

## Các thao tác và truy vấn trên OLAP



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

Các thao tác và truy vấn trên OLAP – TS. Phan Anh Phong

1

1

## Đặt vấn đề



- Một số ứng dụng của OLAP
  - Báo cáo số liệu bán hàng
  - Báo cáo về quản lý
  - Báo cáo tài chính
  - Báo cáo về quá trình kinh doanh
  - Phân tích tiếp thị
  - Phân tích khách hàng và sản phẩm
  - Phân tích cung và cầu
  - Phân tích nguồn nhân lực...

Thống kê  
dữ liệu  
theo chủ  
thể/chủ đề

Các thao tác và truy vấn trên OLAP – TS. Phan Anh Phong

2

2

## Lược đồ kho dữ liệu



- Lược đồ hình sao và các mở rộng (bông tuyết, chòm sao)
  - Quan hệ
  - Cubes

Các thao tác và truy vấn trên OLAP – TS. Phan Anh Phong

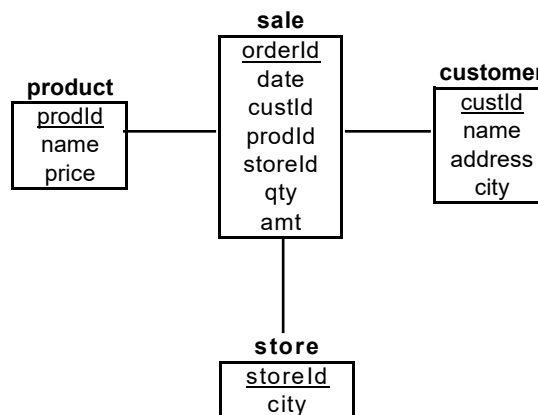
3

3

## Lược đồ hình sao



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

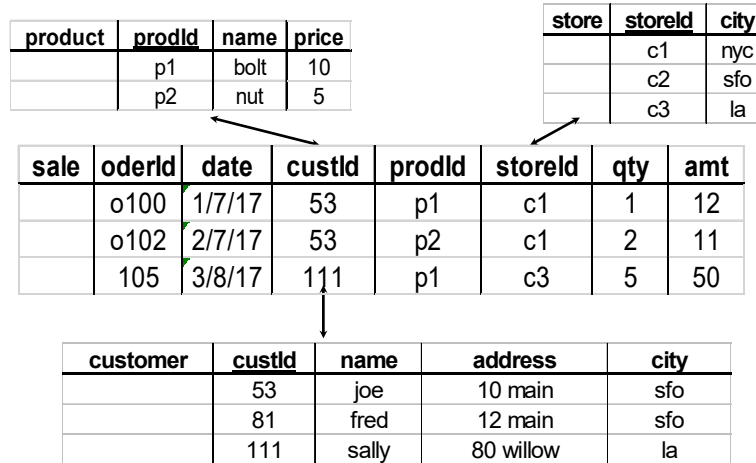


Các thao tác và truy vấn trên OLAP – TS. Phan Anh Phong

4

4

## Lược đồ hình sao



Các thao tác và truy vấn trên OLAP – TS. Phan Anh Phong

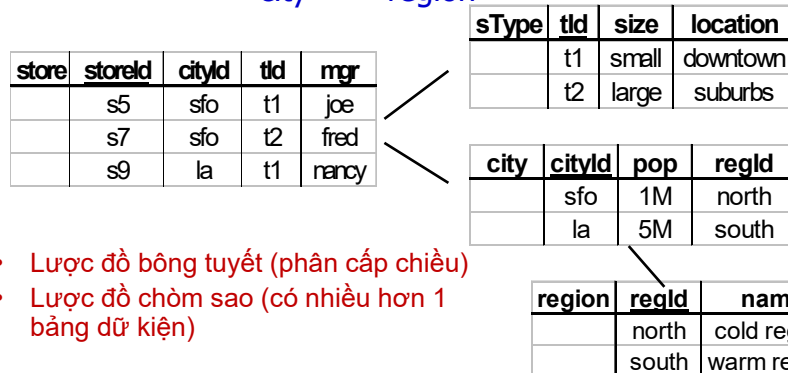
5

5

## Lược đồ hình sao mở rộng



- Phân cấp chiều:



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

Các thao tác và truy vấn trên OLAP – TS. Phan Anh Phong

6

6

## Cubes – khối đa chiều



Bảng dữ kiện:

sale	prodId	storeId	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?

Các thao tác và truy vấn trên OLAP – TS. Phan Anh Phong

7

7

## 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:

	c1	c2	c3	
day 2	p1	44	4	
day 1	p1	12		50
	p2	11	8	

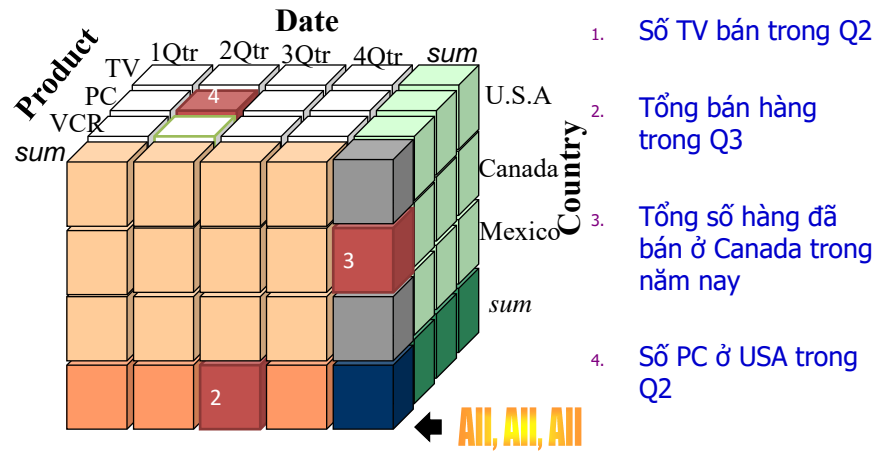
Số chiều = 3

Các thao tác và truy vấn trên OLAP – TS. Phan Anh Phong

8

8

## Cubes



Các thao tác và truy vấn trên OLAP – TS. Phan Anh Phong<sup>9</sup>

9

9

## Các thao tác trên OLAP



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

Các thao tác và truy vấn trên OLAP – TS. Phan Anh Phong

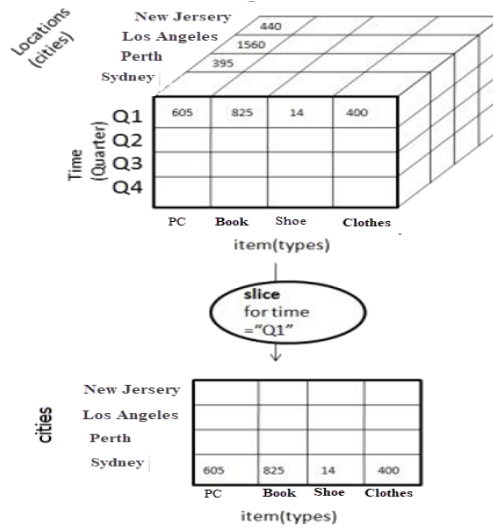
10

10

## Các thao tác trên OLAP



- Slice: cắt cube theo 1 chiều nào đó



Các thao tác và truy vấn trên OLAP – TS. Phan Anh Phong

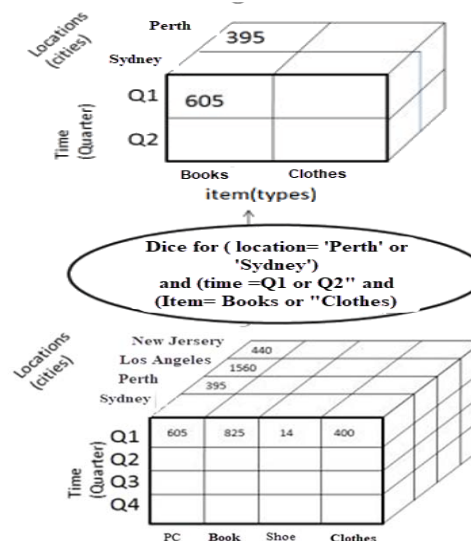
11

11

## Các thao tác trên OLAP



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



Các thao tác và truy vấn trên OLAP – TS. Phan Anh Phong

12

12

## Các thao tác trên OLAP



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

- ?

sale	prodl	storeld	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

Các thao tác và truy vấn trên OLAP – TS. Phan Anh Phong

13

13

## Các thao tác trên OLAP



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

- ?

sale	prodl	storeld	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

Các thao tác và truy vấn trên OLAP – TS. Phan Anh Phong

14

14

## 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 ←

Các thao tác và truy vấn trên OLAP – TS. Phan Anh Phong

15

15

## 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 ...

Các thao tác và truy vấn trên OLAP – TS. Phan Anh Phong

16

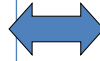
16



## 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#,( ))
```

Các thao tác và truy vấn trên OLAP – TS. Phan Anh Phong

17

17

## 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, ( ))

Các thao tác và truy vấn trên OLAP – TS. Phan Anh Phong

18

18

## Group by ROLL UP



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

## Group by CUBE



- $\text{GROUP BY CUBE}(a, b, c)$   
 $\equiv \text{GROUP BY}(a, b, c)$   
 $\cup \text{GROUP BY}(a, b) \cup \text{GROUP BY}(a, c) \cup \text{GROUP BY}(b, c)$   
 $\cup \text{GROUP BY}(a) \cup \text{GROUP BY}(b) \cup \text{GROUP BY}(c)$   
 $\cup \text{GROUP BY}()$
- $\text{CUBE}(a, b), c) = ?$

## 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) ,
  - ()
  - )

Các thao tác và truy vấn trên OLAP – TS. Phan Anh Phong

21

21

## 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

Các thao tác và truy vấn trên OLAP – TS. Phan Anh Phong

22

22

## PIVOT – Xoay trục dữ liệu



Ban\_hang

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...

Các thao tác và truy vấn trên OLAP – TS. Phan Anh Phong

23

23

## PIVOT trong T-SQL



```

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>
  
```

Các thao tác và truy vấn trên OLAP – TS. Phan Anh Phong

24

24

## PIVOT qua ví dụ



Ban\_hang

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
```

Các thao tác và truy vấn trên OLAP – TS. Phan Anh Phong

25

25

## PIVOT qua ví dụ



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

Ban\_hang

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

Nhan_vien_id	sun	mon	tue	wed	thu	fri	sat
DONG	600	NULL	NULL	NULL	NULL	NULL	NULL
HA	400	NULL	500	NULL	NULL	NULL	NULL
THU	NULL	300	NULL	900	NULL	300	NULL
XUAN	NULL	300	200	500	NULL	500	NULL

```
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
```

Các thao tác và truy vấn trên OLAP – TS. Phan Anh Phong

26

26

## Pivot – một số chú ý



- Không sử dụng nhiều hơn 1 hàm gộp trong mệnh đề PIVOT
- Khi các hàm gộp được sử dụng với PIVOT, giá trị null sẽ không được xem xét khi tính toán, tổng hợp

Các thao tác và truy vấn trên OLAP – TS. Phan Anh Phong

27

27

## Bài tập 4.1



Sales		
Item	Color	Quality
Table	Blue	11
Table	Red	22
Chair	Blue	33
Chair	Red	66

Hãy biết kết quả và giải thích ý nghĩa của các mỗi dòng trong kết quả của các truy vấn sau:

- Select item, color, sum(quality) from sales  
group by ROLLUP(item, color)
- Select item, color, sum(quality) from sales  
group by CUBE(item, color)

Các thao tác và truy vấn trên OLAP – TS. Phan Anh Phong

28

28