AWS Object Storage - S3

Simple Storage Service - S3

TABLE OF **CONTENTS**

01

Introduction to S3

Introduction

A scalable, high-speed, web-based cloud storage service designed to store and retrieve any amount of data from anywhere on the web. Since its inception, S3 has been one of the most fundamental services offered by AWS, providing a simple yet highly reliable object storage solution.

02
Understand Concepts

Understanding Concepts

- Buckets & Objects
- Bucket Naming & Properties
- Data Consistency

2.1

Buckets & Objects

Buckets & Objects

Bucket
A container for storing objects. Every object is stored in a bucket.

Object
The fundamental entity stored in S3, consisting of data, metadata, and a unique identifier (key).

Key
The unique identifier for an object within a bucket.

2.2

Naming & Properties

Naming & Properties

Global Uniqueness

Bucket names must be unique across all of AWS.

DNS Compliance

Bucket names must comply with DNS naming conventions.

Region Specific

Buckets are created in specific AWS regions, and data stored in a bucket never leaves the region unless explicitly configured

2.3

Data Consistency

Data Consistency

For new PUTs of objects.

Read-After-Write Consistency:

Eventual Consistency

For overwrite PUTs and DELETEs, meaning changes may take some time to propagate.

03
Key Features

Key Features

- Storage Classes
- Data Management
- Security & Access Controls
- Data Transfers
- Data Processing
- Logging and Monitoring
- Integration with AWS Services

3.1
Storage Classes

Storage Classes

S3 Standard

General-purpose storage for frequently accessed data

Intelligent-Tiering

Automatically moves data between two access tiers when access patterns change.

Standard-IA

Lower-cost storage for data that is accessed less frequently but requires rapid access when needed.

Storage **Classes**

One Zone-IA

Similar to Standard-IA but stored in a single Availability Zone.

S3 Glacier

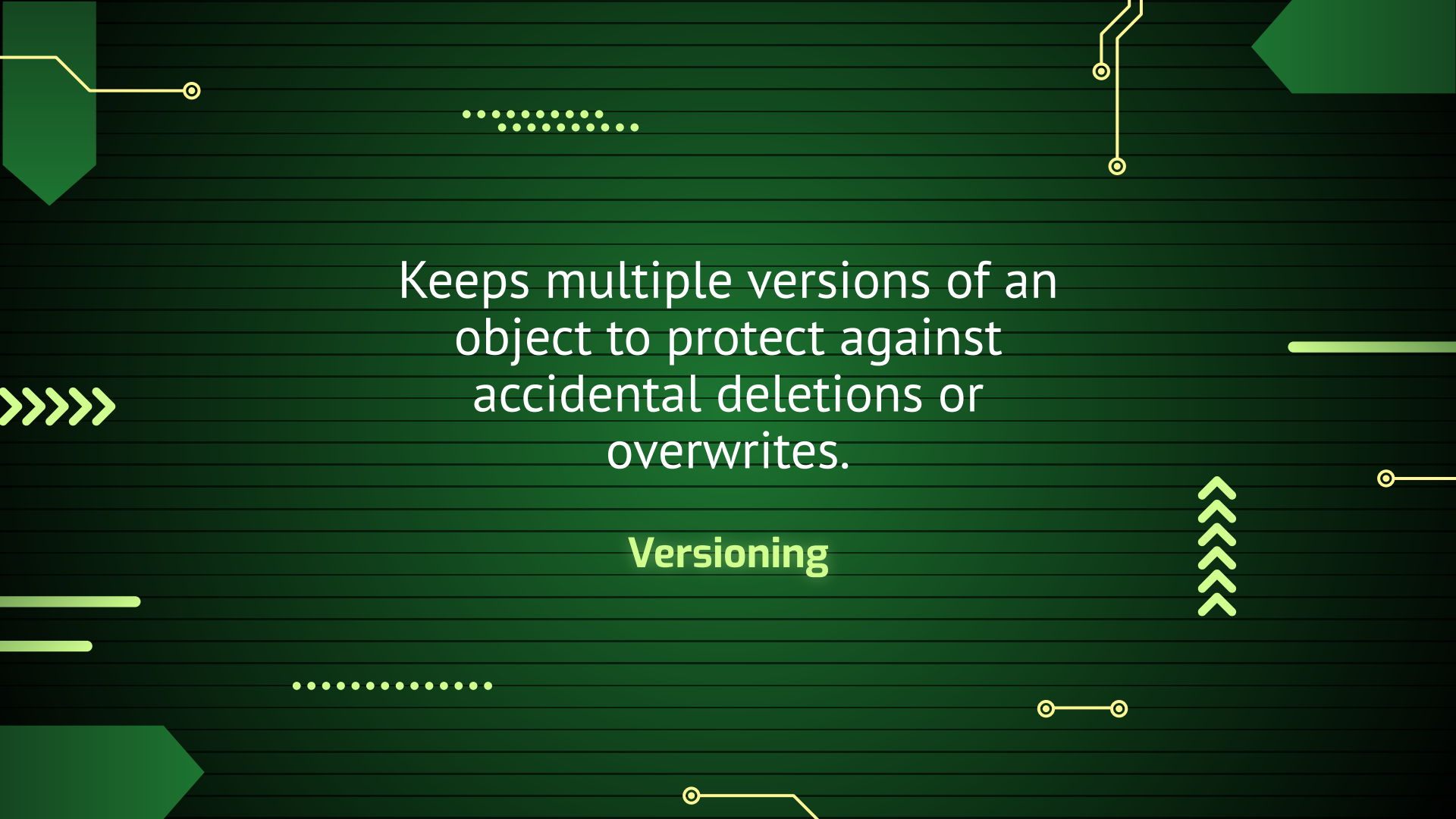Low-cost storage for data archiving where retrieval times of minutes to hours are acceptable.

Glacier Deep Archive

Lowest-cost storage for long-term data archiving with retrieval times of up to 12 hours.
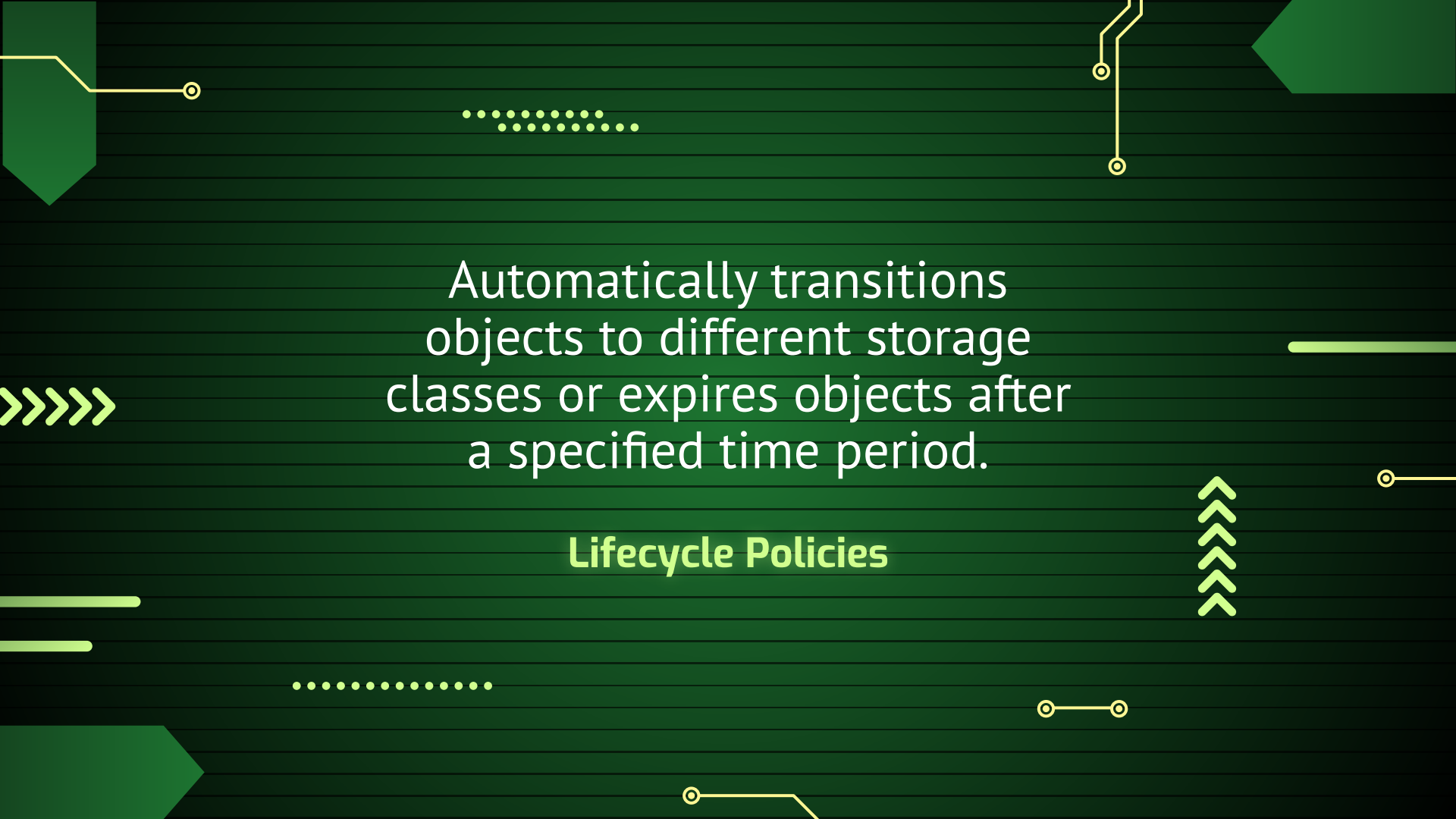
3.2
Data Management

Keeps multiple versions of an object to protect against accidental deletions or overwrites.

Versioning

Automatically transitions objects to different storage classes or expires objects after a specified time period.

Lifecycle Policies

Replicates objects across different AWS regions (Cross-Region Replication) or within the same region (Same-Region Replication).
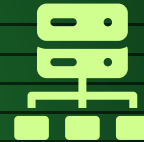
Replication

3.3
Security & Access Control

Security & Access Controls

IAM Policies

Control access to S3 resources using AWS Identity and Access Management (IAM) policies.
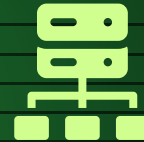
Bucket Policies

Apply fine-grained access control policies at the bucket level.

Security & Access Controls

ACLs

Set permissions for individual objects within a bucket.
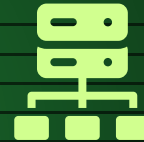
Server-side Encryption

Encrypts data at rest using AWS-managed keys (SSE-S3),
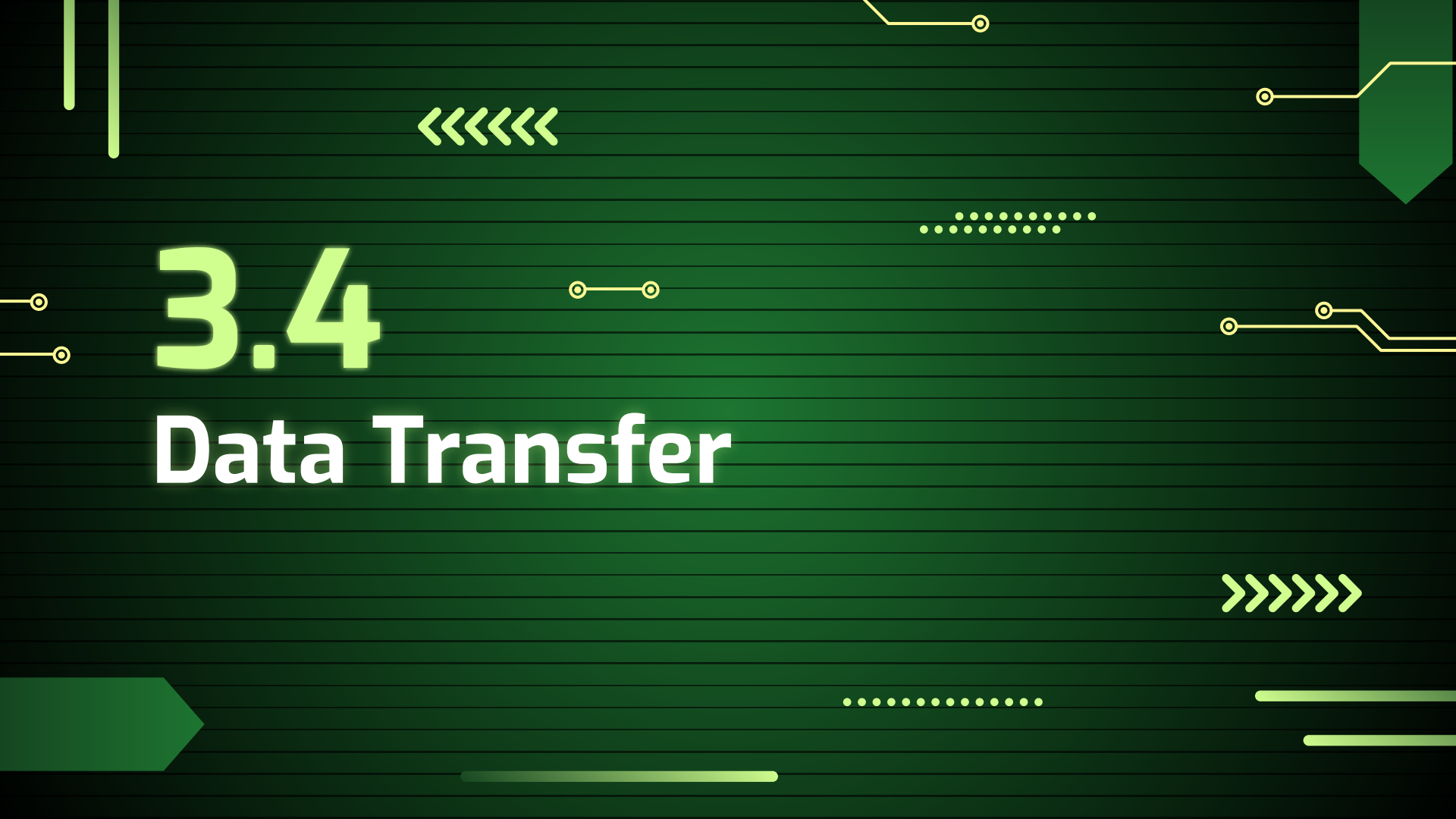
Security & Access Controls

Block Public Access

Prevent public access to buckets and objects unless explicitly granted.

Client-side Encryption

Encrypt data before sending it to S3.

3.4

Data Transfer

Data Transfer

Multipart Upload

Increases upload efficiency by dividing a large object into smaller parts and uploading them concurrently.
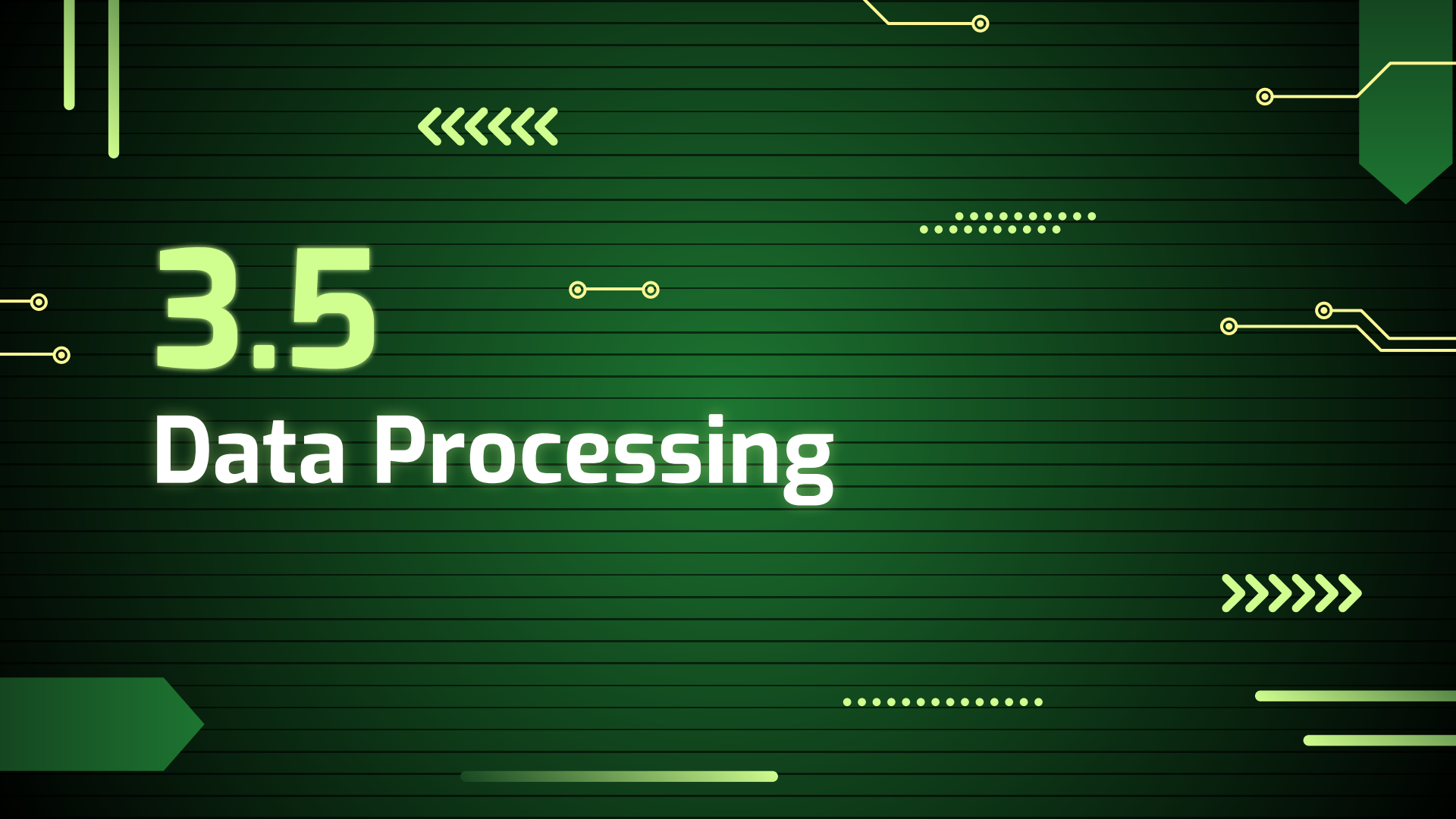
Transfer Acceleration

Speeds up content uploads and downloads by using AWS CloudFront's globally distributed edge locations.

S3 Transfer Manager

High-level library for managing S3 data transfers.

3.5
Data Processing

Data Processing

Retrieve subsets of data from an object using SQL expressions, reducing the amount of data transferred.

S3 Select

S3 Object Lambda

Run code on S3 data as it is being retrieved using AWS Lambda, allowing for data transformation and processing on the fly.

3.6

Logging & Monitoring

Logging & Monitoring

Server Access Logging

Track requests for access to your S3 buckets.

AWS CloudTrail

Capture API calls made to S3 for governance, compliance, and auditing.

AWS CloudWatch

Monitor storage metrics and configure alarms.

3.7
Integration with AWS

Integration with AWS

Trigger serverless functions in response to S3 events.

Query data directly in S3 from Amazon Redshift.

Process big data stored in S3 with Hadoop, Spark, and other big data frameworks.

Lambda

Redshift Spectrum

AWS EMR

Athena

AWS Glue

Query S3 data using standard SQL.

Data cataloging, ETL (Extract, Transform, Load), and data preparation.

04

Use Cases

Use Cases of S3
01
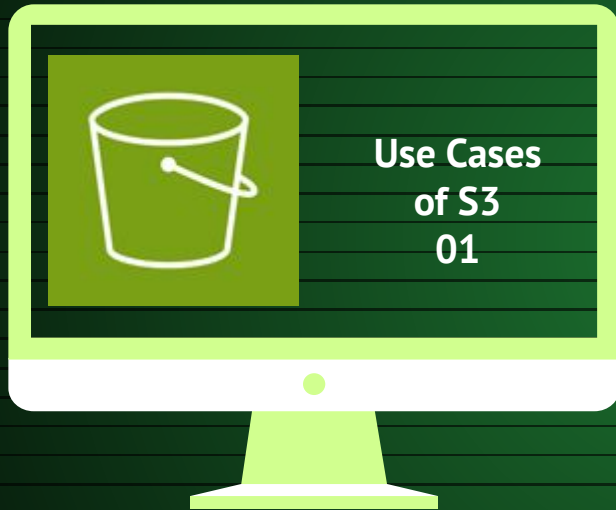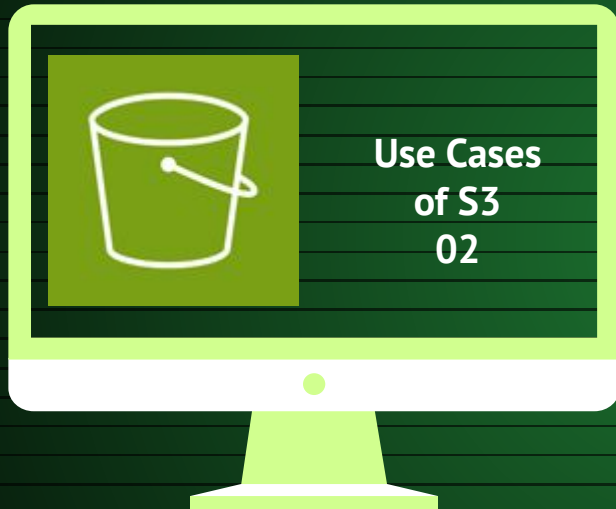
Backup & Restore

- Use S3 to store backups of critical data, leveraging its durability and availability

- Utilize lifecycle policies to transition older backups to Glacier or Glacier Deep Archive.
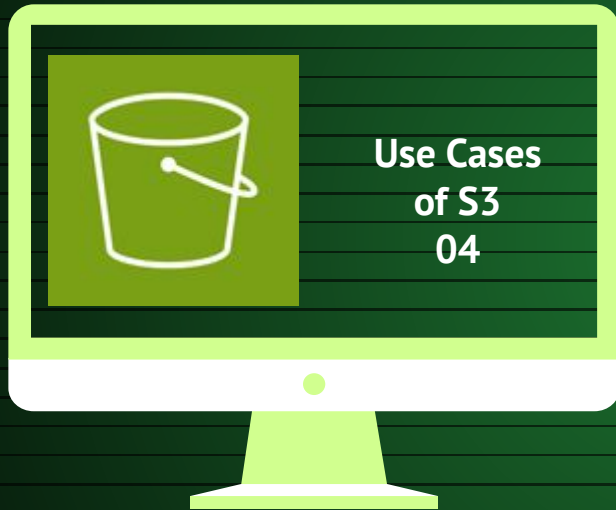
Use Cases
of S3
02

Data Lakes

Store vast amounts of raw data in S3, which can be processed and analyzed using AWS analytics services like Athena, Redshift, and EMR.
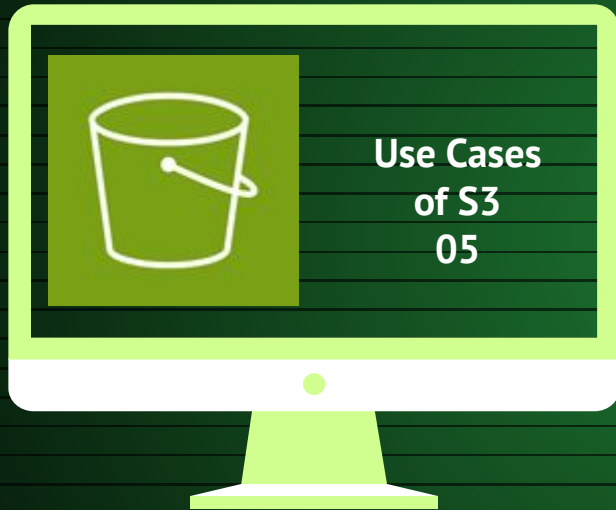
Use Cases of S3

03
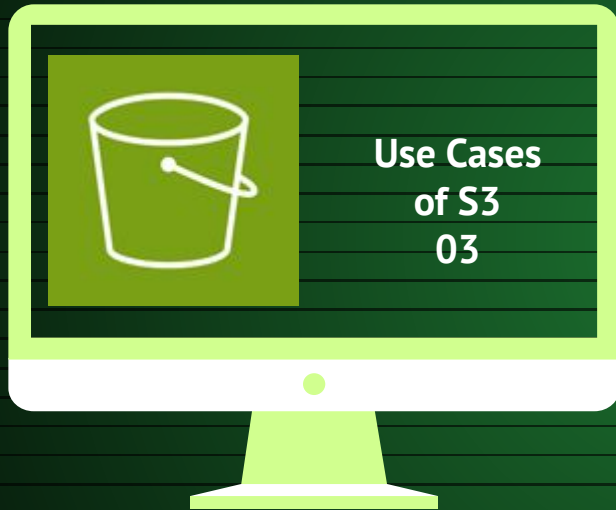
Content Distribution

Store static content (e.g., images, videos) in S3 and distribute it globally using AWS CloudFront.
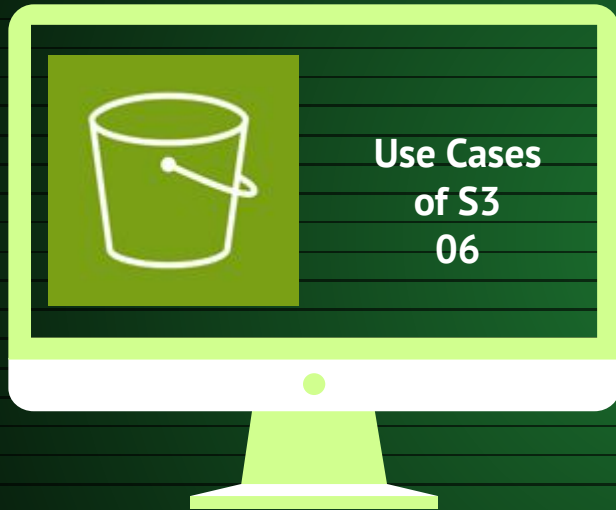
Use Cases of S3
04

Big Data Analytics

Store large datasets in S3 and process them using analytics frameworks and tools such as Hadoop, Spark, and Presto.

Use Cases of S3
05

Archival & Compliance

Archive data that must be retained for long periods for compliance purposes using S3 Glacier or Glacier Deep Archive.

Use Cases of S3
06

Application Hosting

Host static websites directly from S3, utilizing its scalability and cost-effectiveness.

05
Best Practices

Security Best Practice

- Enable versioning to protect against accidental deletions.

- Use server-side encryption and AWS KMS for sensitive data.

- Apply least privilege principle with IAM policies.

- Regularly review and update bucket policies and ACLs.

Cost Best Practice

- Monitor storage usage with AWS Cost Explorer.
- Use lifecycle policies to transition data to cheaper storage classes.
- Analyze access patterns and choose the appropriate storage class.

Performance Best Practice

- Use S3 Transfer Acceleration for faster uploads and downloads.

- Optimize multipart upload for large objects.

- Use S3 Select to retrieve only necessary data.

Durability & Availability Best Practice

- Enable Cross-Region Replication for disaster recovery.

- Use S3 Intelligent-Tiering for automatic cost optimization based on access patterns.

END OF Part 1

Thanks!

Do you have any questions?

mayamnaiel2013@gmail.com

The Tech Stuff