CloudNut
Learn Clouds in a Nutshell

Giới thiệu về S3

Giới thiệu về S3?

CloudNut
Learn Clouds in a Nutshell

- S3 viết tắt của **S**imple **S**torage **S**ervice.
- S3 là dịch vụ lâu đời và quan trọng của AWS
- S3 là dịch vụ lưu trữ dữ liệu dưới dạng **Object (Object Storage)**
- Lưu trữ không giới hạn (unlimited storage)
- Sử dụng để lưu dữ liệu dạng file (text, image, video, media…)

Khái niệm tổng quan - Bucket

CloudNut
Learn Clouds in a Nutshell

- **Bucket** tương tự như một thư mục (directory)
- Tên của **Bucket** phải là duy nhất (không tồn tại 2 Bucket có tên trùng nhau)
- Bucket có scope là **Region** (dữ liệu của một S3 Bucket sẽ được nhân bản trong **Region**)

Khái niệm tổng quan - Objects

CloudNut
Learn Clouds in a Nutshell

- Objects tương tự như các files
- Mỗi Object sẽ bao gồm:
  - Key (Tên của Object)
    - s3://bucket_name/my_folder/another_folder/my_file
      Key
    - s3://bucket_name/my_folder/another_folder/my_file
      Prefix
  - Value (Dữ liệu của Object)
    - Kích thước dữ liệu của 1 **Object** từ 0 Bytes to 5 TB.
    - Trong trường hợp **Object** có kích thước > 5GB thì sử dung **multi-part upload** để upload Object này lên S3

Khái niệm tổng quan - Objects

CloudNut
Learn Clouds in a Nutshell

- Mỗi Object sẽ bao gồm:
  - Metadata
    - Dữ liệu mô tả về Object này (Được thiết lập tại thời điểm upload Object, sau đó không thể thay đổi).
    - Ex: Thời gian Upload Object, Storage class, dữ liệu có được mã hoá (Encryption) hay không?)
  - Version ID
    - Phục vụ tính năng Versioning

CloudNut
Learn Clouds in a Nutshell

S3 Storage Tier

S3 storage tier

CloudNut
Learn Clouds in a Nutshell

	S3 Standard	S3 Intelligent-Tiering*	S3 Standard-IA	S3 One Zone-IA†	S3 Glacier	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)
Designed for availability	99.99%	99.9%	99.9%	99.5%	99.99%	99.99%
Availability SLA	99.9%	99%	99%	99%	99.9%	99.9%
Availability Zones	≥3	≥3	≥3	1	≥3	≥3
Minimum capacity charge per object	N/A	N/A	128KB	128KB	40KB	40KB
Minimum storage duration charge	N/A	30 days	30 days	30 days	90 days	180 days
Retrieval fee	N/A	N/A	per GB retrieved	per GB retrieved	per GB retrieved	per GB retrieved
First byte latency	milliseconds	milliseconds	milliseconds	milliseconds	select minutes or hours	select hours
Storage type	Object	Object	Object	Object	Object	Object
Lifecycle transitions	Yes	Yes	Yes	Yes	Yes	Yes

S3 Standard

CloudNut
Learn Clouds in a Nutshell

- Storage tier mặc định
- 99.99% Availibility (xác suất một yêu cầu lấy dữ liệu của Object được trả về thành công)
- 99.999999999% i.e. 11 9's Durability (Xác suất dữ liệu tồn tại, được đảm bảo lưu trữ ở S3)
- Phục vụ cho các mục đích cơ bản.
  - Phù hợp cho các ứng dụng yêu cầu độ trễ thấp cho tác vụ lưu trữ files
  - Các ứng dung truy cập dữ liệu thường xuyên (Frequently Access) vào S3.

S3 Standard IA (Infrequently Access)

CloudNut
Learn Clouds in a Nutshell

- Dùng để lưu dữ liệu ít thường xuyên truy cập (**Infrequently Access)** (Khoảng 1 tháng 1 lần)
- Chi phí lưu trữ rẻ hơn **Standard Tier**
- Bị mất thêm chi phí cho việc lấy Object (Object retrieving)
- Objects có thể truy cập nhanh (real time).
- Object phải đảm bảo > 128 KB
- Chi phí được tính làm tròn đến 30 ngày (Nếu Object được lưu < 30 ngày thì sẽ được tính tròn lên 30 ngày)

S3 One-Zone IA (Infrequently Access)

CloudNut
Learn Clouds in a Nutshell

- Dùng để lưu dữ liệu ít thường xuyên truy cập (Infrequently Access)
- Dữ liệu của Object sẽ được lưu trư trong một **AZ (Availibility Zone)**
- Object phải đảm bảo > 128 KB
- Dùng cho dữ liệu có thể khôi phục được nếu như một AZ bị sự cố
- Phí được tính làm tròn lên 30 ngày.

S3 One-Zone IA

S3 Intelligent Tier

CloudNut
Learn Clouds in a Nutshell

- Tự động di chuyển dữ liệu giữa các **Storate Tier** để tối ưu chi phí.
- Dựa vào tần suất truy cập dữ liệu để chọn **Storage Tier** phù hợp
- Phù hợp cho các kiểu dữ liệu có tần suất truy cập không đoán trước được

S3 Intelligent-Tiering

S3 Glacier

CloudNut
Learn Clouds in a Nutshell

- Dùng để lưu trữ dữ liệu lâu dài 3 ~ 10 năm (Long term archive)
- Chi phí rẻ
- Lưu trữ tối thiểu trong vòng 90 ngày
- Thời gian lấy dữ liệu Object từ vài phút tới vài giờ

S3 Glacier

S3 Glacier Deep Archive

CloudNut
Learn Clouds in a Nutshell

- Dùng để lưu trữ dữ liệu lâu dài trên 10 năm (Long term archive)
- Chi phí rất rẻ
- Thời gian lưu trữ tối thiểu là 180 ngày
- Thời gian lấy dữ liệu của Object khoảng trên 12 giờ

S3 Glacier Deep Archive

Knowledge Check

CloudNut
Learn Clouds in a Nutshell

Question: You work for a health insurance company that amasses a large number of patients' health records. Each record will be used once when assessing a customer, and will then need to be securely stored for a period of 7 years. In some rare cases, you may need to retrieve this data within 24 hours of a claim being lodged. Given these requirements, which type of AWS storage would deliver the least expensive solution?

A. S3

B. S3-IA

C. S3-1Zone-IA

D. Glacier

Knowledge Check

CloudNut
Learn Clouds in a Nutshell

Question: You work for a major news network in Europe. They have just released a new mobile app that allows users to post their photos of newsworthy events in real-time, which are then reviewed by your editors before being copied to your website and made public. Your organization expects this app to grow very quickly, essentially doubling its user base each month. The app uses S3 to store the images, and you are expecting sudden and sizable increases in traffic to S3 when a major news event takes place (as users will be uploading large amounts of content.) You need to keep your storage costs to a minimum, and it does not matter if some objects are lost. With these factors in mind, which storage media should you use to keep costs as low as possible?

A. S3 Infrequently Accessed

B. S3 One Zone-Infrequent Access

C. Glacier

D. S3 Provisioned IOPS

Exam tips

CloudNut
Learn Clouds in a Nutshell

- Dữ liệu cần độ trễ truy cập dữ liệu thấp (latency sensitive), truy cập thường xuyên (frequently access) => **Standard Tier**
- Dữ liệu ít truy cập (Infrequently Access) => **IA**
- Dữ liệu ít truy cập (Infrequently Access), dữ liệu có thể khôi phục nếu một AZ bị sự cố => **One-Zone IA**
- Dữ liệu lưu trữ lâu dài (1 ~ 10 năm), thời gian lấy dữ liệu từ phút tới < 12 giờ => **Glacier**
- Dữ liệu lưu trữ lâu dài (> 10 năm), thời gian lấy dữ liệu > 12 hours => **Deep Archived**

CloudNut
Learn Clouds in a Nutshell

Security and Encryption

S3 security

CloudNut
Learn Clouds in a Nutshell

- **Identity-based policy**
  - IAM policies – Sử dụng IAM policies để đinh nghĩa quyền cho các thực thể (IAM identities – user, group, role) được gắn policy này.
- **Resource-based policy**
  - Bucket Policies – Sử dụng để định nghĩa ai có quyền truy cập Bucket này
  - Access Control Lists – Định nghĩa quyền truy cập cho các **Objects**

NOTE: Sử dụng **Policy Evaluation Logic** để xác định một thực thể có quyền truy cập vào resource hay không.

S3 security (cont.)

CloudNut
Learn Clouds in a Nutshell

- **Networking**
  - Hỗ trợ **VPC Endpoint** cho các kết nối cần bảo mật riêng tư
- **Logging and Audit**
  - Các lời gọi API tới S3 được ghi lại bởi dịch vụ **Cloudtrail**
  - Các Logs liên quan tới truy cập S3 được lưu lại
- **User Security**
  - MFA Delete: Yêu cầu MFA cho hành động xoá Object. Nhằm tránh việc không may xoá Object
  - Pre-Signed URLs: Tạo URL cho việc truy cập Object trong một thời gian giới hạn

S3 encryption

CloudNut
Learn Clouds in a Nutshell

- Mã hoá đường truyền (Encryption in Transit)
  - SSL/TLS
- Mã hoá lưu trữ (Encryption at Rest)
  - S3 Managed Keys – SSE-S3
  - KMS Managed Keys – SSE-KMS
  - Customer Managed Keys – SSE-C
- Mã hoá phía Client (Client side encryption)
  - Object được mã hoá trước khi upload lên S3

CloudNut
Learn Clouds in a Nutshell

S3 versioning

S3 versioning

CloudNut
Learn Clouds in a Nutshell

- Lưu lại tất cả các phiên bản của một Object (Bao gồm các hành động update/tạo mới Objects)
- Khi tính năng được **enable**, thì không thể **disable** được. Chỉ có thể **suspended**
- Versioning có thể kết hợp với các luật của quản lý vòng đời của Object (**Life Cycle Managemen)**

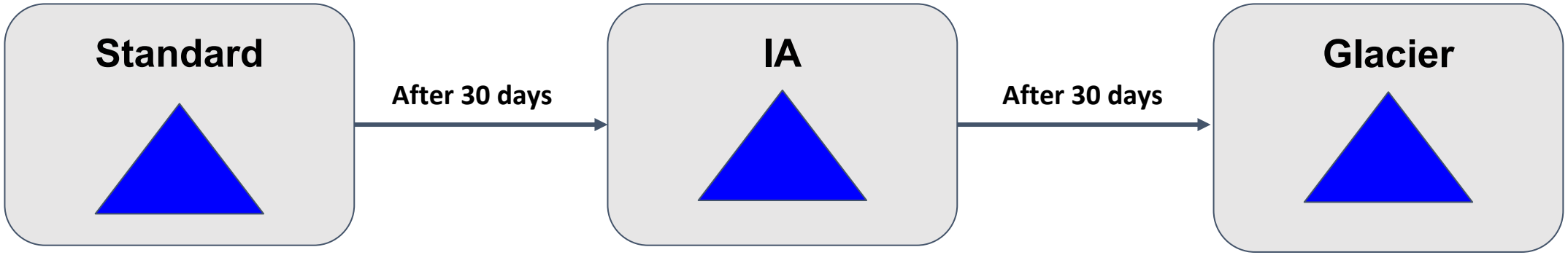
CloudNut
Learn Clouds in a Nutshell

Quản lý vòng đời của S3 Object (S3 Lifecycle Management)

S3 Lifecycle Management

CloudNut
Learn Clouds in a Nutshell

- Tự động di chuyển Objects giữa các Storaga Tier (Class) khác nhau
- Đảm bảo chi phí cho việc lưu trữ được sử dụng hiệu quả (Cost effective)
- Có thể kết hợp sử dụng với tính năng **Versioning**



CloudNut
Learn Clouds in a Nutshell

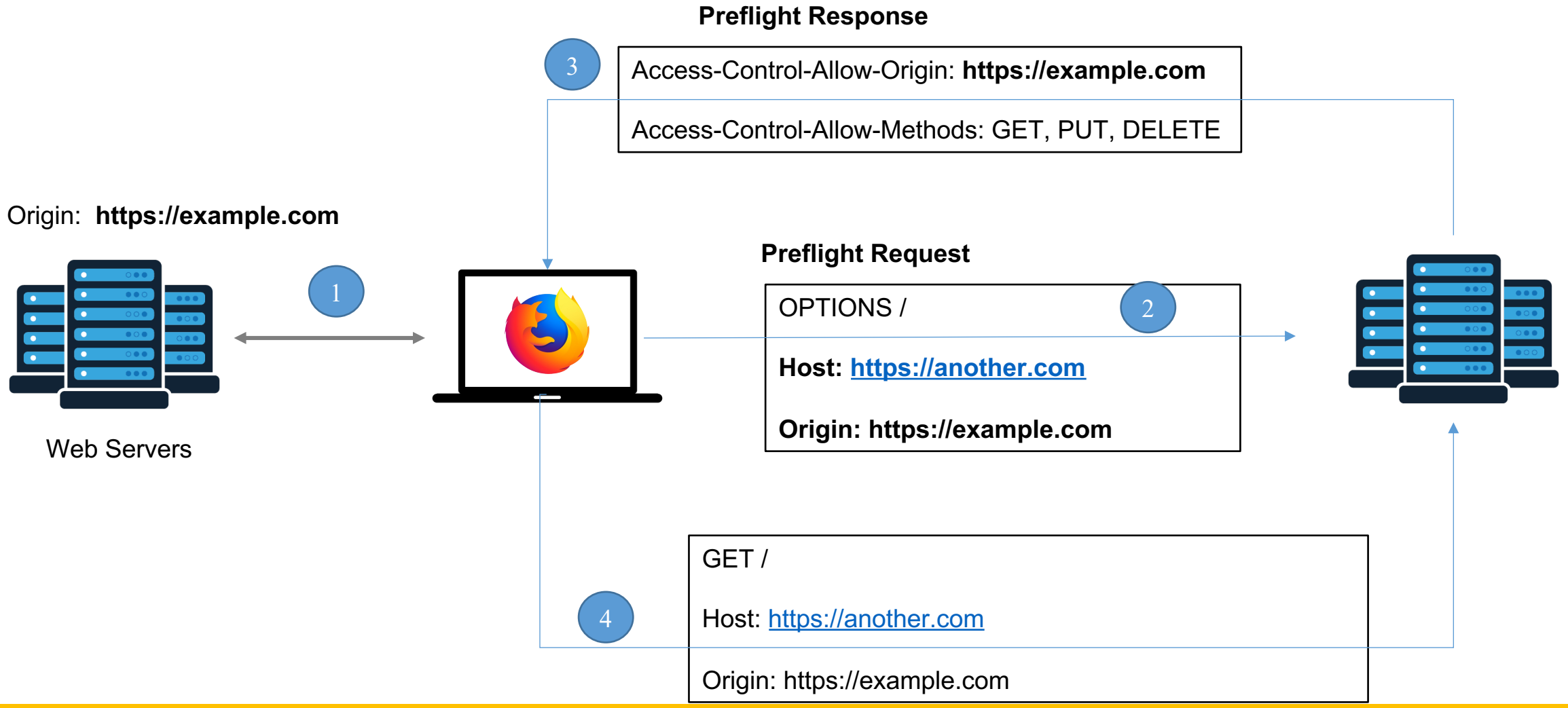
S3 CORS

CORS

CloudNut
Learn Clouds in a Nutshell

- Origin: <scheme> "://" <hostname> ":" <port> (Ex: https://example.com)
- CORS viết tắt của **C**ross-**O**rigin **R**esource **S**haring
- Cùng Origin: https://example.com/site1, https://example.com/site2
- Khác Origin: https://example.com/site1, https://another.com/app1
- Các requests phải được cho phép bởi Target Origin sử dụng CORS header (Ex: **Access-Controll-Allow-Origin**)

CORS (cont.)


Preflight Response
3
Access-Control-Allow-Origin: https://example.com
Access-Control-Allow-Methods: GET, PUT, DELETE
Origin: https://example.com
Preflight Request
1
OPTIONS /
2
Host: https://another.com
Origin: https://example.com
Web Servers
4
GET /
Host: https://another.com
Origin: https://example.com

S3 CORS

CloudNut
Learn Clouds in a Nutshell

- Sử dụng khi gọi tới S3 API endpoint cho các yêu cầu GET/PUT từ Webpage

`http://website.s3-website.us-east-1.amazonaws.com`

`website.s3.us-east-1.amazonaws.com`

CloudNut
Learn Clouds in a Nutshell

S3 performance

S3 baseline performance

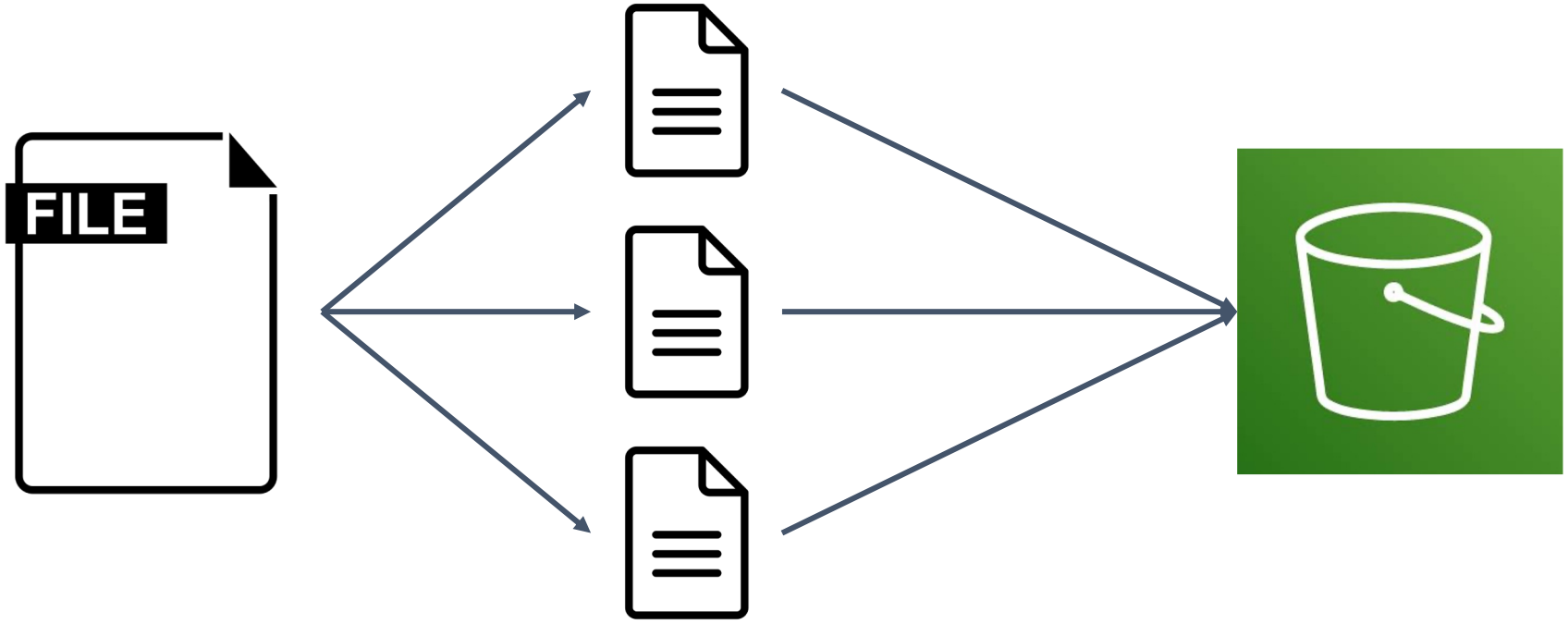
CloudNut
Learn Clouds in a Nutshell

- Mỗi **prefix** cho phép 3500 PUT/COPY/POST/DELETE và 5500 GET/HEAD request trong một **Bucket**
- Không giới hạn số lượng prefix trong một **Bucket**
- Ví dụ về **prefix**.
  - s3://bucket/**folder2/sub2/**file2.txt
- Càng nhiều **prefix** thì tốc độ hiệu năng đọc ghi dữ liệu của **Bucket** càng cao

S3 - Multipart upload

CloudNut
Learn Clouds in a Nutshell

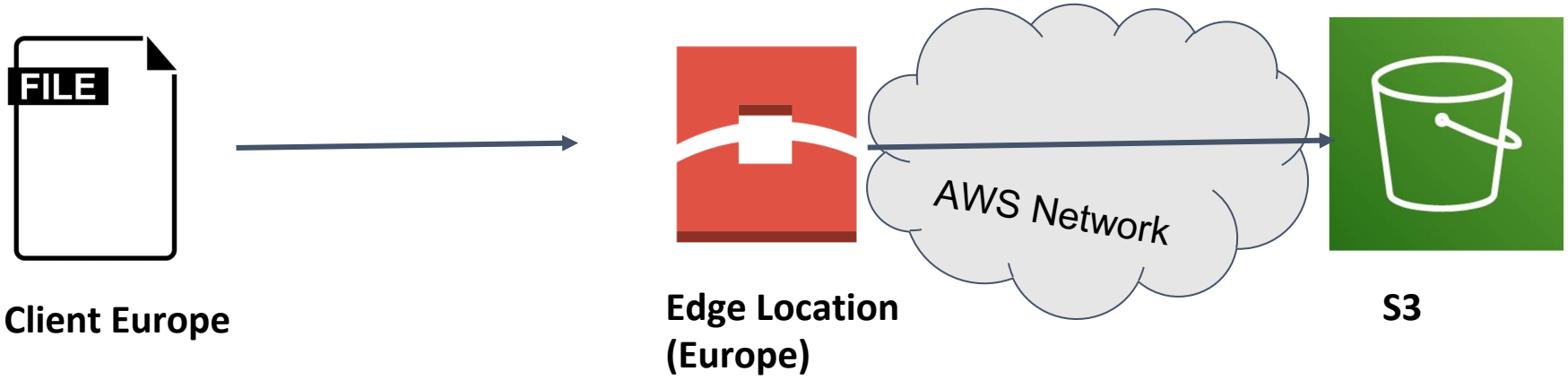
- Bắt buộc phải sử dụng cho Object có dung lượng > 5GB
- Giúp tối ưu thông lượng nhờ việc upload song song



S3 - Transfer Acceleration

CloudNut
Learn Clouds in a Nutshell

- Sử dụng cho việc upload Objects lên S3
- Sử dụng **Edge Location** như là các điểm trung gian giữa S3 và Client

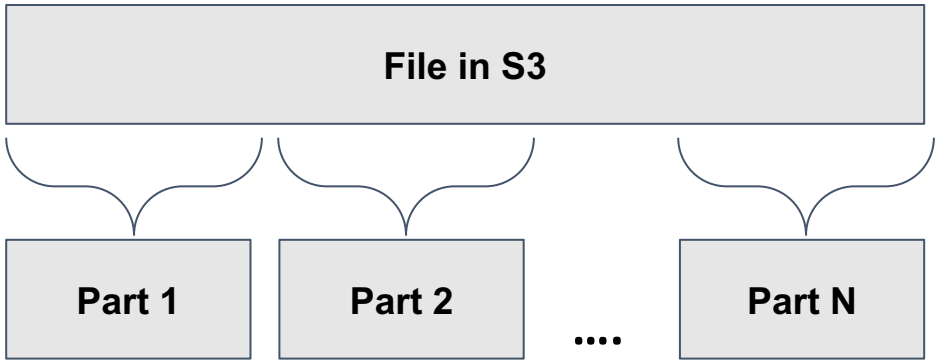


S3 - Byte-Range Fetches

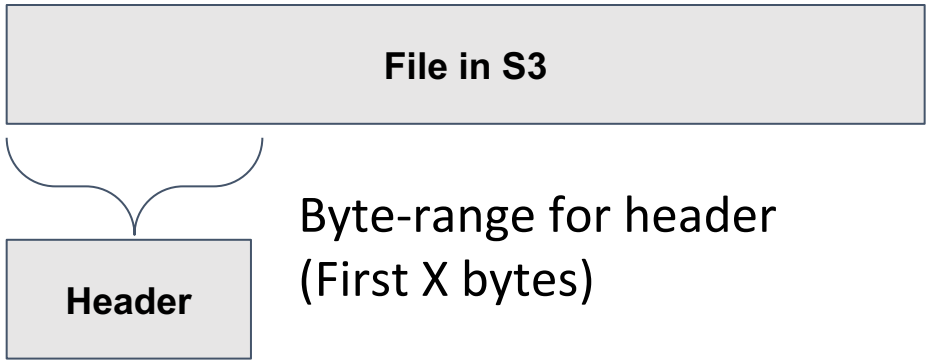
CloudNut
Learn Clouds in a Nutshell

- Sử dụng Range HTTP header trong request để download một phần byte-range của Object
- Tạo đồng thời nhiều kết nối để download song song nhiều byte-range của Object
- Giúp tăng tốc độ download Object

Speed up downloads



Download partial data



CloudNut
Learn Clouds in a Nutshell

Knowledge Check

Knowledge Check

CloudNut
Learn Clouds in a Nutshell

Which of the following default settings are INCORRECT for a newly created S3 bucket? (choose 2 options)

A. Encryption is not enabled

B. Transfer Acceleration is enabled.

C. No bucket policy exists

D. Versioning is enabled.

Knowledge Check

CloudNut
Learn Clouds in a Nutshell

You have created an S3 bucket in the us-east-1 region with default configuration. Versioning is not enabled. You are located in Asia and deleted an object in the bucket using AWS CLI. What may happen when you try to list the objects in the bucket?

A. The object is still there.

B. The object is deleted completely.

C. The object may still be there or deleted, depending on whether the deletion is finished in AWS.

D. The object is attached with a deletion mark.

Ref: https://aws.amazon.com/blogs/aws/amazon-s3-update-strong-read-after-write-consistency/

Knowledge Check

CloudNut
Learn Clouds in a Nutshell

You have an application running on EC2. When the application trying to upload a 7 GB file to S3, the operation fails. What could be the reason for failure, and what would be the solution?

A. With a single PUT operation, you can upload objects up to 5 GB in size. Use multi-part upload for larger file uploads

B. EC2 is designed to work best with EBS volumes. Use EBS Provisioned IOPs and use an Amazon EBS- optimized instance.

C. NAT gateway only supports data transfers going out upto 5 GB. Use EBS Provisioned IOPs and use an Amazon EBS-optimized instance.

D. VPC Endpoints only supports data transfers going out upto 5 GB. Use EBS Provisioned IOPs and use an Amazon EBS-optimized instances

Knowledge Check

CloudNut
Learn Clouds in a Nutshell

You work for a busy digital marketing company who currently store their data on-premise. They are looking to migrate to AWS S3 and to store their data in buckets. Each bucket will be named after their individual customers, followed by a random series of letters and numbers. Once written to S3 the data is rarely changed, as it has already been sent to the end customer for them to use as they see fit. However, on some occasions, customers may need certain files updated quickly, and this may be for work that has been done months or even years ago. You would need to be able to access this data immediately to make changes in that case, but you must also keep your storage costs extremely low. The data is not easily reproducible if lost. Which S3 storage class should you choose to minimize costs and to maximize retrieval times?

A. S3

B. S3-IA

C. S3-1Zone-IA

D. Glacier