

PIVOT và UNPIVOT (Phần 2)

13:24 POSTED BY NTP ANALYSIS , DATABASE NO COMMENTS

Trong quá trình Pivot hoặc Unpivot ta có thể sẽ gặp thêm các trường hợp sau đây:

1. Pivot nhiều cột:

```
1 SELECT *
2 FROM cust_sales PIVOT (SUM (sales) AS SUM, COUNT (sales) AS COUNT
3 FOR month
4 IN ('Jan' AS jan, 'Feb' AS feb, 'Mar' AS mar));
```

(Câu lệnh 1)

Kết quả:

| | A | B | C | D | E | F | G | H |
|---|-------------|------------|---------|-----------|---------|-----------|---------|-----------|
| 1 | CUSTOMER_ID | PRODUCT_ID | JAN_SUM | JAN_COUNT | FEB_SUM | FEB_COUNT | MAR_SUM | MAR_COUNT |
| 2 | Kanes | Prod A | | 0 | | 0 | 39 | 1 |
| 3 | Jones | Prod A | | 0 | 92 | 1 | 34 | 1 |
| 4 | Kanes | Prod C | | 0 | 83 | 1 | | 0 |
| 5 | Kanes | Prod B | 24 | 1 | | 0 | | 0 |
| 6 | Adam | Prod B | | 0 | 49 | 1 | | 0 |
| 7 | Adam | Prod A | 34 | 1 | | 0 | | 0 |
| 8 | Adam | Prod C | | 0 | | 0 | 134 | 1 |

(Hình 1)

2. Xử lý giá trị NULL khi PIVOT:

Khi thực hiện PIVOT, với các ô không có giá trị Oracle sẽ trả về giá trị NULL. Để thay giá trị NULL bằng 1 giá trị khác ta dùng hàm **COALESCE**.

```
1 SELECT customer_id,
2 product_id,
3 COALESCE (jan, 0) jan,
```

```

4      COALESCE (feb, 0) feb,
5      COALESCE (mar, 0) mar
6  FROM cust_sales PIVOT (SUM (sales)
7                          FOR month
8                          IN ('Jan' AS jan, 'Feb' AS feb, 'Mar' AS mar));

```

(Câu lệnh 2)

Kết quả:

| | A | B | C | D | E |
|---|-------------|------------|-----|-----|-----|
| 1 | CUSTOMER_ID | PRODUCT_ID | JAN | FEB | MAR |
| 2 | Kanes | Prod A | 0 | 0 | 39 |
| 3 | Jones | Prod A | 0 | 92 | 34 |
| 4 | Kanes | Prod C | 0 | 83 | 0 |
| 5 | Kanes | Prod B | 24 | 0 | 0 |
| 6 | Adam | Prod B | 0 | 49 | 0 |
| 7 | Adam | Prod A | 34 | 0 | 0 |
| 8 | Adam | Prod C | 0 | 0 | 134 |

(Hình 2)

3. Hạn chế của Pivot:

Chúng ta thấy rằng khi thực hiện Pivot thì pivot_in_clause bắt buộc chúng ta phải nhập tay các giá trị chúng ta muốn pivot. Như ví dụ trên giả sử có thêm tháng April thì chúng ta phải sửa lại mệnh đề pivot_in_clause. Vậy có cách nào để ta lấy tự động khi có các tháng mới phát sinh không?

Oracle có hỗ trợ cho ta cách lấy động này nhưng khi đó kết quả trả về của các cột được PIVOT sẽ thể hiện trong 1 cột và có định dạng là XML.

```

1 SELECT *
2 FROM cust_sales PIVOT XML (SUM (sales)
3                             FOR month
4                             IN (SELECT DISTINCT month FROM cust_sales));

```

(Câu lệnh 3)

Kết quả:

| CUSTOMER_ID | PRODUCT_ID | MONTH_XML |
|-------------|------------|--|
| | | <PivotSet><item><column name = "MONTH">Feb</column><column name = "SUM(SALES)"></column></item><item><column name = "MONTH">Jan</column> |

| | | |
|-------|--------|--|
| Adam | Prod A | <column name = "SUM(SALES)">34</column></item><item><column name = "MONTH">Mar</column><column name = "SUM(SALES)"></column></item></PivotSet> |
| Adam | Prod B | <PivotSet><item><column name = "MONTH">Feb</column><column name = "SUM(SALES)">49</column></item><item><column name = "MONTH">Jan</column><column name = "SUM(SALES)"></column></item><item><column name = "MONTH">Mar</column><column name = "SUM(SALES)"></column></item></PivotSet> |
| Adam | Prod C | <PivotSet><item><column name = "MONTH">Feb</column><column name = "SUM(SALES)"></column></item><item><column name = "MONTH">Jan</column><column name = "SUM(SALES)"></column></item><item><column name = "MONTH">Mar</column><column name = "SUM(SALES)">134</column></item></PivotSet> |
| Jones | Prod A | <PivotSet><item><column name = "MONTH">Feb</column><column name = "SUM(SALES)">92</column></item><item><column name = "MONTH">Jan</column><column name = "SUM(SALES)"></column></item><item><column name = "MONTH">Mar</column><column name = "SUM(SALES)">34</column></item></PivotSet> |
| Kanes | Prod A | <PivotSet><item><column name = "MONTH">Feb</column><column name = "SUM(SALES)"></column></item><item><column name = "MONTH">Jan</column><column name = "SUM(SALES)"></column></item><item><column name = "MONTH">Mar</column><column name = "SUM(SALES)">39</column></item></PivotSet> |
| Kanes | Prod B | <PivotSet><item><column name = "MONTH">Feb</column><column name = "SUM(SALES)"></column></item><item><column name = "MONTH">Jan</column><column name = "SUM(SALES)">24</column></item><item><column name = "MONTH">Mar</column><column name = "SUM(SALES)"></column></item></PivotSet> |
| Kanes | Prod C | <PivotSet><item><column name = "MONTH">Feb</column><column name = "SUM(SALES)">83</column></item><item><column name = "MONTH">Jan</column><column name = "SUM(SALES)"></column></item><item><column name = "MONTH">Mar</column><column name = "SUM(SALES)"></column></item></PivotSet> |

(Bảng kết quả 1)

Ta có thể thay câu lệnh **SELECT DISTINCT month FROM cust_sales** bằng từ khóa ANY.

Với kết quả nhận được của cột MONTH_XML ta có thể xử lý thêm 1 phần nữa để lấy được kết quả mong muốn.

Phần tiếp theo là phần bổ sung để xử lý dữ liệu định dạng XML khi lấy được ở trên. Đây là dữ liệu XML của dòng đầu tiên:

```
1 <PivotSet>
2  <item>
```

```

3  <column name = "CUSTOMER_ID">Adam</column>
4  <column name = "SUM(SALES)"/>
5  </item>
6  <item>
7    <column name = "CUSTOMER_ID">Jones</column>
8    <column name = "SUM(SALES)">92</column>
9  </item>
10 <item>
11   <column name = "CUSTOMER_ID">Kanes</column>
12   <column name = "SUM(SALES)"/>
13 </item>
14</PivotSet>

```

Nhìn vào dữ liệu XML trên, ta thấy rằng có thể duyệt qua các phần tử bằng cách dùng hàm **EXTRACTVALUE** kết hợp với cấu trúc XPATH.

```

1 SELECT month,
2        product_id,
3        EXTRACTVALUE (customer_id_xml, '/PivotSet/item[1]/column[2]') person1,
4        EXTRACTVALUE (customer_id_xml, '/PivotSet/item[2]/column[2]') person2,
5        EXTRACTVALUE (customer_id_xml, '/PivotSet/item[3]/column[2]') person3
6 FROM v_cust_sales t;

```

(Câu lệnh 4)

Kết quả:

| | A | B | C | D | E |
|---|-------|------------|---------|---------|---------|
| 1 | MONTH | PRODUCT_ID | PERSON1 | PERSON2 | PERSON3 |
| 2 | Feb | Prod A | | 92 | |
| 3 | Feb | Prod B | 49 | | |
| 4 | Feb | Prod C | | | 83 |
| 5 | Jan | Prod A | 34 | | |
| 6 | Jan | Prod B | | | 24 |
| 7 | Mar | Prod A | | 34 | 39 |
| 8 | Mar | Prod C | 134 | | |

(Hình 3)

Ta thấy rằng với kết quả trên thì các cột PERSON1, PERSON2 và PERSON3 vẫn không xác định được là của Customer_ID nào. Để xác định được các cột đó là của Person nào ta có thể dùng XPATH lấy column 1 và chỉ lấy 1 dòng đầu tiên (ROWNUM=1) vì các dòng còn lại đều cho kết quả như nhau.

| | A | B | C | D | E |
|---|---------|--------------|------|-------|-------|
| 1 | 'MONTH' | 'PRODUCT_ID' | P1 | P2 | P3 |
| 2 | Month | Product_id | Adam | Jones | Kanes |
| 3 | Feb | Prod A | | 92 | |
| 4 | Feb | Prod B | 49 | | |
| 5 | Feb | Prod C | | | 83 |
| 6 | Jan | Prod A | 34 | | |
| 7 | Jan | Prod B | | | 24 |
| 8 | Mar | Prod A | | 34 | 39 |
| 9 | Mar | Prod C | 134 | | |

(Hình 4)

4. Sắp xếp kết quả trả về khi UNPIVOT:

Ta xét ví dụ sau:

```

1 SELECT
2   FROM cust_sales_month UNPIVOT (yes
3                                   FOR month
4                                   IN("JANUARY", "FEBRUARY", "MARCH"))
5 ORDER BY customer_id, product_id;
```

(Câu lệnh 5)

| CUSTOMER_ID | PRODUCT_ID | MONTH | YES |
|-------------|------------|----------|-----|
| Adam | Prod A | JANUARY | 1 |
| Adam | Prod A | FEBRUARY | 0 |
| Adam | Prod A | MARCH | 0 |
| Adam | Prod B | FEBRUARY | 1 |
| Adam | Prod B | JANUARY | 0 |
| Adam | Prod B | MARCH | 0 |
| Adam | Prod C | FEBRUARY | 0 |
| Adam | Prod C | JANUARY | 0 |
| Adam | Prod C | MARCH | 1 |

| | | | |
|-------|--------|----------|---|
| Jones | Prod A | FEBRUARY | 1 |
| Jones | Prod A | JANUARY | 0 |
| Jones | Prod A | MARCH | 1 |
| Kanes | Prod A | MARCH | 1 |
| Kanes | Prod A | JANUARY | 0 |
| Kanes | Prod A | FEBRUARY | 0 |
| Kanes | Prod B | FEBRUARY | 0 |
| Kanes | Prod B | JANUARY | 1 |
| Kanes | Prod B | MARCH | 0 |
| Kanes | Prod C | MARCH | 0 |
| Kanes | Prod C | FEBRUARY | 1 |
| Kanes | Prod C | JANUARY | 0 |

(Bảng kết quả 2)

Với kết quả trên ta thấy rằng cột MONTH sau khi Unpivot thì không được sắp xếp theo thứ tự từng tháng, để thực hiện việc sắp xếp cho cột MONTH ta thực hiện như sau:

```

1 SELECT *
2 FROM cust_sales_month UNPIVOT (yes
3                                FOR month
4                                IN("JANUARY", "FEBRUARY", "MARCH"))
5 ORDER BY customer_id,
6          product_id,
7          DECODE (month, 'JANUARY', 1, 'FEBRUARY', 2, 'MARCH', 3);

```

(Câu lệnh 6)

| CUSTOMER_ID | PRODUCT_ID | MONTH | YES |
|-------------|------------|----------|-----|
| Adam | Prod A | JANUARY | 1 |
| Adam | Prod A | FEBRUARY | 0 |
| Adam | Prod A | MARCH | 0 |
| Adam | Prod B | JANUARY | 0 |
| Adam | Prod B | FEBRUARY | 1 |
| Adam | Prod B | MARCH | 0 |
| Adam | Prod C | JANUARY | 0 |
| Adam | Prod C | FEBRUARY | 0 |
| | | | |

(Bảng kết quả 3)