Data Life-cycles

Reza Salehi
CLOUD CONSULTANT

@zaalion linkedin.com/in/rezasalehi2008

Overview

Globomantics needs to

- Keep track of S3 object versions
- Reduce storage cost for cold data
- Long time archiving for compliance

Introducing Amazon Glacier

- Data model, using Glacier

Versioning of data objects

- Enable it, fetch object versions

Use life-cycle management to save big on storage costs

Demo: Glacier, versioning, and object life-cycle

Summary

Globomantics Needs to:

Have change history for their S3 objects and recover older versions

Reduce S3 storage cost for infrequently accessed objects (cold data)

Archive old objects for years for compliance, cheapest storage possible

Using Amazon S3 Glacier to Archive and Backup Data

Amazon S3 Glacier is a secure, durable, and extremely low-cost cloud storage service for data archiving and long-term backup.

It is designed to deliver 99.999999999% durability

Security and compliance capabilities to help meet regulatory requirements

Store data cost effectively for months, years, or even decades

Offload the administrative burdens of operating and scaling storage to AWS

- Automatic capacity planning, hardware provisioning
- Automatic hardware failure detection and recovery

Working with S3 Glacier

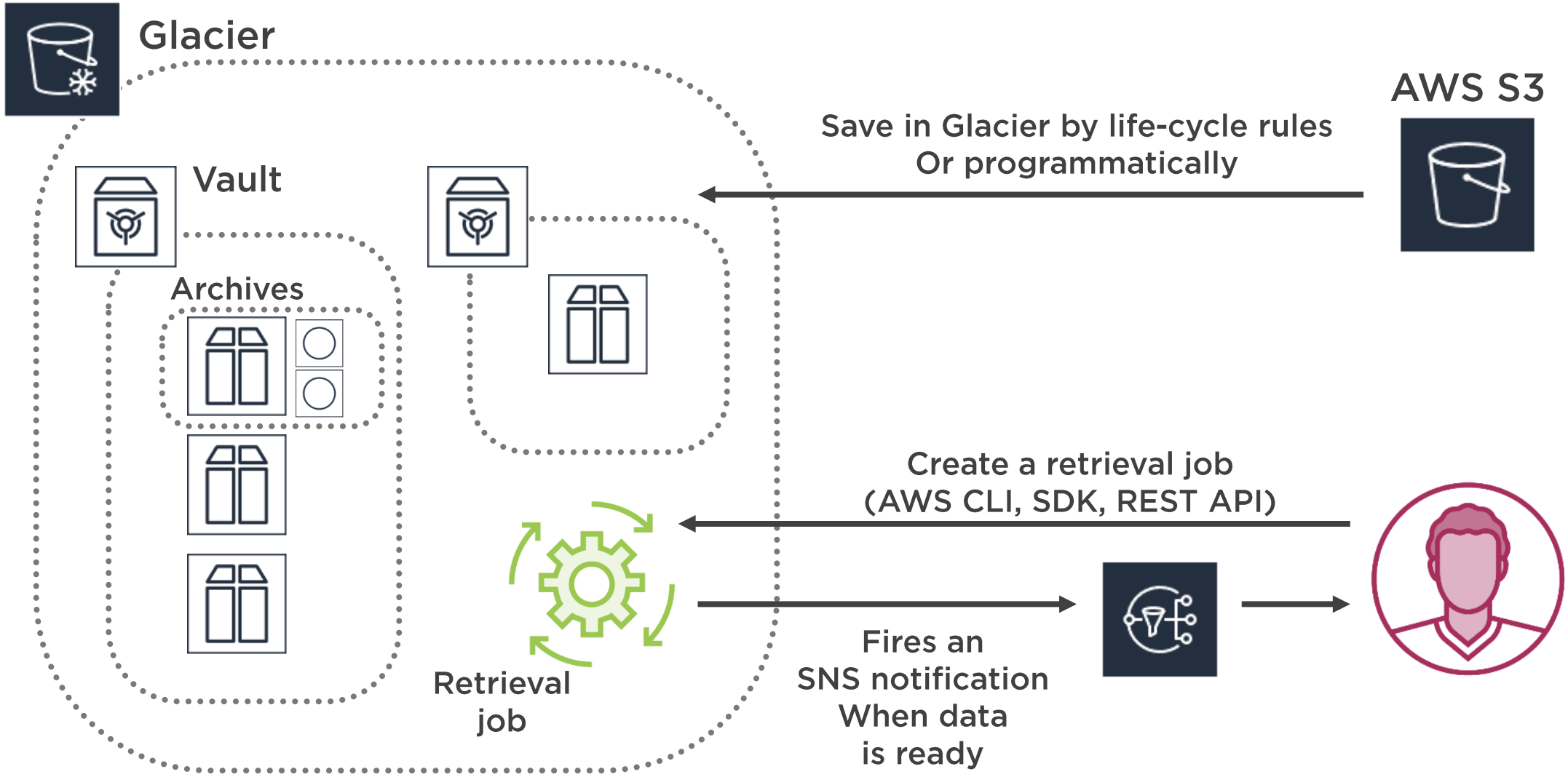
Glacier provides a console, which you can use to create and delete vaults

Other interactions with Glacier require that you use the AWS Command Line Interface (AWS CLI) or write code

To upload data, you must either use the AWS CLI or write code to make requests (REST API directly or the AWS SDKs)

Or by setting up object life-cycle rules

S3 Glacier Data Model



S3 Glacier Data Model

In Glacier, a vault is a container for storing archives

An archive can contain any data and is a base unit of storage in Glacier

Glacier jobs perform select query on an archive, retrieve an archive, or can get an inventory of a vault

Jobs take time. Configure a vault to send notification to an Amazon SNS topic when jobs complete

Vaults

In Glacier, a vault is a container for storing archives

When you create a vault, you specify a name and choose an AWS Region where you want to create the vault

Within a Region, an account must use unique vault names but can create same-named vaults in different Regions

https://<region-specific endpoint>/<account-id>/vaults/<vaultname>

Archives

An archive can be any data such as a photo, video, or document and is a base unit of storage in Glacier

Each archive has a unique ID and an optional description. Id is auto-assigned, optional description only during the upload of an archive

You can store an unlimited number of archives in a vault

https://<region-specific endpoint>/<account-id>/vaults/<vault-name>/archives/<archive-id>

Jobs

Glacier jobs perform a select query on an archive, retrieve an archive, or get an inventory of a vault

A vault can have multiple jobs in progress at any point in time

When you initiate a job, Glacier returns to you a job ID to track the job

Each job is uniquely identified by a URI, call it to get the progress status

Job Completion

Glacier supports a notification mechanism to notify you when a job is complete

You can configure a vault to send notification to an Amazon SNS topic when a job completes

You can specify one SNS topic per vault in the notification configuration

Each notification configuration resource is uniquely identified by a URI

To Retrieve an S3 Archive

Get the ID of the archive that you want to retrieve

Initiate a job to prepare an archive for subsequent download

Download the bytes after the job completes

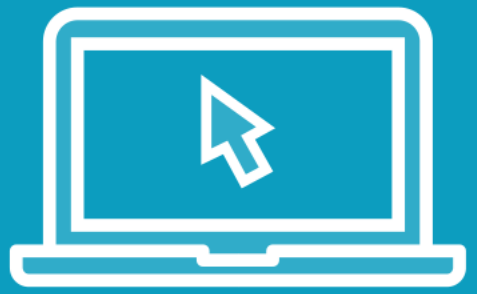
Archive Retrieval Options

Expedited retrievals allow you to quickly access your data when urgent requests for a subset of archives are required. Under 250 MB, within 1–5 minutes

Standard retrievals allow you to access any of your archives within several hours. within 3–5 hours

Bulk retrievals are Glacier's lowest-cost retrieval option, which you can use to retrieve large amounts, even petabytes of data inexpensively. within 5–12 hours

Demo



- Creating a Glacier vault (AWS Console)
- Create an archive
- Upload objects into the archive
- Create a retrieval job
- Retrieve the archive from Glacier

Enable Versioning for S3 Objects

S3 Versioning is a means of keeping multiple variants of an object in the same bucket.

Introducing S3 Versioning

Preserve and restore every version of objects

Recover from user actions and application failures

Delete an object, results in creation of a delete marker

Overwrite an object, results in a new object version in the bucket

Amazon S3 assigns each object a unique version ID

You need to explicitly enable versioning on your bucket. Once enabled, it can only be suspended.

Demo

Working with S3 versioning

- Enable versioning on a bucket
- Update an object
- Get different versions of an object using versionId
- Delete an object

Managing the Life-cycle of Data Objects

With life-cycle configuration rules, you can tell Amazon S3 to transition objects to less expensive storage classes, or archive, or delete them.

Introducing Object Life-cycle Rules

Transition actions

Transition objects to the STANDARD_IA class, 30 days after creation, or archive in GLACIER, one year after

Expiration actions

Define when objects expire. Amazon S3 deletes expired objects automatically on your behalf

Life-cycle Configuration

A lifecycle configuration is a set of rules that define actions that Amazon S3 applies to a group of objects.

Define when objects transition to another storage class or expire.

You can use AWS Console, CLI, SDK and rest API to set lifecycle rules

STANDARD_IA is cheaper for storage but more expensive when retrieving frequently.

When to Use Life-cycle Configuration

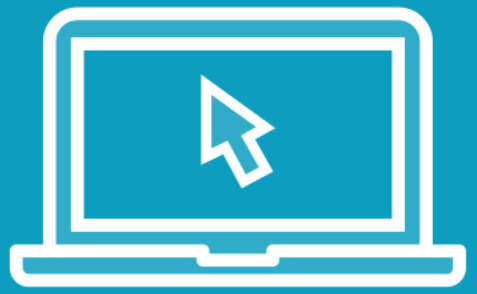
Define lifecycle configuration rules for objects that have a well-defined life-cycle

Your application needs logs for a month. After that, you might want to delete them

Documents frequently accessed for a limited time. After that, they are infrequently accessed

Some types of data are primarily kept for archival purposes (regulatory compliance)

Demo



Configure life-cycle rules for objects in an AWS S3 bucket

Summary

Globomantics needs to

- Keep track of S3 object versions
- Reduce storage cost for cold data
- Long time archiving for compliance

Amazon Glacier data model

Using Glacier

Versioning of data objects

Life-cycle management to save big on storage costs

Demo: Glacier, versioning and object life-cycle