**Exporting a subset of CloudTrail Lake events to an Amazon S3 bucket**

Introduction

Monitoring and managing your AWS environment is critical to maintaining security and operational excellence. With the availability of [AWS CloudTrail Lake](https://docs.aws.amazon.com/awscloudtrail/latest/userguide/cloudtrail-lake.html) data for zero-ETL analysis in Athena, you can use Athena to query your activity logs in CloudTrail Lake without the operational complexity of moving data or building data processing pipelines. CloudTrail Lake is a managed data lake that lets you aggregate, immutably store, and analyze activity logs for audit, security, and operational investigations. Athena is an interactive query service that makes it simple to analyze data in Amazon S3 and [other data stores](https://docs.aws.amazon.com/athena/latest/ug/connectors-available.html) using SQL. Using Athena, security engineers can correlate activity logs in CloudTrail Lake with application and traffic logs in data stores such as Amazon S3 for security incident investigations.

AWS CloudTrail offers a robust logging and continuous monitoring solution for account activity. Reviewing all CloudTrail log data to identify relevant events can be time-consuming. For example, consider a scenario where an organization wants to monitor changes to their IAM policies for security and compliance purposes. They can set up a CloudTrail to capture all events and then create a mechanism to export only IAM policy change events to an Amazon S3 bucket. This subset can then be monitored, analyzed and stored for compliance reporting or to ingest CloudTrail data to a third-party. Exporting a subset of CloudTrail events to an Amazon S3 bucket enables focused analysis of specific activities, streamlining event management.

In this blog post, we will walk you through the process of exporting a filtered set of CloudTrail Lake events to an Amazon S3 bucket. There may be a certain use case where you only want to ingest a subset of CloudTrail Lake events to a third-party whether you're focusing on security-related activities, compliance requirements or operational monitoring this step-by-step tutorial will help you harness the full potential of CloudTrail Lake without drowning in data. Join us as we delve into the prerequisites, configurations and the mechanism to make your AWS environment more manageable and secure.

**Unified Management**: Using native AWS services allows you to manage all aspects of the solution from a single console, simplifying operations and reducing the need for multiple management interfaces.

Overview of the Solution

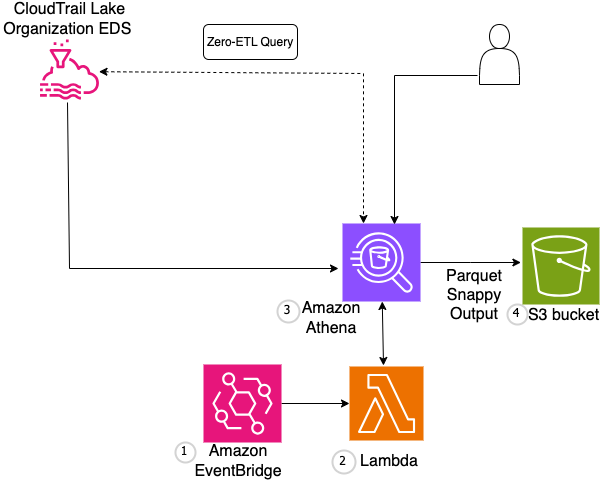


Figure 1: Export a subset of filtered events within CloudTrail Lake to an Amazon S3 bucket in parquet format

ALT Text: Export a subset of filtered events within CloudTrail Lake to an Amazon S3 bucket in parquet format

1. **EventBridge Rule**: Schedules a job to trigger the Lambda function at defined intervals.

2. **Lambda Function**:

* + Executes the Athena query.
  + Convert the results to Parquet format.
  + Save the results to the Amazon S3 bucket.

3. **Athena Database and Table**: Stores the organizational CloudTrail Lake event data.

4. **Amazon S3 Bucket**: Stores the Parquet format output from the Athena query.

**CloudFormation Template**: Defines all resources required including Athena workgroup, Athena query execution, Amazon S3 bucket, Lambda function and EventBridge rule.

We will use CloudFormation template to define the necessary AWS resources.

This CloudFormation template creates an Event Bridge scheduled event that triggers an AWS Lambda function to run an Athena query. It defines parameters for scheduling, Athena configuration and S3 bucket details. The template creates resources including an EventBridge rule, IAM role for Lambda execution with necessary permissions, and a Lambda function. The Lambda function reads a custom SQL query from an S3 bucket, combines it with a create table statement, executes the query in Athena, and stores the results in a specified S3 bucket. The function also manages execution time tracking using AWS Systems Manager Parameter Store. The template provides a comprehensive setup for automating scheduled Athena queries with proper IAM permissions and error handling.

Pre-requisites

For this walkthrough, you should have the following prerequisites:

* Create an event data store with [CloudTrail Lake query federation](https://docs.aws.amazon.com/awscloudtrail/latest/userguide/query-enable-federation.html) enabled. This option would allow the ability to query against your event data using Athena
* Create CustomQueryBucketname ahead of deploying the CFN template in Amazon S3 to store the subset data

Walkthrough

Download the CFN template & demo.sql from this [location](https://amazon.awsapps.com/workdocs-preview/index.html#/folder/f567032f2b61c0f8c3d4c14cc2abc5a886e8ae46d2132ee670c14f9bd93b7801)

The demo.sql in this example has the query to filter the logs with eventName = ‘GetBucketACL’ to export only the ‘GetBucketACL’ action performed on the Amazon S3 buckets. This can be customized to any query based on the requirement. Some example queries are mentioned at the end of the “Walkthrough” section.

1. Upload the demo.sql file to the “Custom Query Bucket name” bucket created in the pre-requisites.
2. Navigate to the CloudFormation > Stack from your AWS Console and select “Create Stack with new resources”

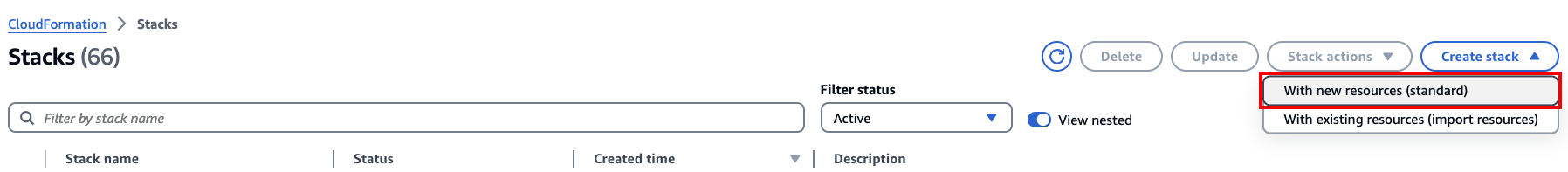


Figure 2: Create Stack with new resources

1. Put in the following parameters in the CFN template

a) “Specify Stack details” by providing the stack name

b) Account number & region for “Athena Query Output Location”

Athena Configuration

c) Athena Query Output Location

c) “Athena Database” - The default Athena database

d) Provide the “Event Data Store ID” – Get the event data store ID from the CloudTrail Lake Console

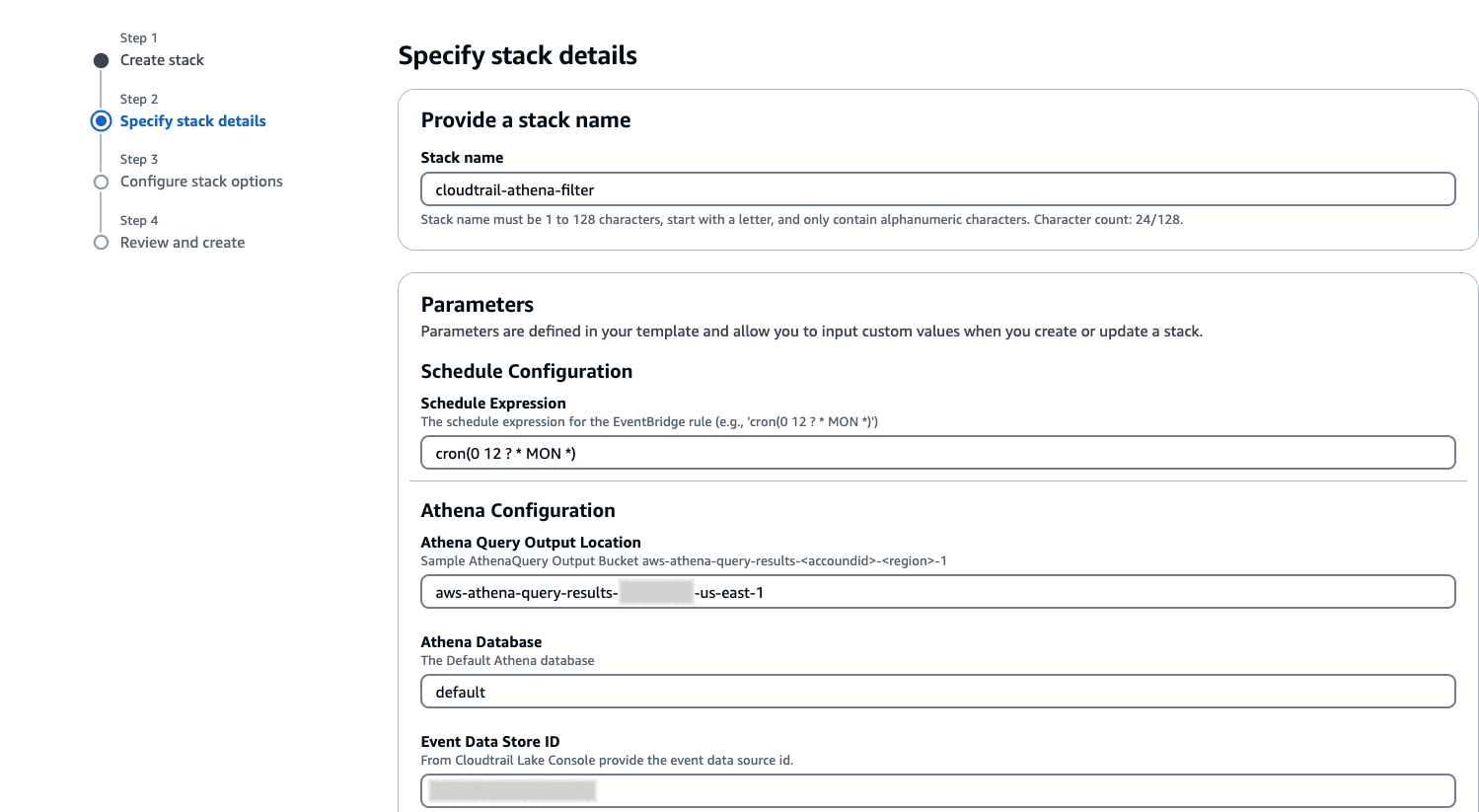
S3 Configuration

e) Fill in the “Export S3 Bucket Name” details

Provide the s3 bucket details name to store the subset of CloudTrail events to store. Sample Export S3 Bucket name, where the query results are stored

Please note this bucket is different from the “Custom Query Bucket” used for demo.sql

f) Also, the “Custom Query Bucket Name” where you’ve uploaded the demo.sql file.



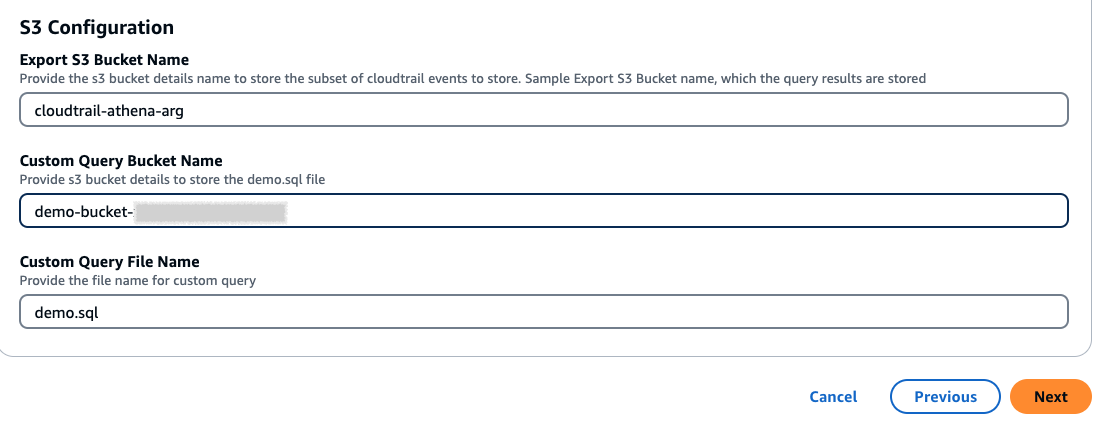
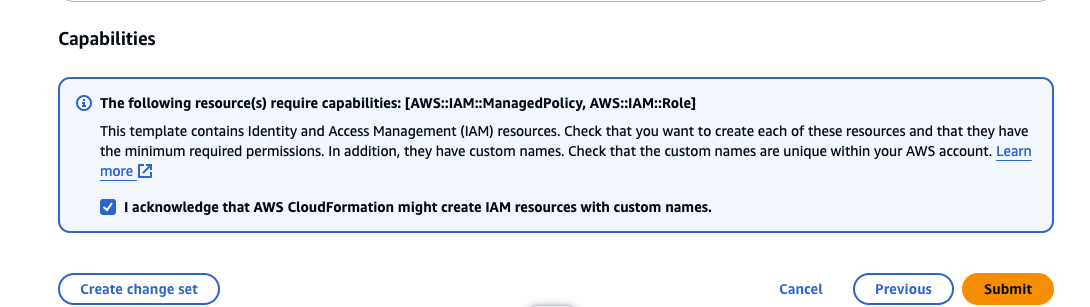


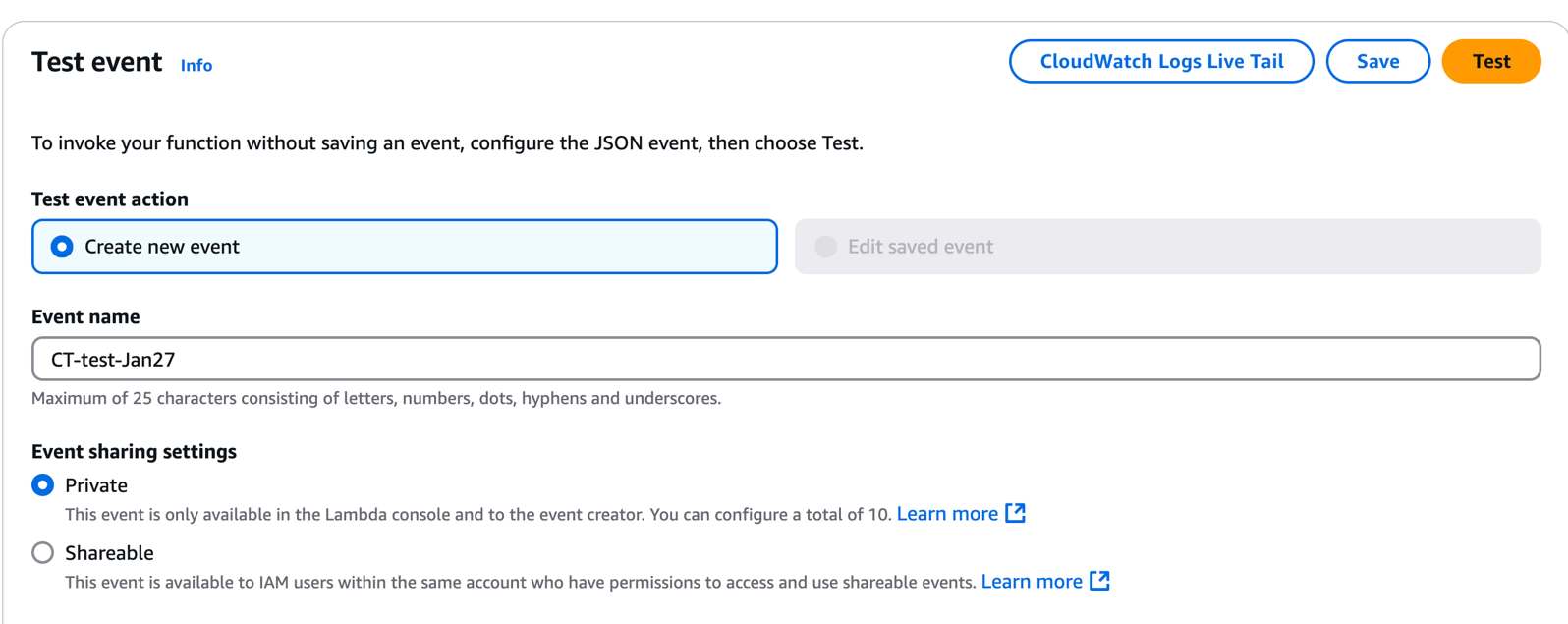
Figure 3: Parameters to be filled out

1. Leave the default settings for the “configure stack options”
2. Review and create by acknowledging the IAM resources creation

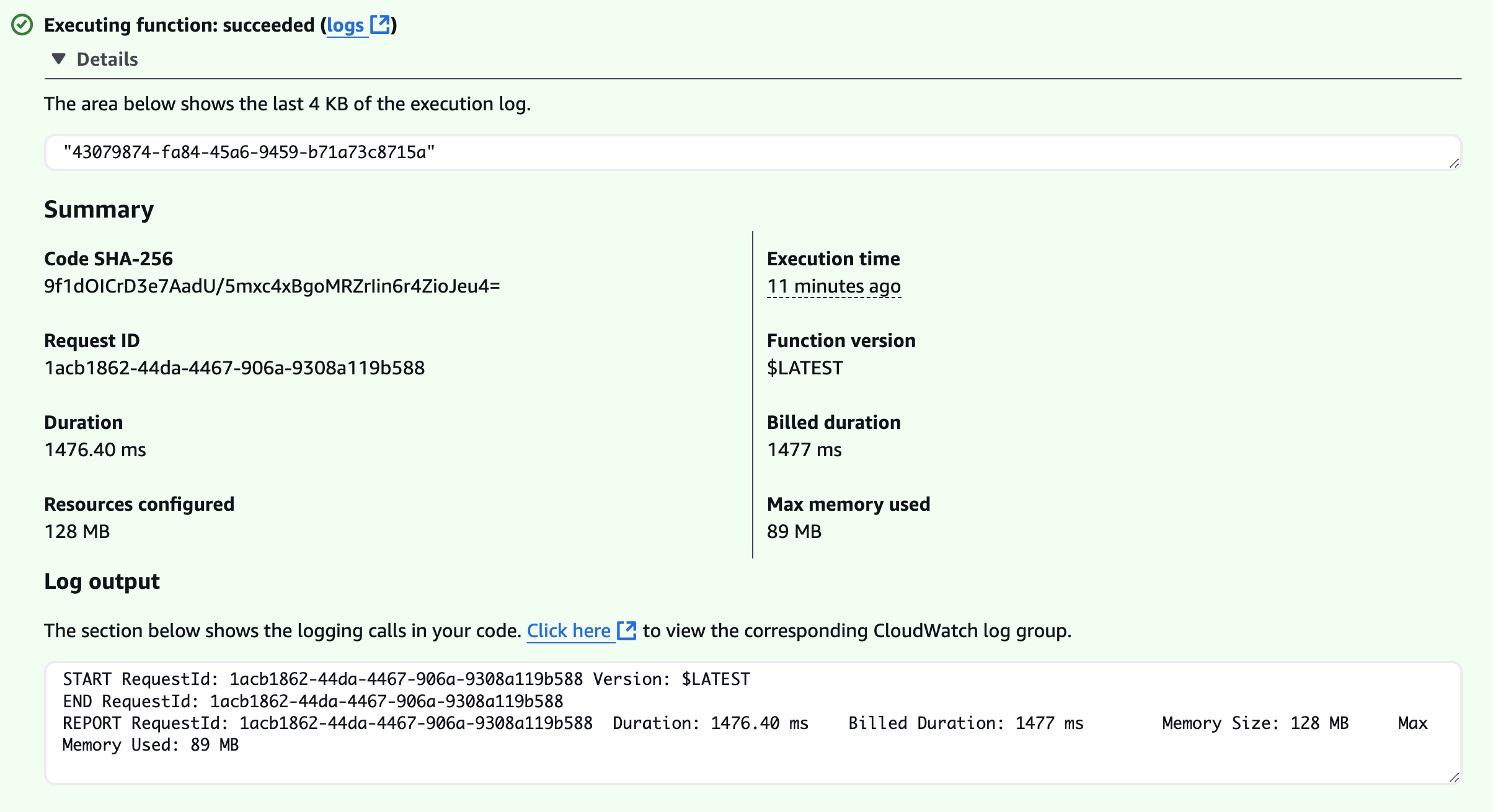


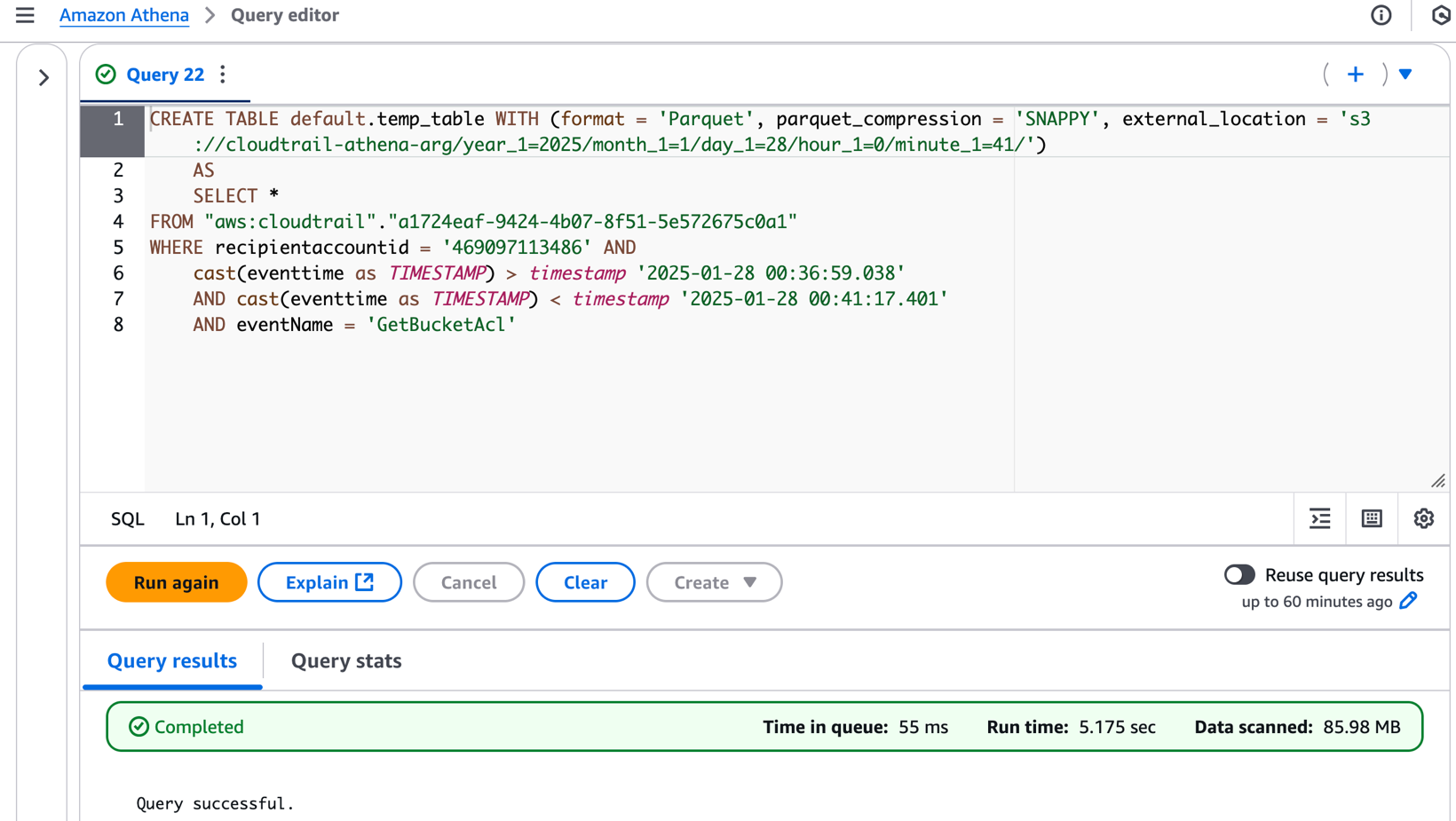
1. You should be able to see the query output from the Athena > Query Editor based on the Lambda function triggered.



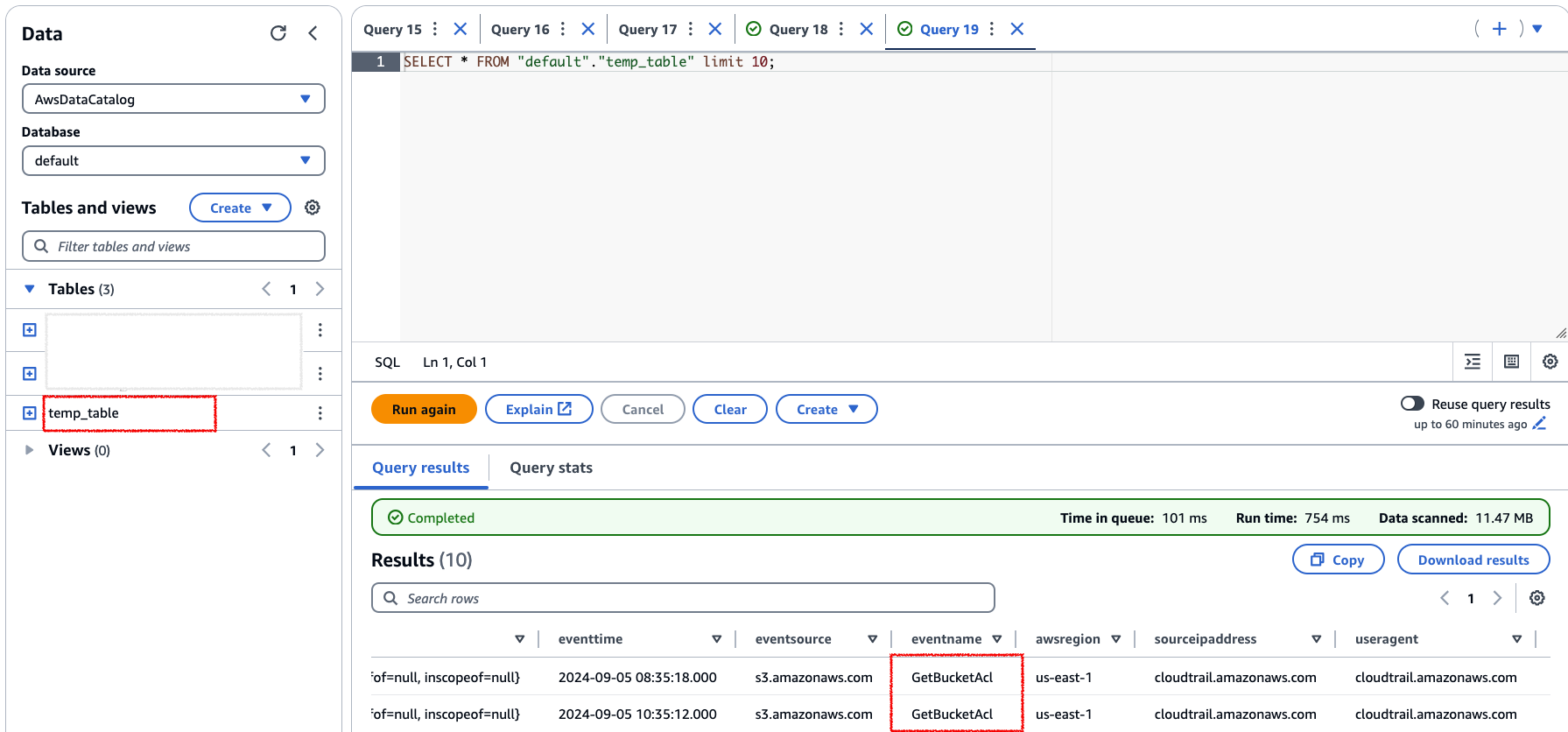
You can also test the query execution instead of waiting for the cron job to run by creating a “Test” event in the Lambda function

Capture the execution ID and check in Athena too for successful execution.





1. CloudTrail logs filtered output specific to the query given in demo.sql can be found under the temp table of the default Database as shown below. In this example we are looking at “GetBucketACL”



You can provide more SQL queries based on the data you would like to export. Few more examples to change your demo.sql file are here below.

To filter the logs only for the Event Source type “dynamodb”

SELECT \*  
FROM "aws:cloudtrail"."<EventDataStore\_ID>" WHERE recipientaccountid = '469097113486' and

cast(eventtime as TIMESTAMP) > cast('2024-06-01 12:11:22.000' as TIMESTAMP)  
and cast(eventtime as TIMESTAMP) < cast('2024-06-04 12:11:22.000' as TIMESTAMP)  
and eventSource = 'dynamodb.aws.com'

Filter all the logs “for the list of users who have turned off multi-factor authentication

SELECT userIdentity.arn,

userIdentity.userName,

userIdentity.accountId,

useridentity.principalId

FROM "aws:cloudtrail"."<EventDataStore\_ID>"

WHERE eventSource = 'iam.aws.com'

AND eventName in ('DeactivateMFADevice', 'DeleteVirtualMFADevice')

cast(eventtime as TIMESTAMP) > timestamp '{start\_time}' AND cast(eventtime as TIMESTAMP) < timestamp

GROUP BY userIdentity.arn,

userIdentity.userName,

userIdentity.accountId,

useridentity.principalId

Cleaning up

To avoid incurring future charges, you would want to delete the resources created in this demonstration, including the IAM policies, IAM Roles, CloudFormation stack, CloudTrail Lake data store.

Conclusion

In summary, this solution simplifies the task of extracting valuable insights from CloudTrail logs and enhances your ability to monitor and secure your AWS environment. By following the steps outlined in this blog post, you can set up an automated, scalable, and cost-efficient system to handle your CloudTrail Lake event analysis needs. Whether you are a cloud architect, a security professional, or a DevOps engineer, this method empowers you to gain deeper visibility and control over your AWS operations.