Amazon Glacier Tutorial – Glacier Storage Service

**Amazon Glacier:**

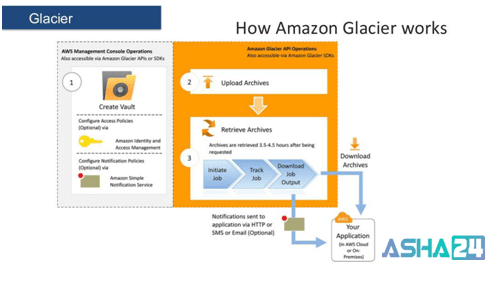
**1) Introduction to Amazon Glacier:**

**What Is Amazon Glacier?**

[Amazon Glacier](https://aws.amazon.com/documentation/glacier/) is a storage service optimized for infrequently used data, or “cold data.”

It is an extremely low-cost storage service that provides durable storage with security features for data archiving and backup. Customers can store their data cost-effectively for months, years, or even decades.

It also enables customers to reduce the administrative burdens of operating and scaling storage to AWS, so they don’t have to worry about capacity planning, hardware provisioning, data replication, hardware failure detection and recovery, or time-consuming hardware migrations.



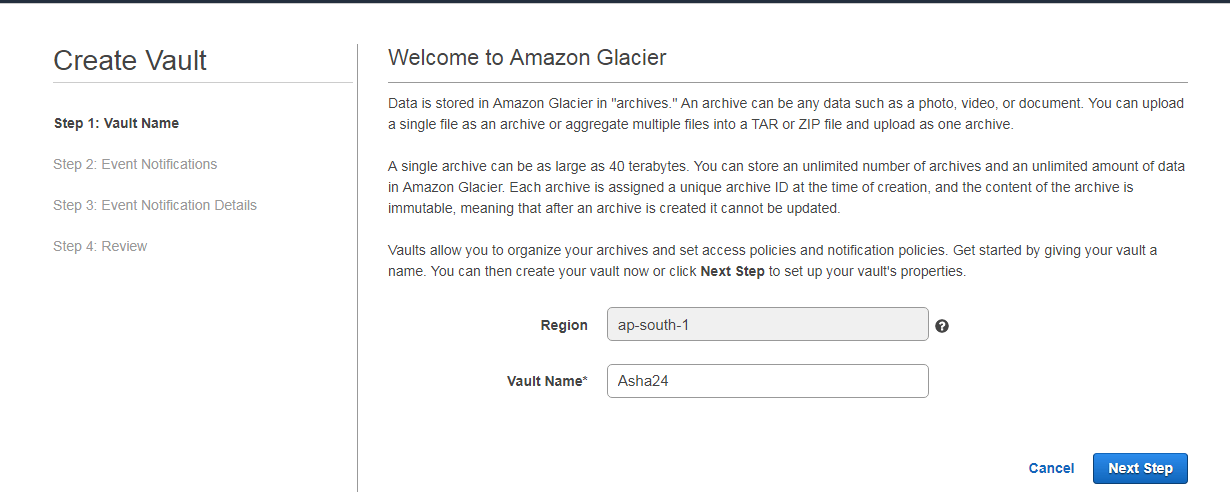
**2) Amazon Glacier Data Model:**

The Amazon Glacier data model concepts include vaults, archives, jobs and Notification Configuration. Glacier is a REST-based web service where vaults and archives are the resources.

You can create a vault in just 4 steps. This is how it can be done:-

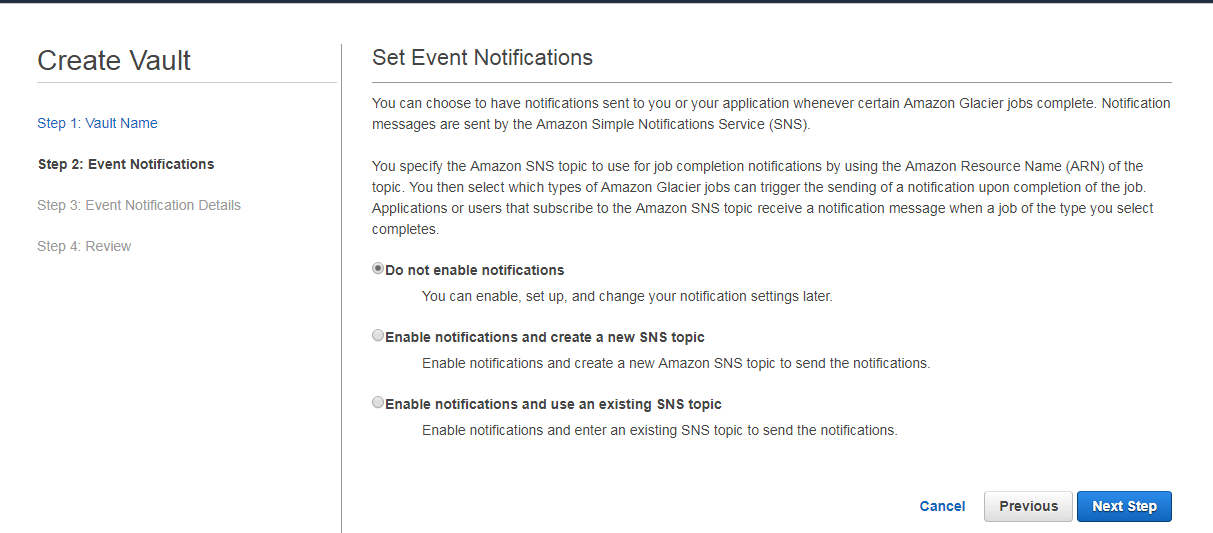
**Step-1: Vault Name**

In Amazon 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.

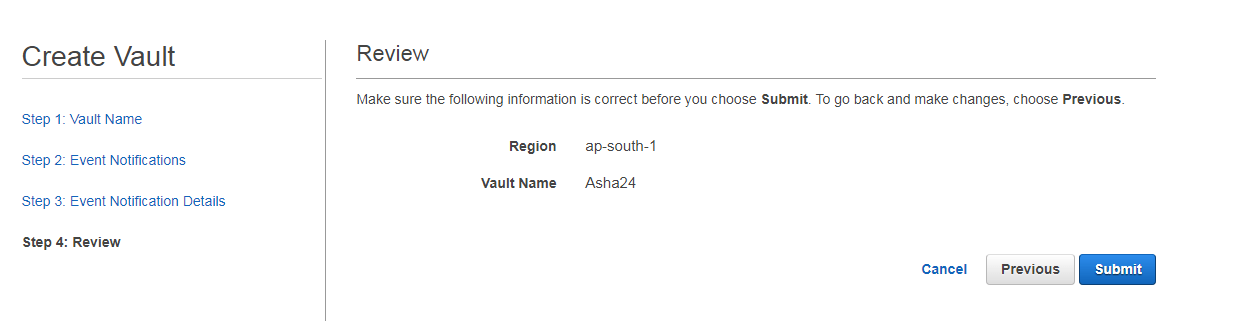


**Step-2: Event Notifications**

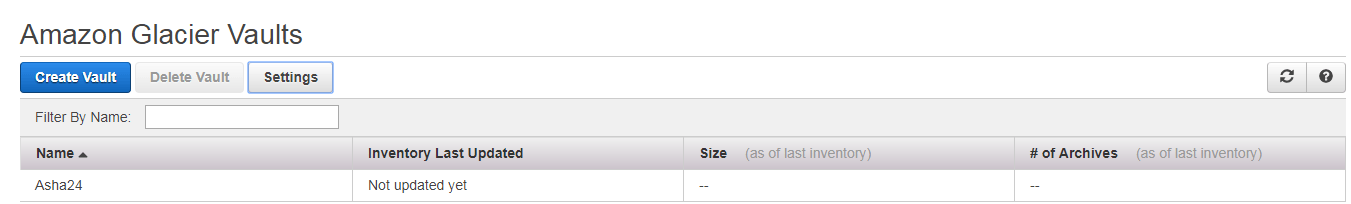
You can choose to have notifications sent to you or your application whenever certain Amazon Glacier jobs complete. Notification messages are sent by the Amazon Simple Notifications Service (SNS).



**Step-3: Event Notification Details**



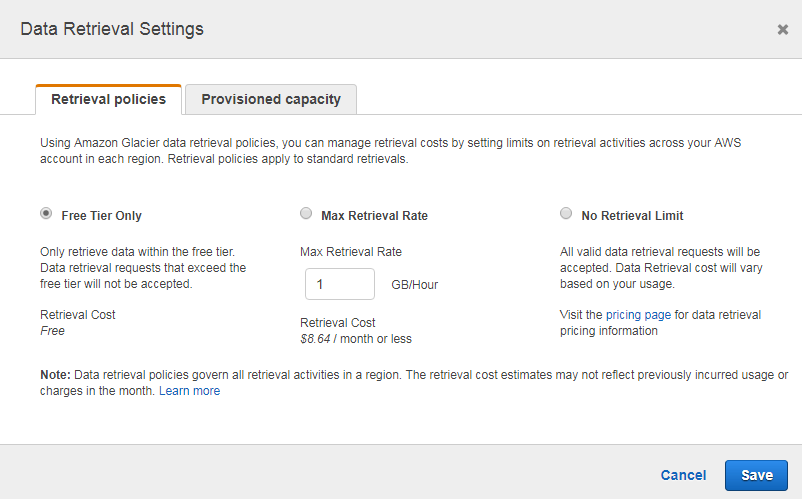
**Step-4: Review**



**After completing this process you are needed to set Up a Data Retrieval Policy.**

This is how you can go about it:

You can view and update data retrieval policies in the Amazon Glacier console or by using the Amazon Glacier API. To setup, a data retrieval policy in the console, choose an AWS region then click **Settings** .



**2.Archive**

An archive can be any kind of 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.

**Note:**

You can only specify the optional description during the upload of an archive. Amazon Glacier assigns the archive an ID, which is unique in the AWS Region in which it is stored.

**3. Jobs**

Amazon Glacier jobs can perform a select query on an archive, retrieve an archive, or get an inventory of a vault. When performing a query on an archive, you initiate a job providing a SQL query and list of Amazon Glacier archive objects. Amazon Glacier select runs the query in place and writes the output results to Amazon S3.

**Note:**

Amazon Glacier offers a cold storage data archival solution. If your application needs a storage solution that requires real-time data retrieval, you might consider using Amazon S3.

**4. Notification Configuration**

You can choose to have notifications sent to you or your application whenever certain Amazon Glacier jobs complete. Notification messages are sent by the Amazon Simple Notifications Service (SNS).

You specify the Amazon SNS topic to use for job completion notifications by using the Amazon Resource Name (ARN) of the topic. You then select which types of Amazon Glacier jobs can trigger the sending of a notification upon completion of the job. Applications or users that subscribe to the Amazon SNS topic receive a notification message when a job of the type you select completes.

Amazon Glacier supports operations to set, get, and delete a notification configuration. When you delete a notification configuration, no notifications are sent when any data retrieval operation on the vault is complete.

**3) Amazon Glacier Benefits:**

**1. Unmatched Durability, Reliability, & Scalability**

Amazon Glacier runs on the world’s largest global cloud infrastructure and was designed for cent percent durability. Data is automatically distributed across a minimum of three physical facilities that are geographically separated within an AWS Region, and Amazon Glacier can also automatically replicate data to any other AWS Region.

**2. Most Comprehensive Security & Compliance Capabilities**

Amazon Glacier also supports security standards and compliance capabilities including SEC Rule 17a-4, PCI-DSS, HIPAA/HITECH, Fed RAMP, EU Data Protection Directive, and FISMA, helping satisfy compliance requirements for virtually every regulatory agency around the globe.

**3. Query in Place**

Amazon Glacier is the only cloud archive storage service that allows you to query data in place and retrieve only the subset of data you need from within an archive. Amazon Glacier helps you reduce the total cost of ownership by extending your data lake into cost-effective archive storage.

**4. Low Cost**

Starting at a minimal $0.004 per gigabyte per month, Amazon Glacier allows you to archive large amounts of data at a very low cost. You pay only for what you need, with no minimum commitments or up-front fees.

**5. Most Supported Platform with the Largest Ecosystem**

The Amazon object storage ecosystem includes tens of thousands of consulting, systems integrators, and independent software vendor partners, with more joining every month. And the AWS Marketplace offers 35 categories and more than 3,500 software listings from over 1,100 ISVs that are pre-configured to deploy on the AWS Cloud. AWS Partner Network partners who have adapted their services and software to work with Amazon S3 and Amazon Glacier for solutions like Backup & Recovery, Archiving, and Disaster Recovery.

**4) Amazon Glacier usages:**

**Media Asset Archiving**

Media assets archiving such as video and news footage require durable storage and can grow to many pet bytes over time. Glacier allows you to archive older media content reliably and then move it to Amazon S3 for distribution when needed.

**Healthcare Information Archiving**

Hospital systems need to retain pet bytes of patient records (LIS, PACS, EHR, etc.) for decades to meet regulatory requirements. Glacier helps you reliably archive patient record data securely at a very low cost.

**Regulatory and Compliance Archiving**

Many enterprises must retain regulatory and compliance archives for extended durations. Its Vault Lock helps you set compliance controls to meet your compliance objectives, such as SEC Rule 17a-4(f).

**Scientific Data Storage**

Research organizations generate, analyze, and archive vast amounts of data. With Glacier, you avoid the complexities of hardware and facility management and capacity planning.

**Digital Preservation**

Libraries and government agencies face data-integrity challenges in their digital preservation efforts. Unlike traditional systems, which can require laborious data verification and manual repair, Glacier performs regular, systematic data integrity checks and is built to be automatically self-healing.

**Magnetic Tape Replacement**

On-premises or offsite tape libraries can lower storage costs but require large upfront investments and specialized maintenance. Amazon Glacier has no upfront cost and eliminates the cost and burden of maintenance.

**5) Amazon Glacier Key Features:**

**Data Retrieval Features**

It provides three retrieval features for your archives to meet varying access time and cost requirements: Expedited, Standard, and Bulk retrievals.

**Amazon Glacier Select**

It allows queries to run directly on data stored in Amazon Glacier without having to retrieve the entire archive. Amazon Glacier Select changes the value of archive storage by allowing you to process and find only the bytes you need out of the archive to use for analytics.

**AWS Snowball and Direct Connect Integration**

AWS Snowball can accelerate moving large amounts of data into and out of AWS using portable storages devices for transport. AWS transfers your data directly onto and off of storage devices using Amazon’s high-speed internal network and bypassing the Internet.

**Vault Lock**

Glacier’s Vault Lock allows you to easily deploy and enforce compliance controls on individual vaults via a lockable policy.

**Access Control**

It uses AWS Identity and Access Management (IAM) to help you securely control access to AWS and your Amazon Glacier data.

**Tagging Support**

It allows you to tag your vaults for easier resource and cost management. Tags are labels that you can define and associate with your vaults, and using tags adds filtering capabilities to operations such as AWS cost reports.

**Audit Logs**

It supports audit logging with AWS Cloud Trail, which records Amazon Glacier API calls for your account and delivers these log files to you.

**Integrated lifecycle management with Amazon S3**

Amazon Glacier works together with Amazon S3 lifecycle rules to help you automate archiving of Amazon S3 data and reduce your overall storage costs.