

مقدمه (شرح مساله)

چندی پیش در تالار T-SQL سوالی مطرح شد راجع به مساله ای که معروف است به top N per group. تنها موضوعی که باعث شد من مطلبی راجع به آن بنویسم محدودیتی بود که کاربر مورد نظر داشت؛ که آن محدودیت چیزی نبود جز: query بایستی در نسخه 2000 جوابگو باشد.

قطعا شده است که بخواهید مثلا به ازای هر مشتری آخرین سفارش آن را انتخاب کنید. این مساله Top N نامیده می‌شود.

فرض کنید جدولی داریم که حاوی سفارشات مشتریان می‌باشد. هر مشتری می‌تواند چندین سفارش داشته باشد؛ هر سفارش دارای حداقل دو مقدار "تاریخ سفارش" و "مبلغ سفارش است". هدف پیدا کردن آخرین سفارشات هر مشتری می‌باشد. نکته: اگر چند تاریخ برای آخرین سفارش مشتری وجود داشت آنگاه بایستی بر اساس مبلغ سفارش مرتب سازی نزولی صورت بگیرد. یا به عبارت دیگر ابتدا باید مرتب سازی نزولی بر اساس ستون تاریخ سفارش انجام شود و سپس مرتب سازی نزولی بر اساس ستون مبلغ سفارش.

فرض می‌گیریم داده‌های جدول ما چیزیست شبیه به این:

Row_id	Customer_id	Order_date	Order_value
6	1	10	15
9	1	10	5
4	1	9	10
3	1	8	20
10	2	12	15
7	2	10	15
11	3	15	5
5	3	10	20
8	3	10	10
1	3	5	15
2	3	5	15

سطرهایی از جدول که رنگی شده اند سطرهای مورد نظر ما هستند که باید در خروجی ظاهر شوند. داده‌های جدول با کمک قابلیت Sort نرم افزار word مرتب سازی شده اند، این تصویر را به این خاطر در اینجا قرار دادم چون که دیدم می‌تواند در شفاف سازی مساله به من کمک کند. ابتدا مرتب سازی نزولی بر اساس ستون order_date انجام گرفته و سپس مرتب سازی نزولی بر اساس ستون order_value. و در پایان اولین سطر مربوط به هر مشتری به عنوان خروجی مورد نظر انتخاب می‌شوند.

The screenshot shows the 'Sort' dialog box in SQL Server. It has three sections for sorting criteria:

- Sort by:** 'Customer_id' (Type: Number, Using: Paragraphs). The 'Ascending' radio button is selected.
- Then by:** 'Order_date' (Type: Number, Using: Paragraphs). The 'Descending' radio button is selected.
- Then by:** 'Order_value' (Type: Number, Using: Paragraphs). The 'Descending' radio button is selected.

At the bottom, the 'My list has' section has 'Header row' selected. There are 'Options...', 'OK', and 'Cancel' buttons.

راه حل ها

خب پر واضح است که در نسخه 2005 و بعد از آن ساده ترین و بهینه ترین راه حل استفاده از تابع row_number می باشد.

```
SELECT row_id, customer_id, order_date, order_value
FROM (SELECT *,
      ROW_NUMBER() OVER(PARTITION BY customer_id
                        ORDER BY order_date DESC, order_value DESC) AS rnk
FROM table_name
)t
WHERE rnk = 1;
```

اما با محدودیتی که در نسخه 2000 وجود دارد راه حلی بهتر از این پیدا نخواهیم کرد:

```
SELECT *
FROM table_name t
WHERE row_id = (SELECT TOP 1 row_id
               FROM table_name
               WHERE customer_id = t.customer_id
               ORDER BY order_date DESC, order_value DESC);
```

حالا چه میشود راه حلی بخواهیم مستقل از هر یک از نسخه های SQL Server:

```
SELECT MIN(row_id) AS row_id, customer_id, order_date, order_value
FROM table_name t
WHERE order_date =
```

```
(SELECT MAX(order_date)
FROM table_name
WHERE customer_id = t.customer_id)
AND order_value =
(SELECT MAX(order_value)
FROM table_name
WHERE customer_id = t.customer_id
AND order_date =
(SELECT MAX(order_date)
FROM table_name
WHERE customer_id = t.customer_id))
GROUP BY customer_id, order_date, order_value;
```