

Marc Lucas (Con't)

The success of the warehouse project resulted in presentations at a local Saskatoon SUG meeting, Canadian (CIRPA) and U.S.A. (AIR) research association conferences and at SUGI 30.

I have helped re-establish the Saskatoon SAS users group and have presented at each meeting since 2002. I am a firm believer in the power and flexibility of SAS, and often demonstrate SAS products/features to other University of Saskatchewan staff to help them get up and running with SAS.

Partner/Family:

I married Lise 14 years ago, and we are currently busy raising three energetic boys aged 7, 8 and 10. Lise is from the ethnically rich French community of Ponteix, Saskatchewan, near Gravelbourg. We are keeping her family's French heritage alive through the French Immersion Program in Saskatoon.

Pets:

Goldfish, although half of the family would like a dog.

Sports/Hobbies:

Camping, fishing, hunting, soccer, coaching the boys' soccer teams and – when I get a chance – reading.

What your ideal weekend would be:

Fishing, boating and sitting around a fire with friends and family at an isolated northern lake.

If I could be anything at all (besides a SAS programmer), I would be....

A retired Powerball lottery winner. ☺

When I'm not programming in SAS, I like to

Enjoy one of my sports/hobbies or just spend time with friends and family.

SAS Techie Tip:

Using PROC SQL, Nested Queries and SQL functions to update/modify fields

PROC SQL is very powerful tool that, once understood, can be a powerful addition to every SAS programmer's toolset.

The basic structure of the PROC UPDATE SQL command is

```
PROC SQL;  
  Update libname.dataset1 as d1
```

Set

```
Field1 = (select fieldname from libname.dataset2  
Where key1 = d1.key1);
```

Quit;

Two important considerations:

- This form of the query works great if the relationship between dataset1 and dataset2 is a 0 to 1 relationship. In other words, if the subquery finds more than one row in dataset 2, the update query will fail.
- If the subquery does not find at least one row in dataset2, the field in dataset 1 will be set to NULL (missing).

The issue of a 0 to Many relationship between dataset1 and dataset2 can be solved (possibly) though the use of a SQL function like DISTINCT, MIN or MAX. For example, say that dataset2 has multiple rows that match the fields that you are joining on and that the rows in table 2 are time stamped. You could then pick either the earliest or the latest value from table two using the following structure:

PROC SQL;

```
Update libname.dataset1 as d1  
Set field = (select field2 from libname.dataset2  
Where key1 = d1.key1  
and key2 = (select max(key2) from libname.dataset2  
where key1 = d1.key1)  
);
```

Quit;

Additional notes:

- The MIN/MAX functions appear to work on Text values. You can make use of the distinct function to help weed out duplicates. Performance on queries can be improved by defining appropriate indexes on libname.dataset2.