AWS EFS

Elastic File System

Managed NFS for Linux workloads with auto-scaling, high availability, and pay-per-use pricing

Managed Storage

NFS Compatible

Auto-Scaling

Security & Compliance

Tổng quan về EFS

Định nghĩa

EFS (Elastic File System) là dịch vụ lưu trữ **Network File System (NFS)** được quản lý hoàn toàn bởi AWS.

NFS là giao thức cho phép nhiều máy chủ (như EC2) truy cập và chia sẻ cùng một hệ thống tệp qua mạng, tương tự như việc nhiều người cùng mở và chỉnh sửa file trên một ổ đĩa mạng.

Khả năng gắn kết (Mount)

- Có thể gắn kết (mount) lên nhiều EC2 Instances trong cùng một VPC
- Có thể truy cập từ máy chủ on-premises thông qua AWS Direct Connect (DX) hoặc AWS VPN
- Cho phép các kịch bản đám mây lai và di chuyển dữ liệu

Tự động mở rộng

EFS tự động mở rộng và thu nhỏ dung lượng lưu trữ khi bạn thêm hoặc xóa tệp

Sẵn sàng cao

Lưu trữ dữ liệu trên nhiều Availability Zones (AZs) trong cùng một Region

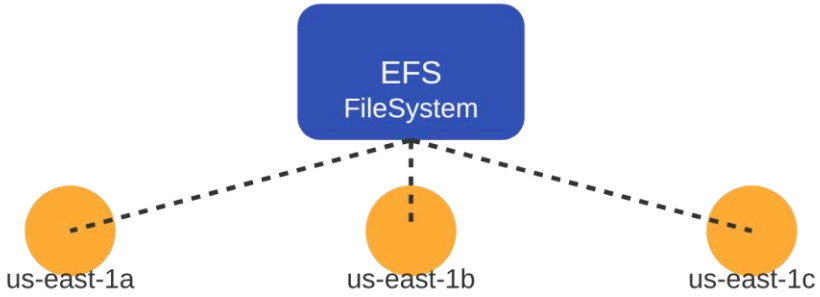
Pay-per-use

Chỉ thanh toán cho dung lượng lưu trữ thực tế mà bạn sử dụng

Chi phí

Chi phí của EFS thường cao hơn khoảng 3 lần so với gp2 EBS

Kiến trúc EFS với 3 Availability Zones



Kiến trúc và đặc điểm của EFS

Trường hợp sử dụng

Web serving
WordPress, CMS

Chia sẻ dữ liệu
Giữa nhiều EC2

Quản lý nội dung
Ứng dụng phân tán

Big Data Analytics
Phân tích dữ liệu lớn

Bảo mật

Security Group
Kiểm soát truy cập

KMS Encryption
Mã hóa dữ liệu lưu trữ

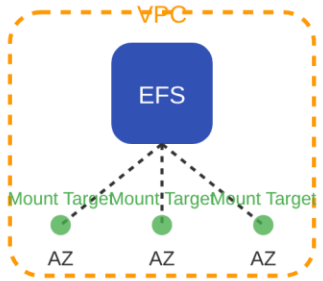
TLS 1.2
Mã hóa dữ liệu truyền tải

IAM Policies
Kiểm soát quyền truy cập

Tính tương thích

- **Linux AMI:** Chỉ hoạt động với Linux AMI (POSIX-compliant)
- **NFS Protocol:** Sử dụng giao thức NFSv4.1 và NFSv4.0

Hạ tầng



Ưu điểm

Hệ thống file POSIX
Tiêu chuẩn Linux

Tự động mở rộng
Không cần hoạch định dung lượng

Hiệu năng và Throughput

Khả năng mở rộng (Scale)

- Hàng nghìn client NFS truy cập đồng thời
- Tốc độ throughput lên đến hơn 10 GB/s
- Quy mô lưu trữ mở rộng đến hàng Petabyte
- Tự động tăng/giảm dung lượng khi thêm/xóa tệp

Hỗ trợ tối đa: **3 GiB/s**

cho các thao tác đọc

Hỗ trợ tối đa: **1 GiB/s**

cho các thao tác ghi

Performance Mode

General Purpose

Chế độ mặc định, khuyến nghị cho hầu hết các trường hợp sử dụng

Độ trễ thấp nhất cho mỗi thao tác

Lý tưởng cho: web serving, CMS, thư mục người dùng

Max I/O

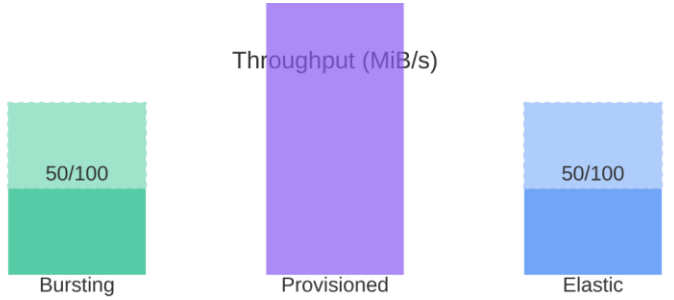
Tối ưu cho các khối lượng công việc có tính song song cao

Tổng throughput và IOPS cao

Lý tưởng cho: phân tích dữ liệu lớn, xử lý media

Chế độ này có thể có độ trễ mỗi thao tác cao hơn General Purpose

Throughput Mode



Bursting

Throughput tăng theo dung lượng lưu trữ

1TB = 50MiB/s + burst lên 100MiB/s

Provisioned

Cấp phát throughput cố định (MiB/s)

VD: 1 GiB/s cho 1TB dữ liệu

Elastic

Chế độ mặc định, tự động điều chỉnh throughput

Thanh toán chỉ cho dữ liệu thực tế

Storage Classes & Lifecycle

Các lớp lưu trữ (Storage Tiers)

EFS Standard

- Xây dựng trên bộ nhớ SSD tốc độ cao
- Độ trễ dưới mili giây
- Lớp mặc định cho dữ liệu hoạt động

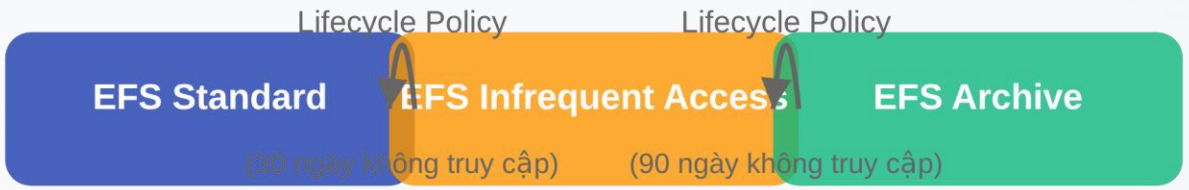
EFS Infrequent Access

- Chi phí lưu trữ thấp hơn 95%
- Độ trễ cao hơn (hàng chục mili giây)
- Dùng cho dữ liệu ít truy cập (vài lần mỗi quý)

EFS Archive

- Chi phí lưu trữ thấp hơn 97% so với Standard
- Chi phí thấp hơn 50% so với EFS-IA
- Dùng cho dữ liệu hiếm khi truy cập

Lifecycle Management



Chính sách chuyển đổi vào IA

Di chuyển tệp vào lớp Infrequent Access khi không được truy cập trong khoảng thời gian nhất định (30 ngày mặc định)

Chính sách chuyển đổi vào Archive

Di chuyển tệp vào lớp Archive khi không được truy cập trong khoảng thời gian dài hơn (90 ngày mặc định)

Tùy chọn khả dụng

Regional (Multi-AZ)

Lưu trữ dữ liệu dự phòng trên nhiều AZs, cung cấp độ bền cao nhất

One Zone

Lưu trữ trong một AZ duy nhất, giảm chi phí (rẻ hơn 47% so với Regional)

One Zone-IA

Kết hợp One Zone với Infrequent Access, tiết kiệm chi phí đáng kể

EFS Access Points

Khái niệm

EFS Access Points là tính năng của Amazon EFS cung cấp các điểm truy cập tùy chỉnh vào hệ thống tệp của bạn, giúp quản lý quyền truy cập NFS cho từng ứng dụng hoặc nhóm người dùng một cách dễ dàng và hiệu quả.

Cấu hình

- Mỗi Access Point có thể được cấu hình để thực thi một danh tính người dùng/nhóm POSIX (UID/GID) cụ thể
- Mỗi Access Point có thể được cấu hình với một thư mục gốc (root directory) riêng biệt
- Khi một client kết nối thông qua Access Point, nó sẽ truy cập vào thư mục gốc với quyền của người dùng/được chỉ định
- Có thể kết hợp với IAM Policies để kiểm soát truy cập chi tiết hơn

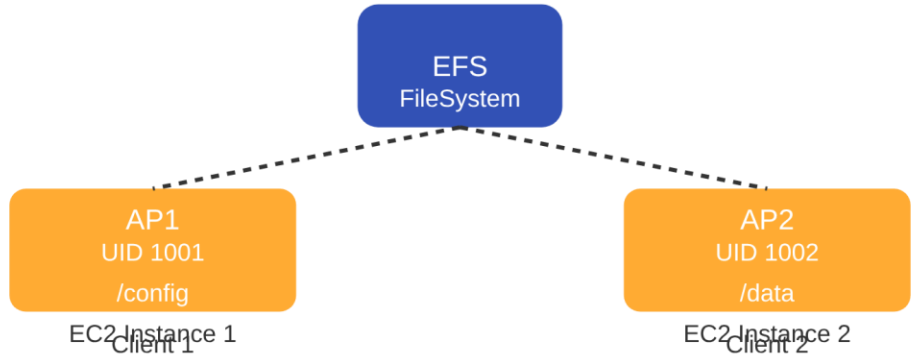
Ví dụ thực tế

Access Point 1

UID: 1001

Thư mục gốc: /config

Cách hoạt động của Access Points



Ứng dụng thực tế

Môi trường đa người dùng

Phân quyền rõ ràng cho các nhóm khác nhau (developers, analysts) khi họ chia sẻ cùng một hệ thống tệp

Môi trường đa ứng dụng

Các ứng dụng khác nhau có thể truy cập cùng một hệ thống tệp nhưng với các quyền và đường dẫn khác nhau

Bảo mật nâng cao

Kết hợp với IAM Policies để kiểm soát truy cập chi tiết hơn, cho phép xác định IAM

File System Policies

Định nghĩa

File System Policies trong Amazon EFS là các chính sách dựa trên tài nguyên (resource-based policy) cho phép kiểm soát quyền truy cập vào hệ thống tệp EFS.

Chúng hoạt động tương tự như S3 Bucket Policy, cho phép bạn định nghĩa chi tiết ai có thể thực hiện hành động gì trên tài nguyên EFS của bạn.

Cơ chế hoạt động

- Mặc định: Amazon EFS cho phép tất cả các client trong cùng một VPC truy cập vào hệ thống tệp
- Tùy chỉnh: Bạn có thể giới hạn quyền đọc/ghi cho từng IAM User hoặc Role cụ thể
- Thêm lớp kiểm soát truy cập dựa trên danh tính, bổ sung cho các kiểm soát mạng thông qua Security Group

Ví dụ Policy

Dưới đây là một ví dụ về File System Policy cho phép một IAM user cụ thể thực hiện các hành động ClientMount

Cross-Region Replication

Khái niệm

EFS Cross-Region Replication là tính năng cho phép tự động sao chép dữ liệu và metadata từ một hệ thống tệp EFS nguồn sang một hệ thống tệp EFS đích ở một AWS Region khác.

Đặc điểm

- Áp dụng linh hoạt: Có thể thiết lập cho cả các hệ thống tệp EFS mới hoặc hiện có
- RPO/RTO thấp: Cung cấp Recovery Point Objective (RPO) và Recovery Time Objective (RTO) chỉ trong vài phút
- Không ảnh hưởng hiệu năng: Quá trình sao chép không tiêu tốn burst credits và không tính vào throughput đã provisioned
- Đồng bộ liên tục: EFS tự động duy trì đồng bộ hóa giữa hệ thống tệp nguồn và đích

Trường hợp sử dụng

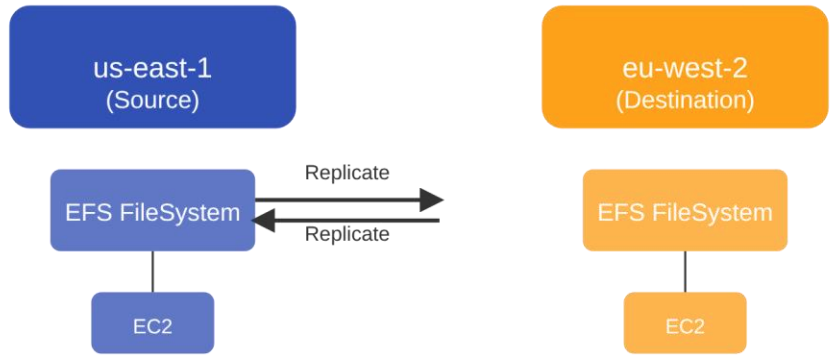
Tuân thủ bảo mật & dự phòng

Đảm bảo dữ liệu được sao lưu ở một

Kế hoạch khôi phục thảm họa

Trong trường hợp xảy ra thảm họa tại

Sơ đồ minh họa



Lợi ích

- Tăng cường khả năng phục hồi
- Bảo vệ dữ liệu trước thảm họa
- Đáp ứng yêu cầu tuân thủ

Cấu hình đơn giản

- Chọn Region nguồn và đích
- Cấu hình chính sách sao chép
- Quá trình tự động hóa

Tổng kết

EFS = Managed NFS cho Linux workloads

EFS hoạt động như một Network File System (NFS) được quản lý, tương thích với các ứng dụng Linux (POSIX-compliant), cho phép nhiều EC2 Instances truy cập đồng thời vào cùng một hệ thống tệp.

Hỗ trợ multi-AZ, auto scaling, encryption

EFS được thiết kế để có tính sẵn sàng cao bằng cách lưu trữ dữ liệu trên nhiều Availability Zones, tự động mở rộng dung lượng và hiệu suất, hỗ trợ mã hóa dữ liệu bằng AWS KMS và kiểm soát truy cập thông qua IAM policies.

Nhiều chế độ hiệu năng và lớp lưu trữ

EFS cung cấp các chế độ hiệu năng như General Purpose và Max I/O, cùng với các chế độ throughput như Elastic, Provisioned và Bursting. Các lớp lưu trữ (Standard, Infrequent Access, Archive) giúp tối ưu chi phí dựa trên tần suất truy cập.

Ứng dụng thực tế

EFS lý tưởng cho web serving (WordPress, CMS), chia sẻ dữ liệu giữa nhiều EC2, quản lý nội dung, ứng dụng phân tán, phân tích dữ liệu lớn và tính toán hiệu năng cao (HPC).

Chiến lược tối ưu chi phí

- Mặc dù EFS mang nhiều lợi ích nhưng chi phí có thể cao hơn gp2 EBS (khoảng 3 lần)
- Sử dụng Lifecycle Policies để tự động chuyển dữ liệu ít truy cập sang các lớp lưu trữ rẻ hơn
- Cân nhắc sử dụng One Zone file systems cho các khối lượng công việc không yêu cầu tính sẵn sàng đa AZ
- One Zone-IA kết hợp One Zone với Infrequent Access, tiết kiệm chi phí đáng kể