HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

Bùi Minh Tuấn

THIẾT KẾ VÀ TRIỂN KHAI HỆ THỐNG SAO LƯU VÀ PHỤC HỒI DỮ LIỆU PHÂN TÁN CHO DOANH NGHIỆP ĐA NỀN TẢNG

Chuyên ngành: Hệ thống thông tin

Mã số: 8.48.01.04

TÓM TẮT ĐỀ ÁN TỐT NGHIỆP THẠC SĨ

HÀ NỘI - 2026

Đề án tốt nghiệp được hoàn thành tại:

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

Người hướng dẫn khoa học: PGS.TS. Đỗ Trung Tuấn

Phản biện 1: TS. Dương Trần Đức

Phản biện 2: PGS.TS. Nguyễn Đức Dũng

Đề án tốt nghiệp được bảo vệ trước Hội đồng chấm đề án tốt nghiệp thạc sĩ tại Học viện Công nghệ Bưu chính Viễn thông

Vào lúc: 08 giờ 30 phút ngày 24 tháng 01 năm 2026

Có thể tìm hiểu đề án tốt nghiệp tại:

- Thư viện của Học viện Công nghệ Bưu chính Viễn thông

LỜI MỞ ĐẦU

Trong bối cảnh chuyển đổi số diễn ra mạnh mẽ, dữ liệu ngày càng trở thành yếu tố quyết định hoạt động và năng lực cạnh tranh của doanh nghiệp. Sự phụ thuộc lớn vào thông tin cùng với nguy cơ mất mát dữ liệu do lỗi hệ thống, con người hay tấn công mạng khiến yêu cầu xây dựng một giải pháp sao lưu an toàn, ổn định và dễ vận hành trở nên cấp thiết, đặc biệt đối với doanh nghiệp vừa và nhỏ – nơi nguồn lực kỹ thuật còn hạn chế. Từ thực tiễn đó, đề tài tập trung nghiên cứu mô hình sao lưu phân tán dựa trên công nghệ NAS và snapshot nhằm cung cấp một giải pháp khả thi, chi phí hợp lý nhưng vẫn đảm bảo mức độ bảo vệ dữ liệu cao.

1. Lý do chọn đề tài

Dữ liệu hiện đóng vai trò tương đương tài sản cốt lõi. Mọi gián đoạn hay mất mát dữ liệu đều có thể gây thiệt hại nghiêm trọng. Ransomware và các sự cố vận hành cho thấy nhiều doanh nghiệp có sao lưu nhưng vẫn không thể phục hồi. Vì vậy, cần một mô hình sao lưu phân tán có khả năng chống chịu tốt và phục hồi nhanh.

2. Tổng quan về vấn đề nghiên cứu

Các giải pháp bảo vệ dữ liệu hiện nay hướng tới snapshot hệ thống tệp, sao lưu phân tán, tối ưu RTO/RPO và tăng tính sẵn sàng. Đề tài kế thừa các hướng tiếp cận này để xây dựng mô hình phù hợp môi trường doanh nghiệp nhỏ và vừa.

3. Mục đích nghiên cứu

Mục tiêu khoa học: Làm rõ cơ sở lý luận về an toàn dữ liệu, snapshot Btrfs, replication và các phương pháp sao lưu.

Mục tiêu thực tiễn: Xây dựng mô hình sao lưu hai site độc lập, hoạt động ổn định, dễ triển khai và tiết kiệm chi phí.

4. Đối tượng và phạm vi nghiên cứu

Đối tượng là dữ liệu doanh nghiệp cùng các cơ chế snapshot, replication và sao lưu đa nền tảng. Phạm vi tập trung vào mô hình hai site tại Hà Nội – TP.HCM sử dụng NAS Synology và Active Backup for Business.

5. Phương pháp nghiên cứu

Đề tài kết hợp nghiên cứu lý thuyết và thực nghiệm: triển khai mô hình thật gồm hai cụm NAS, kiểm thử sao lưu Windows – Linux – máy ảo và mô phỏng sự cố để đánh giá khả năng phục hồi.

6. Cấu trúc luận văn

Ngoài phần Mở đầu và Kết luận, luận văn được tổ chức thành bốn chương chính, mỗi chương đảm nhiệm một vai trò riêng trong quá trình hình thành và đánh giá mô hình sao lưu phân tán.

Chương 1: Cơ sở lý luận về sao lưu và phục hồi dữ liệu trong doanh nghiệp đa nền tảng.

Chương 2: Thực trạng và phân tích hệ thống sao lưu dữ liệu.

Chương 3: Thiết kế và triển khai hệ thống sao lưu phân tán.

Chương 4: Đánh giá hiệu quả, so sánh và đề xuất cải tiến.

CHƯƠNG 1

CƠ SỞ LÝ LUẬN VỀ SAO LƯU VÀ PHỤC HỒI DỮ LIỆU TRONG DOANH NGHIỆP ĐA NỀN TẢNG

1.1. Tổng quan về dữ liệu và yêu cầu bảo vệ dữ liệu trong doanh nghiệp

1.1.1. Khái niệm và tầm quan trọng của dữ liệu

Trong bối cảnh số hóa mạnh mẽ, dữ liệu ngày nay được xem như một dạng "vốn" quan trọng, ảnh hưởng trực tiếp đến sự vận hành và khả năng phát triển của doanh nghiệp. Ở góc độ cơ bản, dữ liệu bao gồm những thông tin, số liệu hay sự kiện được ghi lại trong quá trình hoạt động thường nhật. Chúng có thể xuất hiện dưới nhiều hình thức khác nhau, từ báo cáo tài chính, hồ sơ khách hàng, tài liệu nội bộ cho đến mã nguồn phần mềm. Khi được thu thập, xử lý và phân tích đúng cách, dữ liệu trở thành tri thức, hỗ trợ doanh nghiệp nhìn nhận vấn đề chính xác hơn và củng cố năng lực ra quyết định.

Vai trò của dữ liệu thể hiện rõ trong hoạt động quản trị hiện đại. Thay vì dựa vào trực giác hay kinh nghiệm, các nhà quản lý có thể khai thác thông tin từ dữ liệu lịch sử và dữ liệu thời gian thực để hình thành chiến lược kinh doanh và tối ưu hóa vận hành. Ví dụ, phân tích hành vi khách hàng cho phép doanh nghiệp nâng cao mức độ cá nhân hóa dịch vụ, cải thiện trải nghiệm và tăng doanh thu.

Bên cạnh đó, dữ liệu còn góp phần tạo lợi thế cạnh tranh lâu dài. Một tổ chức biết khai thác và bảo vệ dữ liệu hiệu quả thường có khả năng thích ứng nhanh trước thay đổi của thị trường, dự đoán nhu cầu và phát triển sản phẩm hoặc dịch vụ phù hợp hơn so với đối thủ. Các chỉ số về hiệu suất cũng giúp doanh nghiệp giảm lãng phí và nâng cao hiệu quả quy trình.

Không chỉ phục vụ chiến lược, dữ liệu còn là yếu tố thiết yếu trong các hoạt động hằng ngày. Hầu hết quy trình như vận hành chuỗi cung ứng, xử lý giao dịch hay chăm sóc khách hàng đều cần dữ liệu chính xác và sẵn sàng. Việc mất quyền truy cập dữ liệu hoặc để xảy ra thất thoát có thể khiến hoạt động bị gián đoạn, gây tổn thất đáng kể cho doanh nghiệp.

Trong xu hướng công nghiệp 4.0, các công nghệ như trí tuệ nhân tạo, học máy hay Internet vạn vật phụ thuộc hoàn toàn vào nguồn dữ liệu lớn và chất lượng cao để vận hành. Có thể xem dữ liệu như "nhiên liệu" của các hệ thống thông minh, giúp doanh nghiệp mở rộng mô hình kinh doanh và tạo ra những giá trị mới.

1.1.2. Các nguy cơ tác động đến dữ liệu trong doanh nghiệp

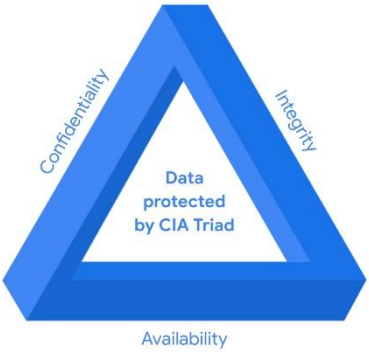
Trong quá trình vận hành, dữ liệu của doanh nghiệp phải đối mặt với nhiều rủi ro ảnh hưởng trực tiếp đến tính chính xác, khả năng sẵn sàng và độ tin cậy. Nhận diện đúng các nguy cơ này là cơ sở để xây dựng cơ chế bảo vệ phù hợp.

- Nguy cơ mất dữ liệu
- Nguy cơ thiếu dữ liệu
- Nguy cơ dữ liệu sai lệch
- Nguy cơ dữ liệu ngoại lai

1.1.3. Yêu cầu về bảo vệ dữ liệu trong doanh nghiệp

Nền tảng của mọi hệ thống an toàn thông tin hiện đại đều được xây dựng dựa trên ba trụ cột: Tính bí mật

(Confidentiality), Tính toàn vẹn (Integrity) và Tính sẵn sàng (Availability), thường được gọi tắt là CIA. Ba nguyên tắc này tạo thành nền tảng cho việc đánh giá mức độ an toàn dữ liệu và định hướng các biện pháp bảo vệ phù hợp trong doanh nghiệp.



Hình 1.1. Tam giác CIA an toàn dữ liệu

a. Tính Bí mật (Confidentiality)

b. Tính Toàn vẹn (Integrity)

c. Tính Sẵn sàng (Availability)

Ba nguyên tắc của mô hình CIA có mối quan hệ chặt chẽ và cần được cân bằng phù hợp với nhu cầu thực tế của doanh nghiệp. Việc tập trung quá mức vào một yếu tố có thể ảnh hưởng đến các yếu tố còn lại. Do đó, doanh nghiệp

cần xác định rõ các yêu cầu bảo vệ dữ liệu phù hợp với đặc thù hoạt động, trong đó đặc biệt chú trọng đến khả năng sao lưu và phục hồi dữ liệu nhằm đảm bảo tính sẵn sàng và giảm thiểu rủi ro khi xảy ra sự cố.

1.2. Sao lưu và phục hồi dữ liệu

1.2.1. Khái niệm sao lưu và phục hồi dữ liệu

Trong môi trường doanh nghiệp đa nền tảng, sao lưu và phục hồi dữ liệu không chỉ nhằm mục tiêu bảo toàn dữ liệu mà còn phải đảm bảo thời gian khôi phục nhanh, đáp ứng các yêu cầu về RPO và RTO đã đề cập ở mục trước.

- Sao lưu toàn phần (Full Backup)
- Sao lưu gia tăng (Incremental Backup)
- Sao lưu vi sai (Differential Backup)

1.2.3. Các phương pháp phục hồi dữ liệu

Tùy theo mức độ sự cố và yêu cầu khôi phục, doanh nghiệp có thể áp dụng nhiều phương pháp phục hồi dữ liệu khác nhau. Các phương pháp phục hồi phổ biến hiện nay bao gồm phục hồi mức tệp, phục hồi toàn hệ thống và phục hồi tức thời.

Các phương pháp sao lưu và phục hồi dữ liệu nêu trên là nền tảng để xây dựng các hệ thống sao lưu hiện đại, trong đó sao lưu gia tăng kết hợp với phục hồi nhanh đóng vai trò then chốt trong việc đảm bảo tính sẵn sàng cho doanh nghiệp đa nền tảng.

1.3. Sao lưu phân tán và bất đồng bộ

1.3.1. Khái niệm và các đặc điểm

- Sao lưu phân tán (Distributed Backup)
- Sao lưu bất đồng bộ (Asynchronous Backup)

1.3.2. Ưu điểm của sao lưu bất đồng bộ và phân tán trong doanh nghiệp đa nền tảng

- Ưu điểm
- Nhược điểm

1.4. Các chỉ tiêu đánh giá hệ thống sao lưu và phục hồi dữ liệu

Để đánh giá hiệu quả của một hệ thống sao lưu và phục hồi dữ liệu, cần xác định các chỉ tiêu kỹ thuật phản ánh khả năng bảo vệ dữ liệu và duy trì hoạt động liên tục

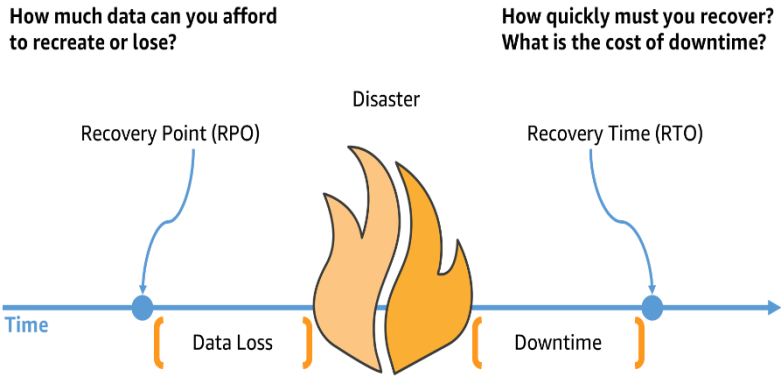
của hệ thống. Trong đó, các chỉ tiêu thường được sử dụng bao gồm RPO, RTO, khả năng mở rộng và tính sẵn sàng của hệ thống.

1.4.1. RPO (Recovery Point Objective)

RPO là chỉ tiêu xác định mức độ mất dữ liệu tối đa cho phép, được tính bằng khoảng thời gian từ thời điểm xảy ra sự cố đến thời điểm bản sao lưu gần nhất. Chỉ tiêu này phản ánh tần suất sao lưu dữ liệu và mức độ chấp nhận rủi ro mất dữ liệu của doanh nghiệp.

1.4.2. RTO (Recovery Time Objective)

RTO là chỉ tiêu xác định thời gian tối đa cho phép để hệ thống khôi phục và hoạt động trở lại sau khi xảy ra sự cố. RTO phản ánh khả năng phục hồi nhanh của hệ thống sao lưu và ảnh hưởng trực tiếp đến mức độ gián đoạn hoạt động của doanh nghiệp.



Hình 1.3. Chỉ số RPO/RTO đánh giá hệ thống sao lưu và phục hồi dữ liệu

1.4.3. Khả năng mở rộng và tính sẵn sàng

- Khả năng mở rộng
- Tính sẵn sàng

Các chỉ tiêu đánh giá nêu trên là cơ sở để phân tích thực trạng hệ thống sao lưu hiện tại trong doanh nghiệp và đánh giá mức độ đáp ứng yêu cầu thực tế, được trình bày trong Chương 2.

1.5. Các biện pháp an toàn hỗ trợ cho hệ thống sao lưu dữ liệu

1.5.1. Mã hoá dữ liệu sao lưu

Việc áp dụng mã hóa cần được thực hiện ở cả hai trạng thái then chốt của dữ liệu: khi dữ liệu đang được truyền tải và khi dữ liệu đang được lưu trữ.

Hai loại mã hóa chính được sử dụng rộng rãi hiện nay bao gồm:

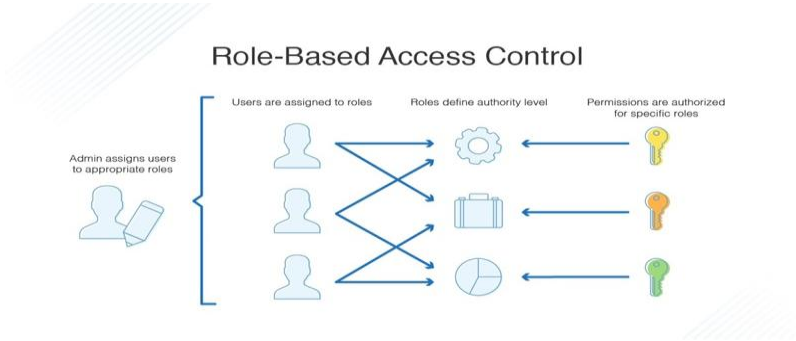
- Mã hóa khoá đối xứng
- Mã hóa khoá bất đối xứng

Mã hóa dữ liệu sao lưu giúp bảo vệ dữ liệu khỏi nguy cơ bị truy cập trái phép trong quá trình lưu trữ và truyền tải. Việc áp dụng cơ chế mã hóa đảm bảo rằng ngay cả khi dữ liệu sao lưu bị đánh cắp, nội dung bên trong vẫn không thể bị khai thác nếu không có khóa giải mã phù hợp.

1.5.2. Xác thực và phân quyền truy cập

Xác thực là quá trình xác minh danh tính của một người dùng hoặc hệ thống.

Sau khi xác thực danh tính, hệ thống cần xác định những gì người dùng đó được phép làm. Đây là quá trình phân quyền (authorization). RBAC (Role-Based Access Control - Kiểm soát truy cập dựa trên vai trò) là mô hình phân quyền phổ biến và hiệu quả.



Hình 1.6. Khái niệm về RBAC

Cơ chế xác thực và phân quyền truy cập đóng vai trò quan trọng trong việc kiểm soát quyền thao tác trên hệ thống sao lưu. Việc áp dụng phân quyền hợp lý giúp hạn chế rủi ro người dùng không có thẩm quyền truy cập, chỉnh sửa hoặc xóa dữ liệu sao lưu, từ đó giảm nguy cơ mất mát dữ liệu do lỗi con người hoặc hành vi cố ý.

1.5.3. Giám sát, phát hiện và phòng chống tấn công ransomware đối với hệ thống sao lưu dữ liệu

Các biện pháp giám sát, phát hiện và phòng chống tấn công ransomware không chỉ góp phần nâng cao mức độ an toàn của hệ thống thông tin mà còn đóng vai trò hỗ trợ trực tiếp cho hệ thống sao lưu và phục hồi dữ liệu, đảm bảo khả năng khôi phục nhanh và duy trì tính sẵn sàng của hệ thống doanh nghiệp.

1.6. Kết luận chương 1

CHƯƠNG 2

THỰC TRẠNG VÀ PHÂN TÍCH HỆ THỐNG SAO LƯU DỮ LIỆU

2.1. Đánh giá các phương pháp sao lưu dữ liệu hiện tại trong doanh nghiệp

Trong các doanh nghiệp nhỏ hiện nay, công tác sao lưu dữ liệu thường được triển khai dựa trên các phương pháp truyền thống với mục tiêu đảm bảo an toàn dữ liệu ở mức cơ bản. Các phương pháp này chủ yếu tập trung vào việc tạo bản sao dữ liệu định kỳ và lưu trữ tại một vị trí tập trung trong nội bộ doanh nghiệp. Mặc dù cách tiếp cận này tương đối đơn giản và dễ triển khai, tuy nhiên nó bộc lộ nhiều hạn chế khi khối lượng dữ liệu ngày càng tăng và hệ thống CNTT phát triển theo hướng đa nền tảng.

2.1.1. Các phương pháp sao lưu dữ liệu đang được sử dụng

Phương pháp sao lưu toàn bộ theo hình thức sao chép dữ liệu trực tiếp tuy dễ triển khai và phù hợp với quy mô nhỏ, nhưng chưa đáp ứng tốt yêu cầu về chi phí hạ tầng

và vận hành, hiệu năng, khả năng mở rộng và tính sẵn sàng trong môi trường doanh nghiệp đa nền tảng. Những hạn chế này là nguyên nhân chính dẫn đến nhu cầu cải tiến và áp dụng các phương pháp sao lưu hiệu quả hơn trong giai đoạn tiếp theo.

2.1.2. Những hạn chế của phương pháp sao lưu hiện tại

a) Dữ liệu thống kê và điều kiện tính toán

b) Tính toán thời gian sao lưu trong kịch bản sao lưu đồng thời của hệ thống hiện tại

c) Áp dụng công thức để tính toán theo từng nhân sự

d) Hạn chế về chi phí lưu trữ và vận hành do sao lưu toàn phần

Bên cạnh các hạn chế về thời gian sao lưu và khả năng phục hồi, phương pháp sao lưu toàn bộ còn gây áp lực lớn lên hạ tầng lưu trữ và công tác vận hành, làm gia tăng chi phí và giảm tính linh hoạt của hệ thống sao lưu hiện tại.

2.2. Phân tích mức độ phù hợp của hệ thống sao lưu hiện tại trong môi trường doanh nghiệp đa nền tảng

Dựa trên các số liệu phân tích ở mục 2.1, có thể thấy rằng khi hệ thống thực hiện sao lưu đồng thời trên nhiều nền tảng khác nhau, băng thông mạng bị chia sẻ giữa nhiều luồng sao lưu, khiến tốc độ sao lưu hiệu dụng trên mỗi hệ thống giảm đáng kể. Thời gian hoàn tất chu kỳ sao lưu phụ thuộc vào nền tảng có dung lượng dữ liệu lớn nhất, làm cho quá trình sao lưu kéo dài và khó kiểm soát trong môi trường có nhiều loại hệ thống cùng tồn tại. Thực tế ở mục 2.1 cho thấy chu kỳ sao lưu có thể kéo dài 6–6,5 giờ, gây khó khăn khi cần tăng tần suất sao lưu để giảm RPO.

2.3. Các yêu cầu đặt ra đối với hệ thống sao lưu mới

2.3.1. Yêu cầu về hiệu năng và thời gian sao lưu

2.3.2. Yêu cầu về RPO và RTO

2.3.3. Yêu cầu về kiến trúc hệ thống và độ tin cậy

2.3.4. Yêu cầu về khả năng mở rộng

2.3.5. Yêu cầu về tối ưu dung lượng lưu trữ và chi phí vận hành

Bảng 2.4. Tóm tắt yêu cầu định lượng của hệ thống sao lưu mới

Nhóm chỉ tiêu	Hiện tại (mục 2.1)	Yêu cầu hệ thống mới
Thời gian sao lưu toàn hệ thống	6–6,5 giờ	~2–3 giờ (giảm 50–70%)
RPO	~24 giờ	< 3 giờ
RTO phục hồi toàn hệ thống	4–6 giờ	< 2–3 giờ (giảm ≥50%)

Điểm hỏng đơn	Có	Loại bỏ/giảm thiểu
Tiêu thụ dung lượng lưu trữ	20TB ~ 10–12 ngày	Giảm 60–80% tốc độ tăng

2.4. Kết luận chương 2

CHƯƠNG 3

THIẾT KẾ VÀ TRIỂN KHAI

HỆ THỐNG SAO LƯU DỮ LIỆU

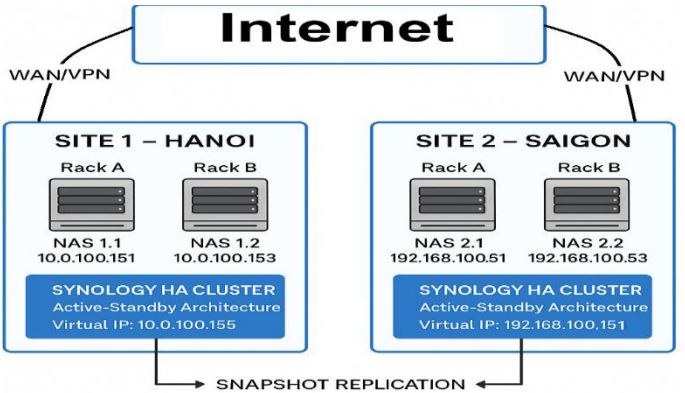
3.1. Phát biểu bài toán

Bài toán đặt ra trong chương này là thiết kế một hệ thống sao lưu mới thay thế phương pháp sao lưu hiện tại, phù hợp điều kiện doanh nghiệp vừa và nhỏ, đồng thời đáp ứng các yêu cầu định lượng đã tổng hợp ở bảng 2.4, gồm: (i) rút ngắn thời gian sao lưu toàn hệ thống xuống khoảng 2–3 giờ; (ii) giảm RPO xuống dưới 3 giờ; (iii) rút ngắn RTO phục hồi toàn hệ thống xuống dưới 2–3 giờ; (iv) loại bỏ/giảm thiểu điểm hỏng đơn; và (v) tối ưu dung lượng để giảm 60–80% tốc độ gia tăng dung lượng lưu trữ so với phương pháp sao lưu toàn phần.

3.2. Mô hình tổng thể hệ thống sao lưu doanh nghiệp

3.2.1. Kiến trúc hai địa điểm (Site 1 – Hà Nội và Site 2 – Hồ Chí Minh)

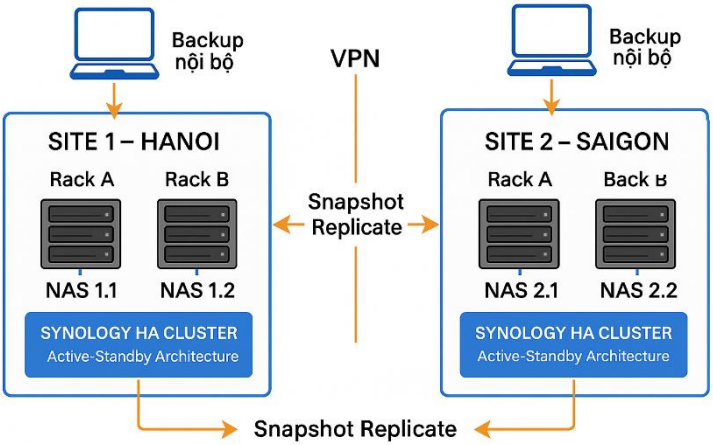
Hệ thống sao lưu được thiết kế theo mô hình phân tán địa lý với hai site đặt tại Hà Nội và HCM



Hình 3.1. Kiến trúc tổng thể hệ thống sao lưu phân tán tại hai địa điểm

3.2.2. Bố trí thiết bị tại mỗi Site

3.2.3. Luồng dữ liệu sao lưu và đồng bộ giữa hai Site



Hình 3.3. Sơ đồ luồng sao lưu nội bộ từng site và cơ chế sao chép snapshot hai chiều giữa Site 1 và Site 2

3.3. Thiết kế hạ tầng lưu trữ tại hai Site

3.3.1. Thiết kế và triển khai cụm NAS tại Site 1

Cụm lưu trữ tại Site 1 được triển khai dựa trên hai thiết bị NAS vật lý, đặt tại hai tủ rack độc lập (Rack A và Rack B) nhằm bảo đảm khả năng chịu lỗi phần cứng và nguồn điện. Hai thiết bị này được cấu hình hoạt động theo mô hình Synology High Availability (HA), trong đó một node giữ vai trò *Active* và node còn lại giữ vai trò *Standby*,

sẵn sàng tiếp quản khi xảy ra sự cố. Việc triển khai cụm NAS tại Site 1 được thực hiện qua các bước: chuẩn bị bộ cài đặt DSM, triển khai máy ảo NAS trên ESXi, điều chỉnh tương thích phần cứng, cài đặt hệ điều hành DSM và hoàn thiện cấu hình HA.

3.3.2. Thiết kế và triển khai cụm NAS tại Site 2

Cụm NAS tại Site 2 được thiết kế và triển khai theo cùng một kiến trúc với cụm NAS tại Site 1. Hai thiết bị NAS 2.1 và NAS 2.2 được cài đặt DSM, cấu hình hệ thống tập tin Btrfs và tham gia cụm Synology High Availability với địa chỉ IP ảo 192.168.100.155. Các bước cấu hình cơ bản bao gồm khai báo thông số hệ thống, thiết lập mạng, tạo volume và tham gia cụm HA đều tương tự như đối với NAS 1.1 và NAS 1.2. Nhờ duy trì kiến trúc đồng nhất giữa hai site, việc sao chép snapshot, giám sát và vận hành hệ thống trở nên đơn giản và dễ chuẩn hóa.

3.4. Thiết kế hệ thống sao lưu nội bộ bằng Active Backup for Business (ABB)

Kiến trúc truyền thông của ABB sử dụng giao thức **gRPC kết hợp TLS** trên các cổng **5510/5511**, giúp đảm

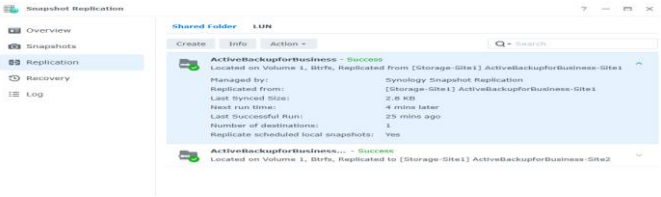
bảo tính bảo mật cũng như hiệu quả truyền tải. Đồng thời, ABB áp dụng cơ chế **Changed Block Tracking (CBT)** để ghi nhận chỉ các khối dữ liệu thay đổi, nhờ đó tối ưu hóa dung lượng lưu trữ và giảm thời gian thực hiện phiên backup.

3.5. Thiết kế cơ chế Replication giữa hai Site

3.5.1. Mục tiêu và nguyên tắc thiết kế

Nguyên tắc vận hành của replication dựa trên snapshot Btrfs: hệ thống chỉ gửi các block thay đổi so với snapshot trước đó, giảm chi phí băng thông và tăng hiệu quả đồng bộ giữa hai site.

3.5.2. Replication snapshot từ Site 1 sang Site 2



Hình 3.42. Kết quả replicate thành công

Tại Site 2, hệ thống tự động hiển thị chuỗi snapshot đã được nhận từ Site 1. Điều này cho phép người quản trị kiểm tra toàn bộ điểm khôi phục (restore points) tương ứng.



Hình 3.43. Các snapshot từ Site 1 hiển thị trong mục Recovery tại Site 2

3.5.4. Phân tích an toàn và khả năng chịu lỗi

Từ các thử nghiệm trên có thể kết luận ngắn gọn như sau:

- Snapshot Btrfs đảm bảo an toàn dữ liệu khỏi lỗi logic và mã hóa ransomware.
- Cơ chế replicate hai chiều duy trì dữ liệu toàn vẹn tại cả hai site.
- Cụm HA bảo đảm dịch vụ không bị gián đoạn khi một NAS gặp sự cố.
- Khi một site mất hoàn toàn, site còn lại vẫn giữ đầy đủ dữ liệu backup của toàn doanh nghiệp.

3.6. Thiết kế cơ chế chống Ransomware cho dữ liệu doanh nghiệp bằng Snapshot Btrfs

3.6.3. Phân tích cơ chế chống ransomware của Snapshot Btrfs

Snapshot được lưu trữ hoàn toàn ở tầng hệ thống tập tin và tồn tại ngầm dưới dạng metadata:

- Snapshot không xuất hiện trong File Station và không thể được truy cập bởi bất kỳ ứng dụng người dùng nào.
- Ransomware mã hóa dữ liệu thông qua SMB chỉ có thể thao tác trên các tệp trong thư mục đang mở, không thể tiếp cận hoặc xóa snapshot nằm trong **/volume/@sharesnap**.
- Snapshot không bị ghi đè trong quá trình mã hóa do cơ chế **copy-on-write** của Btrfs.

3.7. Kết luận chương 3

Chương 3 đã triển khai chi tiết việc thiết kế và xây dựng một hệ thống sao lưu và phục hồi dữ liệu phân tán, bất đồng bộ dành cho doanh nghiệp vừa và nhỏ, dựa trên hạ tầng Synology NAS, công nghệ Btrfs và các cơ chế HA,

replication và backup đa nền tảng. Thông qua mô hình hai site đặt tại hai địa điểm địa lý khác nhau, hệ thống đạt được mức độ sẵn sàng cao và khả năng chịu lỗi mạnh mẽ, bảo đảm dữ liệu luôn tồn tại ở ít nhất hai điểm độc lập.

CHƯƠNG 4

ĐÁNH GIÁ HIỆU QUẢ VÀ SO SÁNH

4.1. Đánh giá hiệu quả hệ thống sao lưu đề xuất

4.1.1. Đánh giá thời gian sao lưu và hiệu năng hệ thống

Kết quả cho thấy thời gian hoàn tất một chu kỳ sao lưu tại mỗi site đã được rút ngắn đáng kể so với hệ thống cũ (6–6,5 giờ), đáp ứng yêu cầu về thời gian sao lưu khoảng 2–3 giờ đã đặt ra ở Chương 2. Việc sao lưu gia tăng giúp giảm tải cho hệ thống mạng và thiết bị lưu trữ, đồng thời hạn chế ảnh hưởng đến hoạt động của người dùng cuối.

4.1.2. Đánh giá RPO và RTO của hệ thống

Các kết quả cho thấy hệ thống mới đã cải thiện rõ rệt khả năng phục hồi so với hệ thống cũ, đặc biệt trong các kịch bản chỉ cần khôi phục một phần dữ liệu. Việc duy trì RPO dưới 3 giờ và RTO dưới 2–3 giờ đáp ứng tốt yêu cầu đảm bảo tính liên tục hoạt động trong môi trường doanh nghiệp đa nền tảng.

4.1.3. Đánh giá hiệu quả sử dụng dung lượng lưu trữ

Kết quả cho thấy hệ thống đề xuất giúp giảm khoảng 60-70% tốc độ gia tăng dung lượng lưu trữ, phù hợp với mục tiêu tối ưu lưu trữ đã xác định ở Chương 2. Điều này góp phần giảm chi phí đầu tư mở rộng lưu trữ và đơn giản hóa công tác vận hành trong dài hạn.

4.1.4. Đánh giá khả năng chống chịu sự cố và ransomware

Từ các kết quả đánh giá trên, có thể thấy hệ thống sao lưu phân tán được đề xuất đã đáp ứng đầy đủ các yêu cầu về thời gian sao lưu, RPO, RTO, tối ưu dung lượng và khả năng chống chịu sự cố đã được đặt ra ở Chương 2. Việc triển khai hai site với quy mô tương đương giúp tăng mức độ dự phòng mà không làm gia tăng đáng kể tải xử lý tại từng site, qua đó nâng cao hiệu quả và độ tin cậy của hệ thống sao lưu trong môi trường doanh nghiệp vừa và nhỏ.

4.2. So sánh hệ thống sao lưu cũ và hệ thống đề xuất mới

4.2.2. Nhận xét kết quả so sánh

Kết quả so sánh cho thấy hệ thống sao lưu đề xuất không chỉ khắc phục được các hạn chế của hệ thống hiện tại mà còn đáp ứng đầy đủ các yêu cầu định lượng đã đặt ra ở Chương 2. Những cải thiện về thời gian sao lưu, RPO, RTO, khả năng dự phòng và hiệu quả sử dụng dung lượng là cơ sở để khẳng định tính khả thi và hiệu quả của giải pháp trong bối cảnh doanh nghiệp vừa và nhỏ.

4.3. So sánh hệ thống sao lưu đề xuất mới với các giải pháp sao lưu khác

- **So sánh với các giải pháp sao lưu thương mại chuyên dụng**
- **So sánh với các giải pháp sao lưu cục bộ truyền thống**
- **So sánh với các giải pháp sao lưu thuần đám mây**

4.4. Nhận xét và đề xuất cải tiến

4.4.1. Nhận xét tổng quan

- **Về độ tin cậy**
- **Về hiệu năng**
- **Về an toàn trước ransomware**
- **Về chi phí**

4.4.2. Các hạn chế còn tồn tại

- **Phụ thuộc nền tảng Synology**
- **Tính năng nâng cao cho doanh nghiệp lớn chưa đầy đủ**
- **Phụ thuộc băng thông WAN giữa hai site**
- **Thiếu lớp sao lưu "offline"**

4.4.3. Đề xuất cải tiến

- **Bổ sung lớp sao lưu thứ ba theo nguyên tắc 3–2–1**
- **Tăng cường giám sát và cảnh báo**
- **Tối ưu băng thông WAN**
- **Mở rộng phạm vi bảo vệ ứng dụng**

- **Hoàn thiện quy trình DR và diễn tập định kỳ**

Những đề xuất này không làm thay đổi kiến trúc nền tảng của hệ thống nhưng sẽ góp phần nâng cao đáng kể mức độ an toàn, khả năng mở rộng cũng như tính sẵn sàng vận hành trong thực tế.

KẾT LUẬN

Luận văn đã làm rõ vai trò thiết yếu của dữ liệu trong hoạt động của doanh nghiệp vừa và nhỏ, đồng thời khẳng định yêu cầu bảo vệ dữ liệu cần được tiếp cận một cách hệ thống, dựa trên ba nguyên tắc nền tảng: tính bảo mật, tính toàn vẹn và tính sẵn sàng. Trên cơ sở phân tích cơ sở lý luận và thực trạng hệ thống sao lưu truyền thống, đề tài đã chỉ ra những hạn chế phổ biến như thời gian sao lưu kéo dài, RPO và RTO cao, rủi ro điểm hỏng đơn và chi phí lưu trữ gia tăng nhanh theo thời gian.

Xuất phát từ các yêu cầu định lượng được xác định, luận văn đã đề xuất và triển khai mô hình hệ thống sao lưu phân tán phù hợp với điều kiện thực tế của doanh nghiệp vừa và nhỏ. Giải pháp dựa trên Synology NAS kết hợp cơ chế High Availability và Snapshot Replication hai chiều giữa hai site độc lập đã góp phần loại bỏ điểm hỏng đơn và nâng cao khả năng chịu lỗi của hệ thống. Việc áp dụng Active Backup for Business cho phép hợp nhất sao lưu các nền tảng Windows, Linux và máy ảo trong một hệ thống quản lý tập trung, đơn giản hóa vận hành và tối ưu chi phí đầu tư.

Kết quả đánh giá và so sánh cho thấy hệ thống sao lưu đề xuất đã đáp ứng tốt các mục tiêu đặt ra: thời gian hoàn tất chu kỳ sao lưu được rút ngắn xuống khoảng 2–3 giờ, RPO được duy trì dưới 3 giờ, RTO phục hồi toàn hệ thống dưới 2–3 giờ, đồng thời tốc độ gia tăng dung lượng lưu trữ giảm khoảng 60–70% so với phương pháp sao lưu toàn bộ truyền thống. Các thử nghiệm mô phỏng sự cố và tấn công ransomware cũng cho thấy cơ chế snapshot trên hệ thống file Btrfs cho phép khôi phục dữ liệu nhanh chóng và hiệu quả, góp phần đảm bảo tính liên tục của hoạt động doanh nghiệp.

Từ góc độ ứng dụng, luận văn đã chứng minh rằng một kiến trúc sao lưu phân tán với mức độ phức tạp vừa phải vẫn có thể đáp ứng các yêu cầu quan trọng về hiệu năng, độ tin cậy và an toàn dữ liệu, đồng thời phù hợp với khả năng triển khai và vận hành của doanh nghiệp vừa và nhỏ. Trong tương lai, hệ thống có thể tiếp tục được hoàn thiện thông qua việc tăng cường tự động hóa, mở rộng tích hợp với các nền tảng đám mây và bổ sung các cơ chế bảo mật nâng cao nhằm đáp ứng tốt hơn nhu cầu phát triển và chuyển đổi số của doanh nghiệp.