Managing S3 Buckets

Reza Salehi
CLOUD CONSULTANT

@zaalion linkedin.com/in/rezasalehi2008

Overview

Globomantics need to manage their S3 buckets

- Categorize, monitor, audit and minimize costs

Categorizing S3 Storage Objects Using Tagging

Introducing AWS CloudTrail

- Configure and save S3 logs
- Analyze S3 logs

Monitoring S3 metrics and cost

Demo: tags, CloudTrail and CloudWatch

Summary

Globomantics Needs to Manage S3 Buckets

Categorize S3 objects for fine-grained reporting and control of permissions

Log S3 audit information and analyze them to enable governance

Monitor bucket metrics such as “storage taken” and get event notifications

Control S3 storage cost

Categorizing S3 Storage Objects Using Tagging

Object tagging allows you
to categorize storage.
Each tag is a key-value pair.

Few Use Cases for Tagging

Tags can be used to categorize S3 buckets and objects for management purposes.

Object tags enable fine-grained permission control. For instance, you can grant a user, “read only” permissions to objects with specific tags.

You can use tags to “mark” target object for other AWS services. For example, tag your EBS volume to create automated snapshots from it using DLM.

In bucket lifecycle configuration, you can specify a filter to select a subset of objects to which the rule applies.

A Few Rules around Tags

Add tags to new objects or existing objects

A tag key can be up to 128 Unicode characters

Keys and values are case sensitive

Up to 10 tags with an object

A tag value can be up to 256 Unicode characters

Dept =accounting

Project =ODM

Demo

AWS Console

- Add tags to S3 buckets and objects
- See existing tags

AWS CLI

- Assign a new tag to S3 bucket

Configuring S3 Logs with AWS CloudTrail

AWS CloudTrail is a service to log, monitor, and retain account activity related to your AWS infrastructure.

Introducing CloudTrail

Actions taken by a user on a resource are recorded in CloudTrail

What resources were modified

who or what took which action

Actions taken in AWS Console, AWS CLI, AWS SDKs and APIs

Not to be confused with CloudWatch

Introducing CloudTrail

CloudTrail is enabled on your AWS account when you create it.

For an ongoing record of activity and events in your AWS account, create a “trail”.

CloudTrail saves logs as JSON text files in compressed gzip format in Amazon S3.

Analyzing S3 Logs with AWS Athena

Amazon Athena is an interactive query service that enables you to analyze data in Amazon S3 using standard SQL.

Introducing Amazon Athena

Athena is easy to use. Simply point to your data in Amazon S3, define the schema, and start querying using standard SQL.

Using Athena with CloudTrail logs is a powerful way to enhance your analysis of AWS service activity.

You can automatically create schemas and data tables for querying CloudTrail logs, directly from the CloudTrail console.

Make sure Amazon Athena is available in your region.

Demo

See the default CloudTrail audit logs for the account

Create a "trail" to record activity on an specific S3 bucket

Take a look at generated JSON log file by this "trail"

Use AWS Athena to analyze the collected CloudTrail logs

- Creating schema and tables
- Using SQL to query the table

Monitoring S3 Metrics Using CloudWatch

Amazon CloudWatch and S3

Amazon CloudWatch is a performance monitoring and management service.

Watch a single metric over a time period and perform actions based on the value of the metric relative to a given threshold.

Not to be confused with Amazon CloudTrail. CloudWatch is not an auditing service.

Amazon CloudWatch metrics for Amazon S3 can help you understand and improve the performance of applications that use Amazon S3.

S3 Metrics Collected by CloudWatch

Daily storage metrics for buckets

Free, once per day (*BucketSizeBytes*, *NumberOfObjects*)

Request metrics

Not free, 1 minute intervals (GetRequests, PostRequests, AllRequests, etc.)

Demo

Collecting bucket storage metrics for our S3 bucket

Getting notified when a metric value passes a threshold

Controlling S3 Cost

Three Major Costs Associated with Amazon S3

Storage cost

Charged hourly, per GB (Glacier is the cheapest)

API cost for operation of files

Charged per request, read & write requests (10 times more)

Data transfer outside of AWS region

Charged per GB

Saving S3 Cost

Pick the right AWS region for lower transfer cost

Choose by storage size and access frequency

Use lifecycle feature to move and delete data

Compress Data if possible (BMP vs. PNG)

Use One Zone-IA for non-critical data

Batch read and write requests when possible

Summary

Globomantics needs to categorize, monitor and audit S3 storage

Categorizing S3 Storage Objects Using Tagging

Introducing AWS CloudTrail to configure and analyze S3 logs

Monitoring S3 metrics

Controlling cost

Demo:

- Categorizing S3 objects using tags
- Audit logs using CloudTrail
- Monitoring using CloudWatch