

CHƯƠNG 7

CSDL PHÂN TÁN

NỘI DUNG

- Sự Phát Triển Của HQTCSDL Phân Tán
- Các Thành Phần Của Hệ Thống CSDL Phân Tán
- Quản Lý Dữ Liệu Phân Tán
- Tính Trong Suốt Của CSDL Phân Tán

- **Đường Link you tube**

- <https://www.youtube.com/watch?v=IMoBFJnkDm4&list=PL67CJL04EcjN5PQrippgbVWz9L06a3Atz&index=18>

HỆ QUẢN TRỊ CSDL PHÂN TÁN

1 - Sự Phát Triển Của HQTCSDL Phân Tán

* Từ những năm 1970, DBMS giải quyết các yêu cầu về thông tin có cấu trúc và ngôn ngữ lập trình thế hệ thứ 3 để đáp ứng các yêu cầu của người sử dụng.

* Những năm 1980 sự phát triển của Xã hội và Kỹ thuật ảnh hưởng đến sự phát triển và thiết kế CSDL như:

- Các hoạt động kinh doanh trở nên phân tán nhiều hơn về mặt địa lý.
- Cạnh tranh gia tăng ở mức toàn cục.
- Yêu cầu của khách hàng và thị trường cần 1 kiểu quản lý phân tán.
- Sự phát triển nhanh chóng về mặt kỹ thuật là giảm chi phí các thiết bị với chất lượng cao.

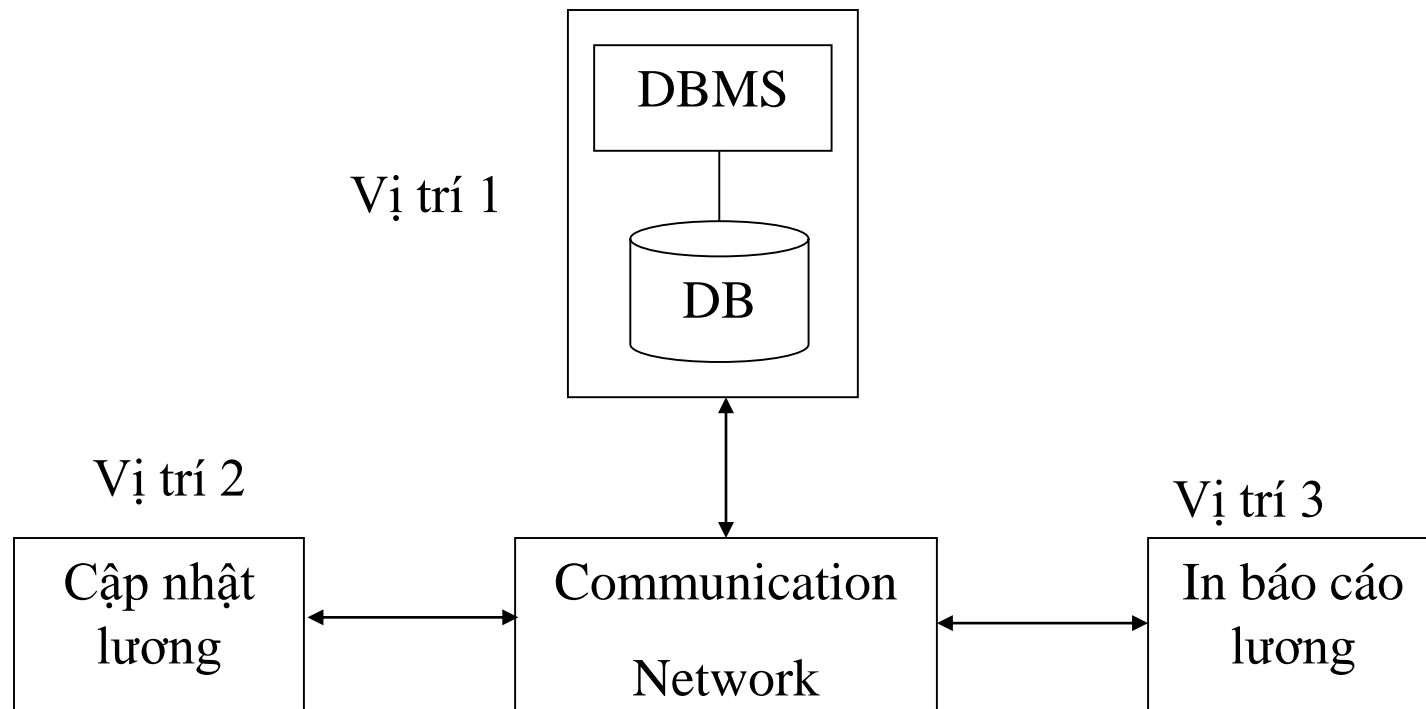
→ Mạng LAN là 1 giải pháp cơ bản.

Sự phân tán của cấu trúc quản lý dựa trên sự phân tán của đơn vị kinh doanh, hình thành sự phân tán về truy xuất và lưu trữ dữ liệu.

2 - Các Khái Niệm

▪ Xử lý phân tán

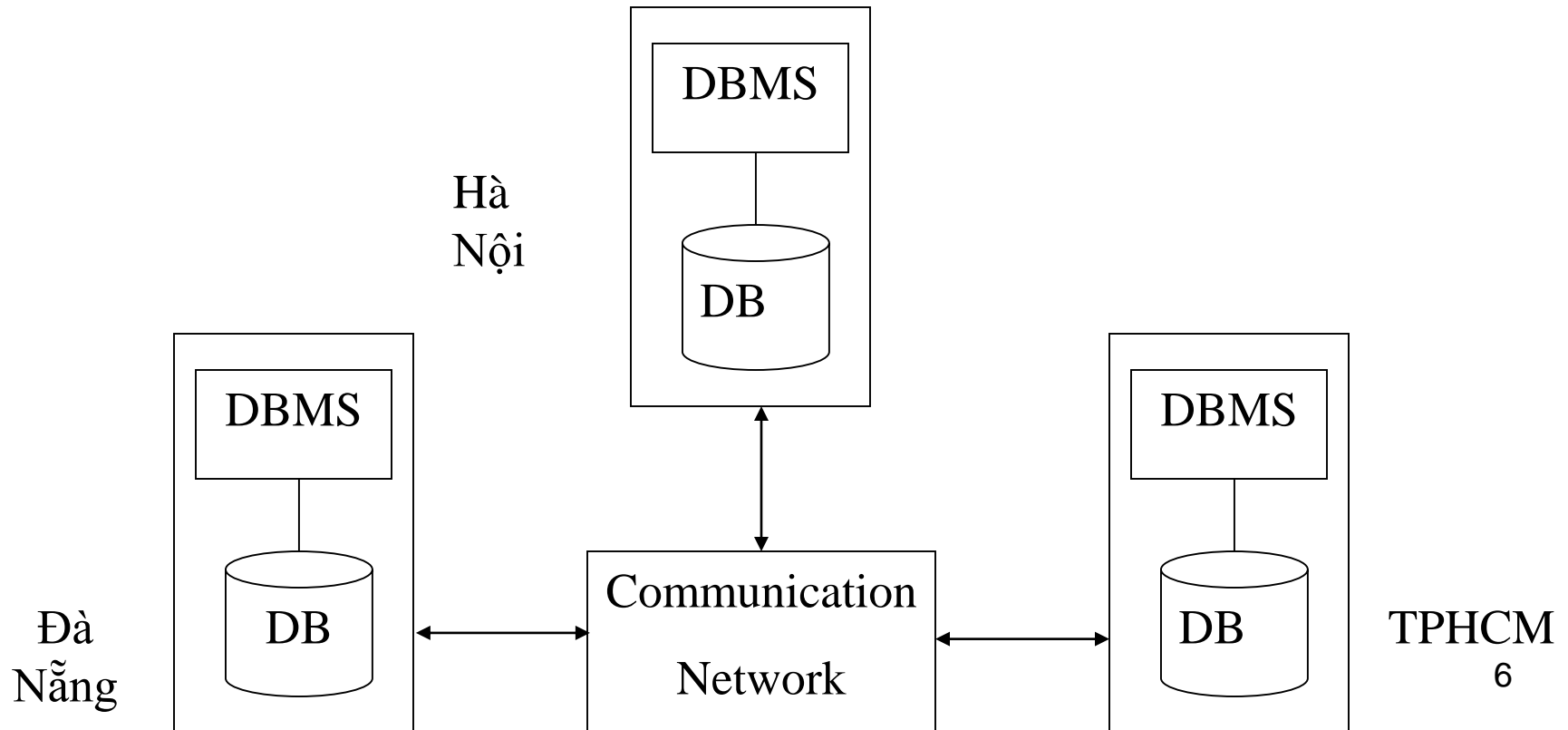
Xử lý phân tán chia sẻ xử lý trên CSDL cho 2 hay nhiều vị trí độc lập vật lý được liên kết với nhau qua hệ thống mạng.



2 - Các Khái Niệm

▪ Dữ liệu phân tán

Dữ liệu phân tán lưu trữ CSDL có quan hệ về mặt logic trên 2 hay nhiều vị trí khác nhau thông qua hệ thống mạng. Các CSDL được chia nhỏ ở nhiều vị trí khác nhau được gọi là database - fragment.



2 - Các Khái Niệm

Dữ liệu phân tán

*** Nhận xét về xử lý phân tán và CSDL phân tán :**

Xử lý phân tán không đòi hỏi có CSDL phân tán nhưng 1 CSDL phân tán đòi hỏi có xử lý phân tán.

Cả hai xử lý phân tán và CSDL phân tán đòi hỏi 1 mạng máy tính để kết nối các thành phần với nhau.

2 - Các Khái Niệm

▪ CSDL Phân Tán

CSDL phân tán là một tập dữ liệu được phân phối trong các máy tính khác nhau trong 1 mạng máy tính được gọi là các nút (vị trí) nhưng vẫn đảm bảo tính nhất quán trên toàn bộ CSDL và được khai thác giống như trên CSDL tập trung.

▪ Hệ Quản Trị CSDL Phân Tán

Để quản lý CSDL phân tán người ta dùng hệ quản trị CSDL phân tán (Distributed Database Management System - DDBMS). DDBMS cũng chính là DBMS tập trung nhưng được bổ sung 1 số thành phần để hỗ trợ cho việc liên lạc và phối hợp hoạt động giữa các DBMS được cài đặt tại những nút khác nhau trong mạng máy tính.

3 - Những Thuận Lợi Và Hạn Chế Của Môi Trường CSDL Phân Tán

▪ Thuận lợi

1. Dữ liệu được đặt gần nơi có yêu cầu lớn nhất
2. Truy xuất dữ liệu nhanh
3. Xử lý dữ liệu nhanh
4. Dễ dàng tăng trưởng
5. Giảm chi phí hoạt động
6. Ít chi phí hơn khi có sự cố
7. Xử lý độc lập

3 - Những Thuận Lợi Và Hạn Chế Của Môi Trường CSDL Phân Tán

▪ Hạn chế

1. Phức tạp trong việc thiết kế, quản lý và điều khiển
2. Các chức năng điều khiển truy xuất đồng thời, an toàn dữ liệu phải thích hợp với môi trường dữ liệu phân tán.
3. Thiếu sự chuẩn hoá, CSDL phân tán phụ thuộc vào hiệu quả của việc truyền thông.

4 - Các Thành Phần Của Hệ Thống CSDL Phân Tán

- ✓ Hệ thống máy tính (Node /site): hình thành hệ thống.
- ✓ Mạng máy tính (Network): gồm cả phần cứng và phần mềm liên quan đến mạng, dùng để liên kết các máy tính
- ✓ Thiết bị truyền thông (Communication media): chuyển dữ liệu, lệnh giữa các máy tính.
- ✓ Bộ xử lý giao dịch (Transaction processor - TP): thành phần phần mềm nằm trên các máy tính có yêu cầu dữ liệu, làm nhiệm vụ nhận và xử lý dữ liệu nhận được từ bộ xử lý của chương trình ứng dụng (Application processor)
- ✓ Bộ xử lý dữ liệu (Data processor - DP): thành phần phần mềm trên các vị trí lưu trữ dữ liệu. DP có thể là hệ quản trị CSDL(DBMS).

5 - Quản Lý Dữ Liệu Phân Tán

Phân mảnh dữ liệu (Fragmentation)

* Lý do phân mảnh:

- Đơn vị truy xuất không phải là toàn bộ quan hệ nhưng chỉ là các tập con của quan hệ
- Cho phép thực hiện nhiều giao dịch đồng thời
- Cho phép thực hiện song song một câu vấn tin bằng cách chia nó thành một tập các câu vấn tin con hoạt tác trên các mảnh, vì thế làm tăng lưu lượng hoạt động của hệ thống.

IV.5. Quản Lý Dữ Liệu Phân Tán

IV.5.1. Phân mảnh dữ liệu (Fragmentation)

IV.5.1.1. Phân mảnh theo chiều ngang (Horizontal fragmentation)

Định nghĩa: Phân mảnh ngang được định nghĩa bằng 1 phép toán chọn trên các quan hệ chủ nhân của 1 lược đồ CSDL. Vì thế nếu cho biết quan hệ R, các mảnh ngang của R_i là:

$$R_i = \delta_{F(i)} (R), 1 \leq i \leq z$$

Trong đó, F_i là công thức chọn được sử dụng để có mảnh R_i

Như vậy, chia 1 bảng dữ liệu thành các bảng con, mỗi bảng được lưu trữ ở 1 vị trí khác nhau, việc phân mảnh theo chiều ngang tương đương 1 câu lệnh select với điều kiện đặt trên 1 số thuộc tính nào đó.

Câu lệnh SQL: Select * From R Where F(i)

IV.5.1.1. Phân mảnh theo chiều ngang (Horizontal fragmentation)

SOKH	ĐỊA CHỈ	TỈNH	SỐ TIỀN NỢ
1	123 Phan Thanh	Đà Nẵng	200000
2	12 CMT8	HCM	300000
3	50 Trần Đình Xu	HCM	400000
4	20 Nguyễn văn Linh	Đà Nẵng	600000

Phân mảnh theo chiều ngang theo tiêu chuẩn $TỈNH = "HCM"$.

Định nghĩa như sau:

$$KHACHHANG1 = \delta_{TINH="HCM"} (KHACHHANG)$$

$$KHACHHANG2 = \delta_{TINH \neq "HCM"} (KHACHHANG)$$

Câu lệnh SQL :

Select * from KHACHHANG where TINH = "HCM"

Select * from KHACHHANG where TINH <> "HCM"

IV.5.1.1. Phân mảnh theo chiều ngang (Horizontal fragmentation)

KHACHHANG1:

SOKH	ĐỊA CHỈ	TỈNH	SỐ TIỀN NỢ
2	12 CMT8	HCM	300000
3	50 Trần Đình Xu	HCM	400000

KHACHHANG2:

SOKH	ĐỊA CHỈ	TỈNH	SỐ TIỀN NỢ
1	123 Phan Thanh	Đà Nẵng	200000
4	20 Nguyễn văn Linh	Đà Nẵng	600000

IV.5. Quản Lý Dữ Liệu Phân Tán

IV.5.1. Phân mảnh dữ liệu (Fragmentation)

IV.5.1.2. Phân mảnh theo chiều dọc (Vertical fragmentation)

Định nghĩa: Một quan hệ R sinh ra các mảnh R1, R2, ... Rr, mỗi mảnh chứa một tập con thuộc tính của R và cả khoá của R.

Công thức: $R_i = \pi_{F(i)}(R)$, $1 \leq i \leq z$

Như vậy, chia 1 bảng dữ liệu thành các bảng con với tập thuộc tính khác nhau, mỗi bảng con được lưu trữ ở 1 nút khác nhau, ngoại trừ thuộc tính khoá, nó xuất hiện ở tất cả các bảng con. Phân mảnh theo chiều ngang tương đương với việc chiếu lên 1 số thuộc tính nào đó.

Cau Lenh: `SELECT <CAC THUOC TINH> FROM BANG`

IV.5.1.2. Phân mảnh theo chiều dọc (Vertical fragmentation)

SOKH	ĐỊA CHỈ	TỈNH	SỐ TIỀN NỢ
1	123 Phan Thanh	Đà Nẵng	200000
2	12 CMT8	HCM	300000
3	50 Trần Đình Xu	HCM	400000
4	20 Nguyễn văn Linh	Đà Nẵng	600000

Ta định nghĩa 2 mảnh như sau:

$KHACHHANG1 = \pi_{SOKH, DCHI, TINH} (KHACHHANG)$

$KHACHHANG2 = \pi_{SOKH, SOTIENNO} (KHACHHANG)$

Phân mảnh quan hệ KHACHHANG thành 2 phân mảnh dọc như sau:

$KHACHHANG1: SELECT SOKH, DCHI, TINH FROM KHACHHANG$

$KHACHHANG2: SELECT SOKH, SOTIENNO FROM KHACHHANG$

IV.5.1.2. Phân mảnh theo chiều dọc (Vertical fragmentation)

KHACHHANG1:

SOKH	ĐỊA CHỈ	TỈNH
1	123 Phan Thanh	Đà Nẵng
2	12 CMT8	HCM
3	50 Trần Đình Xu	HCM
4	20 Nguyễn văn Linh	Đà Nẵng

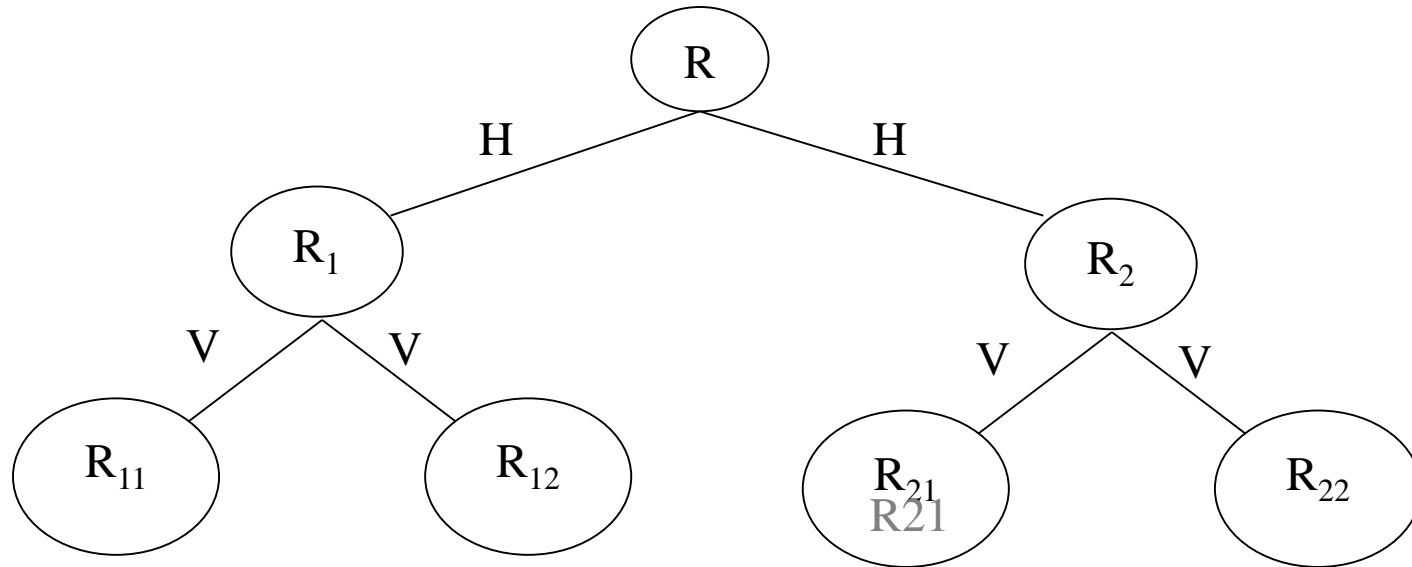
KHACHHANG2:

SOKH	SỐ TIỀN NỢ
1	200000
2	300000
3	400000
4	600000

IV.5. Quản Lý Dữ Liệu Phân Tán

IV.5.1. Phân mảnh dữ liệu (Fragmentation)

IV.5.1.3. Phân mảnh hỗn hợp (Mixed fragmentation)



Nhận xét: Trong phân mạch hỗn hợp:

Đối với phân mảnh ngang, quá trình sẽ dừng khi mỗi mảnh còn 1 bộ

Đối với phân mảnh dọc, điểm ngừng là mỗi thuộc tính mỗi mảnh

$KHACHHANG1 = \pi_{SOKH, DCHI, TINH} (\delta_{TINH="HCM"} (KHACHHANG))$

$KHACHHANG2 = \pi_{SOKH, SOTIENGUI} (\delta_{TINH="HCM"} (KHACHHANG))$

IV.5. Quản Lý Dữ Liệu Phân Tán

IV.5.1. Phân mảnh dữ liệu (Fragmentation)

Nhận xét: Trong phân mảnh:

Ưu điểm: Đảm bảo độ tin cậy và hiệu quả cho các câu vấn tin chỉ đọc

Nhược điểm: Câu vấn tin cập nhật phải được cập nhật tại tất cả các mảnh.

IV.5. Quản Lý Dữ Liệu Phân Tán

IV.5.2. Nhân Bản Dữ Liệu (Replicated data)

Lưu trữ các bảng được sao chép của dữ liệu ở nhiều vị trí khác nhau. Việc nhân bản dữ liệu phải đảm bảo việc nhất quán dữ liệu đòi hỏi tất cả các dữ liệu ở các vị trí khác nhau mà được nhân bản phải giống nhau.

Khi cập nhật ở 1 nơi nào đó thì ở những vị trí khác cũng phải được cập nhật.

*** Các loại nhân bản dữ liệu:**

Nhân bản đầy đủ(fully replicate): tất cả các mảnh dữ liệu được nhân bản.

Nhân bản từng phần: chỉ 1 vài bảng dữ liệu được nhân bản.

Không nhân bản: mỗi mảnh dữ liệu được lưu trữ ở 1 vị trí khác nhau.

IV.5. Quản Lý Dữ Liệu Phân Tán

IV.5.1. Phân mảnh dữ liệu (Fragmentation)

IV.5.2. Nhân Bản Dữ Liệu (Replicated data)

Nhận xét chung:

Nếu CSDL được phân mảnh thì phải phân tích câu hỏi thành các câu hỏi con để truy xuất đến các mảnh thích hợp

Nếu CSDL được nhân bản, DDBMS phải quyết định bảng nào để truy xuất. Thao tác đọc sẽ chọn bảng gần nhất để thỏa mãn yêu cầu, thao tác viết đòi hỏi tất cả các bản phải được cập nhật

IV.6. Tính Trong Suốt Của CSDL Phân Tán

Khái niệm trong suốt: Người sử dụng khi thao tác trong môi trường phân tán mà thấy giống như là đang thao tác trong môi trường tập trung.

IV.6.1. Trong suốt về mặt phân tán dữ liệu

IV.6.1.1. Trong suốt về mặt phân mảnh dữ liệu

Người sử dụng hay người lập trình không cần quan tâm đến tên mảnh dữ liệu và vị trí của nó. Người sử dụng không cần biết

- Dữ liệu nào được phân mảnh, dữ liệu nào được nhân bản
- Dữ liệu được chứa ở những vị trí nào trên môi trường phân tán

Q & A

1. So sánh xử lý phân tán và dữ liệu phân tán?
2. Phân mảnh, nhân bản?
3. Các loại trọng số trong môi trường CSDL phân tán?
4. Thuận lợi và hạn chế của CSDL phân tán?

- **Bài tập nhóm**

- Các sinh viên chia nhóm. Mỗi nhóm 3 sinh viên thực hiện yêu cầu sau:
- Sinh viên 1: *Cài hệ thống phân tán trên SQL server?*
- Sinh viên 2: *Cài hệ thống phân tán trên Oracle?*
- Sinh viên 3: *Xử lý dữ liệu phân tán trên SQL ?*

- **Tóm tắt**
 - **Các ưu điểm của HQTCSDL phân tán:**
- **Dữ liệu được đặt gần nơi có yêu cầu lớn nhất**
- **Truy xuất dữ liệu nhanh**
- **Xử lý dữ liệu nhanh**
- **Dễ dàng tăng trưởng**
- **Giảm chi phí hoạt động**
- **Ít chi phí hơn khi có sự cố**
- **Xử lý độc lập**