

# The Query Builder: The Swiss Army Knife of SAS® Enterprise Guide®

Jennifer First-Kluge and Steven First, Systems Seminar Consultants, Inc.

## ABSTRACT

The SAS® Enterprise Guide® Query Builder is one of the most powerful components of the software. It enables a user to bring in data, join, drop and add columns, compute new columns, sort, filter data, leverage the advanced expression builder, change column attributes, and more! This presentation provides an overview of the major features of this powerful tool and how to leverage it every day.

## WHAT IS THE QUERY BUILDER?

Enterprise Guide is a user interface for SAS. EG passes SAS code to a server to run the SAS code (whether you are writing code or using the tasks in EG). The server could be your local machine, a server on your network, it could even be on another platform. A task is a point and click user interface in Enterprise Guide that generates SAS code. Each task has an analogous PROC (although not 1 to 1 relationship).

One of the most powerful tasks is the Query Builder. The Query Builder allows you to manipulate data tables in a variety of ways, including joining tables, selecting variables, filtering data, sorting data, changing data sources, computing columns, setting up prompts, de-duping observations, adding titles and footnotes, limiting output, changing query options, and more.

Quite simply, the query builder will quickly bring together data (joining up to 32 tables per query). It will allow the user to manipulate and modify the data easily. These queries will be quicker and less error prone than coding from scratch.

## SQL IN THE QUERY BUILDER

SQL is a query language used by many software packages. The Query Builder in Enterprise Guide generates PROC SQL code behind the scenes. The user doesn't need to code SQL because the query builder will do the heavy lifting.

Example of SQL code in SAS:

```
proc sql options;
  create table|view as          /* output table to create */
  select column(s)             /* Select/create columns */
  from table-name | view-name /* Name input sources */
  where expression             /* Sub-set rows from table*/
  group by column(s)          /* Group rows for summary */
  having expression            /* Subset GROUP BY results*/
  order by column(s)          /* Sort resulting rows */
  ;
quit;
```

An EG user can take the shell SQL code that EG generates and modify with special options, etc.

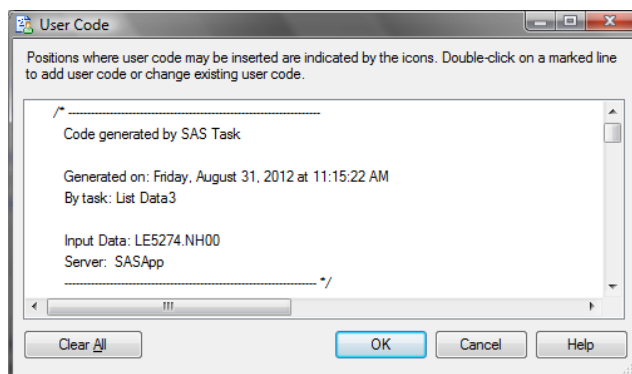





Figure 1: SQL Code From the Query Builder

## CREATING A NEW QUERY

To create a new query in Enterprise Guide (EG), in the Process Flow or Project Window, highlight a data set. On the tasks menu, select 'Query Builder'. You can also do this by right-clicking on the data.

The Tables list shows the columns that make up query table. The Select Data tab lists the columns included in the query (the user must add the columns to the Select Data tab to have them show up on the output. These columns can be added to the Select Data tab by double clicking the columns or selecting them over to the Select Data tab. The  and  buttons move columns up/down. The  button deletes columns.

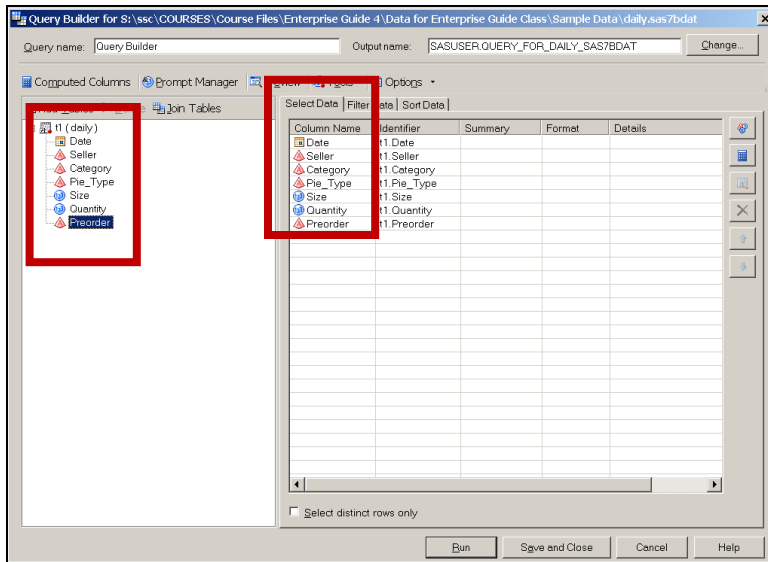


Figure 2: Selecting Columns in the Query Builder

## QUERY RESULTS

There are three formats for query results/output: data table, data view, or report. A Data Table is a static table of values. It is not updated unless the query is rerun. It can also have tasks run against it. A data view is a dynamic table. It stores logic to carry out on data rather than the data itself. It is updated as the source data changes. It can also have tasks run against it. A Report is a static HTML document. It is not updated unless the query is rerun. It cannot have tasks run against it.

To globally select the query results format, go to the Tools menu, select Options, Query. Select save query result set as, Select Data Table, Data View or Report.

After the query runs, objects for the query and its results appear in the Process Flow and Project Tree. The original data set has not changed.

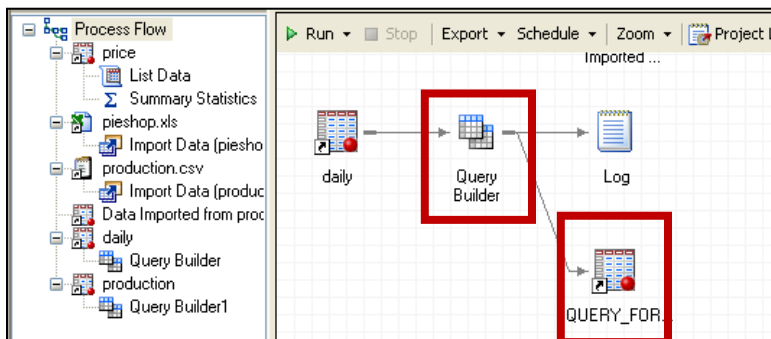


Figure 3: Query Results

The user can set the output location (a file location, server, or library) of the query, under the options. The default setting is to save the output in a temporary library. This means that the output will have to be recreated each time that Enterprise Guide is closed and reopened.

Tasks may be run on query results (if the results are a data table or data view). To run a task on query results, in the Process Flow or Project Tree, highlight the query results. From the Task List, select the Query Builder and proceed as usual.

## SORTING IN THE QUERY BUILDER

Sorting changes the order of the observations in a table by ranking the observations based on the values of one or more variables. This can be done on the Sort Data tab in the Query Builder by dragging columns onto the Sort Data tab. The user can specify an ascending or descending sort.

Queries can also be sorted by multiple variables. All of the sort variables should be in the Sort Data Tab. Clicking the 'Up' and 'Down' buttons will arrange the variables in order of sort priority.

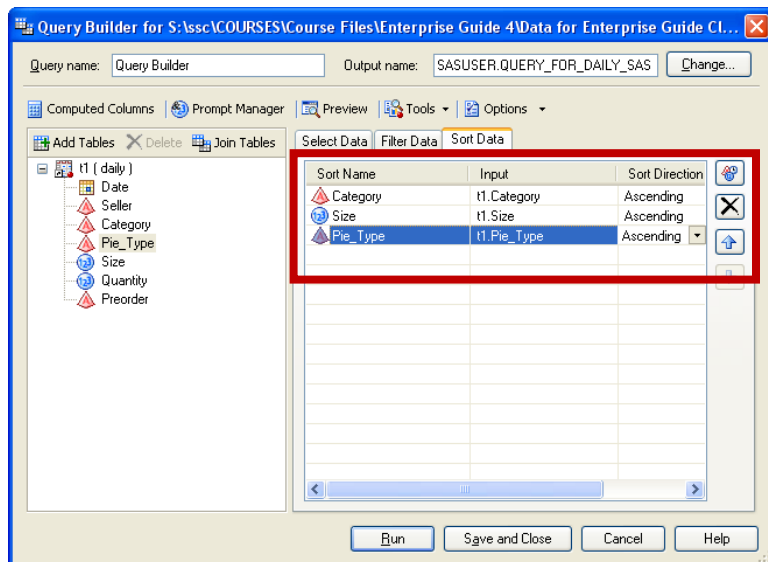


Figure 4: Sorting in the Query Builder

## FILTERING DATA IN THE QUERY BUILDER

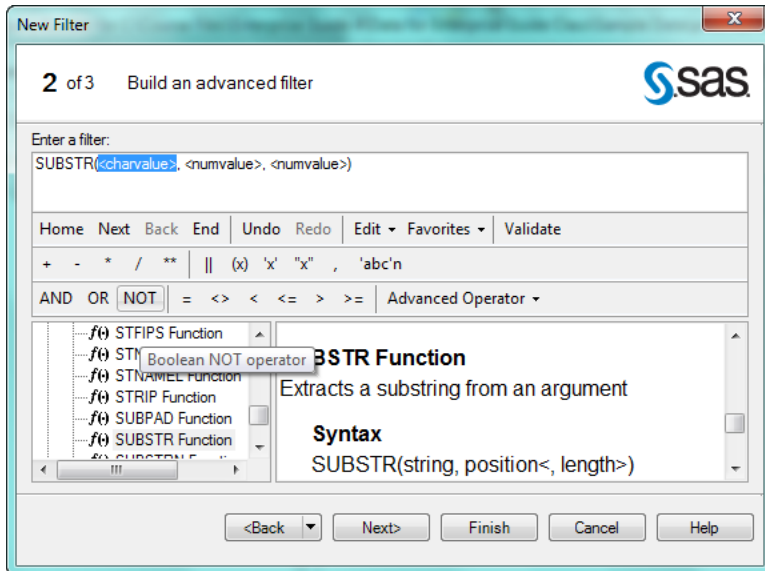
Filtering can also be carried out in the Query Builder. A filter specifies a condition that determines which rows will be included in the query results. It does not affect which columns are in the results. For example: Inventory is less than 200, Age is greater than 65, Importance equals 'urgent'.

To create a filter, click the Filter Data tab in the Query Builder. Double-click a variable to open the Edit Filter window (or click the filter button). This variable will be the left side of the filter condition.

An operator states the relationship between the two sides of the filter condition. Examples of operators are greater than, less than, contains, equal to, and many more. These can all be selected from the Operator drop-down menu.

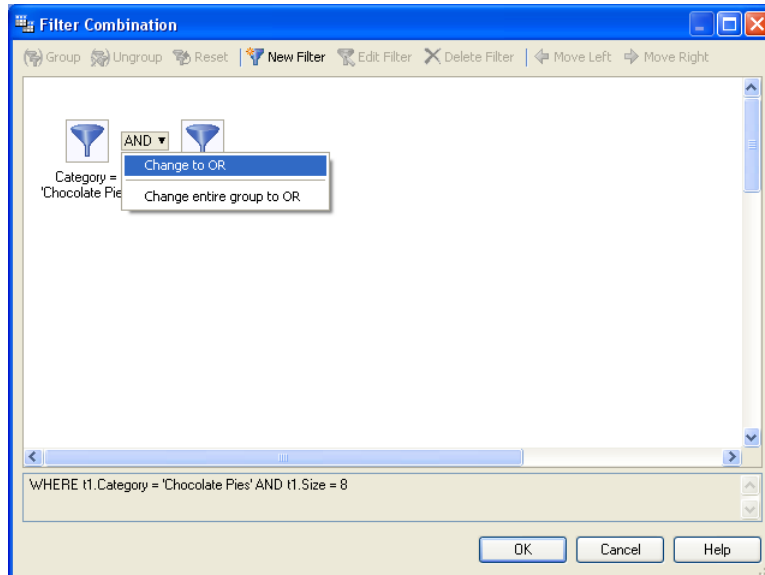
The right side of the filter condition may be a value, a value selected from a list, a group of values, a variable, a user entered value, or a prompt.

Advanced filters can contain expressions or functions on either side of the operator. These filters are created using the Advanced Expression Builder. These are built by typing in the expression box, double-clicking variables, values, or functions, or single-clicking operators. Over 500 SAS functions are available in a filter, and can be used with point and click options.



**Figure 5: Advanced Filters with SAS Functions**

Multiple filters can be combined. By default, if two or more filters are added to a query, the filters are related by AND. AND means an observation will pass through only if both conditions are true. OR means an observation will pass through if either condition is true or if both are true. To change an AND Filter to an OR Filter, on the Filter Data tab, click Combine Filters to open the Filter Combination window. Click AND between the two filters. Click Change to OR.



**Figure 6: Changing filters From AND to OR**

If filters are grouped together, the user must be careful about the group logic. Note the difference between: (Filter 1 AND Filter 2) OR Filter 3 and Filter 1 AND (Filter 2 OR Filter 3). In the Filter Combination window, hold down the CTRL key and click on filters to select filters for grouping. Use the Group and Ungroup buttons to change the way the selected filters are grouped.

Filters can also be deleted. On the Filter Data tab, highlight a filter. Click the X button to delete the filter.

## JOINING DATA TABLES

Joining is the process of merging two data sets by fusing together certain observations based on the values of variables. A join can be carried out using the Query Builder.

### TYPES OF JOINS

An automatic join occurs when two data sets in a query have a variable with the same name and type. EG searches for a suitable join variable for an automatic join when the second data table is added to the query. An automatic join is an equijoin by default.

In an inner join, a row is included only if the join variable value is common to both tables. Rows without a match are omitted.

In a full outer join, all rows from both tables are included. In a right outer join or a left outer join, all rows from one of the tables are included, along with the corresponding rows from the other table.

### SETTING UP JOINS

To add an additional data set to a query, click Add Tables. Click Local Computer, Servers, SAS Folders, or Project. Highlight the name of the data set. Click Open.

EG will try to do an automatic join (if it can find variables with the same name and variable type). To view the join, click the Join Tables button. By default, none of the columns from the new table will be included in the query. Double-click columns on the Tables list to add them to the query.

A manual join requires the user to specify which variables will be used to join the tables. It is necessary when joining by more than one variable or variables with different names. To perform a manual join, add the second data set. Click OK when a message tells you that you need to join the tables manually. Right-click the join variable from the first table. Select Join <variable> with. Select the name of the second table. Select the name of the join variable from the second table. A window will pop up for you to define what type of join. Click OK to close the Join Properties screen. Click CLOSE to close the Tables and Join screen.

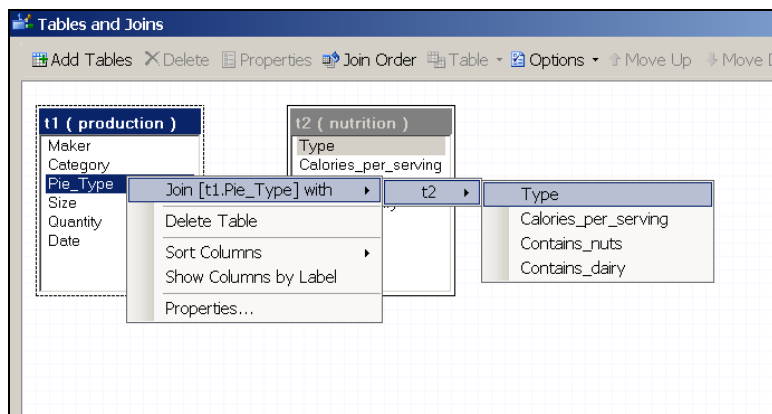
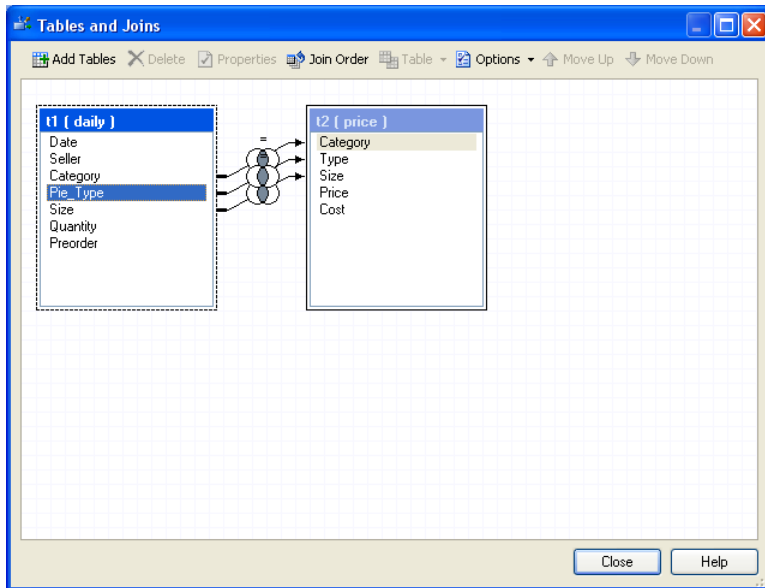


Figure 7: A Manual Join in EG

If data sets have multiple variables in common, an automatic join uses only one variable to match the data sets. It may be necessary to manually join the tables using other variables the data sets have in common. Right-click the next join variable from the first table. Select Join <variable> with. Select the name of the second table. Select the name of the join variable from the second table. Select the type of join from the pop-up menu. Repeat for additional join variables. Click Close.

To remove a join variable, right-click the join symbol connecting the variables. Click Delete Join.




**Figure 8: Joining By Multiple Variables**

To join more than two tables, continue adding data sets. Modify joins or make manual joins as necessary. Enterprise Guide can join up to 32 tables in a single query.

Date	Category	Pie_Type	Size	Quantity	Price	Cost	Calories per Serving
06/25/2006	Chocolate	Columns from Daily data		2	Columns from Price data		Columns from Nutrition data
06/25/2006	Chocolate Pies	Death By Chocolate	8	1			
06/25/2006	Chocolate Pies	Peanut Butter Cup	8	2			
06/25/2006	Chocolate Pies	Peanut Butter Cup	10	1	11.99	6	750
06/25/2006	Chocolate Pies	Black Forest	8	5	7.99	3.95	700
06/25/2006	Fruit Pies	Raspberry	10	2	8.99	4.2	400
06/25/2006	Fruit Pies	Strawberry	10	2	6.99	3.7	485
06/25/2006	Fruit Pies	Blueberry	10	3	7.99	4	375
06/25/2006	Fruit Pies	Lemon Meringue	10	5	10.99	5.65	525
06/25/2006	Fruit Pies	Banana Cream	8	2	7.99	4.2	490
06/25/2006	Fruit Pies	Apple	8	4	6.99	3.7	350
06/25/2006	Fruit Pies	Apple	10	1	8.99	4.5	350

**Figure 9: Output from Joining 3 Tables**

By default, all joins are inner joins. To change the join type, click the Join Tables button in the Query Builder. Right-click the join symbol (  ). Click Properties. Under Join Type, select the new join type.

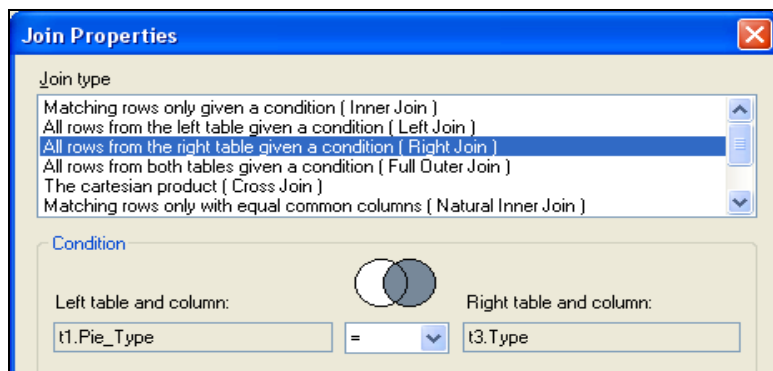


Figure 10: Changing Join Properties

## CONCLUSION

This paper overviews just some of the capabilities of Enterprise Guide's Query Builder. The Query Builder has much more that it can do. It is intuitive and easy to do so much of what we need in SAS right in this one task. And it's easy!

## RECOMMENDED READING <HEADING 1>

- *The Missing Semicolon:* [www.sys-seminar.com](http://www.sys-seminar.com) – Enterprise Guide Tips, Papers, and Webinars
- *Chris Hemedinger-The SAS Dummy* <http://blogs.sas.com/content/sasdummy/>

## CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name:	Jennifer First-Kluge and Steven First
Organization:	Systems Seminar Consultants, Inc.
Address:	2997 Yarmouth Greenway DR
City, State ZIP:	Madison, WI 53711
Work Phone:	608-278-9964
Fax:	608-237-1081
Email:	<a href="mailto:train@sys-seminar.com">train@sys-seminar.com</a>
Web:	<a href="http://www.sys-seminar.com">www.sys-seminar.com</a>

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. © indicates USA registration.

Other brand and product names are trademarks of their respective companies.