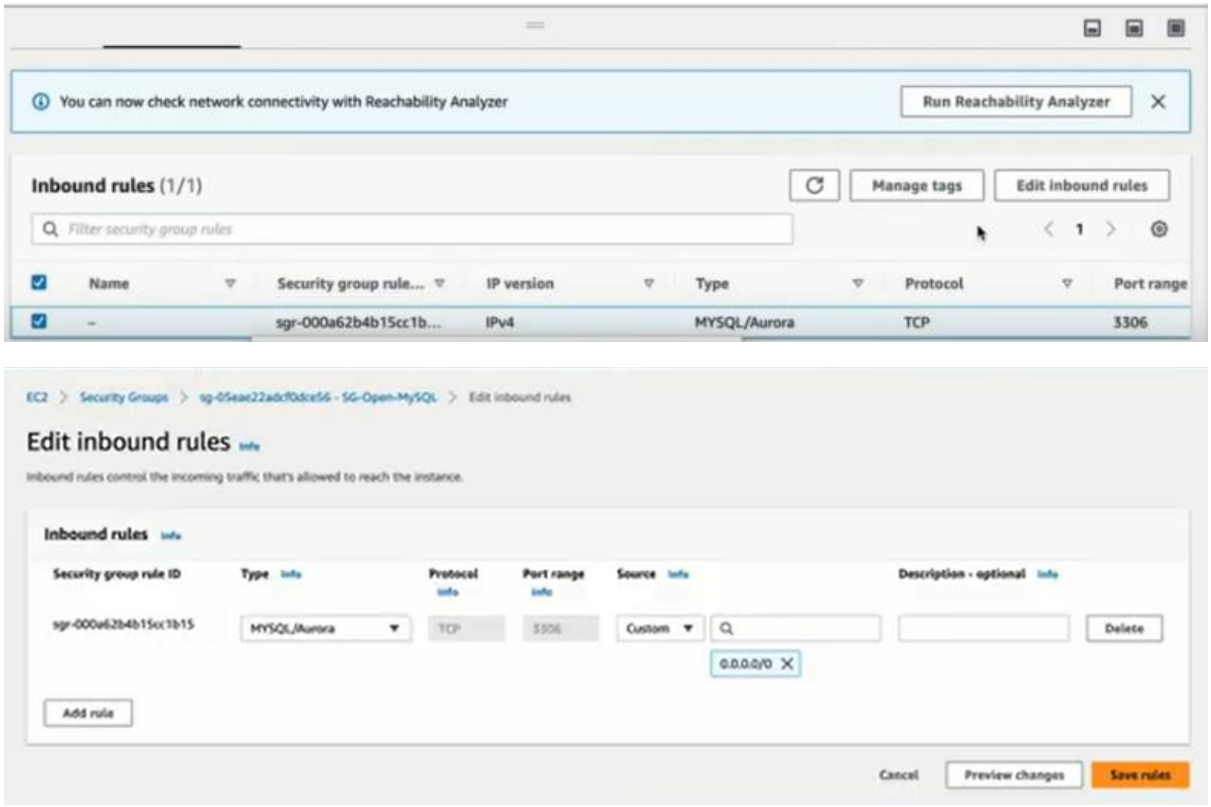
90. **GO BACK TO THE RDS DATABASE INSTANCE. IN THE SECURITY FIELD, YOU WILL SEE THOSE TWO (2) VPC SECURITY GROUPS, THE ONE THAT WE`VE CREATED, SG-OPEN-MYSQL, AND THE DEFAULT ONE. BASED ON THE PREVIOUS TROUBLESHOOT SUGGESTIONS, WE NEED TO OPEN AT LEAST ONE SECURITY GROUP OF ALL INGRESS PORTS. SO, INSTEAD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

91. 

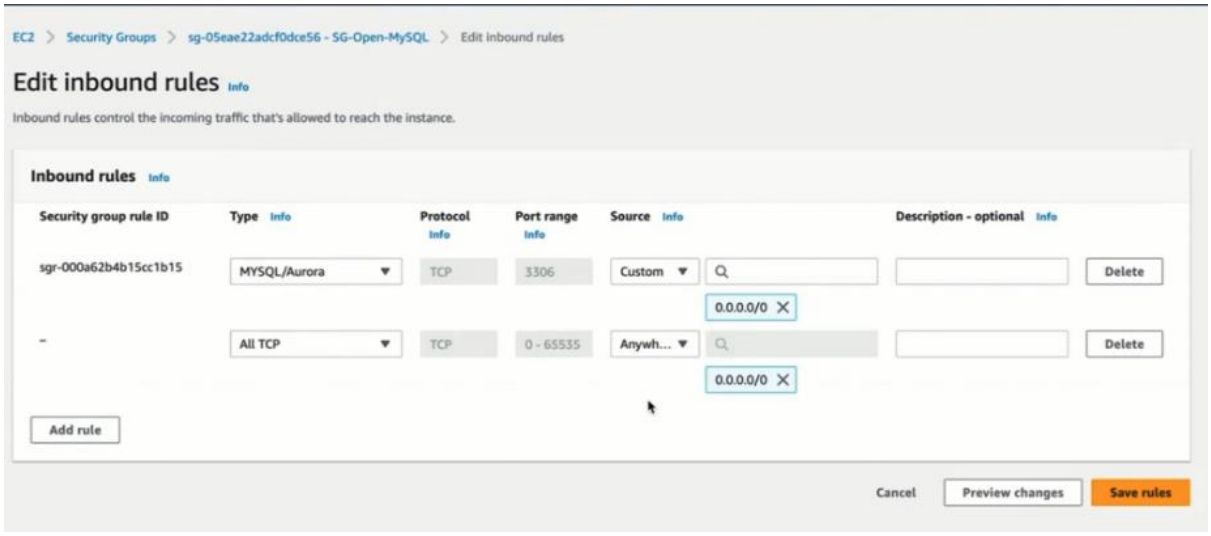

92. **GO BACK TO THE EC2 INSTANCE, SECURITY GROUPS. SELECT SG-OPEN-MYSQL. GO TO INBOUND RULES. THERE, WE ARE GOING TO ADD THE NEW RULE WHICH OPENS ALL PORTS.**



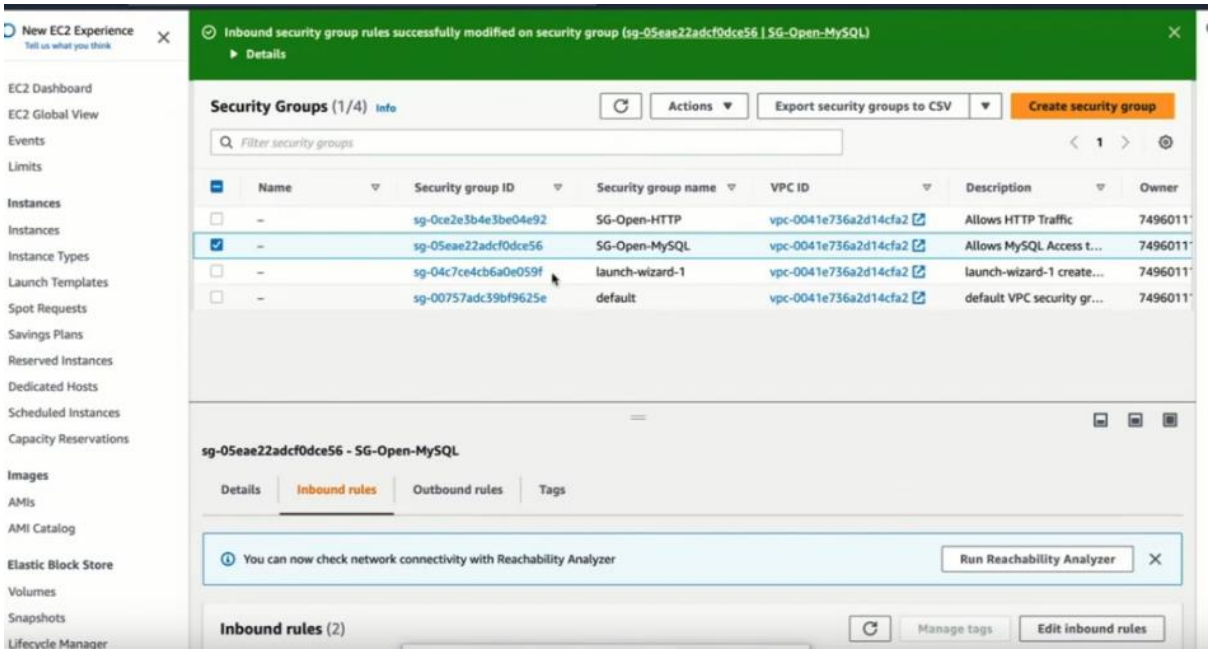
93. **IN THE INBOUND RULES, CLICK EDIT INBOUND RULES. THEN ADD RULE.**



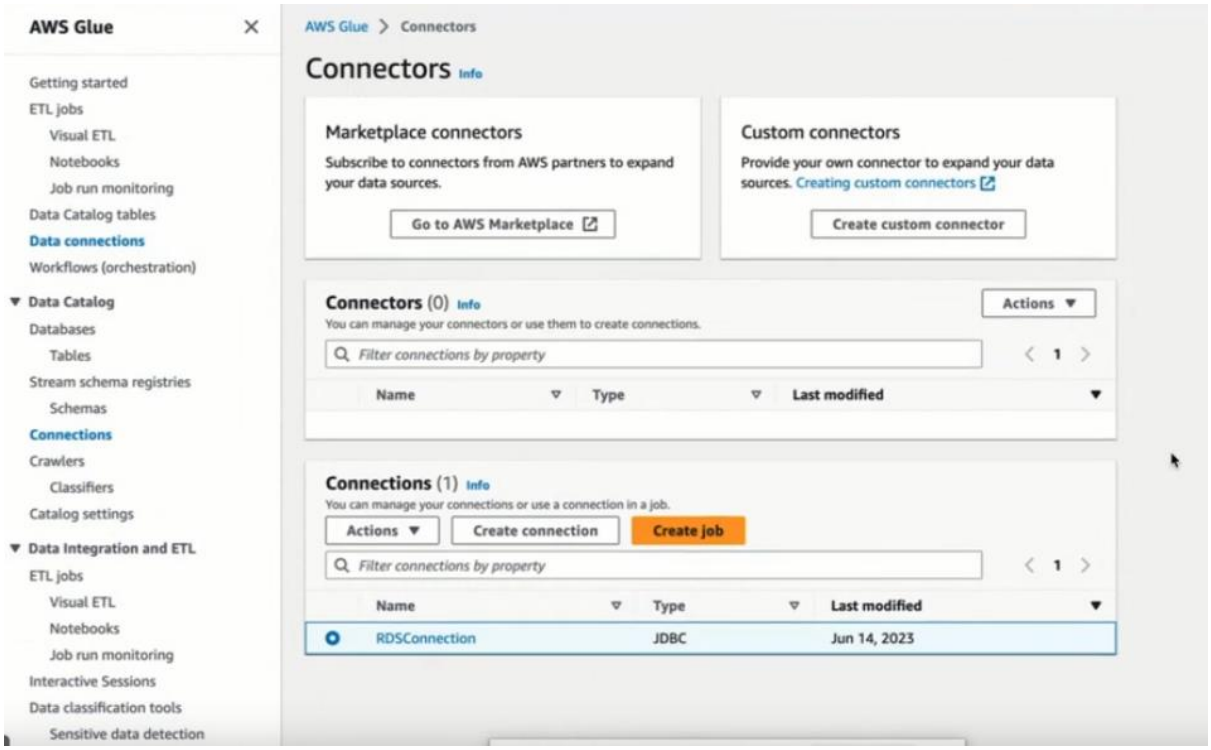

94. **IN THE ADD RULE, CLICK ALL TCP IN THE TYPE TO OPEN ALL PORTS FROM 0 – 65535. IN THE SOURCE, SELECT ANYWEHERE-IPV4. CLICK SAVE RULES.**



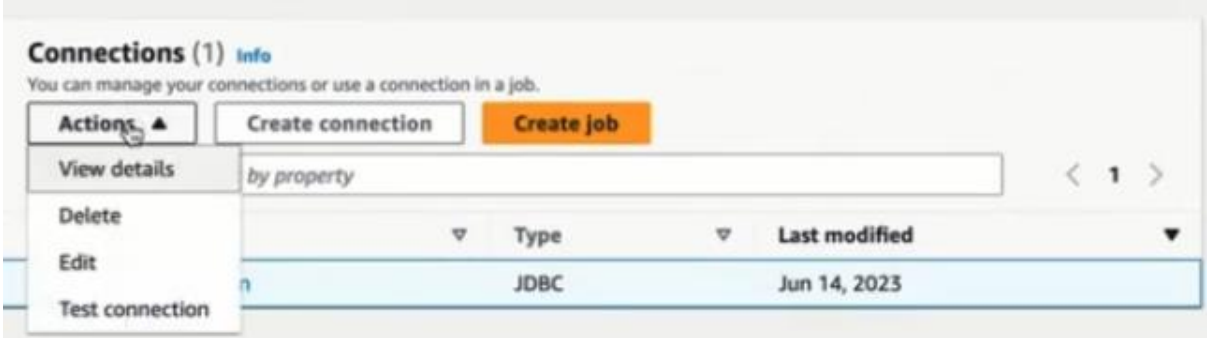

95. **WE ARE NOW GOING TO SEE IF IT FIXES THE ISSUE. HERE, THE SECURITY GROUP IS MODIFIED.**



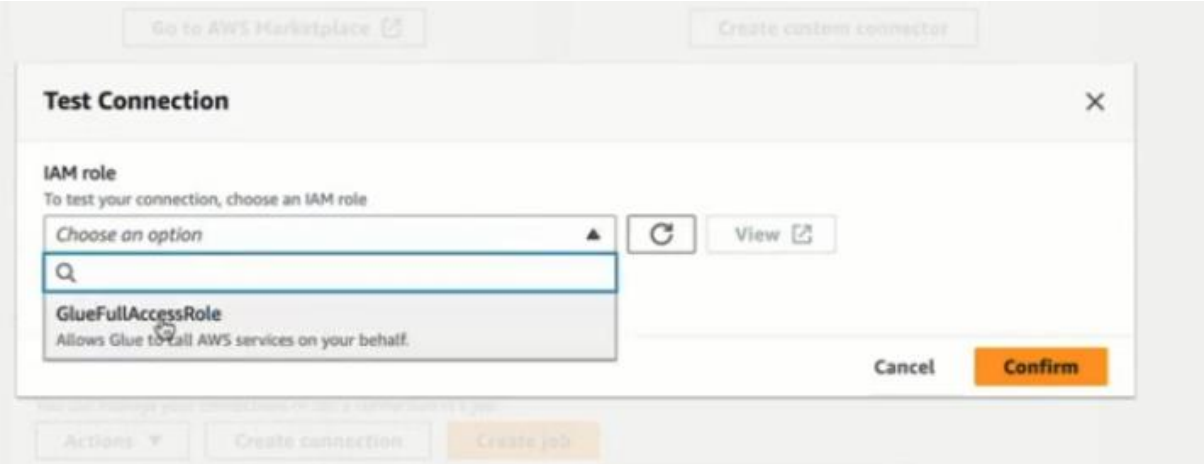

**96. GO BACK TO CONNECTORS TO TEST THE CONNECTION AGAIN. IN THE CONNECTIONS, CLICK RDSCONNECTION, CLICK ACTIONS, CHOOSE TEST CONNECTION.**



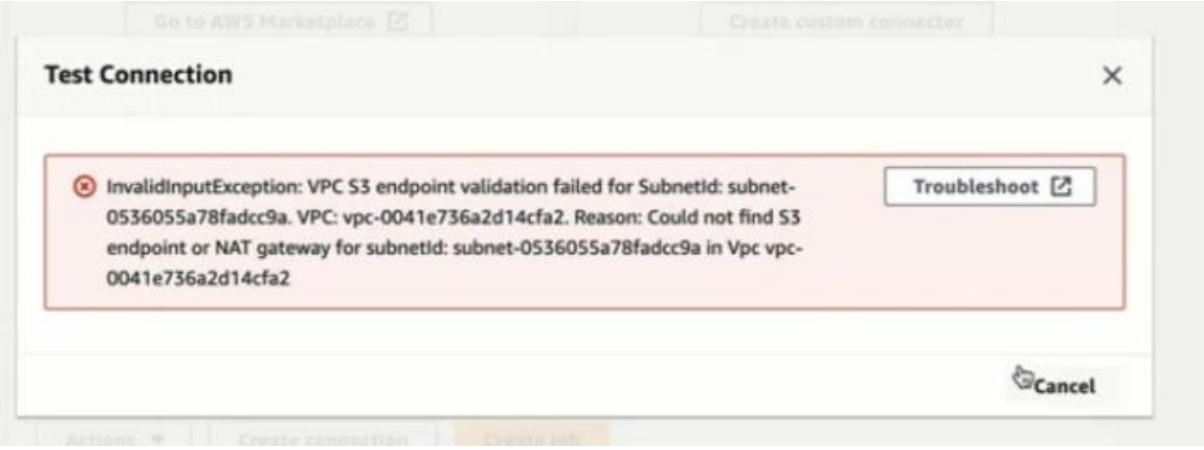
**SELECT TEST CONNECTION.**



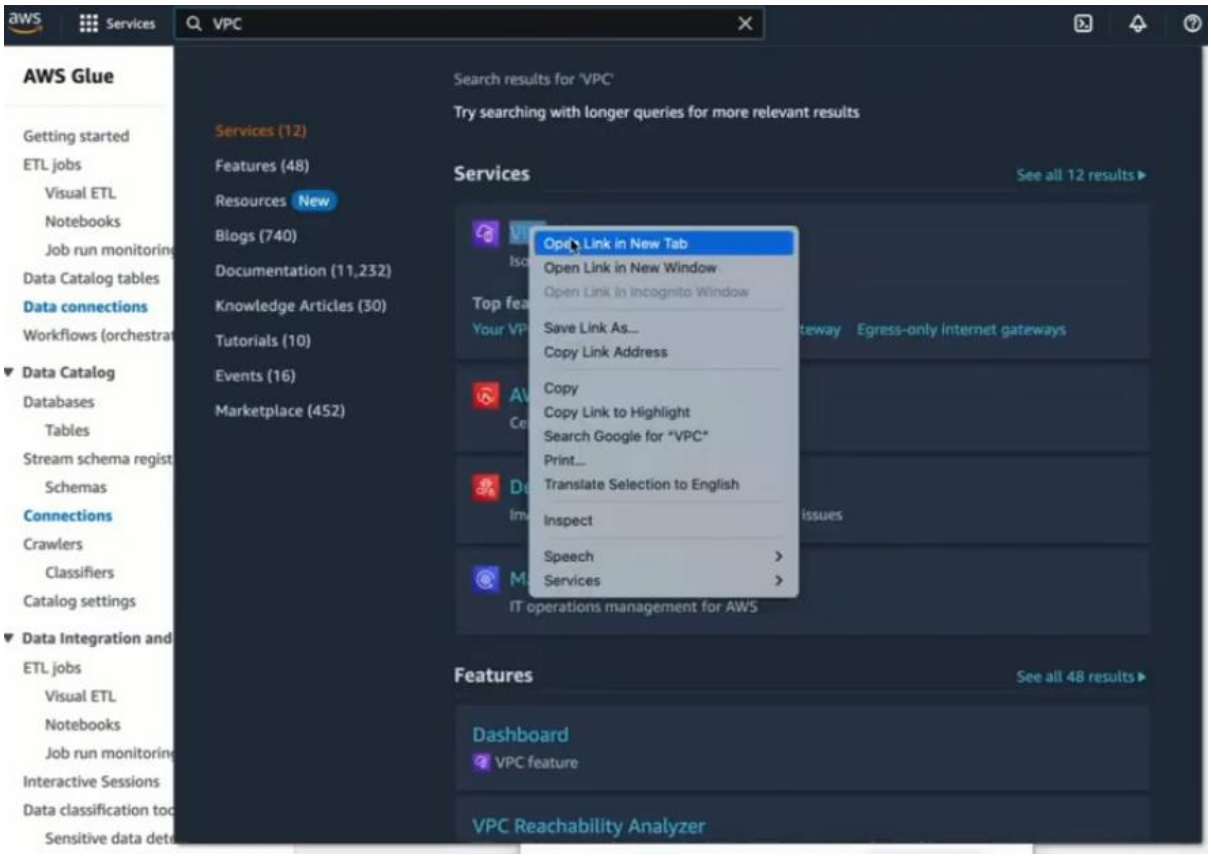

**CHOOSE GLUEFULLACCESSROLE. THEN, CONFIRM.**



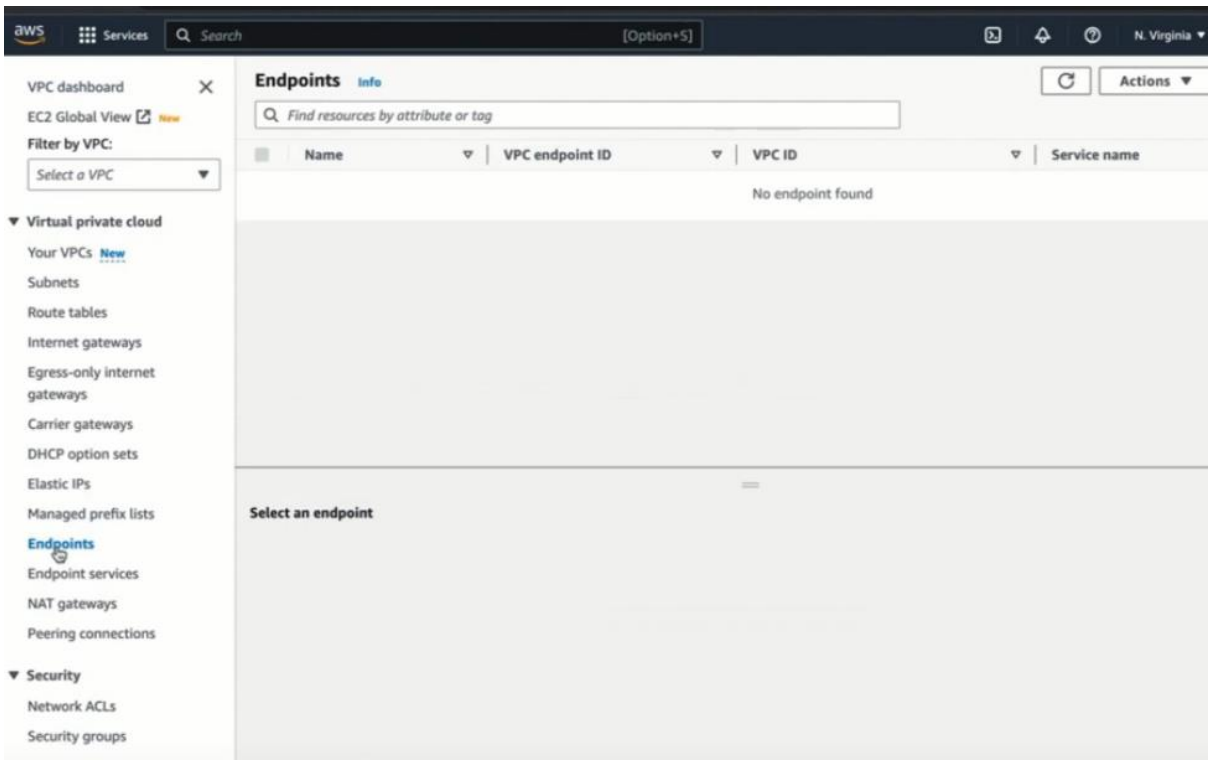

**UNFORTUNATELY, WE HAVE ANOTHER MESSAGE ERROR. BASED ON THE TROUBLESHOOT MESSAGE, WE NEED TO CREATE A VPC S3 ENDPOINT.**



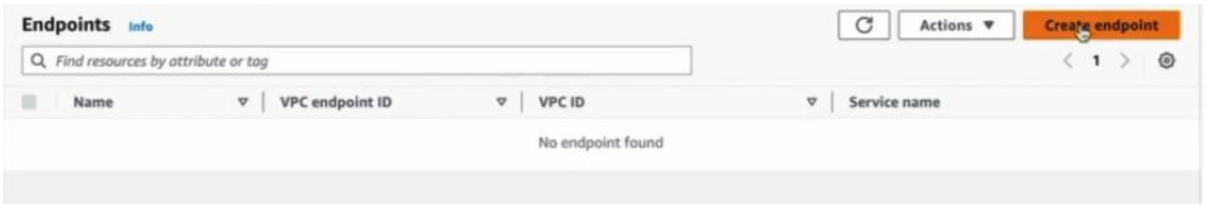
**97. GO BACK TO THE SEARCH BAR, TYPE VPC. THEN, RIGHT CLICK, OPEN LINK IN NEW TAB.**



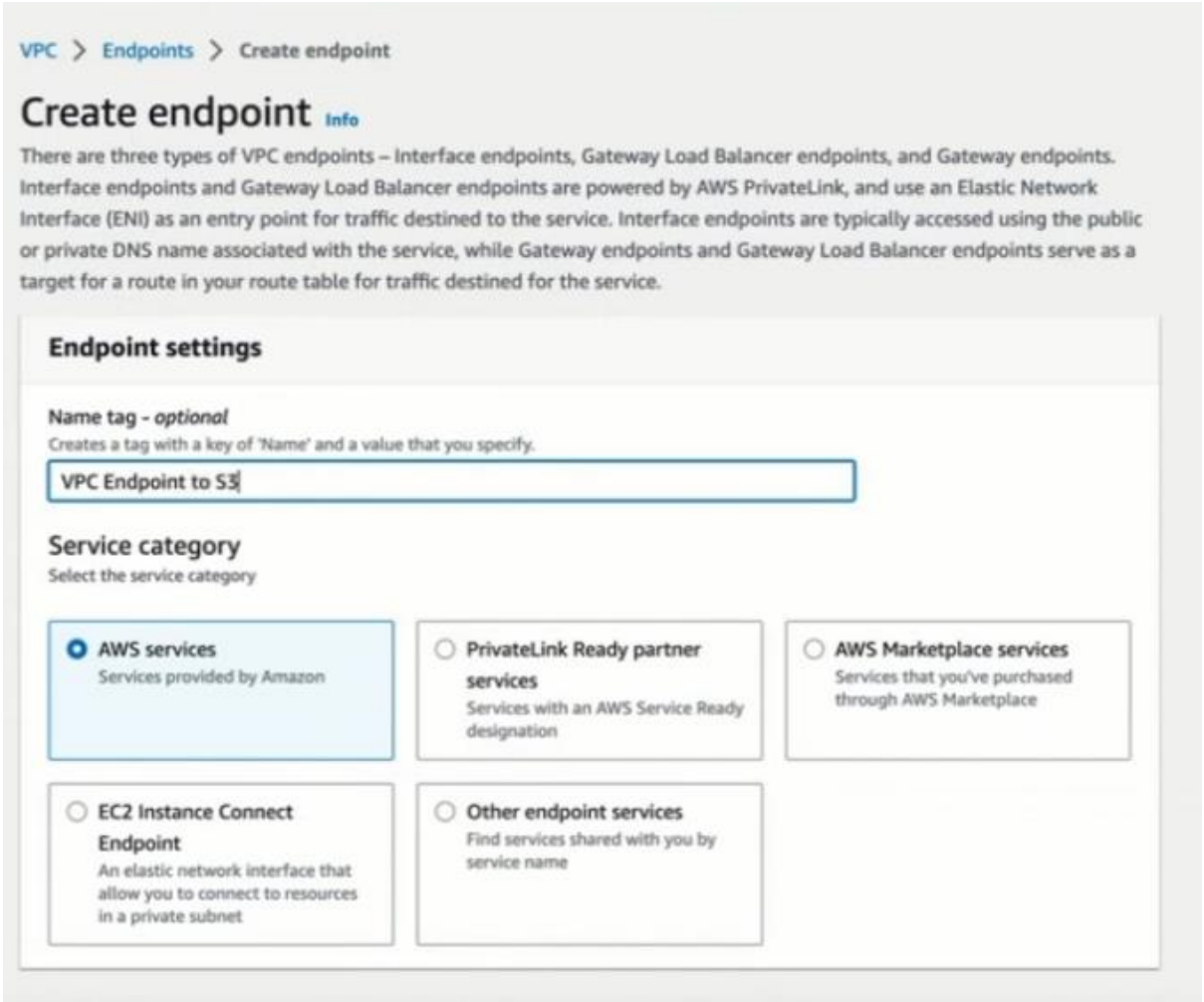

**98. IN THE VCPC, GO TO THE ENDPOINTS. IT IS NORMAL NOT TO SEE ANY ENDPOINTS. BUT AS A SOLUTION TO THE CONNECTION, WE ARE GOING TO CREATE AN ENDPOINT.**



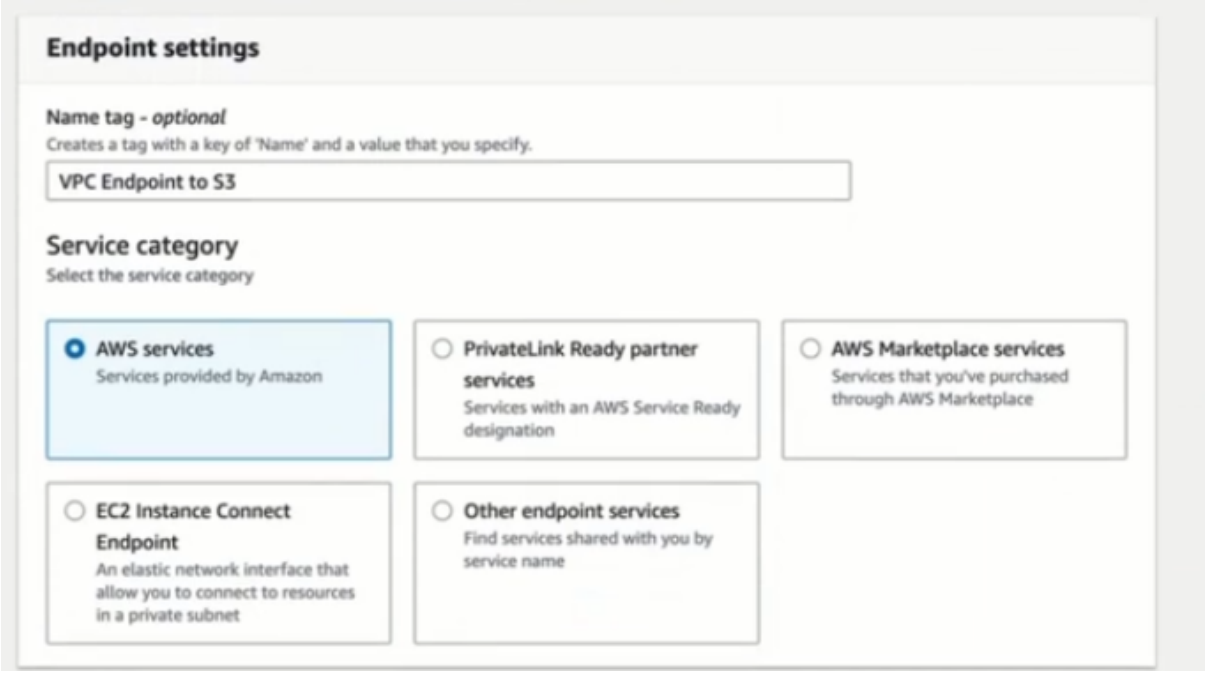
**99. CLICK CREATE ENDPOINTS**



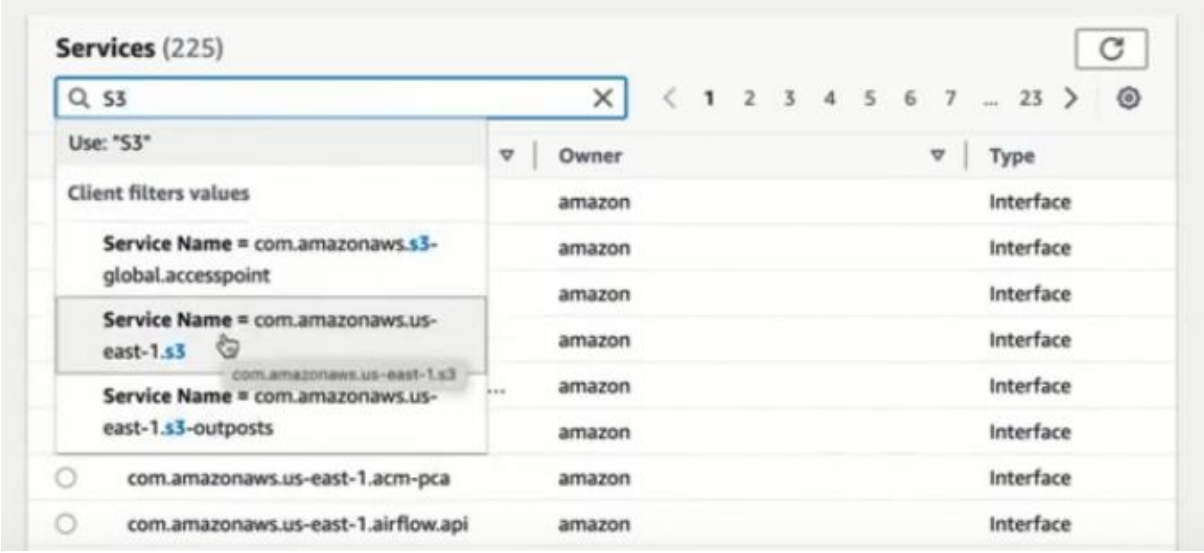

**100. THE ENDPOINT SHOULD BE THE VPC ENDPOINT TO THE S3. SO, WE ARE GOING TO CALL IT VPC ENDPOINT TO S3.**



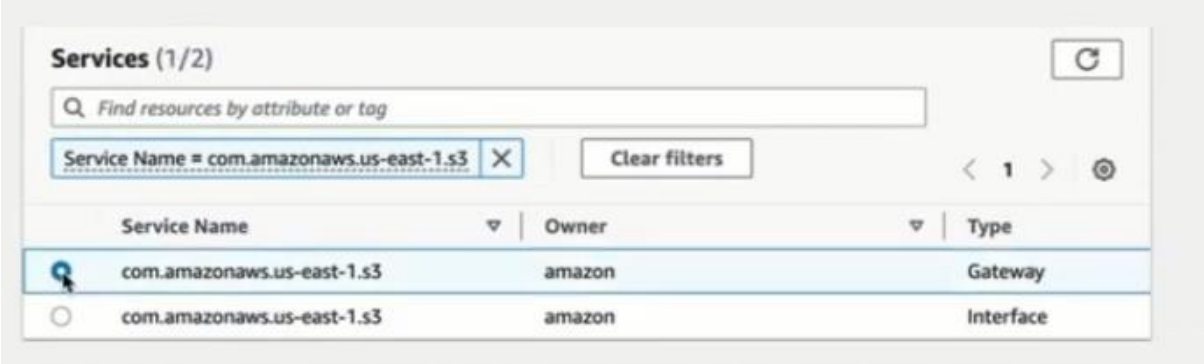

**101. IN THE SERVCE CATERGORY, SELECT AWS SERVICES.**



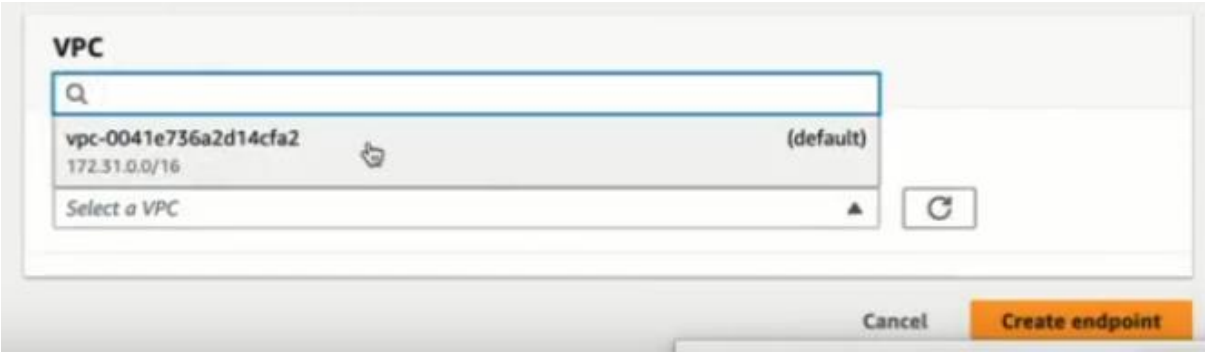

**102. IN THE SERVICES, SELECT SERVICE NAME = COM.AMAZONAWS.US-EAST-1.S3**



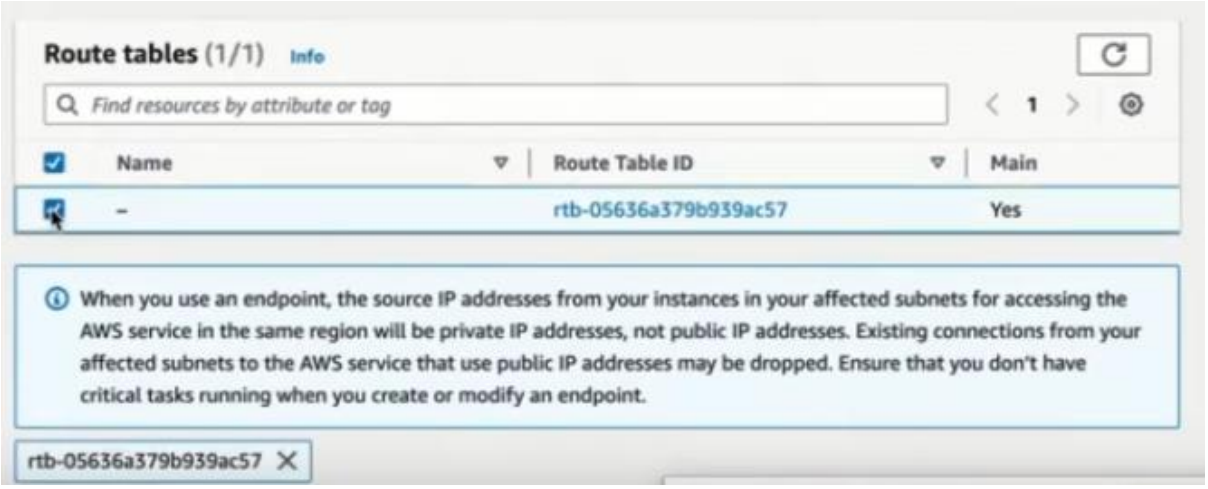

**103. SELECT GATEWAY**



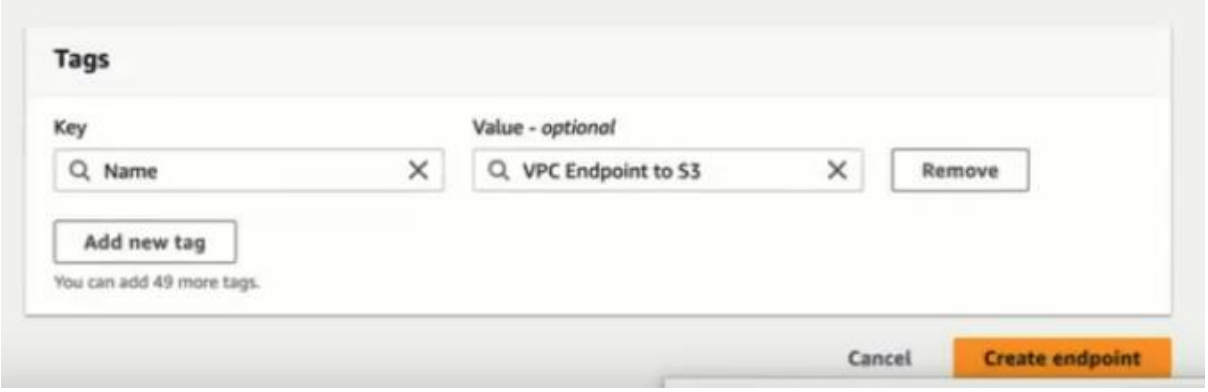

**104. SELECT THE VPC. REMEMBER TO CONNECT IT TO THE SUBNET.**



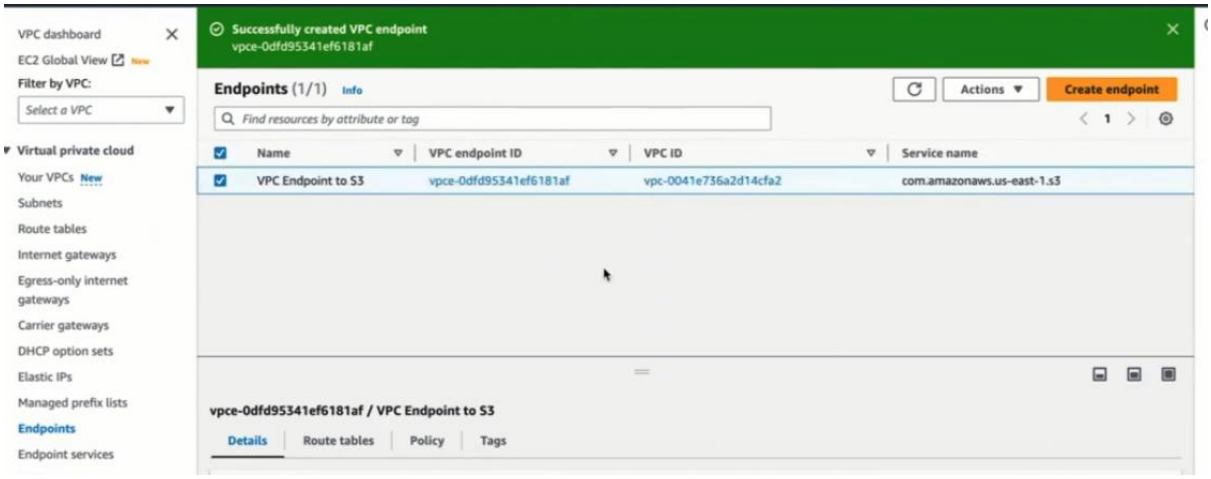

**105. WE WILL CONNECT THE SUBNET THROUGH THE ROUTE TABLES BECAUSE THAT IS WHAT CONTAINS THE CONNECTION TO THE SUBNETS.**




**106. CLICK CREATE ENDPOINT.**

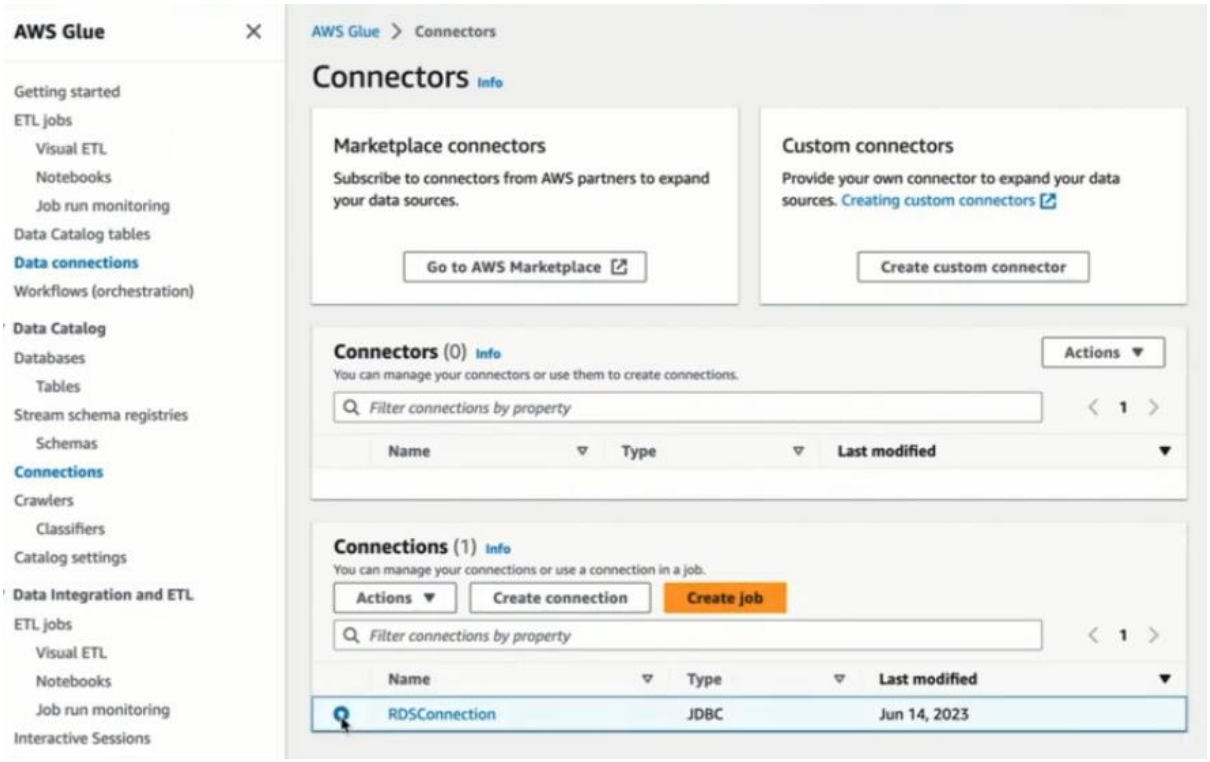



**107. GO CHECK THE ENDPOINTS. HERE, WE SHOULD SEE THE ONE THAT WE HAVE CREATED.**

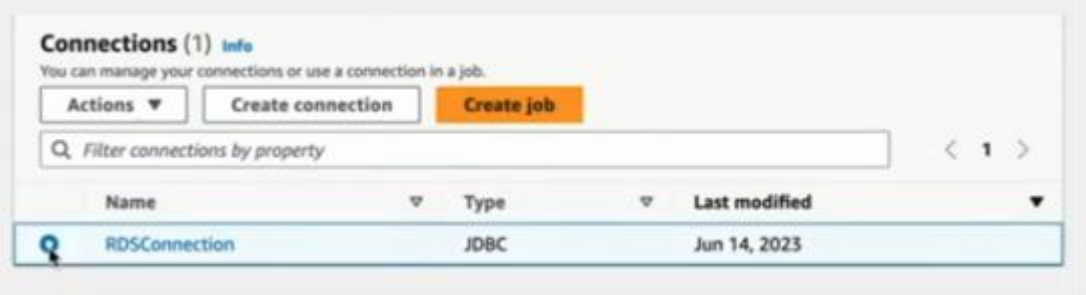


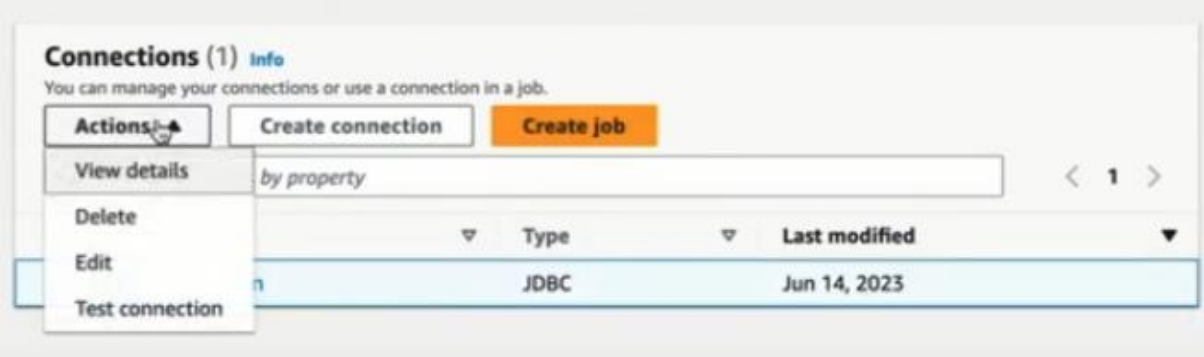

**108. CHECK IF IT FIXES THE ISSUE. GO BACK TO THE CONNECTORS**

.

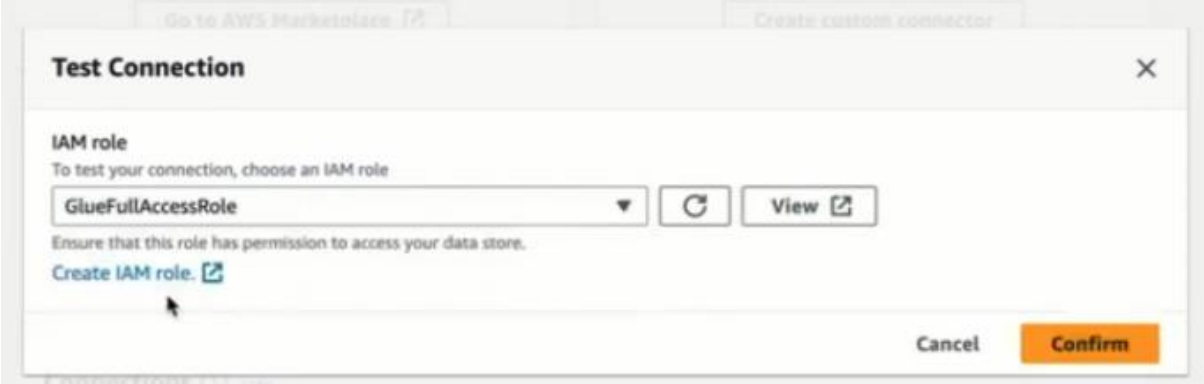



**109. CLICK RDSCONNECTION, ACTION THE TEST CONNECTION.**

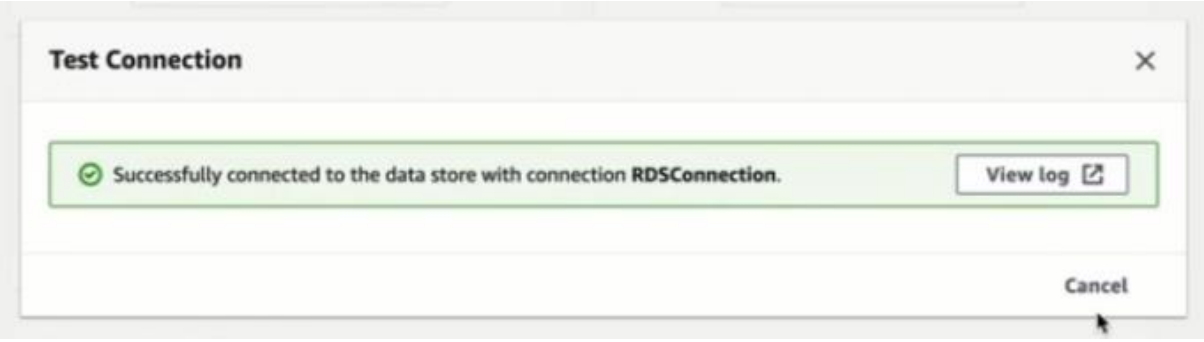


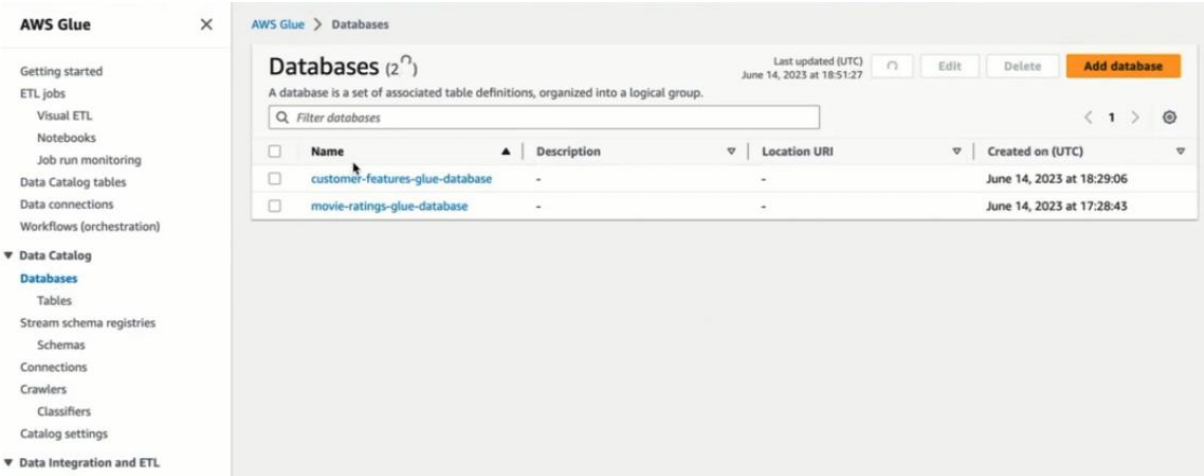

**110. SELECT GLUEFULLACCESS, THEN, CONFIRM.**



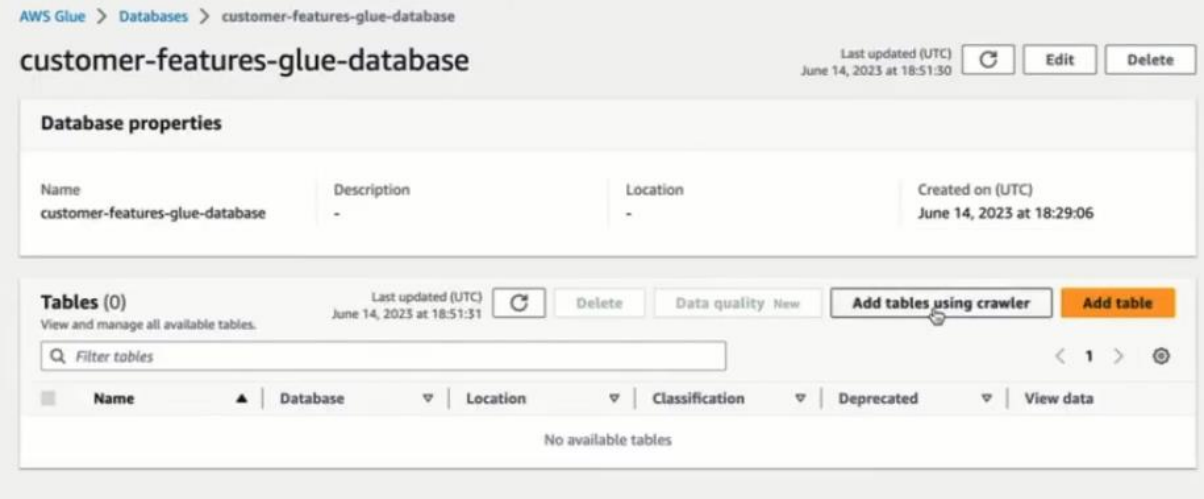

**111. IT SHOULD BE SUCCESSFULLY CONNECTED BY NOW. CONGRATULATIONS YOU DID THE HARD PART! WE CAN NOW PROCEED TO THE NEXT PROCEDURE WHICH IS TO HANDLE THE REDSHIFT.**



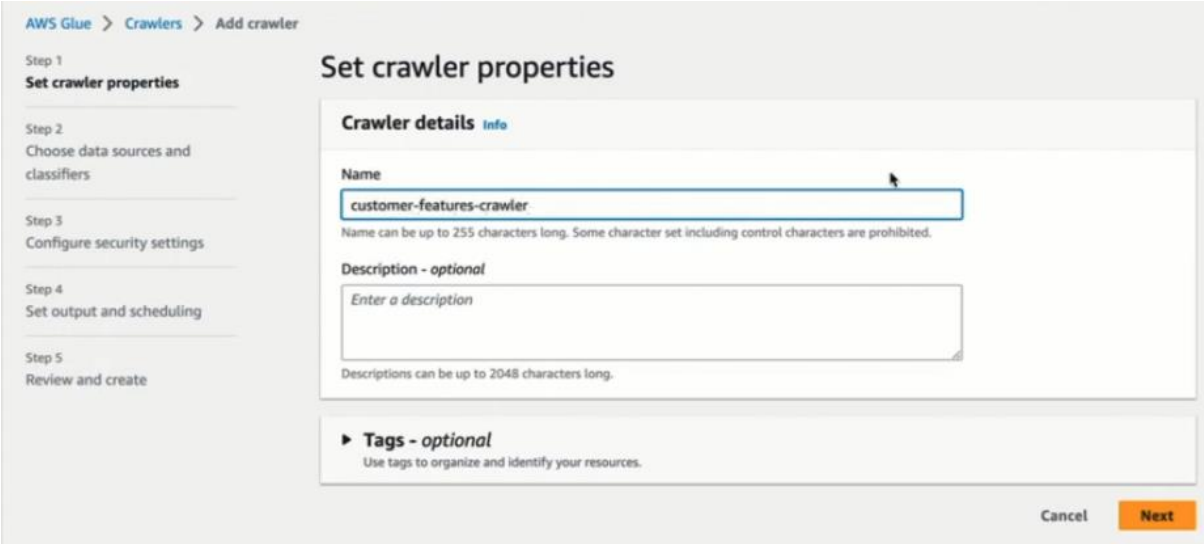

**112. IN THE AWS GLUE, SELECT DATA CATALOG, THEN, DATABASES. IN THE DATABASES, SELECT CUSTOMER-FEATURES-GLUE-DATABASE.**



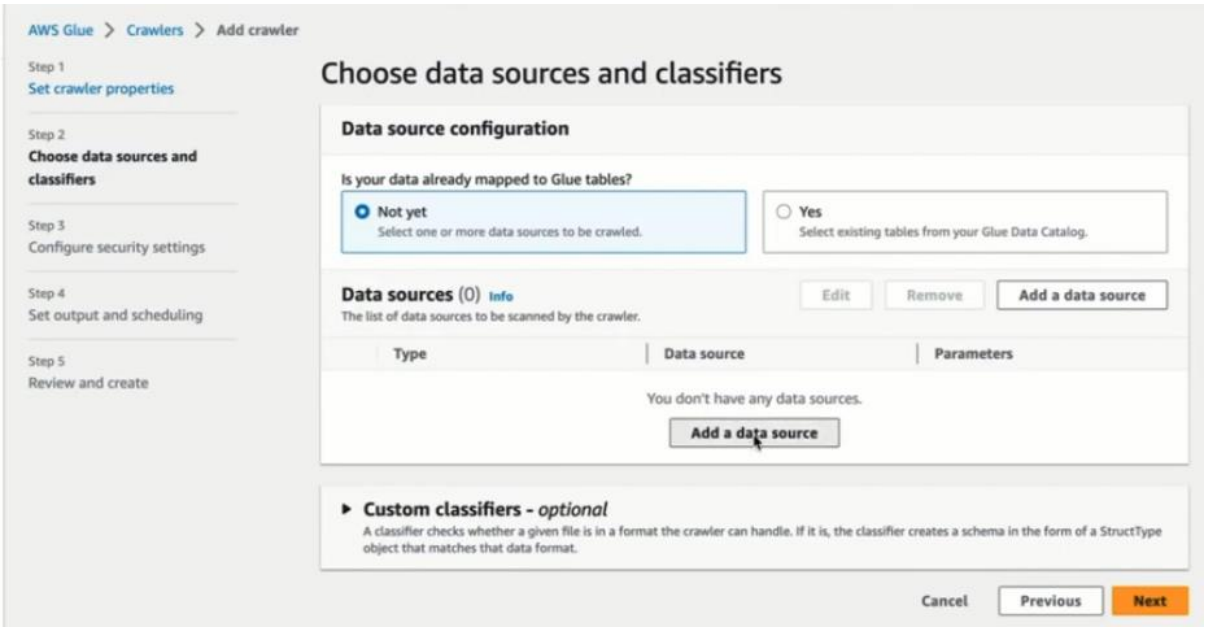

**113. IN THE CUSTOMER-FEATURES-GLUE-DATABASE, CLICK, ADD TABLES USING CRAWLER.**



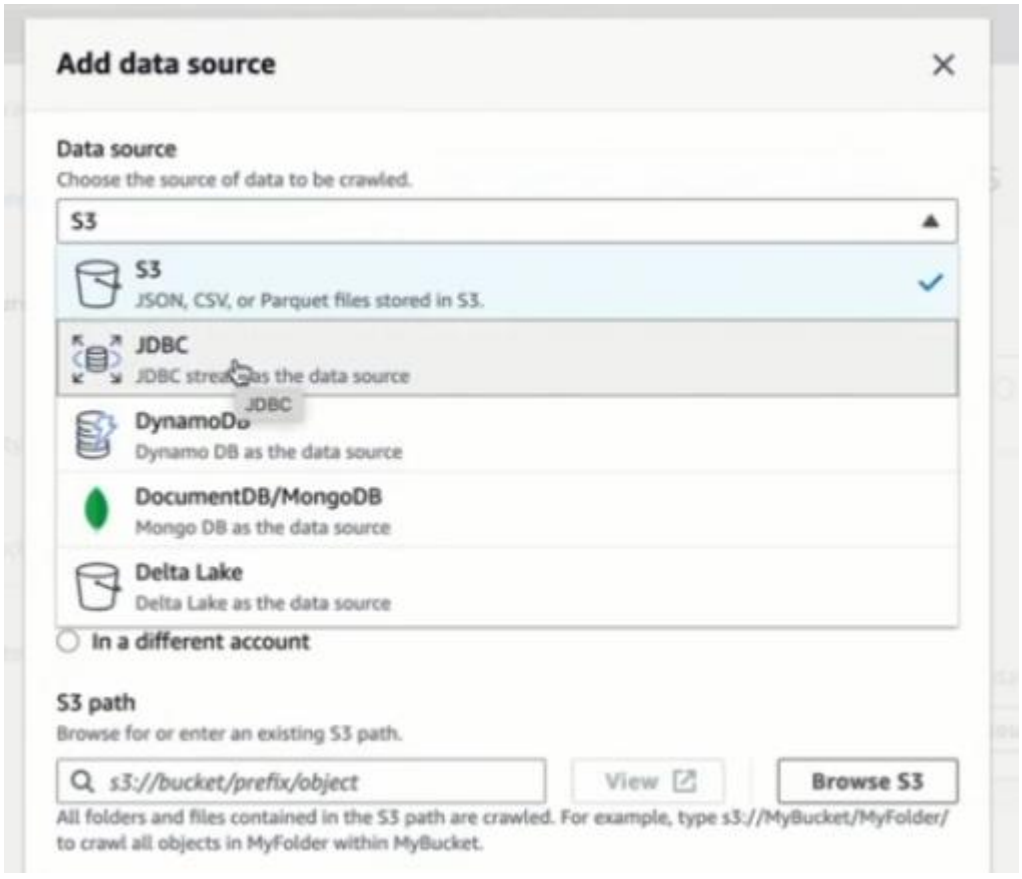

**114. ENTER UNIQUE CRAWLER NAME. CLICK, NEXT.**




**115. CLICK NOT YET IN THE "IS YOUR DATA ALREADTY MAPPED TO GLUE TABLES?". THEN, ADD A DATA SOURCE.**

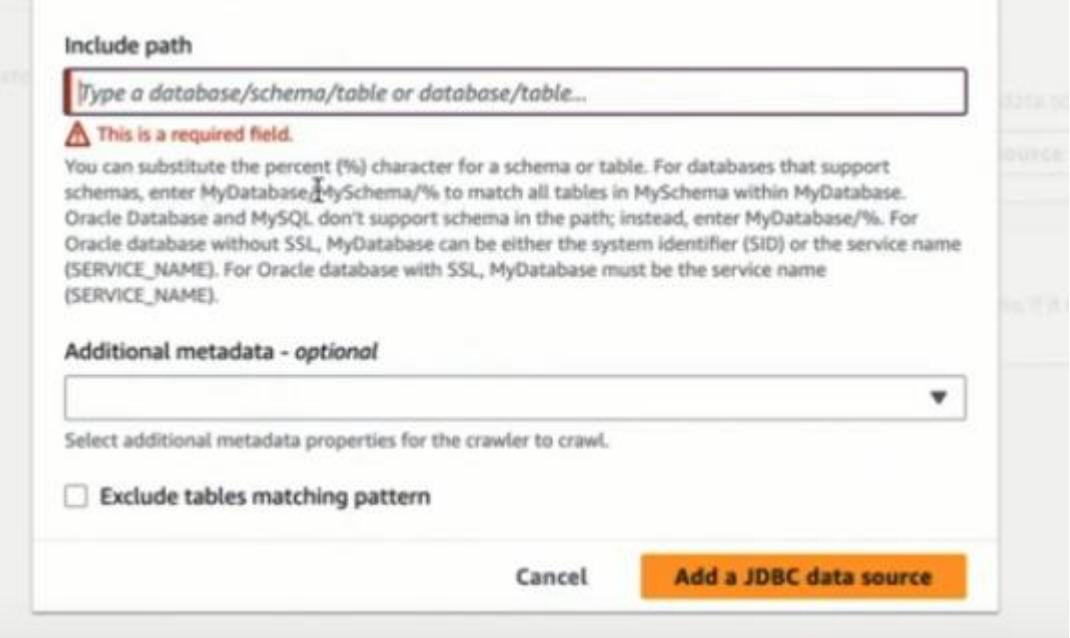



**116. IN THE DATA SOURCE, SELECT JDBC WHICH MEANS JAVA DATABASE CONNECTION.**



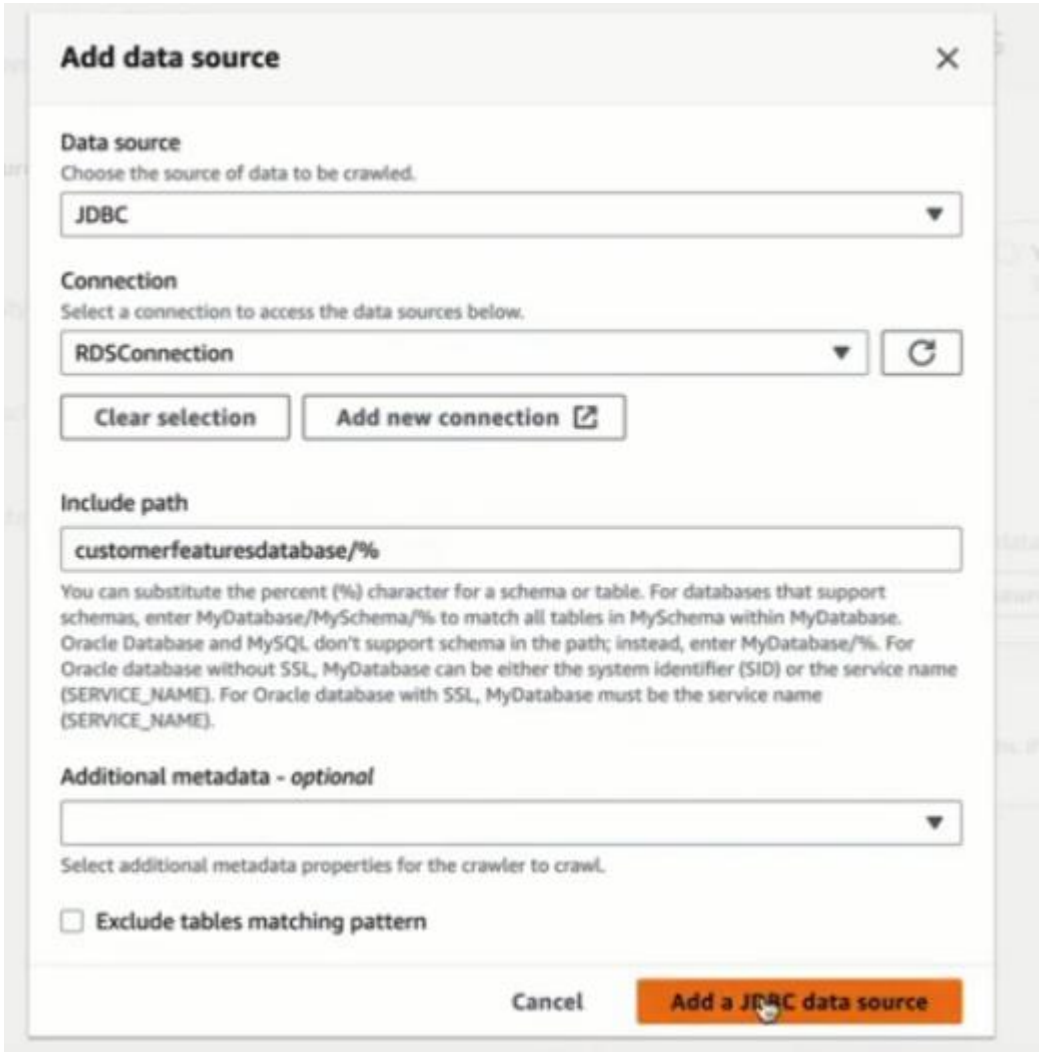

**117. IN THE CONNECTION, SELECT RDSCONNECTION THAT WE HAVE JUST CREATED AND TESTED SUCCESSFULLY.**


**118. 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.**

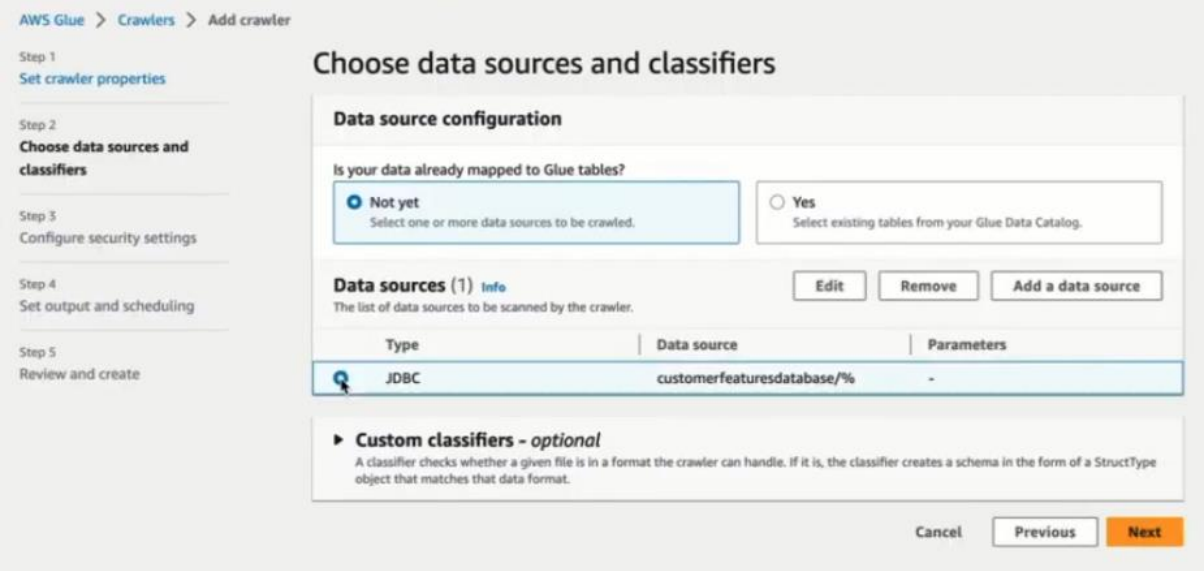



**119. GO BACK TO THE MYSQL WORKBENCH TO DOUBLE CHECK THE DATABASE NAME THAT WE HAVE JUST CREATED PREVIOUSLY.**

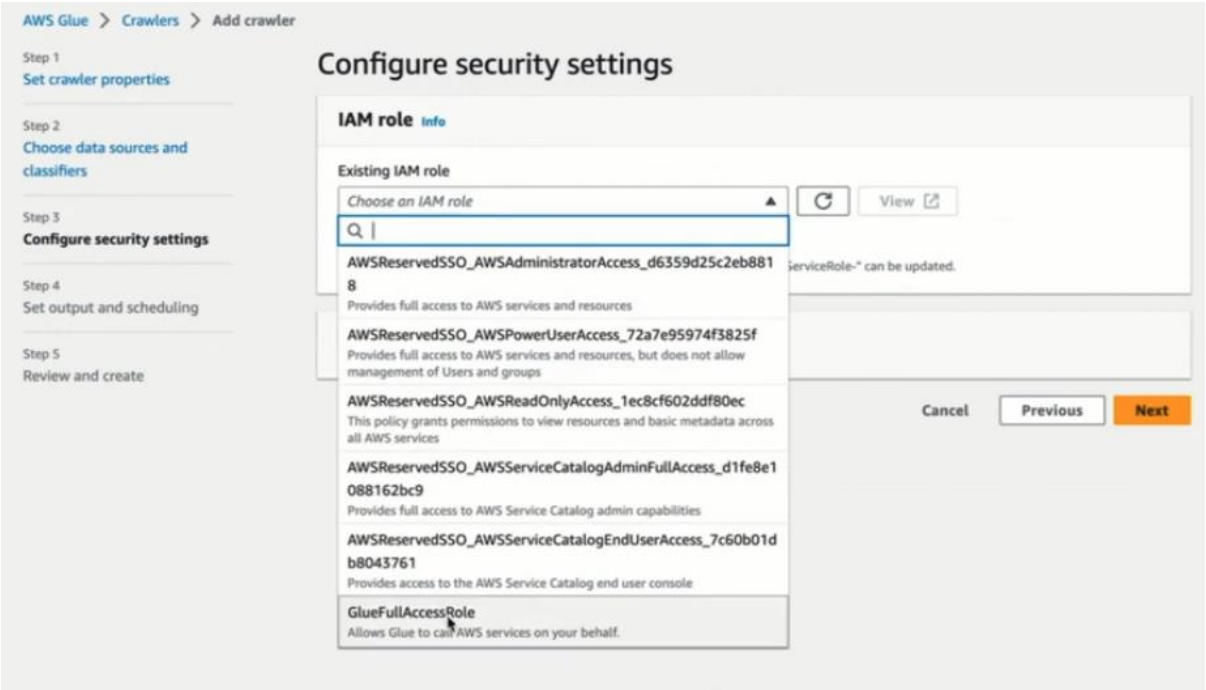

**120. FOLLOW THE DATABASE NAME IN THE MYSQL AND THE INSTRUCTION GIVEN IN PUTTING THE NAME IN THE 'INCLUDE PATH'. FINALLY, CLICK ADD A JDBC DATA SOURCE.**

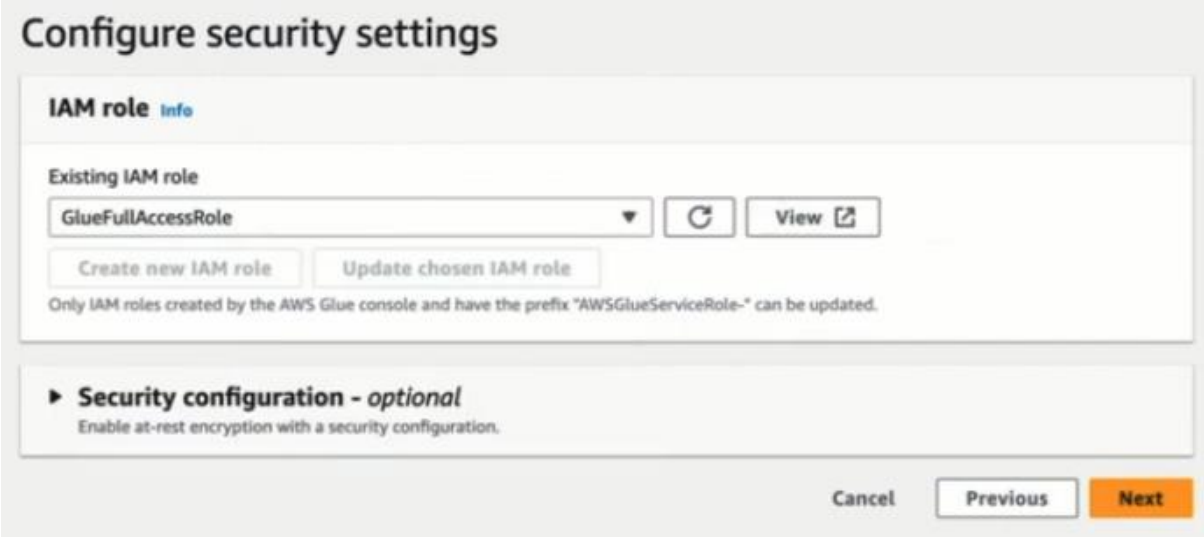



**121. SELECT JDBC. THEN, CLICK NEXT.**

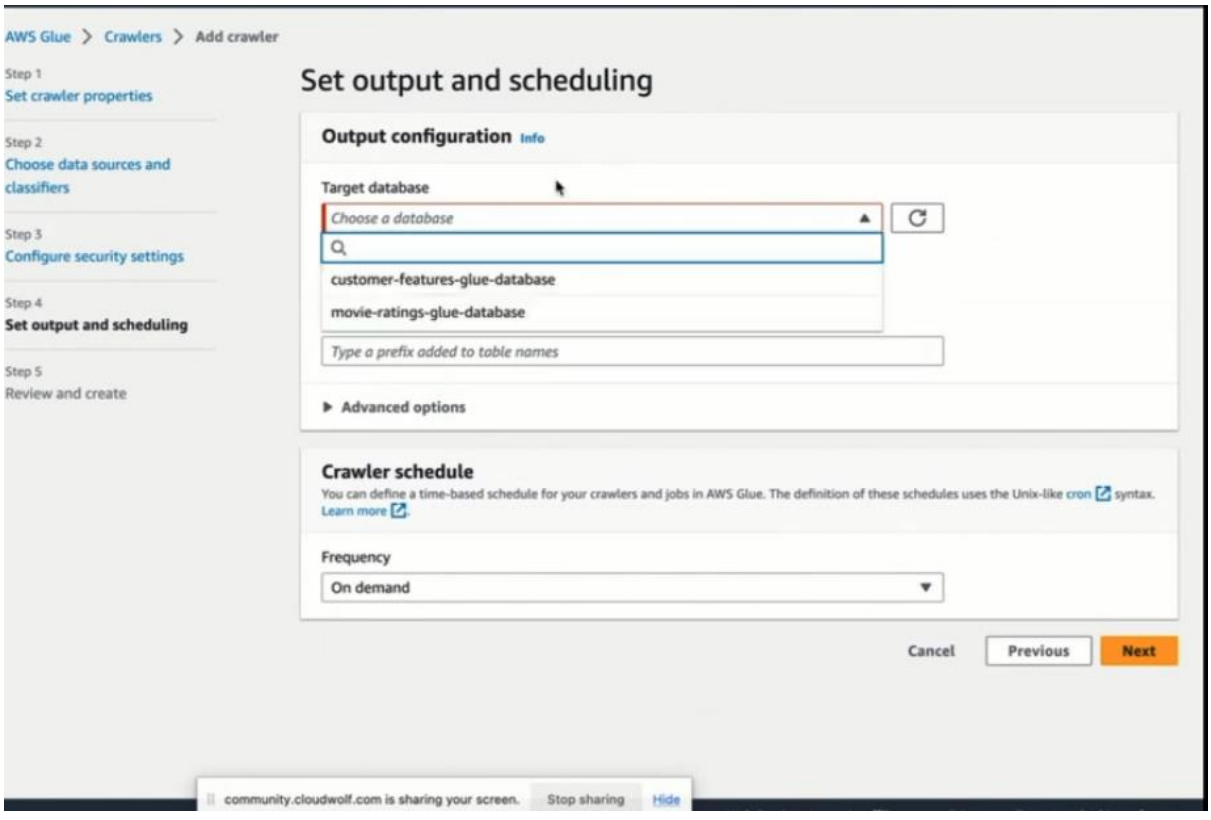



**122. IN THE CONFIGURE SECURITY SETTINGS, EXISTING IAM ROLE, SELECT GLUEFULLACCESSROLE. THEN, CLICK NEXT.**

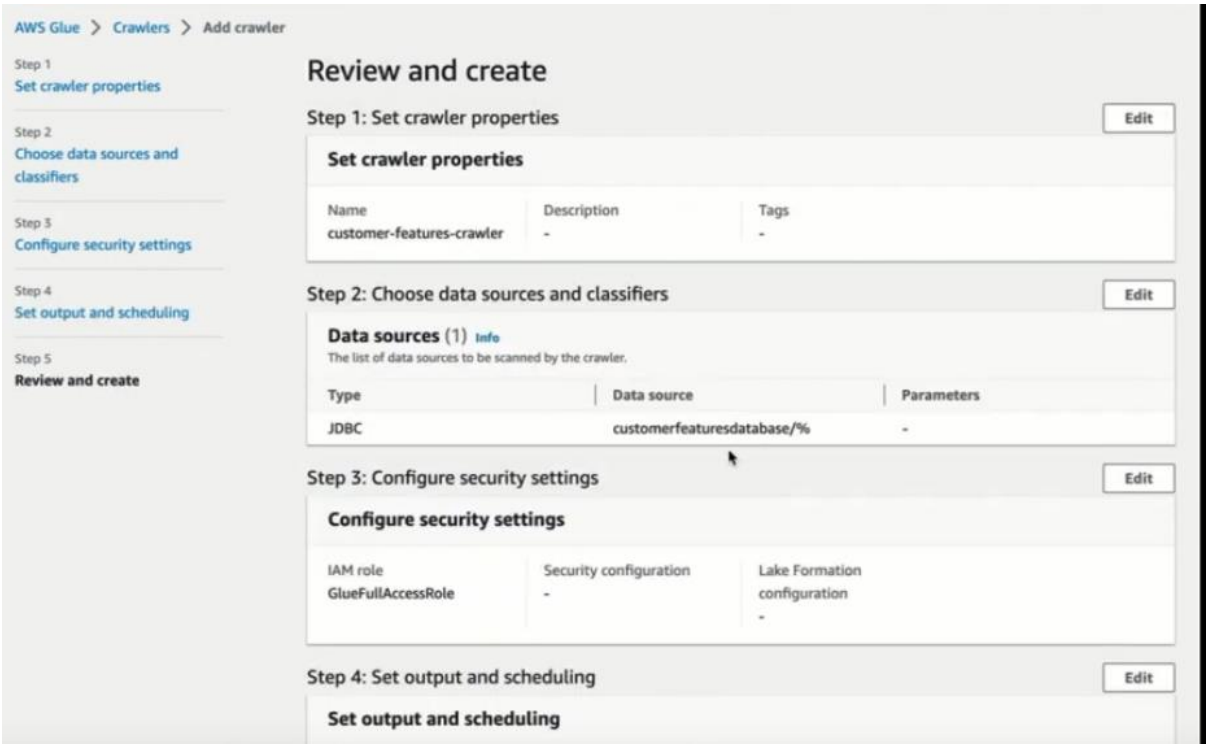


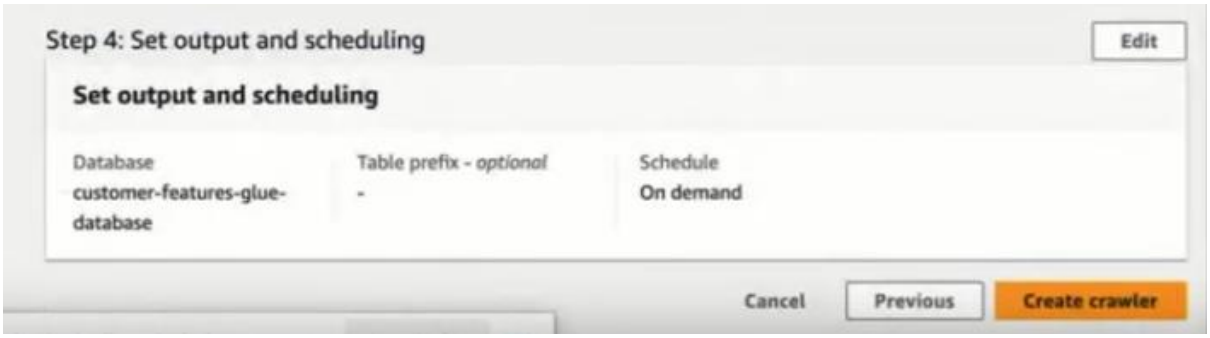



**123. IN THE TARGET DATABASE, SELECT CUSTOMER-FEATURES-GLUE-DATABASE. KEEP THE ON-DEMAND FREQUENCY OF THE CRAWLER SCHEDULE. IT MEANS THAT IT IS GOING TO CRAWL WHEN WE CLICK THE RUN BUTTON. THEN, CLICK NEXT.**

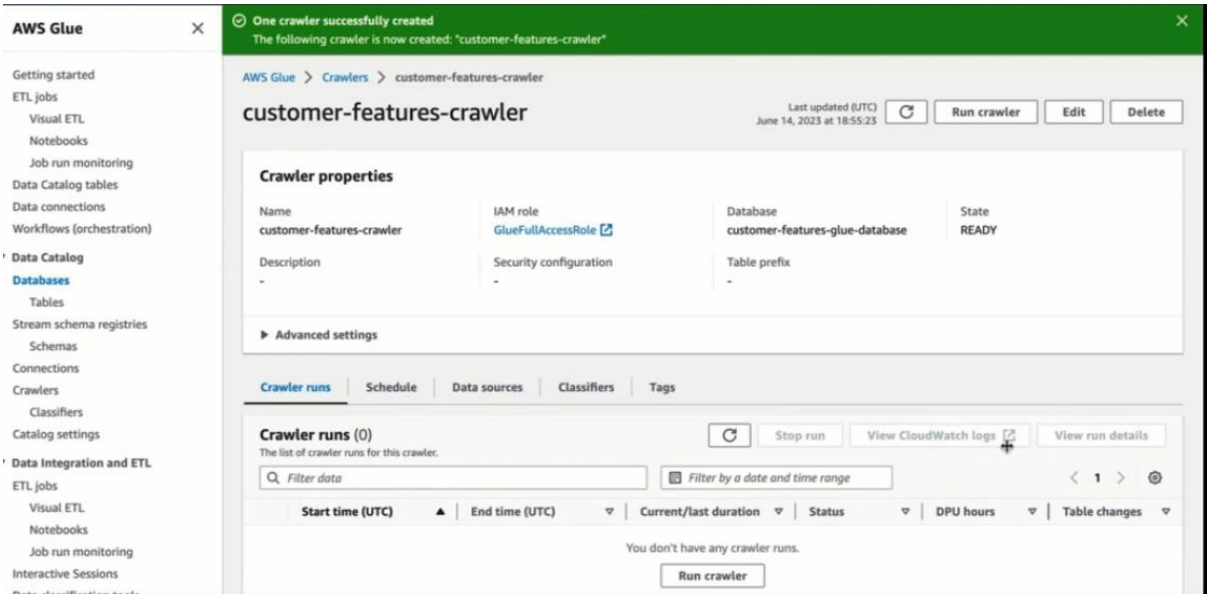



**124. REVIEW, THEN, CLICK CREATE CRAWLER.**

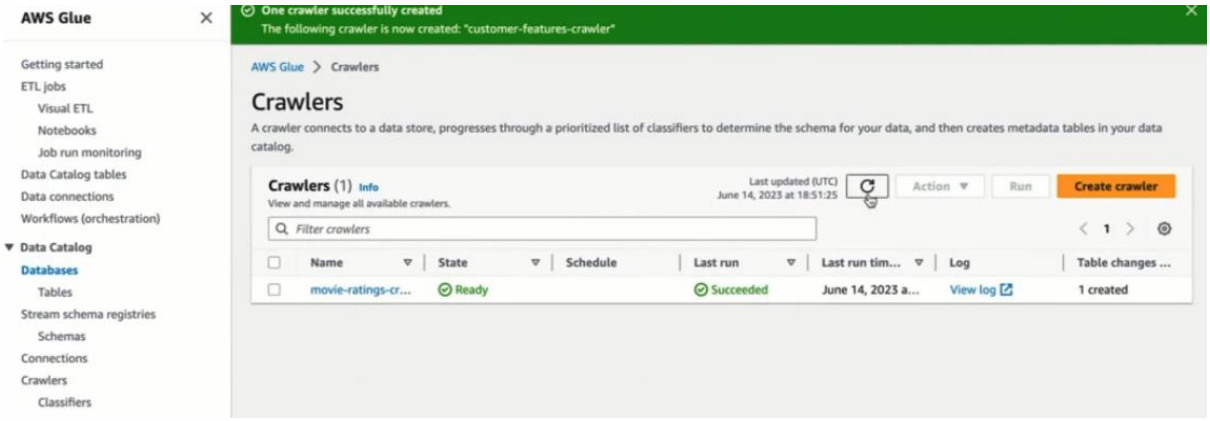


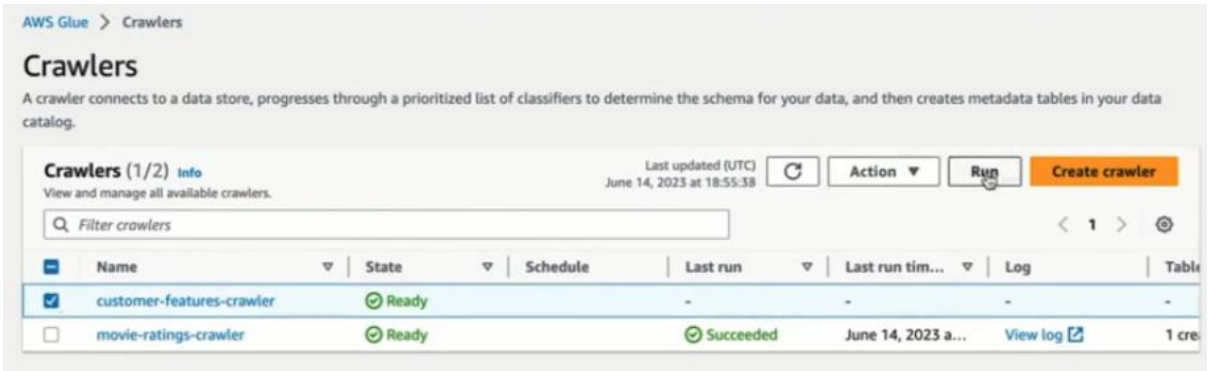

**125. CRAWLER SUCCESSFULLY CREATED, BUT, WE NEED TO RUN IT. ON THE UPPER RIGHT SIDE, CLICK RUN CRAWLER.**




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

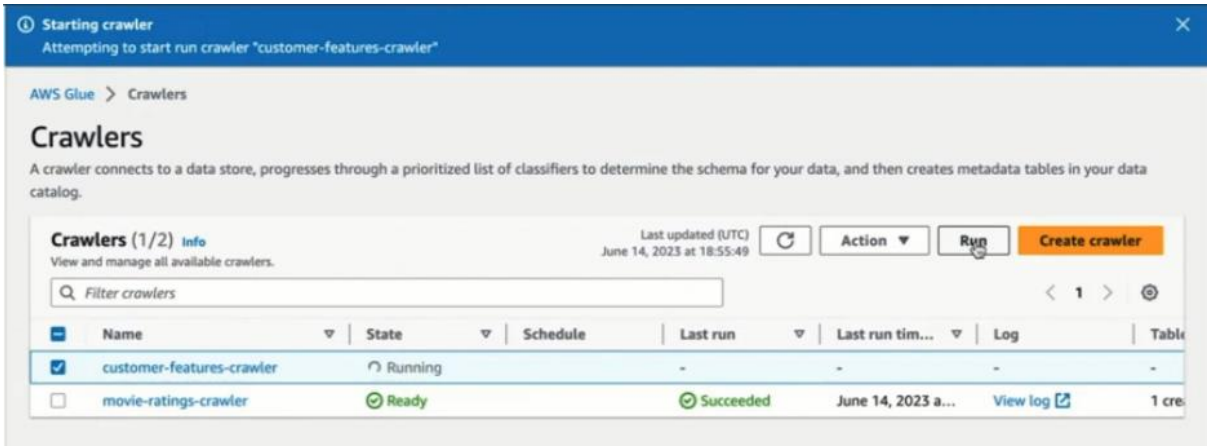



**127. SELECT THE NEW CRAWLER, CUSTOMER-FEATURES-CRAWLER. THEN,CLICK RUN.**

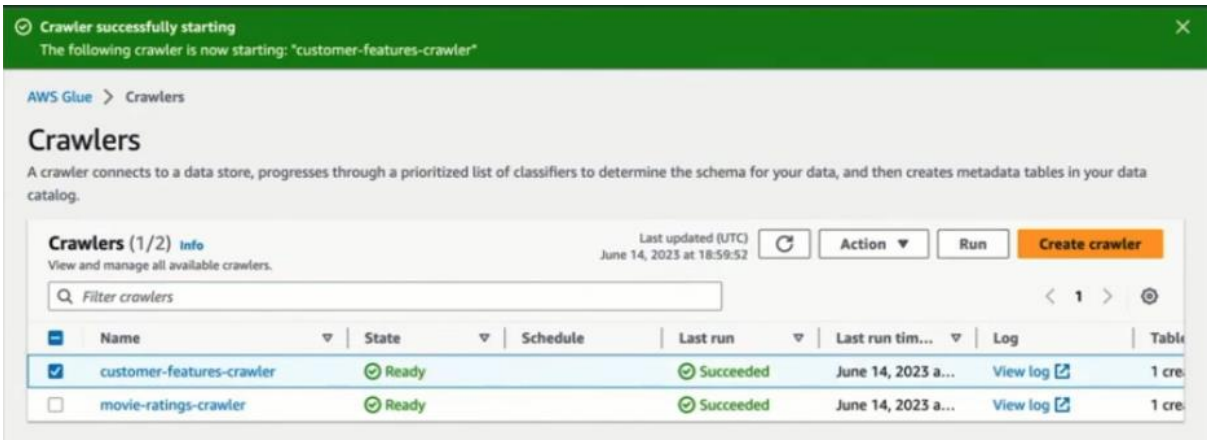



**128. AT THE TOP, YOU WILL SEE THE BLUE FLAG WHICH IS STATING THAT THE CRAWLER IS STARTING. IT IS GOING TO CRAWL THE DATA FROM THE RDS DATABASE TABLE THAT IS**

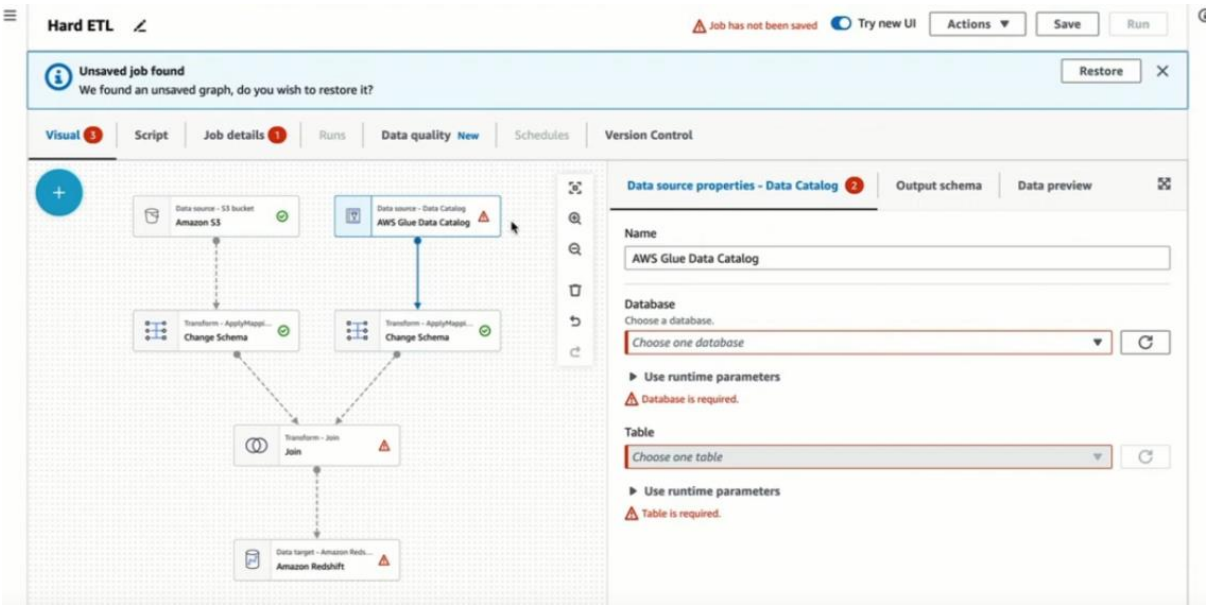
**IN THE MYSQL WORKBENCH. IT IS NOW RUNNING AND EXTRACTING THE DATA AS IN THE FIRST STEP OF THE ETL PROCESS WHICH MEANS THAT IT IS IN THE EXTRACT STEP.**



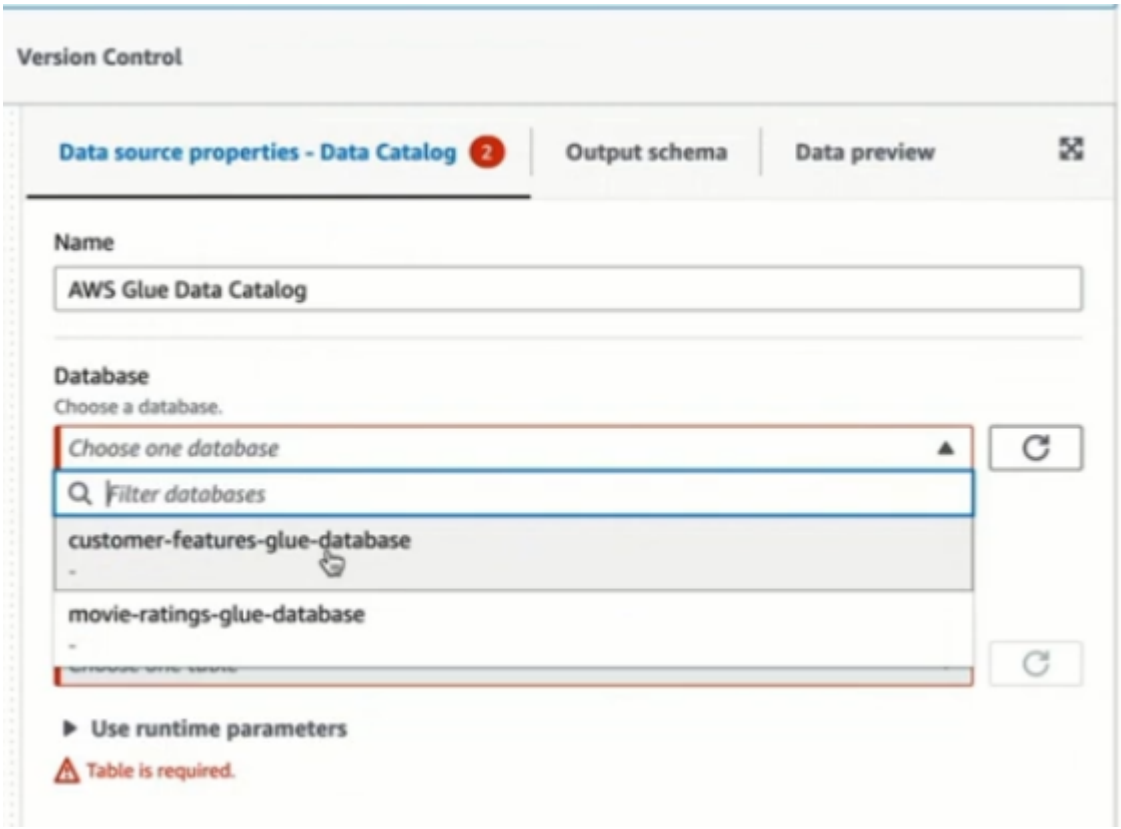

**129. CRAWLER SUCCESSFULLY STARTING. THE NEXT STEP IS TO GO BACK TO THE VISUAL EDITOR.**



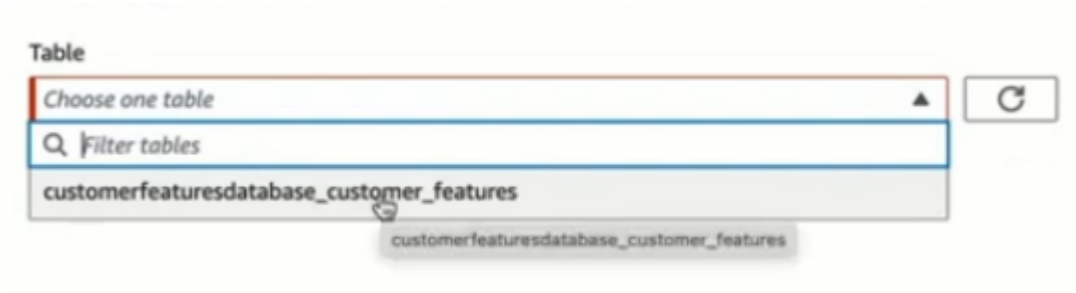

**130. HERE, WE ARE GOING TO CONNECT THE DATA CATALOG ELEMENT RELATED TO OUR DS DATABASE IN THE ETL PROCESS.**



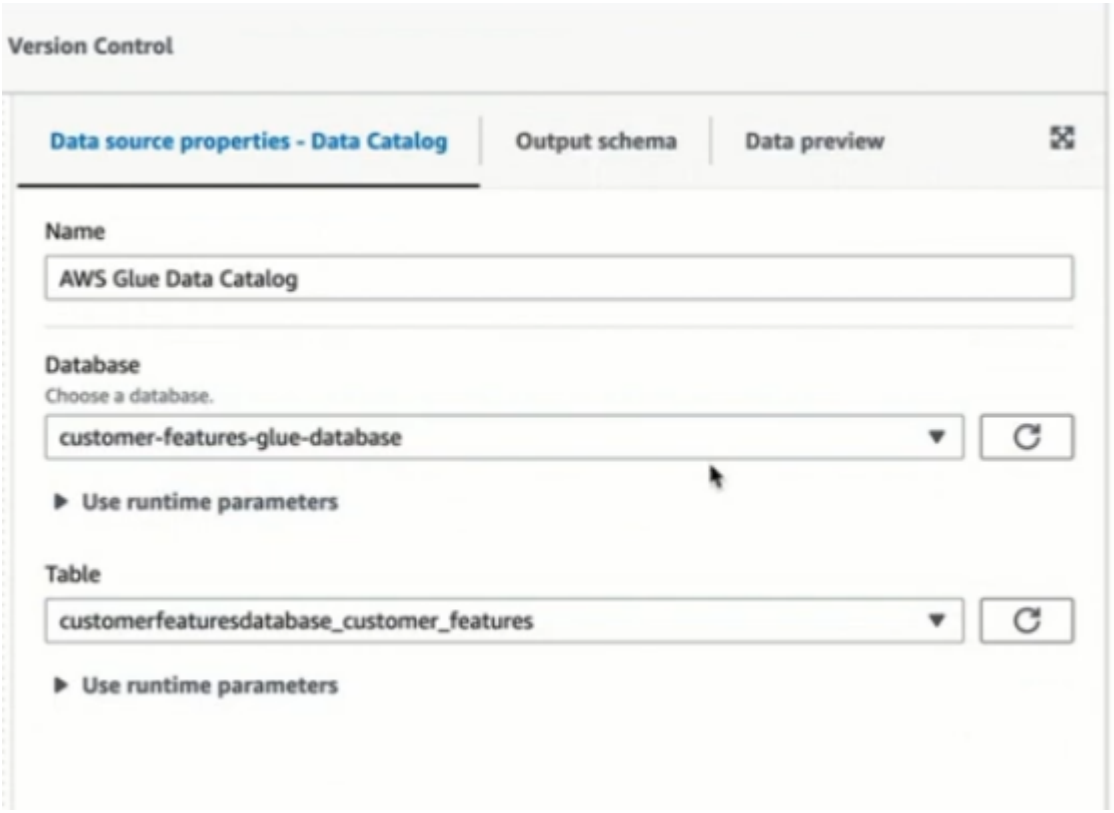

**131. IN THE NAME PORTION, MAKE SURE THAT THE AWS GLUE DATA CATALOG IS SELECTED. IN THE DATABASE, SELECT CUSTOMER-FEATURE-GLUE-DATABASE THAT WE CREATED.**



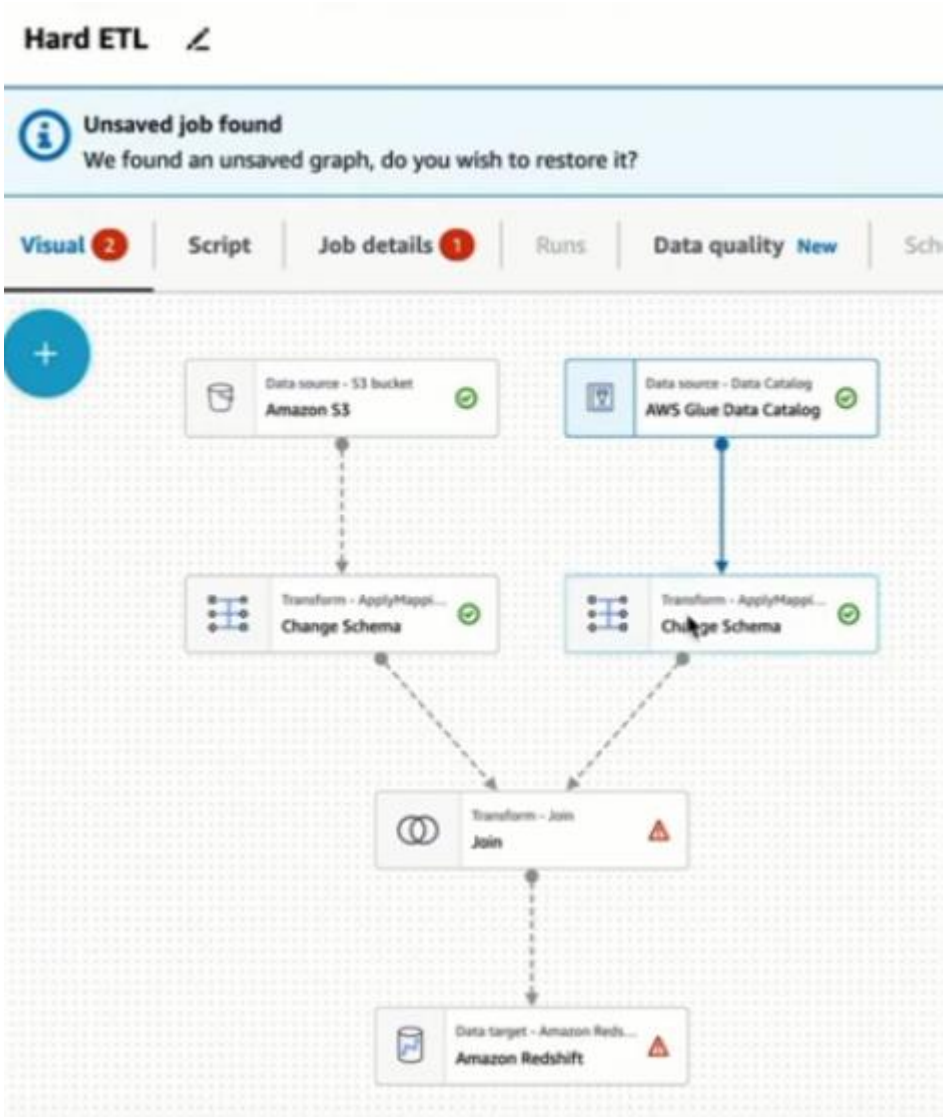

**132. IN THE TABLE, CHOOSE CUSTOMERFEATURESDATABASE_CUSTOMER_FEATURES.**



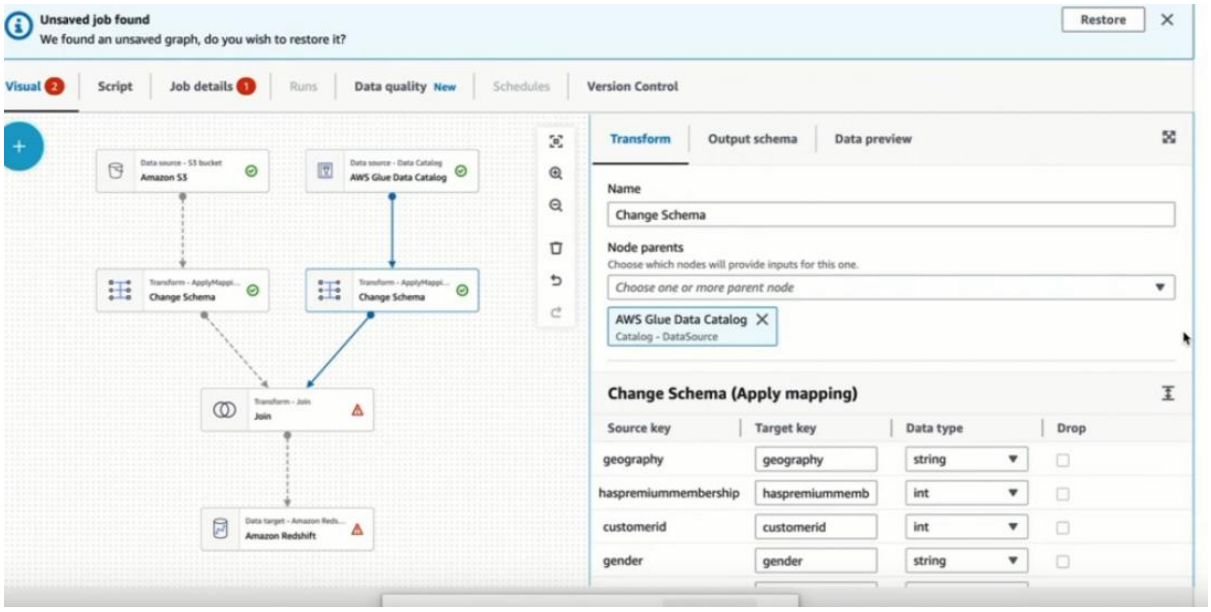

**133. IT SHOULD THEN LOOK LIKE THIS. THIS IS THE STEP OF THE EXTRACTION OF OUR SECOND DATA SOURCE RELATED TO RDS.**



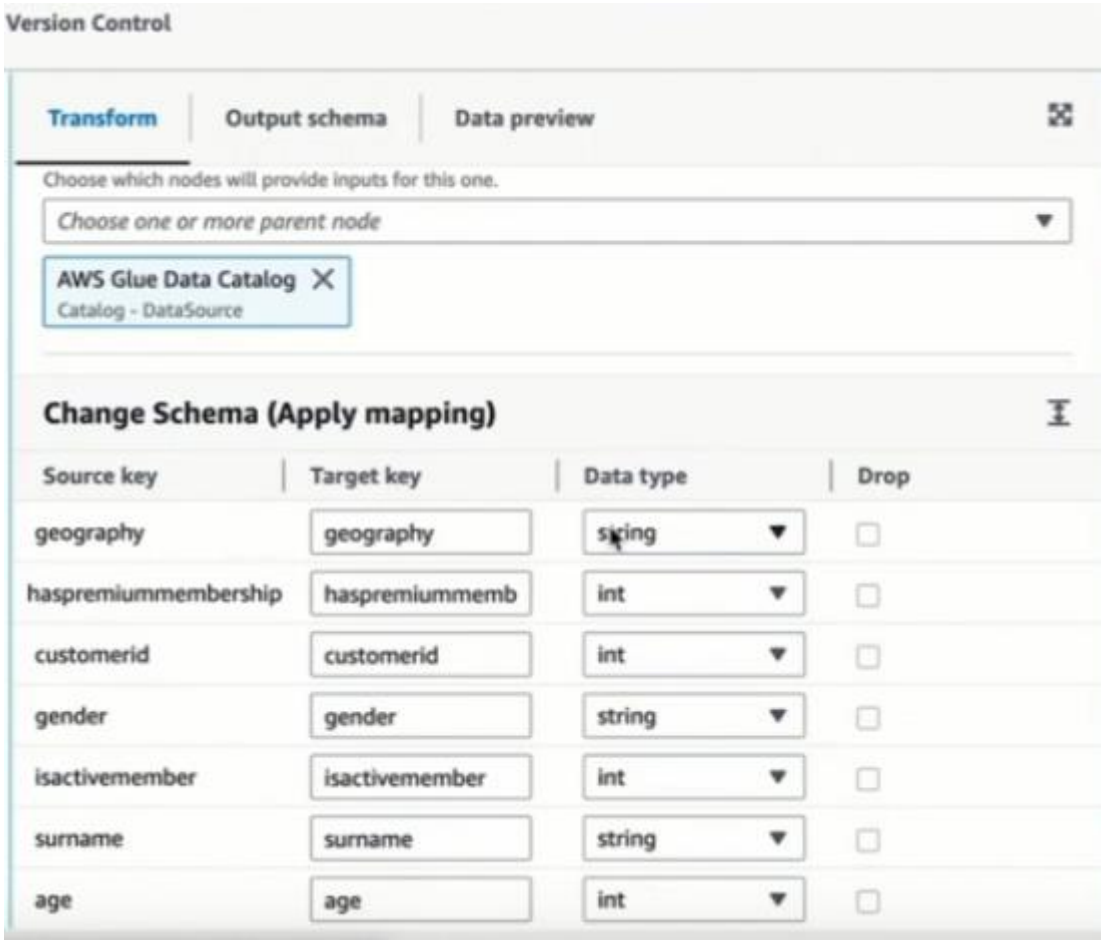

**134. WE, THEN, MOVE ON TO THE TRANSFORM CHANGE SCHEMA PHASE.**



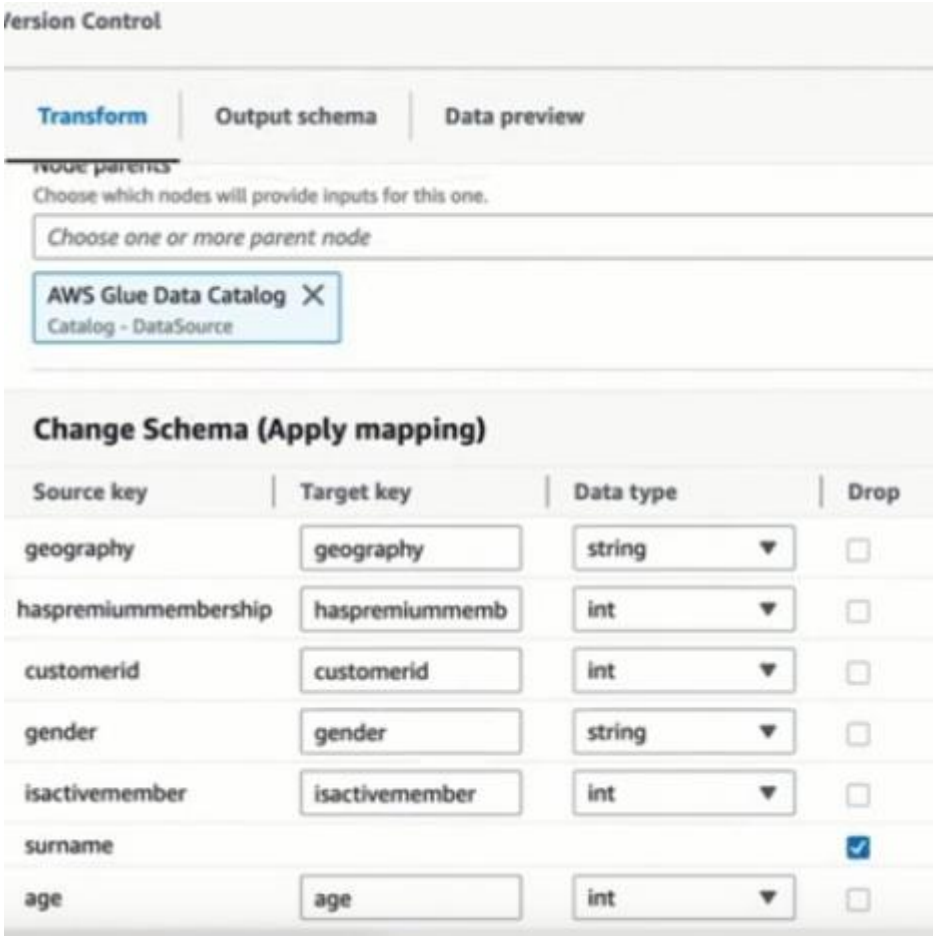

**LET`S CHECK IF WE ARE GOING TO TRANSFORM A BIT OUR DATA RELATED TO THE CUSTOMER FEATURES.**



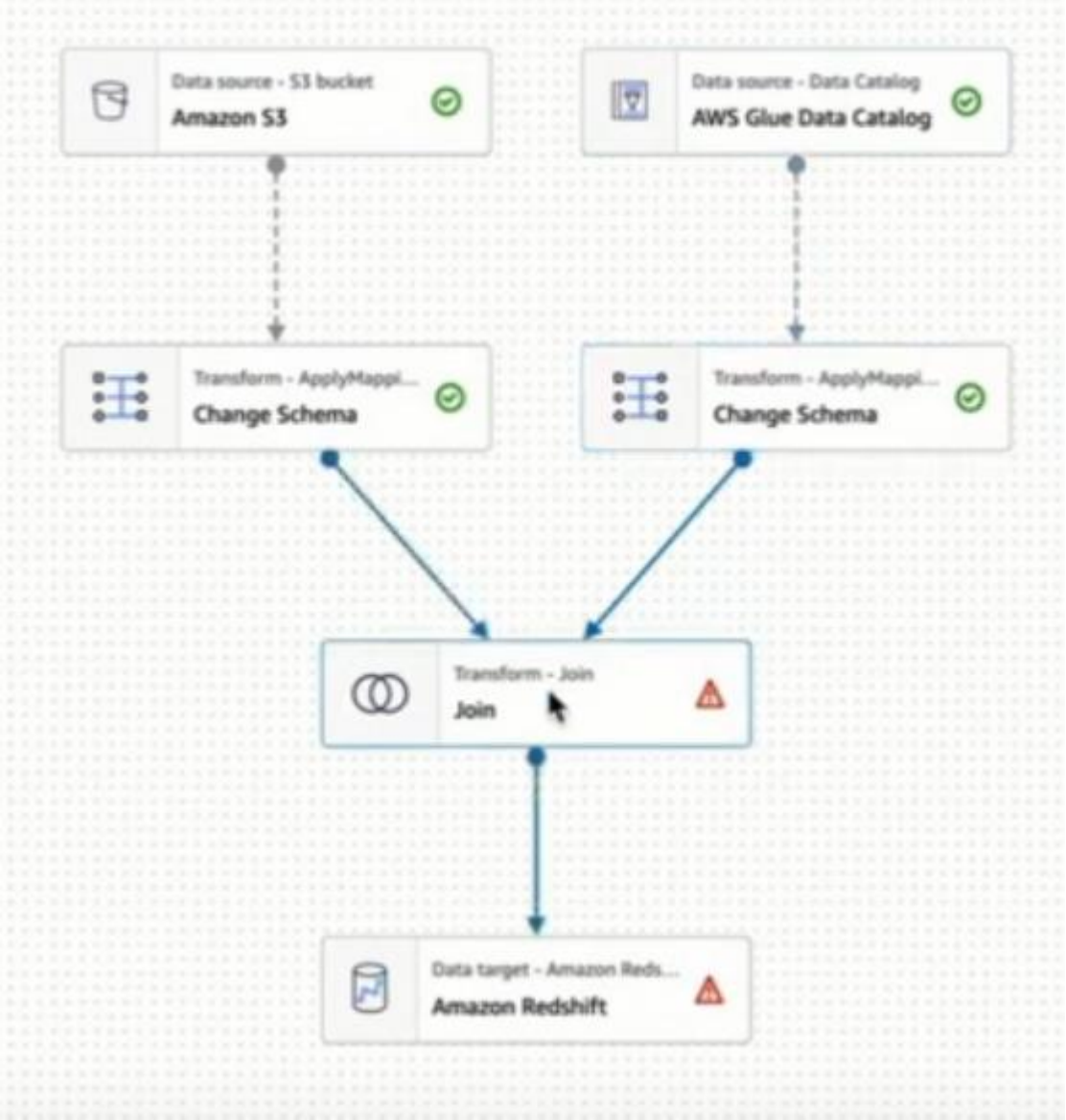

**HERE, WE CAN CHANGE THE TARGET KEY AND DATA TYPES BASED ON THE PREFERENCE. BUT FOR THIS TRAINING PURPOSES, WE WILL PUT IT AS IT IS.**



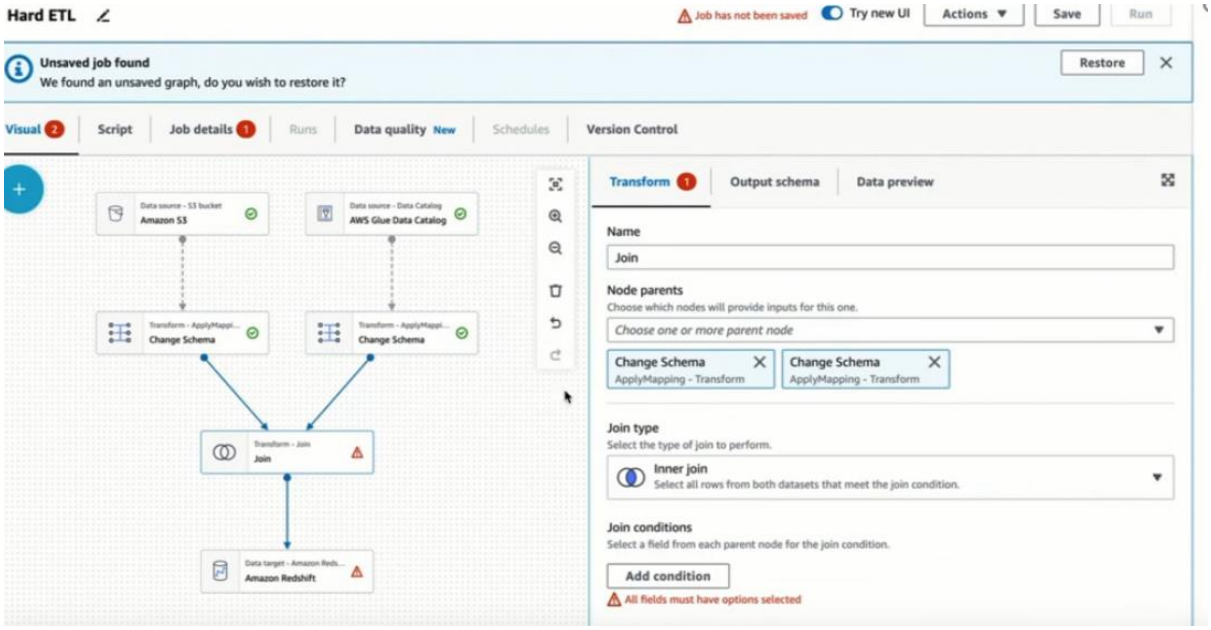
135. **BUT TO TAKE INTO CONSIDERATION IN ORDER TO BUILD THE RECOMMENDER SYSTEM, WE CAN CHECK IF THERE IS ANY FEATURE OR VARIABLES THAT WE CAN DROP BECAUSE THEY ARE NOT RELEVANT TO MAKING A RECOMMENDER SYSTEM FOR MOVIES. FOR EXAMPLE, THE SURNAME MIGHT NOT BE RELEVANT, SO, WE ARE GOING TO DROP THIS VARIABLE.**



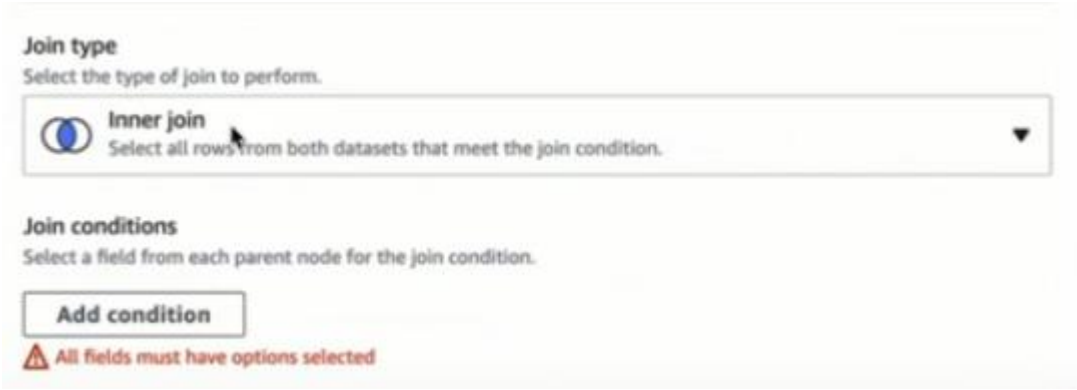

136. **WE WILL THEN PROCEED TO THE NEXT STEP, JOIN, WHICH IS STILL PART OF THE TRANSFORM STEP IN THE ETL PROCESS.**




**137. 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.**

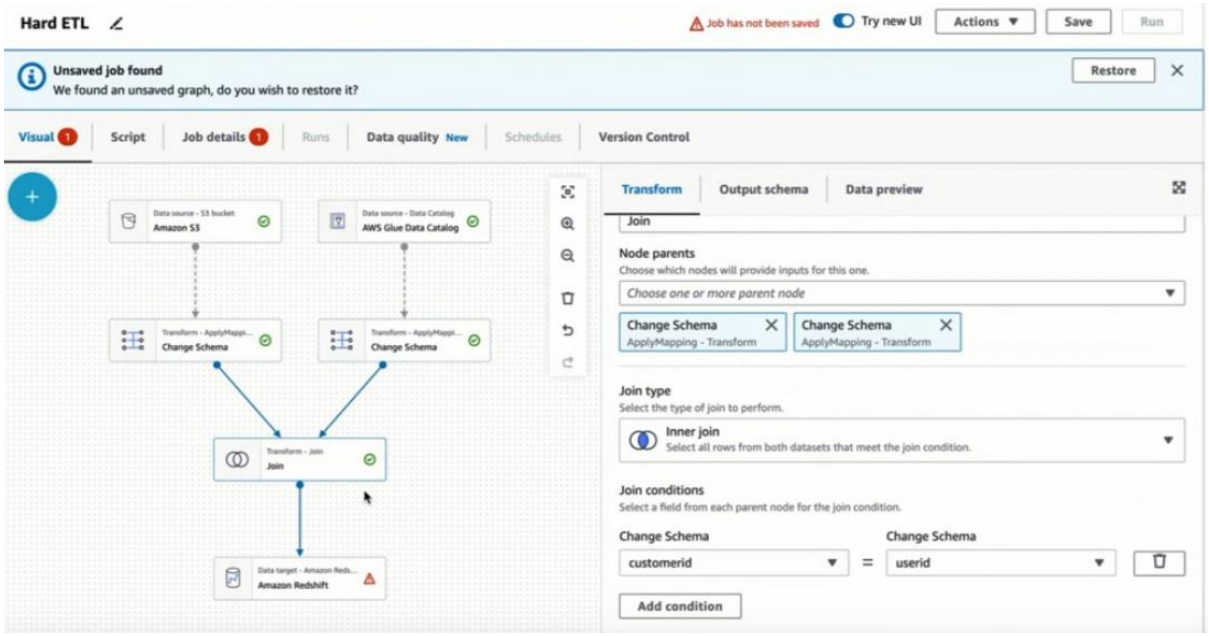



**138. TO CHOOSE THE COMMON VARIABLE, WE NEED TO CLICK THE ADD CONDITION IN THE JOIN CONDITIONS FIELD.**

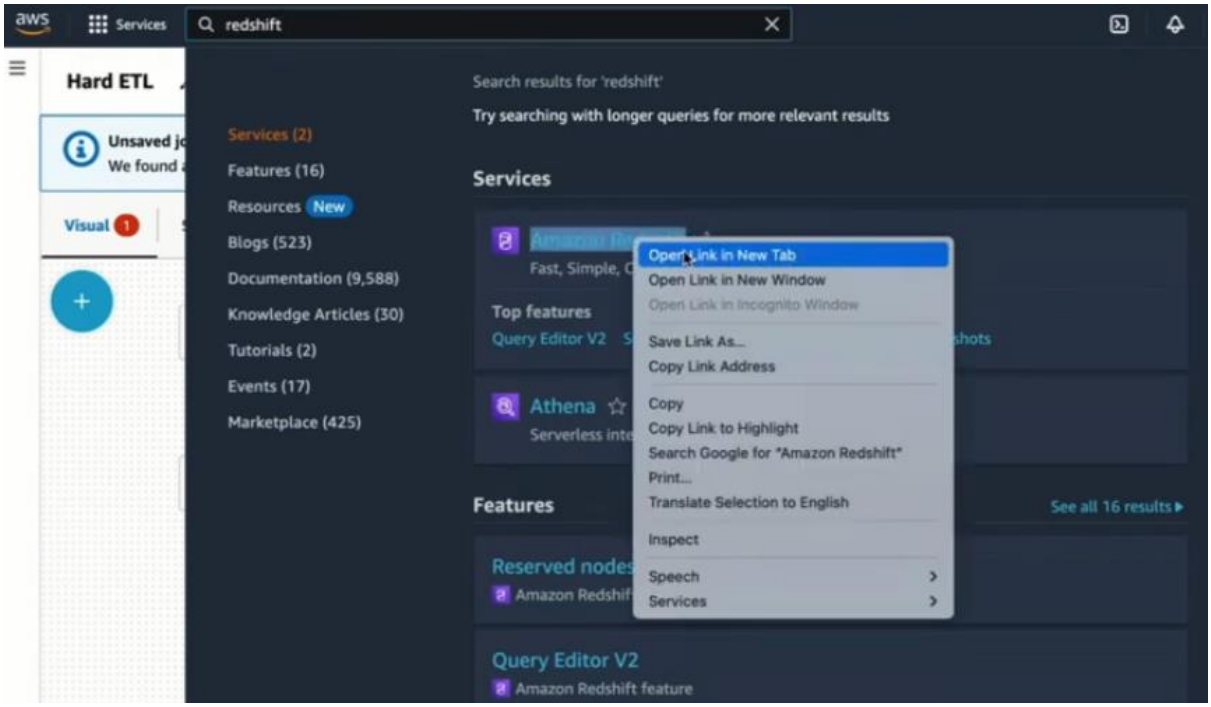



**139. IN THE CHANGE SCHEMA, CHOOSE CUSTOMERID AND USERID. WE ARE GOING TO JOIN THEM THROUGH THIS COMMON DENOMINATOR OR VARIABLE WHICH ARE THE IDS OF THE CUSTOMERS OR USERS OF THIS MOVIE STREAMING PLATFORM.**

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

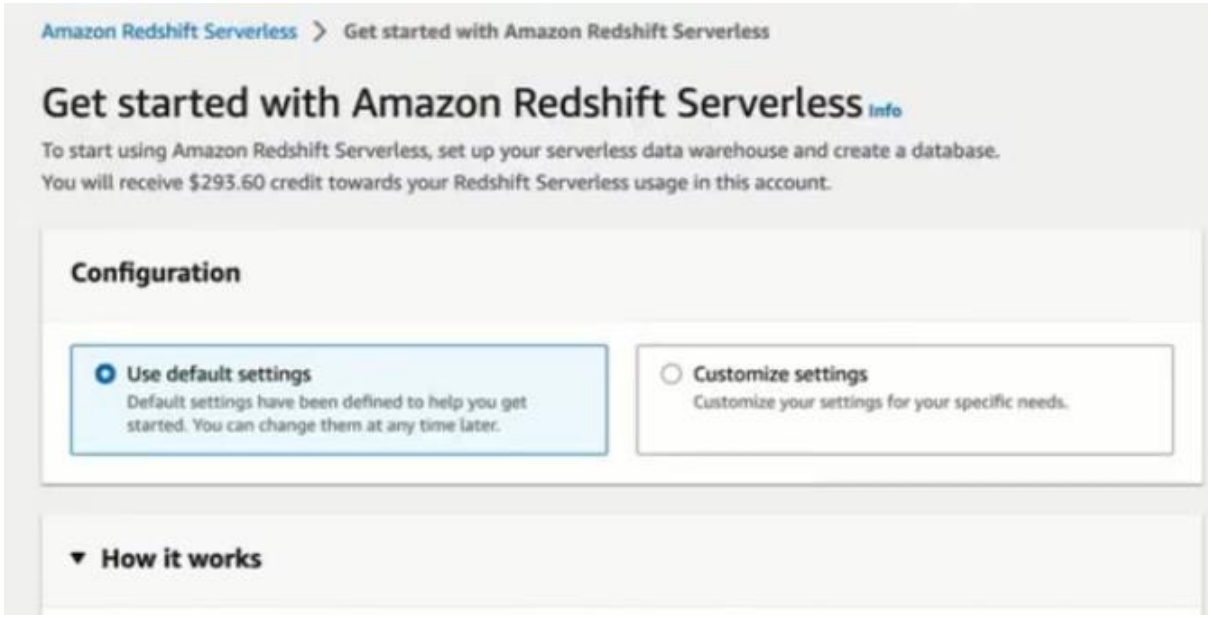



**141. MOVING ON TO THE FINAL STEP IS THE TARGET AMAZON REDSHIFT. IN THE SERVICES, TYPE REDSHIFT. OPEN IT IN THE NEW TAB.**

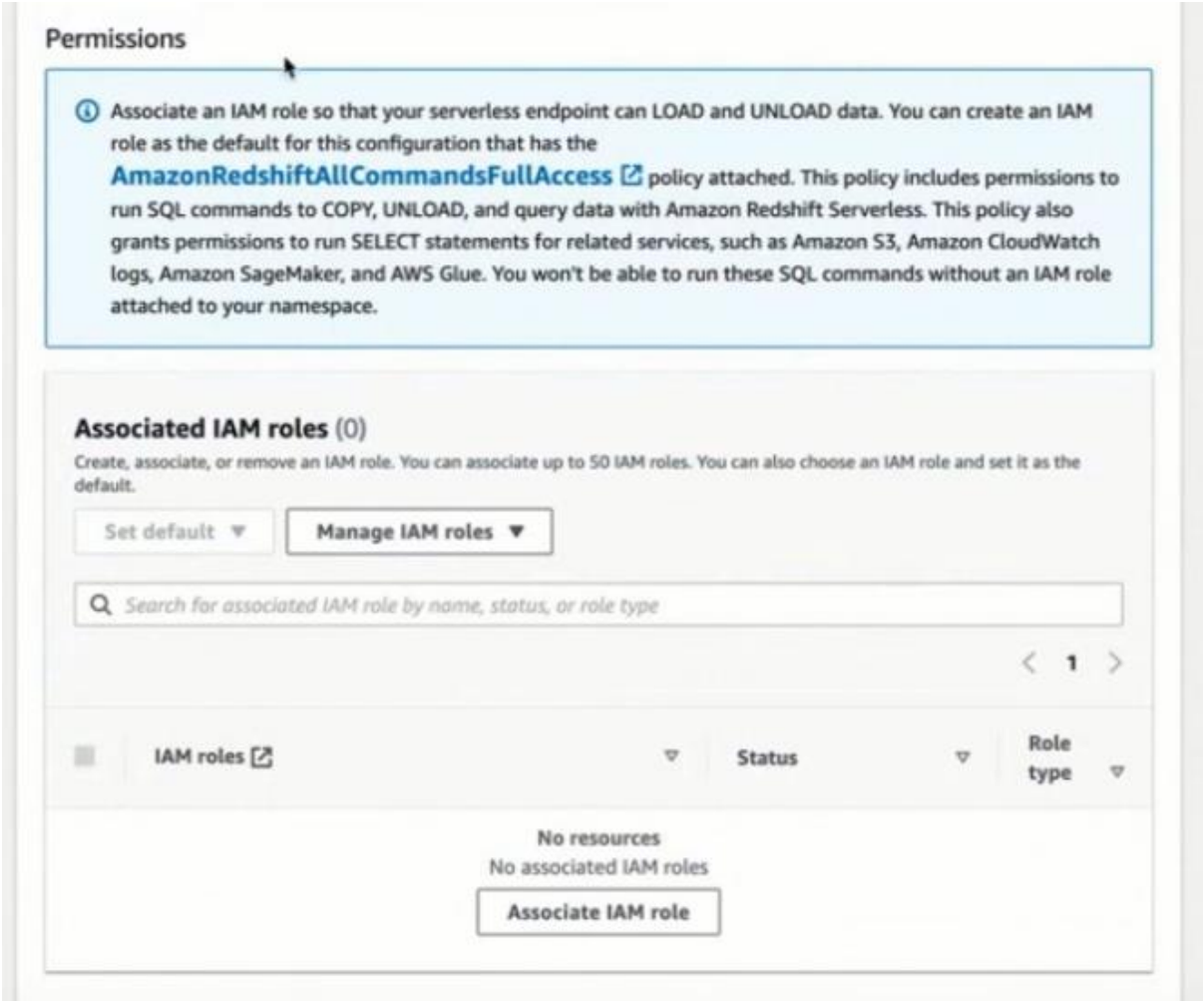



**142. CLICK THE TRY REDSHIFT SERVERLESS FREE TRIAL.**


**143. CLICK DEFAULT SETTINGS.**




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


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 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

**145.**