

# AWS Cloud for beginner

Instructor: **Linh Nguyen**

(Engineering Consultant, AWS Cloud Solution Architect)

**Level: Beginner**

# Simple Storage Service – S3

Copyright@Linh Nguyen on Udemy

# Target

Hiểu được S3 là gì, các tính năng của S3.

Hiểu được S3 resource policy, Access Control List (ACL), cách vận dụng.

Thành thạo các thao tác cơ bản với S3 (Console/CLI).

Hiểu được S3 storage classes và trường hợp sử dụng.

Kết hợp S3 trigger với Lambda thông qua bài lab đơn giản.

Sử dụng S3 để host 1 website tĩnh (html, css, js).

Copyright@Linh Nguyen on Udemy

# What is S3

## Viết tắt của **Simple Storage Service**

Là một dịch vụ lưu trữ dạng Object cung cấp khả năng mở rộng, availability, performance.

KH có thể sử dụng S3 để lưu trữ và bảo vệ nhiều loại data cho các usecase như: data lake, website, mobile, backup & restore, archive, enterprise application, IoT device, Big Data & Analytic.

S3 cung cấp nhiều managed feature giúp tối ưu, tổ chức và cấu hình access tới data đáp ứng nhu cầu về business, organization & complicate.



Amazon Simple Storage Service (Amazon S3)

Copyright@Linh Nguyen on Udemy

# Đặc trưng cơ bản của S3

- Là một Managed Service. User không cần quan tâm tới hạ tầng ở bên dưới.
- Cho phép lưu file dưới dạng object với size từ 0 - 5TB
- High Durability (11 9s), Scalability, High Availability (99.99%), High performance.
- Usecase đa dạng (mọi bài toán về lưu trữ từ lớn tới nhỏ đều có thể sd S3).
- Cung cấp nhiều class lưu trữ để tiết kiệm chi phí cho từng loại data.
- Cung cấp khả năng phân quyền và giới hạn truy cập một cách chi tiết.
- Dễ sử dụng, có thể kết hợp với nhiều service khác cho bài toán automation và data processing.

# Features of S3

S3 cung cấp các tính năng cơ bản sau:

- Storage classes: cung cấp nhiều hình thức lưu trữ phù hợp cho nhiều loại data khác nhau về nhu cầu access, yêu cầu về durability, thời gian lưu trữ khác nhau giúp KH tùy chọn được class lưu trữ phù hợp từ đó tối ưu chi phí.
- Storage management: Cung cấp nhiều tính năng liên quan quản lý như: Life Cycle, Object Lock, Replication, Batch Operation
- Access Management: quản lý truy cập đến bucket và các thư mục thông qua cơ chế resource permission & access list. Block public access, control access via IAM, bucket policy, S3 access point, Access Control List, Ownership, Access Analyzer.
- Data processing: kết hợp với lambda, SNS, SQS để hỗ trợ xử lý data 1 cách nhanh chóng

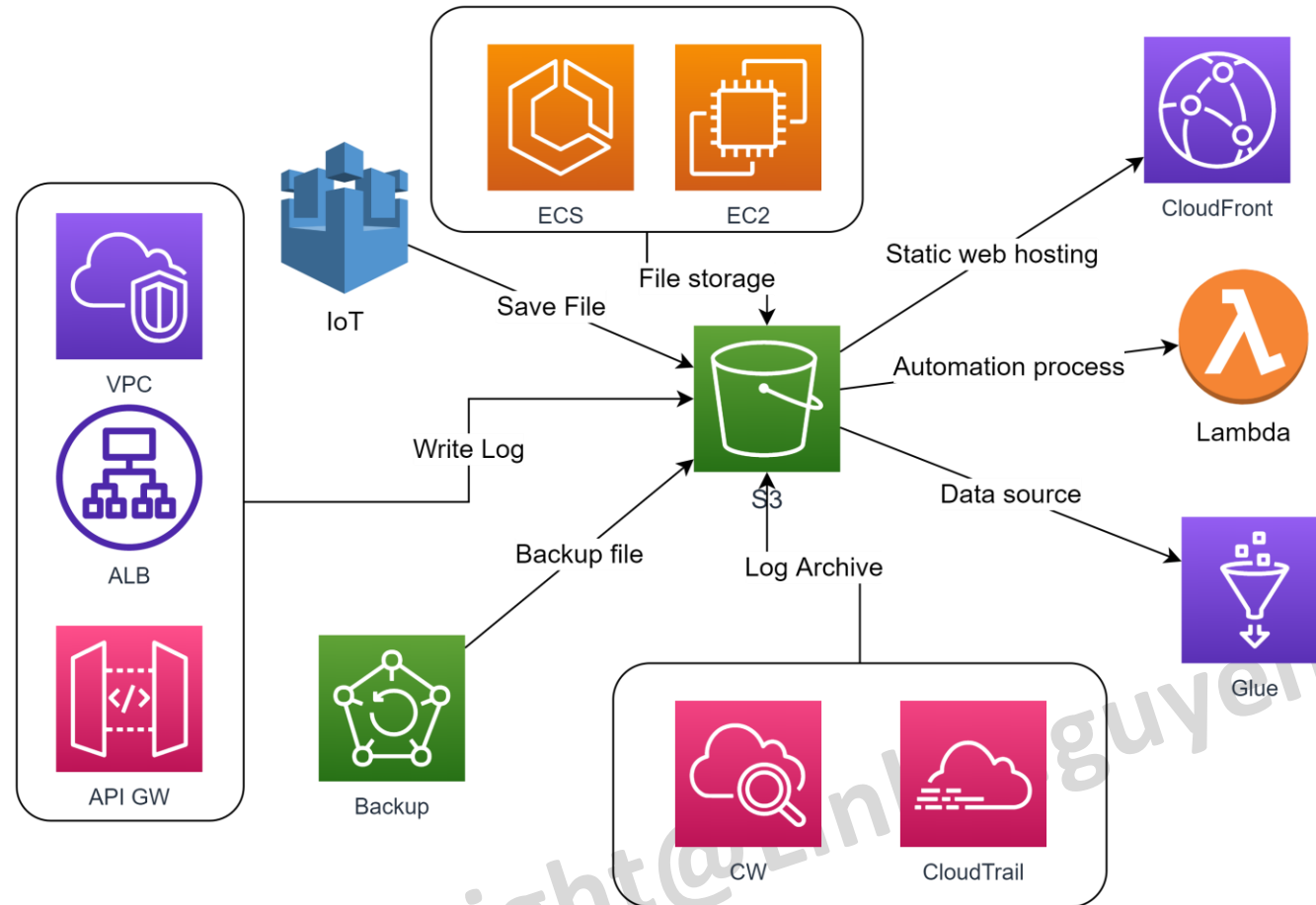
# Features of S3

S3 cung cấp các tính năng cơ bản sau:

- Auto Logging and Monitoring: cung cấp công cụ monitor S3 bucket và truy vết sử dụng CloudTrail.
- Manual Monitoring Tool: Log lại từng record thực hiện trên bucket.
- Analytic and insight: phân tích storage để optimize.
- Strong consistency: Provide strong read-after-write consistency for PUT and DELETE object.

Copyright@Linh Nguyen on Udemy

# S3 có thể kết hợp với các dịch vụ nào?



\*Lưu ý: trên đây chỉ là những ví dụ nổi bật thường được sử dụng



# S3 có thể kết hợp với các dịch vụ nào?

- Dùng làm nơi lưu trữ file cho các ứng dụng chạy trên EC2, Container, Lambda. Các file có thể đa dạng về loại & kích thước (Image, Video, Document,...)
- Dùng làm nơi chứa/archive log cho hầu hết các dịch vụ khác của AWS (VPC, ALB, APIGateway,...)
- Dùng làm data source cho các bài toán big data & data warehouse
- Nơi lưu dữ liệu gửi lên từ các thiết bị IoT
- Vùng lưu trữ tạm thời cho bài toán ETL (Extract – Transform - Load) khi kết hợp với lambda
- Host 1 website tĩnh (html,css,js) khi kết hợp với CloudFront

# Lab 1 – S3 Basic Operation

1. Login to AWS console, navigate to S3
2. Tạo 1 bucket \*Lưu ý tên s3 bucket là global unique.
3. Thực hiện tạo folder
4. Thực hiện upload file/folder
5. Thực hiện move file
6. Download file
7. Chỉnh sửa object metadata để trình duyệt quyết định hành động default với file khi open (view trực tiếp hoặc popup download)
8. Thực hiện xóa file

# S3 Bucket Policy and Access Control List

- S3 là một trong số các resource hỗ trợ Resource Level Policy để giới hạn quyền truy cập bên cạnh IAM Policy (đã học ở bài IAM).
- Bản chất S3 bucket policy hoạt động như 1 IAM Policy nhưng chỉ trong phạm vi bucket và những resource bên trong nó (folder/file).
- S3 bucket policy sẽ cho phép (allow) hoặc chặn (deny) truy cập tới bucket hoặc các resource bên trong.
- Áp dụng cho những bucket yêu cầu security cao, cần được setting giới hạn truy cập một cách chặt chẽ.

Copyright@Linh Nguyen on Udemy

# S3 Bucket Policy and Access Control List

- S3 bucket policy example

Ví dụ về 1 bucket policy cho phép user “riccardo” có quyền GetObject trên tất cả resource của bucket “abc-bucket-sample-001”

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "Allow_riccardo_read",
      "Effect": "Allow",
      "Principal": {
        "AWS": [
          "arn:aws:iam::614073292094:user/riccardo"
        ]
      },
      "Action": [
        "s3:GetObject"
      ],
      "Resource": "arn:aws:s3:::abc-bucket-sample-001/*"
    }
  ]
}
```

# S3 Bucket Policy and Access Control List

- S3 bucket policy example

Ví dụ về 1 bucket policy deny user “ronaldo” không được phép làm bất kì action nào trên tất cả resource của bucket “abc-bucket-sample-001”

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "Deny_ronaldo_all",
      "Effect": "Deny",
      "Principal": {
        "AWS": [
          "arn:aws:iam::614073292094:user/ronaldo"
        ]
      },
      "Action": [
        "s3:*"
      ],
      "Resource": "arn:aws:s3:::abc-bucket-sample-001/*"
    }
  ]
}
```

# S3 Bucket Policy and Access Control List





- Access Control List: quy định quyền access của một AWS Account hoặc nhóm user (Group) đến bucket hoặc resource bên trong.
- Thường dùng trong trường hợp muốn cấp access cho một resource cụ thể bên trong bucket mà không muốn thay đổi bucket policy.
- \*Gần đây AWS khuyến nghị người dùng KHÔNG nên xài ACL trừ khi có yêu cầu đặc biệt, thay vào đó hay sử dụng bucket policy, iam policy & pre-sign URL là đủ để cover hầu hết các usecase.

Copyright@Linh Nguyen on Udemy

# S3 ACL

## Access control list (ACL)

Grant basic read/write permissions to other AWS accounts. [Learn more](#)

Grantee	Objects	Bucket ACL
Bucket owner (your AWS account) Canonical ID:  d6a77a72ac70c0ff7e1b0353b3597331c62bddd3dc70f09d7aa5e244d6cdcd4e	<input checked="" type="checkbox"/> List <input checked="" type="checkbox"/> Write	<input checked="" type="checkbox"/> Read <input checked="" type="checkbox"/> Write
Everyone (public access) Group:  http://acs.amazonaws.com/groups/global/AllUsers	<input type="checkbox"/> List <input type="checkbox"/> Write	<input type="checkbox"/> Read <input type="checkbox"/> Write
Authenticated users group (anyone with an AWS account) Group:  http://acs.amazonaws.com/groups/global/AuthenticatedUsers	<input type="checkbox"/> List <input type="checkbox"/> Write	<input type="checkbox"/> Read <input type="checkbox"/> Write
S3 log delivery group Group:  http://acs.amazonaws.com/groups/s3/LogDelivery	<input type="checkbox"/> List <input type="checkbox"/> Write	<input type="checkbox"/> Read <input type="checkbox"/> Write

## Access for other AWS accounts

No other AWS accounts associated with the resource.

Copy

## Lab 2 – S3 Access Control List

1. Upload một file bất kỳ (vd image jpg/png).
2. Modify Access List của một object thành public (Allow all).
3. Thử access object mà không cần login -> kết luận ACL có thể cho phép access public tới một object cụ thể

Copyright@Linh Nguyen on Udemy



# S3 Versioning

- Sử dụng khi có nhu cầu lưu trữ nhiều version của cùng 1 object.
- Tránh được việc mất mát khi thao tác xóa nhầm hoặc ghi đè (có thể lấy lại version trước đó).
- Chi phí theo đó sẽ tăng lên so với khi không bật versioning.
- Sau khi bật versioning, nếu tắt versioning thì những object trước khi tắt vẫn sẽ có nhiều version, những object sinh ra sau khi tắt sẽ không có version.

Copyright@Linh Nguyen on Udemy

## Lab 3 – Versioning

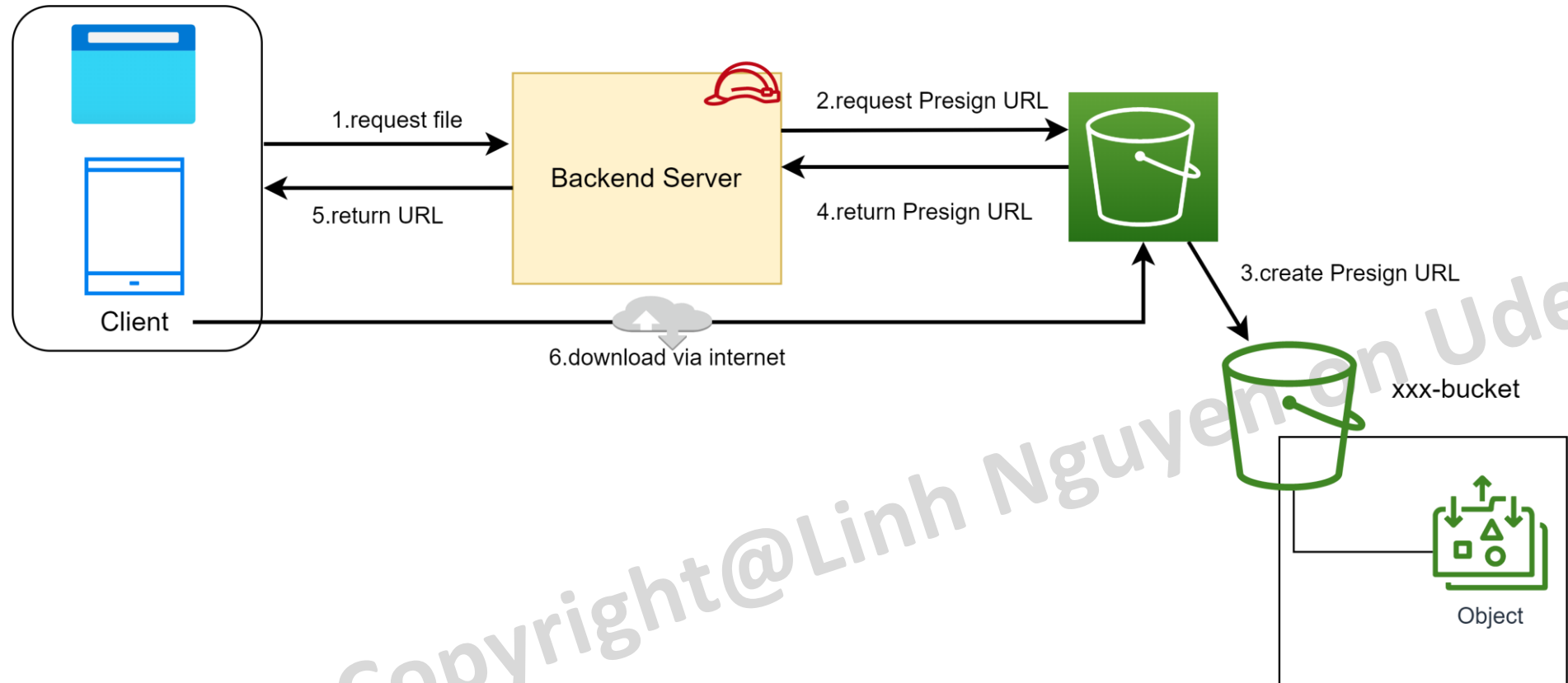
1. Yêu cầu chuẩn bị một file text có nội dung bất kỳ
2. Bật tính năng versioning của Bucket lên.
3. Upload file lên S3.
4. Chỉnh sửa nội dung, upload file với cùng tên. Confirm xem có một version mới được tạo ra?
5. Tiến hành xóa file. Kiểm tra versioning với Delete flag.
6. Phục hồi file đã bị xóa bằng cách xóa Delete marker.
7. Tắt versioning.
8. Chỉnh sửa nội dung, upload file với cùng tên. Confirm xem file có bị ghi đè? Các version trước khi tắt versioning có còn không?

## S3 Presign URL

- Khi muốn cấp access tạm thời cho người dùng public tới một object trên S3, AWS cung cấp cơ chế Presign URL.
- User có thể dùng Presign URL để download/upload object trên s3 trong thời gian quy định (setting lúc phát hành URL).
- Usecase
  - Muốn cấp access public cho 1 object nhưng không muốn thay đổi ACL hoặc tạo thêm bucket policy.
  - Cần authen người dùng hoặc yêu cầu họ làm gì đó trước khi được download file (vd xem quảng cáo).
  - Ngăn chặn resource để public vô thời hạn khiến cho tài nguyên bị khai thác bởi bên khác.

# S3 Presign URL

Flow for Presign URL (download/upload)



## Lab 4 – Pre-sign URL (sử dụng CLI)

1. Yêu cầu đã cài sẵn AWS CLI và thiết lập profile tại máy local
2. Upload 1 file bất kì lên S3
3. Tham khảo câu lệnh sau để phát hành 1 presign url cho object trên s3  
<https://docs.aws.amazon.com/cli/latest/reference/s3/presign.html>
4. Sử dụng URL để download file trên trình duyệt.
5. Đợi hết thời gian hiệu lực của url, thử lại => expired.

Copyright@Linh Nguyen on Udemy

# S3 Storage Classes

S3 cung cấp nhiều storage class khác nhau nhằm giúp người dùng linh động trong việc lựa chọn class phù hợp với nhu cầu, tiết kiệm chi phí.

Việc lựa chọn class phụ thuộc vào các yếu tố như:

- Durability, High Availability
- Thời gian lưu trữ (1 tháng, 1 năm, 5 năm...)
- Tần suất truy cập, thời gian cần có file khi có yêu cầu
- Mục đích sử dụng: document, image, log file, backup file, archive

Copyright@Linh Nguyen on Udemy

# S3 Storage Classes

- **S3 Standard:** loại mặc định khi tạo object mà không chỉ định classes. Phù hợp cho hầu hết các usecase.
- **S3 Intelligent Tiering:** Monitor tần suất access của các object một cách tự động để move xuống các class rẻ tiền hơn giúp tiết kiệm chi phí. Chỉ apply cho object  $\geq 128\text{KB}$ . KH phải chịu thêm chi phí monitor.
- S3 standard infrequently access (**Standard IA**): phù hợp cho các data ít khi được access nhưng khi request cần có ngay. Availability 99.9% (nhỏ hơn standard 99.99%).
- S3 One-zone infrequently access (**One zone IA**): rẻ hơn standard IA 20% do chỉ lưu trữ trên 1 AZ. Phù hợp cho các data có thể dễ dàng tạo ra nếu không may bị mất (report, file image resized). Availability 99.5%

# S3 Storage Classes

- **S3 Glacier:** phù hợp cho việc lưu trữ những data có yêu cầu thời gian lưu trữ lên tới vài năm nhưng ít khi được sử dụng. Tùy theo nhu cầu khi access mà Glacier lại chia ra 1 số sub class:
  - **Glacier Instant Retrieval:** rẻ hơn tới 68% so với S3 Standard IA. Cho phép access 1 file với thời gian ngắn khi có nhu cầu. VD hồ sơ phim chụp của bệnh nhân ở bệnh viện rất ít khi cần lục lại nhưng muốn xem phải có ngay.
  - **Glacier Flexible Retrieval (Normal Glacier):** phù hợp cho data không yêu cầu access ngay hoặc chưa rõ, thời gian cần để access file có thể từ vài phút tới vài giờ. Phù hợp cho việc lưu data backup hoặc archive.
  - **S3 Glacier Deep Archive:** phù hợp cho việc lưu trữ lâu dài lên tới 7-10 năm tùy theo tiêu chuẩn ngành như tài chính, y tế,... Data được lưu trên các băng đĩa từ (magnetic tap). AWS cam kết có thể access data trong vòng 12h khi cần.
- **S3 on Outposts:** Cho phép sử dụng S3 ở on-premise.



# S3 Storage Classes

So sánh các Storage Classes

Copyright@Linh Nguyen on Udemy

	S3 Standard	S3 Intelligent-Tiering*	S3 Standard-IA	S3 One Zone-IA†	S3 Glacier Instant Retrieval	S3 Glacier Flexible Retrieval	S3 Glacier Deep Archive
Designed for durability	99.999999999% (11 9's)	99.999999999% (11 9's)	99.999999999% (11 9's)	99.999999999% (11 9's)	99.999999999% (11 9's)	99.999999999% (11 9's)	99.999999999% (11 9's)
Designed for availability	99.99%	99.9%	99.9%	99.5%	99.9%	99.99%	99.99%
Availability SLA	99.9%	99%	99%	99%	99%	99.9%	99.9%
Availability Zones	≥3	≥3	≥3	1	≥3	≥3	≥3
Minimum capacity charge per object	N/A	N/A	128 KB	128 KB	128 KB	40 KB	40 KB
Minimum storage duration charge	N/A	N/A	30 days	30 days	90 days	90 days	180 days
Retrieval charge	N/A	N/A	per GB retrieved	per GB retrieved	per GB retrieved	per GB retrieved	per GB retrieved
First byte latency	milliseconds	milliseconds	milliseconds	milliseconds	milliseconds	minutes or hours	hours

# Lab 5 – Practice Storage Class

*\*Lưu ý không upload file quá lớn.*

1. Upload 1 file và chọn storage class là Standard IA
2. Upload 1 file và chọn storage class là OneZone IA
3. Upload 1 file và chọn storage class là Glacier Instant Retrieval
4. Upload 1 file và chọn storage class là Glacier Flexible Retrieval
5. Access 1 file được lưu trữ dưới class Glacier Deep Archive
6. Xóa các file đã upload

Copyright@Linh Nguyen on Udemy

## S3 Lyfe Cycle

- Tính năng cho phép tự động move object xuống các class lưu trữ thấp hơn hoặc xoá luôn sau một khoảng thời gian nhằm tiết kiệm chi phí.
- Khác với Intelligent Tiering, người dùng sẽ tự quyết định life cycle cho objects (hoặc 1 thư mục), vd sau 90 ngày thì cho xuống Glacier, sau 270 ngày thì xoá hoàn toàn.
- Phù hợp cho các bài toán lưu trữ Log đã biết trước thời gian thường xuyên access và thời gian có thể xoá.

Copyright@Linh Nguyen on Udemy

## Lab 4 - S3 Lyfe Cycle

1. Tạo 1 life cycle rule move object trong 1 folder vd /log xuống Glacier sau 90 ngày.
2. Tạo 1 life cycle rule xoá hoàn toàn object trong /log sau 270 ngày.

Copyright@Linh Nguyen on Udemy

# S3 Static Website Hosting

- S3 có hỗ trợ người dùng host 1 website tĩnh (chỉ bao gồm html, css, js, image...)
- Được thừa hưởng toàn bộ đặc tính của S3 (Durability, HA)
- Không cần duy trì server, giảm effort Administration.
- Hỗ trợ setting CORS nhằm tránh tài nguyên bị khai thác bởi website khác.
- Kết hợp với dịch vụ CDN (CloudFront) có thể giúp tăng tốc độ truy cập khi user nằm ở các region khác nhau.

*\*Hầu hết các framework frontend hiện nay như Angular, Vue, NodeJS đều hỗ trợ build ra 1 website tĩnh để có thể deploy lên S3 sau khi code xong.*

## Lab 6 – S3 Static Website Hosting

1. Chuẩn bị 1 bộ source HTML tĩnh (có thể là static files build ra từ Angular/NodeJS...).
2. Upload source lên s3 (lưu ý file index.html phải ở level gốc)
3. Bật s3 static web hosting
4. Cấu hình open public access cho bucket
5. Cấu hình bucket policy cho phép mọi người access.
6. Test truy cập từ trình duyệt.

Copyright@Linh Nguyen on Udemy

# S3 event trigger

- S3 cung cấp cơ chế trigger 1 event sang dịch vụ khác khi có thay đổi đối với object (upload, delete)
- Target của trigger có thể là Lambda Function, SNS, SQS.
- Sample usecase
  - Resize image khi có người upload image lên s3 bucket, lưu vào các thư mục size khác nhau
  - Giải nén 1 file zip khi có người upload.
  - Extract csv file, xử lý data rồi lưu vào database
  - Notification tới Operator khi có ai xóa 1 file
  - ...

Copyright@Linh Nguyen on Udemy



# Lab 7 – Combine S3 event with Lambda

1. Tạo 1 lambda function đơn giản với chức năng in ra event nhận được từ S3 bao gồm tên bucket, object key (code python)  
\*Lưu ý vì chưa học tới bài lambda nên code simple nhất có thể, chủ yếu để các bạn nắm concept.
2. Thiết lập s3 trigger sang lambda khi có ai đó upload file.
3. Test upload 1 file bất kì.
4. Check xem lambda có được kích hoạt và chạy thành công không.

Copyright@Linh Nguyen on Udemy

# Best practices for S3

- Chọn region của S3 cùng region với application (EC2, ECS) để tối ưu performance.
- Sử dụng bucket policy cho những data quan trọng. Cấp quyền vừa đủ cho user/role, hạn chế cấp S3FullAccess.
- Bật versioning để bảo vệ data tránh bị mất, xóa nhầm.
- Mã hoá data nhạy cảm (client side or server side).
- Enforce TLS để yêu cầu sd HTTPS khi truyền nhận file (chống hack).
- Sử dụng VPC endpoint để tăng tốc truy cập từ application (sẽ học ở bài VPC).
- Khi host static web, nên kết hợp với CloudFront để tối ưu chi phí và tăng trải nghiệm người dùng.

# Clear resources

Trên AWS console, thực hiện nội dung sau:

1. Xóa Toàn bộ S3 bucket đã tạo ra trong lab (hoặc chỉ xóa những file lớn trong bucket).
2. Terminate EC2 Instance.
3. Lambda có thể giữ lại nhằm mục đích tham khảo trong tương lai (vì nếu không chạy thì không tốn tiền)

Copyright@Linh Nguyen on Udemy