

عنوان: توابع Window و مساله های آماری running total و running average

نویسنده: محمد سلیم آبادی

تاریخ: ۹:۵۱۳۹۱/۱۱/۱۵

آدرس: www.dotnettips.info

برچسب‌ها: sql server 2005, window function, running total, running average

مقدمه و شرح مساله

توسط ویژگی‌های جدیدی که در نسخه 2012 به بحث window افزوده شد می‌توانیم مساله‌های running total و running average را به شکل بهینه‌ای حل کنیم.

ابتدا این دو مساله را بدون بکارگیری ویژگی‌های جدید، حل نموده و سپس سراغ توابع جدید خواهیم رفت.

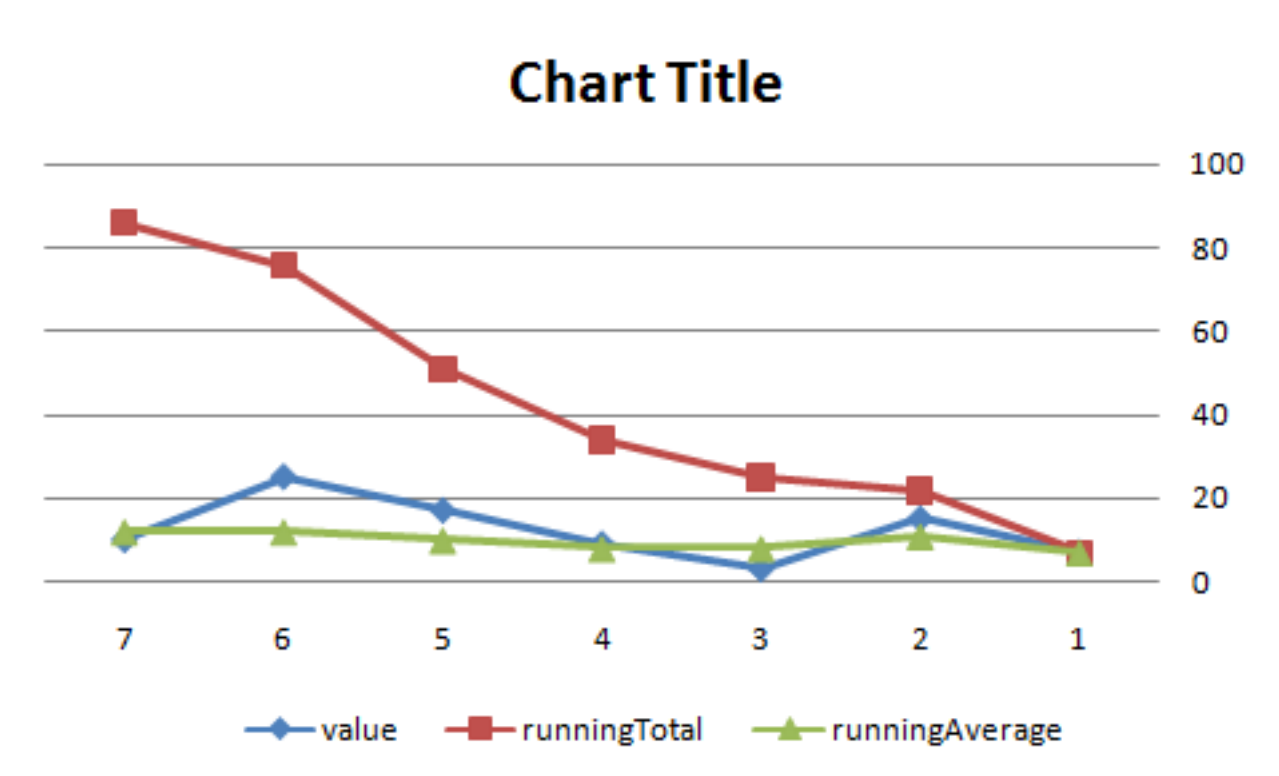
قبل از هر چیزی لازم است جدول زیر ساخته شود و داده‌های نمونه در آن درج شود:

```
create table testTable
(
    day_nbr integer not null primary key clustered,
    value integer not null check (value > 0)
);
insert into testTable
values (10, 7), (20, 15), (30, 3), (40, 9), (50, 17), (60, 25), (70, 10);
```

مساله running total بسیار ساده است، یعنی جمع مقدار سطر جاری با مقادیر سطرهای قبلی (بر اساس یک ترتیب معین) running average هم مشابه به running total هست با این تفاوت که میانگین مقادیر سطر جاری و سطرهای قبلی محاسبه می‌شود.

| | day_nbr | value | runningTotal | runningAverage |
|---|---------|-------|--------------|----------------|
| 1 | 10 | 7 | 7 | 7 |
| 2 | 20 | 15 | 22 | 11 |
| 3 | 30 | 3 | 25 | 8 |
| 4 | 40 | 9 | 34 | 8 |
| 5 | 50 | 17 | 51 | 10 |
| 6 | 60 | 25 | 76 | 12 |
| 7 | 70 | 10 | 86 | 12 |

و نتیجه به صورت نمودار:

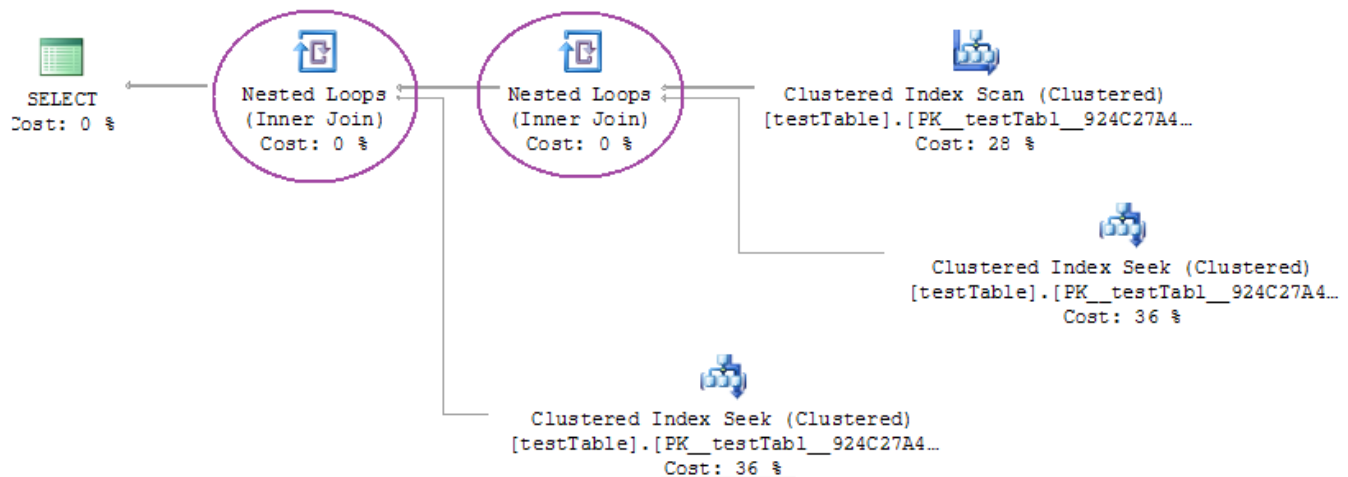


راه حل در SQL Server 2000

توسط دو correlated scalar subquery می‌توانیم مقادیر دو ستون مورد نظر با محاسبه کنیم:

```
select *,
    runningTotal = (select sum(value)
                    from testTable
                    where day_nbr <= t.day_nbr),
    runningAverage = (select avg(value)
                     from testTable
                     where day_nbr <= t.day_nbr)
from testTable t;
```

اگر به نقشه اجرای این query نگاه کنید گره (عملگر) inner join دو بار بکار رفته است (به وجود دو subquery)، که این عدد در روش توابع تجمعی window به صفر کاهش پیدا خواهد کرد

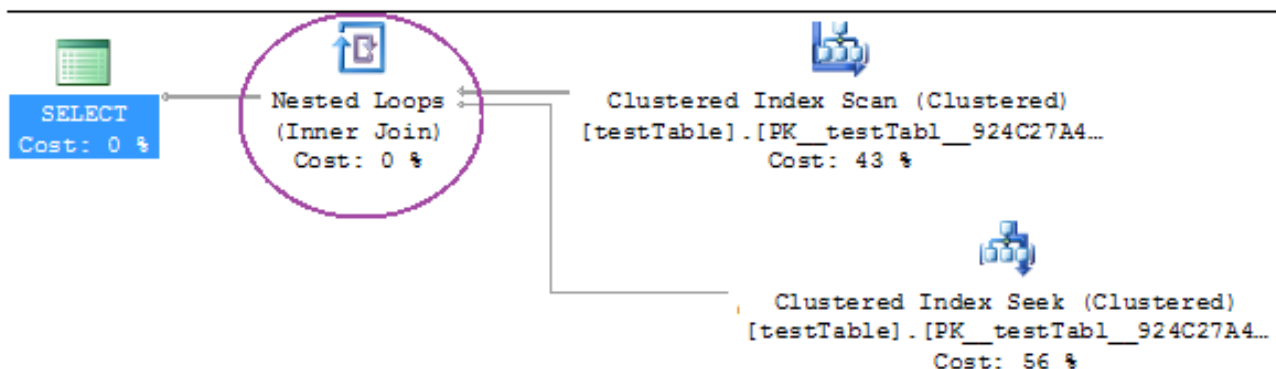


راه حل در SQL Server 2005

توسط cross apply به سادگی می توانیم دو subquery که در روش قبل بود را به یکی کاهش دهیم:

```
select *
from testTable t
cross apply (select sum(value) as runningTotal,
                avg(value) as runningAverage
            from testTable
            where day_nbr <= t.day_nbr)d;
```

این بار تنها یک عملگر inner join در نقشه اجرای query مشاهده می شود:



راه حل در SQL Server 2012

با اضافه شدن برخی از ویژگی های استاندارد به ماده OVER مثل rows و range شاهد بهبودی در عملکرد query هستیم.

یکی از کاربردهای توابع تجمعی window حل مساله running total و running average است.

به تصویر زیر توجه کنید، همانطور که در قبل توضیح دادم ما به سطر جاری و سطرها پیشین نیاز داریم تا اعمال تجمعی (جمع و میانگین) را روی مقادیر بدست آمده انجام دهیم. در تصویر زیر سطر جاری و سطرها قبلی به ازای هر سطر به وضوح قابل مشاهده است، مثلاً هنگامی که سطر جاری برابر با روز 30 است ما خود سطر جاری (current row) و تمام سطرها پیشین و قبلی (unbounded preceding) را نیاز داریم.

| Day | value |
|-----|-------|
| 10 | 7 |
| 20 | 15 |
| 30 | 3 |
| 40 | 9 |
| 50 | 17 |
| 60 | 25 |
| 70 | 10 |

Unbounded preceding ←

Current row →

و اکنون query مورد نظر

```
select *, sum(value) over(order by day_nbr rows between unbounded preceding and current row) as
runningTotal,
avg(value) over(order by day_nbr rows between unbounded preceding and current row) as
runningAverage
from testTable
```

در نقشه اجرای این query دیگر خبری از عملگر inner join نخواهد بود که به معنای عملکرد بهتر query است.

