Initial Experiences with the SAS System for Minicomputers

Edwin D. Suckstorff, Lilly Research Laboratories
C. David Hardison, Lilly Research Laboratories

ABSTRACT

This paper describes the experiences with the SAS System on a DEC VAX-11/780 at Lilly Research Laboratories. Some differences between the SAS System under VAX/VMS and the SAS System under VM/CMS or MVS are described, and examples are given which illustrate the use of the SAS System:

1. to analyze VAX system performance data,
2. to analyze and generate reports from data collected on remote DECnet nodes,
3. to move SAS applications between the VAX/VMS environment and the IBM VM/CMS environment with PROC XCOPY.

THE SAS SYSTEM FOR MINICOMPUTERS PACKAGE

The Program Products available for minicomputers differ from those available for IBM computer systems. The major differences between the SAS System for minicomputers and the "standard" SAS System are summarized in Figures 1 and 2. At the time of this writing, the SAS System for minicomputers only includes SAS/BASE and SAS/GRAPH, plus Display Manager, which is a new development tool available only with the SAS System for minicomputers. SAS/BASE for minicomputers is complete, with the exception of the pre-processing facilities (%INCLUDE and the MACRO language) and the interactive HELP panels.

A COMPUTING ENVIRONMENT FOR THE SAS SYSTEM FOR MINICOMPUTERS

Eli Lilly and Company is a multi-national pharmaceutical manufacturer with corporate headquarters in Indianapolis, IN. Lilly Research Laboratories (LRL) is the research arm of the corporation, with its primary mission being the development of new pharmaceuticals. Within LRL scientists from diverse fields such as chemistry, biology, statistics, and engineering collaborate in the process of drug design, chemical synthesis, biological and clinical testing, scale-up and manufacturing. This research process produces huge amounts of data which currently are being managed primarily by several Digital Equipment Corporation VAX 11-780 mini-computers, with some additional processing being supplied by IBM mainframes. Figure 3 is a simplified diagram of the current LRL computing network.

Before 1980 (ie, before LRL acquired the VAXs) most of the large-scale scientific data analysis was performed using the SAS System on the IBM mainframes, with real-time data collection being performed by DEC PDP-11 and Hewlett Packard HP-1000 mini-computers. Most scientists were uncomfortable with the batch computing environment supplied by the IBM systems, and so either developed their own data analysis programs on the DEC and HP minicomputers (or had someone do it for them) or performed the analyses by hand. The introduction of the VAXs to the LRL research community brought easier-to-use interactive computing to the scientists; unfortunately, the best data analysis tool -- the SAS System -- still only existed on the IBM mainframes. The recent introduction of the SAS System for minicomputers now offers two significant opportunities to the scientist:

* data can be collected and fully analyzed on the same computer system, and
* the SAS System (with SAS/GRAPH) offers a complete, uniform, integrated package for statistical analysis, data management, graphics and report writing.

Several examples are given to illustrate some of the uses of the SAS System on the VAX.
EXAMPLE 1: ANALYZING VAX SYSTEM PERFORMANCE

Managing computing resources and insuring adequate computing support for over 200 VAX users is an extremely important task. One of the uses of the SAS System on the VAX at LRL is in the analysis and reporting of VAX system performance data. The MONITOR system utility (which is an integral part of the VAX/VMS operating system) enables the collection of raw resource utilization data of a VAX system. The MONITOR Utility operates in primarily two modes: Record mode, in which system resource utilization is sampled and the data recorded in binary format on disk, and Display mode, in which the resource utilization data is displayed on either a CRT terminal or in a file (suitable for printing). The interval between samples, as well as the duration of the monitoring, can be specified by the user. Data can be collected simultaneously on a number of resources: IO usage (both terminal and disk), memory (paging rates and utilization), CPU modes, and others. Unfortunately, the MONITOR Utility (in Display mode) does not provide any mechanism for graphically displaying resource utilization as a function of time. The SAS System and SAS/GRAPH, however, offers an excellent means for both graphically displaying resource utilization vs time (hour, day, week, month) and for doing the inevitable data manipulations. An example of a plot of resource utilization vs time is given in Figure 4. Reports and plots of this type are invaluable in system management and capacity planning.

A feature missing in the SAS System for minicomputers became apparent very quickly: the lack of INFORMATS for reading binary values (eg, IBw.d, PDw.d, RBw.d). Since the MONITOR Utility records its data in binary form, these INFORMATS would have been very helpful in reading the recorded data directly. To get around this problem a small FORTRAN program was written which translated the binary data into ASCII (suitable for input into the SAS System).

EXAMPLE 2: MOVEMENT OF SAS APPLICATIONS BETWEEN IBM AND NON-IBM SYSTEMS

With the SAS System on both the IBM systems and the VAX, it is possible to move an entire SAS application (both SAS code and SAS datasets) from one system to another. Transferring the SAS code is relatively simple: the file (or files) containing the SAS code can be copied from one system to another either via RJE link or magnetic tape. Transferring SAS datasets is more complex, however, since the data within the dataset files are in a system-dependent binary format. The transfer of a SAS dataset can be generalized as a three-step process:

1. transformation of the SAS dataset into “transport” format, a format which can be read by SAS on any system,
2. the physical transfer of the dataset from one system to another,
3. transformation of the dataset from transport format back into the “normal” format.

The method which LRL uses for migrating a SAS dataset from VM (IBM) to VMS (VAX) is shown in figure 5(a). PROC XCOPY is used to transform the SAS dataset into transport format; this “flat file” is then copied to tape and physically carried to the VAX system. Once the tape is loaded on a VAX tape drive, the SAS dataset (as a flat file) is copied and transformed into normal format directly via a SAS DATA step.

Migrating a SAS dataset from VMS (VAX) to VM (IBM) is very similar, as shown in Figure 5(b). A SAS dataset on the VAX (in transport format) is first copied to tape. The tape is then physically carried to the IBM system, where the SAS dataset is copied to disk and transformed into normal format by PROC XCOPY.

These methods work fine when transferring one SAS dataset; unfortunately, when there is more than one SAS dataset on a tape, PROC XCOPY does not automatically advance the tape to the next dataset - this must be performed explicitly in the code.

The mechanism of inter-system transference of SAS datasets between systems seems rather clumsy now, but we expect this mechanism to improve as we gain experience. The ability to move SAS datasets between computer systems offers great potential; when the Full Screen Product, SAS/FSP, becomes available for minicomputers, entire full-screen SAS applications can be migrated.
EXAMPLE 3: ANALYSIS OF EXPERIMENTAL DATA

LRL routinely screens antibiotic cultures against target microorganisms, the purpose being to identify cultures with greater antibiotic activity than the control cultures. A frequently used screening method is one in which the level of inhibition of microorganism growth is measured optically, with on-line data collection being performed by a DEC PDP-11/24. Before the introduction of the SAS System for minicomputers, programs would have been needed on the PDP-11/24 to analyze the test culture data against the control data, generate printed output and charts and archive the results. Now, however, the data are collected on-line by the PDP-11/24, stored in files, and shipped to the VAX via DECnet, DEC's inter-system communication link. Once on the VAX, the SAS System is used to perform the analysis, report generation, and data management. An example output report is shown in Figure 6. By using the SAS System the program development time and complexity has been reduced dramatically.

COMMENTS ON THE SAS SYSTEM ON THE VAX

Since LRL was an Alpha and Beta test site for the SAS System for minicomputers, we inevitably found bugs, and SAS Institute has been very cooperative in correcting them. But perhaps the most important question is "How well does the SAS System actually perform?" The SAS System was installed on one of the VAX-11/780 minicomputers at LRL in the first half of 1983. Prior to the installation of the SAS System, this VAX was very heavily used, with the primary performance bottleneck being contention for main memory (2 megabytes), i.e., most processes spent too much time paging. Understandably, the SAS System runs very slowly on this VAX, but it is unclear whether the slow execution is a result of the VAX system configuration, or a lack of efficiency within SAS itself. At the time of this writing, plans are to increase main memory to 8 megabytes (thereby creating a memory rich VAX environment) which should increase the performance of the SAS System significantly. However, making the SAS System as streamlined and as efficient as possible should be a priority issue for SAS Institute.

Display Manager is a very innovative and exciting tool for users of the SAS System. It provides an interactive session manager with utilities to retrieve previously executed commands and to scroll throughout the SAS log and SAS output. Also an interactive session can be saved by the SAS System on an external file. Two limitations of Display Manager are:

1. the requirement to "ASSIGN" an external file for saving a session prior to invoking the SAS System, and

2. the lack of notification to the user on the display terminal that SAS is processing a request, i.e., "WORKING" or "RUNNING" should appear on the display after a user enters a command or request.

Lilly Research Laboratories has a large base of computer users whose computing needs range from generating graphs and plots, to report writing, to database management, to data collection and analysis. Now with the SAS System for minicomputers, the LRL computer user can meet all these needs, regardless of the computer environment.

SAS and SAS/FSP are registered trademarks of SAS Institute Inc., Cary, NC, USA. SAS/GRAPH is a trademark of SAS Institute Inc. DEC, DECnet, PDP, VAX and VMS are trademarks of Digital Equipment Corporation.
Figure 1
SAS Institute Program Product Availability

<table>
<thead>
<tr>
<th>Institute Program Product</th>
<th>IBM SAS</th>
<th>SAS for Minicomputers</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAS/Base</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>SAS/GRAPH</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>SAS/ETS</td>
<td>YES</td>
<td>NO</td>
</tr>
<tr>
<td>SAS/OR</td>
<td>YES</td>
<td>NO</td>
</tr>
</tbody>
</table>

Figure 2
Available Facilities in IBM SAS and Portable SAS

<table>
<thead>
<tr>
<th>Facility</th>
<th>IBM SAS</th>
<th>SAS for Minicomputers</th>
</tr>
</thead>
<tbody>
<tr>
<td>MACRO Facility</td>
<td>YES</td>
<td>NO</td>
</tr>
<tr>
<td>%INCLUDE</td>
<td>YES</td>
<td>NO</td>
</tr>
<tr>
<td>HELP</td>
<td>YES</td>
<td>NO</td>
</tr>
<tr>
<td>Display Manager</td>
<td>NO</td>
<td>YES</td>
</tr>
</tbody>
</table>

Figure 3
Lilly Research Laboratories Computing Environment

Diagram showing IBM 3033, 3083 MVS, RJE (planned), VAX-11/780, tape, cluster controller, and IBM 3033 VM.
Figure 4
Example of a VAX System Performance Plot

VAX CPU Utilization

Figure 5
Migration of SAS datasets between IBM and VAX Systems

(a)
From VM (IBM) to VMS (VAX)
From VMS (VAX) to VM (IBM)

**Figure 6**
Example of a Data Analysis Output Report

**DESCRIPTIVE ANALYSIS OF FERMENTATION RESULTS**

<table>
<thead>
<tr>
<th>MODNUM</th>
<th>FERM</th>
<th>N</th>
<th>MEAN</th>
<th>STD</th>
<th>CV</th>
<th>MIN</th>
<th>MAX</th>
<th>RANGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>CONTROL</td>
<td>10</td>
<td>19.70</td>
<td>0.79</td>
<td>4.00</td>
<td>18.0</td>
<td>21.0</td>
<td>3.0</td>
</tr>
<tr>
<td>01</td>
<td>ISOLATE</td>
<td>96</td>
<td>18.60</td>
<td>1.33</td>
<td>7.15</td>
<td>10.5</td>
<td>20.5</td>
<td>10.0</td>
</tr>
<tr>
<td>02</td>
<td>CONTROL</td>
<td>10</td>
<td>19.05</td>
<td>1.28</td>
<td>6.71</td>
<td>17.0</td>
<td>21.5</td>
<td>4.5</td>
</tr>
<tr>
<td>02</td>
<td>ISOLATE</td>
<td>74</td>
<td>18.15</td>
<td>1.34</td>
<td>7.41</td>
<td>11.5</td>
<td>20.0</td>
<td>8.5</td>
</tr>
<tr>
<td>03</td>
<td>CONTROL</td>
<td>10</td>
<td>20.85</td>
<td>2.78</td>
<td>13.3</td>
<td>18.5</td>
<td>28.5</td>
<td>10.0</td>
</tr>
<tr>
<td>03</td>
<td>ISOLATE</td>
<td>54</td>
<td>19.33</td>
<td>0.95</td>
<td>4.90</td>
<td>17.0</td>
<td>24.0</td>
<td>7.0</td>
</tr>
</tbody>
</table>

724