

Tom Cole, SAS Institute Inc., Cary NC

1 Introduction

This paper is an overview of the current performance and features of the SAS product under the VAX/VMS operating system. Many performance improvements, enhanced features, and new samples have been incorporated in the newest release of the SAS System under VMS.

1.1 Product History

The SAS System has seen a steady progression of improvements in its performance and capabilities on minicomputers since the inception of its development over five years ago.

Early releases concentrated on bringing most of the functionality and "feel" of the SAS System to minicomputers. Emphasis was on creating a framework in which additional features and products could be added at a later date. Performance was a concern but it was not the main priority during the first releases of the SAS System.

The user community began to use these versions of the SAS System for Minicomputers and provided the Institute with extensive feedback on both the features and the relatively low level of performance. The Institute shifted gears and began an extensive program of revision and enhancement to bring the system "up to speed". The goal was to make the SAS System a tool that was an acceptable balance of sophisticated software and resource utilization.

Each subsequent release of the SAS System provided performance improvements and more efficient resource utilization, combined with a steadily growing level of compatibility with the existing SAS Systems. Improvements in execution speed and associated resource utilization grew from twenty to one hundred percent between each successive release of the system.

Of particular note is release 5.03 of the SAS System. This release has been the best performing version of the SAS System under VMS, and became the groundwork for the newest release of the SAS System. However, this release became "infamous" because of a problem with compatibility between versions of the VAX/VMS operating system. The problems that caused that dependence have been solved, and special checks are included in our development cycle to prevent that particular problem from occurring again. SAS Institute recognizes the extreme importance of being as independent of operating system versions as possible. Digital Equipment Corporation policy regarding releases of VMS is somewhat beyond our domain. However, every effort is being made with the current and future releases to eliminate version dependencies.

The version completing Beta Test cycle is SAS Release 5.16. This release is currently available to licensed sites as a test release. Sites that are interested in helping with final testing of this product should contact their Software Sales Representative.

1.2 Version Synchronization

The SAS System for minicomputers has always consisted of two distinct components. The first is the base supervisor and DATA step, which was developed for the minicomputers. The second is the PROCs, which in most instances are the same procedures provided in the mainframe versions of the SAS System.

Because the PROCs are the same source code across all operating systems and are ported to every system, the Institute has elected to begin numbering versions of the SAS System consistently across all operating systems. The next production release to be shipped is SAS 5.16. This means that the procedures included in this software system are the same procedures provided with the SAS System under OS and the SAS System under CMS. There have not been thirteen releases since SAS Release 5.03 was shipped; we are simply synchronizing version numbers to avoid confusion about the status of the procedures and their compatibility between minicomputer and mainframe products.

1.3 What's Special About SAS Release 5.16?

This paper describes Release 5.16 in some detail, and covers both feature and performance enhancements to the product. There have been extensive rewrites of some major subsystems which provide significant improvements for most SAS users.

This new release of the SAS System incorporates additional support for new VAX processors such as the new 8000 series processors and MicroVAX-based systems. It also has the capability of licensing a single copy of SAS software to run on a multi-cpu cluster using a shared disk.

SAS Release 5.16 is the first version of the SAS System under VMS with the new CPE ToolKit. This first release of the ToolKit consists of a collection of SAS programs, user written functions and formats, and VMS batch jobs for collecting and organizing VMS performance data using VMS data gathering utilities. The CPE ToolKit is provided in subdirectories of the SAS sample library.

1.4 Summary of New Features and Performance Enhancements

This list describes briefly the features of the SAS System which are available with SAS 5.16 that contribute towards its enhanced performance and flexibility.

Performance Improvements

- o DATA Step Code Generation
- o Dynamic Loading And Unloading Capabilities
- o Dataset I/O redesign

Feature Enhancements

- o User Exit Support
- o New Software Products Supported Since Last Release
- o PROC PRINTTO
- o New Full-Screen Installation Procedure
- o SAS Options Files
- o Command Line Options
- o User-written Extensions To SAS
- o Complete Support Of Control-Y And Control-C
- o Command Line Recall
- o New SETINIT Support For Clusters
- o Miscellaneous "Ease-Of-Use" Enhancements

Many features described here will also be available in future releases of SAS under AOS/VS and PRIMOS operating systems. More information and details are available from developers from these groups and the Software Sales Representatives.

The following sections describe each of the above improvements and how each one affects the SAS System's performance and usability under the VMS operating system.

2 DATA Step Code Generation

In previous versions of SAS for minicomputers, the SAS DATA step was interpreted. This means that DATA steps were compiled to pseudo code which was then "executed" by the SAS System. With release 5.16, the DATA step compiler generates native VAX machine code. Some of the features of this enhancement are:

- o Increased execution speed.
The DATA step may execute up to ten times faster depending upon the application. DATA steps that are very compute intensive or perform many string manipulations will show the greatest performance improvement.
- o Minimal impact on compilation time.
The additional work being done to generate the executable machine code is very small, and typically accounts for less than ten percent of the compile time for the step.
- o Better handling of precision, DO-loops, etc.
One effect of this enhancement is that jobs that are dependent on extremely small numbers for calculations and comparison operations will run with greater accuracy. In addition, special enhancements have been made to the DO-LOOP logic to ensure that it will run the correct number of times regardless of the precision of the control variables.
- o All existing jobs should run with the code generator without change.
It is important to note that the new code generator is completely transparent. No changes are needed in SAS programs to use this new feature.

3. Dynamic Loading Package

This release of the SAS System under VMS incorporates a major redesign of the way that dynamic loading occurs. Dynamic loading means that the executable code needed to run a particular SAS session is loaded into memory on demand.

Before VMS SAS Release 4.10, the SAS System was shipped as a single large image. At that time, the only additional product available was SAS/GRAPH. However, an image with the complete base SAS System plus SAS/GRAPH was almost eighteen thousand disk blocks in size. This entire image had to be loaded and consumed huge amounts of memory.

With release 4.10 of the SAS System under VMS, most PRDCS were dynamically loaded. The code for a procedure was only loaded into memory if the procedure was used. No unloading was provided; a PROC remained in memory for the duration of the remaining session. Dynamic loading allowed the Institute to ship additional software products separately as they became available. It also paved the way for user-written procedures on the VAX.

In SAS Release 5.16, all major optional system components are dynamically loaded. This means that if DATA step functions or graphics device drivers, or even the SAS Display Manager System are not used, then the code for these parts of the SAS System are not loaded into memory. This is especially helpful for batch jobs which need not pay for the overhead of the display manager when it cannot be used.

New with SAS 5.16 is the concept of optional unloading. This feature is controlled by either the user or system manager. This option frees memory that contained image code when it is no longer being used. For example, when the GLM procedure completes, the memory used by the procedure can be freed up to be used for other steps.

The benefit of this feature is that the SAS System uses less physical and virtual memory resources. Acceptable performance can be achieved with a smaller working set (amount of physical memory allocated to each process). Less virtual memory and page file space is consumed for a normal job.

Because less code must be resident for the SAS System to run, the feature reduces page faulting for image activation at startup. This means the SAS System activates faster, and the first prompt appears much more quickly than in previous releases.

Because the components of the system are individual images, the new enhancement allows the system manager to install frequently used parts as sharable code pieces. This improves page faulting performance when multiple users are accessing the SAS System.

Another consequence of unbundling the system is that the base SAS image has steadily grown smaller with each successive release of the system, while the functionality of the entire system has grown. Because it is unbundled, system managers may elect to remove unused parts of the SAS System that you do not use from online disk storage. This may be useful on computer configurations with small amounts of disk space, such as smaller MicroVAX systems. The new full-screen installation procedure shipped with SAS Release 5.16 includes features for selecting components of the SAS System to be installed at each site.

Finally, because each kind of SAS System component is dynamically loaded instead of being an intrinsic part of the base SAS image itself, support of user-written extensions to the SAS System is possible. In addition to user-written procedures, the Institute will support user-written functions, formats and informats, and device drivers in the near future. The sample library for the CPE Toolkit that is shipped with SAS 5.16 contains sample user-written functions, formats, and informats.

4. Dataset I/O Redesign

SAS has long been known as a heavy resource user on any operating system, and this is particularly noticeable on minicomputer systems. Because of the software's use of SAS datasets for storage of user data as well as internal work files, the Institute concentrated on improvements to the dataset I/O subsystem of the SAS System. This involved a significant rewrite of major portions of the code for handling all kinds of SAS-specific datasets.

The redesign consisted of two phases. The first was a rewrite of the "portable" subsystem. This is the code that is common to all minicomputer versions of the SAS System. The second phase was a host-specific rewrite of I/O support on each minicomputer. Many layers of the system software were compressed to decrease procedure overhead and better use was made of the VMS I/O functions. Additionally, the lowest level I/O was converted from RMS record I/O to block I/O.

As a result of these performance improvements, sample jobs show that dataset I/O performance has improved 40-150% depending on the SAS program. Areas that show the most improvement are jobs that perform large amounts of sequential I/O.

These improvements do not affect support for previously created datasets or utility files. Datasets and utility files created in previous versions can be read and updated with SAS 5.16. However, please note that utility files (such as PROFILE.SFS or catalog datasets) created under SAS 5.16 cannot be read by previous versions of SAS. Only SAS datasets created by SAS 5.16 can be manipulated using previous releases.

In response to user requests for more complete tape handling, all "sequential" dataset operations on tape are fully supported. However, tape operations not normally provided by VMS cannot be performed on SAS datasets. Datasets on tape cannot be renamed, deleted, or opened for update.

In order to optimize disk and tape operations, new SAS options can be used to control internal buffer sizes for dataset operations. A new option, BUFSIZE=, controls the internal buffer size used by the I/O subsystem for datasets. For operations that are sequential in nature, a large BUFSIZE results in the best performance. For operations involving lots of random access to datasets, a BUFSIZE that is a small multiple of the observation size is best. The default is 4096 bytes, but can range from 512 to 32768 bytes.

For tape datasets on unlabeled tapes, the user must use the DCL MOUNT command qualifier /BLOCKSIZE along with the BUFSIZE option to control the blocking on the tape. The default is 4096 bytes. Datasets can be written to tape using a block size from 512 to 32768, resulting in much more efficient usage of tape.

Another feature of the new version is a redesign of the way in which SAS software accesses the WORK library. In previous versions, files written to the work library had a special name to identify them as temporary files, and to prevent file collisions. However, there were still contention and session cleanup problems with this scheme. In SAS 5.16, the

work library is a unique subdirectory created by SAS at startup. There is one subdirectory for each process that is running SAS. When the SAS System exits, it deletes the subdirectory and its contents. As a result, there are no longer "temporary" file extensions like .SSW for datasets in the WORK library. All datasets have the extension .SSD regardless of the library in which they reside.

New options NOERASE and NOWORKINIT control the disposition of the work library at shutdown and startup. A user may elect to retain the contents of the WORK library after his session completes, or to have the SAS software reuse an existing WORK library.

5 User Exit Support

SAS Release 5.16 under VMS includes a feature designed to allow the user or site manager to exercise control over SAS software usage or perform individual accounting about SAS software usage. This feature is known as the User-Written Exit system.

A user-written exit is a VMS executable image written by the system manager or an individual user that is loaded and executed at specific points during the execution of the SAS System. The exit routine is provided with the action being performed (for example, starting a PROC), the user's name, the time of day, and the name of the object the action is being performed upon (for example, the name of the PROC).

This feature is provided primarily to address three areas of need. First, sites may wish to control access to individual components of the SAS System. For example, universities may wish to allow users to execute SAS software interactively, but not allow students to use the full-screen software products. Another example would be sites that wish to control the use of the SAS system in batch versus interactive mode for particular users. An exit can be written to disallow the user access to the SAS software or specific procedures based on the user's name.

A second area of interest is user accounting. Exits can be written that store accounting data for each PROC run by a user, or for an entire SAS session. This is especially useful for sites that perform timesharing accounting for individual software products and sites that have SAS RCS licenses.

A third area of interest is in monitoring SAS usage for the purposes of system tuning. Modules that are loaded frequently and used for long periods of time by multiple users are candidates for installation as VMS known images. Exits can be written to provide this kind of usage information. For example, during the alpha test cycle for SAS 5.16 the Institute used exits to determine which components should be dynamically loaded and which should be linked into the image because they are used often. These sample exit routines are provided in a subdirectory of the SAS Release 5.16 sample library.

6 CPE ToolKit Provided In Sample Library

The SAS 5.16 sample library includes a toolkit for CPE applications. This is the first part of an ongoing commitment on the part of SAS Institute to provide a complete set of tools for solving CPE problems using the SAS System.

The samples are divided into four major areas based on the tool used to gather the CPE data from the VMS system. Performance data can be gathered using the VMS Monitor Utility or the VMS layered product SPM (Release 3.0 or later). Accounting data can be read directly from VMS accounting files and stored in SAS datasets for further analysis. VMS diskquota information can be gathered from the running system and stored in SAS datasets.

Sample VMS command files for capturing each kind of data are provided. SAS programs for reading this raw data and producing SAS datasets containing the descriptive information about the system are also provided. Further processing of the data and generation of line printer plots and SAS/GRAPH displays are performed by additional sample programs.

7 New Institute Program Products And Features

New SAS software products and base SAS software features have been made available since the last release of SAS under VMS. These are summarized below. For more information, contact a Software Sales Representative.

7.1 SAS/QC Software, A Statistical Quality Control Package.

This product has many applications in the minicomputer environment. It provides tools for analyzing data gathered from manufacturing or other processes using statistical quality control methods.

SAS/QC includes a full interface to SAS/GRAPH, permitting users to plot graphs of their data using either line-printer plots or graphics devices. Note that a SAS/GRAPH license is required to use the graphics device interface in SAS/QC.

User-written sample functions TTOPEN, TTREAD, TTWRITE, and TTCLOSE were developed to assist in collecting data from process control equipment for use with SAS/QC software. These procedures serve a dual purpose as samples of user-written functions (written in the C programming language) as well as being general purpose tools for reading data from an asynchronous serial port on a VAX or MicroVAX system. These samples show methods for gathering data directly from a SAS program. Some modifications may be necessary for use in a particular application. These sample functions are provided with the base product.

7.2 SAS/IML Software, The Interactive Matrix Language.

SAS/IML software has become a production product for SAS under the VMS operating system since the last release of the SAS System. This product provides a conversational language environment for performing complex matrix and array calculations using English-like commands.

7.3 The Micro-To-Host Link.

This release of SAS under VMS supports the Micro-To-Host Link as a standard part of the base SAS System. This feature allows PC SAS users to make use of the computing power of the VAX from a remote session. Users of the SAS System on a PC can transparently execute SAS programs on either the local system or transmit them to the VAX for remote processing. Output from each step can be transmitted back to the PC for review by the user.

PROC UPLOAD and PROC DOWNLOAD are provided for transferring entire SAS datasets and external files between VAX and PC systems.

7.4 CBT Training.

Release SAS 5.16 under VMS supports SAS Computer-Based Training (CBT) courses. CBT001, "Fundamentals of the SAS Software System" will be available soon.

8 PROC PRINTTO Supported

The Institute is responsive to the needs of the user community, which helps us make the SAS product as functional and usable as possible. Based on requests from many users, minicomputer SAS now supports PROC PRINTTO with release 5.16. This procedure is used to redirect SAS output to disk files for later processing.

The minicomputer version of PROC PRINTTO supports the existing syntax of the mainframe versions of PRINTTO for SAS programs migrated from other environments.

PROC PRINTTO on the minicomputers supports extensions appropriate for path-oriented access to files. The new FILE= option identifies the output file as one referenced in a previous FILENAME statement. The NAME= option allow the user to specify the full path name using VMS syntax in quotation marks.

9 Full-Screen Installation Procedure

The installation procedure for loading a SAS System tape onto a VAX system has changed dramatically with the release of SAS 5.16. The procedure now supports a full-screen menu-driven interface for ANSI-compatible terminals. The procedure may still be run on hardcopy terminals if desired.

The full-screen implementation features pop-up help windows, menu selections, and records a complete log of all operations performed during installation. The hard-copy implementation functions in a fashion similar to the previous versions of the installation. SAS Institute strongly recommends using the full-screen version of the procedure. It is much easier to use, and keeps a more complete record of the operations performed during installation. A new comprehensive "Installation Guide" describes both forms of the installation procedure.

10 SAS Options Files

SAS 5.16 for minicomputers supports a facility referred to as the "SAS Options File" facility. Using VMS logical names, the system manager and users may specify files containing SAS statements which are executed automatically at the start of the session. The system file is processed first, followed by a user-specified file, if any, and then command line options.

This file can contain any SAS statements. However, it is specifically intended to contain statements for configuring the SAS session, such as option statements, libnames, libsearch lists, and filename statements.

The first prompt the user receives is still the "1?" prompt. Option files are processed in a fashion similar to %INCLUDE files, and error messages and other diagnostics resulting from execution of an options file are displayed to the user. OPTION LTYPE can be used in the options file to display its text on the SAS log.

11 Command Line Options

In previous releases of the SAS System, the FSD option could be specified on the DCL command line to invoke the SAS Display Manager System. This capability has been extended with SAS Release 5.16 to permit any valid SAS System option to be specified as a qualifier on the command line.

In the VAX environment, double quotes must be used for option strings instead of single quotes. The option name must be completely spelled out as in all SAS OPTION statements. The SAS System does not support the DCL abbreviation of qualifiers.

Options specified on the command line override options specified in the SAS options file(s) if option files are being used. Options take effect before the specified SAS program (if any) executes, or before the 1? interactive prompt.

12 User-Written Extensions To SAS

Because of the power and flexibility of dynamic loading, user-written extensions to the SAS System can be more fully supported. Dynamic loading allows users to write some kinds of loadable modules themselves.

It is possible to write user-written components in most VAX/VMS languages. SAS Institute currently provides support for PL/I. The sample library includes some examples in the C programming language as well as PL/I for use as models in developing extensions to the SAS system. In the future the Institute may support development in other VMS languages.

SAS Release 5.16 allows the user to write user-written procedures, DATA step functions, formats, and informats. The installation tape includes samples of each of these written in either PL/I or C. Other examples and documentation may be provided in the future.

13 Extensions For Compatibility With The VMS Environment.

The SAS System is designed for use in a wide variety of computing environments on a wide variety of operating systems. However, the Institute recognizes the need to support the unique features of each operating system. The SAS System under VMS Release 5.16 takes several steps towards this goal with additional VMS-specific features.

13.1 Complete support of Control-C, Control-Y.

These command keys are used in the VMS environment to interrupt the current flow of control. In the SAS environment, Control-Y allows one to exit from a SAS session at any time, with correct session cleanup.

Control-C behaves like Control-Y when no DATA step or PROC is running. However, Control-C will terminate the current DATA step or PROC and proceed to the program boundary or prompt. It can also be used to cancel a currently running X command.

13.2 Command line recall

SAS supports VMS command line editing for both hardcopy and CRT terminals. This means SAS conforms to the Digital Equipment Corporation "standard" for VMS system software, tools and utilities.

SAS uses arrow and control keys as documented in DCL manual. Unlike DCL, there is no limit to the number of lines that can be recalled. Any statement executed in your SAS session may be recalled. The %INCLUDE statement can also be used to recall previous program lines.

14 New SETINIT Features

The SETINIT syntax has been modified, and new multiple processor support has been added. The syntax has been changed to support additional features.

New support for up to sixteen CPUs in a single license is provided. This allows a single copy of SAS to be licensed for all processors in a VAX Cluster. A single license can support different product mixes for each machine in the SETINIT license.

Additionally, the SETINIT licensing process is about four times faster, and does not create a new copy of the base SAS image. This means that the installation procedure takes less time and consumes less disk space.

15 Other Miscellaneous Changes

The SAS System supports the DCL command "ATTACH" via the X command. There are no longer any restrictions on X commands in batch mode.

SAS permits %INCLUDE with a quoted file name string just like the FILE and INFILE statements, in addition to a fileref.

16 Summary

The SAS 5.16 System under VMS takes significant steps towards completing the unification of the SAS System across many computer systems, while adding many features that improve the usability and performance of the software in each environment.

- o Better performance. Faster in both DATA step and I/O bound applications.
- o Many enhancements to promote ease of use and better integration into the VMS environment.
- o Additional support for SAS Software Products.
- o More complete system, with a strong emphasis on quality.