S3

Simple Storage Service

INTRODUCTION

S3, which stands for Simple Storage Service, is an object storage service provided by Amazon Web Services (AWS). It is designed to store and retrieve large amounts of data from anywhere on the web. S3 is widely used by individuals, businesses, and organizations for various purposes, including data backup, content storage, application hosting, and data archiving.

Key Features of S3

Scalability: S3 is built to handle any amount of data, ranging from a few gigabytes to petabytes or more. It automatically scales to meet the demands of applications and users.

Durability: S3 provides high durability for stored data. It replicates data across multiple facilities and automatically detects and repairs any data inconsistencies. AWS guarantees 99.999999999% (11 nines) durability for S3 objects.

Availability: S3 offers high availability, ensuring that stored data is accessible when needed. It provides a service level agreement (SLA) for uptime and is designed to provide consistent and low-latency performance.

Security: S3 offers robust security features to protect stored data. It supports encryption at rest, allowing you to encrypt data using AWS Key Management Service (KMS) keys. Access to S3 buckets and objects can be controlled using AWS Identity and Access Management (IAM) policies and bucket policies.

Storage Classes

S3 Standard:

- Designed for frequently accessed data.
- Offers high performance, low latency, and high throughput.
- Ideal for dynamic website content, mobile applications, and real-time analytics.
- Provides high availability and durability.

S3 Intelligent-Tiering:

- Automatically optimizes costs by moving data between frequent and infrequent access tiers.
- Uses machine learning to analyze access patterns and adjusts storage automatically.
- Suitable for workloads with unpredictable or changing access patterns.
- Ensures frequently accessed data remains in the frequent access tier.

Storage Classes

S3 Standard-IA (Infrequent Access):

- Designed for data that is accessed less frequently but requires rapid access when needed.
- Offers lower storage costs compared to S3 Standard.
- Slightly higher retrieval cost than S3 Standard.
- Suitable for backups, long-term storage, and disaster recovery scenarios.

S3 One Zone-IA:

- Similar to S3 Standard-IA but stores data in a single availability zone.
- Provides cost savings compared to S3 Standard-IA.
- Less resilient to availability zone failures.
- Suitable for secondary backups, easily reproducible data, and infrequently accessed data that can be recreated if lost.

Storage Classes

S3 Glacier:

- Archival storage class for long-term data retention.
- Secure, scalable, and low-cost storage.
- Retrieval times range from minutes to hours.
- Suitable for data that is rarely accessed but requires long-term retention.

S3 Glacier Deep Archive:

- Most cost-effective storage class in S3.
- Designed for long-term archival data.
- Retrieval times range from hours to days.
- Suitable for data that is seldom accessed and has strict budget constraints.

Components of S3

- **Bucket:** A bucket is a container for objects stored in Amazon S3. You can store any number of objects in a bucket and can have up to 100 buckets in your account. Every object is contained in a bucket. For example, if the object named photos/puppy.jpg is stored in the DOC-EXAMPLE-BUCKET bucket in the US West (Oregon) Region, then it is addressable by using the URL https://DOC-EXAMPLE-BUCKET.s3.us-west-2.amazonaws.com/photos/puppy.jpg.
- **Objects:** Objects are the fundamental entities stored in Amazon S3. Objects consist of object data and metadata. An object is uniquely identified within a bucket by a key (name) and a version ID (if S3 Versioning is enabled on the bucket).
- **S3 Versioning:** You can use S3 Versioning to keep multiple variants of an object in the same bucket. With S3 Versioning, you can preserve, retrieve, and restore every version of every object stored in your buckets.
- **Bucket Policy:** A bucket policy is a resource-based AWS Identity and Access Management (IAM) policy that you can use to grant access permissions to your bucket and the objects in it. Only the bucket owner can associate a policy with a bucket. The permissions attached to the bucket apply to all of the objects in the bucket that are owned by the bucket owner.

Components of S3

- **S3 Access Points:** Amazon S3 Access Points are named network endpoints with dedicated access policies that describe how data can be accessed using that endpoint. Access Points are attached to buckets that you can use to perform S3 object operations, such as GetObject and PutObject. Access Points simplify managing data access at scale for shared datasets in Amazon S3.
- **Access Control List:** You can use ACLs to grant read and write permissions to authorized users for individual buckets and objects. Each bucket and object has an ACL attached to it as a subresource. The ACL defines which AWS accounts or groups are granted access and the type of access. ACLs are an access control mechanism that predates IAM.
- **Regions:** You can choose the geographical AWS Region where Amazon S3 stores the buckets that you create. You might choose a Region to optimize latency, minimize costs, or address regulatory requirements. Objects stored in an AWS Region never leave the Region unless you explicitly transfer or replicate them to another Region. For example, objects stored in the Europe (Ireland) Region never leave it.

Accessing AWS S3

- AWS Management Console: The console is a web-based user interface for managing Amazon S3 and AWS resources. If you've signed up for an AWS account, you can access the Amazon S3 console by signing into the AWS Management Console and choosing S3 from the AWS Management Console home page.
- AWS CLI: You can use the AWS command line tools to issue commands or build scripts at your system's command line to perform AWS (including S3) tasks.The AWS Command Line Interface (AWS CLI) provides commands for a broad set of AWS services. The AWS CLI is supported on Windows, macOS, and Linux.
- AWS SDK: AWS provides SDKs (software development kits) that consist of libraries and sample code for various programming languages and platforms (Java, Python, Ruby, .NET, iOS, Android, and so on). The AWS SDKs provide a convenient way to create programmatic access to S3 and AWS. Amazon S3 is a REST service. You can send requests to Amazon S3 using the AWS SDK libraries, which wrap the underlying Amazon S3 REST API and simplify your programming tasks. For example, the SDKs take care of tasks such as calculating signatures, cryptographically signing requests, managing errors, and retrying requests automatically. For information about the AWS SDKs, including how to download and install them

Storage Logging & Monitoring

- **Automated monitoring tools:**

Amazon CloudWatch metrics for Amazon S3 – Track the operational health of your S3 resources and configure billing alerts when estimated charges reach a user-defined threshold.

AWS CloudTrail – Record actions taken by a user, a role, or an AWS service in Amazon S3. CloudTrail logs provide you with detailed API tracking for S3 bucket-level and object-level operations.

Manual monitoring tools

Server access logging – Get detailed records for the requests that are made to a bucket. You can use server access logs for many use cases, such as conducting security and access audits, learning about your customer base, and understanding your Amazon S3 bill.

AWS Trusted Advisor – Evaluate your account by using AWS best practice checks to identify ways to optimize your AWS infrastructure, improve security and performance, reduce costs, and monitor service quotas. You can then follow the recommendations to optimize your services and resources.

Access Management & Security

- Amazon S3 provides features for auditing and managing access to your buckets and objects. By default, S3 buckets and the objects in them are private. You have access only to the S3 resources that you create. To grant granular resource permissions that support your specific use case or to audit the permissions of your Amazon S3 resources, you can use the following features.
- S3 Block Public Access – Block public access to S3 buckets and objects. By default, Block Public Access settings are turned on at the bucket level. We recommend that you keep all Block Public Access settings enabled unless you know that you need to turn off one or more of them for your specific use case. For more information, see Configuring block public access settings for your S3 buckets.
- AWS Identity and Access Management (IAM) – IAM is a web service that helps you securely control access to AWS resources, including your Amazon S3 resources. With IAM, you can centrally manage permissions that control which AWS resources users can access. You use IAM to control who is authenticated (signed in) and authorized (has permissions) to use resources.
- Bucket policies – Use IAM-based policy language to configure resource-based permissions for your S3 buckets and the objects in them.
- Amazon S3 access points – Configure named network endpoints with dedicated access policies to manage data access at scale for shared datasets in Amazon S3.

Access Management & Security

- Access control lists (ACLs) – Grant read and write permissions for individual buckets and objects to authorized users. As a general rule, we recommend using S3 resource-based policies (bucket policies and access point policies) or IAM user policies for access control instead of ACLs. Policies are a simplified and more flexible access control option. With bucket policies and access point policies, you can define rules that apply broadly across all requests to your Amazon S3 resources. For more information about the specific cases when you'd use ACLs instead of resource-based policies or IAM user policies, see Access policy guidelines.
- S3 Object Ownership – Take ownership of every object in your bucket, simplifying access management for data stored in Amazon S3. S3 Object Ownership is an Amazon S3 bucket-level setting that you can use to disable or enable ACLs. By default, ACLs are disabled. With ACLs disabled, the bucket owner owns all the objects in the bucket and manages access to data exclusively by using access-management policies.
- IAM Access Analyzer for S3 – Evaluate and monitor your S3 bucket access policies, ensuring that the policies provide only the intended access to your S3 resources.