Assignment 05 Assignment: Customer Debit Card Purchase Aggregation

1)created mock generator lambda

2)given iam role for that lambda

3)created rds schema for the given columns

4)created s3 bucket for data to be pushed from lambda into s3 in hive style partitioning

5)made connection to mysql workbench through rds

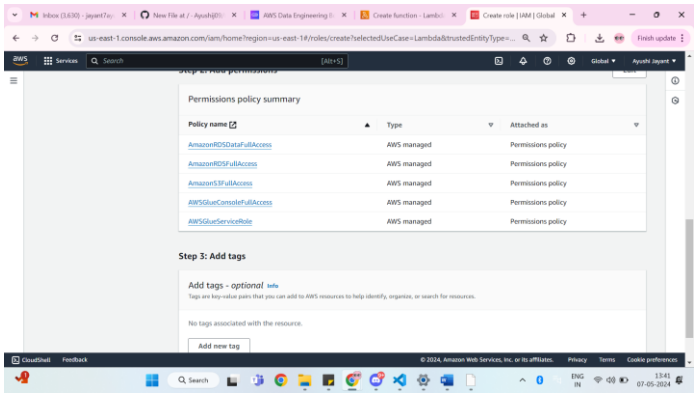
6)created database and crawler in glue

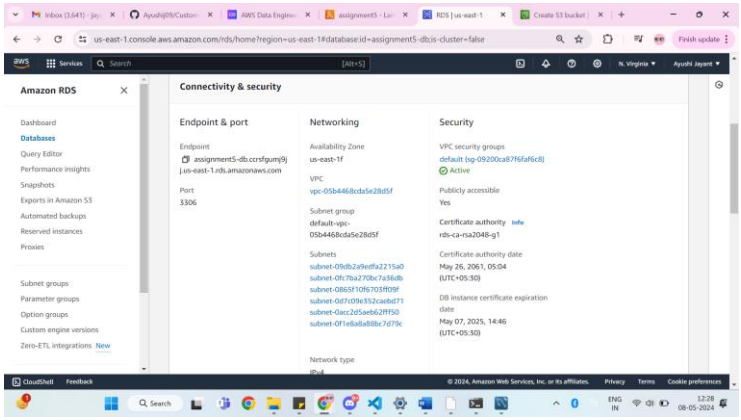
7)made connections, vpc endpoint for s3 bucket

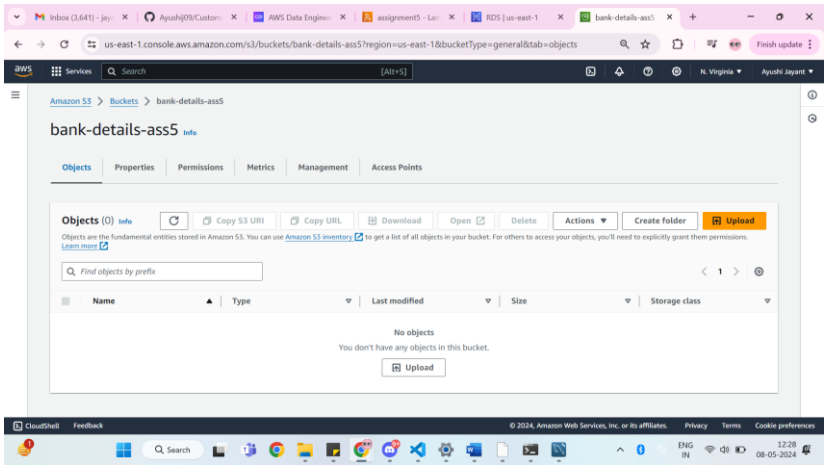
8)created crawler for mysql

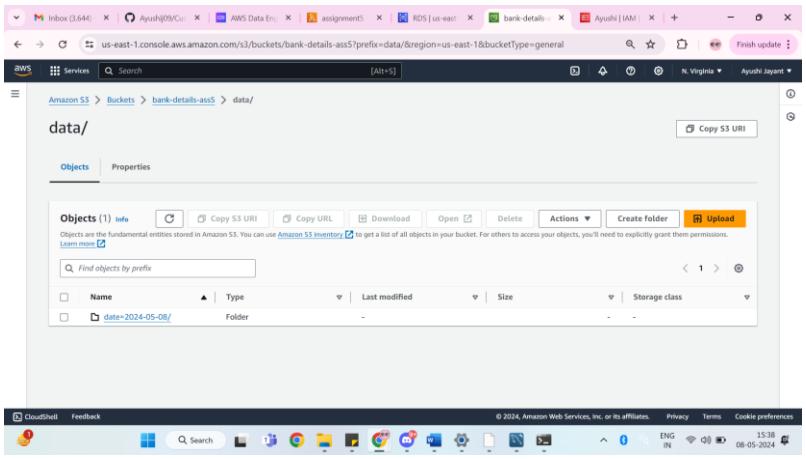
9)after the two crawlers have been made we created a visual etl glue job and applied all the relevant functions happening then inserted the data into mysql rds by changing schema

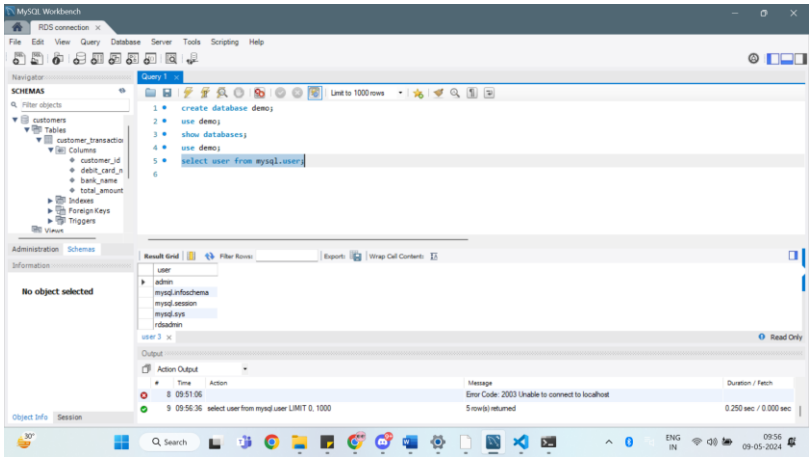
10)The job ran successfully

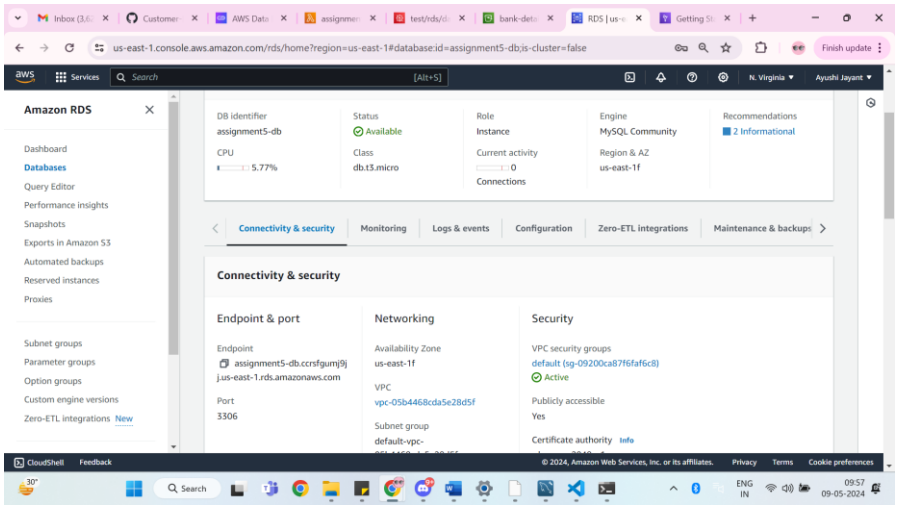


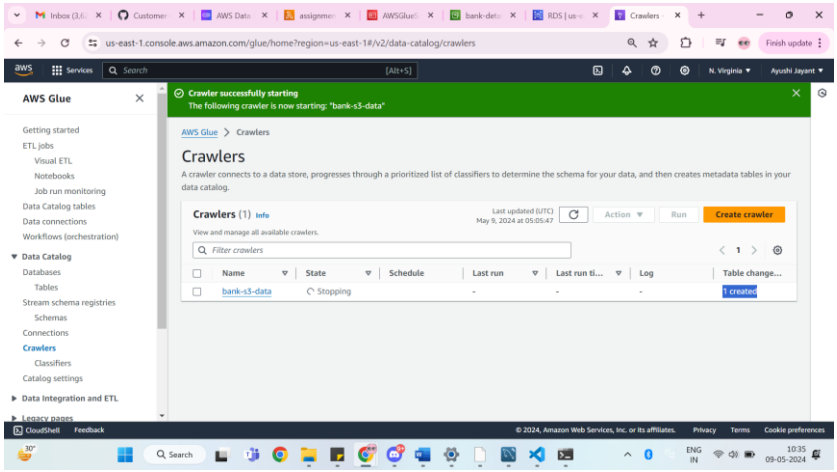


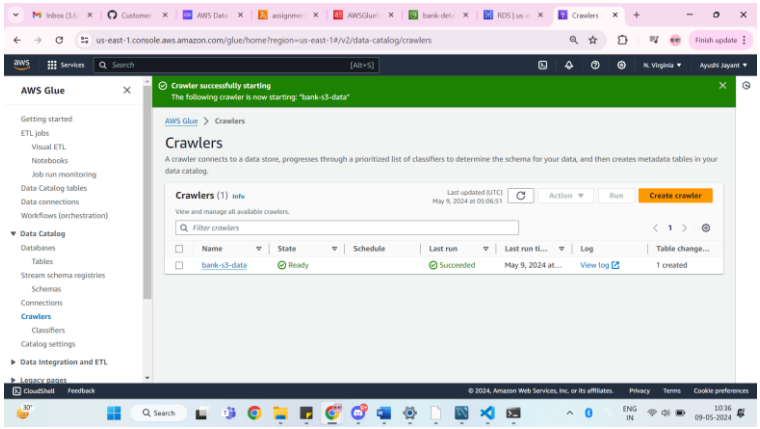

us-east-1.console.aws.amazon.com/s3/buckets/bank-details-ass5?region=us-east-1&bucketType=general&tab=objects
Amazon S3 > Buckets > bank-details-ass5
bank-details-ass5 Info
Objects
Properties
Permissions
Metrics
Management
Access Points
Objects (0) Info
Copy S3 URI
Copy URL
Download
Open
Delete
Actions
Create folder
Upload
Find objects by prefix
Name
Type
Last modified
Size
Storage class
No objects
You don't have any objects in this bucket.
Upload
CloudShell
Feedback
© 2024, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

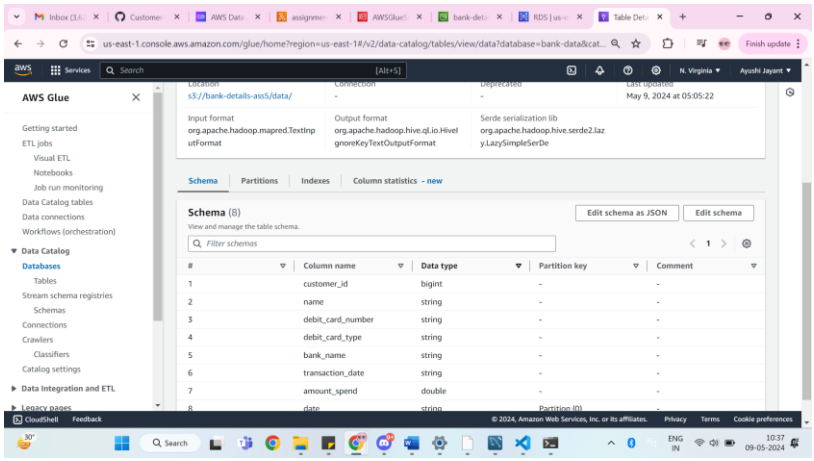

us-east-1.console.aws.amazon.com/s3/buckets/bank-details-ass5?prefix=data/&region=us-east-1&bucketType=general
Amazon S3 > Buckets > bank-details-ass5 > data/
data/
Copy S3 URI
Objects
Properties
Objects (1) Info
Copy S3 URI
Copy URL
Download
Open
Delete
Actions
Create folder
Upload
Find objects by prefix
Name
Type
Last modified
Size
Storage class
date=2024-05-08/
Folder
CloudShell
Feedback
© 2024, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

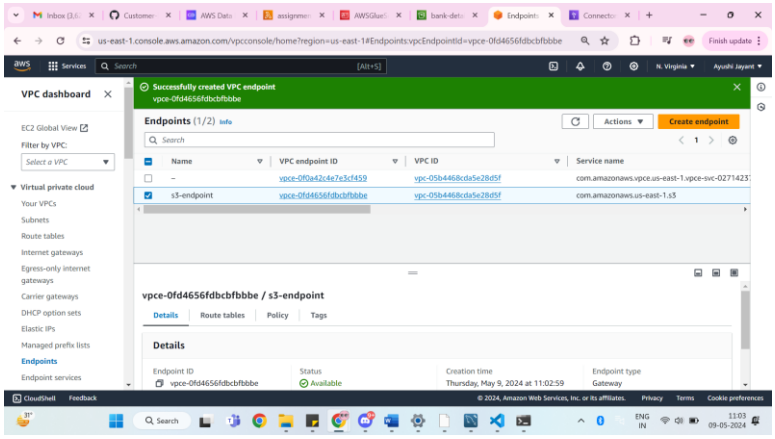

MySQL Workbench
RDS connection
File Edit View Query Database Server Tools Scripting Help
SCHEMAS
customers
Tables
customer_transactio
Columns
customer_id
debit_card_n
bank_name
total_amount
Indexes
Foreign Keys
Triggers
Views
Administration
Schemas
Information
No object selected
Query 1
1 create database demo;
2 use demo;
3 show databases;
4 use demo;
5 select user from mysql.user;
Result Grid
user
admin
mysql.infoschema
mysql.session
mysql.sys
rdsadmin
Read Only
Output
Action Output
Time Action Message Duration / Fetch
8 09:51:06 Error Code: 2003 Unable to connect to localhost
9 09:56:36 select user from mysql.user LIMIT 0, 1000 5 row(s) returned 0.250 sec / 0.000 sec

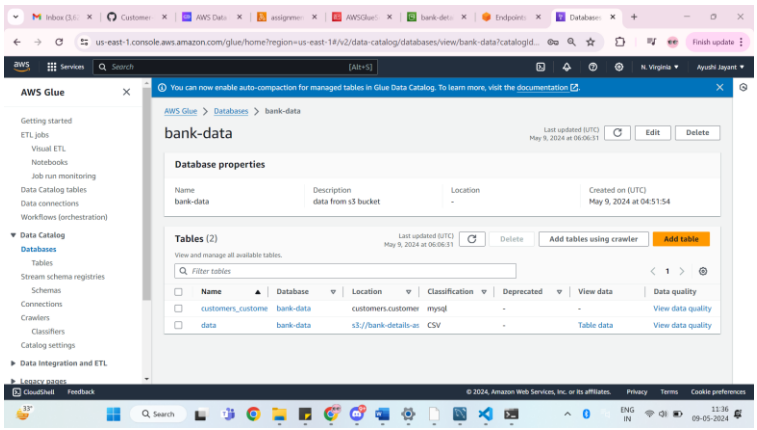

us-east-1.console.aws.amazon.com/rds/home?region=us-east-1#database:id=assignment5-db;is-cluster=false
Amazon RDS
Dashboard
Databases
Query Editor
Performance insights
Snapshots
Exports in Amazon S3
Automated backups
Reserved instances
Proxies
Subnet groups
Parameter groups
Option groups
Custom engine versions
Zero-ETL integrations New
DB identifier
assignment5-db
Status
Available
Role
Instance
Engine
MySQL Community
Recommendations
2 Informational
CPU
5.77%
Class
db.t3.micro
Current activity
0
Connections
Region & AZ
us-east-1f
Connectivity & security
Monitoring
Logs & events
Configuration
Zero-ETL integrations
Maintenance & backups
Connectivity & security
Endpoint & port
Endpoint
assignment5-db.ccrsfgumj9j
j.us-east-1.rds.amazonaws.com
Port
3306
Networking
Availability Zone
us-east-1f
VPC
vpc-05b4468cda5e28d5f
Subnet group
Security
VPC security groups
default (sg-09200ca87f6faf6c8)
Active
Publicly accessible
Yes
Certificate authority Info

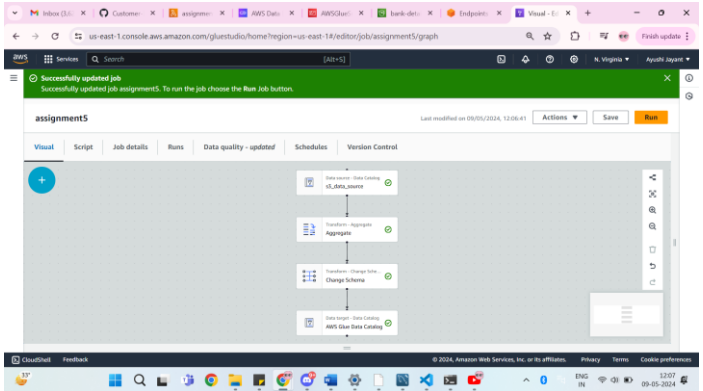

us-east-1.console.aws.amazon.com/glue/home?region=us-east-1#/v2/data-catalog/crawlers
Crawler successfully starting
The following crawler is now starting: "bank-s3-data"
AWS Glue
Getting started
ETL jobs
Visual ETL
Notebooks
Job run monitoring
Data Catalog tables
Data connections
Workflows (orchestration)
Data Catalog
Databases
Tables
Stream schema registries
Schemas
Connections
Crawlers
Classifiers
Catalog settings
Data Integration and ETL
Legacy pages
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.
Crawlers (1) Info
Last updated (UTC)
May 9, 2024 at 05:05:47
Action
Run
Create crawler
View and manage all available crawlers.
Filter crawlers
Name
State
Schedule
Last run
Last run ti...
Log
Table change...
bank-s3-data
Stopping
1 created

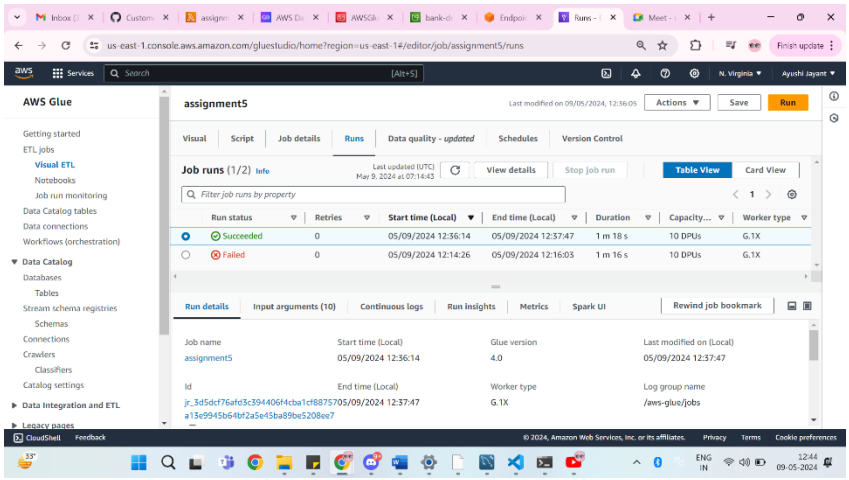

us-east-1.console.aws.amazon.com/glue/home?region=us-east-1#/v2/data-catalog/crawlers
Crawler successfully starting
The following crawler is now starting: "bank-s3-data"
AWS Glue
Getting started
ETL jobs
Visual ETL
Notebooks
Job run monitoring
Data Catalog tables
Data connections
Workflows (orchestration)
Data Catalog
Databases
Tables
Stream schema registries
Schemas
Connections
Crawlers
Classifiers
Catalog settings
Data Integration and ETL
Legacy pages
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.
Crawlers (1) Info
Last updated (UTC)
May 9, 2024 at 05:06:51
Action
Run
Create crawler
View and manage all available crawlers.
Filter crawlers
Name
State
Schedule
Last run
Last run ti...
Log
Table change...
bank-s3-data
Ready
Succeeded
May 9, 2024 at...
View log
1 created

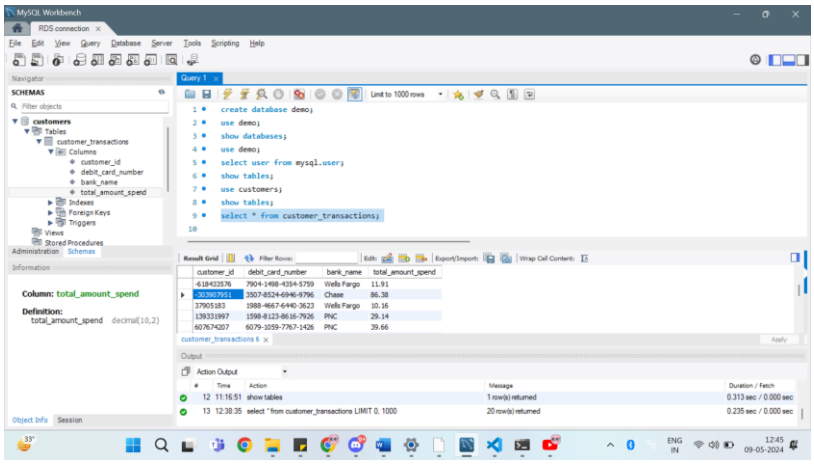

AWS Glue
Schema (8)
Edit schema as JSON
Edit schema


VPC dashboard
Endpoints (1/2)
Create endpoint
vpce-0fd4656fdbcbfbbbe / s3-endpoint


AWS Glue
bank-data
Database properties
Tables (2)


assignment5
Successfully updated job


AWS Glue
assignment5
Visual
Script
Job details
Runs
Data quality - updated
Schedules
Version Control
Job runs (1/2)
View details
Stop job run
Table View
Card View
Run status
Retries
Start time (Local)
End time (Local)
Duration
Capacity...
Worker type
Succeeded
Failed
Run details
Input arguments (10)
Continuous logs
Run insights
Metrics
Spark UI
Rewind job bookmark
Job name
assignment5
Start time (Local)
05/09/2024 12:36:14
Glue version
4.0
Last modified on (Local)
05/09/2024 12:37:47
Id
End time (Local)
Worker type
G.1X
Log group name
/aws-glue/jobs


MySQL Workbench
SCHEMAS
customers
customer_transactions
create database demo;
use demo;
show databases;
use demo;
select user from mysql.user;
show tables;
use customers;
show tables;
select * from customer_transactions;
Column: total_amount_spend
Definition:
total_amount_spend decimal(10,2)