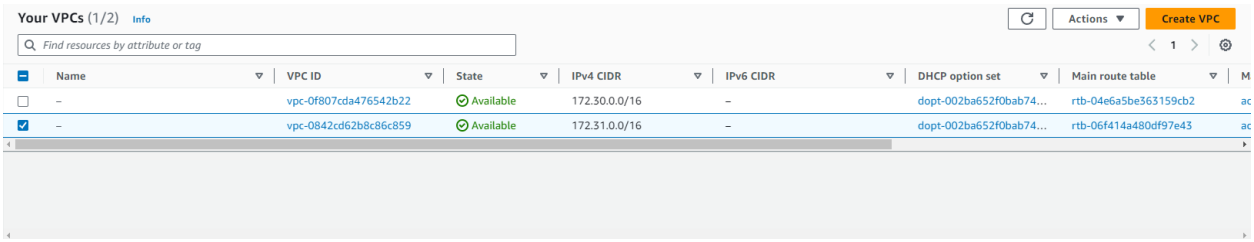
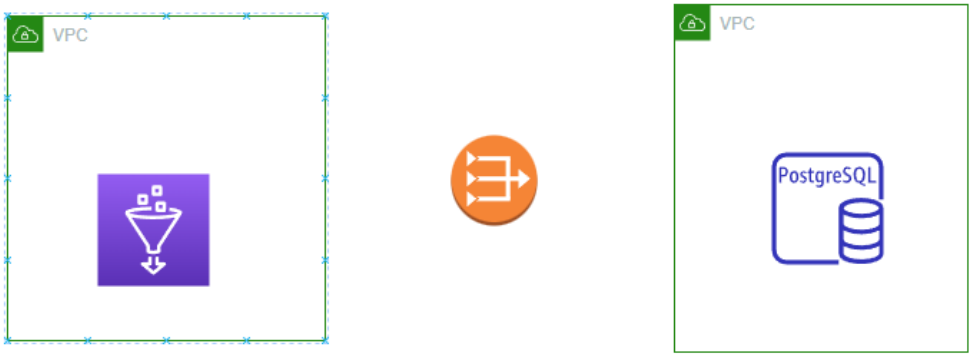
Basic Environment Setup:

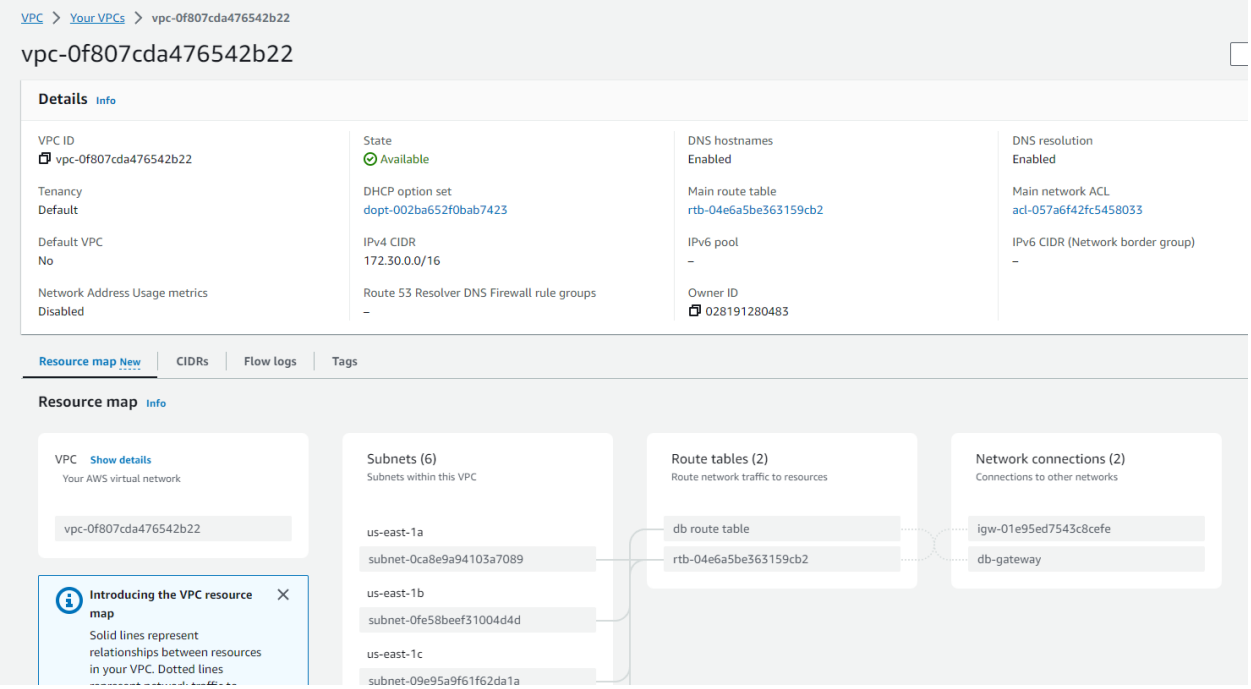
Two VPCs created with one Glue Service in one VPC and the Database in another.

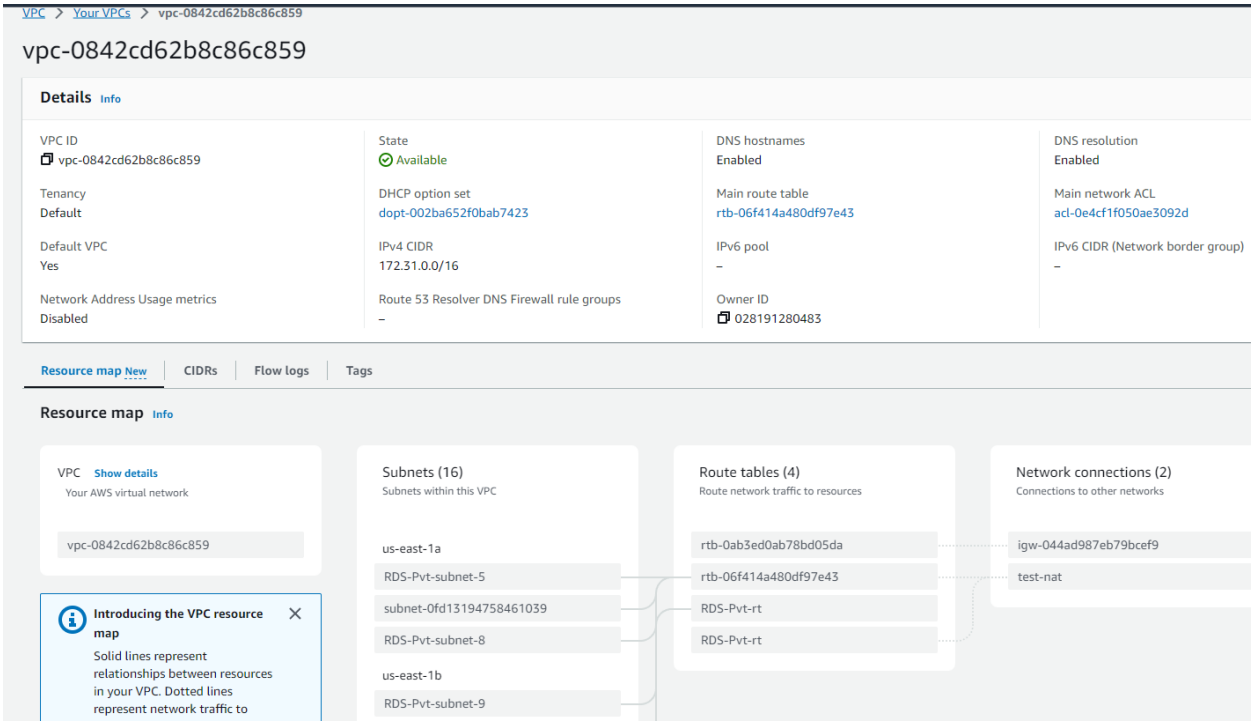
(Had trouble with using just one vpc)



Link used to get the 2 connected: https://repost.aws/knowledge-center/glue-s3-endpoint-validation-failed







Database Setup:

Create one RDS Postgres DB. This will mimic the sara sync db

RDS > Databases > database-13e

database-13e

Summary

DB identifier database-13e	CPU 4.57%	Status Available	Class db.t3.micro
Role Instance	Current activity 0 Connections	Engine PostgreSQL	Region & AZ us-east-1a

Connectivity & security | Monitoring | Logs & events | Configuration | Maintenance & backups | Tags

Connectivity & security

Endpoint & port

Endpoint
database-13e.czhqbccm3aqo.us-east-1.rds.amazonaws.com

Port
5432

Networking

Availability Zone
us-east-1a

VPC
vpc-0f807cda476542b22

Subnet group
default-vpc-0f807cda476542b22

Subnets
subnet-0fe58beef31004d4d
subnet-0bebc40291688c2ef
subnet-09e95a9f61f62da1a
subnet-02e06ac226b1e7a76

Security

VPC security groups
default (sg-02e85989bae844b19)
Active

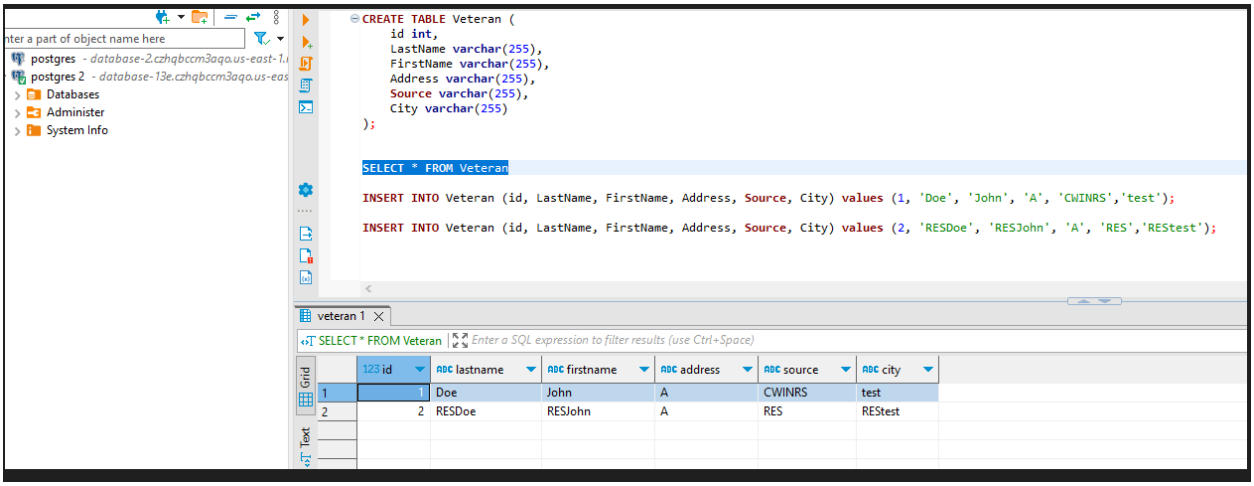
Publicly accessible
Yes

Certificate authority Info
rds-ca-2019

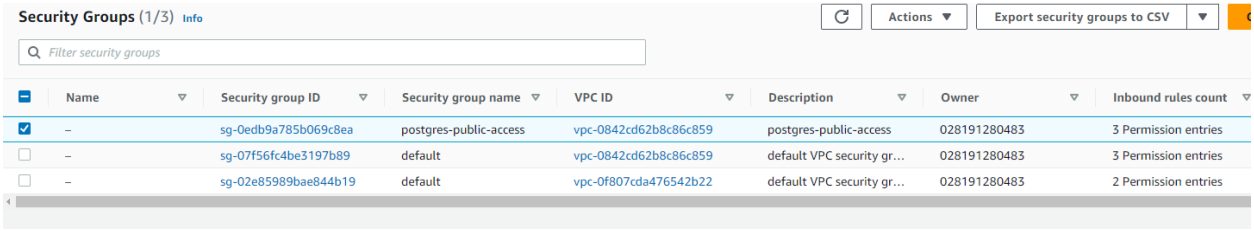
Certificate authority date
August 22, 2024, 11:08 (UTC-06:00)

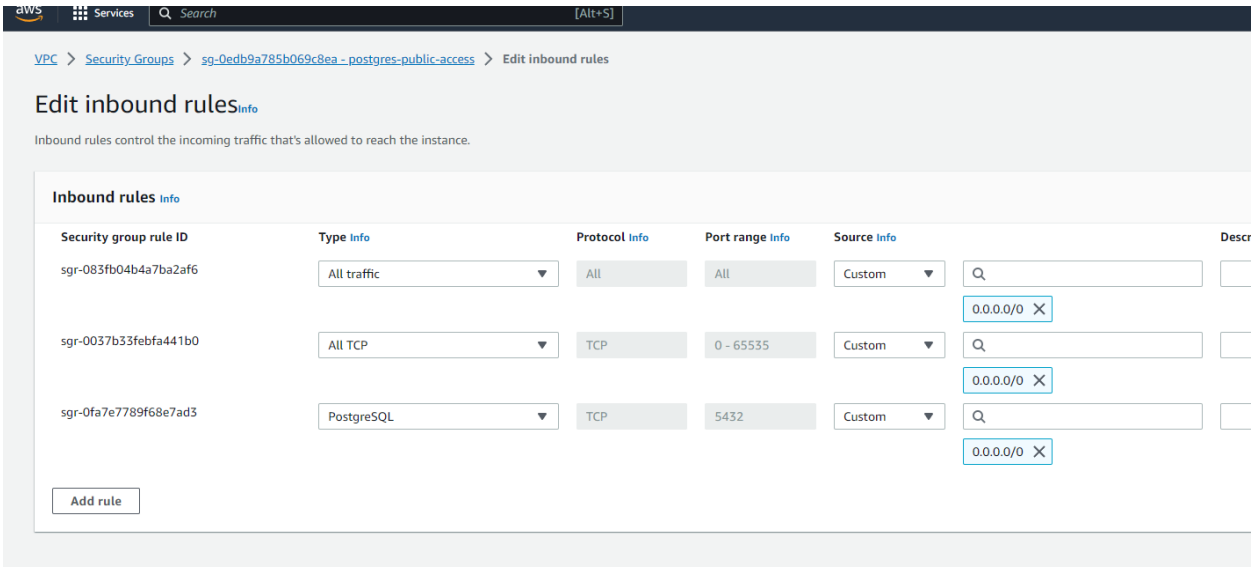
DB instance certificate expiration date

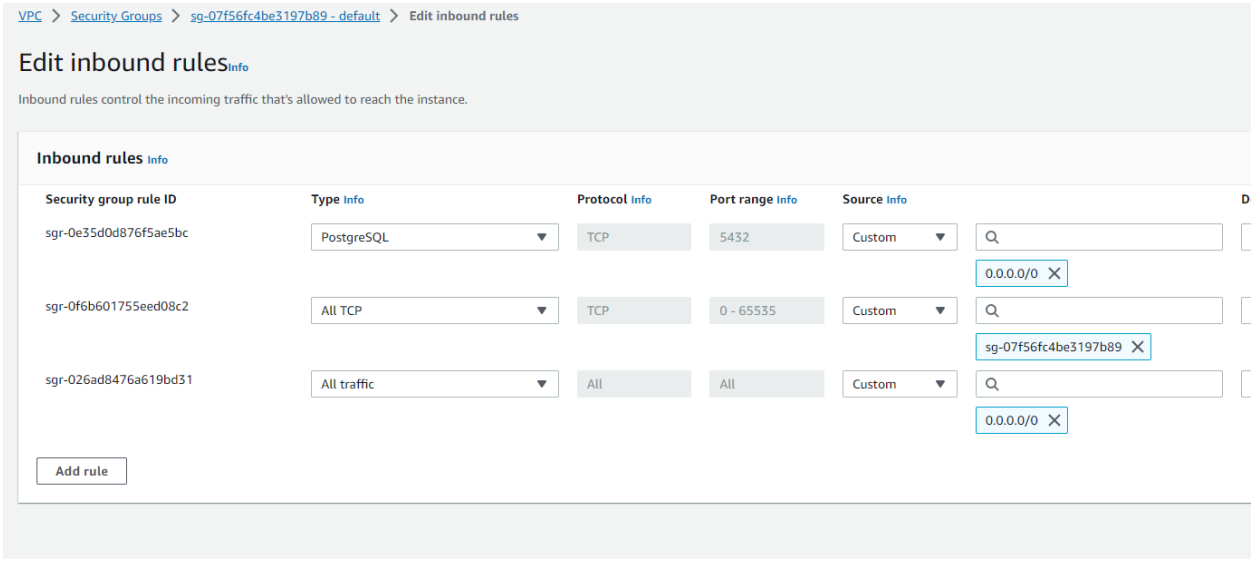
Using a DB client, Created a Veteran table and populated, schema simplified for spike purposes.



Security Setup:







IAM > Roles > glueadmin

glueadmin

Allows Glue to call AWS services on your behalf.

Summary

Creation date
August 10, 2023, 13:09 (UTC-06:00)

ARN
arn:aws:iam::028191280483:role/glueadmin

Last activity
17 hours ago

Maximum session duration
1 hour

Permissions | Trust relationships | Tags | Access Advisor | Revoke sessions

Permissions policies (9) Info

You can attach up to 10 managed policies.

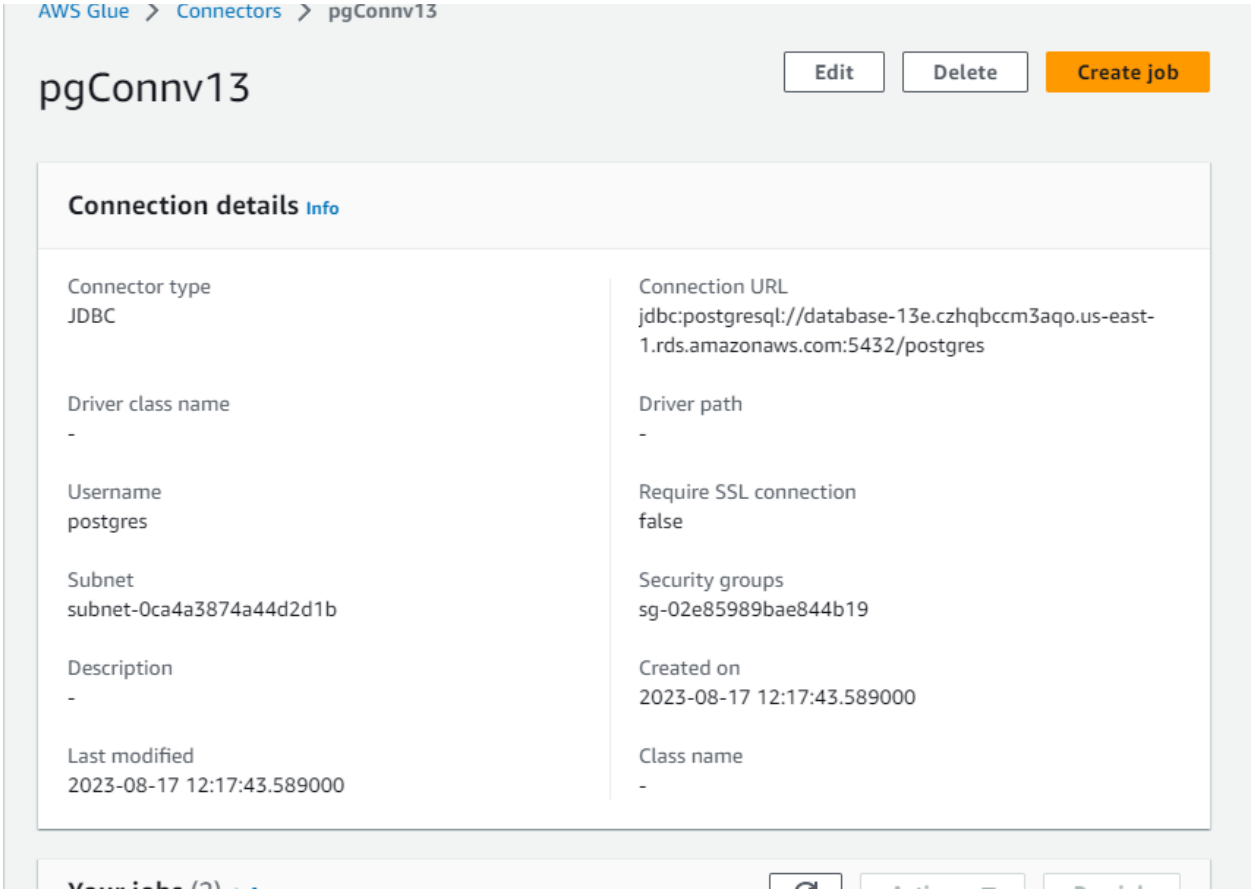
Simulate | Remove

Filter policies by property or policy name and press enter.

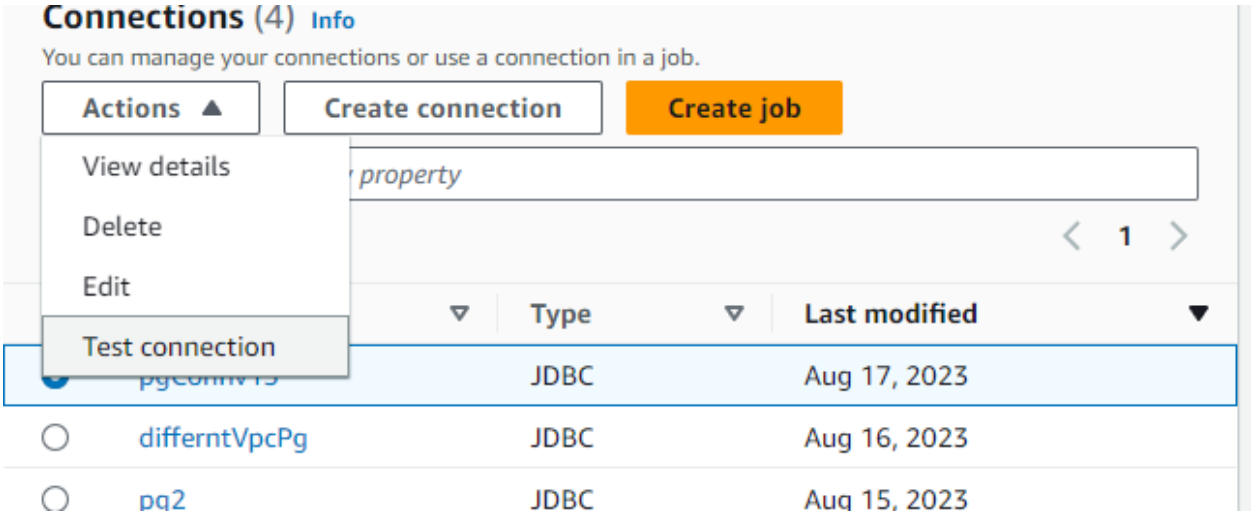
Policy name	Type
AdministratorAccess	AWS managed - job function
AmazonEC2FullAccess	AWS managed
AmazonRDSReadOnlyAccess	AWS managed
AmazonS3FullAccess	AWS managed
AmazonS3ReadOnlyAccess	AWS managed
AWSGlueServiceRole	AWS managed

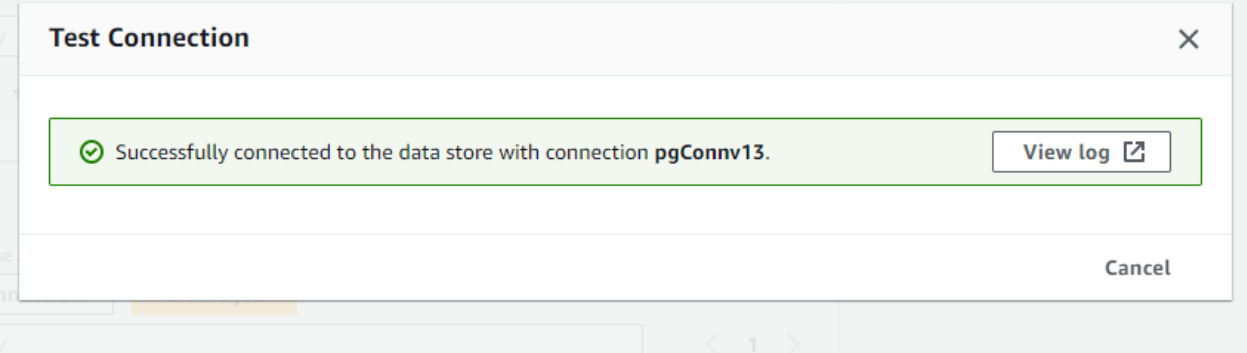
Glue Setup:

Create a Connector using db connections, security groups, VPC from earlier steps

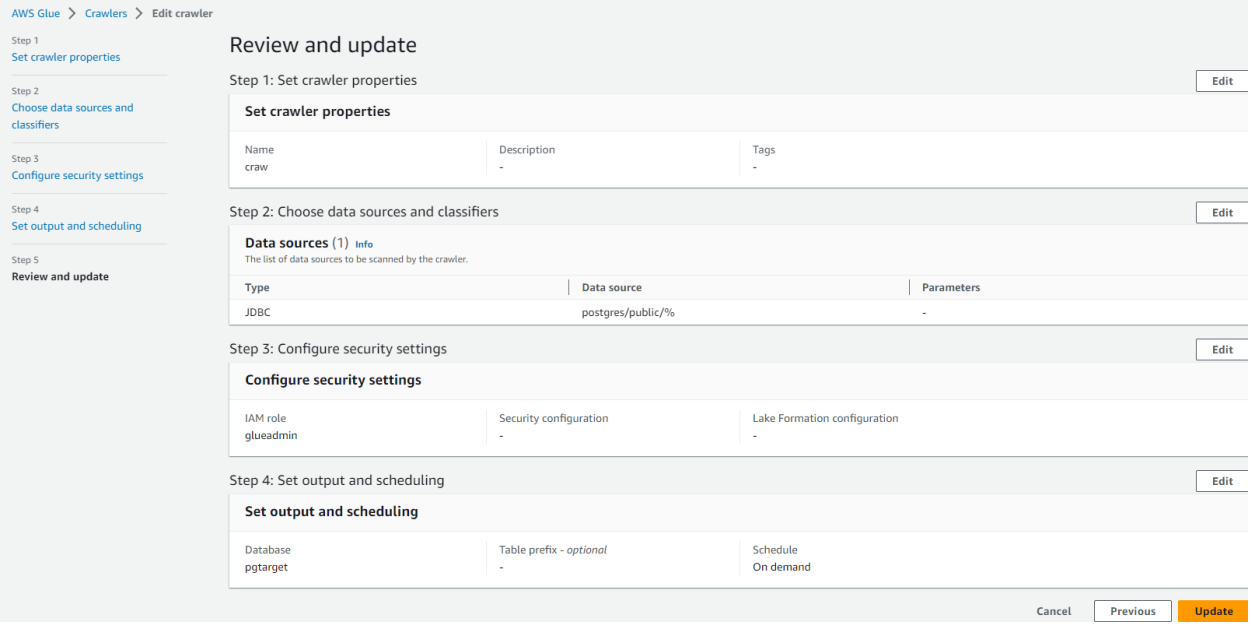


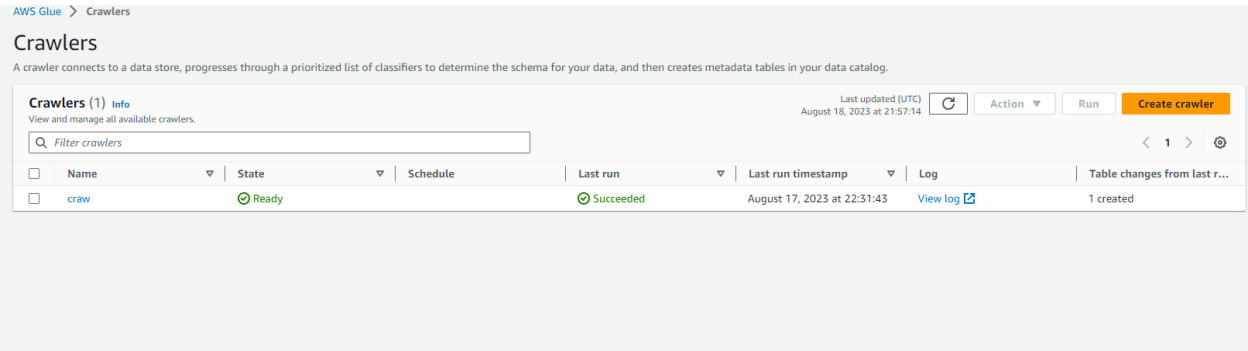
Test the connector



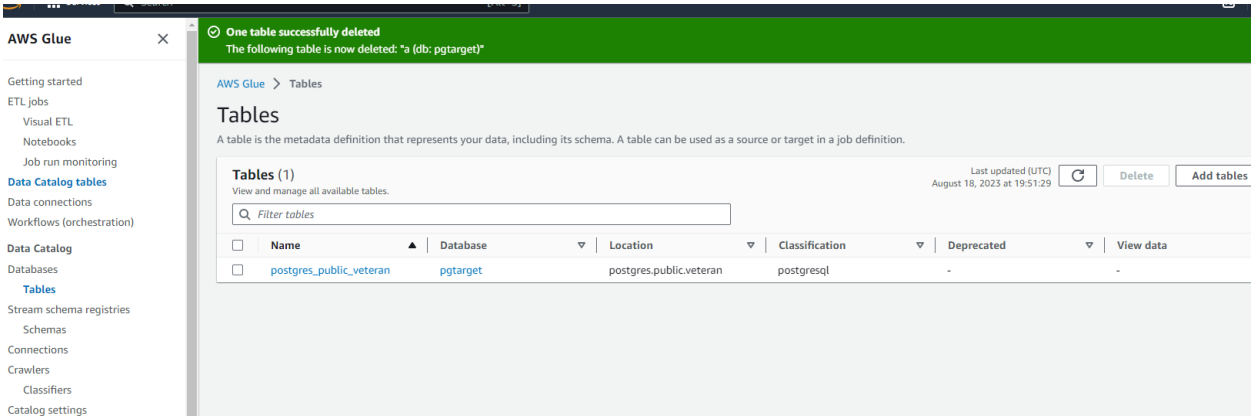


Create a Crawler using data sources from earlier steps

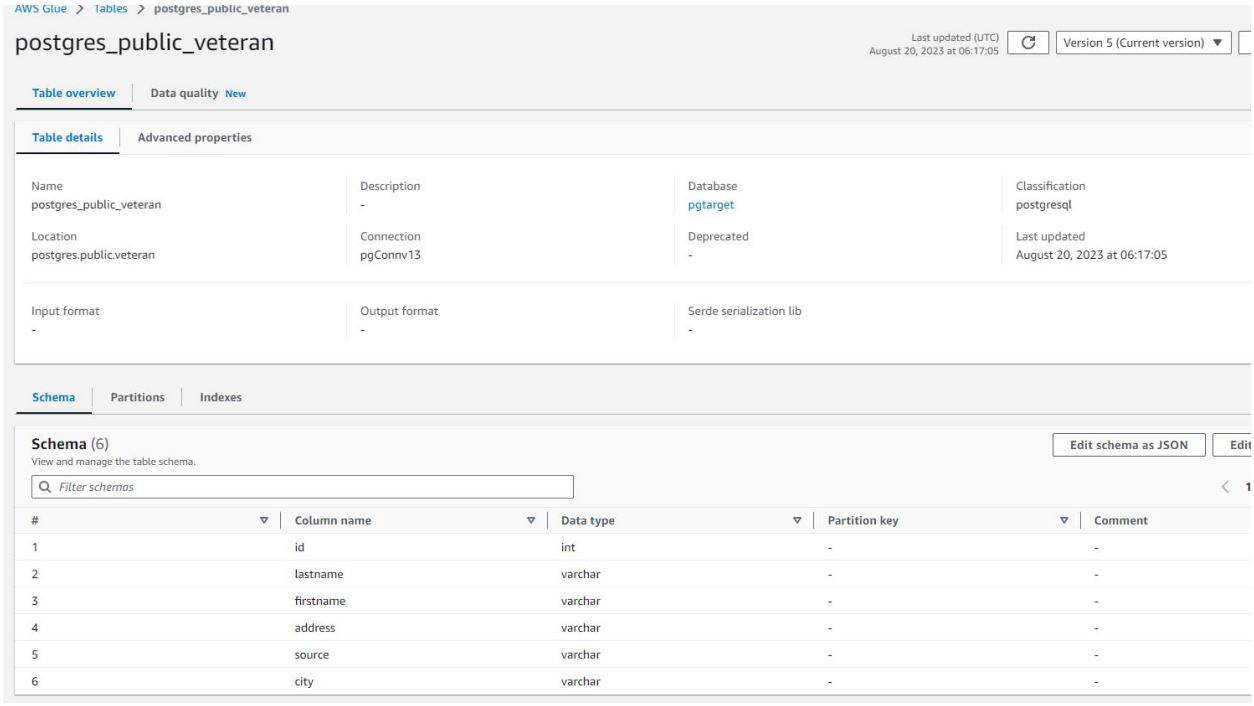




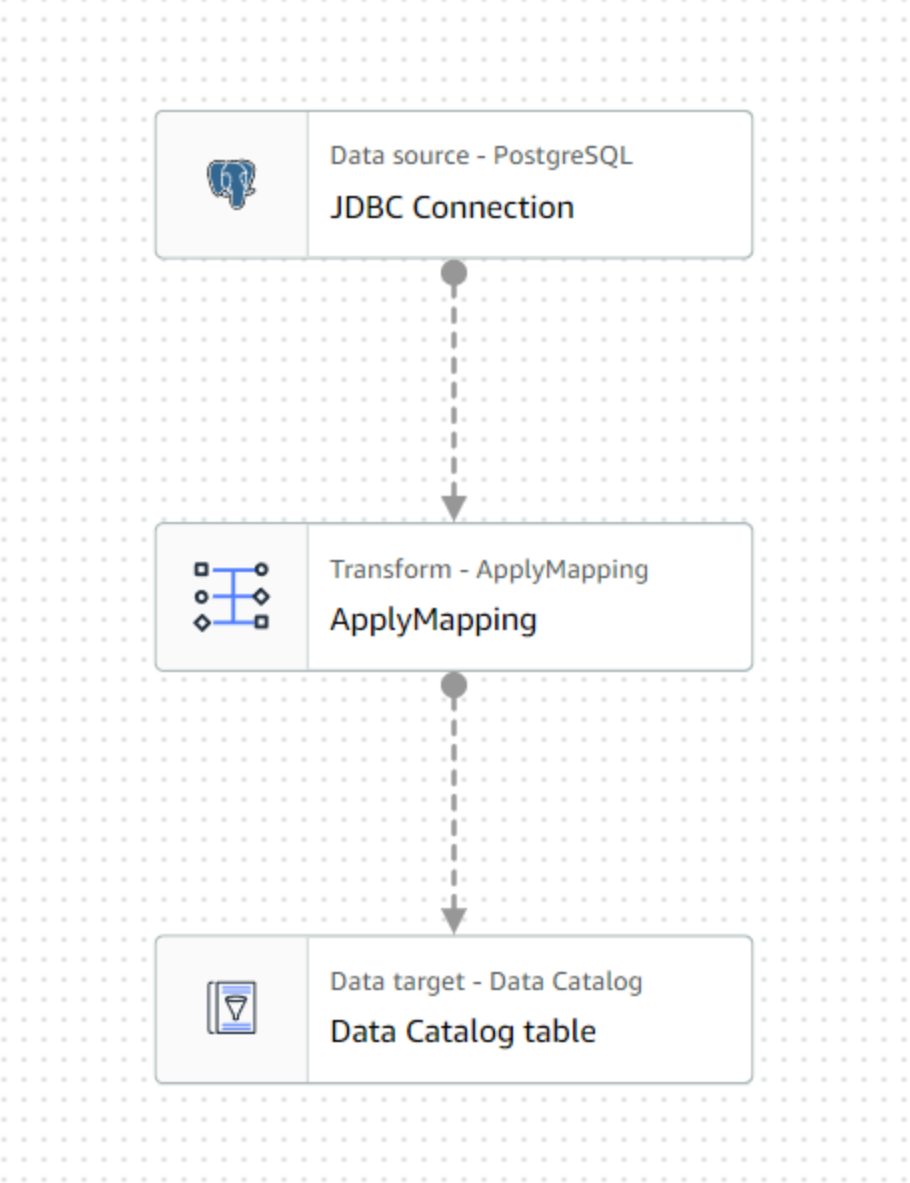
Create a Catalog, (this will mimic the CWINRS in/out tables)



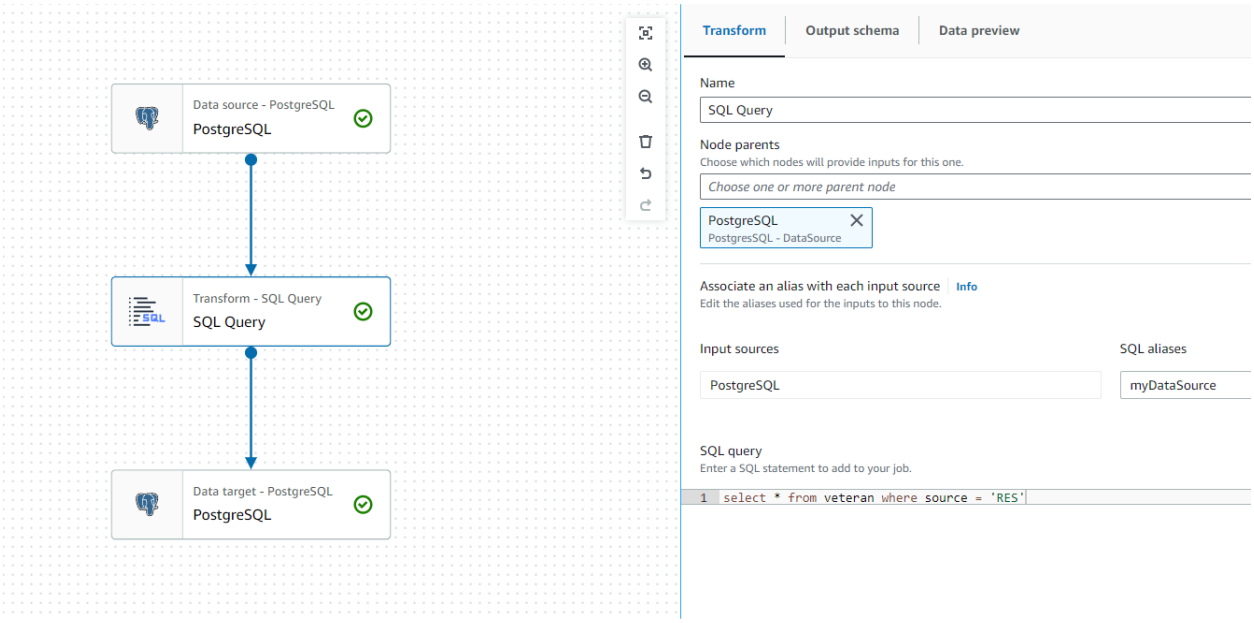
Clicking into the category table



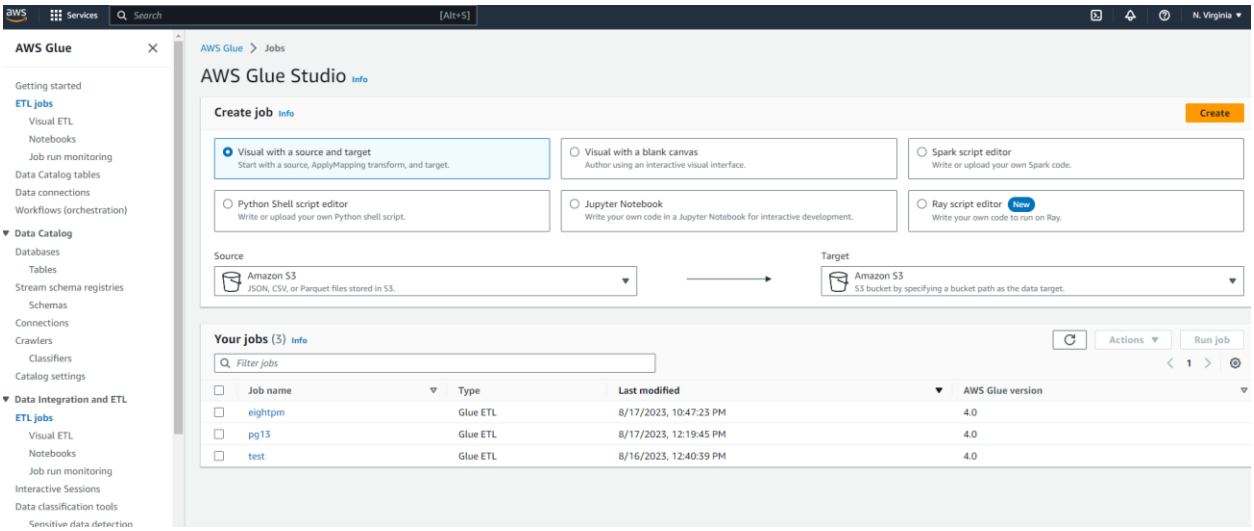
Create a ETL job using the connectors, crawlers, data catalog from earlier steps



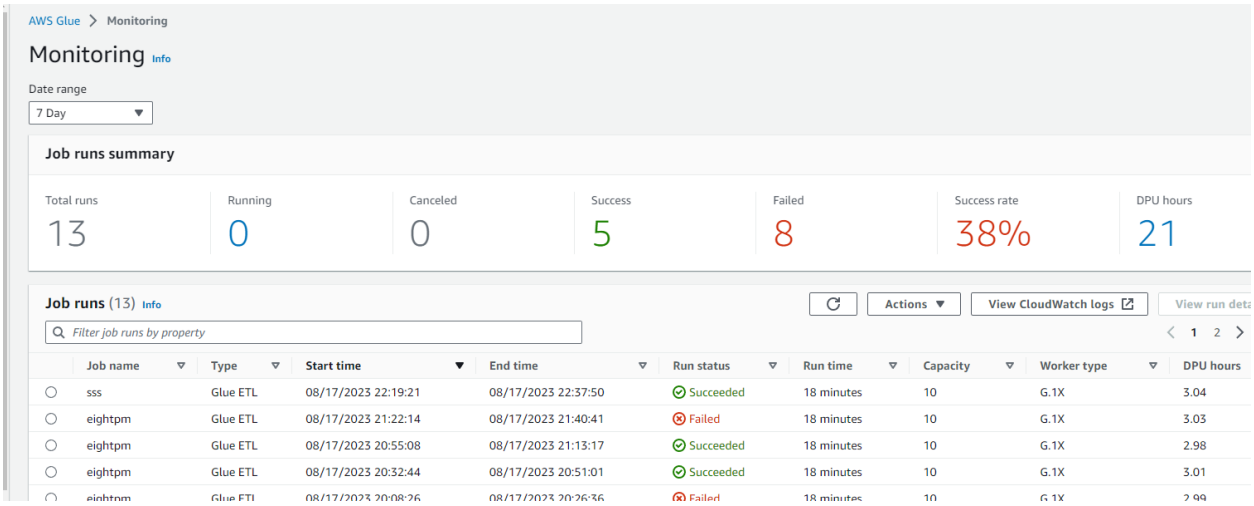
Another ETL job version where we are filtering with the Source column



ETL job list



ETL Job runs view



This demonstrates a spiked ETL job getting data from the eVA Sara Sync tables, filtering it out and sending the records into respective CWINRS/RES systems