AMAZON S3

By VIGNESH

Object-based storage (OBS)

is a data storage architecture that manages data as objects, rather than as blocks within a traditional file system.

Each object typically consists of the data itself, metadata, and a unique identifier.

This architecture is often used in cloud storage and distributed storage systems due to its scalability, ease of management, and suitability for large-scale data environments.

Amazon S3 (Simple Storage Service): Amazon S3 is a widely used object storage service provided by Amazon Web Services (AWS).

Microsoft Azure Blob Storage: Azure Blob Storage is Microsoft's object storage solution within the Azure cloud platform.

Google Cloud Storage: Google Cloud Storage is a part of the Goog

s3

Amazon S3 is a cloud or object storage service, started in 2006 as a first service.

By the end of 2012, 1.3 trillion objects were stored in Amazon S3, the world's largest and most widely known object storage system. Now, that number was growing faster, so the 2 trillion mark is right around the corner.

S3 is a Global service.

S3 enables a customer to upload, store and download practically any file or object .

Amazon Simple Storage Service (Amazon S3)

An object storage service that offers industry-leading scalability, data availability, security, and performance.

Amazon S3 provides management features so that you can optimize, organize, and configure access to your data to meet your specific business, organizational, and compliance requirements.

Scalability, Durability and Availability, Global availability, Data Protection, Data Management, Access Control, Data Transfer Acceleration, Storage Classes

Use Cases:

Backup and Restore: S3 is commonly used for backing up critical data and enabling disaster recovery strategies.

Data Archiving: The Glacier storage class is suitable for long-term data archival at a lower cost.

Content Distribution: S3 can be used to store and serve static website content, images, videos, and other media files.

Data Lakes: S3 is often used as a storage layer for building data lakes, where large volumes of structured and unstructured data are stored for analytics and data processing.

Use Cases :

Big Data Analytics: S3 can be integrated with various AWS services, such as Amazon EMR and Amazon Redshift, to store and analyze large datasets.

Application Data Storage: Many applications use S3 as a reliable and scalable storage solution for user-generated content, such as profile images and document uploads.

Media Storage and Distribution: S3 is used by media and entertainment companies to store and distribute digital assets like videos, music, and multimedia content.

STORAGE CLASSES :

Standard: This is the default storage class and offers high durability, availability, and performance. It's designed for frequently accessed data.

Intelligent-Tiering: This storage class uses machine learning to automatically move objects between two access tiers: frequent and infrequent access.

One Zone-Infrequent Access (IA): Data stored in this class is stored in a single availability zone, providing a lower-cost option for infrequently accessed data. However, it doesn't provide the same level of durability as the Standard storage class.

STORAGE CLASSES :

Glacier: Glacier is designed for long-term archival of data. It offers very low-cost storage but with longer retrieval times (minutes to hours).

Glacier Deep Archive: This is the lowest-cost storage option in S3 and is intended for data that you rarely need to access. Retrieval times are longer (12 hours or more), making it suitable for true archival data.

Outposts: This storage class is used with AWS Outposts, which extends AWS infrastructure to on-premises locations. It provides object storage for data residing on-premises.

BUCKET:

Think of a bucket as a logical unit for storing objects, which are the individual files or pieces of data you upload to S3.

Instead of organizing files in a directory hierarchy it stores files in flat organization of containers called “Buckets”. It stores data, meta-data.

There is no limit for objects stored in the bucket but each account can only create 100 buckets.

Bucket Naming: Bucket names must be globally unique across all of Amazon S3.

Region: When you create a bucket, you specify a region where the bucket will be physically located. This region is important because it affects data latency and compliance. Objects stored in a specific region do not automatically replicate to other regions.

Access Control: You can control access to your S3 buckets and objects using bucket policies, IAM (Identity and Access Management) policies, and access control lists (ACLs). This allows you to define who can upload, download, and manage objects within the bucket.

Versioning: Versioning allows you to keep multiple versions of an object in a bucket. This is useful for data protection and recovery scenarios.

Static Website Hosting: S3 buckets can be used to host static websites by enabling the static website hosting feature. This is a cost-effective way to host simple websites and web applications.

Versioning:

S3 Versioning is a feature within S3 that allows you to keep multiple versions of an object (file) in a single S3 bucket. This feature provides a way to preserve, retrieve, and manage different versions of objects over time.

When you enable versioning on an S3 bucket, every object (file) you upload to that bucket will automatically have a unique version ID assigned to it. This version ID distinguishes different iterations of the same object.

Whenever you upload a new version of an object to a versioned bucket, S3 retains the existing version and assigns a new version ID to the new object. This way, you have access to the complete history of changes made to an object over time.

Deletion and Lifecycle Policies: When you delete an object from a versioned bucket, S3 does not actually remove the object. Instead, it marks the latest version as a "delete marker." If you try to retrieve the deleted object, S3 returns the most recent non-deleted version. You can also configure S3 Lifecycle policies to automatically transition older versions to different storage classes or delete them after a certain period.

Cost Implications: It's important to note that S3 Versioning can impact storage costs, as each version of an object consumes storage space. Storing multiple versions of the same object can increase the storage usage, so you need to consider the cost implications when using this feature.

Data Consistency: When you retrieve an object without specifying a version, S3 returns the latest version. If you request a specific version, S3 returns that exact version, even if newer versions exist.

Cross Region Replication

Cross-Region Replication (CRR) is a feature provided by Amazon S3 (Simple Storage Service) that allows you to automatically and asynchronously replicate objects from one S3 bucket in one AWS region to another S3 bucket in a different AWS region. This feature is particularly useful for improving data durability, availability, and disaster recovery.

Source and Destination Buckets: You'll have two S3 buckets involved—the source bucket in one AWS region and the destination bucket in another AWS region. The source bucket contains the objects you want to replicate, while the destination bucket will receive the replicated copies.

Configuration: To set up Cross-Region Replication, you need to configure replication rules on the source bucket. These rules specify which objects to replicate, the destination bucket, and other options such as encryption settings, storage class for replicated objects, and more.

Cross-Account Replication: You can also set up Cross-Region Replication between AWS accounts. This might be useful for scenarios where you want to replicate data from a source account to a destination account in a different region.

Event Notifications

Event notifications allow you to set up automated triggers or notifications in response to specific events that occur within an S3 bucket.

These events could include object creation, deletion, or other changes. When a specified event occurs, S3 can send a notification to an Amazon Simple Notification Service (SNS) topic, an AWS Lambda function, or an Amazon Simple Queue Service (SQS) queue.

Common use cases for event notifications include:

- Automatically triggering data processing tasks, such as image resizing or transcoding, using AWS Lambda.
- Notifying relevant parties or systems when new objects are uploaded or when objects are deleted.
- Keeping track of changes in your S3 bucket for auditing and compliance purposes.

Lifecycle Policies

Lifecycle policies allow you to define rules for transitioning objects between different storage classes or deleting objects after a specified period of time.

This helps you optimize storage costs by ensuring that objects are stored in the most cost-effective storage class based on their access patterns and retention requirements.

A lifecycle policy consists of rules that define the actions to be taken on objects based on certain criteria, such as object age or size.

- **You can create a rule that transitions objects older than 30 days to the S3 Glacier storage class for cost savings.**

- **You can create another rule that permanently deletes objects that are older than a year.**

The available storage classes for transitioning objects include S3 Standard, S3 Intelligent-Tiering, S3 Glacier, and S3 Glacier Deep Archive. Lifecycle policies help you ensure that your data is stored efficiently and that you're not overpaying for storage that isn't frequently accessed.