

Amazon S3 - Dịch vụ lưu trữ đối tượng của AWS

Amazon Simple Storage Service: Lưu trữ không giới hạn, an toàn, sẵn sàng cao



Scalability



Security



Performance



Durability

Khái niệm cơ bản về Amazon S3



Amazon S3 (Amazon Simple Storage Service) là dịch vụ lưu trữ đối tượng (object storage) của AWS.

Được thiết kế để lưu trữ mọi loại dữ liệu không cấu trúc như ảnh, video, tài liệu, log, dữ liệu máy học, backup, v.v.



Scalability

Tự động mở rộng quy mô để lưu trữ không giới hạn dữ liệu



Durability

99.99999999% (11 số 9) — dữ liệu gần như không mất



Availability

Dữ liệu có thể truy cập 24/7, được lưu trữ đa AZ



Security

Nhiều lớp mã hóa, quản lý quyền truy cập



Performance

Truy cập nhanh, tối ưu cho big data, AI/ML

English: Amazon S3 (Amazon Simple Storage Service) is an object storage service offering industry-leading scalability, data availability, security, and performance.

Các đặc tính chính của Amazon S3



Khả năng mở rộng

- Tự động mở rộng quy mô để lưu trữ không giới hạn dữ liệu
- Xử lý khối lượng dữ liệu lớn mà không cần can thiệp thủ công
- Scale ngang hàng qua nhiều máy chủ

English: Automatically scales to store unlimited amounts of data



Độ bền dữ liệu

- 99.999999999% **999999999** (11 số 9)
- Dữ liệu gần như không bao giờ bị mất
- Được thiết kế để giữ dữ liệu lâu dài

English: 99.999999999% (11 nines) durability, data is virtually never lost



Tính sẵn sàng cao

- Dữ liệu có thể truy cập 24/7
- Lưu trữ đa Availability Zones (AZ)
- Đảm bảo sẵn sàng cho các ứng dụng quan trọng

English: Data accessible 24/7, stored across multiple AZs



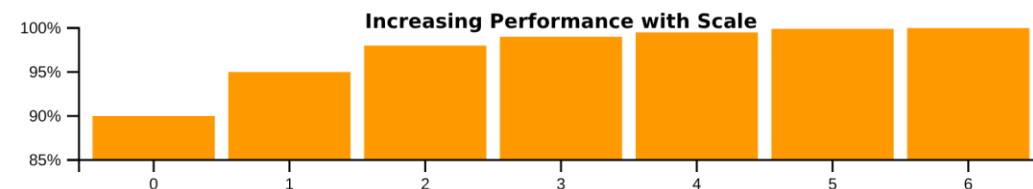
Bảo mật đa lớp

- Mã hóa dữ liệu
- Quản lý quyền truy cập
- Đa dạng hóa dữ liệu
- Bảo vệ khỏi mất mát
- Kiểm soát truy cập
- Compliance

English: Multiple layers of encryption and access management controls



Hiệu năng



- Truy cập nhanh đến dữ liệu
- Tối ưu cho big data




Thành phần cốt lõi - Buckets và Objects



Bucket (Thùng chứa)

Là thư mục gốc hoặc container logic để chứa dữ liệu.

Thuộc tính chính:

-  **Tài nguyên cấp vùng:** Được tạo trong một AWS Region cụ thể.
-  **Tên duy nhất toàn cầu:** Tên bucket phải là duy nhất trên toàn bộ AWS.
-  **Cấu trúc phẳng:** Dữ liệu bên trong bucket không có cấu trúc thư mục vật lý.

Ví dụ:

Cấu trúc URL của Amazon S3



`https://` `notes-bucket` `.s3.` `us-west-2` `.amazonaws.com/` `awsservice/notes.txt`

Bucket Name

Tên của bucket chứa object, phải là duy nhất toàn cầu trong AWS.

Region

Vùng AWS nơi bucket được tạo ra, ví dụ: us-west-2, eu-central-1.

Key

Đường dẫn hoặc tên duy nhất của object trong bucket.

English: Every object in S3 has a unique URL that includes the bucket name, region, and object key.

Tổ chức dữ liệu trong S3 - Prefix và Delimiter

Mặc dù S3 có **cấu trúc lưu trữ phẳng** (flat structure), không có thư mục vật lý thực sự, nhưng nó cho phép người dùng mô phỏng cấu trúc thư mục logic thông qua việc sử dụng **Prefix** và **Delimiter**.

Prefix & Delimiter

Prefix (Tiền tố): Chuỗi ký tự đứng đầu Key của object, giúp nhóm các object có liên quan lại với nhau.

Delimiter (Ký tự phân cách): Thường là dấu gạch chéo (/), được sử dụng để tạo ra một hệ thống phân cấp logic.

Ví dụ

Bucket: notes-bucket

Dev/awsservice/dynamodb/notes.txt

Dev/awsservice/polly/sam.mp3

Dev/awsservice/summary.txt

readMe.txt

English: S3 has a flat structure, but allows users to simulate a hierarchical structure through Prefix and Delimiter.

Cấu trúc thư mục logic

Dev

awsservice

dynamodb

notes.txt

polly

sam.mp3

summary.txt

readMe.txt

Lợi ích



Tổ chức dữ liệu hợp lý



Dễ lọc và tìm kiếm



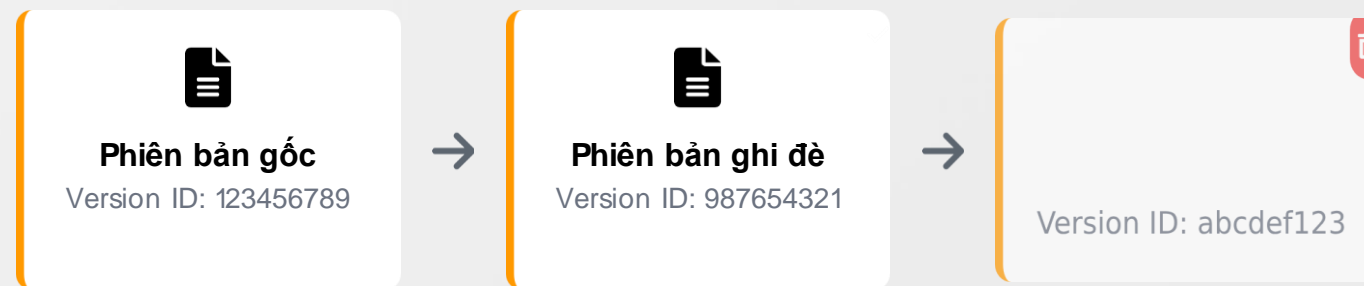
Phân quyền chi tiết

Object Versioning - Quản lý phiên bản đối tượng



Object Versioning là tính năng cho phép lưu trữ nhiều phiên bản của cùng một đối tượng trong bucket.

Khi Versioning được bật, mỗi khi một đối tượng bị ghi đè hoặc xóa, S3 sẽ giữ lại phiên bản cũ thay vì xóa vĩnh viễn. Mỗi phiên bản có một **Version ID** riêng biệt.



Phục hồi dữ liệu

Dễ dàng phục hồi các file bị xóa nhầm hoặc ghi đè, đảm bảo không mất dữ liệu quan trọng.









Kiểm toán và kiểm soát thay đổi

Cung cấp khả năng theo dõi lịch sử thay đổi của đối tượng, hữu ích cho mục đích kiểm toán và tuân thủ.

English: When Versioning is enabled, each time an object is overwritten or deleted, S3 keeps the previous version instead of deleting it permanently. Each version has a unique Version ID.

Các lớp lưu trữ trong Amazon S3

Amazon S3 cung cấp nhiều lớp lưu trữ được thiết kế để đáp ứng các yêu cầu khác nhau về chi phí, hiệu suất và tần suất truy cập dữ liệu.

Storage Class	Mục đích sử dụng	Độ bền	Độ sẵn sàng	Chi phí	Thời gian truy cập
 S3 Standard	Frequently accessed data	99.999999999%	99.99%	\$ \$ \$	Milliseconds
 S3 Intelligent-Tiering	Changing/unknown access patterns	99.999999999%	99.9%		Milliseconds
 S3 Standard-IA	Infrequently accessed data	99.999999999%	99.9%	\$ \$	Milliseconds
 S3 One Zone-IA	Infrequently accessed, 1 AZ	99.999999999%	99.5%	\$	Milliseconds
 S3 Glacier IR	Long-term archive, rare, instant	99.999999999%	99.9%	\$	Milliseconds
 S3 Glacier Flexible	Long-term archive, rare, slow	99.999999999%	99.99%	\$	Minutes to hours
 S3 Glacier Deep	Longest-term archive, rare	99.999999999%	99.99%	\$	12-48 hours
 S3 Express One Zone	Frequent, low-latency, 1 AZ	99.999999999%	99.95%	\$ \$	Fastest Milliseconds

S3 Lifecycle - Quản lý vòng đời đối tượng



S3 Lifecycle (Chính sách vòng đời) cho phép tự động quản lý các đối tượng trong S3 bucket theo các quy tắc định sẵn.

Các chính sách này giúp tối ưu hóa chi phí lưu trữ và tuân thủ các yêu cầu về lưu giữ dữ liệu bằng cách tự động chuyển đổi đối tượng giữa các lớp lưu trữ hoặc xóa chúng sau một khoảng thời gian nhất định.



Ví dụ về chính sách Lifecycle

Một chính sách Lifecycle có thể được cấu hình để tự động chuyển các file log từ **S3 Standard** sang **S3 Standard-IA** sau 30 ngày, sau đó chuyển sang **S3 Glacier Flexible Retrieval** sau 90 ngày, và cuối cùng xóa chúng sau 365 ngày.



Giảm chi phí lưu trữ cho dữ liệu ít được truy cập theo thời gian



Tuân thủ yêu cầu về lưu giữ dữ liệu



Tự động hóa quản lý không cần can thiệp



Lưu trữ đa lớp phù hợp với các yêu cầu khác nhau

English: S3 Lifecycle policies help optimize storage costs and comply with data retention requirements by automatically transitioning objects between storage classes or deleting them after a specified period.

Bảo mật trong Amazon S3

Amazon S3 cung cấp nhiều lớp bảo mật hoạt động ở các cấp độ khác nhau để đảm bảo quyền truy cập được kiểm soát chặt chẽ và dữ liệu được bảo vệ.

Bucket-level Security



Bucket Policy

Chính sách JSON xác định ai có thể truy cập bucket và thực hiện các hành động gì



Access Points

Entry points đặc biệt cho ứng dụng với policy riêng



User-level Security



IAM Policy

Gắn cho user/role để giới hạn truy cập vào S3 resources



Object-level Security



ACL (Access Control List)

Quyền riêng cho từng object (ít dùng)



Data Encryption



SSE-S3

Mã hóa server-side với khóa quản lý bởi AWS



SSE-KMS

Mã hóa server-side với khóa quản lý bởi AWS KMS



SSE-C

Mã hóa client-side trước khi upload

Phương thức kết nối với Amazon S3

Amazon S3 cung cấp nhiều phương thức kết nối để người dùng và ứng dụng tương tác với dịch vụ lưu trữ:



AWS Management Console

Giao diện web trực quan để quản lý bucket và object



AWS CLI

Công cụ dòng lệnh để tự động hóa tác vụ quản lý S3



AWS SDKs

Thư viện lập trình cho Python, Java, Node.js, .NET



REST API

Giao diện HTTP chuẩn để tạo, đọc, cập nhật và xóa

Ví dụ truy cập bằng AWS SDK (Python Boto3)

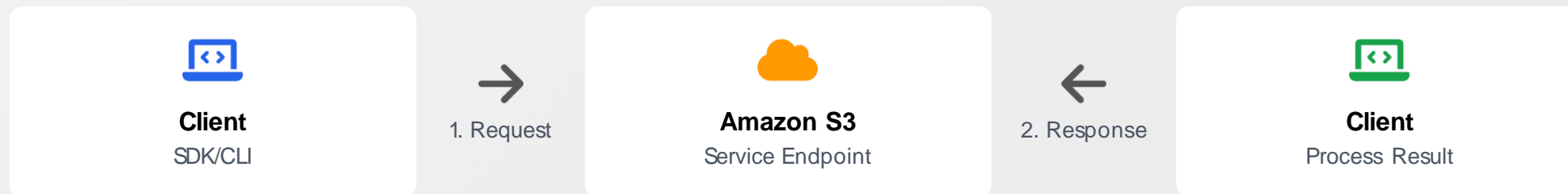
```
import boto3

s3 = boto3.client('s3')
response = s3.list_buckets()
print([bucket['Name'] for bucket in response['Buckets']])
```

English: Every request to S3 goes through a Service Endpoint (default per region). SDKs & CLI automatically choose the correct endpoint.

Request và Response Objects

Khi gọi API đến Amazon S3 (qua SDK hoặc CLI):



➤ PutObject Request

PUT /myphoto.jpg **HTTP/1.1**

Host: my-bucket.s3.us-east-1.amazonaws.com

x-amz-storage-class: STANDARD

➤ Response

HTTP/1.1 200 OK

ETag: "9b2cf535f27731c974343645a3985328"

English: When calling API, Request contains action info and parameters. Response returns HTTP status, metadata, or data.

Các trường hợp sử dụng của Amazon S3

Amazon S3 là một dịch vụ lưu trữ đa năng, được ứng dụng rộng rãi trong nhiều lĩnh vực và trường hợp sử dụng khác nhau nhờ vào khả năng mở rộng, độ bền và tính sẵn sàng cao.



Static Website Hosting

Lưu trữ trang web tĩnh (HTML, CSS, JS) với khả năng truy cập toàn cầu.



Backup & Archiving

Lưu trữ dự phòng và snapshot dữ liệu quan trọng với chi phí lưu trữ thấp.



Content Distribution

Lưu media và tài liệu cho người dùng cuối với tốc độ truy cập nhanh.

Big Data Analytics

Lưu dữ liệu nguồn cho Athena, Glue, EMR và các dịch vụ phân tích dữ liệu.



Disaster Recovery

Khôi phục dữ liệu khi sự cố xảy ra nhờ vào độ bền cao của S3.

English: Amazon S3 is a versatile storage service that can be applied in various scenarios including static website hosting, backup and archiving, content distribution, big data analytics, and disaster recovery.

SDK Dependencies - Thư viện phụ thuộc

Để làm việc với Amazon S3 qua code, bạn cần cài đặt các thư viện SDK sau:



Python

Package: boto3

Thư viện chuẩn để tương tác với S3 trong Python, hỗ trợ mọi chức năng API của S3.



Node.js

Package: aws-sdk

Thư viện JavaScript để tương tác với S3, hỗ trợ cả Node.js và trình duyệt.



Java

Package: aws-java-sdk-s3

Thư viện Java để tương tác với S3, là một phần của bộ SDK Java AWS đầy đủ.



.NET

Package: AWSSDK.S3

Thư viện .NET để tương tác với S3, hỗ trợ cả .NET Framework và .NET Core.



Lợi ích của việc sử dụng SDK:

SDK tự động xử lý xác thực, ký request, và chọn endpoint phù hợp với khu vực AWS của bạn.

Tổng kết về Amazon S3

Amazon S3 là dịch vụ lưu trữ đối tượng mạnh mẽ và linh hoạt, cung cấp khả năng lưu trữ không giới hạn, an toàn, sẵn sàng cao và hiệu suất vượt trội.



Tổ chức dữ liệu

Dữ liệu được tổ chức thông qua Buckets (thùng chứa) và Objects (đối tượng), với khả năng mô phỏng cấu trúc thư mục logic.



Bảo mật mạnh mẽ

Cung cấp nhiều lớp bảo mật, bao gồm mã hóa dữ liệu, kiểm soát truy cập, và các cơ chế bảo vệ dữ liệu.



Khả năng mở rộng

Tự động mở rộng quy mô để lưu trữ không giới hạn dữ liệu, với độ bền 99.999999999% và sẵn sàng 24/7.



Hiệu năng cao

Truy cập nhanh, tối ưu cho big data, AI/ML, với nhiều lớp lưu trữ khác nhau để tối ưu hóa chi phí.



Dễ quản lý

Quản lý thông qua AWS Console, CLI, SDK với nhiều ngôn ngữ, và Lifecycle policies để tự động quản lý vòng đời dữ liệu.



Ứng dụng đa dạng

Dùng cho website hosting, backup, phân tích dữ liệu lớn, disaster recovery, và nhiều trường hợp sử dụng khác.

Amazon S3 - Dịch vụ lưu trữ đối tượng hàng đầu của AWS

Cung cấp giải pháp lưu trữ dữ liệu mạnh mẽ và linh hoạt cho nhiều trường hợp sử dụng