Justin Slade

G00409678

CS 435

Fall 2024

Big Data & Cloud Security

AWS Lab Report 12

Date: 11/26/2024

Grambling State University

Big Data and Cloud Security

**Purpose or Objective:** Learn how to use AWS Step Functions to build an ETL pipeline (extract, transform, and load) the uses Amazon S3 and AWS Glue Data Catalog and Amazon Athena to process a large data set.

**Procedure/s:**

* Create Step Functions
* Create AWS Glue for database and tables
* Store data on Amazon S3
* Create an Athena view to work with Step Functions
* Create and ETL pipeline that molds the previous steps together

**Task 1.**

I will first analyze the IAMs and S3 bucket for relevant information. This is important for the following steps and needed for Step Functions as well.

A screenshot of a computer program

Description automatically generatedI checked the policy for StepLabRole and It shows I have specific permissions allowed for the ETL pipeline creation.

I will repeat this confirmation with each service. This being Athena, IAM, Step Functions, Cloud 9, Amazon S3, and AWS Glue.   
I need to copy this Bucket name for later steps (**gluelab--f95ef0a0) #note bucket name changed due to this taking multiple attempts to get working.**

When moving on to Cloud 9 to copy some commands I noticed the steps told me to copy the name of my bucket and paste it into the terminal.

A black background with white text

Description automatically generatedHowever, you don’t want to do this.  
A black background with white text

Description automatically generatedYou want this instead. This has no < > included with the name of the bucket and will cause issue when trying to do the following commands.

A black screen with white text

Description automatically generatedThis command will prompt the server to download a file into my bucket that I posted earlier.

A screenshot of a computer

Description automatically generated This is the results of the following commands. This is the files that will appear in my bucket. I don’t have access to whats inside but I know this is going good because of their exsitence.

**Task 2.**

Now I need to set up the automation of AWS Glue databases utilizing Step Functions.

When I made my way over to step functions there were a lot of steps required to set up, but I essentially paired Athena with my S3 bucket and I’m now testing the automation.

A diagram of a project

Description automatically generatedThis is the result of 5 mins of steps but this is supposed to automate the creation of Databases and store it within S3 utilizing Athena logging systems.

A screenshot of a computer

Description automatically generatedThis showing at my S3 bucket tells me that I did this correctly.

**Task 3.**

Now I need to create a function that will check to see if tables exist within the AWS Glue database.

A diagram of a project

Description automatically generatedAfter following more steps I got my setup to look like this However that part was red because I forgot to change the location to my bucket name.

A diagram of a project

Description automatically generatedThis is the result of the doing the results properly!

This meta data file lets me know things are going smoothly. A screenshot of a computer code

Description automatically generated

**Task 4.**

Now I will create routing logic that will determine if AWS Glue tables exist. (sounds similar to task 3 I know)  
so after following the steps I was met with this red box of doom. I followed the instructions properly but I’m not sure why this is an issue. I spent a lot of time working around this issue but I’m not sure how to get around this issue. A few of my other classmates have also run into this so I’m not sure what is wrong with this assignment.

A red sign with white text

Description automatically generated  
  
This was what I was able to get up to. A diagram of a work flow

Description automatically generated

**Conclusions:**

This lab was actually very involved with a lot of moving parts, and it was really cool to automate some services just from using AWS Step Functions. It was cool seeing my S3 buckets and cloud 9 communicate with each other and using Step functions to do every other task I needed to do so I did not need to do the busy work. Overall this was a cool lab, unfortunately I was not able to go further.

**Knowledge Check (12) Screenshot:**

A screenshot of a test results

Description automatically generated