

Practical 2 - Storage as a Service using AWS

1] Storage as a Service - S3 :

S3 (Simple Storage Service) is a scalable object storage service offered by AWS designed for storing and retrieving any amount of data from anywhere on the web. It is known for its durability, scalability, security & performance.

• Key Features -

1. **Scalability** → Automatically scales storage capacity to accommodate growing data needs without any user intervention.
2. **Durability and Availability** → Amazon S3 is designed to provide 99.999999999% (11 9's) durability ensuring data is safe and ensures 99.99% availability over a given year.
3. **Storage Classes** → Standard (General purpose), Intelligent-Tiering, Standard-IA (Infrequent Access), One Zone-IA, Glacier, Glacier Deep Archive.
4. **Security** → Supports server-side encryption (SSE) with AWS managed keys (SSE-S3), AWS KMS managed keys etc.
5. **Logging & Monitoring** tracks access and usage with AWS CloudTrail and S3 access logs.
6. **Data Transfer** → Speeds up data transfers by using Amazon CloudFront's globally distributed edge locations.

• Benefits -

1. **Cost-effective** → Pay-as-you-go pricing model with no upfront costs. Multiple storage classes help optimize costs based on access patterns.
2. **Easy Integration** → Integrates with a wide range of AWS services and third party applications, facilitating diverse and complex workflows.

2] S3 usecases:

Simple Storage Service (S3) offers a versatile & scalable storage solution suitable for a wide range of use cases.

1. Backup and Restore → S3 provides highly durable storage that is ideal for backing up critical data.
2. Data Backup and Archiving → S3 is commonly used for data backup and long-term archiving due to its durability and reliability.
3. Content Distribution → It can be used to store static content such as images, videos and downloadable files. Content can be distributed globally to users through Amazon CloudFront, a content delivery network (CDN).
4. Data Lakes → It is often a foundational component of data lake architectures. It allows organizations to store structured and unstructured data in its native format, facilitating data analytics and processing.
5. Big Data & Analytics → S3 is commonly used to store raw data for big data & analytics pipelines. It serves as a landing zone for data before processing it with tools like Apache Spark, Hadoop, or Amazon EMR.
6. Log and Event Storage → Applications & systems can use S3 to store logs, events and telemetry data. This data can later be analyzed for monitoring, troubleshooting & compliance purposes.

3] Steps for S3:

1. Create an AWS account
2. Access the AWS Management Console
 - Login
 - Navigate : Go to S3 dashboard by searching 'S3' in the console or selecting it from list of services.

3. Create an S3 Bucket

→ Click 'Create Bucket'

→ Configure Bucket : • Bucket Name

• Choose the AWS region

• Configure optional settings

4. Setup Bucket Policies and Permissions

→ Access Control

→ Bucket Policy

→ Configure AWS IAM policies

→ ACLs : Set permissions at object level if needed.

5. Upload objects

→ Upload files : Use the S3 console to upload files and folders to your bucket.

• Drag and Drop : Use the S3 web interface to drag & drop files.

6. Manage and Organize Data

→ Use S3 console to create folders to organize objects

→ Assign metadata to objects for easier management

→ Enable versioning to keep multiple versions of objects

7. Setup Data Management Policies

8. Monitor & Optimize

9. Configure Security

10. Integrate with Other AWS services.