

# What is a SAS Index Good for Anyway?

Michael A. Raithel

If there was a SAS performance tool that could drastically reduce your program's I/O's, lower its CPU time, and decrease its run time, would you use it? Of course you would! Such a performance tool exists; it is called a SAS index. SAS Indexes can dramatically improve the performance of programs that access small subsets of observations from large SAS data sets. They do this by only accessing and returning the observations that you specify in a WHERE expression, instead of reading the entire SAS data set.

It is easy to understand how a SAS index can help you to directly access the observations that you need in a particular SAS data set. As an exercise, do the following: Open SAS Online Documentation, Click on the <Search> tab, enter the word "rtrace" into the Search tab's window, and click on <Search>. The SAS Online Documentation search function returns about a dozen links. When you click on any of those links, you get a page in the documentation that discusses the SAS RTRACE facility. This saves you the tedious effort of going through the entire SAS Online documentation, page-by-page, looking for occurrences of the word "rtrace".

A SAS index is analogous to the search function, above. A good index allows your programs to quickly access the subset of SAS observations that you need from a large SAS data set when you specify a key variable value (or values) that must be matched. This can dramatically improve the speed and efficiency of your SAS programs.

Conversely, badly conceived SAS indexes return far too many observations and are no better than reading the entire data set sequentially. In the analogy, above, consider how many pages would be returned and how much longer it would take if you searched the SAS Online Documentation for the word "SAS". That is why it is important to know more about the selection criteria for index variables, as well as the actual creation and use of SAS indexes.

After deciding that an index is appropriate for your subsetting purposes, you have three tools to choose from to create one: the DATASETS procedure, the SQL procedure, and the DATA Option in a Procedure or in the DATA step. When you do so, SAS creates an index file and associates it with your SAS data set. SAS stores additional indexes in that file and deletes the file when all indexes have been removed from the data set.

You can create a Simple index from a single variable, or a Composite index from two or more variables. A single SAS data set can have as many indexes as you think are necessary. However, each index you create increases the size of the SAS index file.

You can exploit indexes with the WHERE statement, the BY statement, or the KEY statement used in either a SET or MODIFY statement. In doing so, you will be increasing the efficiency of your SAS programs that use the index. That is what SAS indexes are good for!

There is enough information about SAS indexes to fill an entire book. If you are interested in learning more, either check the SAS online documentation, or consider my new book: The Complete Guide to SAS Indexes, at: <http://www.sas.com/apps/pubscat/bookdetails.jsp?catid=1&pc=60409>