

Lược đồ và các thao tác trên OLAP



Mục tiêu:

- Lược đồ kho dữ liệu
- Các thao tác cơ bản trên OLAP
- Truy vấn với Grouping sets, Roll up, Cube
- Một số hàm trên OLAP

Lược đồ kho dữ liệu

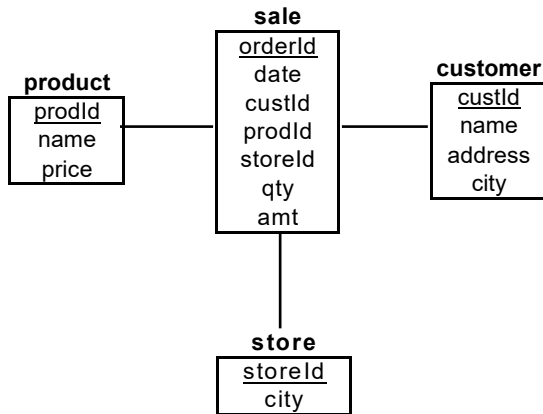


- Bảng tổng hợp
- Lược đồ hình sao và các biến thể (bông tuyết, chòm sao)
 - Quan hệ
 - Cube

Lược đồ hình sao



- Bảng dữ kiện
- Bảng các chiều
- Các độ đo:
sum, avg, max,
min, ...



Kho dữ liệu – TS. Phan Anh Phong

3

3

Lược đồ hình sao



product	prodId	name	price
	p1	bolt	10
	p2	nut	5

store	storeId	city
	c1	nyc
	c2	sfo
	c3	la

sale	orderId	date	custId	prodId	storeId	qty	amt
	o100	1/7/97	53	p1	c1	1	12
	o102	2/7/97	53	p2	c1	2	11
	105	3/8/97	111	p1	c3	5	50

customer	custId	name	address	city
	53	joe	10 main	sfo
	81	fred	12 main	sfo
	111	sally	80 willow	la

Kho dữ liệu – TS. Phan Anh Phong

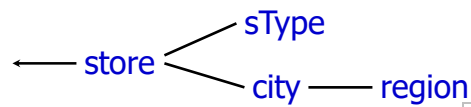
4

4

Lược đồ hình sao mở rộng



- Phân cấp chiều:



store	storeld	cityld	tlid	mgr
	s5	sfo	t1	joe
	s7	sfo	t2	fred
	s9	la	t1	nancy

sType	tlid	size	location
	t1	small	downtown
	t2	large	suburbs

city	cityld	pop	regld
	sfo	1M	north
	la	5M	south

region	regld	name
	north	cold region
	south	warm region

- Lược đồ bông tuyết (phân cấp chiều)
- Lược đồ chòm sao (có nhiều hơn 1 bảng dữ kiện)

Kho dữ liệu – TS. Phan Anh Phong

5

5

Cubes – khối đa chiều



Bảng dữ kiện:

sale	prodld	storeld	amt
	p1	c1	12
	p2	c1	11
	p1	c3	50
	p2	c2	8

Khối đa chiều:

	c1	c2	c3
p1	12		50
p2	11	8	

Số chiều = 2

Là chiều nào?

Kho dữ liệu – TS. Phan Anh Phong

6

6

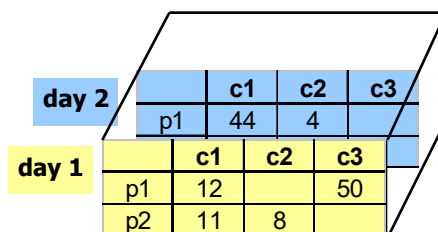
Cubes



Bảng dữ kiện:

sale	prodId	storeId	date	amt
	p1	c1	1	12
	p2	c1	1	11
	p1	c3	1	50
	p2	c2	1	8
	p1	c1	2	44
	p1	c2	2	4

Khối 3-chiều:



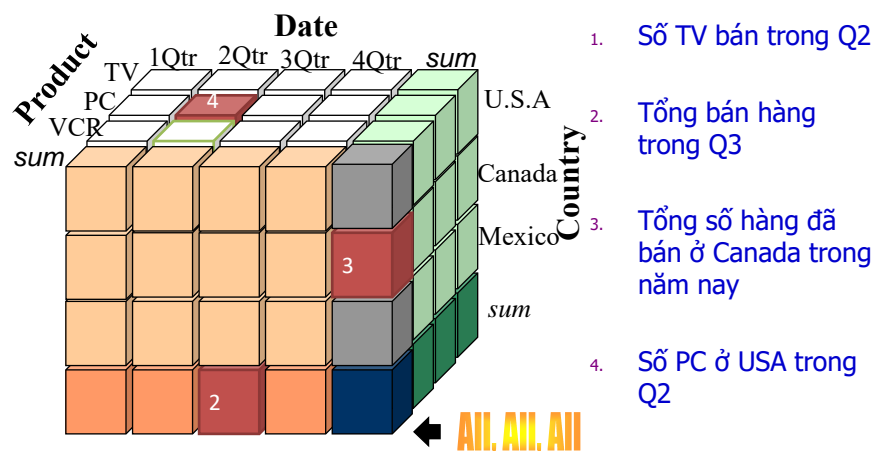
Số chiều = 3

Kho dữ liệu – TS. Phan Anh Phong

7

7

Cubes



Kho dữ liệu – TS. Phan Anh Phong

8

8

8

Các thao tác trên OLAP

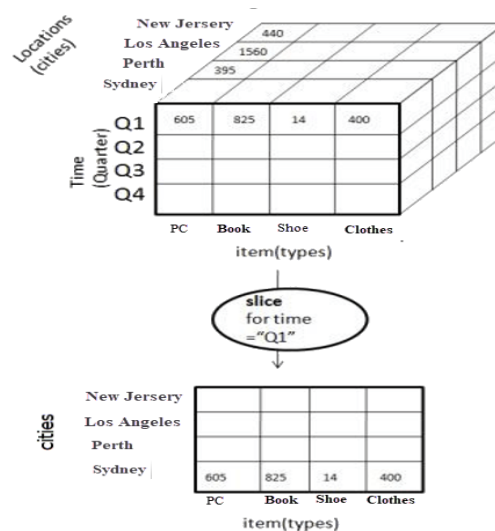


- Slice
- Dice
- Roll-up
- Drill-down
- Pivot (rotate)

Các thao tác trên OLAP



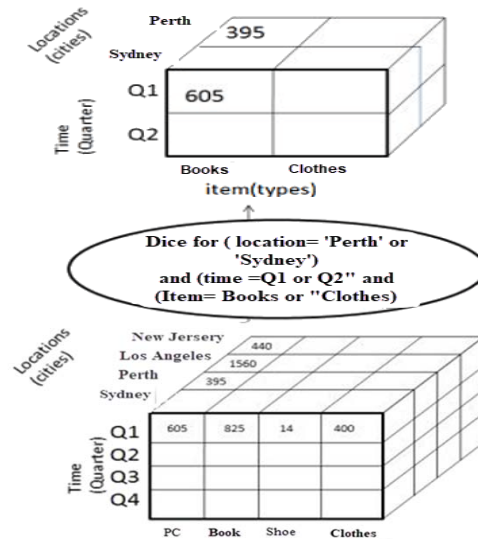
- Slice: cắt cube theo 1 chiều nào đó



Các thao tác trên OLAP



- Dice: tương tự slice nhưng cắt theo hai hay nhiều chiều



Kho dữ liệu – TS. Phan Anh Phong

11

11

11

Các thao tác trên OLAP



- Với SQL: `SELECT sum(amt) FROM SALE
WHERE date = 1`

- ?

sale	prodId	storeId	date	amt
	p1	c1	1	12
	p2	c1	1	11
	p1	c3	1	50
	p2	c2	1	8
	p1	c1	2	44
	p1	c2	2	4



81

Kho dữ liệu – TS. Phan Anh Phong

12

12

Các thao tác trên OLAP



- Với SQL: `SELECT date, sum(amt) FROM SALE GROUP BY date`
- ?

sale	prodId	storeId	date	amt
	p1	c1	1	12
	p2	c1	1	11
	p1	c3	1	50
	p2	c2	1	8
	p1	c1	2	44
	p1	c2	2	4



ans	date	sum
	1	81
	2	48

Roll up và Drill down



Với SQL: `SELECT prodId, date, sum(amt) FROM SALE GROUP BY prodId, date`

sale	prodId	storeId	date	amt
	p1	c1	1	12
	p2	c1	1	11
	p1	c3	1	50
	p2	c2	1	8
	p1	c1	2	44
	p1	c2	2	4



sale	prodId	date	amt
	p1	1	62
	p2	1	19
	p1	2	48

→ rollup →

← drill-down ←

Nhận xét qua ví dụ



a/ SELECT SUM(QTY)
FROM SP
b/ SELECT S#, SUM(QTY)
FROM SP
GROUP BY (S#)
c/ SELECT P#, SUM(QTY)
FROM SP
GROUP BY (P#) ;
d/ SELECT S#, P#, SUM(QTY)
FROM SP
GROUP BY (S#,P#)

SP		
S#	P#	QTY
S1	P1	300
S1	P2	200
S2	P1	300
S2	P2	400
S3	P2	200
S4	P2	200

- OLAP có nhiều truy vấn GROUP BY tương tự, điểm khác biệt thường nằm ở các chiều thống kê, tổng hợp
- Chi phí thực thi các truy vấn GROUP BY là lớn → để tính toán hiệu quả
- SQL 99 bổ sung các truy vấn độc lập: GROUPING SETS, ROLL UP, CUBE
...

Kho dữ liệu – TS. Phan Anh Phong

15

15

Group by grouping sets



a/ SELECT SUM(QTY)
FROM SP
b/ SELECT S#, SUM(QTY)
FROM SP
GROUP BY (S#)
c/ SELECT P#, SUM(QTY)
FROM SP
GROUP BY (P#) ;
d/ SELECT S#, P#,
SUM(QTY)
FROM SP
GROUP BY (S#,P#)



SELECT S#, P#, SUM(QTY)
FROM SP
GROUP BY GROUPING
SETS ((S#,P#), S#, P#,(,))

Kho dữ liệu – TS. Phan Anh Phong

16

16

Group by grouping sets



GROUP BY (a, b, c)
GROUP BY (a, b)
GROUP BY (a)
GROUP BY ()

Tương đương

- GROUP BY GROUPING SETS ((a,b,c),(a,b), a, ())

Group by ROLL UP



- GROUP BY ROLL UP(a, b, c) \equiv GROUP BY (a, b, c) \cup GROUP BY (a, b) \cup GROUP BY (a) \cup GROUP BY ()
- GROUP BY ROLL UP(a, b) không có tính đối xứng theo a và b
- GROUP BY ROLL UP((a, b), c) = ?

Group by CUBE



- GROUP BY CUBE(a, b, c)
 - ≡ GROUP BY(a, b, c)
 - ⊃ GROUP BY(a, b) ⊃ GROUP BY(a, c) ⊃ GROUP BY(b, c)
 - ⊃ GROUP BY(a) ⊃ GROUP BY(b) ⊃ GROUP BY(c)
 - ⊃ GROUP BY()
- CUBE(a, b, c) = ?

CUBE và GROUPING SETS trong T-SQL



- GROUP BY CUBE (C1, C2, ..., Cn-1, Cn)
 - ↕
 - GROUP BY GROUPING SETS (
 - (C1, C2, C3, ..., Cn-2, Cn-1, Cn), --n chiều
 - (C2, C3, ..., Cn-2, Cn-1, Cn),
 - (C1, C3, ..., Cn-2, Cn-1, Cn) ... , (C1, C2, C3, ..., Cn-2, Cn-1),
 - (C3, ..., Cn-2, Cn-1, Cn), (C1 ..., Cn-2, Cn-1, Cn), ... ,
 - (C1, C2), (C1, Cn) , ... , (Cn-1, Cn) , ... ,
 - (C1), (C2) , ... , (Cn-1) , (Cn) ,
 - ()
 -)

Ý nghĩa của Group by Roll up, Cube



- **ROLL UP:** Thực hiện kết nhập dữ liệu trên khối theo một trong các cách sau:
 - Giảm số chiều
 - Cuộn ngược theo phân cấp khái niệm của một chiều (chuyển từ mức thấp lên mức cao theo tính chất chỉ tiết của khái niệm)
- **CUBE:** Tổng hợp theo từng tiêu chí trên tất cả khả năng xảy ra của các chiều

PIVOT – Xoay trục dữ liệu



Student	Subject	Marks	Student	Mathematics	Science	Geography
Jacob	Mathematics	100	Jacob	100	95	90
Jacob	Science	95	Amilee	90	95	100
Jacob	Geography	90				
Amilee	Mathematics	90				
Amilee	Science	95				
Amilee	Geography	100				

Original Records

PIVOT Data

PIVOT – Xoay trục dữ liệu



```

SELECT <non-pivoted column>,
      [first pivoted column] AS <column name>,
      [second pivoted column] AS <column name>,
      ...
FROM <SELECT query that produces the data>/Table/
PIVOT
(
    <aggregation function>(<column being aggregated>)
FOR
    [<column that contains the values that will become column headers>]
    IN ( [first pivoted column], [second pivoted column],
        ... [last pivoted column])
) AS <alias for the pivot table>
    
```

Kho dữ liệu – TS. Phan Anh Phong

23

23

PIVOT – Xoay trục dữ liệu



Nhan_vien_id	ngay_ban	so_tien
XUAN	FRI	100
XUAN	MON	300
HA	SUN	400
XUAN	WED	500
XUAN	TUE	200
THU	WED	900
THU	MON	300
THU	FRI	300
HA	TUE	500
DONG	SUN	600
XUAN	FRI	900

Trung bình tiền bán của mỗi người theo các ngày trong tuần					
NV_ID	Sun	Mon	...	Fri	Sat
Xuan	?	?	?	...	
Ha	?	?	?	...	
Thu	?	?	?		
Dong	?	?			

AVG, MAX, SUM, MIN...

Kho dữ liệu – TS. Phan Anh Phong

24

24

PIVOT qua ví dụ



Nhan_vien_id	ngay_ban	so_tien
XUAN	FRI	100
XUAN	MON	300
HA	SUN	400
XUAN	WED	500
XUAN	TUE	200
THU	WED	900
THU	MON	300
THU	FRI	300
HA	TUE	500
DONG	SUN	600
XUAN	FRI	900

Trung bình tiền bán của mỗi người theo các ngày trong tuần

NV_ID	Sun	Mon	...	Fri	Sat
Xuan	?	?	?	...	
Ha	?	?	?	...	
Thu	?	?	?		
Dong	?	?			

```
SELECT Nhan_vien_id, [sun], [mon], [tue], [wed], [thu], [fri], [sat]
FROM BAN_HANG
PIVOT
(AVG(So_tien)
FOR ngay_ban in ([sun], [mon], [tue], [wed], [thu], [fri], [sat])) AS
bang_pivot
```