Top and the UNION Query

Posted At : August 25, 2006 9:10 AM | Posted By : Mark Kruger
Related Categories: MS SQL Server, Coldfusion & Databases

A "UNION" query can be very useful. Take sorting for example. If I want to select 2 types of data from the same table using different conditions and order *by my conditions* union is a great tool. See my previous post on **Using UNION in your queries**. If you also like using TOP you may run into an unexpected behavior when using UNION. Let's say you had 2 tables, portalUsers and intranetUsers and you wanted to get the most recent entries in each that were unapproved. If you try something like this:

```
<cfquery name="myquery" datasource="mydsn">
    SELECT    TOP 1 fname, lname,
            'Portal' AS sourceTable,
            dateAdded
    FROM    portalUsers
    WHERE   unApproved = 1
  UNION
    SELECT TOP 1 fname, lname,
            'Intranet' AS sourceTable,
            dateAdded
    FROM   intranetUsers
    WHERE   Approved = 0
  ORDER BY dateAdded
</cfquery>
```

You might expect to get 2 records equal to the most recently added portal user and the most recently added intranet user - right? You could take the first record and you would be good to go. Actually, unless the "portalUsers" table has a clustered index ordered by dateAdded, this is *not the case*.

The reason is simple, the condition of top is met for both of the queries before the sort order is applied. If you think about it, this makes sense. To do it any other way would require the order by clause to be run three times - once for each query and then once for the returning result set. Instead, the UNION clause goes out and gets the data for query 1, then the data for query 2 and *then* sorts the data. The result is that you will not be able to depend on getting the records you want from this UNION query.

The Fix: Use a View

One fix would be to use a view. Simply create a view without the TOP key word - like so:

```
<cfquery name="myquery" datasource="mydsn">
CREATE VIEW vwUnApprovedUsers
AS
    SELECT    fname, lname,
            'Portal' AS sourceTable,
            dateAdded
    FROM    portalUsers
    WHERE   unApproved = 1
  UNION
    SELECT fname, lname,
            'Intranet' AS sourceTable,
            dateAdded
    FROM   intranetUsers
```

```
      WHERE    Approved = 0
   ORDER BY dateAdded
</cfquery>
```

 You could then select directly from this view using TOP 1 and get the most recent
record.

The Fix: An "In-line" View?

Still another way is to use an "inline view" and select *from* a subselect using top 1.
That would look something like this:

```
<cfquery name="myquery" datasource="mydsn">
   SELECT TOP 1 *
   FROM
   ( SELECT     fname, lname,
            'Portal' AS sourceTable,
            dateAdded
     FROM    portalUsers
     WHERE   unApproved = 1
   UNION
     SELECT fname, lname,
            'Intranet' AS sourceTable,
            dateAdded
     FROM   intranetUsers
     WHERE   Approved = 0
     ) AS inLineView
   ORDER BY dateAdded
</cfquery>
```

 Ahah! This creates a dataset and then allows you to pull out only the first record.
Make sure and test for performance to see if one or the other of these methods works
better. NOTE: The subselect alias is required ("AS inLineView"). If you do not alias the
subselect the query will error out - but there will be little indication in the error why
it's throwing an exception.