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

ĐỀ ÁN TỐT NGHIỆP THẠC SĨ KỸ THUẬT

(Theo định hướng ứng dụng)

HÀ NỘI – NĂM 2026

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

ĐỀ ÁN THẠC SĨ KỸ THUẬT

(Theo định hướng ứng dụng)

NGƯỜI HƯỚNG DẪN KHOA HỌC

PGS.TS. ĐỖ TRUNG TUẤN

HÀ NỘI – NĂM 2026

LỜI CAM ĐOAN

Tôi cam đoan đây là công trình nghiên cứu của riêng tôi.

Các số liệu, kết quả nêu trong đề án tốt nghiệp là trung thực và chưa từng được ai công bố trong bất kỳ công trình nào khác.

Hà Nội, tháng 02 năm 2026

Tác giả đề án

Bùi Minh Tuấn

LỜI CẢM ƠN

Trong suốt quá trình học tập và thực hiện đề tài này, tôi đã nhận được rất nhiều sự hỗ trợ, khích lệ và tạo điều kiện thuận lợi từ các cơ quan, thầy cô, đồng nghiệp và gia đình. Với lòng biết ơn chân thành nhất, tôi xin được gửi lời cảm ơn sâu sắc đến tất cả những người đã đồng hành cùng tôi trong chặng đường vừa qua.

Trước hết, tôi xin chân thành cảm ơn Học viện Công nghệ Bưu chính Viễn thông và Khoa Đào tạo Sau Đại học – nơi đã trang bị cho tôi nền tảng kiến thức vững chắc, môi trường học tập chuyên nghiệp và các điều kiện tốt nhất để tôi có thể nghiên cứu và hoàn thiện đề tài. Tôi trân trọng biết ơn các thầy cô của Học viện đã tận tâm giảng dạy, hướng dẫn và truyền cảm hứng, giúp tôi trưởng thành hơn trong tư duy, kỹ năng và phương pháp nghiên cứu khoa học.

Tôi cũng xin gửi lời cảm ơn chân thành đến các anh/chị/em đồng nghiệp đã luôn hỗ trợ, chia sẻ tài liệu, kinh nghiệm thực tiễn và đóng góp những ý kiến quý báu giúp tôi hoàn thiện nội dung đề tài một cách đầy đủ và chính xác hơn. Sự đồng hành và tinh thần hợp tác của mọi người là nguồn động lực rất lớn đối với tôi.

Cuối cùng, tôi xin bày tỏ lòng biết ơn sâu sắc đến gia đình đã luôn là điểm tựa tinh thần vững chắc nhất của tôi. Gia đình đã cho tôi sự yêu thương, ủng hộ và niềm tin không điều kiện, giúp tôi vượt qua mọi khó khăn trong quá trình học tập và nghiên cứu.

Một lần nữa, tôi xin gửi lời tri ân chân thành đến tất cả những ai đã hỗ trợ, đồng hành và tạo điều kiện để tôi hoàn thành tốt công việc của mình.

Xin trân trọng cảm ơn !

MỤC LỤC

LỜI CAM ĐOAN i
LỜI CẢM ƠN ii
MỤC LỤC..... iii
DANH SÁCH CÁC THUẬT NGỮ VIẾT TẮT vii
DANH SÁCH CÁC THUẬT NGỮ TIẾNG ANH ix
DANH SÁCH CÁC HÌNH VẼ x
DANH SÁCH CÁC BẢNG BIỂU..... xiii
LỜI MỞ ĐẦU 1
CHƯƠNG 1 5
CƠ SỞ LÝ LUẬN VỀ SAO LƯU VÀ PHỤC HỒI DỮ LIỆU TRONG DOANH NGHIỆP ĐA NỀN TẢNG 5
1.1. Tổng quan về dữ liệu và yêu cầu bảo vệ dữ liệu trong doanh nghiệp ..5
1.1.1. Khái niệm và tầm quan trọng của dữ liệu 5
1.1.2. Các nguy cơ tác động đến dữ liệu trong doanh nghiệp.....6
1.1.3. Yêu cầu về bảo vệ dữ liệu trong doanh nghiệp 7
1.2. Sao lưu và phục hồi dữ liệu 8
1.2.1. Khái niệm sao lưu và phục hồi dữ liệu 9
1.2.2. Các phương pháp sao lưu dữ liệu 9
1.2.3. Các phương pháp phục hồi dữ liệu 10
1.3. Sao lưu phân tán và bất đồng bộ 11
1.3.1. Khái niệm và các đặc điểm 11
1.3.2. Ưu điểm của sao lưu bất đồng bộ và phân tán trong doanh nghiệp đa nền tảng 12
1.4. Các chỉ tiêu đánh giá hệ thống sao lưu và phục hồi dữ liệu 13
1.4.1. RPO (Recovery Point Objective).....13

1.4.2. RTO (Recovery Time Objective).....13
1.4.3. Khả năng mở rộng và tính sẵn sàng 14
1.5. Các biện pháp an toàn hỗ trợ cho hệ thống sao lưu dữ liệu.....14
1.5.1. Mã hoá dữ liệu sao lưu 14
1.5.2. Xác thực và phân quyền truy cập 16
1.5.3. Giám sát, phát hiện và phòng chống tấn công ransomware đối với hệ thống sao lưu dữ liệu 18
1.6. Kết luận chương 1 19
CHƯƠNG 2 20
THỰC TRẠNG VÀ PHÂN TÍCH HỆ THỐNG SAO LƯU DỮ LIỆU 20
2.1. Đánh giá các phương pháp sao lưu dữ liệu hiện tại trong doanh nghiệp 20
2.1.1. Các phương pháp sao lưu dữ liệu đang được sử dụng 20
2.1.2. Những hạn chế của phương pháp sao lưu hiện tại.....21
2.2. Phân tích mức độ phù hợp của hệ thống sao lưu hiện tại trong môi trường doanh nghiệp đa nền tảng.....27
2.3. Các yêu cầu đặt ra đối với hệ thống sao lưu mới 28
2.3.1. Yêu cầu về hiệu năng và thời gian sao lưu 29
2.3.2. Yêu cầu về RPO và RTO 29
2.3.3. Yêu cầu về kiến trúc hệ thống và độ tin cậy.....29
2.3.4. Yêu cầu về khả năng mở rộng 30
2.3.5. Yêu cầu về tối ưu dung lượng lưu trữ và chi phí vận hành 30
2.4. Kết luận chương 2.....31
CHƯƠNG 3 32
THIẾT KẾ VÀ TRIỂN KHAI 32
HỆ THỐNG SAO LƯU DỮ LIỆU.....32

3.1. Phát biểu bài toán 32
3.2. Mô hình tổng thể hệ thống sao lưu doanh nghiệp 33
3.2.1. Kiến trúc hai địa điểm (Site 1 – Hà Nội và Site 2 – Hồ Chí Minh) 33
3.2.2. Bố trí thiết bị tại mỗi Site 34
3.2.3. Luồng dữ liệu sao lưu và đồng bộ giữa hai Site 35
3.3. Thiết kế hạ tầng lưu trữ tại hai Site 36
3.3.1. Thiết kế và triển khai cụm NAS tại Site 1 36
3.3.2. Thiết kế và triển khai cụm NAS tại Site 2 (Thành phố Hồ Chí Minh) 44
3.4. Thiết kế hệ thống sao lưu nội bộ bằng Active Backup for Business (ABB) 44
3.4.1. Cấu trúc sao lưu và giao thức kết nối 44
3.4.2. Cài đặt và cấu hình sao lưu máy Windows 45
3.4.3. Thử nghiệm mất file và khôi phục dữ liệu trên Windows 46
3.4.4. Cài đặt và cấu hình sao lưu máy Linux (Ubuntu) 47
3.4.5. Thử nghiệm mất file và khôi phục dữ liệu trên Linux 48
3.4.6. Thiết kế sao lưu máy chủ ảo (Virtual Machine) 49
3.5. Thiết kế cơ chế Replication giữa hai Site 50
3.5.1. Mục tiêu và nguyên tắc thiết kế 50
3.5.2. Replication snapshot từ Site 1 sang Site 2 51
3.5.3. Replication snapshot từ Site 2 sang Site 1 52
3.5.4. Phân tích an toàn và khả năng chịu lỗi 53
3.6. Thiết kế cơ chế chống Ransomware cho dữ liệu doanh nghiệp bằng Snapshot Btrfs 55
3.6.1. Mô phỏng tấn công Ransomware 55
3.6.2. Khôi phục dữ liệu từ snapshot an toàn 56

3.6.3. Phân tích cơ chế chống ransomware của Snapshot Btrfs 56
3.7. Kết luận 57
CHƯƠNG 4 59
ĐÁNH GIÁ HIỆU QUẢ VÀ SO SÁNH 59
4.1. Đánh giá hiệu quả hệ thống sao lưu đề xuất 59
4.1.1. Đánh giá thời gian sao lưu và hiệu năng hệ thống 59
4.1.2. Đánh giá RPO và RTO của hệ thống 60
4.1.3. Đánh giá hiệu quả sử dụng dung lượng lưu trữ 60
4.1.4. Đánh giá khả năng chống chịu sự cố và ransomware 61
4.2. So sánh hệ thống sao lưu cũ và hệ thống đề xuất mới 62
4.2.1. So sánh các chỉ tiêu kỹ thuật chính 62
4.2.2. Nhận xét kết quả so sánh 62
4.3. So sánh hệ thống sao lưu đề xuất mới với các giải pháp sao lưu khác 63
4.3.1. So sánh với các giải pháp sao lưu thương mại chuyên dụng 63
4.3.2. So sánh với các giải pháp sao lưu cục bộ truyền thống 65
4.3.3. So sánh với các giải pháp sao lưu thuần đám mây 65
4.4. Nhận xét và đề xuất cải tiến 66
4.4.1. Nhận xét tổng quan 66
4.4.2. Các hạn chế còn tồn tại 67
4.4.3. Đề xuất cải tiến 67
KẾT LUẬN 69
TÀI LIỆU THAM KHẢO 70
BẢN CAM ĐOAN 72

DANH SÁCH CÁC THUẬT NGỮ VIẾT TẮT

Từ viết tắt	Giải nghĩa từ theo tiếng Anh	Giải nghĩa từ theo tiếng Việt
GDDT		Giáo dục và đào tạo
ICT	Information & Communication Technology	Công nghệ thông tyin và truyền thông
CIA	Confidentiality – Integrity – Availability	Tam giác bảo mật: Tính bí mật – Toàn vẹn – Sẵn sàng
ABAC	Attribute-Based Access Control	Kiểm soát truy cập dựa trên thuộc tính
RBAC	Role-Based Access Control	Kiểm soát truy cập dựa trên vai trò
NAS	Network Attached Storage	Thiết bị lưu trữ mạng
HA	High Availability	Cụm dự phòng – tính sẵn sàng cao
ABB	Active Backup for Business	Giải pháp sao lưu doanh nghiệp của Synology
IDS	Intrusion Detection System	Hệ thống phát hiện xâm nhập
IPS	Intrusion Prevention System	Hệ thống phòng chống xâm nhập
SIEM	Security Information and Event Management	Hệ thống quản lý sự kiện & thông tin bảo mật
VM/VMX	Virtual Machine	Máy ảo
DR / DRP	DisasterRecovery/Disaster Recovery Plan	Khôi phục thảm họa / Kế hoạch khôi phục thảm họa
AES	Advanced Encryption Standard	Chuẩn mã hóa đối xứng AES
RSA	Rivest–Shamir–Adleman	Thuật toán mã hóa bất đối xứng
TLS/SSL	Transport Layer Security / Secure Sockets Layer	Giao thức bảo mật truyền dữ liệu

MFA/2FA	Multi-Factor Authentication / Two-Factor Authentication	Xác thực đa yếu tố / hai lớp
OTP	One-Time Password	Mật khẩu dùng một lần
ACL	Access Control List	Danh sách kiểm soát truy cập
VPN	Virtual Private Network	Mạng riêng ảo
API	Application Programming Interface	Giao diện lập trình ứng dụng
RPO	Recovery Point Objective	Mức độ mất dữ liệu tối đa cho phép
RTO	Recovery Time Objective	Thời gian phục hồi hệ thống
Btrfs	B-tree File System	Hệ thống tập tin dạng cây B
SSO	Single Sign On	Đăng nhập một lần cho nhiều hệ thống
SaaS	Software as a Service	Phần mềm cung cấp dưới dạng dịch vụ
MSSP	Managed Security Service Provider	Nhà cung cấp dịch vụ bảo mật được quản lý
UTM	Unified Threat Management	Giải pháp quản lý mối đe dọa hợp nhất

DANH SÁCH CÁC THUẬT NGỮ TIẾNG ANH

Thuật ngữ tiếng Anh	Cách giải nghĩa tiếng Việt
Snapshot	Ảnh chụp trạng thái dữ liệu
Replication	Sao chép dữ liệu
Asynchronous Backup	Sao lưu bất đồng bộ
Full Backup	Sao lưu toàn phần
Incremental Backup	Sao lưu gia tăng
Differential Backup	Sao lưu vi sai
File-level Restore	Phục hồi mức tệp
Bare-metal Restore	Phục hồi toàn hệ thống
Instant Restore	Phục hồi tức thời
Load Balancer	Bộ cân bằng tải
Outlier	Giá trị ngoại lai
Hash function (SHA-256, MD5)	Hàm băm để kiểm tra toàn vẹn dữ liệu
Digital Signature	Chữ ký số
File-Level Restore	Khôi phục theo tệp
Volume / Storage Pool	Vùng lưu trữ / nhóm lưu trữ
Cluster	Cụm máy chủ
Copy-on-write	Cơ chế ghi-sao của Btrfs

DANH SÁCH CÁC HÌNH VẼ

Hình 1.1. Tam giác CIA an toàn dữ liệu 7
Hình 1.2. Giải pháp sao lưu dữ liệu cho doanh nghiệp 9
Hình 1.3. Chỉ số RPO/RTO đánh giá hệ thống sao lưu và phục hồi dữ liệu 14
Hình 1.4. Mã hoá khoá đối xứng 15
Hình 1.5. Mã hoá khoá bất đối xứng 16
Hình 1.6. Khái niệm về RBAC 17
Hình 2.1. Mô hình hạ tầng CNTT đa nền tảng trong doanh nghiệp 27
Hình 3.1. Kiến trúc tổng thể hệ thống sao lưu phân tán tại hai địa điểm 34
Hình 3.2. Sơ đồ bố trí phần cứng và kết nối của cụm lưu trữ tại Site 1 35
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 36
Hình 3.4. Giao diện tải ARC Loader trên GitHub 37
Hình 3.5. Minh họa tệp arc-3.0.7.vmdk/ova 37
Hình 3.6. Màn hình “Ready to complete” khi deploy OVA 37
Hình 3.7. Lỗi “Unsupported hardware family vmx-17” 38
Hình 3.8. Màn hình ARC Loader khi boot vào DSM. 39
Hình 3.9. Màn hình khởi động quá trình thiết lập hệ thống 39
Hình 3.10. Lựa chọn các ổ đĩa để tạo Storage Pool với cấu hình RAID 40
Hình 3.11. Xác nhận cấu hình Storage Pool và Volume 40
Hình 3.12. Storage Pool và Volume sau khi khởi tạo 40
Hình 3.13. Trạng thái Storage Pool và Volume trên NAS 1.1 41
Hình 3.14. Cấu hình địa chỉ IP cho từng NAS trong Site 1 41
Hình 3.15. Khởi tạo kết nối High Availability Cluster trên DSM 42

Hình 3.16. Lựa chọn giao diện mạng cho Cluster và Heartbeat 42
Hình 3.17. Xác thực máy chủ Passive trong cụm HA 42
Hình 3.18. Khai báo hostname và IP ảo của cụm HA trong Site 1.....43
Hình 3.19. Kiểm tra các điều kiện để tạo cụm Synology HA.....43
Hình 3.20. Hình thành cụm HA và bắt đầu đồng bộ dữ liệu trong Site 1.....43
Hình 3.21. Cài đặt thành công ABB agent trên Windows 45
Hình 3.22. Thiết lập vùng để sao lưu trên Windows 45
Hình 3.23. Thiết lập thời gian để sao lưu trên Windows 46
Hình 3.24. Sao lưu máy Windows thành công 46
Hình 3.25. Thư mục kết quả các lần sao lưu thành công 46
Hình 3.26. Nhật ký hiển thị trên máy trạm Windows sau khi sao lưu.....46
Hình 3.27. Tệp bị xoá nhầm trên Windows 47
Hình 3.28. Tệp được khôi phục từ NAS 47
Hình 3.29. Tệp sau khi khôi phục thành công bằng ABB 47
Hình 3.30. Cài đặt agent theo dạng tệp .run.....48
Hình 3.31. Cài đặt và kết nối agent trên máy linux thành công.....48
Hình 3.32. Thêm ổ cứng 3G để tiến hành thử nghiệm sao lưu 48
Hình 3.33. Sao lưu thành công máy Linux 48
Hình 3.34. Thư mục tệp kết quả sao lưu 48
Hình 3.35. Trường hợp mất tệp dữ liệu trên linux 49
Hình 3.36. Khôi phục tệp bị mất trên máy linux từ NAS 49
Hình 3.37. Tệp được khôi phục thành công trên máy linux 49
Hình 3.38. Kiểm tra dịch vụ của các máy ảo 50
Hình 3.39. Sao lưu thành công máy chủ ảo ESXi.....50
Hình 3.40. Khai báo máy chủ đích replication 51
Hình 3.41. Cấu hình lịch replicate chu kỳ 30 phút 51

Hình 3.42. Kết quả replicate thành công từ Site 1 sang Site 2.....52
Hình 3.43. Các snapshot từ Site 1 hiển thị trong mục Recovery tại Site 2.....52
Hình 3.44. Replicate thành công từ Site 2 sang Site 1.....52
Hình 3.45. Tạo folder thử nghiệm trên Site 1 53
Hình 3.46. Folder đã xuất hiện trên Site 2 sau quá trình replicate.....53
Hình 3.47. Tệp kiểm thử tại Site 1 replicate sang Site 2.....53
Hình 3.48. NAS Passive chuyển thành Active duy trì dịch vụ 54
Hình 3.49. Replicate vẫn thành công trên cả hai chiều khi sự cố xảy ra 54
Hình 3.50. Danh sách snapshot phục hồi tại Site 2 sau sự cố.....54
Hình 3.51. Tệp từ Site 1 replicate thành công đến Site 2 55
Hình 3.52. Các tệp bị mã hoá ransomware chuyển sang đuôi .docdgm 55
Hình 3.53. Dữ liệu trong thư mục được khôi phục thành công 56

DANH SÁCH CÁC BẢNG BIỂU

Bảng 1.1. Ưu nhược điểm của sao lưu phân tán và sao lưu bất đồng bộ.....12

Bảng 2.1. Bảng thống kê dung lượng và thời gian sao lưu toàn phần theo từng nhân sự áp dụng với hệ thống hiện tại 23

Bảng 2.2. Bảng tổng hợp chỉ tiêu hiệu năng và độ tin cậy của hệ thống sao lưu hiện tại.....25

Bảng 2.3. Ước tính mức tiêu thụ dung lượng lưu trữ theo thời gian 26

Bảng 2.4. Tóm tắt yêu cầu định lượng của hệ thống sao lưu mới 31

Bảng 4.1. Thời gian sao lưu tại mỗi site của hệ thống đề xuất mới.....59

Bảng 4.2. Đánh giá RPO và RTO của hệ thống đề xuất mới.....60

Bảng 4.3. So sánh mức tiêu thụ dung lượng lưu trữ tại mỗi site 61

Bảng 4.4. So sánh hệ thống sao lưu cũ và hệ thống đề xuất mới.....62

LỜI MỞ ĐẦU

Trong bối cảnh chuyển đổi số đang diễn ra mạnh mẽ, dữ liệu ngày càng trở thành tài sản cốt lõi, quyết định hiệu quả vận hành và khả năng cạnh tranh của doanh nghiệp. Sự gia tăng nhanh chóng của khối lượng dữ liệu, cùng với xu hướng triển khai hạ tầng CNTT đa nền tảng gồm máy trạm, máy chủ vật lý và hệ thống ảo hóa, đặt ra yêu cầu ngày càng cao đối với công tác bảo vệ và đảm bảo an toàn dữ liệu. Đặc biệt đối với doanh nghiệp vừa và nhỏ, nơi nguồn lực đầu tư và nhân sự CNTT còn hạn chế, việc xây dựng một hệ thống sao lưu và phục hồi hiệu quả, tin cậy và phù hợp với điều kiện thực tế là một bài toán có ý nghĩa quan trọng.

Bên cạnh các rủi ro truyền thống như lỗi phần cứng, sự cố vận hành hay thiên tai, các mối đe dọa an ninh mạng, đặc biệt là tấn công ransomware, đang gia tăng cả về quy mô và mức độ tinh vi. Nhiều trường hợp cho thấy doanh nghiệp vẫn bị gián đoạn hoạt động nghiêm trọng dù đã triển khai sao lưu, do dữ liệu dự phòng bị mã hóa hoặc không thể phục hồi kịp thời. Trong bối cảnh đó, việc nghiên cứu và xây dựng một mô hình sao lưu phân tán có khả năng chịu lỗi cao, rút ngắn thời gian phục hồi và đảm bảo an toàn dữ liệu mang ý nghĩa thực tiễn rõ rệt.

1. Lý do chọn đề tài

Trong môi trường doanh nghiệp hiện đại, dữ liệu đóng vai trò trung tâm trong hầu hết các hoạt động quản lý, vận hành và ra quyết định. Mất mát dữ liệu hoặc gián đoạn truy cập dữ liệu có thể gây ra những thiệt hại nghiêm trọng về tài chính, uy tín và khả năng duy trì hoạt động. Đối với doanh nghiệp vừa và nhỏ, các sự cố này càng có tác động lớn do hạn chế về nguồn lực kỹ thuật và khả năng đầu tư cho hạ tầng dự phòng.

Thực tế cho thấy nhiều doanh nghiệp vẫn đang áp dụng các phương pháp sao lưu truyền thống, mang tính tập trung và sao lưu toàn bộ, dẫn đến thời gian sao lưu kéo dài, RPO và RTO cao, đồng thời tồn tại rủi ro điểm hỏng đơn. Trong khi đó, các giải pháp sao lưu thương mại cao cấp thường có chi phí đầu tư và vận hành vượt quá khả năng của doanh nghiệp vừa và nhỏ. Xuất phát từ nhu cầu thực tiễn này, đề tài lựa chọn nghiên cứu và đề xuất một mô hình hệ thống sao lưu phân tán dựa trên công nghệ NAS, snapshot hệ thống

file và cơ chế sao chép dữ liệu giữa các site, nhằm cân bằng giữa hiệu quả kỹ thuật và tính khả thi trong triển khai.

2. Tổng quan về vấn đề nghiên cứu

Bài toán bảo vệ dữ liệu trong doanh nghiệp bao gồm nhiều khía cạnh như sao lưu, phục hồi, đảm bảo tính sẵn sàng, phân quyền truy cập và an toàn trước các mối đe dọa an ninh mạng. Trong những năm gần đây, các mô hình sao lưu phân tán kết hợp snapshot ở mức hệ thống file và sao chép dữ liệu bất đồng bộ được xem là hướng tiếp cận hiệu quả nhằm giảm RPO, RTO và tăng khả năng chống chịu trước ransomware.

Nhiều nghiên cứu tập trung vào việc tối ưu hóa thời gian phục hồi, loại bỏ điểm hỏng đơn và giảm chi phí lưu trữ thông qua các kỹ thuật sao lưu gia tăng và khử trùng lặp dữ liệu. Đề tài kế thừa các cơ sở lý luận này để xây dựng một mô hình sao lưu phù hợp với điều kiện vận hành và hạ tầng của doanh nghiệp vừa và nhỏ, đồng thời chú trọng đến khả năng triển khai và quản trị trong thực tế.

3. Mục đích nghiên cứu

Mục tiêu khoa học:

- Làm rõ cơ sở lý luận về dữ liệu, an toàn dữ liệu và các chỉ tiêu đánh giá hệ thống sao lưu như RPO và RTO.
- Nghiên cứu cơ chế snapshot trên hệ thống file, sao chép dữ liệu bất đồng bộ và các mô hình sao lưu trong môi trường đa nền tảng.

Mục tiêu thực tiễn:

- Thiết kế và triển khai mô hình hệ thống sao lưu phân tán giữa hai site độc lập, đáp ứng yêu cầu phục hồi dữ liệu nhanh và đảm bảo tính sẵn sàng.
- Đề xuất giải pháp sao lưu phù hợp với doanh nghiệp vừa và nhỏ, có hiệu quả kỹ thuật, chi phí hợp lý và dễ vận hành.

4. Đối tượng và phạm vi nghiên cứu

Đối tượng nghiên cứu của đề án là hệ thống dữ liệu trong doanh nghiệp, bao gồm dữ liệu máy trạm, máy chủ vật lý và máy ảo; các cơ chế snapshot, sao chép dữ liệu và giải pháp sao lưu phân tán.

Phạm vi nghiên cứu được giới hạn trong mô hình hệ thống sao lưu triển khai tại hai site độc lập, sử dụng hạ tầng lưu trữ NAS và giải pháp Active Backup for Business, với quy mô và tải dữ liệu phù hợp đặc trưng của doanh nghiệp vừa và nhỏ.

5. Phương pháp nghiên cứu

Luận văn sử dụng hai nhóm phương pháp chính:

- Nghiên cứu lý thuyết: tổng hợp các mô hình an toàn dữ liệu, snapshot, replication và các phương pháp sao lưu trong doanh nghiệp.
- Thực nghiệm: triển khai hệ thống gồm hai cụm NAS, kiểm thử sao lưu Windows – Linux – máy ảo, mô phỏng sự cố và 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 này trình bày các khái niệm cơ bản về dữ liệu, các nguyên tắc bảo mật cốt lõi và các yêu cầu bảo vệ dữ liệu trong doanh nghiệp, làm nền tảng cho việc xây dựng mô hình kỹ thuật ở các chương tiếp theo.

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

Chương này phân tích thực trạng hệ thống sao lưu truyền thống trong doanh nghiệp, chỉ ra các hạn chế về hiệu năng, RPO, RTO và chi phí lưu trữ, đồng thời xác định các yêu cầu định lượng đối với hệ thống sao lưu mới.

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

Đây là nội dung trọng tâm của luận văn, mô tả quá trình xây dựng hạ tầng sao lưu tại hai site, cấu hình cụm NAS HA, triển khai Active Backup for Business và thiết lập Snapshot Replication hai chiều. Các thử nghiệm thực tế được trình bày nhằm kiểm chứng khả năng vận hành và phục hồi của hệ thống.

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

Chương này đánh giá hiệu quả hệ thống sao lưu đề xuất dựa trên các chỉ tiêu đã xác định, so sánh với hệ thống sao lưu hiện tại và một số giải pháp khác, đồng thời đưa ra các hướng cải tiến trong tương lai.

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.

Một trong những rủi ro nghiêm trọng nhất là mất dữ liệu. Sự cố này có thể bắt nguồn từ hỏng hóc phần cứng, lỗi hệ thống, thiên tai, mã độc tống tiền hoặc thao tác nhầm của người dùng. Hậu quả thường ảnh hưởng lớn đến hoạt động, gây thiệt hại tài chính và làm giảm uy tín doanh nghiệp. Để hạn chế, cần áp dụng các biện pháp dự phòng như sao lưu định kỳ, giải pháp phục hồi sau thảm họa và đào tạo người dùng.

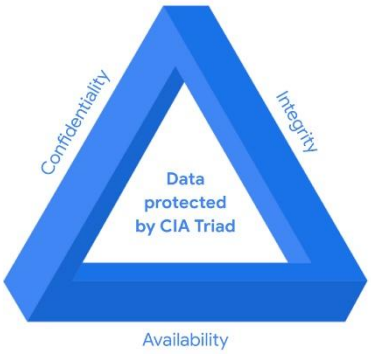
Bên cạnh nguy cơ mất dữ liệu, tình trạng thiếu dữ liệu cũng tác động không nhỏ đến chất lượng báo cáo và phân tích. Việc thu thập không đầy đủ, lỗi truyền tải hoặc hạn chế kỹ thuật trong quá trình tích hợp có thể khiến dữ liệu bị thiếu hụt. Điều này dễ dẫn đến đánh giá sai hoặc quyết định thiếu cơ sở. Các biện pháp khắc phục bao gồm quy trình thu thập rõ ràng, cơ chế giám sát và hệ thống tích hợp đáng tin cậy.

Một vấn đề tinh vi hơn là dữ liệu sai lệch. Dữ liệu vẫn tồn tại nhưng mang giá trị không chính xác do lỗi nhập liệu, sai sót khi xử lý hoặc sự không đồng bộ giữa các hệ thống. Dữ liệu sai rất khó phát hiện và có thể gây ra các quyết định chiến lược không phù hợp. Doanh nghiệp cần áp dụng kiểm tra tính hợp lệ, làm sạch dữ liệu thường xuyên và giám sát tính nhất quán.

Ngoài ra, việc sử dụng dữ liệu từ các nguồn bên ngoài (dữ liệu ngoại lai) cũng tiềm ẩn rủi ro nếu không đánh giá chất lượng trước khi khai thác. Dữ liệu không đảm bảo độ tin cậy có thể dẫn đến các phân tích lệch hướng và chiến lược sai lầm. Vì vậy, cần có quy trình đánh giá nguồn dữ liệu, kiểm tra phương pháp thu thập và đảm bảo tính đại diện trước khi đưa vào sử dụng.

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)

Nguyên tắc này đảm bảo rằng thông tin chỉ có thể được truy cập bởi những người dùng hoặc hệ thống được ủy quyền. Nó ngăn chặn sự tiết lộ thông tin nhạy cảm cho các thực thể không mong muốn, bảo vệ quyền riêng tư và bí mật kinh doanh. Ví dụ: Thông tin khách hàng, bí mật công nghệ, chiến lược kinh doanh.

Các rủi ro làm suy giảm tính bí mật dữ liệu thường xuất phát từ tấn công mạng, phần mềm độc hại, nghe lén trên đường truyền hoặc các sai sót từ phía người dùng nội bộ. Do đó, doanh nghiệp cần đặt ra yêu cầu kiểm soát truy cập chặt chẽ và bảo vệ dữ liệu khỏi nguy cơ bị truy cập trái phép trong suốt vòng đời sử dụng.

b. Tính Toàn vẹn (Integrity)

Tính toàn vẹn đảm bảo rằng dữ liệu là chính xác, đáng tin cậy và không bị sửa đổi một cách trái phép hoặc vô tình trong suốt quá trình lưu trữ, xử lý và truyền tải. Nó duy trì sự tin cậy của dữ liệu, là cơ sở cho các quyết định kinh doanh quan trọng. Ví dụ: Dữ liệu giao dịch tài chính, hồ sơ bệnh án, thông tin kế toán.

Các mối đe dọa đối với tính toàn vẹn có thể đến từ lỗi con người, sự cố phần cứng, lỗi phần mềm hoặc các hành vi tấn công nhằm chỉnh sửa dữ liệu. Vì vậy, hệ thống thông tin cần có cơ chế phát hiện và ngăn chặn các thay đổi bất hợp pháp, đồng thời đảm bảo khả năng kiểm tra và xác minh tính chính xác của dữ liệu khi cần thiết.

c. Tính Sẵn sàng (Availability)

Nguyên tắc này đảm bảo rằng thông tin và các tài nguyên hệ thống luôn có thể truy cập được một cách kịp thời và tin cậy bởi những người dùng được ủy quyền khi họ cần. Sự gián đoạn trong tính sẵn sàng có thể làm tê liệt hoạt động kinh doanh. Ví dụ: Hệ thống email, cơ sở dữ liệu khách hàng, ứng dụng bán hàng trực tuyến.

Các yếu tố ảnh hưởng đến tính sẵn sàng bao gồm sự cố phần cứng, lỗi phần mềm, thảm họa tự nhiên hoặc các cuộc tấn công làm gián đoạn hệ thống. Do đó, doanh nghiệp cần đặt ra yêu cầu cao đối với khả năng khôi phục dữ liệu và duy trì hoạt động liên tục của hệ thống. Trong đó, sao lưu và phục hồi dữ liệu được xem là biện pháp nền tảng nhằm đảm bảo tính sẵn sàng, giúp hệ thống có thể nhanh chóng khôi phục dữ liệu và hoạt động bình thường khi xảy ra sự cố.

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

Sau khi xác định các yêu cầu bảo vệ dữ liệu dựa trên mô hình CIA, trong đó tính sẵn sàng đóng vai trò đặc biệt quan trọng đối với hoạt động liên tục của doanh nghiệp, sao lưu và phục hồi dữ liệu được xem là giải pháp nền tảng nhằm đảm bảo khả năng khôi phục dữ liệu khi xảy ra sự cố. Mục này trình bày các khái niệm cơ bản, các phương pháp sao lưu và các hình thức phục hồi dữ liệu phổ biến đang được áp dụng trong hệ thống thông tin doanh nghiệp.

1.2.1. Khái niệm sao lưu và phục hồi dữ liệu

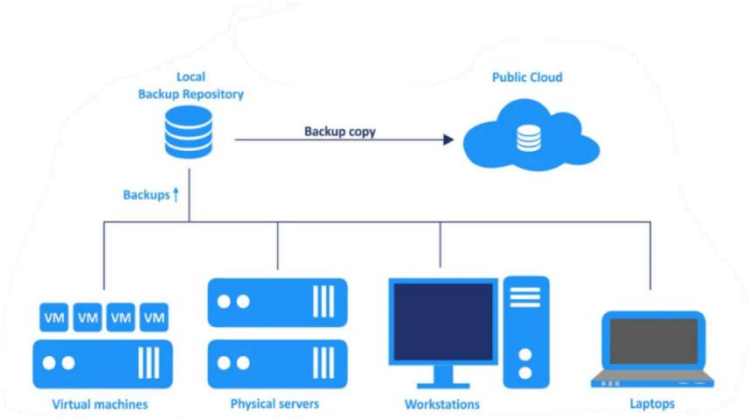
Sao lưu dữ liệu (Backup) là quá trình tạo ra các bản sao của dữ liệu gốc và lưu trữ chúng tại một vị trí an toàn nhằm phục vụ cho việc khôi phục khi dữ liệu chính bị mất mát, hư hỏng hoặc không thể truy cập. Sao lưu dữ liệu giúp doanh nghiệp giảm thiểu rủi ro mất dữ liệu do lỗi phần cứng, lỗi phần mềm, sai sót của con người hoặc các cuộc tấn công mạng.

Phục hồi dữ liệu (Restore) là quá trình khôi phục dữ liệu từ các bản sao lưu để đưa hệ thống trở lại trạng thái hoạt động bình thường sau sự cố. Hiệu quả của quá trình phục hồi phụ thuộc trực tiếp vào phương pháp sao lưu được sử dụng, tần suất sao lưu cũng như khả năng truy xuất dữ liệu từ các bản sao lưu.

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.

1.2.2. Các phương pháp sao lưu dữ liệu

Hiện nay, trong hệ thống thông tin doanh nghiệp, các phương pháp sao lưu được xem là nền tảng cốt lõi trong hầu hết các chiến lược bảo vệ dữ liệu, bao gồm ba loại chính: Sao lưu toàn phần, Sao lưu gia tăng và Sao lưu vi sai.



Hình 1.2. Giải pháp sao lưu dữ liệu cho doanh nghiệp

Sao lưu toàn phần (Full Backup) là phương pháp cơ bản nhất trong đó toàn bộ dữ liệu được chọn sẽ được sao chép đầy đủ trong mỗi lần thực hiện. Phương pháp này tạo ra một bản sao hoàn chỉnh và độc lập của dữ liệu tại một thời điểm. Ưu điểm chính của Full Backup là việc phục hồi dữ liệu rất đơn

giản và nhanh chóng, vì toàn bộ dữ liệu đã nằm trong một bản sao lưu duy nhất. Tuy nhiên, nhược điểm lớn nhất của phương pháp này là tốn nhiều thời gian, tài nguyên hệ thống và dung lượng lưu trữ. Vì vậy, sao lưu toàn phần thường chỉ được thực hiện theo chu kỳ dài (hàng tuần hoặc hàng tháng hoặc hang quý).

Sao lưu gia tăng (Incremental Backup) là phương pháp chỉ sao lưu những dữ liệu đã thay đổi kể từ lần sao lưu gần nhất (có thể là Full Backup hoặc Incremental Backup trước đó). Ưu điểm vượt trội của sao lưu gia tăng là tốc độ sao lưu nhanh, tiết kiệm đáng kể dung lượng lưu trữ và băng thông mạng. Tuy nhiên, nhược điểm của phương pháp này nằm ở quá trình phục hồi, vì để khôi phục dữ liệu hoàn chỉnh, hệ thống cần phải có bản Full Backup gần nhất và tất cả các bản Incremental Backup theo đúng thứ tự từ sau bản Full Backup đó. Quá trình này phức tạp và mất nhiều thời gian hơn so với Full Backup.

Sao lưu vi sai (Differential Backup) là phương pháp trung gian giữa Full Backup và Incremental Backup. Sao lưu vi sai sẽ sao lưu tất cả dữ liệu đã thay đổi kể từ lần Full Backup gần nhất. Ưu điểm của phương pháp này là việc phục hồi tương đối đơn giản, chỉ cần bản Full Backup gần nhất và bản Differential Backup mới nhất. So với Incremental Backup, tốc độ phục hồi của Differential Backup nhanh hơn. Nhược điểm chính là dung lượng lưu trữ và thời gian sao lưu sẽ tăng dần theo thời gian cho đến khi có một bản Full Backup mới, mặc dù vẫn tiết kiệm hơn so với Full Backup.

Một chiến lược sao lưu hiệu quả thường kết hợp nhiều phương pháp. Ví dụ điển hình là mô hình GFS (Grandfather – Father - Son). Sao lưu hàng ngày (Incremental - Son), sao lưu hàng tuần (Full/Differential - Father) và sao lưu hàng tháng (Full - Grandfather).

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.

Phục hồi mức tệp (File-level Restore) cho phép khôi phục các tệp hoặc thư mục riêng lẻ từ bản sao lưu mà không ảnh hưởng đến toàn bộ hệ thống.

Phương pháp này phù hợp trong các trường hợp mất dữ liệu cục bộ hoặc lỗi thao tác của người dùng, giúp giảm thời gian gián đoạn và hạn chế ảnh hưởng đến hệ thống đang vận hành.

Phục hồi toàn hệ thống (Bare-metal Restore) là phương pháp khôi phục toàn bộ hệ điều hành, ứng dụng và dữ liệu từ bản sao lưu lên một máy chủ hoặc máy trạm trống. Phương pháp này thường được sử dụng khi xảy ra sự cố nghiêm trọng như hỏng hóc phần cứng hoặc lỗi hệ điều hành, giúp hệ thống nhanh chóng trở lại trạng thái hoạt động ban đầu.

Phục hồi tức thời (Instant Restore) cho phép truy cập và vận hành dữ liệu hoặc máy ảo trực tiếp từ bản sao lưu mà không cần chờ quá trình phục hồi hoàn tất. Phương pháp này giúp rút ngắn đáng kể thời gian gián đoạn dịch vụ, đặc biệt phù hợp với các hệ thống yêu cầu tính sẵn sàng cao.

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

Để hạn chế rủi ro mất dữ liệu diện rộng và duy trì khả năng sẵn sàng cao, nhiều doanh nghiệp chuyển sang mô hình sao lưu phân tán và sao lưu bất đồng bộ.

1.3.1. Khái niệm và các đặc điểm

Sao lưu phân tán (Distributed Backup) là mô hình sao lưu trong đó các bản sao dữ liệu không được lưu trữ tập trung tại một vị trí duy nhất mà được phân tán trên nhiều node, nhiều thiết bị lưu trữ hoặc nhiều trung tâm dữ liệu khác nhau về mặt địa lý. Kiến trúc này giúp loại bỏ điểm hỏng đơn (Single Point of Failure), nâng cao khả năng chịu lỗi của hệ thống. Ngay cả khi một node hoặc một phần hạ tầng gặp sự cố, dữ liệu vẫn có thể được phục hồi từ các node còn lại. Bên cạnh đó, sao lưu phân tán còn hỗ trợ cân bằng tải và cải thiện hiệu suất truy cập dữ liệu. Tuy nhiên, mô hình này đòi hỏi cơ chế quản lý và đồng bộ phức tạp hơn, đồng thời chi phí triển khai ban đầu thường cao hơn so với các mô hình sao lưu tập trung.

Sao lưu bất đồng bộ (Asynchronous Backup) là cơ chế sao lưu trong đó quá trình ghi dữ liệu trên hệ thống nguồn không cần chờ việc đồng bộ hoàn tất tới kho lưu trữ đích. Dữ liệu được ghi nhận và chuyển tiếp theo cơ chế nền, cho phép hệ thống nguồn tiếp tục hoạt động mà không bị ảnh hưởng đáng kể bởi độ trễ mạng. Nhờ đó, sao lưu bất đồng bộ đặc biệt phù hợp với các hệ thống yêu cầu hiệu suất cao hoặc sao lưu dữ liệu tới các vị trí từ xa. Nhược điểm của cơ chế này là tồn tại khả năng mất dữ liệu trong một khoảng thời gian rất ngắn nếu xảy ra sự cố trước khi dữ liệu được đồng bộ hóa hoàn toàn, tương ứng với giá trị RPO (Recovery Point Objective) lớn hơn 0.

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

Trong môi trường doanh nghiệp đa nền tảng, dữ liệu thường phân tán trên nhiều hệ điều hành, máy chủ vật lý, máy ảo và các hệ thống lưu trữ khác nhau. Việc áp dụng sao lưu bất đồng bộ kết hợp với sao lưu phân tán giúp doanh nghiệp nâng cao khả năng bảo vệ dữ liệu, đảm bảo tính sẵn sàng và khả năng mở rộng của hệ thống sao lưu. Các đặc điểm ưu nhược điểm chính của hai phương pháp này được tổng hợp trong Bảng 1.1 bên dưới đây.

Bảng 1.1. Ưu nhược điểm của sao lưu phân tán và sao lưu bất đồng bộ

Phương pháp	Ưu điểm	Nhược điểm	Sự phù hợp
Sao lưu phân tán	Khả năng chịu lỗi cao, loại bỏ điểm hỏng đơn, tăng tính sẵn sàng cao và khả năng mở rộng	Quản lý phức tạp, chi phí triển khai ban đầu cao	Môi trường yêu cầu dự phòng địa lý và khả năng phục hồi sau thảm hoạ mạnh mẽ
Sao lưu bất đồng bộ	Hiệu suất cao, ít ảnh hưởng đến hệ thống nguồn, phù hợp sao lưu từ xa	Có nguy cơ mất dữ liệu nhỏ (RPO > 0)	Hệ thống nhạy cảm yêu cầu hiệu năng cao và tính liên tục

Việc kết hợp sao lưu bất đồng bộ với kiến trúc sao lưu phân tán cho phép doanh nghiệp tận dụng ưu điểm của cả hai phương pháp: vừa giảm tác

động đến hiệu năng hệ thống nguồn, vừa tăng khả năng chịu lỗi và phục hồi dữ liệu. Đây là cơ sở quan trọng để thiết kế các hệ thống sao lưu hiện đại, đặc biệt phù hợp với các doanh nghiệp đa nền tảng có yêu cầu cao về tính sẵn sàng và khả năng mở rộng.

Đồng thời, để tối ưu, doanh nghiệp nên áp dụng mô hình sao lưu phân tán bất đồng bộ cho các dữ liệu quan trọng, giúp đảm bảo khả năng phục hồi ngay cả khi xảy ra thảm họa tại trung tâm dữ liệu chính, trong khi vẫn duy trì hiệu suất cho hệ thống sản xuất. Sự kết hợp này tạo thành một cơ chế phòng thủ nhiều lớp, giảm thiểu tối đa rủi ro mất mát dữ liệu cho doanh nghiệp.

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.

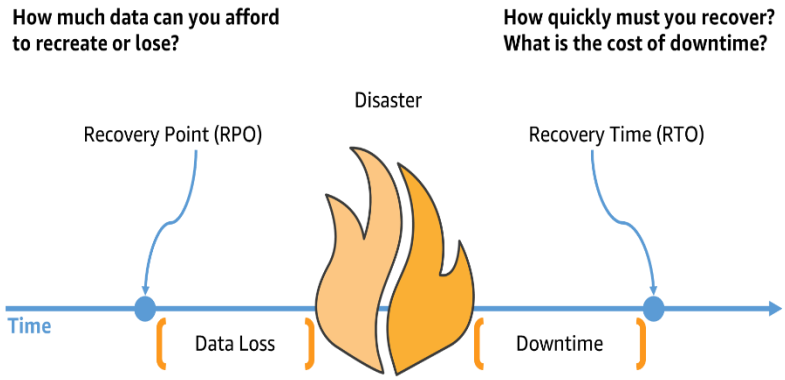
Trong các hệ thống sao lưu truyền thống với tần suất sao lưu thấp, giá trị RPO thường lớn, dẫn đến nguy cơ mất nhiều dữ liệu khi xảy ra sự cố. Ngược lại, các hệ thống sao lưu hiện đại với cơ chế sao lưu gia tăng hoặc sao lưu bất đồng bộ có thể giảm đáng kể giá trị RPO, đáp ứng tốt hơn yêu cầu bảo vệ dữ liệu trong môi trường doanh nghiệp đa nền tảng.

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.

Đối với các doanh nghiệp có yêu cầu cao về tính liên tục của dịch vụ, RTO cần được giảm thiểu ở mức thấp nhất. Điều này đòi hỏi hệ thống sao lưu

không chỉ đảm bảo việc lưu trữ dữ liệu mà còn hỗ trợ các cơ chế phục hồi nhanh, giúp rút ngắn thời gian gián đoạn khi xảy ra sự cố.



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 thể hiện mức độ đáp ứng của hệ thống sao lưu khi quy mô dữ liệu, số lượng máy trạm, máy chủ hoặc nền tảng công nghệ tăng lên theo thời gian. Một hệ thống sao lưu hiệu quả cần cho phép mở rộng linh hoạt mà không làm gián đoạn hoạt động hiện tại.

Tính sẵn sàng phản ánh khả năng hệ thống sao lưu duy trì hoạt động liên tục và đảm bảo dữ liệu luôn có thể được truy cập hoặc phục hồi khi cần thiết. Trong bối cảnh doanh nghiệp đa nền tảng, việc đảm bảo tính sẵn sàng cao là yêu cầu quan trọng, đặc biệt khi hệ thống phải đối mặt với các sự cố phần cứng, phần mềm hoặc tấn công mạng.

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.

Mã hóa dữ liệu khi truyền tải (Encryption in Transit): Biện pháp này bảo vệ dữ liệu khi nó di chuyển qua mạng, chẳng hạn như giữa máy khách và máy chủ, hoặc giữa các dịch vụ đám mây. Các giao thức như TLS/SSL

(Transport Layer Security/Secure Sockets Layer) được sử dụng rộng rãi để tạo một kênh truyền thông an toàn, mã hóa mọi dữ liệu trao đổi qua lại. Điều này ngăn chặn các cuộc tấn công nghe lén (eavesdropping) và "man-in-the-middle". Ví dụ, khi một người dùng truy cập vào website ngân hàng, biểu tượng ổ khóa trên thanh địa chỉ chính là dấu hiệu cho thấy kết nối đang được bảo vệ bởi TLS/SSL.

Mã hóa dữ liệu khi lưu trữ (Encryption at Rest): Biện pháp này bảo vệ dữ liệu khi nó được lưu trữ trên các phương tiện như ổ cứng máy chủ, cơ sở dữ liệu, thiết bị lưu trữ đám mây hoặc các bản sao lưu. Ngay cả khi ai đó đánh cắp được ổ cứng vật lý hoặc xâm nhập được vào hệ thống lưu trữ, họ cũng không thể đọc được dữ liệu. Mã hóa ở trạng thái lưu trữ có thể được thực hiện ở cấp độ tệp (file-level), cấp độ ổ đĩa (disk-level) hoặc cấp độ cơ sở dữ liệu (database-level).

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

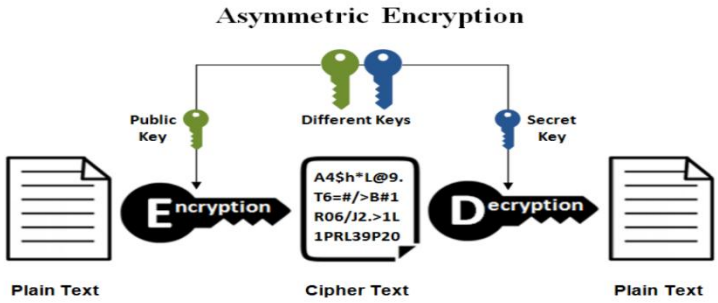
- Mã hóa khoá đối xứng (Symmetric Encryption) sử dụng một khóa duy nhất cho cả hai quá trình mã hóa và giải mã. Phương pháp này rất nhanh và hiệu quả để mã hóa một lượng lớn dữ liệu.



Hình 1.4. Mã hoá khoá đối xứng

AES (Advanced Encryption Standard): Là thuật toán mã hóa đối xứng được sử dụng rộng rãi và an toàn nhất hiện nay, thường được dùng để mã hóa dữ liệu khi lưu trữ (ví dụ: FileVault trên macOS, BitLocker trên Windows) và cả trong nhiều giao thức truyền thông.

- Mã hóa khoá bất đối xứng (Asymmetric Encryption) sử dụng một cặp khóa: Khóa công khai (Public Key) để mã hóa và Khóa bí mật (Private Key) để giải mã. Phương pháp này an toàn hơn cho việc trao đổi khóa nhưng chậm hơn so với mã hóa đối xứng.



Hình 1.5. Mã hoá khoá bất đối xứng

RSA (Rivest–Shamir–Adleman): Là một trong những thuật toán mã hóa bất đối xứng đầu tiên và được sử dụng phổ biến nhất. Nó thường được dùng để thiết lập các kết nối an toàn (như trong TLS/SSL), cho chữ ký số và để mã hóa các khóa phiên (session keys) ngắn hạn dùng trong mã hóa đối xứng.

TLS/SSL (Transport Layer Security/Secure Sockets Layer): Đây không phải là một thuật toán đơn lẻ mà là một giao thức bảo mật sử dụng kết hợp cả mã hóa đối xứng và bất đối xứng. Trong quá trình "bắt tay" TLS, mã hóa bất đối xứng (như RSA) được dùng để xác thực danh tính máy chủ và trao đổi an toàn một khóa phiên (session key). Sau đó, khóa phiên này được dùng cho mã hóa đối xứng (thường là AES) để mã hóa toàn bộ phiên truyền thông tiếp theo, kết hợp giữa tốc độ của AES và tính bảo mật của RSA.

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. Các phương pháp xác thực ngày càng trở nên mạnh mẽ hơn để chống lại việc đánh cắp thông tin đăng nhập. Các phương pháp phổ biến nhất kể đến như là:

- Mật khẩu
- Mật khẩu một lần (OTP - One-Time Password)
- Xác thực hai yếu tố (2FA)

Bằng cách kết hợp, chẳng hạn, mật khẩu (yếu tố 1) và mã OTP từ điện thoại (yếu tố 2), bảo mật được tăng cường đáng kể vì kẻ tấn công cần phải có cả hai yếu tố mới có thể đăng nhập.

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. Trong RBAC, quyền truy cập (permissions) không được gán trực tiếp cho từng người dùng, mà được gán cho các vai trò (roles). Sau đó, người dùng được gán vào các vai trò phù hợp với chức năng công việc của họ.

Hình 1.6. Khái niệm về RBAC

Trong hình 1.5 trên, admin của hệ thống sẽ gán từng người dùng (users) vào một vai trò (role) phù hợp của họ. Mỗi vai trò này quy định mức quyền hạn khác nhau (ví dụ như vai trò kỹ thuật, quản lý và phân tích dữ liệu). Các quyền truy cập (permissions) sẽ không gán trực tiếp cho người dùng mà được gán cho vai trò. Và cuối cùng, người dùng thuộc vai trò nào sẽ tự động có các quyền tương ứng đó.

Bài toán thường được đặt ra cho các quản trị hệ thống trong doanh nghiệp là, trong một công ty, có thể định nghĩa các vai trò như "Nhân viên Kế toán", "Quản lý Nhân sự", "Quản trị viên Hệ thống". Vai trò "Nhân viên Kế toán" được gán quyền "Đọc và Ghi" vào cơ sở dữ liệu lương. Vai trò "Quản lý Nhân sự" được gán quyền "Đọc" cơ sở dữ liệu lương và "Đọc/Ghi" cơ sở dữ liệu nhân viên. Vậy nên, khi một nhân viên mới vào phòng kế toán, họ chỉ cần được gán vai trò "Nhân viên Kế toán" là tự động có các quyền tương ứng.

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

Trong các mối đe dọa an ninh mạng hiện nay, ransomware được xem là nguy cơ đặc biệt nghiêm trọng đối với hệ thống sao lưu dữ liệu, bởi mục tiêu của các cuộc tấn công này không chỉ là dữ liệu đang vận hành mà còn nhắm trực tiếp vào các bản sao lưu nhằm làm mất khả năng phục hồi của doanh nghiệp. Do đó, việc triển khai các cơ chế giám sát và phát hiện sớm đóng vai trò quan trọng trong việc bảo vệ hệ thống sao lưu. Một số cơ chế giám sát phổ biến hiện nay có thể kể đến là:

- Hệ thống phát hiện xâm nhập (IDS - Intrusion Detection System), giúp phân tích lưu lượng mạng (Network-based IDS - NIDS) hoặc hoạt động trên các máy chủ (Host-based IDS - HIDS) để tìm kiếm các dấu hiệu của hoạt động độc hại hoặc vi phạm chính sách.
- Hệ thống phòng chống xâm nhập (IPS - Intrusion Prevention System), giúp không chỉ phát hiện mà còn có khả năng tự động thực hiện hành động để ngăn chặn cuộc tấn công trong thời gian thực, chẳng hạn như chặn lưu lượng từ một địa chỉ IP độc hại hoặc đóng một kết nối nguy hiểm.
- Hệ thống giám sát và quản lý các sự kiện bảo mật (SIEM -) Security Information and Event Management), giúp thu thập và tổng hợp dữ liệu nhật ký bảo mật từ khắp nơi trong tổ chức - bao gồm máy chủ, thiết bị mạng, tường lửa, IDS/IPS, ứng dụng và endpoint. Nó chuẩn hóa và phân tích các sự kiện này, tìm kiếm mối tương quan và các mẫu hành vi đáng ngờ mà các hệ thống độc lập không thể nhìn thấy.

Các giải pháp giám sát và phát hiện xâm nhập giúp theo dõi liên tục hoạt động của hệ thống, phát hiện các hành vi bất thường như truy cập trái

phép, thay đổi dữ liệu không hợp lệ hoặc các dấu hiệu lây nhiễm mã độc. Thông qua việc thu thập và phân tích nhật ký hệ thống, các cơ chế này hỗ trợ đội ngũ quản trị nhanh chóng nhận diện nguy cơ và đưa ra cảnh báo kịp thời trước khi sự cố lan rộng.

Bên cạnh đó, việc tích hợp các biện pháp phòng chống tấn công ransomware cho hệ thống sao lưu giúp hạn chế tối đa thiệt hại khi xảy ra sự cố. Trong trường hợp phát hiện xâm nhập, hệ thống cần có khả năng cô lập thành phần bị ảnh hưởng, ngăn chặn truy cập trái phép và khôi phục dữ liệu từ các bản sao lưu an toàn. Quá trình phục hồi từ bản sao lưu sạch là bước then chốt nhằm đưa hệ thống trở lại trạng thái hoạt động bình thường và giảm thiểu thời gian gián đoạn dịch vụ.

Như vậy, 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.

Các biện pháp an toàn nêu trên đóng vai trò hỗ trợ quan trọng cho hệ thống sao lưu và phục hồi dữ liệu, góp phần đảm bảo dữ liệu sao lưu được bảo vệ an toàn trong suốt quá trình lưu trữ và khai thác. Trên cơ sở các cơ sở lý luận đã trình bày, chương tiếp theo sẽ phân tích thực trạng hệ thống sao lưu dữ liệu trong doanh nghiệp.

1.6. Kết luận chương 1

Chương 1 đã trình bày các cơ sở lý luận liên quan đến sao lưu và phục hồi dữ liệu trong doanh nghiệp đa nền tảng, bao gồm các khái niệm cơ bản, các phương pháp sao lưu và phục hồi dữ liệu, mô hình sao lưu bất đồng bộ và phân tán, cùng các chỉ tiêu đánh giá hiệu quả hệ thống sao lưu. Bên cạnh đó, chương cũng đề cập đến các biện pháp an toàn hỗ trợ nhằm bảo vệ dữ liệu sao lưu trước các mối đe dọa trong quá trình lưu trữ và khai thác.

Những nội dung lý luận này là nền tảng để phân tích thực trạng hệ thống sao lưu dữ liệu hiện tại trong doanh nghiệp, qua đó làm rõ các vấn đề còn tồn tại và yêu cầu cải tiến. Trên cơ sở đó, Chương 2 sẽ tập trung đánh giá thực trạng và phân tích hệ thống sao lưu dữ liệu, làm tiền đề cho việc đề xuất và triển khai giải pháp sao lưu phân tán, bất đồng bộ ở các chương tiếp theo.

CHƯƠNG 2

THỰC TRẠNG VÀ PHÂN TÍCH HỆ THỐNG SAO LƯU DỮ LIỆU

Chương này tập trung phân tích thực trạng hệ thống sao lưu dữ liệu hiện đang được sử dụng trong một doanh nghiệp nhỏ với quy mô khoảng 20 nhân sự, có hạ tầng công nghệ thông tin đa nền tảng. Trên cơ sở đó, các phương pháp sao lưu dữ liệu phổ biến được đánh giá nhằm chỉ ra những hạn chế còn tồn tại như thời gian sao lưu kéo dài, rủi ro mất dữ liệu và khả năng phục hồi chậm khi xảy ra sự cố. Bên cạnh đó, chương cũng phân tích hiệu quả sao lưu và phục hồi dữ liệu trong môi trường doanh nghiệp đa nền tảng, làm rõ các vấn đề về quản lý dữ liệu, tính đồng bộ và khả năng đáp ứng yêu cầu phục hồi nhanh. Các dữ liệu thực nghiệm về hiệu suất hệ thống sao lưu hiện tại được tổng hợp làm cơ sở cho việc đề xuất giải pháp sao lưu bất đồng bộ và phân tán ở các chương tiếp theo.

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

Trong mô hình doanh nghiệp quy mô nhỏ khoảng 20 nhân sự, hệ thống sao lưu dữ liệu hiện tại thường được triển khai theo phương pháp sao lưu toàn phần (Full Backup) với cách tiếp cận đơn giản và thủ công. Phương pháp này chủ yếu tập trung vào việc sao chép toàn bộ dữ liệu từ các hệ thống nguồn sang thiết bị lưu trữ theo định kỳ, phổ biến là một lần mỗi ngày.

Cụ thể, dữ liệu từ các máy trạm người dùng và máy chủ nội bộ được sao chép trực tiếp sang ổ cứng gắn ngoài hoặc một máy chủ lưu trữ duy nhất trong mạng nội bộ. Quá trình sao lưu thường được thực hiện thông qua các công cụ sao chép dữ liệu cơ bản hoặc các tác vụ tự động đơn giản, chưa có cơ chế phân loại hay tối ưu dữ liệu trong quá trình sao lưu.

Việc áp dụng sao lưu toàn bộ giúp đơn giản hóa quy trình quản lý và phục hồi dữ liệu, do mỗi bản sao lưu chứa đầy đủ dữ liệu của hệ thống tại thời điểm sao lưu. Tuy nhiên, do toàn bộ dữ liệu đều được sao chép lại trong mỗi lần sao lưu, phương pháp này tiêu tốn nhiều thời gian, băng thông mạng và dung lượng lưu trữ, đặc biệt khi khối lượng dữ liệu ngày càng tăng theo quy mô hoạt động của doanh nghiệp.

Ngoài ra, mô hình sao lưu hiện tại thường phụ thuộc vào một thiết bị lưu trữ duy nhất, không có cơ chế phân tán hay dự phòng địa lý. Khi thiết bị lưu trữ này gặp sự cố, toàn bộ dữ liệu sao lưu có nguy cơ không thể sử dụng để phục hồi. Bên cạnh đó, quá trình sao lưu toàn bộ diễn ra trong thời gian dài có thể ảnh hưởng đến hiệu năng của hệ thống và hoạt động của người dùng nếu trùng với giờ làm việc.

Nhìn chung, 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

Như đã trình bày ở mục 2.1.1 trên, doanh nghiệp đang áp dụng phương pháp sao lưu toàn phần (Full Backup) theo hình thức sao chép dữ liệu trực tiếp từ các máy trạm về một ổ cứng lưu trữ ngoài hay thiết bị lưu trữ tập trung trong mạng nội bộ. Để đánh giá cụ thể các hạn chế của phương pháp này, phần dưới đây tổng hợp số liệu ghi nhận trong quá trình vận hành hệ thống sao lưu tại một doanh nghiệp nhỏ quy mô khoảng 20 nhân sự, đồng thời thực hiện tính toán thời gian sao lưu trong kịch bản nhiều máy trạm sao lưu đồng thời.

a) Dữ liệu thống kê và điều kiện tính toán

Hệ thống sao lưu được cấu hình thực hiện sao lưu đồng thời trên toàn bộ máy trạm vào khung giờ ngoài giờ làm việc. Kết nối mạng nội bộ sử dụng hạ tầng 1Gbps, tuy nhiên do đặc tính chia sẻ băng thông, overhead giao thức và tải I/O tại thiết bị lưu trữ tập trung, băng thông hiệu dụng quan sát được trong giai đoạn sao lưu vào khoảng xấp xỉ 800 Mbps. Khi các máy trạm sao lưu đồng thời, băng thông hiệu dụng được chia sẻ giữa các luồng sao lưu; trong phạm vi đánh giá này, băng thông được xem là phân bổ tương đối đồng đều giữa các máy trạm trong chu kỳ sao lưu.

- Tổng băng thông hiệu dụng của hệ thống sao lưu hiện tại:

$$\boldsymbol{B_{total} \approx 800\text{Mbps}}$$

- Số máy trạm thực hiện sao lưu đồng thời: **N = 20**

Khi các máy trạm sao lưu đồng thời, băng thông hiệu dụng được chia sẻ giữa các luồng sao lưu. Để đơn giản hóa mô hình tính toán và phù hợp với bối cảnh doanh nghiệp nhỏ, giả định băng thông được phân bổ tương đối đồng đều giữa các máy trạm trong khoảng thời gian sao lưu.

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

Băng thông cấp cho mỗi máy trạm khi sao lưu đồng thời:

$$\boldsymbol{B_i = \frac{B_{total}}{N}}$$

Trong đó:

- $\boldsymbol{B_i}$ (Mbps): băng thông hiệu dụng cấp cho máy trạm thứ i
- $\boldsymbol{B_{total}}$ (Mbps): tổng băng thông hiệu dụng của hệ thống trong thời gian sao lưu
- **N**: số máy trạm sao lưu đồng thời

Thời gian sao lưu của máy trạm i là:

$$\boldsymbol{T_i(giây) = \frac{D_i * 8192}{B_i}}$$

Trong đó:

- $\boldsymbol{T_i}$ (giây): thời gian sao lưu của máy trạm thứ i
- $\boldsymbol{D_i}$ (GB): dung lượng dữ liệu cần sao lưu của máy trạm thứ i
- **8192**: hệ số đổi GB → Mb (1 GB = 1024 MB; 1 MB = 8 Mb)
- $\boldsymbol{B_i}$ (Mbps): băng thông cấp cho máy trạm i

Thời gian theo phút được quy đổi như sau:

$$\boldsymbol{T_i(\text{phút}) = \frac{T_i(\text{giây})}{60}}$$

Do các máy trạm sao lưu đồng thời**,** thời gian hoàn tất toàn bộ chu kỳ sao lưu của hệ thống không phải là tổng thời gian từng máy, mà phụ thuộc vào máy có thời gian sao lưu lớn nhất:

$$\boldsymbol{T_{total} = \max(T_i)}$$

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

Từ các giả định nêu trên:

$$\boldsymbol{B_i = \frac{800}{20} = 40\ \text{Mbps} \approx 5\ \text{MB/s}}$$

Bảng bên dưới đây tổng hợp dung lượng dữ liệu theo từng nhân sự và thời gian sao lưu được tính theo công thức.

Bảng 2.1. Bảng thống kê dung lượng và thời gian sao lưu toàn phần theo từng nhân sự áp dụng với hệ thống hiện tại

STT	Nhân sự	Dữ liệu cần sao lưu D_i (GB)	B_{total} (Mbps)	B_i (Mbps)	Tốc độ hiệu dụng (MB/s)	T_i (phút)
1	User01	95	800	40	5.0	324
2	User02	80	800	40	5.0	273
3	User03	105	800	40	5.0	358
4	User04	70	800	40	5.0	239

5	User05	110	800	40	5.0	375
6	User06	85	800	40	5.0	290
7	User07	90	800	40	5.0	307
8	User08	75	800	40	5.0	256
9	User09	100	800	40	5.0	341
10	User10	65	800	40	5.0	222
11	User11	95	800	40	5.0	324
12	User12	80	800	40	5.0	273
13	User13	105	800	40	5.0	358
14	User14	70	800	40	5.0	239
15	User15	110	800	40	5.0	375
16	User16	85	800	40	5.0	290
17	User17	90	800	40	5.0	307
18	User18	75	800	40	5.0	256
19	User19	100	800	40	5.0	341
20	User20	65	800	40	5.0	222

Nhận xét từ bảng 2.1, ta thấy máy có dung lượng lớn nhất (110 GB) có thời gian sao lưu khoảng 375 phút (xấp xỉ 6,25 giờ). Do hệ thống thực hiện sao lưu đồng thời với các máy trạm nên thời gian hoàn tất chu kỳ sao lưu của hệ thống được xác định theo công thức dưới đây:

$$\boldsymbol{T_{total} = \max(T_i) \approx 375 \text{ phút} \approx 6,25 \text{ giờ}}$$

Bảng 2.2. Bảng tổng hợp chỉ tiêu hiệu năng và độ tin cậy của hệ thống sao lưu hiện tại

Chỉ tiêu	Giá trị ghi nhận
Số máy trạm sao lưu đồng thời	20
Tổng dung lượng dữ liệu (xấp xỉ)	~1.7–1.9 TB
Tổng băng thông hiệu dụng trong giờ sao lưu	~800 Mbps
Băng thông trung bình mỗi máy	~40 Mbps
Thời gian hoàn tất 1 chu kỳ sao lưu	~6–6,5 giờ
Tần suất sao lưu	1 lần/ngày
RPO	~24 giờ
RTO (phục hồi file)	1–2 giờ
RTO (phục hồi toàn hệ thống)	4–6 giờ
Điểm hỏng đơn (thiết bị lưu trữ tập trung)	Có
Dự phòng địa lý	Không

Từ những số liệu phân tích trên, ta có thể thấy một số điểm hạn chế của hệ thống sao lưu hiện thời như:

- Thời gian sao lưu kéo dài do băng thông bị chia sẻ khi nhiều máy sao lưu đồng thời; thời gian hoàn tất phụ thuộc máy có dung lượng lớn nhất.
- RPO cao do tần suất sao lưu thấp (1 lần/ngày), dẫn đến nguy cơ mất dữ liệu của một ngày làm việc khi xảy ra sự cố.
- RTO cao vì phục hồi toàn hệ thống cần khôi phục lượng dữ liệu lớn, phụ thuộc hiệu năng của thiết bị lưu trữ và mạng nội bộ.
- Rủi ro điểm hỏng đơn do dữ liệu sao lưu tập trung tại một thiết bị duy nhất, chưa có cơ chế dự phòng địa lý.

Những hạn chế này cho thấy hệ thống sao lưu hiện tại chưa đáp ứng tốt yêu cầu về hiệu năng và khả năng phục hồi trong môi trường doanh nghiệp đa nền tảng, đồng thời đặt ra nhu cầu cần cải tiến theo hướng giảm thời gian sao lưu, giảm RPO/RTO và tăng độ tin cậy của hạ tầng sao lưu.

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

Với phương pháp sao lưu toàn bộ, mỗi chu kỳ sao lưu tạo ra một bản sao đầy đủ của toàn bộ dữ liệu hệ thống. Điều này dẫn đến mức tiêu thụ dung lượng lưu trữ tăng nhanh theo thời gian, đặc biệt trong bối cảnh dữ liệu phát sinh liên tục trong quá trình vận hành.

Xét hệ thống sao lưu hiện tại với dung lượng lưu trữ khả dụng ban đầu khoảng 20 TB, tổng dung lượng dữ liệu nguồn của doanh nghiệp vào khoảng 1,8 TB. Khi thực hiện sao lưu toàn bộ mỗi ngày một lần, dung lượng lưu trữ tiêu thụ theo thời gian có thể ước tính như sau:

Bảng 2.3. Ước tính mức tiêu thụ dung lượng lưu trữ theo thời gian

Thời gian lưu trữ	Số bản sao lưu toàn phần	Dung lượng tiêu thụ (xấp xỉ)
5 ngày	5	9 TB
7 ngày	7	12,6 TB
10 ngày	10	18 TB
11–12 ngày	11–12	19,8–21,6 TB

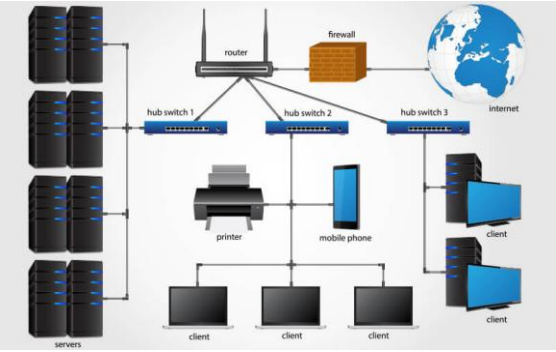
Từ bảng trên có thể thấy rằng, với dung lượng lưu trữ khả dụng ban đầu khoảng 20 TB, hệ thống sao lưu sẽ gần như đầy chỉ sau khoảng 10 đến 12 ngày vận hành. Khi dung lượng lưu trữ đạt ngưỡng, doanh nghiệp buộc phải thực hiện các biện pháp như xóa bớt bản sao lưu cũ, đóng băng dữ liệu, hoặc di chuyển dữ liệu sang thiết bị lưu trữ khác để giải phóng không gian.

Các thao tác này không chỉ làm tăng chi phí đầu tư lưu trữ mà còn phát sinh chi phí vận hành và quản trị, bao gồm thời gian xử lý thủ công, rủi ro thao tác nhầm và nguy cơ ảnh hưởng đến khả năng phục hồi dữ liệu khi cần thiết. Ngoài ra, việc không có cơ chế tối ưu dung lượng như khử trùng lặp hoặc sao lưu gia tăng khiến hệ thống khó mở rộng khi khối lượng dữ liệu tiếp tục tăng trong tương lai.

Như vậy, 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

Trong giai đoạn trước đây, khi hệ thống CNTT của doanh nghiệp chủ yếu bao gồm các máy trạm độc lập và số lượng máy chủ hạn chế, phương pháp sao lưu toàn bộ theo hình thức sao chép dữ liệu trực tiếp có thể đáp ứng được yêu cầu bảo vệ dữ liệu ở mức cơ bản. Tuy nhiên, trong bối cảnh hiện nay, hạ tầng CNTT của doanh nghiệp nhỏ đã phát triển theo hướng đa nền tảng, bao gồm đồng thời máy trạm người dùng, máy chủ vật lý và các hệ thống máy ảo, làm thay đổi đáng kể đặc điểm và yêu cầu của công tác sao lưu dữ liệu.



Hình 2.1. Mô hình hạ tầng CNTT đa nền tảng trong doanh nghiệp

Trong môi trường đa nền tảng, dữ liệu không chỉ gia tăng về dung lượng mà còn đa dạng về loại hình, tần suất thay đổi và mức độ quan trọng. Việc áp dụng một cơ chế sao lưu đồng nhất cho tất cả các nền tảng, như sao lưu toàn bộ theo chu kỳ cố định, dẫn đến tình trạng không phân biệt được đặc thù của từng hệ thống, làm giảm hiệu quả chung của quá trình sao lưu.

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.

Bên cạnh đó, phương pháp sao lưu toàn bộ không có cơ chế nhận biết và xử lý sự khác biệt về mức độ thay đổi dữ liệu giữa các nền tảng. Các máy trạm, máy chủ ứng dụng và hệ thống máy ảo đều phải sao chép lại toàn bộ dữ liệu trong mỗi chu kỳ sao lưu, mặc dù trên thực tế chỉ một phần nhỏ dữ liệu có sự thay đổi. Điều này làm gia tăng đáng kể lưu lượng truyền tải, gây áp lực lên hạ tầng mạng và thiết bị lưu trữ tập trung, đồng thời hạn chế khả năng mở rộng của hệ thống sao lưu khi số lượng nền tảng tiếp tục tăng.

Về khả năng phục hồi, môi trường đa nền tảng yêu cầu hệ thống sao lưu phải hỗ trợ khôi phục nhanh và linh hoạt cho từng loại hệ thống khác nhau. Tuy nhiên, với cách tiếp cận sao lưu toàn bộ hiện tại, thời gian phục hồi phụ thuộc lớn vào dung lượng dữ liệu cần khôi phục và hiệu năng của thiết bị lưu trữ tập trung. Các chỉ tiêu RPO và RTO ghi nhận được vẫn ở mức cao, chưa đáp ứng tốt yêu cầu duy trì hoạt động liên tục trong môi trường đa nền tảng, nơi sự gián đoạn của một hệ thống có thể ảnh hưởng đến toàn bộ chuỗi dịch vụ.

Như vậy, có thể thấy rằng vấn đề cốt lõi của hệ thống sao lưu hiện tại không chỉ nằm ở hiệu năng, mà xuất phát từ việc kiến trúc sao lưu không còn phù hợp với đặc điểm của môi trường doanh nghiệp đa nền tảng hiện nay. Đây chính là cơ sở để đặt ra các yêu cầu đối với một hệ thống sao lưu mới, được trình bày trong phần tiếp theo.

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

Từ kết quả phân tích thực trạng hệ thống sao lưu hiện tại ở mục 2.1, có thể thấy rằng hệ thống sao lưu toàn phần đang được sử dụng tồn tại nhiều hạn chế mang tính hệ thống. Với thời gian hoàn tất một chu kỳ sao lưu khoảng 6 đến 6,5 giờ, RPO xấp xỉ 24 giờ, RTO phục hồi toàn hệ thống từ 4 đến 6 giờ và tốc độ tiêu thụ dung lượng lưu trữ rất nhanh (20 TB chỉ đáp ứng khoảng 10 đến 12 ngày), hệ thống hiện tại không còn phù hợp với yêu cầu vận hành của doanh nghiệp trong môi trường đa nền tảng. Do đó, hệ thống sao lưu mới cần được thiết kế để đáp ứng các yêu cầu định lượng cụ thể sau.

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

Theo số liệu ở mục 2.1, việc sao lưu toàn bộ dữ liệu trong môi trường nhiều máy trạm sao lưu đồng thời khiến thời gian hoàn tất chu kỳ sao lưu kéo dài tới 6 đến 6,5 giờ, chủ yếu do băng thông bị chia sẻ và phải sao chép lại toàn bộ dữ liệu ở mỗi chu kỳ. Thời gian sao lưu dài không chỉ làm giảm khả năng tăng tần suất sao lưu mà còn tiềm ẩn nguy cơ ảnh hưởng đến hoạt động của hệ thống nếu quá trình sao lưu kéo dài sang giờ làm việc.

Vì vậy, hệ thống sao lưu mới cần đáp ứng yêu cầu rút ngắn thời gian hoàn tất chu kỳ sao lưu tối thiểu 50% và hướng tới 70% so với hệ thống hiện tại. Điều này tương đương với việc đưa thời gian sao lưu toàn hệ thống từ mức 6-6,5 giờ xuống còn khoảng 2–3 giờ, ngay cả khi thực hiện sao lưu đồng thời trên nhiều nền tảng. Đây là điều kiện tiên quyết để hệ thống sao lưu có thể vận hành ổn định trong môi trường đa nền tảng.

2.3.2. Yêu cầu về RPO và RTO

Hệ thống sao lưu hiện tại thực hiện sao lưu theo chu kỳ 1 lần/ngày, dẫn đến RPO khoảng 24 giờ, đồng nghĩa với nguy cơ mất dữ liệu của một ngày làm việc khi xảy ra sự cố. Trong môi trường doanh nghiệp đa nền tảng, nơi các hệ thống phụ thuộc lẫn nhau, mức RPO này là không chấp nhận được.

Do đó, hệ thống sao lưu mới cần hướng tới việc giảm RPO từ 24 giờ xuống dưới 3 giờ, cho phép doanh nghiệp hạn chế tối đa lượng dữ liệu có nguy cơ bị mất. Đồng thời, RTO phục hồi toàn hệ thống cần được rút ngắn tối thiểu 50%, từ mức 4–6 giờ hiện tại xuống còn dưới 2–3 giờ, nhằm giảm thiểu thời gian gián đoạn hoạt động và đảm bảo tính liên tục của các dịch vụ trong môi trường đa nền tảng.

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

Phân tích ở mục 2.1 cho thấy hệ thống sao lưu hiện tại phụ thuộc vào một thiết bị lưu trữ tập trung duy nhất, tạo ra điểm hỏng đơn (Single Point of Failure). Khi thiết bị này gặp sự cố, toàn bộ khả năng phục hồi dữ liệu của doanh nghiệp bị ảnh hưởng nghiêm trọng, làm gia tăng rủi ro mất dữ liệu.

Do đó, hệ thống sao lưu mới cần được thiết kế theo hướng loại bỏ điểm hỏng đơn, đảm bảo rằng dữ liệu sao lưu không phụ thuộc vào một thành phần lưu trữ duy nhất. Kiến trúc hệ thống phải duy trì khả năng phục hồi dữ liệu ngay cả khi một phần hạ tầng gặp sự cố, từ đó nâng cao độ tin cậy tổng thể so với mô hình sao lưu tập trung hiện tại.

2.3.4. Yêu cầu về khả năng mở rộng

Với phương pháp sao lưu toàn bộ, thời gian sao lưu và mức tiêu thụ tài nguyên tăng gần như tuyến tính theo số lượng hệ thống và dung lượng dữ liệu, khiến hệ thống hiện tại khó mở rộng khi doanh nghiệp bổ sung thêm máy trạm, máy chủ hoặc hệ thống ảo hóa.

Hệ thống sao lưu mới cần đảm bảo rằng việc mở rộng số lượng nền tảng không làm tăng tương ứng thời gian sao lưu vượt quá ngưỡng 2–3 giờ đã đặt ra, đồng thời không gây áp lực lớn lên băng thông mạng nội bộ. Đây là yêu cầu quan trọng để hệ thống sao lưu có thể đáp ứng nhu cầu phát triển trong dài hạn mà không phải thay đổi toàn bộ kiến trúc.

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

Theo phân tích ở mục 2.1.2, với dung lượng lưu trữ khả dụng khoảng 20 TB, hệ thống sao lưu toàn bộ hiện tại có thể bị lấp đầy chỉ sau 10–12 ngày vận hành, do mỗi chu kỳ sao lưu tạo ra một bản sao đầy đủ của toàn bộ dữ liệu hệ thống. Điều này buộc doanh nghiệp phải thường xuyên xóa bản sao lưu cũ hoặc mở rộng hạ tầng lưu trữ, làm tăng chi phí đầu tư và chi phí vận hành.

Vì vậy, hệ thống sao lưu mới cần giảm tốc độ gia tăng dung lượng lưu trữ theo thời gian từ 60–80% so với phương pháp sao lưu toàn bộ. Việc đạt được mức tối ưu này sẽ giúp kéo dài đáng kể thời gian sử dụng hạ tầng lưu trữ hiện có, giảm tần suất mở rộng dung lượng và đơn giản hóa công tác quản trị hệ thống sao lưu.

Từ các yêu cầu định lượng nêu trên, có thể thấy rằng hệ thống sao lưu mới cần mang lại sự cải thiện rõ rệt so với hệ thống hiện tại về thời gian sao lưu (từ 6–6,5 giờ xuống dưới 2–3 giờ), RPO (từ 24 giờ xuống dưới 1 giờ), RTO (từ 4–6 giờ xuống dưới 2–3 giờ) và hiệu quả sử dụng dung lượng lưu trữ

(giảm 60–80%). Đây là cơ sở kỹ thuật rõ ràng để đề xuất và đánh giá giải pháp sao lưu bất đồng bộ và phân tán trong các chương tiếp theo.

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 2 đã phân tích thực trạng hệ thống sao lưu và phục hồi dữ liệu đang được triển khai trong một doanh nghiệp vừa và nhỏ với môi trường CNTT đa nền tảng. Thông qua việc khảo sát mô hình sao lưu, phương thức vận hành và các số liệu thực tế, chương đã làm rõ đặc điểm của hệ thống sao lưu toàn bộ theo mô hình tập trung đang được sử dụng.

Kết quả phân tích cho thấy hệ thống sao lưu hiện tại tuy đơn giản và dễ triển khai nhưng bộc lộ nhiều hạn chế khi áp dụng trong môi trường đa nền tảng. Cụ thể, thời gian sao lưu toàn hệ thống kéo dài 6–6,5 giờ, RPO khoảng 24 giờ và RTO phục hồi toàn hệ thống từ 4–6 giờ chưa đáp ứng tốt yêu cầu đảm bảo tính liên tục của hoạt động. Bên cạnh đó, việc phụ thuộc vào một thiết bị lưu trữ duy nhất tạo ra rủi ro điểm hỏng đơn, trong khi phương pháp sao lưu toàn bộ làm dung lượng lưu trữ tiêu thụ nhanh, gây áp lực lớn lên hạ tầng và chi phí vận hành.

Từ những hạn chế này, chương đã xác định các yêu cầu đối với hệ thống sao lưu mới, bao gồm việc rút ngắn thời gian sao lưu xuống khoảng 2–3 giờ, giảm RPO xuống dưới 3 giờ, rút ngắn RTO tối thiểu 50%, đồng thời giảm 60–80% tốc độ gia tăng dung lượng lưu trữ so với hệ thống cũ.

CHƯƠNG 3
THIẾT KẾ VÀ TRIỂN KHAI
HỆ THỐNG SAO LƯU DỮ LIỆU

3.1. Phát biểu bài toán

Trong bối cảnh doanh nghiệp vừa và nhỏ vận hành hạ tầng CNTT đa nền tảng (máy trạm người dùng, máy chủ vật lý và hệ thống máy ảo), dữ liệu trở thành tài sản quan trọng cần được bảo vệ liên tục. Tuy nhiên, mô hình sao lưu truyền thống theo hướng sao lưu toàn phần (Full Backup) bằng sao chép trực tiếp về một thiết bị lưu trữ tập trung tuy đơn giản nhưng bộc lộ nhiều hạn chế trong vận hành thực tế.

Kết quả phân tích ở Chương 2 cho thấy hệ thống hiện tại có thời gian sao lưu toàn hệ thống kéo dài khoảng 6–6,5 giờ, RPO xấp xỉ 24 giờ, RTO phục hồi toàn hệ thống 4–6 giờ, đồng thời tồn tại điểm hỏng đơn do phụ thuộc vào một thiết bị lưu trữ tập trung. Về chi phí lưu trữ, với phương pháp sao lưu toàn phần theo chu kỳ ngày, dung lượng khả dụng khoảng 20 TB có thể bị lấp đầy sau 10–12 ngày, gây phát sinh công việc vận hành thủ công và chi phí mở rộng hạ tầng.

Do đó, 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.

Trong phạm vi đề án, bài toán được giải quyết bằng mô hình hệ thống sao lưu phân tán dựa trên thiết bị NAS, kết hợp cơ chế đảm bảo sẵn sàng tại từng địa điểm (HA) và cơ chế sao chép dữ liệu bất đồng bộ giữa hai địa điểm (site) nhằm tăng khả năng chịu lỗi và phục hồi khi xảy ra sự cố nghiêm trọng. Hệ thống cần hỗ trợ bảo vệ dữ liệu cho nhiều nền tảng (máy trạm, máy chủ vật lý, máy ảo) và cho phép phục hồi linh hoạt theo nhiều mức độ (tệp/thư mục, toàn máy, máy ảo).

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

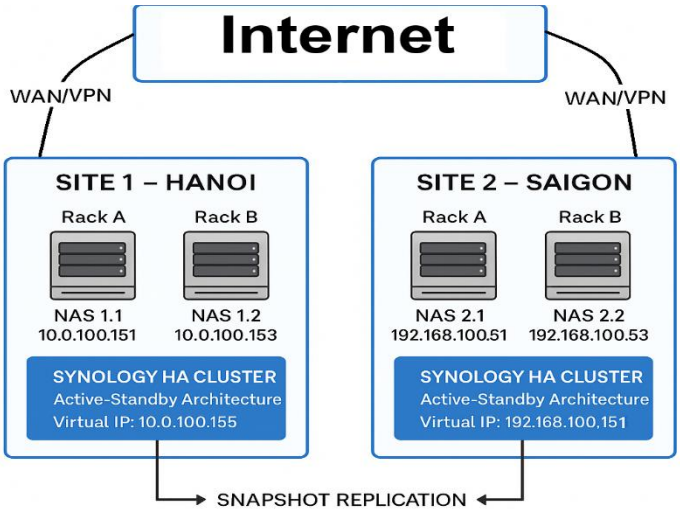
Mô hình đề xuất trong mục này được xây dựng nhằm đáp ứng trực tiếp các yêu cầu định lượng đã nêu ở bảng 2.4, đặc biệt về thời gian sao lưu, RPO/RTO và loại bỏ điểm hỏng đơn

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 Thành phố Hồ Chí Minh. Việc bố trí hạ tầng tại hai khu vực cách biệt nhau mang lại lợi thế lớn về khả năng chống chịu thảm họa (Disaster Resilience), đặc biệt trong các kịch bản như mất điện diện rộng, cháy nổ trung tâm dữ liệu hoặc sự cố hạ tầng mạng tại một khu vực. Khi một site gặp sự cố, site còn lại vẫn có thể duy trì hoạt động, tiếp nhận dữ liệu và cung cấp khả năng phục hồi cho toàn bộ hệ thống.

Tại mỗi site, hai thiết bị NAS được đặt ở hai tủ rack độc lập, sử dụng các nguồn điện khác nhau để giảm thiểu rủi ro mất toàn bộ cụm lưu trữ do sự cố phần cứng. Hai NAS tại mỗi site được cấu hình thành một cụm Synology High Availability (HA) hoạt động theo cơ chế Active–Passive, đảm bảo dịch vụ lưu trữ vẫn tiếp tục ngay cả khi một thiết bị ngừng hoạt động. Hai site kết nối qua đường truyền VPN nội bộ riêng của doanh nghiệp, cho phép thực hiện replication snapshot hai chiều, từ đó duy trì được các bản sao dữ liệu cập nhật của nhau tại mọi thời điểm.

Mô hình hai site song song kèm với mỗi site có khả năng sao lưu và phục hồi độc lập giúp hệ thống đạt được tính liên tục dịch vụ cao, hạn chế tối đa thời gian gián đoạn và đảm bảo an toàn dữ liệu trước các sự cố nghiêm trọng.

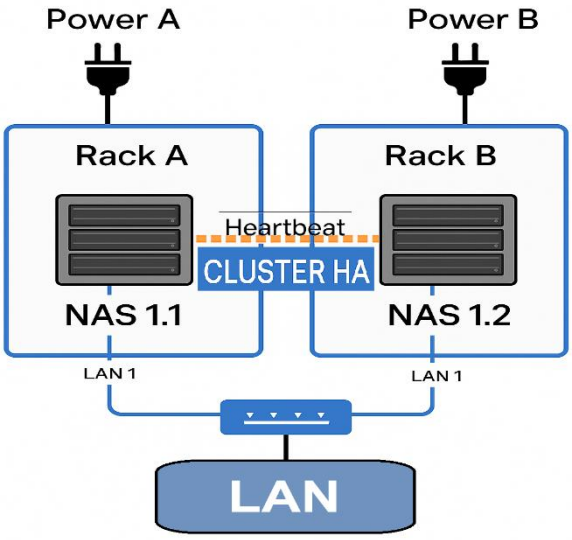


Hình 3.1. Kiến trúc tổng thể hệ thống sao lưu phân tán tại hai địa điểm

3.2.2. Bố trí thiết bị tại mỗi Site

Site 1 (Hà Nội) được bố trí hai thiết bị NAS vật lý đặt tại hai tủ rack độc lập trong cùng một trung tâm dữ liệu. NAS 1.1 được đặt tại tủ Rack A và sử dụng nguồn điện riêng, trong khi NAS 1.2 được đặt tại tủ Rack B với đường điện tách biệt. Hai thiết bị này được cấu hình tham gia cụm Synology High Availability (SHA), với địa chỉ IP riêng cho từng node (10.0.100.151 và 10.0.100.153) và một địa chỉ IP ảo dùng chung cho cụm là 10.0.100.155. Cách bố trí này giúp loại bỏ điểm lỗi đơn tại cấp độ phần cứng NAS hoặc tủ rack.

Site 2 (Thành phố Hồ Chí Minh) được thiết kế theo kiến trúc hoàn toàn tương tự Site 1. Hai thiết bị NAS 2.1 và NAS 2.2 được đặt tại hai tủ rack khác nhau trong cùng hạ tầng trung tâm dữ liệu, mỗi tủ sử dụng một đường cấp nguồn độc lập. Địa chỉ IP giả định của hai node lần lượt là 192.168.100.51 và 192.168.100.53, trong khi địa chỉ IP ảo phục vụ truy cập cụm là 192.168.100.151. Về mặt thiết kế, cluster tại Site 2 hoạt động tương tự cluster tại Site 1, đảm bảo khả năng chịu lỗi ở mức node và tủ rack, đồng thời cung cấp một đầu mối truy cập thống nhất cho các tác vụ sao lưu và phục hồi dữ liệu của các hệ thống thuộc Site 2.

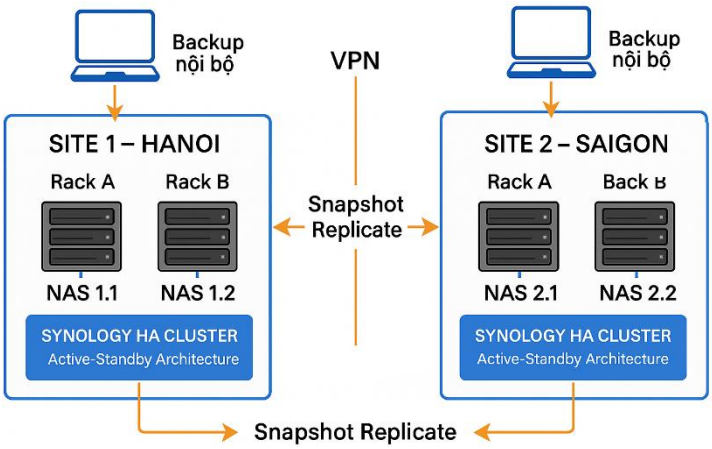


Hình 3.2. Sơ đồ bố trí phần cứng và kết nối của cụm lưu trữ tại Site 1

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

Trong mô hình được đề xuất, mỗi site được xem là một miền độc lập về mặt vận hành, chịu trách nhiệm bảo vệ dữ liệu cho các tài nguyên CNTT đặt tại chính site đó. Các máy trạm và máy chủ thuộc Site 1 sẽ cấu hình agent kết nối đến địa chỉ IP ảo của cluster NAS tại Site 1 (10.0.100.155), trong khi các hệ thống thuộc Site 2 sẽ kết nối đến địa chỉ IP ảo của cluster NAS tại Site 2 (192.168.100.155). Như vậy, việc sao lưu luôn được ưu tiên thực hiện vào hạ tầng lưu trữ tại chỗ, giảm độ trễ truy cập và không phụ thuộc đường truyền liên tỉnh.

Sau khi dữ liệu sao lưu được ghi nhận tại mỗi cụm NAS, hệ thống tiếp tục thực hiện cơ chế sao chép snapshot bất đồng bộ giữa hai site. Các thư mục chứa dữ liệu backup quan trọng tại Site 1 sẽ được replicate định kỳ sang Site 2, và ngược lại, dữ liệu backup tương ứng tại Site 2 cũng được replicate sang Site 1. Kết quả là tại bất kỳ thời điểm nào, mỗi site đều lưu trữ được ảnh chụp dữ liệu sao lưu của cả hai site, đảm bảo rằng nếu một site gặp sự cố nghiêm trọng, site còn lại vẫn đủ dữ liệu để phục hồi cho toàn bộ tập máy trạm và máy chủ của doanh nghiệp.



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

Hạ tầng lưu trữ của hệ thống sao lưu được tổ chức đồng nhất tại cả hai site nhằm đơn giản hóa việc vận hành và đảm bảo tính nhất quán về mặt kiến trúc. Mỗi site sử dụng một cụm NAS cấu hình theo mô hình High Availability, dựa trên hệ điều hành DSM và hệ thống tập tin Btrfs. Trong phạm vi luận văn, quá trình triển khai chi tiết được trình bày tập trung cho Site 1, trong khi các bước cấu hình tại Site 2 được thực hiện tương tự và chỉ được tóm lược ở mức khái quát.

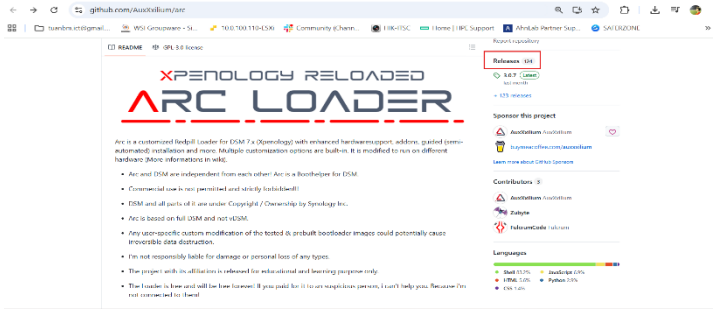
3.3.1. Thiết kế và triển khai cụm NAS tại Site 1

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.

a) Chuẩn bị bộ cài đặt ARC Loader cho DSM

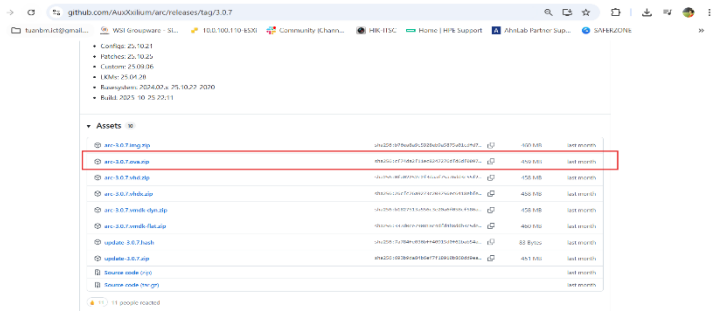
Để triển khai hệ điều hành DSM trên môi trường ảo hóa ESXi, bộ ARC Loader cần được tải xuống từ kho mã nguồn công khai của dự án. ARC

Loader đóng vai trò là bộ nạp khởi động, cho phép cài đặt DSM trên phần cứng không chính thức.



Hình 3.4. Giao diện tải ARC Loader trên GitHub

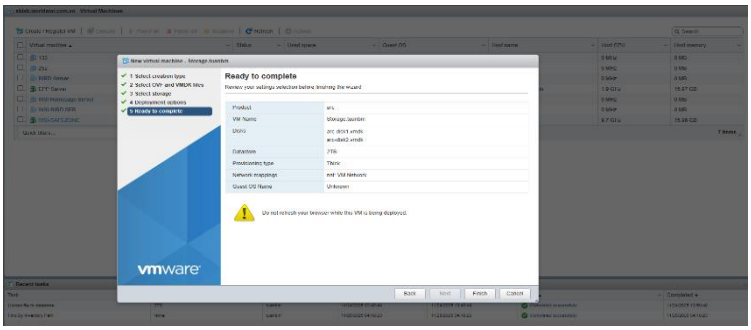
Sau khi truy cập mục *Releases*, tệp cài đặt tương ứng phiên bản 3.0.7 được tải xuống, bao gồm cả tệp OVA/ZIP dùng để triển khai máy ảo trên hypervisor.



Hình 3.5. Minh họa tệp arc-3.0.7.vmdk/ova

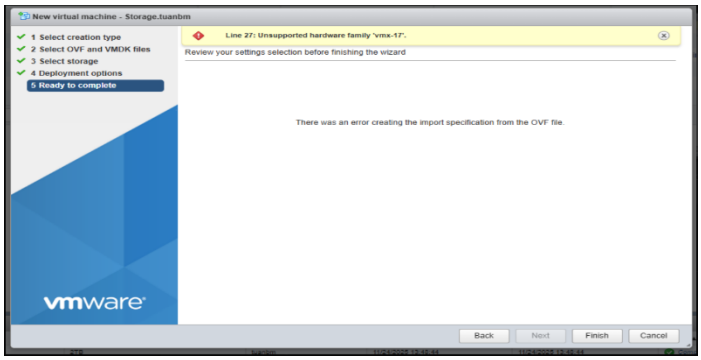
b) Triển khai máy ảo NAS trên ESXi

Quá trình cài đặt NAS 1.1 và NAS 1.2 được thực hiện bằng cách import tệp OVA lên ESXi. Khi thực hiện triển khai, giao diện ESXi hiển thị thông tin cấu hình bao gồm loại VM, dung lượng đĩa, RAM và card mạng.



Hình 3.6. Màn hình "Ready to complete" khi deploy OVA

Tuy nhiên, trong quá trình import, hệ thống ESXi xuất hiện cảnh báo *Unsupported hardware family vmx-17*, cho thấy phiên bản OVA mặc định sử dụng *Virtual Hardware Version 17*, vốn chỉ được hỗ trợ từ ESXi 7.0 trở lên. Do Site 1 chạy trên nền ESXi 6.7, tệp OVA cần được chỉnh sửa để tương thích.



Hình 3.7. Lỗi "Unsupported hardware family vmx-17"

c) Điều chỉnh tương thích phiên bản phần cứng ảo hóa

Để khắc phục lỗi trên, phần cài đặt cần được điều chỉnh:

B1: Đổi phần mở rộng `.ova` → `.tar`.

B2: Giải nén để lấy file `.ovf`.

B3: Mở file `.ovf` và chỉnh:

```
<vmw:HardwareVersion>vmx-17</vmw:HardwareVersion>
→ <vmw:HardwareVersion>vmx-14</vmw:HardwareVersion>
```

B4: Nén lại thành `.ova` và triển khai bình thường.

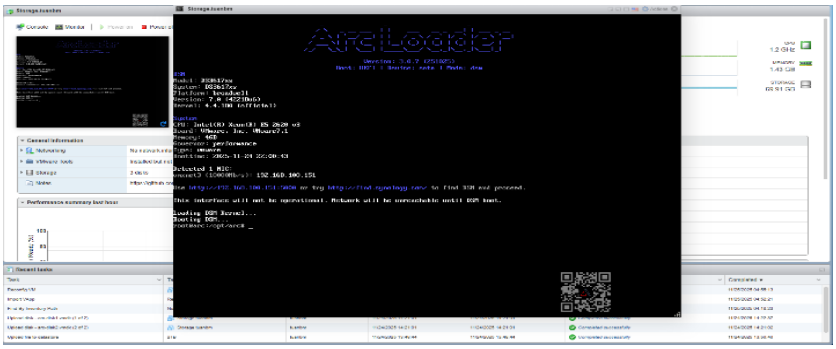
Quy trình này giúp OVA tương thích hoàn toàn với ESXi 6.7, đảm bảo quá trình triển khai diễn ra ổn định.

d) Cài đặt ARC Loader và lựa chọn model NAS

Khi máy ảo khởi động, ARC Loader xuất hiện, cho phép lựa chọn model NAS tương thích với DSM. Người triển khai lựa chọn model **DS3617xs**, phù hợp với phần cứng và hỗ trợ đầy đủ phiên bản DSM dự kiến sử dụng.

Sau đó, phiên bản DSM tương ứng được chọn, trong luận văn sử dụng phiên bản DSM 7.0.1-42218, đảm bảo tương thích với hệ thống và các ứng dụng dự định triển khai.

Khi hoàn tất, ARC Loader chuyển sang chế độ boot DSM, sẵn sàng cho quá trình cài đặt qua trình duyệt.



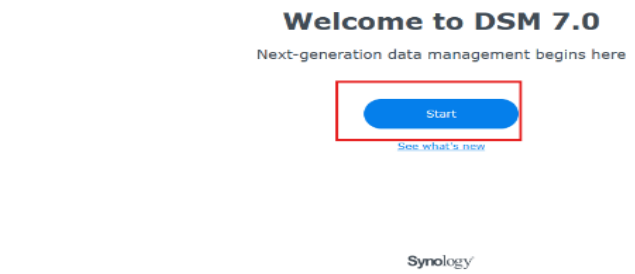
Hình 3.8. Màn hình ARC Loader khi boot vào DSM.

e) Cài đặt DSM và cấu hình hệ thống NAS

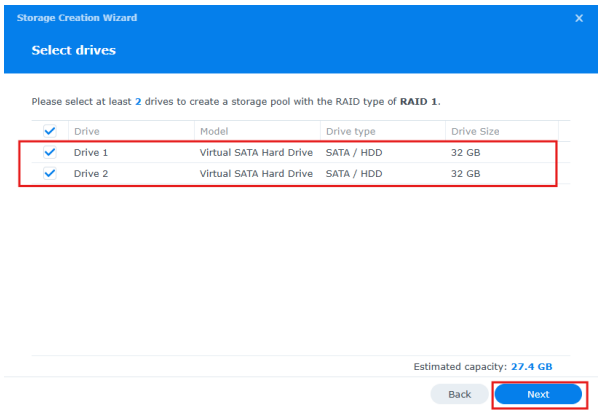
Người triển khai sử dụng địa chỉ hiển thị trên màn hình ARC Loader (ví dụ như http://192.168.100.151:5000) để truy cập giao diện Web Assistant của Synology và tiến hành cài đặt DSM cho từng NAS.

Trong quá trình này, người triển khai sẽ cần thực hiện như sau:

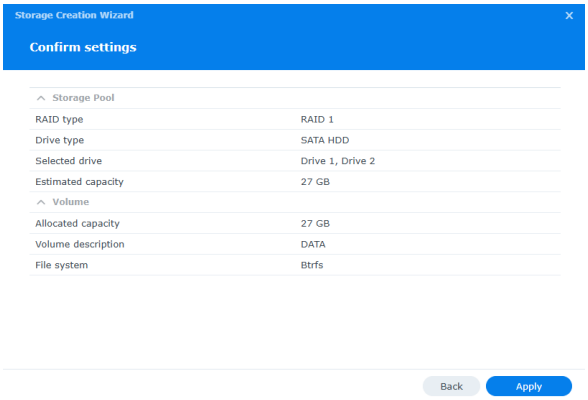
- Tạo tên thiết bị tương ứng: *NAS 1.1* và *NAS 1.2*
- Thiết lập tài khoản quản trị
- Khởi tạo volume Btrfs
- Gán IP tĩnh cho từng NAS theo sơ đồ mạng của Site 1



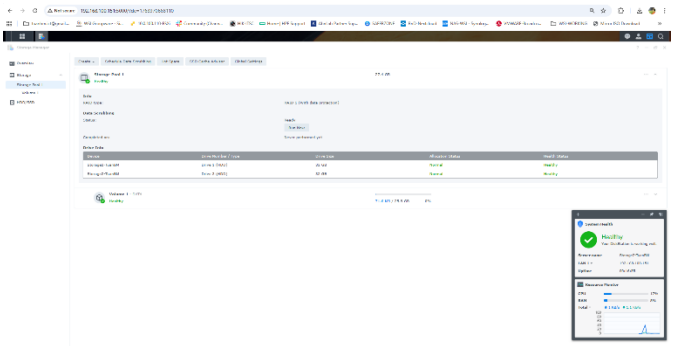
Hình 3.9. Màn hình khởi động quá trình thiết lập hệ thống



Hình 3.10. Lựa chọn các ổ đĩa để tạo Storage Pool với cấu hình RAID



Hình 3.11. Xác nhận cấu hình Storage Pool và Volume



Hình 3.12. Storage Pool và Volume sau khi khởi tạo

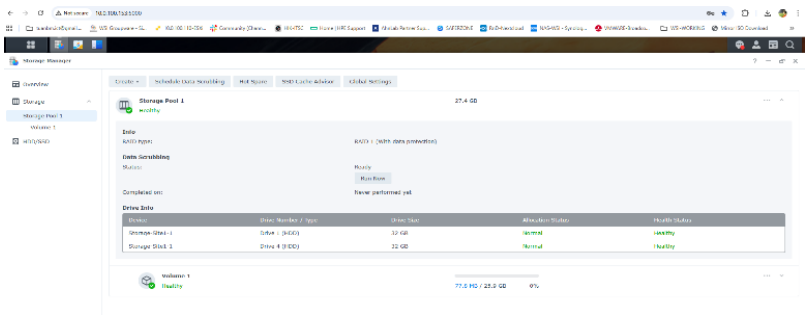
f) Triển khai cụm Synology High Availability (SHA) tại Site 1

Sau khi hoàn tất cài đặt DSM và cấu hình các Volume ban đầu, hai thiết bị NAS 1.1 và NAS 1.2 tại Site 1 được tiến hành ghép cụm Synology High Availability (SHA). Mục tiêu của SHA là đảm bảo hệ thống lưu trữ tại Site 1 luôn duy trì khả năng hoạt động liên tục ngay cả khi một NAS gặp sự cố. Cụm HA sử dụng hai kết nối mạng độc lập:

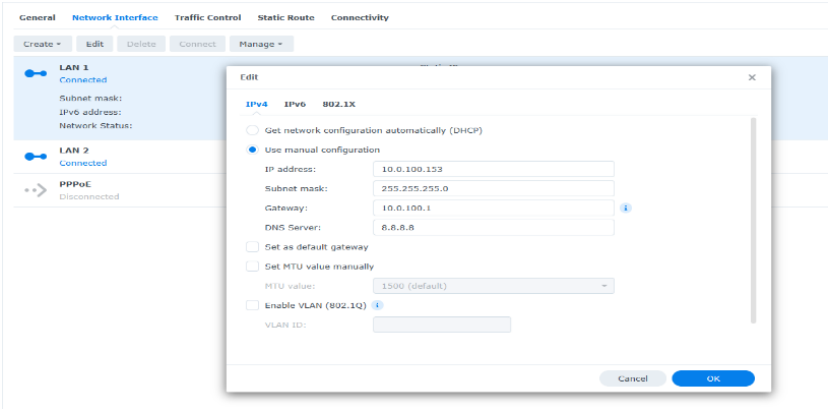
LAN 1: phục vụ truy cập dịch vụ và lưu lượng backup nội bộ.

LAN 2: dành riêng cho **Heartbeat**, đảm bảo đồng bộ block-level giữa hai NAS.

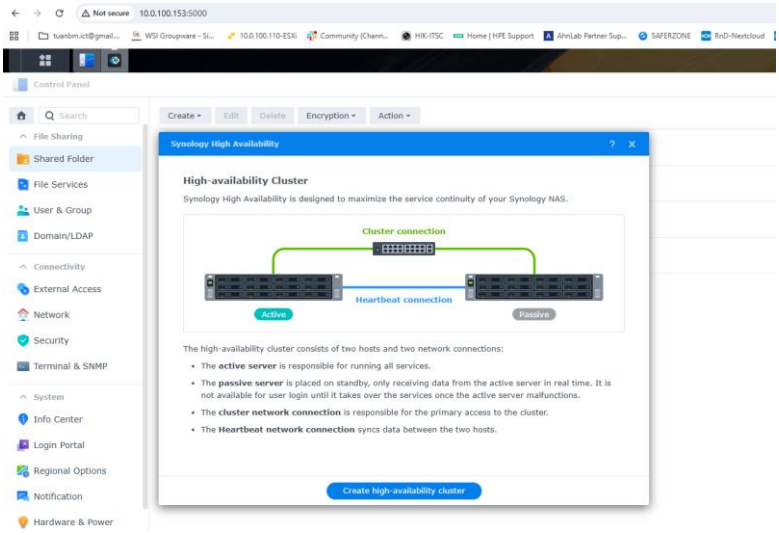
Quy trình thiết lập được thực hiện thông qua giao diện DSM, bao gồm: cấu hình IP tĩnh cho từng NAS, kiểm tra trạng thái Volume, khai báo vai trò Active–Passive, lựa chọn giao diện mạng cho cụm, nhập thông tin xác thực của node Passive và cuối cùng tạo cụm HA với IP ảo 10.0.100.155. Sau khi hoàn tất, hệ thống bắt đầu quá trình đồng bộ dữ liệu giữa NAS Active và NAS Passive. Khi trạng thái chuyển sang *Healthy*, cụm HA sẵn sàng phục vụ các tác vụ sao lưu tại Site 1.



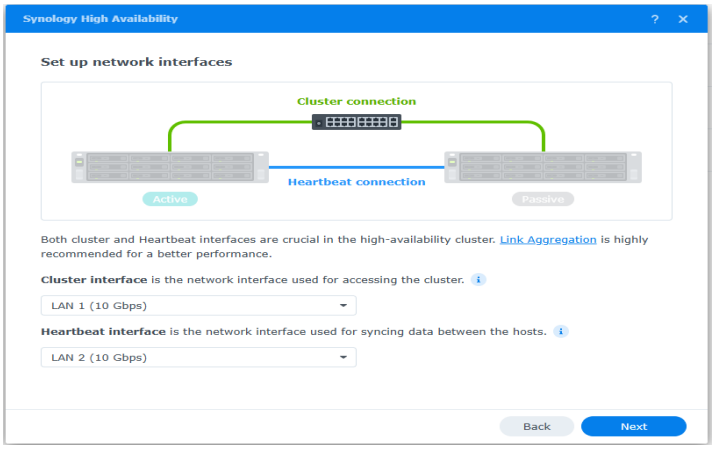
Hình 3.13. Trạng thái Storage Pool và Volume trên NAS 1.1



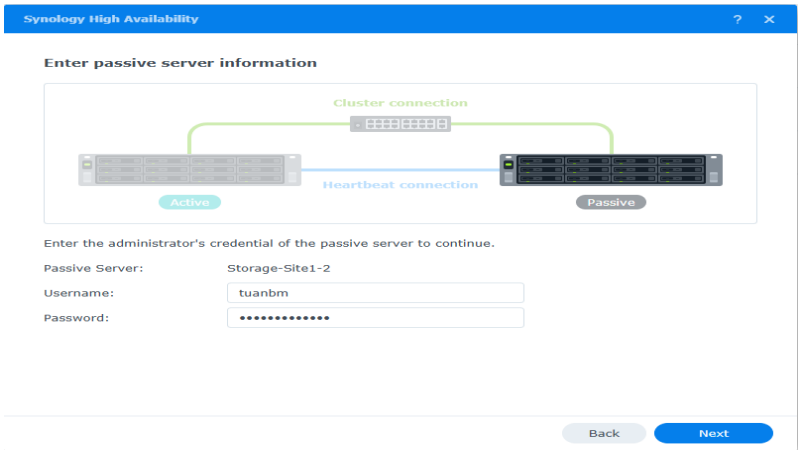
Hình 3.14. Cấu hình địa chỉ IP cho từng NAS trong Site 1



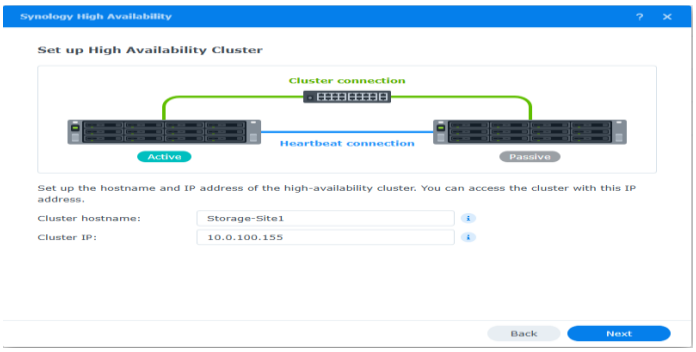
Hình 3.15. Khởi tạo kết nối High Availability Cluster trên DSM



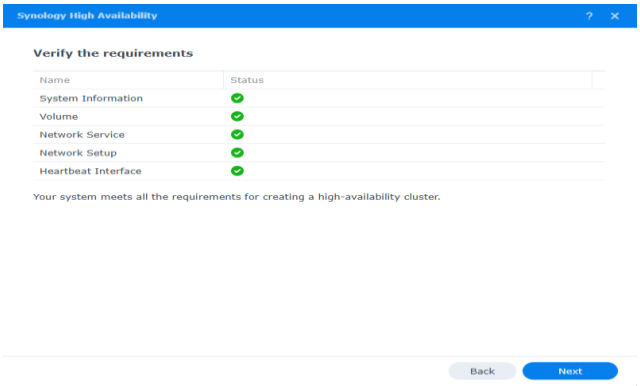
Hình 3.16. Lựa chọn giao diện mạng cho Cluster và Heartbeat



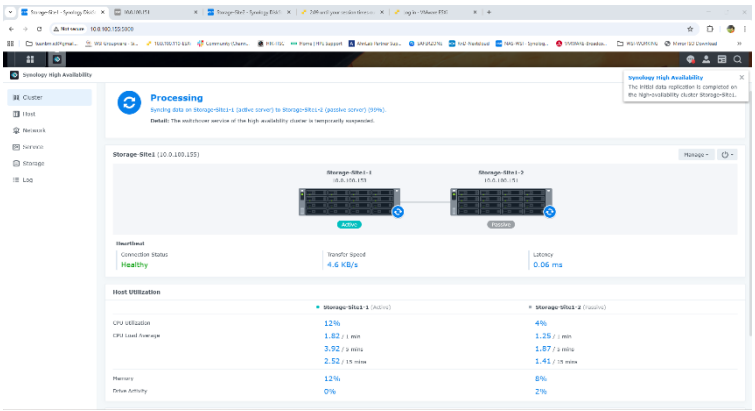
Hình 3.17. Xác thực máy chủ Passive trong cụm HA



Hình 3.18. Khai báo hostname và IP ảo của cụm HA trong Site 1



Hình 3.19. Kiểm tra các điều kiện để tạo cụm Synology HA



Hình 3.20. Hình thành cụm HA và bắt đầu đồng bộ dữ liệu trong Site 1

Cơ chế HA tại từng site giúp đảm bảo hệ thống sao lưu vẫn hoạt động khi một thành phần gặp sự cố, từ đó rút ngắn thời gian gián đoạn và đáp ứng yêu cầu RTO của hệ thống mới.

3.3.2. Thiết kế và triển khai cụm NAS tại Site 2 (Thành phố Hồ Chí Minh)

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.

Bên cạnh đó, toàn bộ quy trình tạo Storage Pool, cấu hình LAN 1 cho dịch vụ và LAN 2 cho Heartbeat, cũng như các bước ghép cụm HA được thực hiện hoàn toàn tương tự Site 1. Khi quá trình đồng bộ hoàn tất và cụm HA tại Site 2 khởi động ổn định với IP ảo **192.168.100.155**, hệ thống lưu trữ tại Site 2 chính thức sẵn sàng tham gia vào hoạt động sao lưu nội bộ và cơ chế replicate hai chiều giữa hai site.

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

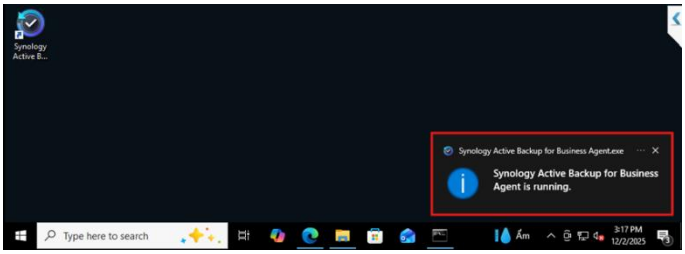
3.4.1. Cấu trúc sao lưu và giao thức kết nối

Hệ thống sao lưu nội bộ được xây dựng dựa trên Active Backup for Business (ABB), giải pháp backup đa nền tảng của Synology cho phép bảo vệ đồng thời máy Windows, Linux và máy chủ ảo. Các tác vụ sao lưu từ các thiết bị trong nội bộ site đều kết nối đến địa chỉ IP ảo của cụm HA, bảo đảm quá trình backup không bị gián đoạn khi một NAS trong cụm gặp sự cố.

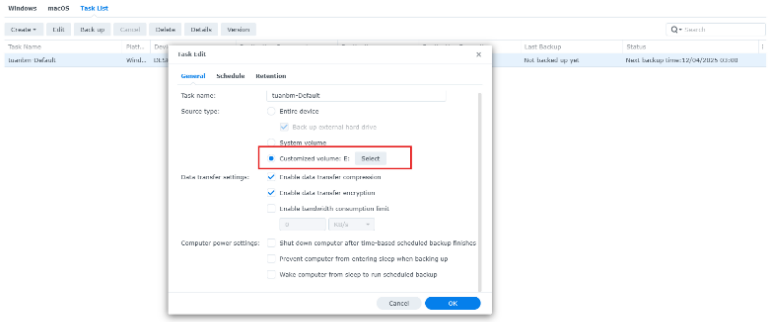
Kiến trúc truyền thông của ABB sử dụng giao thức **gRPC kết hợp TLS** trên các cổng **5510/5511**, giúp đảm bảo tính bảo mật cũng như hiệu quả truyền tải. Đồng thời, ABB áp dụng cơ chế **Changed Block Tracking (CBT)** để ghi nhận chỉ các khối dữ liệu thay đổi, nhờ đó tối ưu hóa dung lượng lưu trữ và giảm thời gian thực hiện phiên backup.

3.4.2. Cài đặt và cấu hình sao lưu máy Windows

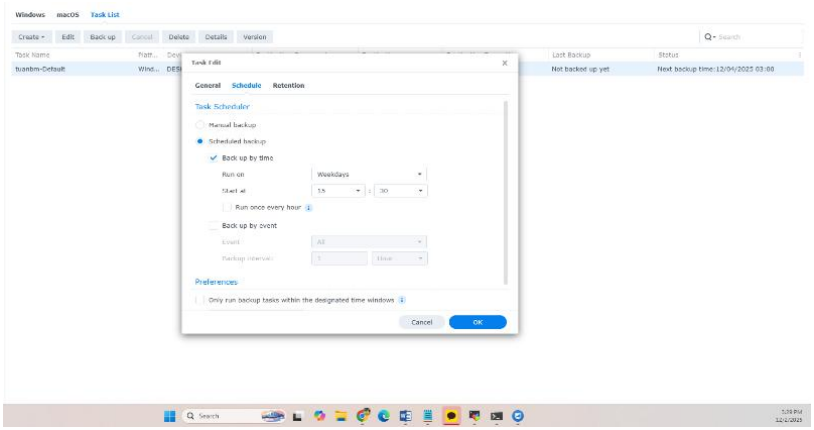
ABB Agent được cài đặt trên các máy Windows để tự động gửi dữ liệu sao lưu về cụm NAS tại Site tương ứng. Trong lần sao lưu đầu tiên, hệ thống tạo bản **Full Backup** nhằm ghi nhận toàn bộ trạng thái của máy. Các phiên sao lưu tiếp theo được thực hiện dưới dạng **Incremental Backup**, chỉ ghi nhận phần dữ liệu thay đổi so với phiên trước đó, giúp giảm tải mạng và tăng tốc độ lưu trữ.



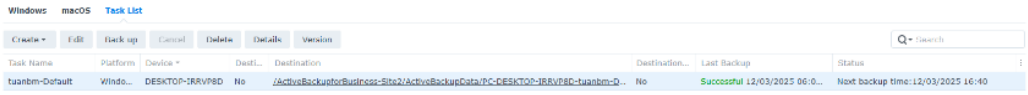
Hình 3.21. Cài đặt thành công ABB agent trên Windows



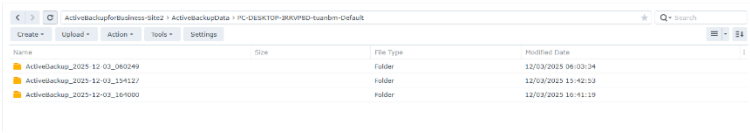
Hình 3.22. Thiết lập vùng để sao lưu trên Windows



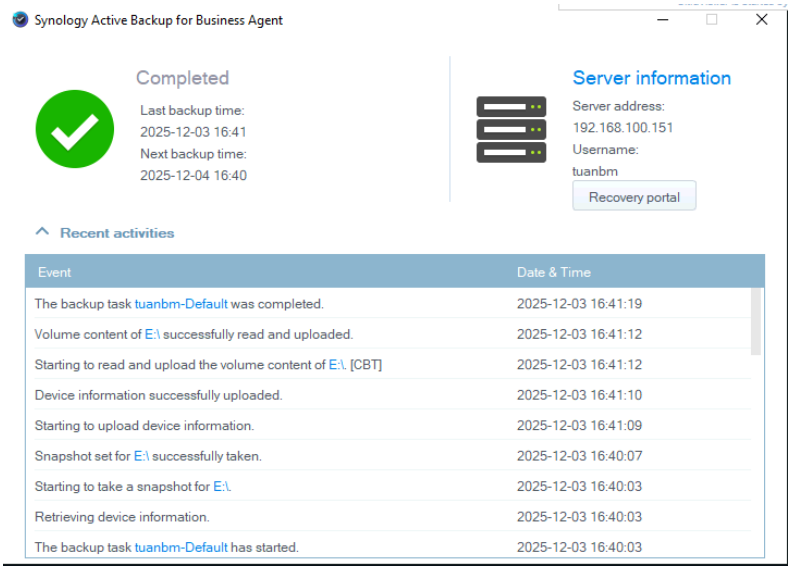
Hình 3.23. Thiết lập thời gian để sao lưu trên Windows



Hình 3.24. Sao lưu máy Windows thành công



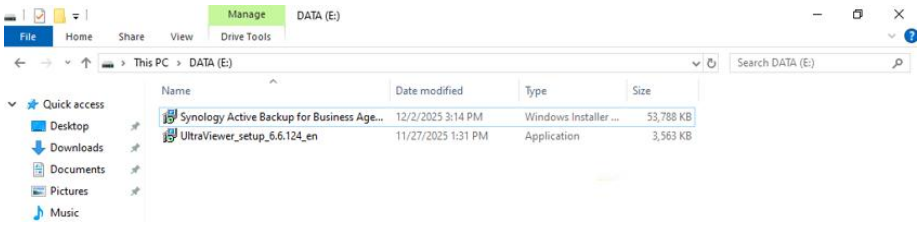
Hình 3.25. Thư mục kết quả các lần sao lưu thành công



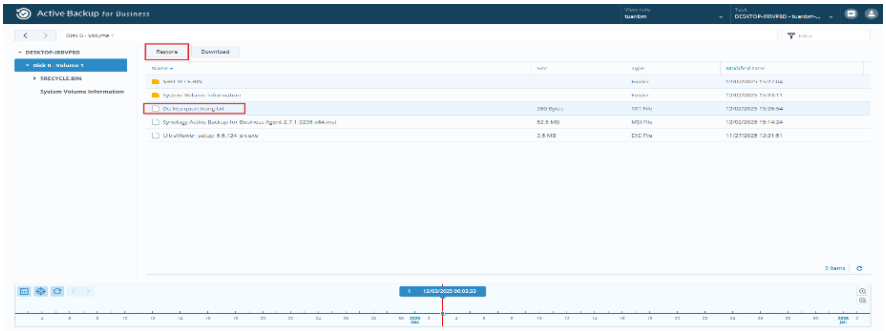
Hình 3.26. Nhật ký hiển thị trên máy trạm Windows sau khi sao lưu

3.4.3. Thử nghiệm mất file và khôi phục dữ liệu trên Windows

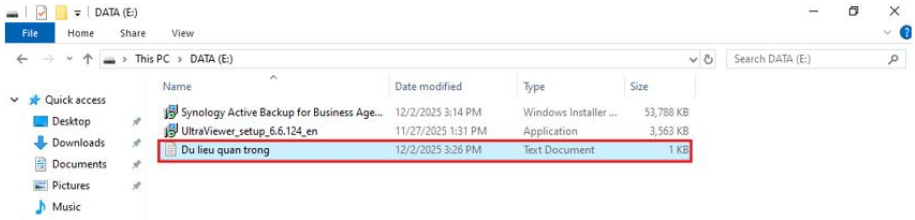
Để kiểm thử khả năng khôi phục, một tệp "Dữ liệu quan trọng" được xóa trên máy Windows. ABB cung cấp cơ chế **File-Level Restore**, cho phép lựa chọn phiên bản snapshot trước thời điểm xảy ra sự cố. Kết quả cho thấy tệp được khôi phục đầy đủ, đúng nội dung, chứng minh tính chính xác và độ tin cậy của giải pháp.



Hình 3.27. Tệp bị xoá nhầm trên Windows



Hình 3.28. Tệp được khôi phục từ NAS



Hình 3.29. Tệp sau khi khôi phục thành công bằng ABB

3.4.4. Cài đặt và cấu hình sao lưu máy Linux (Ubuntu)

Đối với các máy Linux, ABB sử dụng agent dạng `.run`. Quy trình triển khai bao gồm:

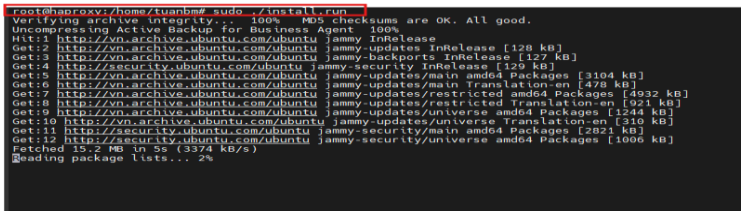
B1: Giải nén và cài đặt agent trên hệ điều hành

B2: Gắn thêm ổ đĩa 3 GB phục vụ thử nghiệm backup block-level

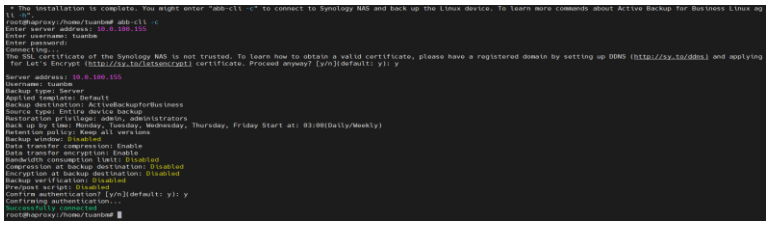
B3: Cập nhật tệp `/etc/fstab` để cấu hình tự động mount ổ đĩa

B4: Tiến hành sao lưu thử nghiệm

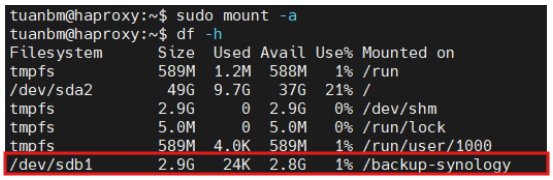
Phiên backup được thực hiện thành công, xác nhận ABB hoạt động ổn định trên môi trường Linux.



Hình 3.30. Cài đặt agent theo dạng tệp .run

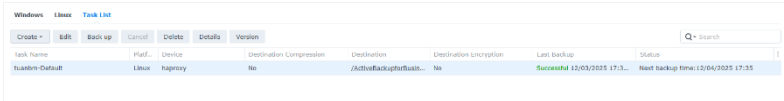


Hình 3.31. Cài đặt và kết nối agent trên máy linux thành công

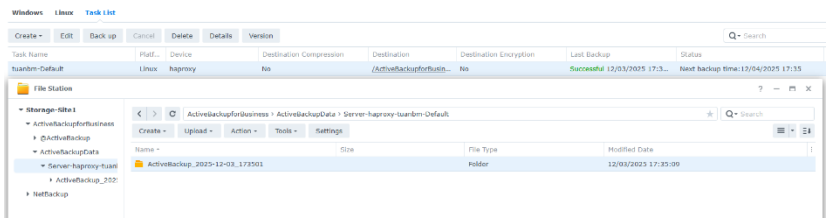


Hình 3.32. Thêm ổ cứng 3G để tiến hành thử nghiệm sao lưu

Tiến hành thiết lập các thông số sao lưu cho máy Linux tương tự như với máy Windows ở trên NAS cluster. Ta thu được kết quả.



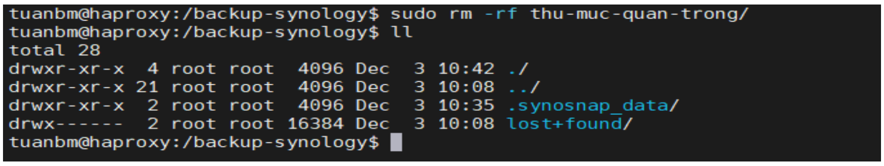
Hình 3.33. Sao lưu thành công máy Linux



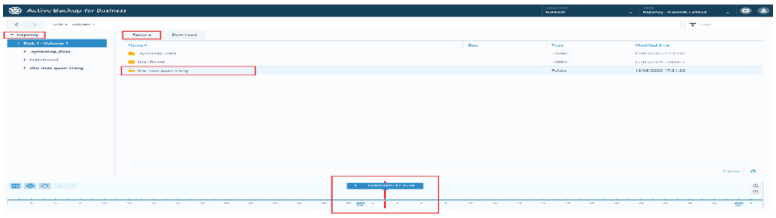
Hình 3.34. Thư mục tệp kết quả sao lưu

3.4.5. Thử nghiệm mất file và khôi phục dữ liệu trên Linux

Một thư mục thử nghiệm được xóa và thực hiện khôi phục từ ABB. Kết quả cho thấy dữ liệu được phục hồi đúng trạng thái ban đầu, đảm bảo khả năng khôi phục granular (từng tệp/thư mục) tương tự Windows.



Hình 3.35. Trường hợp mất tệp dữ liệu trên linux



Hình 3.36. Khôi phục tệp bị mất trên máy linux từ NAS

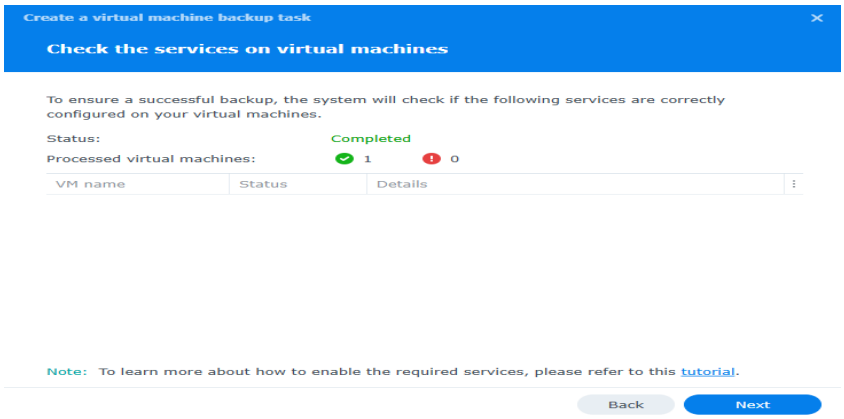


Hình 3.37. Tệp được khôi phục thành công trên máy linux

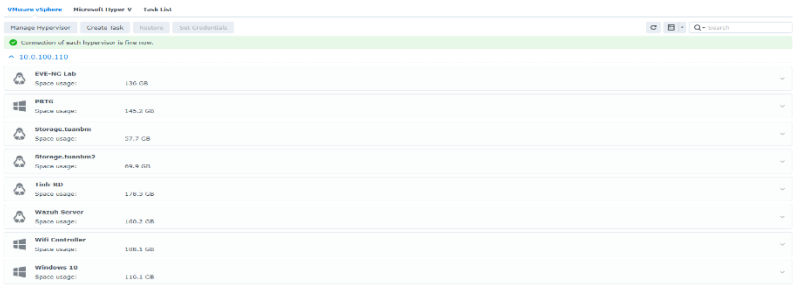
3.4.6. Thiết kế sao lưu máy chủ ảo (Virtual Machine)

Bên cạnh sao lưu máy trạm và máy chủ vật lý, ABB còn cho phép bảo vệ máy chủ ảo chạy trên nền tảng VMware ESXi. Cơ chế sao lưu máy ảo dựa trên snapshot VM được thực hiện trực tiếp từ hypervisor mà không yêu cầu cài đặt agent trong máy ảo. ABB giao tiếp với ESXi thông qua kết nối API và tiến hành sao lưu theo lịch định sẵn, bao gồm cả Full và Incremental Backup.

Ưu điểm của phương pháp này là hạn chế ảnh hưởng đến hiệu năng vận hành của máy ảo, đồng thời đảm bảo khả năng phục hồi nhanh trong trường hợp xảy ra sự cố.



Hình 3.38. Kiểm tra dịch vụ của các máy ảo



Hình 3.39. Sao lưu thành công máy chủ ảo ESXi

3.5. Thiết kế cơ chế Replication giữa hai Site

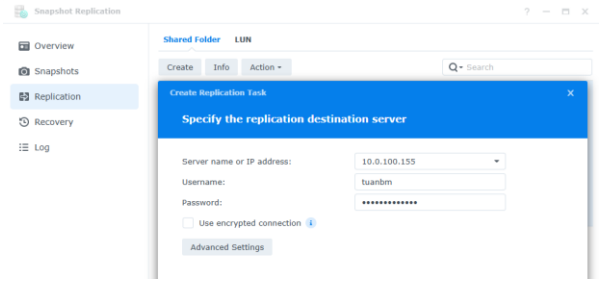
3.5.1. Mục tiêu và nguyên tắc thiết kế

Cơ chế Snapshot Replication được lựa chọn nhằm bảo đảm cho hệ thống sao lưu duy trì khả năng dự phòng địa lý theo mô hình phân tán và bất đồng bộ. Không tồn tại sự phân cấp giữa hai site; Site 1 và Site 2 vận hành như hai nút ngang hàng, mỗi site vừa đóng vai trò nguồn gửi snapshot, vừa đồng thời là đích nhận snapshot của site còn lại. Thiết kế này giúp toàn bộ dữ liệu backup của doanh nghiệp được duy trì đồng thời tại hai vị trí địa lý, bảo đảm khả năng khôi phục trong kịch bản một site mất hoàn toàn vì sự cố hạ tầng hoặc thảm họa.

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.

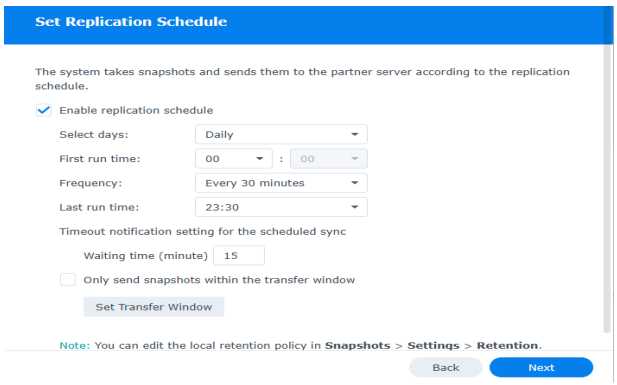
3.5.2. Replication snapshot từ Site 1 sang Site 2

Quá trình cấu hình replicate tại Site 1 bắt đầu bằng việc khai báo địa chỉ IP của cụm HA tại Site 2 cùng với thông tin xác thực của tài khoản quản trị. Sau khi kết nối được thiết lập thành công, hệ thống tiến hành tạo tác vụ replicate và cho phép người quản trị định nghĩa lịch chạy tự động, ví dụ gửi snapshot mỗi 30 phút. Cách làm này giúp tối ưu hóa độ tươi của dữ liệu trong khi vẫn kiểm soát được tải mạng.



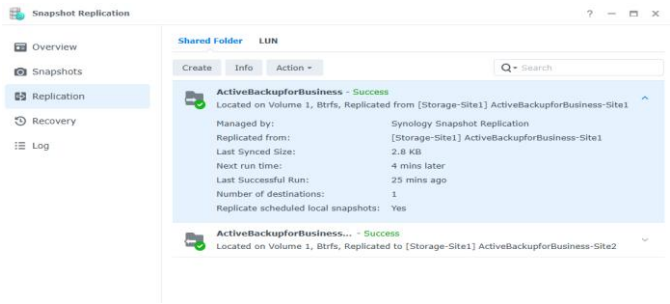
Hình 3.40. Khai báo máy chủ đích replication

Trong bước tiếp theo, lịch chạy replication được cấu hình ở chế độ hằng ngày với tần suất 30 phút/lần. Đây là mức phù hợp với đặc thù khối lượng dữ liệu backup phát sinh tại doanh nghiệp vừa và nhỏ, đồng thời duy trì mức độ rủi ro mất dữ liệu ở mức tối thiểu.



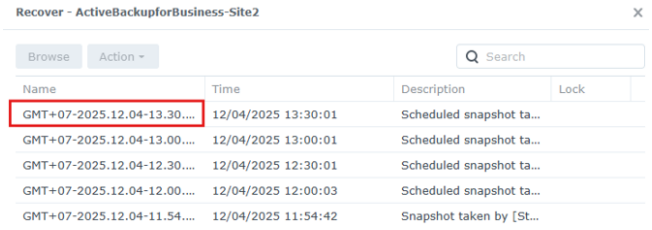
Hình 3.41. Cấu hình lịch replicate chu kỳ 30 phút

Sau khi lịch được kích hoạt, snapshot đầu tiên được tạo và truyền sang Site 2. Khi quá trình replicate hoàn tất, giao diện Snapshot Replication tại Site 1 ghi nhận trạng thái *Success* kèm kích thước dữ liệu đã đồng bộ.



Hình 3.42. Kết quả replicate thành công từ Site 1 sang Site 2

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.

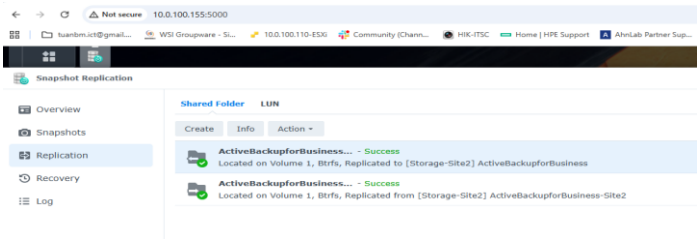


Hình 3.43. Các snapshot từ Site 1 hiển thị trong mục Recovery tại Site 2

3.5.3. Replication snapshot từ Site 2 sang Site 1

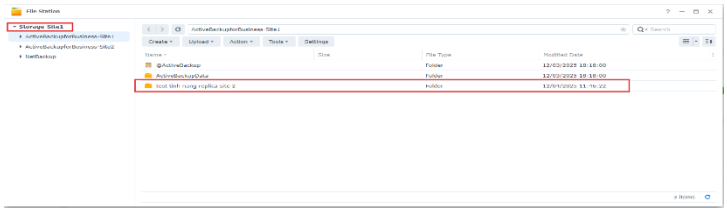
Tương tự quy trình tại Site 1, Site 2 cũng cấu hình replicate snapshot ngược trở lại Site 1. Việc đồng bộ hai chiều tạo ra một vòng lặp dự phòng hoàn chỉnh, bảo đảm mỗi site luôn duy trì dữ liệu mới nhất của site còn lại.

Sau khi kích hoạt replicate, việc kiểm tra trực tiếp trên File Station cho thấy các thư mục được tạo tại Site 2 xuất hiện đầy đủ trên Site 1 sau một chu kỳ replicate. Ngược lại, khi tạo dữ liệu tại Site 1, snapshot replication giúp dữ liệu được đồng bộ đầy đủ sang Site 2 trong thời gian ngắn.

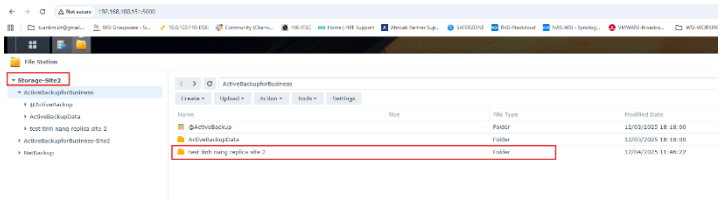


Hình 3.44. Replicate thành công từ Site 2 sang Site 1

Để kiểm thử, trên Site 1 tạo một folder Test tính năng replica, và sau khi quá trình replicate được thực hiện, trên Site 2 đã nhận được folder đó. Như vậy là tính năng hoạt động ổn định cao.



Hình 3.45. Tạo folder thử nghiệm trên Site 1

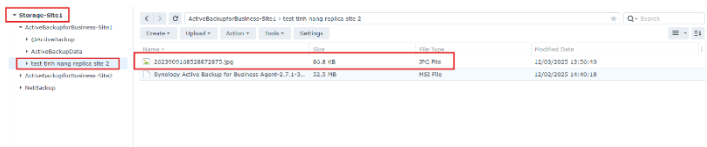


Hình 3.46. Folder đã xuất hiện trên Site 2 sau quá trình replicate

Nhờ cơ chế này, hệ thống luôn duy trì hai bản sao dữ liệu backup ở hai vị trí hoàn toàn tách biệt về địa lý.

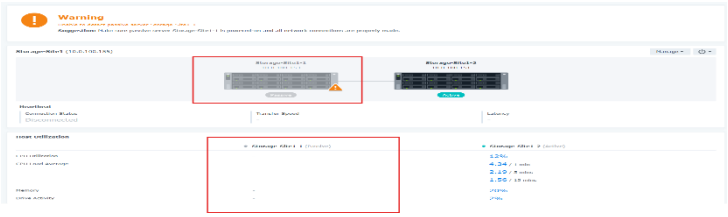
3.5.4. Phân tích an toàn và khả năng chịu lỗi

Cơ chế Snapshot Replication hai chiều giữa Site 1 và Site 2 mang lại khả năng bảo vệ dữ liệu vượt trội nhờ đặc tính lưu trữ snapshot của hệ thống tập tin Btrfs. Snapshot tồn tại dưới dạng cấu trúc metadata và block-level tách biệt khỏi vùng dữ liệu người dùng, vì vậy mọi thao tác chỉnh sửa, xóa nhầm hoặc sự cố mã hóa dữ liệu (ransomware) đều không ảnh hưởng trực tiếp đến snapshot. Đây là cơ sở để hệ thống duy trì các điểm khôi phục an toàn, ngay cả khi dữ liệu vận hành bị hỏng hoặc bị mã độc tấn công. Trong quá trình kiểm thử, thư mục thử nghiệm được tạo tại Site 1 và replicate sang Site 2 thành công theo đúng lịch trình. Việc kiểm tra trên File Station cho thấy các tệp tin đã xuất hiện đầy đủ tại phía đối diện.



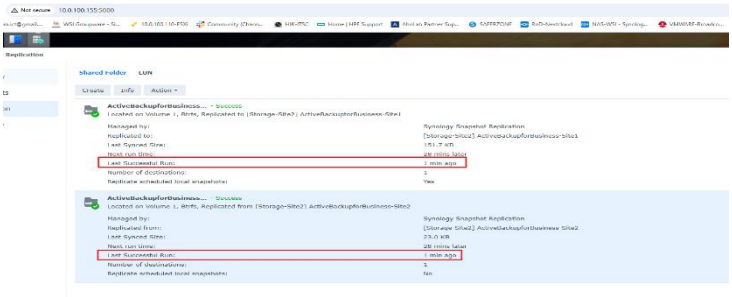
Hình 3.47. Tệp kiểm thử tại Site 1 replicate sang Site 2

Một thử nghiệm sự cố được thực hiện bằng cách tắt NAS đóng vai trò Active tại Site 1. Giao diện HA ghi nhận tình trạng mất kết nối Heartbeat, tuy nhiên node Passive đã nhanh chóng chiếm vai trò của node Active và tiếp tục duy trì dịch vụ và không gây gián đoạn các tác vụ replicate.



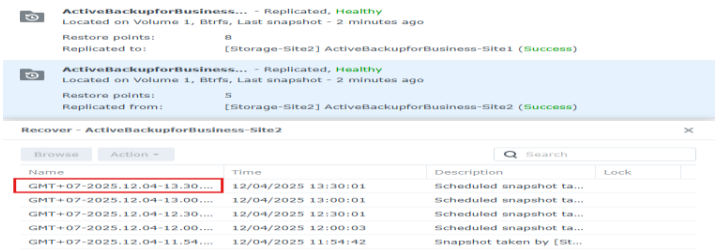
Hình 3.48. NAS Passive chuyển thành Active duy trì dịch vụ

Nhật ký Snapshot Replication tiếp tục ghi nhận các phiên đồng bộ mới thực hiện thành công, chứng minh rằng cụm HA vẫn đảm bảo tính sẵn sàng cho quá trình truyền snapshot ngay cả khi một node gặp lỗi.



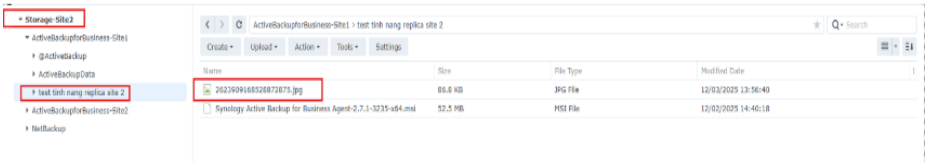
Hình 3.49. Replicate vẫn thành công trên cả hai chiều khi sự cố xảy ra

Tại Site 2, danh sách snapshot nhận từ Site 1 hiển thị đầy đủ, xác nhận rằng sự cố phần cứng tại một NAS không ảnh hưởng đến tính toàn vẹn của dữ liệu xuyên site.



Hình 3.50. Danh sách snapshot phục hồi tại Site 2 sau sự cố

Ngược lại, dữ liệu tạo tại Site 1 cũng replicate đến Site 2 đầy đủ sau chu kỳ, cho thấy cơ chế đồng bộ hai chiều vẫn hoạt động ổn định.



Hình 3.51. Tệp từ Site 1 replicate thành công đến Site 2

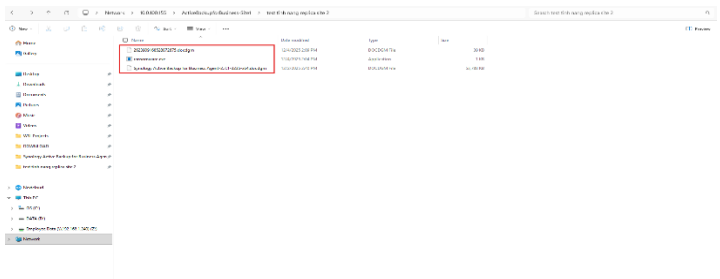
Từ các thử nghiệm trên có thể kết luận ngắn gọn như sau:

- 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

3.6.1. Mô phỏng tấn công Ransomware

Để đánh giá khả năng bảo vệ dữ liệu trong tình huống bị mã hóa, một kịch bản tấn công ransomware được tiến hành trên thư mục dữ liệu đang được bảo vệ bởi Snapshot Replication. Một tệp **ransomware.exe** được đưa vào thư mục và thực thi. Ngay sau đó, toàn bộ các tệp trong thư mục bị mã hóa theo thuật toán AES và đổi đuôi sang *.docdgm*, khiến người dùng không thể truy cập nội dung ban đầu. Hình dưới minh họa trạng thái thư mục sau khi bị mã hóa hoàn toàn:



Hình 3.52. Các tệp bị mã hoá ransomware chuyển sang đuôi .docdgm

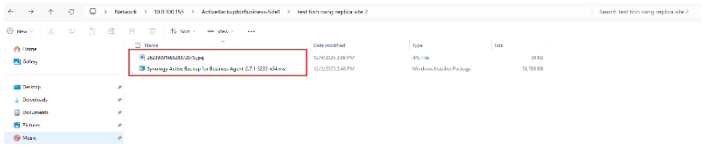
Quá trình này mô phỏng rất sát các kịch bản tấn công ransomware phổ biến trong thực tế, nơi mã độc thường hoạt động thông qua giao thức SMB, mã hóa toàn bộ dữ liệu mà người dùng có quyền truy cập.

3.6.2. Khôi phục dữ liệu từ snapshot an toàn

Sau khi ransomware mã hóa toàn bộ thư mục, tác vụ khôi phục được thực hiện bằng cách chọn snapshot gần nhất trước thời điểm xảy ra tấn công. Việc khôi phục được thực hiện ngay trên giao diện DSM, không cần thao tác dòng lệnh hay ngừng hệ thống.

Quá trình restore diễn ra nhanh nhờ cơ chế copy-on-write của Btrfs, hệ thống chỉ trả lại các block dữ liệu đã ghi nhận trong snapshot mà không ảnh hưởng đến các tệp khác trên volume.

Kết quả khôi phục cho thấy toàn bộ dữ liệu bị mã hóa đã trở lại trạng thái bình thường, đúng với nội dung trước thời điểm tấn công.



Hình 3.53. Dữ liệu trong thư mục được khôi phục thành công

3.6.3. Phân tích cơ chế chống ransomware của Snapshot Btrfs

Hiệu quả khôi phục đạt được trong thử nghiệm xuất phát từ đặc tính bảo vệ tự nhiên của snapshot trên hệ thống tập tin Btrfs. Khác với các thư mục chia sẻ mà người dùng truy cập qua SMB hoặc NFS, 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.

Nhờ đó, mặc dù toàn bộ tệp trong thư mục dữ liệu đã bị mã hóa, snapshot vẫn giữ nguyên vẹn và có thể được dùng để phục hồi toàn bộ trạng thái của thư mục trong vài phút.

Bên cạnh đó, việc snapshot được replicate sang site còn lại giúp tạo thêm một lớp bảo vệ thứ hai. Ngay cả khi toàn bộ NAS tại một site bị tấn công hoặc bị mã hóa, các snapshot tại site đối diện vẫn tồn tại đầy đủ, bảo đảm khả năng phục hồi sau thảm họa.

Cơ chế này giúp hệ thống lưu trữ luôn duy trì được mức độ an toàn rất cao trước các cuộc tấn công ransomware, đồng thời đáp ứng yêu cầu phục hồi dữ liệu nhanh chóng cho doanh nghiệp.

3.7. Kết luận

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.

Cụm High Availability tại mỗi site giúp duy trì dịch vụ ngay cả khi một node phần cứng gặp sự cố, trong khi cơ chế Snapshot Replication hai chiều đảm bảo dữ liệu luôn được đồng bộ liên tục ở cả hai site. Cùng với đó, Active Backup for Business hỗ trợ sao lưu tập trung cho Windows, Linux và máy chủ ảo, tạo thành một giải pháp thống nhất, linh hoạt và phù hợp với môi trường doanh nghiệp đa hệ điều hành.

Kết quả đánh giá hệ thống thông qua các thử nghiệm thực tế, từ backup, restore, replicate cho đến mô phỏng tấn công ransomware đều cho thấy hệ thống không chỉ đáp ứng yêu cầu bảo vệ dữ liệu cơ bản mà còn đảm bảo an toàn trước các mối đe dọa hiện đại. Snapshot Btrfs, vốn tồn tại ở tầng hệ thống tập tin, đã chứng minh khả năng chống mã hóa một cách tự nhiên, cho phép khôi phục dữ liệu nhanh chóng và chính xác ngay cả khi toàn bộ thư mục làm việc bị tấn công.

Tổng thể, chương 3 cho thấy rằng đây là một giải pháp sao lưu và phục hồi phân tán phù hợp đối với doanh nghiệp nhỏ và vừa, nơi nhu cầu bảo vệ dữ

liệu ngày càng lớn nhưng nguồn lực hạ tầng và vận hành còn hạn chế. Hệ thống được xây dựng trong chương này vì vậy không chỉ mang tính khả thi mà còn là mô hình mẫu để áp dụng trong thực tiễn.

CHƯƠNG 4
ĐÁNH GIÁ HIỆU QUẢ VÀ SO SÁNH

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

Việc đánh giá hiệu quả hệ thống sao lưu đề xuất được thực hiện trong bối cảnh hệ thống đã được triển khai tại hai site độc lập, mỗi site có quy mô tương đương hệ thống hiện tại đã phân tích ở Chương 2, với khoảng 20 máy trạm và máy chủ cần được bảo vệ. Mục tiêu của phần này là đánh giá mức độ đáp ứng của hệ thống mới đối với các chỉ tiêu về hiệu năng, độ tin cậy và tối ưu tài nguyên đã được xác định ở Chương 2.

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

Tại mỗi site, hệ thống sao lưu được triển khai trên nền tảng Synology NAS kết hợp Active Backup for Business, thực hiện sao lưu đồng thời cho khoảng 20 máy trạm và máy chủ. Cơ chế sao lưu gia tăng dựa trên Changed Block Tracking giúp hệ thống chỉ sao chép các khối dữ liệu thay đổi sau lần sao lưu đầu tiên, từ đó giảm đáng kể thời gian sao lưu so với phương pháp sao lưu toàn bộ truyền thống.

Bảng 4.1. Thời gian sao lưu tại mỗi site của hệ thống đề xuất mới

Các chỉ tiêu	Giá trị ghi nhận
Tổng số máy trạm/máy chủ	~20
Tổng dung lượng dữ liệu	~1,7–1,9 TB
Thời gian sao lưu lần đầu	~2,5–3 giờ
Thời gian sao lưu gia tăng	20–40 phút
Thời gian hoàn tất chu kỳ sao lưu	~2–3 giờ
Ảnh hưởng đến hiệu năng người dùng	Thấp

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

Hệ thống đề xuất sử dụng cơ chế sao chép dữ liệu bất đồng bộ giữa hai site, đảm bảo dữ liệu sao lưu tại mỗi site được đồng bộ sang site còn lại theo chu kỳ ngắn. Trong các kịch bản thử nghiệm, độ trễ sao chép giữa hai site được duy trì trong giới hạn cho phép, phù hợp với yêu cầu về RPO của hệ thống.

Bảng 4.2. Đánh giá RPO và RTO của hệ thống đề xuất mới

Chỉ tiêu	Kết quả đạt được
RPO giữa hai site	< 3 giờ
RTO phục hồi file/thư mục	5–15 phút
RTO phục hồi máy ảo	15–30 phút
RTO phục hồi toàn site	< 2–3 giờ

Các kết quả trên 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

Nhờ áp dụng cơ chế sao lưu gia tăng kết hợp snapshot trên hệ thống file Btrfs, hệ thống đề xuất làm giảm đáng kể tốc độ gia tăng dung lượng lưu trữ so với phương pháp sao lưu toàn bộ. Việc chỉ lưu trữ các khối dữ liệu thay đổi giúp tiết kiệm không gian lưu trữ và kéo dài thời gian sử dụng hạ tầng hiện có.

Bảng 4.3. So sánh mức tiêu thụ dung lượng lưu trữ tại mỗi site

Thời gian vận hành	Sao lưu toàn bộ (ước tính)	Hệ thống đề xuất
5 ngày	~9 TB	~3 TB
7 ngày	~12,6 TB	~4–5 TB
10 ngày	~18 TB	~5–7 TB

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

Hệ thống được triển khai với cụm NAS HA tại mỗi site và cơ chế Snapshot Replication giữa hai site, giúp loại bỏ điểm hỏng đơn và tăng khả năng chịu lỗi của toàn bộ hạ tầng sao lưu. Trong các kịch bản thử nghiệm mô phỏng sự cố thiết bị hoặc tấn công ransomware, dữ liệu sao lưu được bảo vệ bởi snapshot Btrfs không bị ảnh hưởng bởi quá trình mã hóa dữ liệu tại hệ thống nguồn.

Khi xảy ra sự cố, hệ thống cho phép khôi phục nhanh dữ liệu về trạng thái trước thời điểm tấn công trong khoảng 10–15 phút đối với các tập dữ liệu thông thường. Việc tồn tại bản sao dữ liệu tại site còn lại cũng đảm bảo khả năng phục hồi ngay cả trong trường hợp một site bị gián đoạn hoàn toàn, đáp ứng yêu cầu về độ tin cậy và tính sẵn sàng của hệ thống sao lưu mới.

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

Trên cơ sở các số liệu phân tích ở Chương 2 và kết quả đánh giá hệ thống sao lưu đề xuất ở mục 4.1, phần này tiến hành so sánh trực tiếp giữa hệ thống sao lưu hiện tại và hệ thống sao lưu mới nhằm làm rõ mức độ cải thiện về hiệu năng, độ tin cậy và hiệu quả sử dụng tài nguyên.

4.2.1. So sánh các chỉ tiêu kỹ thuật chính

Bảng 4.4. So sánh hệ thống sao lưu cũ và hệ thống đề xuất mới

Các chỉ tiêu	Hệ thống hiện tại	Hệ thống đề xuất
Số site triển khai	1 site	2 site (mỗi site ~20 máy)
Phương pháp sao lưu	Sao lưu toàn bộ	Sao lưu gia tăng kết hợp snapshot
Thời gian sao lưu toàn hệ thống	6–6,5 giờ	~2–3 giờ
Tần suất sao lưu	1 lần/ngày	Nhiều lần/ngày
RPO	~24 giờ	< 3 giờ
RTO phục hồi toàn hệ thống	4–6 giờ	< 2–3 giờ
Điểm hỏng đơn	Có	Đã loại bỏ
Dự phòng địa lý	Không	Có (2 site độc lập)
Tốc độ tăng dung lượng lưu trữ	Rất nhanh	Giảm ~60–70%
Khả năng chống ransomware	Thấp	Cao (snapshot Btrfs)

4.2.2. Nhận xét kết quả so sánh

Từ Bảng 4.4 có thể thấy hệ thống sao lưu đề xuất mang lại sự cải thiện rõ rệt so với hệ thống hiện tại trên hầu hết các chỉ tiêu kỹ thuật quan trọng.

Thời gian hoàn tất chu kỳ sao lưu được rút ngắn từ 6–6,5 giờ xuống còn khoảng 2–3 giờ, cho phép tăng tần suất sao lưu và trực tiếp làm giảm RPO từ 24 giờ xuống dưới 3 giờ. Đồng thời, RTO phục hồi toàn hệ thống cũng được rút ngắn hơn 50%, giúp giảm đáng kể thời gian gián đoạn dịch vụ khi xảy ra sự cố.

Về kiến trúc, việc chuyển từ mô hình sao lưu tập trung một site sang mô hình hai site độc lập đã loại bỏ điểm hỏng đơn và nâng cao đáng kể khả năng chịu lỗi của hệ thống. Mặc dù mỗi site vẫn bảo vệ số lượng máy trạm tương đương hệ thống cũ (khoảng 20 máy), việc bổ sung site thứ hai giúp tăng mức độ dự phòng mà không làm gia tăng đáng kể tải xử lý tại từng site.

Ngoài ra, nhờ áp dụng cơ chế sao lưu gia tăng và snapshot, hệ thống đề xuất giúp giảm 60–70% tốc độ gia tăng dung lượng lưu trữ so với phương pháp sao lưu toàn bộ. Điều này góp phần giảm chi phí đầu tư mở rộng hạ tầng lưu trữ và đơn giản hóa công tác vận hành trong dài hạn.

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

4.3.1. So sánh với các giải pháp sao lưu thương mại chuyên dụng

Các giải pháp như Veeam Backup & Replication, Acronis Cyber Protect thường cung cấp bộ tính năng rất phong phú: hỗ trợ đa hypervisor, tích hợp sâu với các ứng dụng cơ sở dữ liệu, khả năng orchestration kịch bản DR phức tạp, báo cáo chi tiết và tự động hóa cao. Tuy nhiên, những ưu điểm này đi kèm chi phí bản quyền đáng kể và yêu cầu hạ tầng máy chủ riêng để vận hành phần mềm.

Ngược lại, ABB là thành phần tích hợp sẵn trên Synology NAS, không thu phí bản quyền theo số lượng agent. Điều này đặc biệt phù hợp với doanh nghiệp vừa và nhỏ khi chi phí đầu tư là yếu tố nhạy cảm. Mặc dù không có

đầy đủ các tính năng nâng cao như một số giải pháp thương mại cấp doanh nghiệp, ABB vẫn đáp ứng tốt nhu cầu sao lưu máy trạm, máy chủ ảo ESXi và máy chủ vật lý, đồng thời cung cấp cơ chế Changed Block Tracking giúp giảm thời gian backup và dung lượng lưu trữ.

Trong các thử nghiệm mô phỏng nội bộ, về thời gian sao lưu, hệ thống phân tán giảm đáng kể thời gian thực hiện sao lưu nhờ cơ chế incremental forever và snapshot. Trong khi hệ thống truyền thống cần 8-12 giờ để hoàn thành full backup cho 1TB dữ liệu, hệ thống phân tán chỉ cần 2-4 giờ cho lần backup đầu tiên và 15-30 phút cho các lần incremental tiếp theo.

Về thời gian phục hồi, hệ thống phân tán cho phép khôi phục máy ảo trong vòng 5-15 phút thông qua instant recovery, trong khi hệ thống truyền thống mất 2-6 giờ cho cùng tác vụ. Khả năng granular recovery của hệ thống phân tán giúp tiết kiệm 80-90% thời gian khi chỉ cần khôi phục một số file quan trọng.

Theo đó, thời gian thực hiện bản sao lưu gia tăng (incremental) của ABB đối với máy trạm Windows nằm ở mức chấp nhận được, với tốc độ sao lưu ổn định khi dữ liệu thay đổi vừa phải. Thời gian khôi phục file lẻ và thư mục nhỏ gần như tức thời đối với các tập dữ liệu dưới vài gigabyte, hoàn toàn đáp ứng yêu cầu vận hành của một doanh nghiệp vừa và nhỏ. Các con số đo đạc mang tính tham khảo nhưng cho thấy giải pháp đạt hiệu năng đủ tốt trong phạm vi kịch bản thí điểm.

Về chống chịu sự cố, cách tiếp cận kết hợp cluster HA và Snapshot Replication hai chiều mang lại lợi thế rõ rệt, khác với các giải pháp chỉ lưu dữ liệu backup trên một máy chủ ứng dụng, hệ thống đề xuất duy trì bản sao dữ liệu ở hai site độc lập, cho phép phục hồi kể cả khi một site mất hoàn toàn. Ngoài ra, snapshot Btrfs tồn tại tách biệt với không gian dữ liệu truy cập qua SMB/NFS nên có khả năng chống ransomware tự nhiên mà không cần cấu hình thêm cơ chế “immutable backup” phức tạp.

Nhược điểm chính của giải pháp so với các bộ sản phẩm thương mại là phạm vi tích hợp ứng dụng và công cụ quản trị cao cấp còn hạn chế. Hệ thống hiện mới tập trung vào bảo vệ máy trạm, máy chủ file và máy ảo ESXi; các tính năng như backup nhận thức ứng dụng (application-aware) cho mọi hệ quản trị cơ sở dữ liệu, tự động hóa kịch bản DR phức tạp hay báo cáo tuân thủ

chi tiết chưa được triển khai đầy đủ. Tuy nhiên, xét trong bối cảnh doanh nghiệp vừa và nhỏ, mức tính năng hiện tại vẫn là cân bằng hợp lý giữa chi phí và hiệu quả.

4.3.2. So sánh với các giải pháp sao lưu cục bộ truyền thống

Ở nhiều doanh nghiệp nhỏ, sao lưu dữ liệu vẫn được thực hiện bằng các phương pháp thủ công hoặc bán tự động, chẳng hạn như: sao chép file định kỳ sang ổ cứng ngoài, sử dụng chức năng Backup mặc định của hệ điều hành hoặc các phần mềm sao lưu đơn lẻ không có cơ chế quản lý tập trung. Những cách làm này có ưu điểm là chi phí ban đầu thấp, dễ triển khai. Tuy nhiên, chúng bộc lộ nhiều hạn chế:

- Thiếu quản lý tập trung, khó theo dõi tình trạng backup toàn hệ thống.
- Không có cơ chế snapshot block-level, dẫn đến thời gian sao lưu dài và khó đáp ứng RPO/RTO chặt chẽ.
- Không hỗ trợ dự phòng địa lý; dữ liệu sao lưu thường chỉ nằm trên cùng site.
- Gần như không có khả năng chống ransomware, do dữ liệu backup thường hiển thị trực tiếp trong hệ thống file.

Giải pháp được đề xuất trong luận văn khắc phục phần lớn các hạn chế trên. ABB cung cấp giao diện quản lý tập trung cho toàn bộ tác vụ sao lưu, cho phép theo dõi trạng thái, thống kê dung lượng, lịch chạy backup và cảnh báo lỗi. Cơ chế Changed Block Tracking giúp rút ngắn đáng kể thời gian backup sau lần đầu, trong khi Snapshot Replication hai chiều tạo ra ít nhất hai bản sao dữ liệu tại hai địa điểm độc lập. Nhờ snapshot Btrfs ẩn khỏi không gian SMB, dữ liệu backup cũng ít bị tổn thương hơn trước tấn công ransomware so với cách sao lưu “copy file” truyền thống.

4.3.3. So sánh với các giải pháp sao lưu thuần đám mây

Các dịch vụ backup thuần đám mây (cloud backup) cung cấp ưu điểm rõ rệt về khả năng mở rộng gần như vô hạn, không yêu cầu đầu tư phần cứng tại chỗ và rất phù hợp với môi trường làm việc phân tán. Tuy nhiên, mô hình này phụ thuộc mạnh vào đường truyền Internet và mô hình chi phí dạng thuê

bao theo dung lượng, đối với doanh nghiệp vừa và nhỏ có lượng dữ liệu lớn thì chi phí dài hạn không hẳn thấp.

Giải pháp xây dựng trên Synology NAS có thể xem là mô hình "lai", toàn bộ dữ liệu sao lưu nằm trên hạ tầng lưu trữ riêng, được kiểm soát bởi doanh nghiệp; đồng thời vẫn đạt được lợi ích dự phòng địa lý nhờ triển khai hai site và kết nối WAN/VPN. Trong bối cảnh doanh nghiệp có sẵn hai văn phòng ở hai thành phố, mô hình này tận dụng tối đa hạ tầng hiện có, giảm phụ thuộc vào dịch vụ đám mây bên ngoài, đồng thời vẫn đạt được mục tiêu an toàn dữ liệu.

Hạn chế của mô hình so với cloud thuần túy là khả năng mở rộng vẫn bị giới hạn bởi tài nguyên phần cứng NAS và băng thông WAN giữa hai site. Tuy vậy, đối với quy mô doanh nghiệp mục tiêu của luận văn, dung lượng và băng thông đạt được trong thử nghiệm cho thấy hệ thống vẫn đáp ứng tốt nhu cầu tăng trưởng dữ liệu trong trung hạn.

4.4. Nhận xét và đề xuất cải tiến

4.4.1. Nhận xét tổng quan

Các kết quả triển khai và thử nghiệm cho thấy hệ thống sao lưu phân tán được thiết kế trong luận văn đáp ứng tương đối đầy đủ các yêu cầu ban đầu đặt ra:

- **Về độ tin cậy**: cụm HA tại mỗi site và cơ chế Snapshot Replication hai chiều đảm bảo dữ liệu luôn tồn tại ở hai vị trí vật lý khác nhau, cho phép khôi phục ngay cả khi một site gặp sự cố nghiêm trọng.
- **Về hiệu năng**: với cơ chế Changed Block Tracking, thời gian thực hiện các phiên sao lưu gia tăng ở mức chấp nhận được, không gây ảnh hưởng đáng kể đến hoạt động của máy trạm và máy chủ. Thời gian khôi phục file lẻ và thư mục nhỏ đủ nhanh để đáp ứng nhu cầu vận hành hằng ngày.
- **Về an toàn trước ransomware**: thử nghiệm mô phỏng tấn công cho thấy snapshot Btrfs không bị mã hóa cùng dữ liệu, cho phép khôi phục nhanh toàn bộ thư mục về trạng thái trước khi bị tấn công.

- **Về chi phí**: việc tận dụng ABB, một thành phần không mất phí bản quyền đi kèm NAS kết hợp cùng hạ tầng ảo hóa sẵn có giúp tối ưu chi phí đầu tư cho doanh nghiệp vừa và nhỏ.

Từ góc độ tổng thể, có thể nhận định rằng giải pháp đã đạt được sự cân bằng hợp lý giữa chi phí, mức độ phức tạp và hiệu quả bảo vệ dữ liệu trong phạm vi đối tượng doanh nghiệp mục tiêu.

4.4.2. Các hạn chế còn tồn tại

Bên cạnh những ưu điểm nêu trên, hệ thống vẫn còn một số điểm hạn chế:

- **Phụ thuộc nền tảng Synology**: toàn bộ kiến trúc xoay quanh NAS Synology và Btrfs, do đó khả năng mở rộng sang các nền tảng lưu trữ khác còn hạn chế.
- **Tính năng nâng cao cho doanh nghiệp lớn chưa đầy đủ**: các chức năng như tự động hóa kịch bản DR phức tạp, orchestration đa site, báo cáo tuân thủ chi tiết hoặc tích hợp sâu với tất cả hệ quản trị cơ sở dữ liệu chưa được khai thác.
- **Phụ thuộc băng thông WAN giữa hai site**: trong các kịch bản dữ liệu thay đổi lớn hoặc băng thông hạn chế, thời gian hoàn thành replicate có thể tăng lên, làm nới rộng RPO giữa hai site.
- **Thiếu lớp sao lưu "offline"**: hệ thống hiện tập trung vào bản sao trực tuyến tại hai site; chưa triển khai thêm một bản sao tách biệt hoàn toàn (ví dụ lưu trữ băng từ hoặc object storage với cơ chế "immutable") để đối phó kịch bản tấn công sâu vào hạ tầng NAS.

4.4.3. Đề xuất cải tiến

Để nâng cao hơn nữa mức độ an toàn và khả năng mở rộng của hệ thống, một số hướng phát triển có thể xem xét trong tương lai bao gồm:

- **Bổ sung lớp sao lưu thứ ba theo nguyên tắc 3–2–1**, ví dụ replicate thêm một bản sao snapshot lên dịch vụ lưu trữ đám mây hỗ trợ cơ chế khóa đối tượng (object lock) nhằm phòng ngừa kịch bản cả hai site cùng bị tấn công.

- **Tăng cường giám sát và cảnh báo**, tích hợp log của NAS và ABB vào hệ thống giám sát tập trung hoặc nền tảng SIEM, giúp phát hiện sớm hành vi bất thường liên quan đến truy cập dữ liệu hoặc hoạt động sao lưu.
- **Tối ưu băng thông WAN**, áp dụng nén, lập lịch replicate thông minh theo khung giờ ít tải, đồng thời xem xét nâng cấp đường truyền trong trường hợp dung lượng dữ liệu tăng nhanh.
- **Mở rộng phạm vi bảo vệ ứng dụng**, nghiên cứu tích hợp thêm các cơ chế backup nhận thức ứng dụng cho một số hệ quản trị cơ sở dữ liệu hoặc hệ thống nghiệp vụ quan trọng.
- **Hoàn thiện quy trình DR và diễn tập định kỳ**, xây dựng tài liệu hướng dẫn khôi phục chi tiết và tổ chức các bài diễn tập để kiểm chứng khả năng phục hồi trong các kịch bản sự cố khác nhau.

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.

TÀI LIỆU THAM KHẢO

Tài liệu tiếng Việt

[1] Bộ Thông tin và Truyền thông (2022), *Báo cáo hiện trạng an toàn thông tin Việt Nam*, NXB Thông tin & Truyền thông, Hà Nội.

[2] Cục An toàn thông tin (2023), *Hướng dẫn phòng chống mã độc mã hóa dữ liệu (Ransomware) trong cơ quan, doanh nghiệp*, 85 trang.

[3] Nguyễn Anh Tuấn (2021), *An toàn dữ liệu và sao lưu hệ thống trong doanh nghiệp*, NXB Khoa học và Kỹ thuật, Hà Nội.

[4] Tổng cục Thống kê (2020), *Niên giám thống kê Việt Nam năm 2019*, 450 trang.

[5] Trung tâm Giám sát an toàn không gian mạng quốc gia (NCSC) (2023), *Báo cáo tấn công mạng và xu hướng mã độc tại Việt Nam*, Hà Nội.

Tài liệu tiếng Anh

[6] Bar, R., et al. (2019), "Backup strategies for distributed storage systems: a comparative analysis", Journal of Information Security, 14(3), pp. 110–129.

[7] Corbet, J. (2012), "Btrfs: The Linux B-tree Filesystem", Linux Journal, Issue 219, pp. 35–45.

[8] Synology Inc. (2023), Synology High Availability White Paper, Synology Documentation Center, Taiwan.

[9] Synology Inc. (2023), Active Backup for Business – Technical Overview, Synology Documentation Center, Taiwan.

[10] Patterson, D., Gibson, G. (1988), "A Case for Redundant Arrays of Inexpensive Disks (RAID)", ACM SIGMOD, pp. 109–116.

[11] Scaife, N., Carter, H., Traynor, P., Butler, K. (2016), "CryptoLock (and Drop It): Stopping Ransomware Attacks", USENIX Security Symposium, pp. 1–14.

[12] Hogg, J. (2020, May), Ransomware Defense Strategies in Enterprise Environments, Microsoft Security Blog. Available: https://www.microsoft.com/security/blog — truy cập ngày 05/09/2025.

[13] EMC Corporation (2016), Data Protection and Disaster Recovery Architecture Guide, Dell EMC Technical Report, USA.

Tài liệu Internet

[14] J. Jones. (1991, May 10), Networks (2nd ed.) [Online]. Available: http://www.atm.com - truy cập ngày 12/10/2025.

[15] R. J. Vidmar. (1992, Aug.), "On the use of atmospheric plasmas as electromagnetic reflectors", IEEE Trans. Plasma Sci. [Online]. 21(3), pp. 876–880. Available: http://www.halcyon.com/pub/journals/21ps03-vidmar - truy cập ngày 15/10/2025.

[16] https://www.synology.com, truy cập ngày 20/11/2025.

[17] Synology Inc. (2024), "Active Backup for Business - Technical Overview" [Online]. Available: https://www.synology.com/en-global/dsm/feature/active_backup_business - truy cập ngày 25/01/2026.

[18] NIST. (2018), "Guide for Cybersecurity Event Recovery (SP 800-184)" [Online]. Available: https://csrc.nist.gov/publications/detail/sp/800-184/final - truy cập ngày 26/01/2026.

[19] Veeam Software. (2023), "Understanding RPO and RTO in Data Protection" [Online]. Available: https://www.veeam.com/solutions/data-security/ransomware-recovery.html?ad=menu-solutionsml - truy cập ngày 30/01/2026.

BẢN CAM ĐOAN

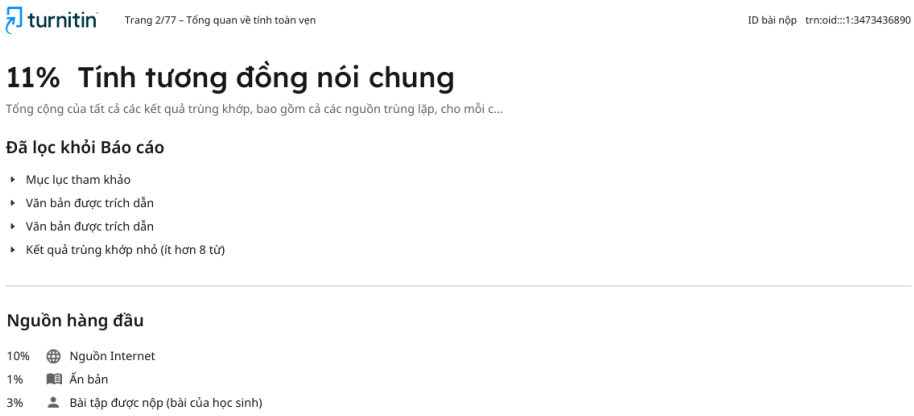
Tôi cam đoan đã thực hiện việc kiểm tra mức độ tương đồng nội dung đề án qua phần mềm Turnitin một cách trung thực và đạt kết quả mức độ tương đồng 11% toàn bộ nội dung đề án. Bản đề án kiểm tra qua phần mềm là bản cứng đề án đã nộp để bảo vệ trước hội đồng. Nếu sai tôi xin chịu các hình thức kỷ luật theo quy định hiện hành của Học viện.

Hà Nội, ngày 05 tháng 02 năm 2026

HỌC VIÊN CAO HỌC

(Ký và ghi rõ họ tên**)**

Bùi Minh Tuấn



turnitin Trang 2/77 – Tổng quan về tính toàn vẹn ID bài nộp trn:oid:::1:3473436890

11% Tính tương đồng nói chung

Tổng cộng của tất cả các kết quả trùng khớp, bao gồm cả các nguồn trùng lặp, cho mỗi c...

Đã lọc khỏi Báo cáo

- Mục lục tham khảo
- Văn bản được trích dẫn
- Văn bản được trích dẫn
- Kết quả trùng khớp nhỏ (ít hơn 8 từ)

Nguồn hàng đầu

10% Nguồn Internet
1% Ấn bản
3% Bài tập được nộp (bài của học sinh)

Học viên	Người hướng dẫn khoa học
Bùi Minh Tuấn	PGS.TS. Đỗ Trung Tuấn