Conversion of Running SAS® Software from OS to VMS™ on a Local Area VAX™ Cluster

M. Gary Haskin & Ronald L. Copp

ABSTRACT

The Boots Company has recently installed a local area VAXcluster for the entry and statistical analysis of clinical data. This system features a relatively small host node with the SAS system installed only on satellite workstations (VAXstation™ 2000's). This paper presents the experiences in designing and establishing the VAXcluster and the benefits which have arisen from it. Some problems and solutions in transferring SAS datasets and data entry screen catalogs over a transatlantic data link and conversion from an IBM to VAX environment are discussed.

*****************************************

The U.S. subsidiary of the Boots Company, PLC in 1985 performed all of its statistical analysis using SAS on a relatively small IBM 4381 system operating under CMS. Clinical data entry and analysis were done using the SAS FSEDIT® procedure and screen catalogs. As part of an ongoing effort to improve R&D computing support, a 56 KB satellite link to SAS on an IBM 3090Q system operating under OS AND TSO at the company's main offices in Nottingham, England was implemented. It was thought that the virtually unlimited computing resources of this system with a high-speed data link would be sufficient for long-term support of clinical data activities. Data entry on this new system was to be done online using SAS FSEDIT.

The preparation for implementation of the link involved the conversion of SAS datasets and data entry screen conversion catalogs to a form which could be transported to the new environment. At the time, all of the clinical data remained in the format of the older release, SAS 82.5. The data and catalogs needed to be converted to the Version 5 format, but other considerations kept postponing the conversion. It was known that the SAS datasets would be converted automatically when read by the XCOPY procedure. A problem which caused XCOPY to fail under certain conditions was corrected with the help of the SAS technical group. There was, however, no way to transport data entry screen catalogs in version 5 format. A solution came when it was discovered that the data entry screens under SAS 82.5 were no more than ordinary SAS datasets which could be read by any normal SAS program. These "normal" datasets were converted to transport format, shipped to the U.K. over the link, and "imported" as any other dataset. The SAS/FSP® procedure FSCNV was then used on the larger system to convert the catalogs simultaneously in to version 5 form under OS. This operation was only possible for catalogs in the pre-version 5 format. In all, over two thousand five hundred files of SAS programs, SAS datasets, and SAS screen catalogs were converted to run on the OS system.

After six months of testing, it was determined that while the link performed well for infrequent file transfers and short-term interactive computing sessions, reliability for day-to-day work on a continuous basis was not adequate. Response time was not a problem for the data entry personnel using SAS/FSEDIT since they were entering data in screen units. Statisticians and programming staff, on the other hand, were spending most of the time in text editors preparing SAS programs for running under the batch system and therefore were very sensitive to increased response time. Problems with the continuity of the satellite link were also noted. It was decided that while such a link was desirable for relatively small amounts of data transfer and irregular interactive sessions, a different system tailored to R&D requirements was needed.

In February, 1988 a VAX system consisting of a relatively small host node (VAX 8250) was installed as a data entry and administrative node in a local area VAXcluster consisting of VAXstation 2000 satellite nodes. Each VAXstation was ordered with the maximum configuration of memory (6 Mb), a large capacity disk (159 Mb), and had the SAS Version 5.16 (since updated to 5.18) system installed on the host node. For a relatively small number of users (six at the beginning), such a configuration was estimated to be as responsive as a much larger system at considerably lower cost for both hardware and software. A benefit was that as each new SAS user was added to the system, an additional VAXstation with a SAS license could be added. This was envisioned as a means to extend the life of the relatively expensive host node since upgrades would consist primarily of disk and network products rather than CPU.
enoughments. The system was configured such that each VAXstation user has the option of using the machine in either clustered or stand-alone mode. In the clustered mode, the internal disks of the satellite machines are "served" to the cluster. The stand-alone VAXstations are accessed through the SET HOST command or by DECnet. While the speed of moving data between machines is faster in the clustered mode, a VAXstation operating in the stand-alone mode is totally independent of the rest of the system. The result is faster run times with less system overhead for larger SAS jobs.

The local area VAXcluster consists of workstations connected to the host 8250 node by thinwire Ethernet. The SAS system is licensed only on the VAX stations with graphics and full-screen products in addition to the basic system. A 9600 baud packet switched line is maintained between the VAXcluster in Shreveport and the VAX and IBM systems network in Nottingham, England. Data are regularly transported among these systems via Network Job Entry (NJE) through JES2 on the IBM and DECnet on the VAXes. SAS datasets are ported to and from the IBM system through the VAX network with the aid of utility functions which allow data transport with and without translation to EBCDIC/ASCII. The VAX system in Shreveport was set up in such a way as to allow the equivalent of RJE to the VAX from the IBM. This was done by using a captive privileged account on the VAX to field a job stream into a special directory, and submit the job to run in batch under the proper user ID. This allows output of SAS jobs run on the IBM system to be imported to the VAX system.

In order to accomplish the VAX equivalent of RJE, it was necessary to modify the EBCDIC/ASCII conversion tables slightly. The logical "not" symbol in EBCDIC was aligned with the caret (shift-6) on the VAX, and the left/right square brackets on the VAX were equated to the curly brackets on the IBM. This choice has caused the loss of the ability to transmit the curly brackets, but the disadvantages of this decision have not been felt.

Binary transfers dealing with SAS datasets in transport format, it was found, must be handled carefully. The transfer of data from the VAX to the IBM is trouble-free, but the transfer from the IBM to the VAX results in a single extra record being added to the end of the file. The record is blank (not null), and sometimes causes a problem when the SAS data library is "imported" using PROC COPY. It was also discovered that the VAX version of PROC COPY suffered from a problem similar to that discovered on the IBM's PROC XCOPY. The problem exhibits itself when importing a dataset with long, character-type variables which are mostly blank resulting in an empty dataset. The problem on the VAX is fixed under SAS Version 5.18, and may be corrected under Version 5.16 with a patch.

This system configuration places increased responsibility on the VAXstation users. The current VAXstation users are SAS programmers and statisticians who had previously relied on a formal systems group for technical support dealing with the computer system. Using the VAXstation requires the user to take an active part in functions normally performed by a computer operator (such as system shutdown, booting, etc). The training of staff in these operations as well as the general care of the equipment emphasizes differences between the VAXstation and either plain terminals or personal computers (PC). The users copy data from the host node to the VAXstation and proceed to manipulate the data using SAS. Backups are done on the host node only. This makes users responsible for moving files from the VAXstation to the host node for this purpose. Performance of the SAS system is enhanced when the VAXstation is booted stand-alone thus eliminating the overhead imposed by the cluster. A windowing capability allows the user to edit and run other programs while a job is running. Even though each VAXstation is capable of running many concurrent jobs, the best performance is obtained when only a single process is actively running a SAS program.

The statisticians and programmers have had to address several items in the conversion of programs to run under SAS version 5.16 and 5.18 under VMS. The most common problem in the programs has been compensating for the lack of a macro processor. Macros had to be implemented by text substitution using the "%INC" facility. Parameter passing has been emulated by including blocks of macro variable setting statements immediately preceding the %INC statement. Conditional execution of sections of macro code has been emulated by breaking the original macro into several parts. Conventional branching is used within "NULL" data steps to accomplish the conditional code enhanced.
execution. (e.g. "IF &ABC - &X THEN &DO" is implemented as "IF "&ABC" - "&X" THEN &INC"). Other items such as multi-dimensional arrays and implicit do loops not available under the VMS versions have posed problems only when they have been used extensively in long programs.

Data entry for new studies is primarily done using a software package running on the host node. Data are converted to ASCII flat files and then to SAS datasets used by the statisticians and programmers. The data entry personnel have the ability to connect to any of the VAXstations and use the SAS/FSP for updating data converted from the OS system to the VAX. This process can be done independent of the VAXstation boot status.

The biggest difference this system arrangement has made apart from the increase in efficiency in programming has been the relationship between R&D and Corporate Information Systems (I/S). The fact that the VAX system has been clearly dedicated to R&D functions and the responsibility placed on the VAXstation users has caused R&D to be viewed as more than a group of end users by the I/S systems group. The net effect has been an increase in understanding and cooperation on the parts of each group with respect to the other groups tasks and responsibilities.

We realize that there are limits to the growth of a system designed in this way; however the maximum size of a local area VAXcluster has increased significantly with the new release of VMS. Since the cluster serves a group that is expected to remain within these practical bounds for the foreseeable future, it is felt that present limits are not a serious problem. Should the limits of the VAXcluster be approached, several alternatives are possible such as breaking up the cluster into several smaller units.

A major goal which is currently being implemented is the linking of a separate office system with the R&D network so that SAS output can be placed directly in reports. The transatlantic link will then be used to link to office automation systems in the U.K. The R&D network with the use of SAS/AF® will be used to support access to clinical data by company clinicians with the idea of developing interactive systems to assist in the submission of data to the FDA.