## Chương 3

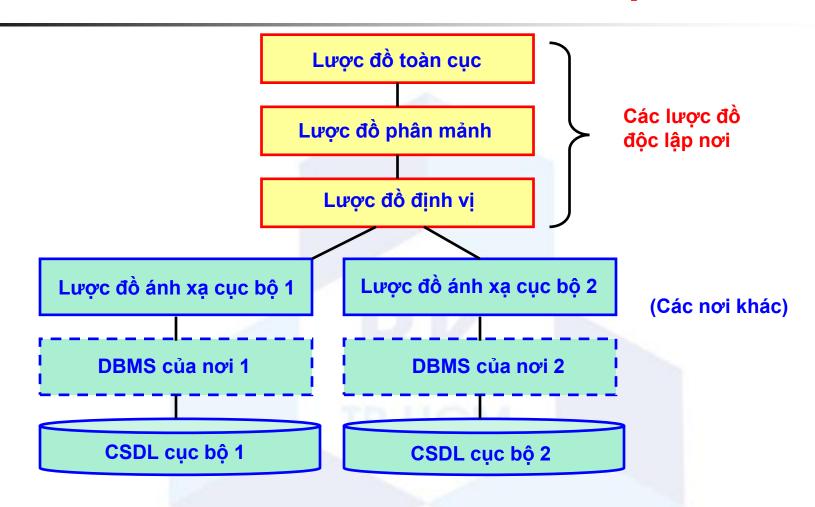
## Các mức trong suốt phân tán

#### Nội dung

- Kiến trúc tham khảo của CSDL phân tán.
- Các loại phân mảnh dữ liệu.
- Các điều kiện đúng đắn để phân mảnh dữ liệu.
- Phân mảnh ngang chính.
- Phân mảnh ngang dẫn xuất.
- Phân mảnh dọc.
- Phân mảnh hỗn hợp.
- Tính trong suốt phân tán dùng cho ứng dụng chỉ đọc.

### Nội dung

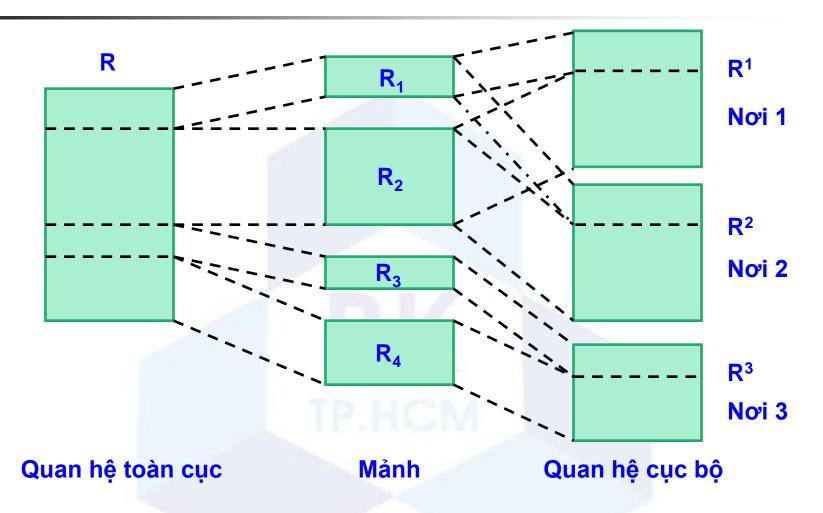
- Tính trong suốt phân tán dùng cho ứng dụng cập nhật.
- Các tác vụ cơ bản truy xuất CSDL phân tán.



Hình 3.1. Kiến trúc tham khảo dùng cho các CSDL phân tán.

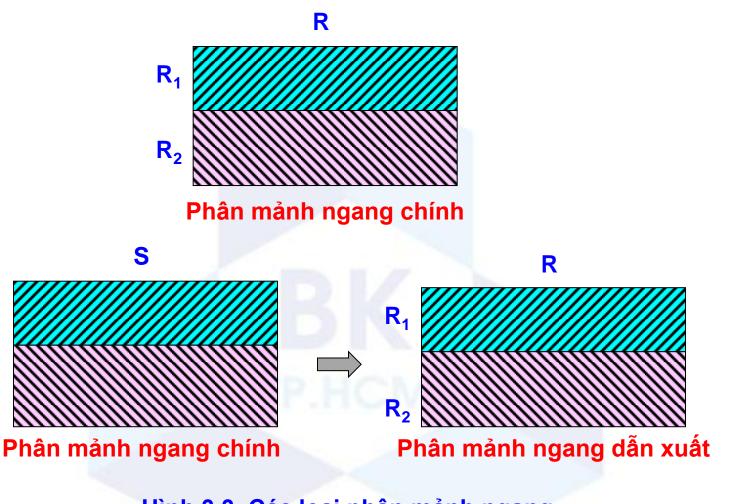
- \* Trong suốt phân mảnh (fragmentation transparency)
  - Không nhìn thấy các mảnh.
  - Nhìn thấy các quan hệ toàn cục (global relation).
  - Nhìn thấy lược đồ toàn cục (global schema).
- Trong suốt vị trí (location transparency)
  - Không nhìn thấy các quan hệ cục bộ.
  - Nhìn thấy các mảnh (fragment).
  - Nhìn thấy lược đồ phân mảnh (fragmentation schema).

- Trong suốt nhân bản (replication transparency)
  - Nhìn thấy các mảnh.
  - Không nhìn thấy sự nhân bản của các mảnh.
- Trong suốt ánh xạ cục bộ (local mapping transparency)
  - Nhìn thấy các quan hệ cục bộ (local relation).
  - Không nhìn thấy CSDL vật lý.
  - Nhìn thấy lược đồ định vị (allocation schema).
- \* Không trong suốt (no transparentcy)
  - Lược đồ ánh xạ cục bộ (local mapping schema)



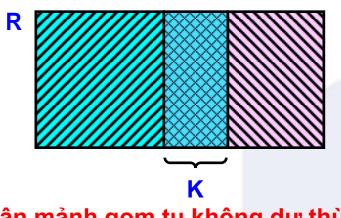
Hình 3.2. Các mảnh và các hình ảnh vật lý của một quan hệ toàn cục.

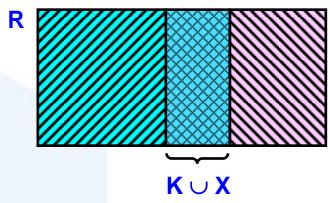
- Phân mảnh ngang (horizontal fragmentation)
  - Phân mảnh ngang chính (primary horizontal fragmentation)
  - Phân mảnh ngang dẫn xuất (derived horizontal fragmentation)



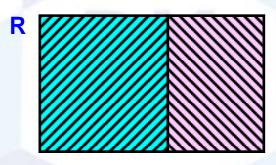
Hình 3.3. Các loại phân mảnh ngang.

- Phân mảnh dọc (vertical fragmentation)
  - Phân mảnh dọc gom tụ (vertical clustering fragmentation)
    - Phân mảnh dư thừa (redundant fragmentation)
    - Phân mảnh không dư thừa (non-redundant fragmentation)
  - Phân mảnh dọc tách biệt (vertical partitioning fragmentation)





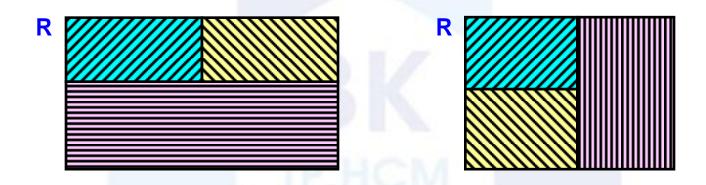
Phân mảnh gom tụ không dư thừa Phân mảnh gom tụ có dư thừa



Phân mảnh dọc tách biệt

Hình 3.4. Các loại phân mảnh dọc.

- \* Phân mảnh hỗn hợp (mixed fragmentation)
  - Kết hợp giữa phân mảnh ngang và phân mảnh dọc.



Hình 3.5. Phân mảnh hỗn hợp.

## Các điều kiện đúng đắn

- ❖ Quan hệ R được phân rã thành các mảnh R₁, R₂, ..., Rn
- \* Điều kiện đầy đủ (completeness condition)
  - Mọi mục dữ liệu trong R phải có trong một hoặc nhiều mảnh R<sub>i</sub>
  - ► Phân mảnh ngang:

```
\forall u \in R, \exists i \in [1, n]: u \in R_i
```

▶ Phân mảnh dọc:

```
\forall A \in Attr(R), \exists i \in [1, n]: A \in Attr(R_i)
```

## Các điều kiện đúng đắn

#### \* Điều kiện tái tạo (reconstruction condition)

Luôn luôn có thể xác định một phép toán quan hệ ∇ sao cho:

$$R = \nabla R_i \quad \forall R_i \in F_R \text{ v\'oi } F_R = \{R_1, R_2, ..., R_n\}$$

▶ Phân mảnh ngang:

$$R = R_1 \cup R_2 \cup ... \cup R_n$$

▶ Phân mảnh dọc:

$$R = R_1 \triangleright \triangleleft R_2 \triangleright \triangleleft ... \triangleright \triangleleft R_n$$

## Các điều kiện đúng đắn

#### Điều kiện tách biệt (disjointness condition)

- Nếu mục dữ liệu  $d_i$  có trong  $R_i$  thì nó không có trong bất kỳ mảnh  $R_k$  khác ( $k \neq i$ ).
- ► Phân mảnh ngang:

```
\forall i \neq k \text{ và } i, k \in [1, n]: R_i \cap R_k = \emptyset
```

hoặc

- Phân mảnh ngang chính (primary horizontal fragmentation) của một quan hệ toàn cục là sự phân chia các bộ của quan hệ này thành các tập hợp con dựa vào các thuộc tính của quan hệ này, mỗi tập hợp con được gọi là mảnh ngang chính (primary horizontal fragment).
- Mỗi mảnh ngang được tạo bởi một phép chọn trên quan hệ toàn cục.
- Vị từ được dùng làm điều kiện chọn của phép chọn để tạo ra một mảnh ngang được gọi là vị từ định tính (qualification) của mảnh ngang đó.

- Ví dụ
  - ▶ Quan hệ toàn cục:

sinhviên (mãsv, họtên, khoa)

► Các mảnh ngang:

```
sinhvi\hat{e}n_1 = \sigma_{khoa = 'CNTT'} sinhvi\hat{e}n

sinhvi\hat{e}n_2 = \sigma_{khoa = 'DIEN'} sinhvi\hat{e}n
```

► Các vị từ định tính:

```
q₁: khoa = 'CNTT'
```

$$q_2$$
: khoa = 'DIEN'

- \* Ví dụ
  - ▶ Xét các điều kiện đúng đắn
    - Điều kiện đầy đủ: được thỏa mãn nếu 'CNTT' và 'DIEN' chỉ là các giá trị có thể có của thuộc tính khoa.
    - Điều kiện tái tạo:
       sinhviên = sinhviên₁ ∪ sinhviên₂
    - Điều kiện tách biệt: được thỏa mãn vì mãsv là khóa của sinhviên.

- Để thỏa mãn điều kiện đầy đủ thì tập hợp các vị từ định tính của tất cả các mảnh ngang phải là đầy đủ, ít nhất là đối với tập hợp các giá trị cho phép.
- Điều kiện tái tạo luôn luôn được thỏa mãn bằng cách sử dụng phép hợp.
- Để thỏa mãn điều kiện tách biệt thì các vị từ định tính phải loại trừ nhau.

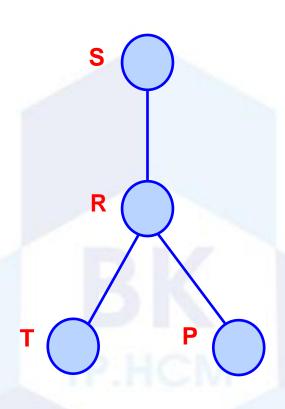
- Phân mảnh ngang dẫn xuất (derived horizontal fragmentation) của một quan hệ toàn cục là sự phân chia các bộ của quan hệ này thành các tập hợp con dựa vào phân mảnh ngang của một quan hệ khác (được gọi là quan hệ chủ), mỗi tập hợp con được gọi là mảnh ngang dẫn xuất (derived horizontal fragment).
- Vị từ định tính của mảnh ngang dẫn xuất bao gồm điều kiện kết và vị từ định tính của mảnh ngang chủ tương ứng.

#### Ví dụ

- Quan hệ toàn cục: điểmthi (mãsv, mãmh, mãbm, điểm)
- ► Các mảnh ngang dẫn xuất:  $\text{diểmthi}_1 = \text{diểmthi} > <_{\text{mãsv} = \text{mãsv}} \text{sinhviên}_1$  $\text{diểmthi}_2 = \text{diểmthi} > <_{\text{mãsv} = \text{mãsv}} \text{sinhviên}_2$
- ► Các vị từ định tính:
- q<sub>1</sub>: sinhviên.mãsv = điểmthi.mãsv AND sinhviên.khoa = 'CNTT'
- q<sub>2</sub>: sinhviên.mãsv = điểmthi.mãsv AND sinhviên.khoa = 'DIEN'

#### Ví dụ

- ► Xét các điều kiện đúng đắn
  - Điều kiện đầy đủ: được thỏa mãn nếu tất cả mã sinh viên mãs v trong quan hệ điểm thi đều phải có trong quan hệ sinh viên (ràng buộc tham chiếu).
  - Điều kiện tái tạo:
     điểmthi = điểmthi₁ ∪ điểmthi₂
  - Điều kiện tách biệt: được thỏa mãn nếu một bộ của quan hệ điểmthi không tương ứng với hai bộ của quan hệ sinhviên thuộc hai mảnh khác nhau, khi đó mãs v là khóa của sinhviên.



Hình 3.6. Cây phân mảnh ngang dẫn xuất.

#### Phân mảnh dọc

Phân mảnh dọc (vertical fragmentation) của một quan hệ toàn cục là sự phân chia tập thuộc tính của quan hệ này thành các tập thuộc tính con; các mảnh dọc (vertical fragment) có được bằng cách chiếu quan hệ toàn cục trên mỗi tập thuộc tính con.

#### Phân mảnh dọc

#### Ví dụ

- Quan hệ toàn cục: giảngviên (mãgv, họtên, phái, mãsếp, mãbm)
- ▶ Phân mảnh dọc không dư thừa:

```
gv_1 = \Pi_{m 	ilde{a}nv, h 	opt 	ilde{e}n, l 	ilde{w} ng} giảng viên gv_2 = \Pi_{m 	ilde{a}nv, m 	ilde{a}s 	ilde{e}p, m 	ilde{a}pb} giảng viên giảng viên gv_1 > 0 m 	ilde{a}gv = m 	ilde{a}gv gv_2
```

▶ Phân mảnh dọc dư thừa:

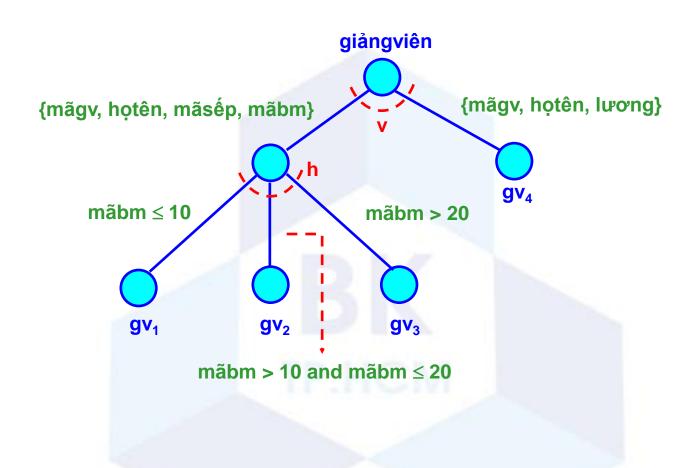
```
gv_1 = \Pi_{m 	ilde{a}gv, hoten, lương} giảng viên gv_2 = \Pi_{m 	ilde{a}gv, hoten, m 	ilde{a}sép, m 	ilde{a}bm} giảng viên gv_1 = gv_1 > 0 m 	ilde{a}gv_1 = m 	ilde{a}gv_1 = gv_1 > 0 m 	ilde{a}gv_1 = m 	i
```

### Phân mảnh hỗn hợp

- Một mảnh ngang được phân mảnh dọc.
- Một mảnh dọc được phân mảnh ngang.
- \* Ví dụ
  - Quan hệ toàn cục: giảngviên (mãgv, họtên, lương, mãsếp, mãbm)
  - ▶ Phân mảnh hỗn hợp:

```
\begin{array}{l} gv_1 = \sigma_{\text{ mãbm} \leq 10} \; \Pi_{\text{ mãgv, họtên, mãsép, mãbm}} \; giảng viên \\ gv_2 = \sigma_{10 < \text{ mãbm} \leq 20} \; \Pi_{\text{ mãgv, họtên, mãsép, mãbm}} \; giảng viên \\ gv_3 = \sigma_{\text{ mãbm} > 20} \; \Pi_{\text{ mãgv, họtên, mãsép, mãbm}} \; giảng viên \\ gv_4 = \Pi_{\text{ mãgv, họtên, lương}} \; nhân viên \\ giảng viên = (gv_1 \cup gv_2 \cup gv_3) \; \triangleright \lhd_{\text{ mãgv} = mãgv}} \; \Pi_{\text{ mãgv, phái}} \; gv_4 \end{array}
```

## Phân mảnh hỗn hợp



Hình 3.7. Cây phân mảnh của quan hệ giảng viên.

### Phân mảnh hỗn hợp

#### Cây phân mảnh (fragmentation tree)

- Nút gốc tương ứng với một quan hệ toàn cục.
- Các nút lá tương ứng với các mảnh cuối cùng.
- Các nút trung gian tương ứng với các kết quả trung gian của các biểu thức định nghĩa mảnh.
- Tập hợp các nút con của một nút biểu diễn sự phân mảnh của nút này (phân mảnh ngang hoặc phân mảnh dọc).

#### **Example.DDB**

#### \* Lược đồ toàn cục

giảngviên (<u>mãgv</u>, họtên, lương, mãsếp, mãbm) bộmôn (<u>mãbm</u>, tênbm, dãynhà, mãsếp) sinhviên (<u>mãsv</u>, họtên, khoa) điểmthi (mãsv, mãmh, mãbm, điểm)

#### Lược đồ phân mảnh

```
giảngviên_1 = \sigma_{mãbm \le 10} \Pi_{mãnv, họtên, mãsép, mãbm} giảngviên giảngviên_2 = \sigma_{10 < mãbm \le 20} \Pi_{mãnv, họtên, mãsép, mãbm} giảngviên giảngviên_3 = \sigma_{mãbm > 20} \Pi_{mãnv, họtên, mãsép, mãbm} giảngviên giảngviên_4 = \Pi_{mãnv, hotên, lương} giảngviên
```

#### **Example.DDB**

#### Lược đồ phân mảnh

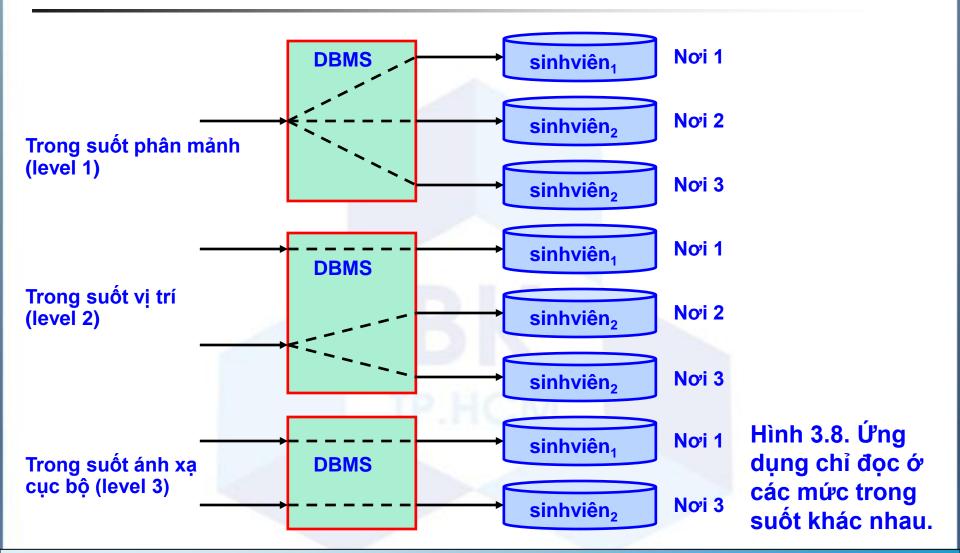
bộmôn<sub>1</sub> = 
$$\sigma_{\text{mãbm} \le 10}$$
 bộmôn  
bộmôn<sub>2</sub> =  $\sigma_{10 < \text{mãbm} \le 20}$  bộmôn  
bộmôn<sub>3</sub> =  $\sigma_{\text{mãbm} > 20}$  bộmôn

$$sinhvi\hat{e}n_1 = \sigma_{khoa = 'CNTT'} sinhvi\hat{e}n$$
  
 $sinhvi\hat{e}n_2 = \sigma_{khoa = 'DIEN'} sinhvi\hat{e}n$ 

```
điểmthi_1 = điểmthi > <_{masv = masv} sinhviên_1
điểmthi_2 = điểmthi > <_{masv = masv} sinhviên_2
```

#### Ví dụ 1

- Cho biết tên của sinh viên có mã được nhập từ thiết bị đầu cuối.
- ► Mức 1 Trong suốt phân mảnh read (terminal, \$mãsv); select họtên into \$họtên from sinhviên where mãsv = \$mãsv; if #FOUND then write (terminal, \$họtên) else write (terminal, 'Not found');



► Mức 2 – Trong suốt vị trí read (terminal, \$masv); select họtên into \$họtên from sinhviên1 where *mãsv* = \$mãsv; if not #FOUND then select *họtên* into \$họtên from sinhviên2 where *mãsv* = \$mãsv; if #FOUND then write (terminal, \$hoten) else write (terminal,'Not found');

► Mức 3 – Trong suốt ánh xạ cục bộ read (terminal, \$masv); select họtên into \$họtên from sinhviên1 at site 1 where *mãsv* = \$mãsv; if not #FOUND then select *họtên* into \$họtên from sinhviên2 at site 3 where *mãsv* = \$mãsv; if #FOUND then write (terminal, \$hoten) else write (terminal,'Not found');

 Trường hợp dữ liệu nhập có liên quan đến vị từ định tính của mảnh

```
read (terminal, $masv);
read (terminal, $khoa);
case $khoa of
  'CNTT': select họtên into $họtên
          from sinhviên1
          where mãsv = $mãsv;
  'DIEN': select họtên into $họtên
         from sinhviên2
         where mãsv = $mãsv;
end;
if #FOUND then write (terminal, $hoten)
else write (terminal, 'Not found');
```

#### Ví dụ 2

- Cho biết tên của bộ môn phụ trách môn học có mã môn học được nhập từ thiết bị đầu cuối.
- Một môn học chỉ thuộc một bộ môn phụ trách.

► Mức 1 – Trong suốt phân mảnh read (terminal, \$mãmh); select tênbm into \$tênbm from điểmthi, bộmôn where điểmthi.mãbm = bộmôn.mãbm and điểmthi.mãmh = \$mãmh; if #FOUND then write (terminal, \$tênbm) else write (terminal, 'Not found');

► Mức 2 – Trong suốt vị trí read (terminal, \$mãmh); select tênbm into \$tênbm from điểmthi1, bộmôn1 where điểmthi1.mãbm = bộmôn1.mãbm and điểmthi1.mãmh = \$mãmh; if not #FOUND then select tênbm into \$tênbm from điểmthi2, bộmôn2

where *diểmthi2.mãbm* = *bộmôn2.mãbm* and *diểmthi2.mãmh* = \$mãmh;

if #FOUND then write (terminal, \$tênbm) else write (terminal, 'Not found');

► Mức 3 – Trong suốt ánh xạ cục bộ

*bộmôn₁* đặt tại nơi 1.

bộmôn, đặt tại nơi 2.

điểmthi₁ đặt tại nơi 3.

điểmthi<sub>2</sub> đặt tại nơi 4.

```
read (terminal, $mamh);
select mãbm into $mãbm
from điểmthi1 at site 3
where mãmh = $mãmh;
if #FOUND then
 begin
    send $mãbm from site 3 to site 1;
    select tênbm into $tênbm
    from bộmôn1 at site 1
    where mãbm = $mãbm;
 end
```

```
else
 begin
    select mãbm into $mãbm
    from điểmthi2 at site 4
    where mãmh = $mãmh;
    if #FOUND then
      begin
           send $mãbm from site 4 to site 2;
           select tênbm into $tênbm
           from bộmôn2 at site 2
           where mãbm = $mãbm;
     end
 end;
if #FOUND then write (terminal, $tênbm)
else write (terminal, 'Not found');
```

- Cập nhật dữ liệu (thêm, sửa, xóa) phải bảo đảm các ràng buộc toàn vẹn về khóa chính, khóa ngoại, phụ thuộc hàm, ràng buộc nghiệp vụ ...
- Qui tắc read-one write-all
  - Việc lấy dữ liệu có thể được thực hiện từ bất kỳ một bản sao nào.
  - Việc cập nhật một mục dữ liệu phải được thực hiện trên tất cả các bản sao của nó.

- Các trường hợp sửa dữ liệu trong CSDL phân tán
  - Nếu mục dữ liệu bị sửa không có trong vị từ định tính: sửa dữ liệu giống như trong CSDL tập trung.
  - Nếu mục dữ liệu bị sửa có trong vị từ định tính và giá trị của vị từ định tính không bị thay đổi khi thay thế dữ liệu cũ và dữ liệu mới: sửa dữ liệu giống như trong CSDL tập trung.
  - Nếu mục dữ liệu bị sửa có trong vị từ định tính và giá trị của vị từ định tính bị thay đổi khi thay thế dữ liệu cũ và dữ liệu mới: di chuyển dữ liệu giữa các mảnh.

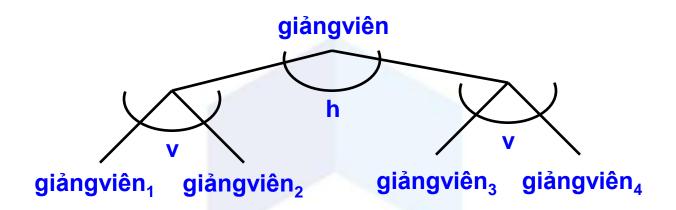
#### Qui tắc owner - member

- Thỏa mãn ràng buộc toàn vẹn tham chiếu.
- Thêm vào quan hệ chủ trước, thêm vào quan hệ bộ phận sau.
- Xóa trong quan hệ bộ phận trước, xóa trong quan hệ chủ sau.

### Ví dụ

- Sửa dữ liệu của giảng viên có mã 100: mã bộ môn 3 thành mã bộ môn 15.
- ▶ Các mảnh:

```
giảngviên<sub>1</sub> được đặt tại nơi 1 và 5. giảngviên<sub>2</sub> được đặt tại nơi 2 và 6. giảngviên<sub>3</sub> được đặt tại nơi 3 và 7. giảngviên<sub>4</sub> được đặt tại nơi 4 và 8.
```



```
giảngviên_1 = \Pi_{\text{mãgv, họtên, lương}} \sigma_{\text{mãbm}} <= 10 giảngviên giảngviên_2 = \Pi_{\text{mãgv, mãsép, mãbm}} \sigma_{\text{mãbm}} <= 10 giảngviên giảngviên_3 = \Pi_{\text{mãgv, mãsép, mãbm}} \sigma_{\text{mãbm} > 10} giảngviên giảngviên_4 = \Pi_{\text{mãgv, mãsép, mãbm}} \sigma_{\text{mãbm} > 10} giảngviên
```

Hình 3.9. Cây phân mảnh của quan hệ giảng viên.

#### giảngviên<sub>1</sub>

mãgv	họtên	lương
100	Hùng	10000

giảngviên<sub>2</sub>

mãgv	mãsếp	mãbm
100	20	3

Trước cập nhật

Sau cập nhật giảngviên<sub>3</sub>

mãgv	họtên	mãbm
100	Hùng	15

giảngviên<sub>4</sub>

mãgv	lương	mãsếp
100	10000	20

Hình 3.10. Ứng dụng cập nhật.

► Mức 1 – Trong suốt phân mảnh

update giảngviên

set *mãbm* = 15

where  $m\tilde{a}gv = 100$ ;

► Mức 2 – Trong suốt vị trí select họtên, lương into \$họtên, \$lương from giảng*viên1* where  $m\tilde{a}gv = 100$ ; if #FOUND then begin select *mãsép* into \$mãsép from giảng*viên*2 where  $m\tilde{a}gv = 100$ ; insert into giảng viên 3 (mãg v, họ tên, mã bm) values (100, \$hoten, 15);

```
insert into giảng viên 4 (mãg v, lương, mã sếp)
  values (100, $lương, $mãsếp);
  delete from giảng viên 1
  where m\tilde{a}gv = 100;
  delete from giảng viên 2
  where m\tilde{a}gv = 100
end;
```

► Mức 3 – Trong suốt ánh xạ cục bộ select họtên, lương into \$họtên, \$lương from giảng viên 1 at site 1 where  $m\tilde{a}gv = 100$ ; if #FOUND then begin select *mãsép* into \$mãsép from giảng *viên2* at site 2 where  $m\tilde{a}gv = 100$ ; insert into giảng viên3 (mãgv, họtên, mãbm) at site 3 values (100, \$hoten, 15);

```
insert into giảng viên3 (mãgv, họtên, mãbm) at site 7
values (100, $hoten, 15);
insert into giảng viên 4 (mãg v, lương, mã sếp) at site 4
values (100, $lương, $mãsếp);
insert into giảng viên 4 (mãg v, lương, mã sếp) at site 8
values (100, $ lương, $ mãsếp);
delete from giảng viên 1 at site 1
where m\tilde{a}gv = 100;
delete from giảng viên 1 at site 5
where m\tilde{a}gv = 100;
```

```
delete from giảng viên2 at site 2
where m\tilde{a}gv = 100;
delete from giảng viên2 at site 6
where m\tilde{a}gv = 100
end;
```

Lưu ý: Nếu tồn tại phụ thuộc hàm mãbm → mãsếp thì chương trình này chưa sửa mãsếp của giảng viên này.

### Ví dụ

Cho biết tất cả mã môn học của các sinh viên cho trước.

► Truy xuất CSDL ứng với mỗi mã sinh viên

```
repeat
  read (terminal, $masv);
  delete from $mamh_rel;
 insert into $mamh_rel ($mamh)
  select mãmh
 from điểmthi
 where mãsv = $mãsv;
  repeat
   read ($mamh_rel, $mamh);
   write (terminal, $mãsv, $mãmh)
  until END-OF-$mamh rel
until END-OF-TERMINAL-INPUT;
```

► Truy xuất CSDL sau khi nhập vào tất cả mã sinh viên

```
repeat
  read (terminal, $masv);
  insert into $mãsv_rel ($mãsv)
  values ($mãsv)
until END-OF-TERMINAL-INPUT;
insert into $mamh_rel ($masv, $mamh)
select $mãsv, mãmh
from điểmthi, $mãsv_rel
where điểmthi. mãsv = $mãsv_rel.$mãsv;
repeat
  read ($mamh_rel, $masv, $mamh);
  write (terminal, $mãsv, $mãmh)
until END-OF-$mamh rel;
```

► Truy xuất CSDL trước khi nhập vào tất cả mã sinh viên

```
insert into $temp_rel ($temp_mamh, $temp_masv)
select mãmh, mãsv
from điểmthi;
repeat
  read (terminal, $masv);
  delete from $temp2_rel;
  insert into $temp2_rel ($temp2_mamh)
  select $temp_mamh
  from $temp_rel
  where $temp_masv = $masv;
```

```
repeat
read ($temp2_rel, $temp2_mamh);
write (terminal, $masv, $temp2_mamh)
until END-OF-$temp2_rel
until END-OF-TERMINAL-INPUT;
```