**KHOA CÔNG NGHỆ THÔNG TIN CÁC CÔNG NGHỆ ẢO HÓA**

**LAB 02: STORAGE**

**Câu 01:** DAS (Direct Attached Storage), NAS (Network Attached Storage), SAN (Storage Area Network) là gì và có công dụng gì trong một Data Center?

DAS là hệ thống lưu trữ được gắn trực tiếp vào máy chủ hoặc máy tính thông qua các kết nối như SATA, SAS, SCSI. Không có mạng lưới trung gian, dữ liệu được truyền tải trực tiếp giữa lưu trữ và máy chủ. Công dụng: có hiệu suất cao vì không có mạng trung gian, việc truyền dữ liệu ít bị trễ, cấu hình đơn giản và không cần đầu tư vào thiết bị mạng phức tạp. Phù hợp cho các ứng dụng yêu cầu hiệu suất cao nhưng không cần chia sẻ dữ liệu qua mạng, như cơ sở dữ liệu, hệ thống xử lý giao dịch trực tuyến (OLTP)

NAS là hệ thống lưu trữ kết nối với mạng thông qua Ethernet và cung cấp các dịch vụ lưu trữ file qua mạng LAN. NAS hoạt động như thiết bị độc lập, có hệ điều hành riêng và thường hỗ trợ các giao thức như NFS, SMB/CIFS. Công dụng: người dùng trong mạng có thể truy cập và chua sẻ dữ liệu dễ dàng, quản lý lưu trữ tập trung và có thể mở rộng dung lượng bằng cách thêm thiết bị NAS mới, phù hợp cho việc lưu trữ, chia sẻ file, sao lưu và khôi phục dữ liệu.

SAN là một mạng tốc độ cao dành riêng cho việc kết nối các thiết bị lưu trữ với máy chủ. SAN sử dụng các giao thức như Fibre Channel hoặc iSCSI để truyền dữ liệu và thường tách biệt hoàn toàn với mạng LAN. Công dụng: Có hiệu suất cao và khả năng mở rộng tốt, các tính năng miroring, replication, snapshot giúp quản lý dữ liệu hiệu quả hơn.

**Câu 02:** Sự khác biệt giữa DAS, NAS và SAN là gì?

|  |  |  |  |
| --- | --- | --- | --- |
|  | DAS | NAS | SAN |
| Cấu trúc và kết nối | Lưu trữ gắn trực tiếp và máy chủ hoặc máy tính. Sử dụng các kết nối trực tiếp như SATA, SAS, SCSI. | Thiết bị lưu trữ kết nối vào mạng LAN. Thông qua kết nối mạng Ethernet. | Mạng lưu trữ chuyên dùng kết nối nhiều thiết bị lưu trữ và máy chủ. Sử dụng giao thức Fibre Channel hoặc iSCSI |
| Hiệu suất và độ trễ | Hiệu suất cao do không có mạng trung gian. Độ trễ thấp vì được truyền trực tiếp | Hiệu suất phụ thuộc vào băng thông mạng LAN. Độ trễ cao hơn DAS do phụ thuộc vào các yếu tố như mạng LAN hay các lưu lượng mạng | Hiệu suất cao, sử dụng mạng chuyên dụng tốc độ cao. Độ trễ thấp tối ưu cho ứng dụng đòi hỏi độ trễ thấp và thông lượng cao |
| Khả năng mở rộng và quản lý | Khả năng mở rộng bị hạn chế, phụ thuộc vào kết nối máy chủ. Quản lý dễ dàng chủ yếu thông qua máy chủ mà nó kết nối | Khả năng mở rộng tốt hơn DAS, có thể thêm nhiều thiết bị NAS vào mạng. Quản lý tập trung dễ dàng quản lý qua giao diện mạng | Khả năng mở rộng rất tố, có thể dễ dàng mở rộng bằng thêm thiết bị mạng SAN. Quản lý thì phức tạp hơn, yêu cầu phần mềm quản lý lưu trữ chuyên dụng và kiến thực chuyên môn cao. |
| Ứng dụng | Phù hợp cho các ứng dụng yêu cầu hiệu suất cao nhưng không cần chua sẻ dữ liệu qua mạng, như cơ sở dữ liệu các nhân hoặc hệ thống giao dịch trực tuyến. | Thích hợp cho môi trường làm việc nhóm, chia sẻ file, sao lưu dữ liệu và các ứng dụng không yêu cầu độ trễ thấp | Thích hợp cho dianh nghiệp lớn, hệ thống cơ sở dữ liệu lớn, môi trường ảo hoá và các ứng dụng đòi hỏi hiệu suất và độ tin cậy cao. |

**Câu 03:** Khái niệm về ảo hóa lưu trữ (storage virtualization) liên quan đến DAS, NAS và SAN như thế nào?

Ảo hoá lưu trữ (storage virtualization) là quá trình tách biệt các dịch vụ lưu trữ vật lý khỏi các dịch vụ lưu trữ logic. Nó cung cập một lớp trừu tượng giữa phần cứng lưu trữ và hệ thống máy chủ, cho phép người quản trị quản lý không gian lưu trữ như một tài nguyên hợp nhất bất kể thiết bị vật lý nào được sử dụng.

Mối quan hệ giữa ảo hoá lưu trữ đối với DAS, NAS, SAN

DAS: Ảo hoá lưu trữ có thể biến nhiều hệ thống DAS thành một nhóm lưu trữ hợp nhất. Điều này cho phép quản lý tập trung và sử dungh tài nguyên hiệu quả hơn, mặc dù việc thự hiện có thể phức tạp hơn do sự phân tán của các thiết bị DAS.

NAS: Ảo hoá lưu trữ có thể hợp nhất nhiều thiết bị NAS thành một không gian lưu trữ duy nhất, cho phép các quản lý tập trung và tối ưu hoá việc sử dụng không gian lưu trữ. Các hệ thống NAS ảo hoá cũng có thể cung cấp các dịch vụ nâng cao như chia sẻ file động và di chuyển dữ liệu liền mạch.

SAN: Ảo hoá lưu trữ trong SAN thường dễ dàng hơn và mang lại hiệu quả cao. Nó cho phép quản lý linh hoạt hơn, tạo ra các vùng lưu trữ hợp nhất từ nhiều thiết bị lưu trữ vật lý khác nhau, cung cấp các dịch vụ lưu trữ động, tối ưu hoá hiệu suất và đảm bảo tính sẵn sàng cao.

**Câu 04:** Các giao thức thông dụng như iSCSI và NFS được sử dụng trong NAS. Hãy giải thích cách hoạt động của chúng và khi nào nên sử dụng mỗi giao thức?

Cách hoạt động của iSCSI: iSCSL là một giao thức lưu trữ mạng dựa trên IP, cho phép truyền tải các lệnh SCSI thông qua TCP/IP. iSCSL tạo ra một kết nối giữa máy chủ và thiết bị lưu trữ. Máy chủ gửi các lệnh SCSI như đọc ghi dữ liệu tới thiết bị lưu trữ từ xa giống như đang tương tác với thiết bị lưu trữ cục bộ.

Nên sử dụng iSCSL trong trường hợp khi mà cần một giao thức lưu trữ mạng có hiệu suất cao và có thể truyền tải khối lượng dữ liệu lớn với độ trễ thấp. Trong các môi trường đã có hạ tầng mạng IP. iSCSL hỗ trọ rộng rãi bởi các nền tảng áo hoá như Vmware, Hyper-V

Cách hoạt động của NFS: NFS là một giao thức chia sẻ file qua mạng, cho phép các hệ thống tệp trên máy chủ được truy cập từ xa bởi các máy tính khách như chúng là hệ thống tệp cục bộ. NFS hoạt động dựa trên mô hình server-client, nơi mà máy chủ chia sẻ hệ thống tệp qua mạng và máy khách được gắn kết hệ thống tệp để sử dụng.

Nên sử dụng NFS trong khi cần chia sẻ file giữa nhiều máy khách trong mạng. NFS hỗ trợ cả UNIX, Linux và các hệ điều hành khác.

**Câu 05:** Các yếu tố cần xem xét khi lựa chọn hệ thống lưu trữ dữ liệu?

Khi lựa chọn một hệ thống lưu trữ dữ liệu cần lưu ý về hiệu suất, dung lượng lưu trữ, quản lý và vận hành, chi phí, khả năng tương thích, bảo mật, tính linh hoạt và hiệu quả

Yêu cầu về hiệu suất: hệ thống lưu trữ cần phải cả khả năng read và write phù hợp với ứng dụng, độ trễ phải thấp.

Dung lượng lưu trữ: Phải xem được lưu lượng hiện tại và phải có khả năng mở rộng thêm tránh tình trạng quá tải.

Quản lý và vận hành: Hệ thống phải có công cụ quản lý hiệu quả, giao diện người dùng thân thiện và khả năng tự động hoá các tác vụ quản lý.

Chi phí: Bao gồm chi phí phần cứng và phần mềm, bảo trì

Khả năng tương thích: Tương thích với các phần mềm, phần cứng và mạng đang sử dụng, đảm bảo hệ thống hoạt động với các hệ điều hành và môi trường ảo hoá khác nhau

Bảo mật: đảm bảo hệ thống lưu trữ tuân thủ các quy định pháp lý và tiêu chuẩn bảo mật

Tính linh hoạt và hiệu quả: Vì sử dụng các giao thức lưu trữ như iSCSI, NFS, SMB/CIFS hỗ trợ ảo hoá lưu trữ để tăng hiệu suất sử dụng và quản lý tài nguyên. Tiết kiệm năng lượng và có độ bền vững cao.

**Câu 06:** So sánh các dạng lưu trữ dữ liệu Block Storage, File Storage, Object Storage.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Block Storage | File Storage | Object Storage |
| Hiệu suất | Hiệu suất cao, độ trễ thấp | Hiệu suất trung bình | Hiệu suất thấp hơn, độ trễ cao |
| Quản lý | Phức tạp hơn | Dễ quản lý hơn | Linh hoạt nhưng cần kỹ năng để quản lý |
| Chia sẻ dữ liệu | Hạn chế | Tốt, dễ chia sẻ | Rất tốt |
| Chi phí | Cao hơn | Thấp hơn so với Block Storage | Thấp |
| Ứng dụng | Cơ sở dữ liệu, hệ thống giao dịch, ảo hoá | Lưu trữ và chia sẻ file môi trường làm việc nhóm | Lưu trữ đám mấy, big data |

|  |  |  |
| --- | --- | --- |
|  | Nguyễn Văn Huy Dũng | 1 |

**Câu 07:** RAID là gì? Tìm hiểu các dạng RAID 0, 1, 5, 6, 10…

Raid là công cụ lưu trữ dữ liệu cho phép kết hợp nhiều ổ đĩa cứng thành một hệ thống lưu trữ duy nhất để cải thiện hiệu suất. Raid có nhiều cấp độ khác nhau, mỗi cấp độ cung cấp các đặc tính riêng về hiệu suất, bảo mật dữ liệu.

Raid 0: hay còn gọi là Striping. Dữ liệu được chia thành các khối và phân phối qua tất cả các ổ đĩa trong mảng Raid. Striping có hiệu suất đọc và ghi nhiều đĩa một cách đồng thời. Không có khả năng bảo vệ dữ liệu, nếu một ổ đĩa bị lỗi thì toàn bộ dữ liệu sẽ bị mất

Raid 1: hay còn gọi là Mirroring. Dữ liệu được sao chép đồng nhất giữa hai hoặc nhiều ổ đĩa. Dữ liệu đọc tăng do nó có thể đọc từ bất kỳ ổ đĩa nào, nhưng hiệu suất ghi thì không được cao. Cung cấp bảo vệ dữ liệu cao, không như Raid 0, khi ổ đĩa bị lỗi, dữ liệu vẫn có sẵn trên các ổ đĩa khác.

Raid 5: Dữ liệu và thông tin parity được phân phối đều qua tất cả các ổ đĩa trong mảng. Hiệu suất đọc tốt nhưng hiệu suất ghi bị giảm do việc tính toán và ghi parity

Raid 6: Tương tự như Raid 5, nhưng có thêm một khối parity thứ hai. Hiệu suất đọc tốt, hiệu suất ghi giảm hơn Raid 5 do phải tính toán và ghi thêm 2 khối parity.

Raid 10 hay Raid 1 + 0 (là sự kết hợp giữa Raid 1 và Raid 0. Kết hợp giữa morriroring và striping. Dữ liệu được phân đoạn (striping) và sau đó mỗi đoạn được sao chép (mirroring). Có thể chịu được lỗi của nhiều ổ đĩa miễn là không cung một cặp mirroring.

**Câu 08:** Cloud storage là gì? Kể tên một số Cloud storage mà anh/chị biết và đã sử dụng

Cloud storage là dịch vụ lưu trữ dữ liệu trên các máy chủ từ xa thay vì trên ổ đĩa cục bộ. Dữ liệu được lưu trữ, quản lý và sao lưu bởi các nhà cung cấp đám mây và có thể được truy cập qua internet từ bất kỳ đâu.

Một số dịch vụ storage: Google Drive, Dropbox, Microsoft OneDrive, Amazon S3