POWER SAS: A Survival Guide

Library of Congress Cataloging-in-Publication Data

Lafler, Kirk Paul, 1956-
Power SAS : a survival guide / Kirk Paul Lafler.
p. cm.
ISBN 1-59059-066-X
1. SAS (Computer file) I. Title.

QA276.4 .L34 2002
005.3'042--dc21

2002013258

Printed and bound in the United States of America 12345678910

Technical Reviewers: Charles Shipp, Mike DiGregorio, and Sunil Kumar Gupta

Editorial Directors: Dan Appleman, Gary Cornell, Jason Gilmore, Simon Hayes, Karen Watterson, John Zukowski

Managing Editor: Grace Wong

Project Manager and Development Editor: Tracy Brown Collins

Copy Editor: Corbin Collins

Production Manager: Kari Brooks

Compositor: Diana Van Winkle, Van Winkle Design

Artist: Cara Brunk, Blue Mud Productions

Indexer: Rebecca Plunkett

Cover Designer: Kurt Krames

Manufacturing Manager: Tom Debolski

Marketing Manager: Stephanie Rodriguez

Distributed to the book trade in the United States by Springer-Verlag New York, Inc., 175 Fifth Avenue, New York, NY, 10010 and outside the United States by Springer-Verlag GmbH & Co. KG, Tiergartenstr. 17, 69112 Heidelberg, Germany.
In the United States, phone 1-800-SPRINGER, email orders@springer-ny.com, or visit http://www.springer-ny.com.
Outside the United States, fax +49 6221 345229, email orders@springer.de, or visit http://www.springer.de.

For information on translations, please contact Apress directly at 2560 9th Street, Suite 219, Berkeley, CA 94710. Phone 510-549-5930, fax 510-549-5939, email info@apress.com, or visit http://www.apress.com.

The source code for this book is available to readers at http://www.apress.com in the Downloads section. You will need to answer questions pertaining to this book in order to successfully download the code.

CHAPTER 7

Efficiency and Performance

EFFICIENCY AND PERFORMANCE are important concepts that involve improving the operation of the SAS software and your other programs. This chapter presents basic techniques on tuning a SAS program. You'll learn valuable tips on performance planning, CPU, I/O, memory, storage, and best-practices programming approaches to make your programs perform better. Space constraints prevent me from doing more than whetting your appetite on this topic—there are entire books written on the subject of tuning, and you should consult them for more minute details on this important aspect of running your software.

In this chapter, you'll learn how to

- Develop a basic planning checklist
- Apply basic CPU-performance techniques
- Avoid unnecessary sorting
- Determine whether an index is needed or not
- Use temporary arrays
- Apply basic I/O performance techniques
- Copy data sets with indexes
- Consolidate program steps
- Apply basic memory performance techniques
- Delete temporary WORK data sets
- Handle large data sets
- Apply basic storage performance techniques
- Assign lengths to numeric variables
- Compress SAS data sets
- Create user-defined formats for coded data
- Apply best-practices programming techniques
- Save labels and formats in SAS data sets
- Create views to reduce data redundancy
- Set system options to control messaging

Planning

This section offers valuable tips related to developing a plan for handling program design issues, constraints, and goals.

1 Planning is everything

It has been said that most people do not plan to fail—they just fail to plan. This adage could help millions of users throughout the world.

13 Using LIBNAME engines

Use existing LIBNAME engines to read from and write to a DBMS object (such as an Oracle table) as if it were a SAS data set.

14 Using temporary arrays

Assigning constants to elements in _TEMPORARY_ arrays reduces CPU usage, as opposed to creating and dropping variables. This not only allows array elements to be accessed directly because they are stored contiguously in memory, it also conserves data storage space.

15 Using the Stored Program Facility

By using the Stored Program Facility for complex or long DATA steps, you eliminate the CPU time that would be spent compiling the DATA step code.

Be sure to store a copy of the DATA step source code.

I/O Techniques

This section offers a number of useful tips that can be used to reduce the amount of input/output (I/O) processes that a program performs.

1 Storing data as data sets

Store data as SAS data sets, not external files. A SAS data set is a very efficient form of storing data in the SAS System. When data is stored as an external file, it must first be read as raw data before each analysis. This generally increases I/O requirements and should be avoided.

2 Subsetting observations

Processing only the observations that are needed during the analysis reduces the amount of I/O resources required. This is accomplished by subsetting observations (records) that are processed with a subsetting IF statement when dealing with external data, a WHERE statement, or WHERE= dataset option when dealing with a SAS data set.

3 Creating subsets earlier than later

Performing subsets from large SAS data sets or external files as early as possible in a single DATA step and at one time can reduce I/O requirements.

4 Reducing the number of steps

Use a WHERE statement or WHERE= dataset option in procedures instead of creating a temporary subset data set in a DATA step and then running the procedure. This saves I/O by eliminating the DATA step used to create the subset.

5 Copying indexes

Using the DATASETS procedure COPY statement to copy data sets with indexes will save I/O and enable both the data sets and any indexes to be copied. In some legacy code, and prior to the advent of indexes, SAS data sets were frequently copied using a DATA – SET statement convention, which doesn't copy indexes.

CHAPTER 8

Configuration and Support

CONFIGURATION AND SUPPORT are areas that all too often don't get the attention they deserve. This chapter presents valuable tips to help you obtain answers to many important SAS System installation and configuration areas, and, once the software is up and running, find where to get ongoing technical support assistance. You'll learn techniques of controlling how SAS software behaves, as well as how it works. You'll also learn about the many services and the various ways SAS Technical Support can help answer your ongoing technical questions and needs.

In this chapter, you'll learn how to

- Install and remove Internet Explorer
- Change SAS System library locations
- Customize the SAS Configuration file
- Access SAS Institute's Service & Support Web site
- Perform searches for technical information
- Find and apply Hot Fixes
- Apply SAS System SETINITs
- Access a library of sample SAS programs and demos
- Exchange information and knowledge with other users with SAS-L
- Unsubscribe to SAS-L
- Share real-world experiences and expand your knowledge by attending user groups

Installation and Configuration

This section offers valuable tips related to the installation and configuration of the SAS System. You'll be able to enhance your SAS experience with customizable options and features.

1 Exploring Release 8.2 system requirements

Release 8.2 provides users with exciting new and enhanced capabilities. Before installing Release 8.2, you need to verify that your system hardware and software resources are adequate. For complete details about the system requirements for Release 8.2 under a variety of different operating environments, see the following Web sites.

To view these documents, you'll need Adobe Acrobat Reader, which is a free download from `www.adobe.com`.

WORK Library: Windows 95 and Windows 98

C:\Windows\Temp\SAS Temporary Files\

WORK Library: Windows NT

%TEMP\SAS Temporary Files

6 Choosing between different installation configurations

As a licensed SAS site, you can choose which type of installation you want. The SAS Setup permits three types of installation configurations for Personal or Client use.

- **Complete**: Installs all components that have been registered by your SAS site.
- **Custom**: Installs only those components that you select and that have been registered by your SAS site.
- **Client**: Installs all components that have been registered by your SAS site to a network location allowing local machines to access the network version.

7 Invoking the SAS System

Once the SAS System is successfully installed, you're ready to invoke it for the first time. Here's how it's done.

1. Click the **Start** button on the Windows desktop, select **Programs**, and choose **The SAS System**.
2. Click The SAS System for Windows Release x.x to invoke the SAS System.

8 Discovering the SAS System's Configuration startup file

When the SAS System initializes, it processes a startup file called the Configuration file (SASV8.CFG). This file establishes important system options that control how the SAS System behaves. Options are available to control such things as communications, environment control, files, input control, graphics, log and procedure output control, Macro, sort, and system administration settings. For example, settings for linesize and pagesize, turning the DMS on or off, size of SAS work and sort space, number of buffers, printer options, location of SASUSER and WORK libraries, YEARCUTOFF value, and more can be specified here.

Under the Windows operating system, the SASV8.CFG file is located in the folder C:\Program Files\SAS Institute\SAS\V8. It's an ASCII file, so it can be opened and viewed with any editor that can handle and save ASCII files (for example, WordPad). The file is shown in Figure 8-1.

9 Customizing the SAS Configuration file

The SAS Configuration file is customizable, meaning you can modify specific settings. Using an ASCII editor (such as WordPad), users may modify the Configuration file to satisfy specific needs. You can specify two types of options: ones with a value assigned to them and ones that toggle between on and off.

Customizing the SAS Configuration file isn't difficult. You need to make your edits above the Warning box in the SASV8.CFG file. Before changing anything, though, you should be aware of the options specific to your operating system (see the following tip for a list of options under the Windows operating system). These can be found in the Help facility by entering "SAS System options."

CHAPTER 9

New Version 9 Features

THE RELEASE OF VERSION 9 introduces many new and exciting features, further enhancing SAS' reputation and reach in the fast growing business-intelligence software marketplace.Introduced as Project Mercury at the 27th annual SAS Users Group International Conference in the spring of 2002, Version 9 boasts unprecedented performance and scalability. This chapter presents tips on a number of the new features introduced in Version 9. With a focus on creating open and adaptable technologies, as well as reducing IT costs, SAS is committed to improving its support in the area of Web services by providing software that can be integrated into all IT environments.

In this chapter, you'll learn many new Version 9 features, including

- DATA step Perl Regular Expressions (PRXs) for performing faster searches
- Text search, retrieval, and replacement capabilities with regular expressions
- Output Delivery System (ODS) features
- ODS MARKUP destination capabilities for creating Web content
- HTML and XML options used with the ODS MARKUP destination
- Macro statements and options
- The SAS Learning Edition

The DATA Step

This section provides an early look at the number of useful new DATA step features found in Version 9.

1 Performing faster searches and text replacement with Perl Regular Expressions (PRXs)

Performing text operations related to matching, changing, or translating a string of characters is popular in text-processing applications. *Perl Regular Expressions* (PRXs) enable users to perform complex word matching and text replacement where a set of characters determines a pattern found in the text. UNIX users have used regular expressions for some time, and now many new functions such as PRXPARSE, PRXMATCH, PRXPARSE, PRXPOSN, PRXCHANGE and more are available in Version 9 of the SAS System.

2 Describing PRXs and metacharacters

If you need to know something or want to identify patterns in text, PRXs provide the tools to get the job done. They provide a powerful way to search and match strings and retrieve and validate data. Although unique in their syntax, regular expressions are essentially a string of characters and special characters called *metacharacters* (for example,

2 Exploring ODS MARKUP file types

You write output to one or more files by substituting the specific option and filename in the ODS MARKUP statement. Here is a brief list of the purpose of each file, followed by more detailed information. You can specify the following types of files with the ODS MARKUP destination:

- Body= file
- Contents= file
- Page= file
- Frame= file
- Code= XSL (Extensible Stylesheet Language)
- Stylesheet= option

The body file

The body file consists of output created by your SAS job. It takes on the characteristics of whatever tagset was specified (such as HTML, XML, and so on). When output is routed to the HTML destination, it is placed within one of three HTML tags:

- TABLE
- IMG
- As an HTML table

The nature of the output object determines which of the three tags is used for displaying with a Web browser. When creating the body file, ODS handles output objects differently depending on the nature of the output. If the output object consists of tabular data without any graphics, ODS inserts the object within TABLE tags. When the object contains a graphic image, ODS places it within IMG tags. And when the object does not contain tabular data or a graphic image, ODS tags it as an HTML table.

The contents file

The contents file consists of a link to each output object in the body file. It uses an anchor tag to link to each table. With your browser, you can view the contents file directly or as part of the frame file.

The page file

The page file consists of a link to each page of output in the body file. With your browser, you can view the page file directly or as part of the frame file.

The frame file

The frame file integrates the body, contents, and page files into a cohesive Web page when you specify the HTML tagset.

3 Obtaining a list of tagset names

The SAS System lets you request a list of the various tagsets stored in the read-only SASHELP.TMPLMST library and, if available, the SASUSER.TEMPLAT library. To obtain a list of tagset names, run the following TEMPLATE procedure. The results illustrate a partial listing of the available tagset names.

Code:

```
PROC TEMPLATE;
 LIST TAGSETS;
RUN;
```