

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

CƠ SỞ DỮ LIỆU PHÂN TÁN

THIẾT KẾ CƠ SỞ DỮ LIỆU PHÂN TÁN

Ts. Phan Thị Hà

Nội dung

- ☐ Phương pháp phân mảnh ngang
- ☐ Phân mảnh ngang dẫn xuất
- ☐ Phân mảnh dọc
- ☐ Phương pháp phân mảnh hỗn hợp
- ☐ Cấp phát và mô hình cấp phát

Phân mảnh ngang

- ❑ Phân mảnh ngang là việc tách ngang một quan hệ toàn cục thành nhiều nhiều mảnh. Mỗi một mảnh là một quan hệ khả hợp, chứa một số bộ và các bộ trong các quan hệ con là tách biệt nhau.
- ❑ Phân mảnh ngang thực chất là phép chọn quan hệ thỏa mãn một biểu thức điều kiện cho trước.
- ❑ Có hai phương pháp phân mảnh ngang:
 - *Phân mảnh ngang nguyên thủy*: Là phân mảnh ngang được thực hiện trên các vị từ của chính quan hệ.
 - *Phân mảnh ngang dẫn xuất*: Là phân mảnh một quan hệ dựa trên các vị từ của quan hệ khác.

Thông tin Phân mảnh ngang

- ❑ Thông tin về CSDL bao gồm tập các quan hệ, mỗi quan hệ, tập các thuộc tính và tập các phụ thuộc hàm
- ❑ Thông tin về các ứng dụng gồm các câu truy vấn trên các quan hệ, vị trí các truy vấn....
- ❑ (Ko)Thông tin về mạng máy tính, cấu trúc, băng thông...
- ❑ (Ko)Thông tin về hệ thống máy tính, bộ nhớ lưu trữ...

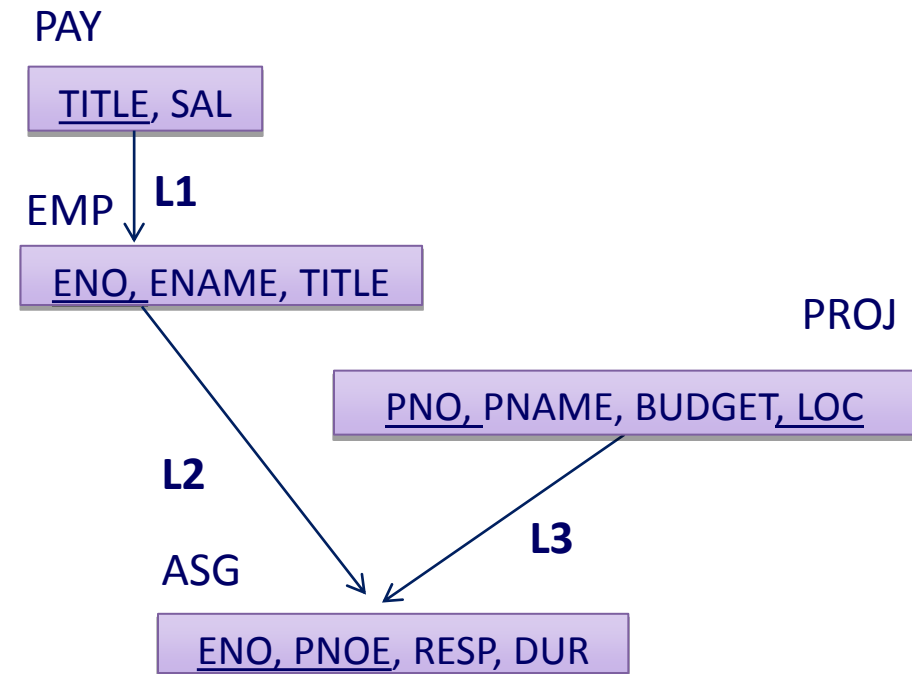
Yêu cầu thông tin về mạng và thông tin về hệ thống máy tính chỉ được sử dụng trong các mô hình cấp phát, không sử dụng trong các thuật toán phân mảnh dữ liệu

Phân mảnh ngang: Dựa trên thông tin về CSDL

- ❑ Thông tin về CSDL: Là thông tin về mối quan hệ một - một, một - nhiều và nhiều - nhiều giữa các quan hệ (bảng), được liên kết bằng các đường nối (Link) có hướng, kết nối bằng

Thông tin vđịnh lượng cần thiết về cơ sở dl là lực lượng của quan hệ R, ký hiệu $|R|$ là Card(R)

- Mỗi quan hệ một - nhiều trở từ các quan hệ PAY đến quan hệ EMP bằng đường nối L1
- Mỗi quan hệ nhiều - nhiều trở từ các quan hệ EMP và PROJ đến quan hệ ASG bằng hai đường nối L2 và L3.



Phân mảnh ngang: Dựa trên thông tin ứng dụng

Thông tin về ứng dụng: Thông tin định lượng và Thông tin định tính

Thông tin định lượng: chủ yếu sử dụng trong các mô hình cấp phát (ko sử dụng ở đây)

Thông tin định tính là cơ bản: hướng dẫn hoạt động phân mảnh .

a) Vị từ đơn giản

□ Ký hiệu: $p_j: A_i \theta$ “value”,

Trong đó:

A_i là thuộc tính của $R(A_1, \dots, A_n)$, $\theta \in \{=, <, \neq, \leq, >, \geq\}$

“Value” là một giá trị A_i

□ Ký hiệu Pr là tập tất cả các vị từ đơn giản được định nghĩa trên quan hệ R : $Pr = \{p_1, p_2, \dots, p_m\}$.

Phân mảnh ngang: Dựa trên thông tin ứng dụng

b). Vị từ hội sơ cấp

□ $P_r = \{p_1, p_2, \dots, p_m\}$ là một tập các vị từ đơn giản

$$M = \left\{ m_i \mid m_i = \bigwedge_{p_k \in P_r} p_k^* \right\}$$

□ Trong đó, $p_k^* = p_k$ hoặc $p_k^* = \neg p_k$. Như vậy, vị từ đơn giản xuất hiện trong vị từ hội sơ cấp dưới dạng tự nhiên hoặc dạng phủ định của nó.

Phân mảnh ngang: Dựa trên thông tin ứng dụng

Ví dụ:

- p_1 : TITLE = “Elect.Eng”
- p_2 : TITLE = “Syst. Anal”
- p_3 : TITLE = “Mech. Eng”
- p_4 : TITLE = “Programmer”
- p_5 : SAL \leq 30000
- p_6 : SAL $>$ 30000

PAY

TITLE	SAL
Elect.Eng	40000
Mech.Eng	27000
Programmer	24000
Syst.Anal	34000

Phân mảnh ngang: Dựa trên thông tin ứng dụng

Sau đây là 1 số vị từ hội

$m_1: \text{TITLE} = \text{"Elect.Eng"} \wedge \text{SAL} \leq 30000$

$m_2: \text{TITLE} = \text{"Elect.Eng"} \wedge \text{SAL} > 30000$

$m_3: \neg(\text{TITLE} = \text{"Elect.Eng"}) \wedge \text{SAL} \leq 30000$

$m_4: \neg(\text{TITLE} = \text{"Elect.Eng"}) \wedge \text{SAL} > 30000$

$m_5: \text{TITLE} = \text{"Programmer"} \wedge \text{SAL} \leq 30000$

$m_6: \text{TITLE} = \text{"Programmer"} \wedge \text{SAL} > 30000$

Trên đây là 1 số vị từ hội đc tạo ra từ tập vị từ cơ sở trên, Các vị từ này đã đc viết đơn giản hóa của các hội. Định nghĩa hội đòi hỏi mỗi vị từ ở dạng tự nhiên hoặc phủ định của nó, bởi vậy m_1 có thể viết là

$$m_1: \text{TITLE} = \text{"Elect. Eng."} \wedge \text{TITLE} \neq \text{"Syst. Anal."} \wedge \text{TITLE} \neq \text{"Mech. Eng."} \\ \wedge \text{TITLE} \neq \text{"Programmer"} \wedge \text{SAL} \leq 30000$$

Phân mảnh ngang: Dựa trên thông tin ứng dụng

Ký hiệu

- ❑ Độ tuyển hội sơ cấp (Minterm Selectivity): số bộ của quan hệ kết quả được chọn theo vị từ hội sơ cấp cho trước. Ký hiệu là $sel(m)$. Ví dụ, $sel(m1)=0$. $sel(m2)=1$.
- ❑ Tần số ứng dụng người dùng truy nhập dữ liệu.
Nếu $Q = \{q1, q2, \dots, qk\}$ là tập truy vấn, tần số truy nhập của truy vấn qi trong một khoảng thời gian đã cho, ký hiệu là $acc(qi)$
- ❑ Tần số truy nhập hội sơ cấp hội sơ cấp m , ký hiệu là $acc(m)$.

Phân mảnh ngang cơ sở

- ❑ Phân mảnh ngang cơ sở được định nghĩa bằng phép chọn trên quan hệ toàn R: $f_i = \sigma_{m_i}(R)$
 $i=1 \dots n$; trong đó m_i là vị từ hội sơ cấp.
- ❑ f_i được gọi là mảnh hội sơ cấp (Minterm Fragment).
- ❑ Một tập M các vị từ hội sơ cấp, số lượng phân mảnh ngang của quan hệ R bằng số lượng các vị từ hội sơ cấp.

Phân mảnh ngang cơ sở

Ví dụ: Giả sử tập các vị từ hội sơ cấp:

$$m_1: \{ \text{BUDGET} \leq 200000 \}$$

$$m_2: \{ 200000 < \text{BUDGET} \leq 400000 \}$$

$$m_3: \{ 400000 < \text{BUDGET} \leq 600000 \}$$

$$m_4: \{ 600000 < \text{BUDGET} \}$$

Khi đó quan hệ PROJ được phân rã thành các mảnh ngang như sau:

$$\text{PROJ}_1 = \sigma_{\text{BUDGET} \leq 200000}(\text{PROJ})$$

$$\text{PROJ}_2 = \sigma_{200000 < \text{BUDGET} \leq 400000}(\text{PROJ})$$

$$\text{PROJ}_3 = \sigma_{400000 < \text{BUDGET} \leq 600000}(\text{PROJ})$$

$$\text{PROJ}_4 = \sigma_{600000 < \text{BUDGET}}(\text{PROJ})$$

- Cho trước: Quan hệ R , tập các vị từ đơn giản Pr
- Đầu ra: Tập các mảnh của $R = \{R_1, R_2, \dots, R_w\}$ tuân theo các luật phân mảnh
- Điều kiện:
 - Pr phải đầy đủ
 - Pr phải tối thiểu

Tính đầy đủ của vị từ đơn giản

Pr là đầy đủ khi và chỉ khi xác suất truy nhập của mỗi ứng dụng đến bộ bất kỳ của mảnh hội sơ cấp bất kỳ được định nghĩa theo Pr là như nhau.

- Vị từ đầy đủ sẽ đảm bảo cho các mảnh sơ cấp nhất quán về mặt logic. Đồng nhất về mặt thống kê theo cách ứng dụng truy nhập. Vì vậy, một tập vị từ đầy đủ là cơ sở cho việc phân mảnh ngang cơ sở.

VD(1). Cho 1 quan hệ EMP

EmpID	Name	Loc	Sal	DOB	Dept
283948	Joe	LA	25,000	2/6/43	Maintenance
109288	Larry	New York	35,200	12/3/52	Payroll
284003	Moe	LA	43,000	7/12/56	Maintenance
320021	Sam	New York	53,500	8/30/47	Production
123456	Steve	Minneapolis	67,000	5/14/78	Management
334456	Jack	New York	55,000	5/30/67	Production
222222	Saeed	Minneapolis	34,000	4/27/59	Management

EMP Table

UD1: Giả sử ứng dụng AP1 truy vấn vào quan hệ EMP để tìm kiếm những nhân viên làm việc ở in Los Angeles (LA),

Tập vị từ đơn giản là $Pr = \{p1: Loc = "LA"\}$

Tập vị từ hội là $\{m1: Loc = "LA", m2: Loc \neq "LA"\}$, vậy tập Pr là cực tiểu và đầy đủ, các mảnh được phân ra:

Fragment F1: Create table LA_EMPS as Select * from EMP Where Loc = "LA";

Fragment F2: Create table NON_LA_EMPS as Select * from EMP Where Loc \neq "LA";

UD2: Giả sử ứng dụng 1 thêm đk loại bỏ tất cả những người có salary ≤ 30000 thì khi đó Pr không còn cực tiểu và đầy đủ nữa, khi đó phải sửa thành

$Pr = \{p1: Loc = "LA", p2: salary > 30000\}$

Tập các vị từ hội

- $\{m1: Loc = "LA" \text{ Sal} > 30000,$
- $m2: Loc = "LA" \text{ Sal} \leq 30000,$
- $m3: Loc \neq "LA" \text{ Sal} > 30000,$
- $m4: Loc \neq "LA" \text{ Sal} \leq 30000\}$

Completeness (4)

JNO	JNAME	BUDGET	LOC
J1	Instrumental	150,000	Montreal
J2	Database Dev.	135,000	New York
J3	CAD/CAM	250,000	New York
J4	Maintenance	350,000	Orlando

$$J_1 = \sigma_{LOC="MONTREAL"}(J)$$

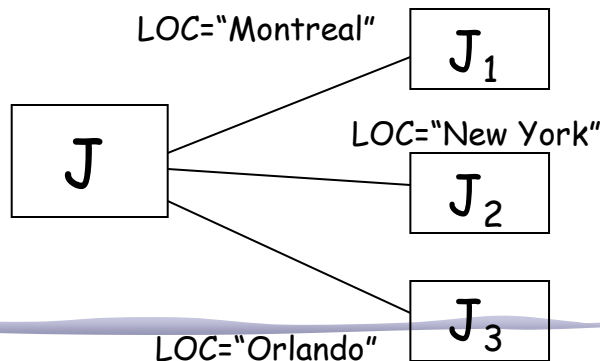
$$J_2 = \sigma_{LOC="NewYork"}(J)$$

$$J_3 = \sigma_{LOC="Orlando"}(J)$$

Case 1: The only application that accesses J wants to access the tuples according to the location.

The set of simple predicates

$$Pr = \begin{cases} LOC="Montreal", \\ LOC="New York", \\ LOC="Orlando" \end{cases}$$



Là đầy đủ vì mỗi bộ trong 1 mảnh dduwwocj phân ra có xác suất truy nhập như sau

Completeness (5)

Example:

Pr = { LOC="Montreal",
LOC="New York",
LOC="Orlando" }

J1	JNO	JNAME	BUDGET	LOC
	001	Instrumental	150,000	Montreal

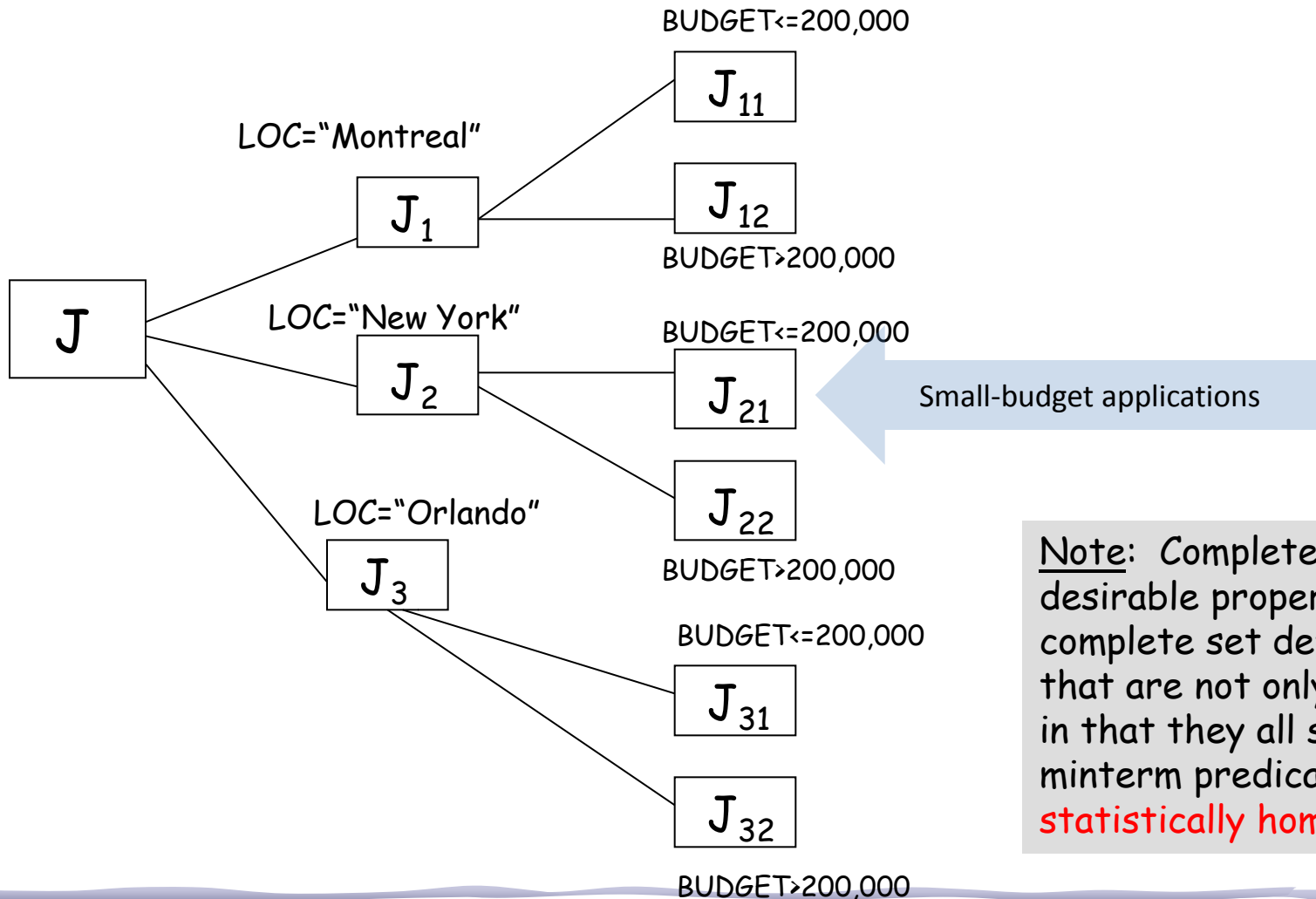
J2	JNO	JNAME	BUDGET	LOC
	004	GUI	135,000	New York
	007	CAD/CAM	250,000	New York

J3	JNO	JNAME	BUDGET	LOC
	003	Database Dev.	310,000	Orlando

Case 2: There is a second application which accesses only those project tuples where the budget is less than \$200,000.

- Since tuple "004" is accessed more frequently than tuple "007", **Pr** is not complete.
- To make the the set complete, we need to add (BUDGET < 200,000) to **Pr**.

Completeness (6)



Note: Completeness is a desirable property because a complete set defines fragments that are not only **logically uniform** in that they all satisfy the minterm predicate, but **statistically homogeneous**.

Tính cực tiểu của vị từ đơn giản

- ❑ Nếu một vị từ ảnh hưởng đến cách phân mảnh được thực hiện (vd: gây ra việc mảnh f được phân thành các mảnh f_i và f_j) thì có ít nhất một ứng dụng truy cập f_i và f_j một cách khác nhau
- ❑ Nói cách khác, vị từ đơn giản phải liên quan đến quyết định phân mảnh
- ❑ Nếu tất cả các vị từ của một tập Pr là liên quan, Pr là **tối thiểu**

$$\frac{acc(m_i)}{card(f_i)} \neq \frac{acc(m_j)}{card(f_j)} \quad \begin{array}{l} card(f) \text{ là lực lượng } |f|. \\ acc(m) \text{ là tần số truy nhập hội sơ cấp } m \end{array}$$

Minimality

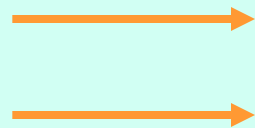
Relevant:

Đã m_i và m_j là các vị từ hội xác định be two almost identical minterm predicates:

$$m_i = p_1 \wedge p_2 \wedge p_3$$

$$m_j = p_1 \wedge \neg p_2 \wedge p_3$$

p_2 là liên đới nếu và chỉ nếu



fragment f_i

fragment f_j

$$\frac{acc(m_i)}{card(f_i)} \neq \frac{acc(m_j)}{card(f_j)}$$

Access frequency

Cardinality

Tức là, phải có ít nhất 1 ứng dụng truy nhập khác nhau vào f_i và f_j

i.e., Vị từ đơn giản p_i phải liên đới trong việc xác định 1 mảnh

Cực tiểu: Nếu tất cả các vị từ của tập P_r là liên đới

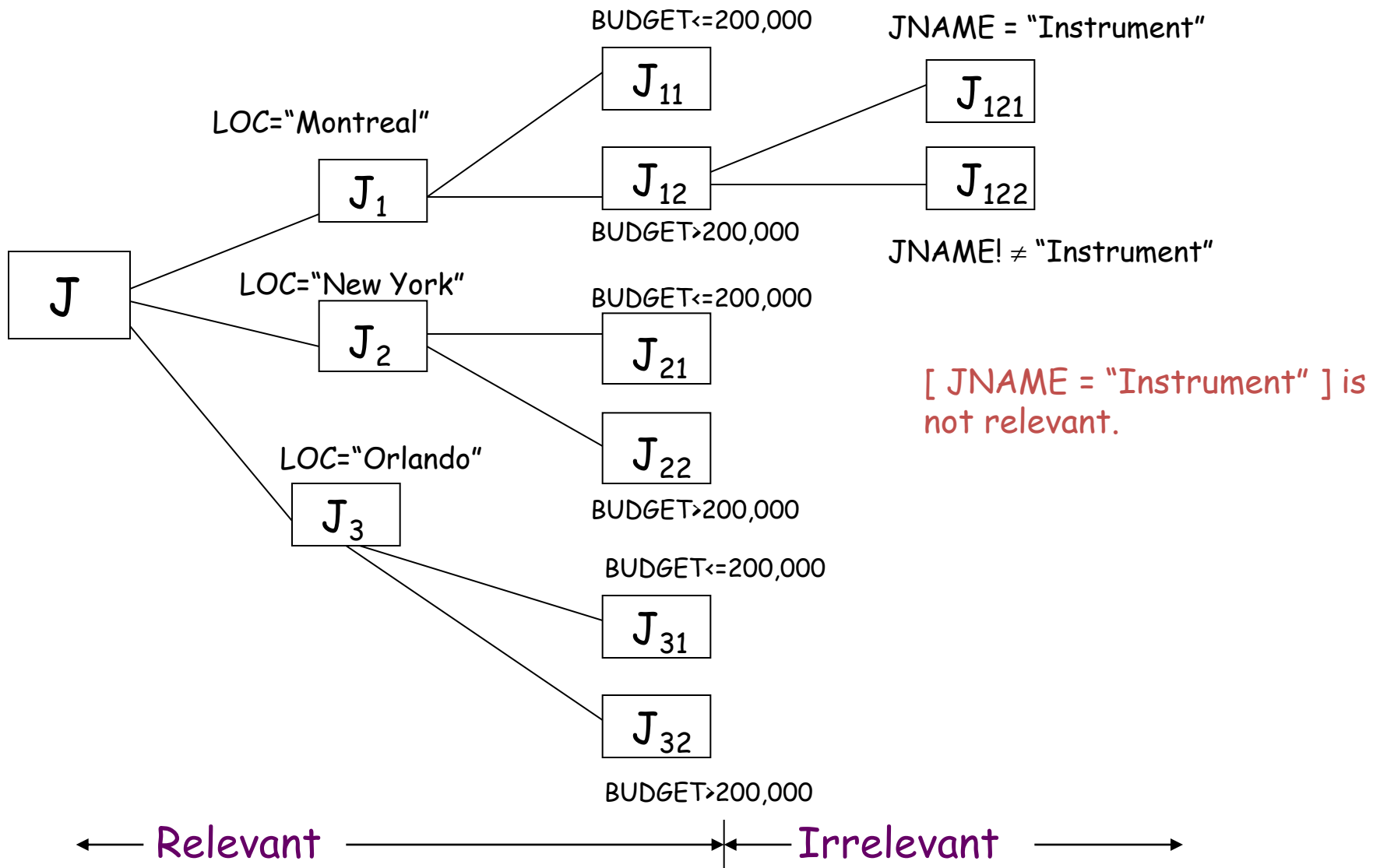
Vd về Complete and Minimal

2 ứng dụng:

1. UD1 truy cập theo vị trí (Loc)
2. UD2 chỉ truy cập vào những bộ có kinh phí (BUDGET) nhỏ hơn \$200,000.

TH1: $Pr = \{Loc = \text{"Montreal"}, Loc = \text{"New York"}, Loc = \text{"Orlando"}, BUDGET \leq 200\}$ is
Đầy đủ và cực tiểu.

TH2: Nếu đưa thêm vị từ $JNAME = \text{"Instrumentation"}$ vào Pr ,
thì vị từ mới sẽ không liên đới gì tới các ứng dụng, $\Rightarrow Pr$
không cực tiểu



Thuật toán xác định tập vị từ đầy đủ và cực tiểu

- ❑ Thuật toán COM_MIN tạo ra một tập vị từ đầy đủ và cực tiểu các vị Pr' từ một tập các vị từ đơn giản Pr cho trước theo quy tắc:
- ❑ Một quan hệ hoặc một mảnh “được phân hoạch thành ít nhất hai phần và chúng được truy nhập khác nhau bởi ít nhất một ứng dụng”.
- ❑ Mảnh f_i được phân hoạch theo vị từ sơ cấp $p_i \in Pr'$, qui ước là f_i của Pr'

Tóm tắt Giải thuật COM_MIN (1)

- Đầu vào: Một quan hệ R và một tập các vị từ đơn giản Pr
- Đầu ra: Một tập đầy đủ và tối thiểu các vị từ đơn giản Pr' cho Pr
- *Luật 1*: Một quan hệ hoặc một mảnh được phân chia thành ít nhất hai phần được truy cập khác nhau bởi ít nhất một ứng dụng.

- Khởi tạo :
 - Tìm một $p_i \in Pr$ sao cho p_i phân chia R theo luật 1
 - Thiết lập $Pr' = p_i$; $Pr \leftarrow Pr - p_i$; $F \leftarrow f_i$
- Lặp lại việc thêm các vị từ vào Pr' cho đến khi hoàn thành.
 - Tìm một $p_j \in Pr$ sao cho p_j phân chia một mảnh f_k nào đó được định nghĩa theo vị từ minterm trên Pr' theo luật 1
 - Thiết lập $Pr' = Pr' \cup p_j$; $Pr \leftarrow Pr - p_j$; $F \leftarrow F \cup f_j$
 - Nếu $\exists p_k \in Pr'$ không liên quan thì
 - $Pr' \leftarrow Pr' - p_k$
 - $F \leftarrow F - f_k$

Thuật toán COM_MIN

Input: R là quan hệ và Pr là tập các vị từ đơn giản

Output: Pr' là tập các vị từ đơn giản.

Khai báo: F là tập các mệnh hội sơ cấp.

Begin

 Tìm một vị từ $p_i \in Pr$ sao cho p_i phân hoạch R theo qui tắc

$Pr' \leftarrow p_i$

$Pr \leftarrow Pr - p_i$

$F \leftarrow f_i$ $\{f_i \text{ là mệnh hội sơ cấp theo } p_i\}$

 do

 begin

 Tìm một $p_j \in Pr$ sao cho p_j phân hoạch một mệnh f_k của Pr' theo qui tắc

$Pr' \leftarrow Pr' \cup p_j$

$Pr \leftarrow Pr - p_j$

$F \leftarrow F \cup f_i$

 If $\exists p_k \in Pr'$, một vị từ không có liên đới then

 begin

$Pr' \leftarrow Pr' - p_k$

$F \leftarrow F - p_k$

 end-if

 end-begin

 until Pr đầy đủ

End. { COM_MIN }

Một số nhận xét về phân mảnh ngang nguyên thủy

- ❑ Tìm tập các vị từ đơn giản đầy đủ và cực tiểu là bước đầu tiên trong thiết kế phân mảnh ngang nguyên thủy
- ❑ Thuật toán bắt đầu bằng cách tìm một vị từ có liên đới và phân hoạch quan hệ đã cho. Vòng lặp do-until thêm các vị từ vào Pr' , đảm bảo Pr' có tính cực tiểu tại mỗi bước.
- ❑ Bước tiếp theo của thiết kế phân mảnh ngang nguyên thủy là tìm tập các vị từ hội sơ cấp được định nghĩa trên các vị từ trong Pr' . Các vị từ hội sơ cấp xác định các mảnh cho cấp phát.

Một số nhận xét về phân mảnh ngang nguyên thủy

- ❑ Tuy nhiên, các vị từ hội sơ cấp có thể rất lớn, tỷ lệ hàm mũ theo số lượng các vị từ đơn giản . Vì vậy cần phải loại bỏ những mảnh không có ý nghĩa, bằng cách xác định những vị từ mâu thuẫn với tập các phép kéo theo (Implication).
- ❑ Thuật toán ***PHORIZONTAL*** tìm cách làm giảm số lượng vị từ hội sơ cấp cần được định nghĩa trong phân mảnh, bằng cách loại bỏ một số mảnh vô nghĩa. Điều này được thực hiện bằng cách xác định những vị từ mâu thuẫn dựa trên tập các liên quan tương đương I.

VD

□ Nếu $Pr' = \{p_1, p_2\}$, trong đó:

$p_1: att = value_1$

$p_2: att = value_2$

Miền biến thiên của $att = (value_1, value_2)$

□ Khi đó có hai phép liên quan tương đương với khẳng định:

$i_1: (att = value_1) \Rightarrow \neg(att = value_2)$

$i_2: \neg(att = value_1) \Rightarrow (att = value_2)$

Một số nhận xét về phân mảnh ngang nguyên thủy

- ❑ Bốn vị từ hội sơ cấp sau đây được tính theo Pr:

$$m_1: (att=value_1) \wedge (att=value_2)$$

$$m_2: (att=value_1) \wedge \neg(att=value_2)$$

$$m_3: \neg(att=value_1) \wedge (att=value_2)$$

$$m_4: \neg(att=value_1) \wedge \neg(att=value_2)$$

- ❑ Các vị từ hội sơ cấp m_1 và m_4 mâu thuẫn với các phép liên quan tương đương I và vì thế bị loại ra khỏi M. Để thực hiện bước này ta sẽ áp dụng thuật toán phân mảnh ngang nguyên thủy.

Thuật toán PHORIZONTAL

Input: R là quan hệ cần phân mảnh ngang cơ sở.

Pr là tập các vị từ đơn giản

Output: M là tập các vị từ hội sơ cấp

Begin

Pr' \leftarrow COM_MIN(R, Pr)

Xác định tập M các vị từ hội sơ cấp

Xác định tập I các phép kéo theo giữa các $p_i \in \text{Pr}'$

For mỗi $m_i \in M$ do

 If m_i mâu thuẫn với I then

 M \leftarrow M - m_i

 End_if

End_for

End. {PHORIZONTAL}

Ví dụ về phân mảnh ngang nguyên thủy - PAY

- ❑ Giả sử có một ứng dụng kiểm tra thông tin lương và xác định số lương sẽ tăng trên quan hệ PAY. Gs thông tin các nhân viên đc lưu trữ tại 2 vị trí
 - Vị trí thứ nhất $SAL \leq 30000$
 - Vị trí thứ hai: $SAL > 30000$

=> Câu truy vấn sẽ thực hiện trên cả hai vị trí. Tập vị từ đơn giản sử dụng để phân hoạch quan hệ PAY là:

$$p_1: SAL \leq 30000$$

$$p_2: SAL > 30000$$

Ví dụ về phân mảnh ngang nguyên thủy – PAY

=> Tập vị từ đơn giản khởi đầu là $Pr = \{p1, p2\}$.

=> Áp dụng COM_MIN với $i=1$ làm giá trị khởi đầu tạo ra $Pr'=\{p1\}$ là đầy đủ và cực tiểu vì $p2$ ($f1$ là mảnh hội sơ cấp được tạo ra ứng với $p1$) theo quy tắc.

=> Các vị từ hội sơ cấp sau đây là các phần tử của M :

$$m_1: (SAL \leq 30000)$$

$$m_2: \neg (SAL \leq 30000) = SAL > 30000$$

□ Khi đó, hai mảnh $F = \{PAY_1, PAY_2\}$ theo M là:

PAY ₁	
TITLE	SAL
Mech. Eng.	27000
Programmer	24000

PAY ₂	
TITLE	SAL
Elect. Eng.	40000
Syst. Anal.	34000

Ví dụ về phân mảnh ngang nguyên thủy – PROJ

Giả sử có hai ứng dụng trên quan hệ PROJ

- ❑ Ứng dụng 1: Tên và kinh phí của dự án trên ba vùng thực hiện dự án.

```
SELECT      PNAME, BUDGET
FROM        PROJ
WHERE       LOC = Value
```

Các vị từ đơn giản sử dụng cho ứng dụng này là:

p₁: LOC = “Montreal”
p₂: LOC = “New York”
p₃: LOC = “Paris”

.

Ví dụ về phân mảnh ngang nguyên thủy – PROJ

- ❑ Ứng dụng 2: Liên quan đến các dự án có kinh phí nhỏ hơn hoặc bằng 200000\$ được quản lý tại một vị trí và các dự án có kinh phí lớn hơn 200000 được quản lý tại vị trí thứ hai. Vì vậy, các vị từ đơn giản được sử dụng để phân mảnh ứng dụng thứ hai là:

p_4 : BUDGET \leq 200000

p_5 : BUDGET $>$ 200000

- ❑ Sử dụng thuật toán COM_MIN kiểm tra tập $Pr' = \{p_1, p_2, p_3, p_4, p_5\}$ là đầy đủ và cực tiểu.

☛ Các phép kéo theo của p_1, p_2, p_3, p_4, p_5 là

$$i_1 : p_1 \Rightarrow \neg p_2 \wedge \neg p_3$$

$$i_2 : p_2 \Rightarrow \neg p_1 \wedge \neg p_3$$

$$i_3 : p_3 \Rightarrow \neg p_1 \wedge \neg p_2$$

$$i_4 : p_4 \Rightarrow \neg p_5$$

$$i_5 : p_5 \Rightarrow \neg p_4$$

$$i_6 : \neg p_4 \Rightarrow p_5$$

$$i_7 : \neg p_5 \Rightarrow p_4$$

Ví dụ về phân mảnh ngang nguyên thủy – PROJ

- ❑ Tập M các vị từ hội sơ cấp tạo ra M dựa trên Pr' như sau:

$m_1: (LOC="Montreal") \wedge (BUDGET \leq 20000)$

$m_2: (LOC="Montreal") \wedge (BUDGET > 20000)$

$m_3: (LOC="New York") \wedge (BUDGET \leq 20000)$

$m_4: (LOC="New York") \wedge (BUDGET > 20000)$

$m_5: (LOC="Paris") \wedge (BUDGET \leq 20000)$

$m_6: (LOC="Paris") \wedge (BUDGET > 20000)$

- ❑ Kết quả phân mảnh ngang cơ sở PROJ tạo ra sáu mảnh $FPROJ = \{PROJ_1, PROJ_2, PROJ_3, PROJ_4, PROJ_5, PROJ_6\}$ theo các vị từ hội sơ cấp M.

Ví dụ về phân mảnh ngang nguyên thủy – PROJ

- ❑ Các mảnh PROJ₂, PROJ₅ rỗng.

PROJ₁

PNO	PNAME	BUDGET	LOC
P1	Instrumentation	150000	Montreal

PROJ₃

PNO	PNAME	BUDGET	LOC
P2	Database Develop.	135000	New York

PROJ₄

PNO	PNAME	BUDGET	LOC
P3	CAD/CAM	250000	New York

PROJ₆

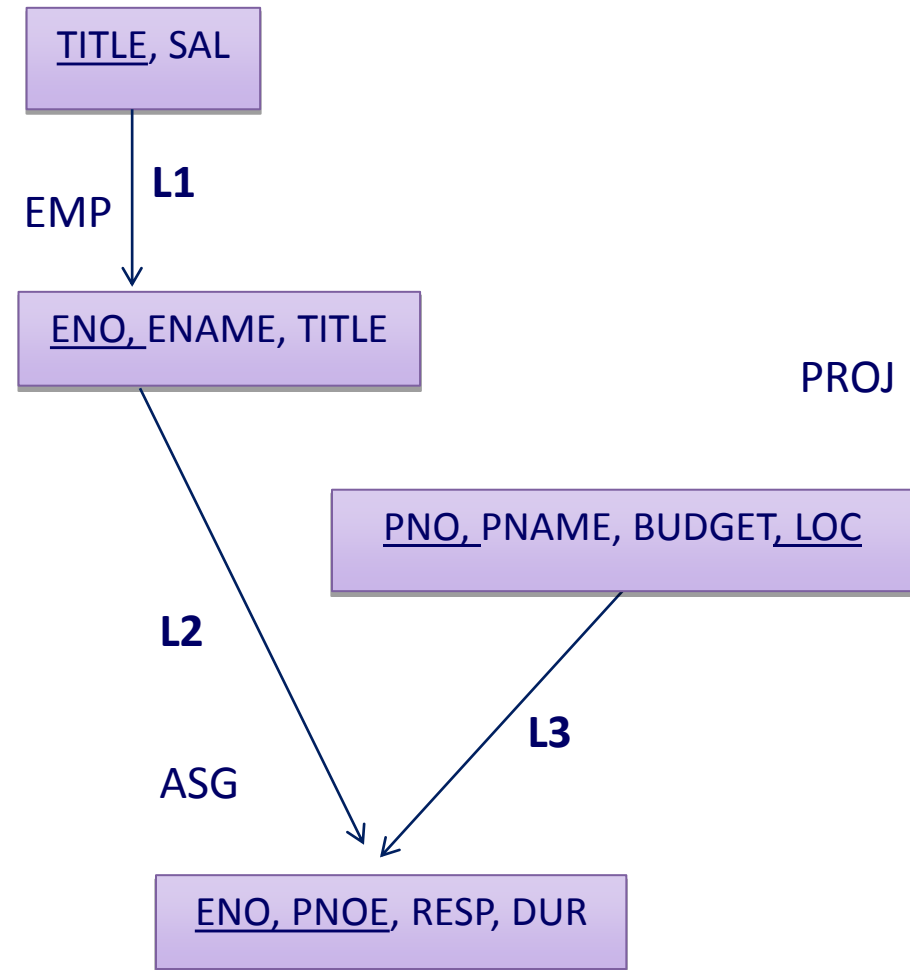
PNO	PNAME	BUDGET	LOC
P4	Maintenance	31000	Paris

Phân mảnh ngang dẫn xuất

PAY

Phân mảnh ngang dẫn xuất là việc phân mảnh một quan hệ theo kết nối bằng nhau (Equijoin) hoặc kết nối nửa bằng nhau (Semijoin) đến các quan hệ khác trong cơ sở dữ liệu. Việc quyết định chọn phân mảnh nào tối ưu hơn cần dựa trên hai tiêu chuẩn sau:

1. Phân mảnh có đặc tính kết nối tốt hơn
2. Phân mảnh sử dụng cho nhiều ứng dụng hơn



Phân mảnh ngang dẫn xuất

Ví dụ: Phân mảnh dẫn xuất quan hệ EMP theo quan hệ PAY.
Nhóm người tham gia dự án thành hai nhóm theo lương

$$SAL \leq 30000 \quad \text{và} \quad SAL > 30000$$

$$PAY_1 = \sigma_{SAL \leq 30000}(PAY)$$

$$PAY_2 = \sigma_{SAL > 30000}(PAY)$$

$$EMP_1 = EMP \triangleright \leq PAY_1$$

$$EMP_2 = EMP \triangleright \leq PAY_2$$

Phân mảnh ngang dẫn xuất

Ví dụ: Phân mảnh dẫn xuất quan hệ ASG theo quan hệ PROJ và EMP

Xét quan hệ ASG có hai ứng dụng trên nó:

Ứng dụng 1: Danh sách các kỹ sư làm việc trong các dự án tại chỗ. Ứng dụng này thực hiện trên các vị trí (thí dụ ba vị trí)

$$\text{PROJ}_1 = \sigma_{\text{LOC}=\text{"Montreal"}}(\text{PROJ})$$

$$\text{PROJ}_2 = \sigma_{\text{LOC}=\text{"New York"}}(\text{PROJ})$$

$$\text{PROJ}_3 = \sigma_{\text{LOC}=\text{"Paris"}}(\text{PROJ})$$

Phân mảnh ngang dẫn xuất

Phân mảnh dẫn xuất ASG theo các mảnh $PROJ_1$, $PROJ_2$ và $PROJ_3$ như sau:

$$ASG_1 = ASG \triangleright \leq PROJ_1$$

$$ASG_2 = ASG \triangleright \leq PROJ_2$$

$$ASG_3 = ASG \triangleright \leq PROJ_3$$

Ứng dụng 2: Tại 2 vị trí quản lý nhân viên, yêu cầu thông tin về kinh phí và thời gian thực hiện các dự án của các nhân viên.

$$EMP_1 = \sigma_{TITLE = \text{“ Syst. Anal”}} (EMP)$$

$$EMP_2 = \sigma_{TITLE = \text{“ Programmer”}} (EMP)$$

Phân mảnh ngang dẫn xuất

Phân mảnh dẫn xuất ASG theo EMP_1 và EMP_2 như sau:

$$ASG_{11} = ASG_1 \triangleright \leq EMP_1$$

$$ASG_{12} = ASG_1 \triangleright \leq EMP_2$$

$$ASG_{21} = ASG_2 \triangleright \leq EMP_1$$

$$ASG_{22} = ASG_2 \triangleright \leq EMP_2$$

$$ASG_{31} = ASG_3 \triangleright \leq EMP_1$$

$$ASG_{32} = ASG_3 \triangleright \leq EMP_2$$

Phân mảnh ngang dẫn xuất

Nhận xét

- ❑ Phân mảnh dẫn xuất có thể xảy ra dây chuyền, trong đó một quan hệ được phân mảnh như là hệ quả của một phân mảnh cho một quan hệ khác, và đến lượt nó lại làm cho các quan hệ khác phải phân mảnh (như dây chuyền PAY-EMP-ASG).
- ❑ Một quan hệ có thể có nhiều cách phân mảnh. Chọn lựa một lược đồ phân mảnh nào cho tối ưu phụ thuộc vào ứng dụng và cấp phát.

KIỂM TRA

Cho quan hệ QLSV(MA, HT, QQ,NS,GT, DT, TB)

Trong đó MA: Mã sinh viên; HT: Họ và tên sinh viên, QQ: Quê quán; NS: Năm sinh; GT: Giới tính; DT: Dân tộc; TB: Điểm trung bình.

- 1) Tập vị từ đơn giản có tính đầy đủ và tính cực tiểu
- 2) Tập các vị từ hội sơ cấp M
- 3) Phân mảnh ngang cơ sở trên tập các vị từ hội sơ cấp