Eating Our Own Cooking: The Role of SAS/CPE® Software in Managing Information Systems Resources at SAS Institute

Virginia M. Dineley, Robert N. Bonham, David A. Fahey, and Daniel J. Squillace
SAS Institute Inc., Cary, North Carolina

Abstract

This paper describes how the SAS/CPE family of software products is used in managing SAS Institute's own complex networked environment, which includes IBM mainframes, VAX midrange systems and clustered workstations, and over 1,000 HP workstations and file servers. Specific examples will be drawn from the areas of ad-hoc performance studies, service level monitoring, and resource usage reporting.

Introduction

SAS Institute now provides MVS, Open Systems, and OpenVMS sites with SAS/CPE software, a comprehensive application for collecting, managing, reporting, and analyzing system and network performance and utilization data. In today's complex, multi-vendor, and distributed environment, system managers, information systems managers, application developers, resource planners, and end users all require timely information on end user service, applications performance, and resource utilization.

- Systems managers try to maintain expected service levels, identify problems and bottlenecks before they seriously impact users, and rapidly resolve serious problems when they do occur.
- Developers of large applications are concerned with efficiency and whether currently available resources will be sufficient to run the application. They also want to identify serious performance and resource utilization problems early in the development cycle when they can be more easily fixed.
- In today's fiercely competitive business environment, information systems managers are under increasing pressure to deliver more services for less money. They need to monitor application resource utilization and workload growth trends, see that resources are being used consistent with Information Systems policy, and anticipate timely and cost-effective resource upgrades.

The purpose of SAS/CPE software is to help you address these challenges by providing tools which

- Collect, reduce, analyze, correlate, and provide reporting on data from a wide variety of heterogenous sources.
- Extract and retain relevant information from potentially overwhelming volumes of data. A flexible and easily configurable Performance Data Base architecture is the key vehicle for addressing this requirement.
- Operate on multiple platforms in a consistent and intuitive manner. A standardized and tailorable menu system is provided for accessing SAS/CPE functions. As with all SAS software, a user familiar with SAS/CPE on one platform can easily learn to work with it on other platforms.
- Allow easy customization and extension to meet unique site or user requirements. An example of this is the ability to readily incorporate new or unique data sources.

We recognized, very early on, that a practical and highly useful test of the capabilities of SAS/CPE software would be to see how well it meets the needs of our own Information Systems Division (ISD). ISD is responsible for managing a highly complex, leading edge networked environment which incorporates IBM mainframes, mid-range VAX systems and clustered workstations, and well over 1000 HP workstations and file servers. Add to this a "Noah's ark" of at least of pair of workstations from every vendor of significance in the Unix marketplace.

During the past few years, SAS/CPE software has been used by the MVS, OpenVMS, and Network organizations within ISD for analysis of system performance, network activity, and disk storage allocation. This paper highlights a few of the applications of SAS/CPE using ISD data sources.

SAS/CPE and the Networking Department

Over the past several years the Institute has been pioneering methods and technologies for analyzing complex network environments simply out of necessity. Our current distributed net-

work environment consists of a centralized, highly segmented communications infrastructure composed of:

- 200 Ethernet segments
- 5 FDDI segments
- 50 Routers
- 200 Hubs
- 10 56Kb Line
- 1 T1 Network Link

From this complex environment, 300 Mbytes of performance measurement data are gathered daily. Most of the data are defined by Management Information Bases (MIBs) and are collected by using the Simple Network Management Protocol (SNMP).

Two MIBs used extensively by the Institute are the MIB-II (RFC 1213) and the RMON MIB (RFC 1271). MIB-II defines data structures to measure the usage of a device's network interface. Most devices supporting TCP/IP support the MIB-II standard including many routers, hubs, and Unix workstations.

MIB Groups	Description
MIB-II System group	The Systems group contains general information about managed objects at the interfaces layer.
MIB-II Interfaces group	The Interfaces group describes the type of interface, such as Ethernet, and provides statistics on the operations occurring at each interface.
MIB-II Internet Protocol group	The Internet Protocol group purpose is to provide information on IP operations, address tables and routing tables.
RMON Statistics group	The statistics group contains statistics measured by the probe for each monitored interface on this device. This group currentiy consists of the etherStatsTable.

Table 1: MIB-II and RMON Groups

RMON MIB describes measures for monitoring Ethernet network segments. Typically, RMON is implemented by a dedicated hardware device designed to listen to activity occurring on the wire. Often these remote probes are stand-alone devices and devote significant internal resources for the sole purpose of managing a network or may be contained as a component of a Ethernet network Hub.

Once the data have been collected, the Networking Department in ISD uses the SAS/CPE product for OpenSystems to analyze and manage network data. Typically reports are generated for the groups shown in Table 1.

Table 2 shows a configuration report of the type and quantity of interfaces for each router on the Cary network and was obtained from the MIB-II interfaces group.

Router Configuration

Router Interfaces	Interface type			
	ethernet csmacd	iso88025 tokenRi-ng	fddi	propPoi-ntToPoi-ntSerial
Machine				
Chevis.net	2	.	.	10
b-decker.net	12	.	2	.
craftsman.net	12	.	2	.
dewalt.net	12	.	2	.
hitachi.net	12	.	2	.
leo.net	12	.	2	.
makita.net	12	.	2	.
milwaukee.net	12	.	2	.
pancho.net	12	.	2	.
porter.net	12	.	2	.
ryobi.net	12	.	2	.
sears.net	12	.	2	.
trigger.net	6	4	.	.
villa.net	12	.	2	.
ALL	152	4	24	10

Table 2: Router Configuration

Table 3 displays a summary report of fourteen routers providing service to an Ethernet LAN. Several interesting trends are shown. Peak network utilization of 79% for router 'sears', output error rates > 100% for router 'pancho' and a fairly good distribution of bytes transmitted and received amongst the fourteen routers.

The SAS/CPE product is also used by the Network administrators for exception analysis. One method of exception analysis is to produce an exception matrix based on a set of user definable exception tests. An exception matrix is a count of the total number of failures for each test specified. Classification variables can be specified to further categorize where the failures are occurring. This procedure provides an excellent data exploration capability by evaluating the quality of exception tests by examining the counts of failures occurring within the data population. If a failure criteria results in 50% failures, either the criteria has been improperly selected, or there are significant performance problems being highlighted by the large number of failures.

A good example of an exception report is shown in Table 4 using several metrics derived from MIB-II interfaces group. This report analyzes data from fourteen routers providing service to a number of Ethernet LANs.

Test1 through Test3 are defined as follows:

- Test1 - Network utilization > 30%
- Test2 - Network input error pct > 5%
- Test3 - Network Output Error pct > 5%

Router Summary Performance Report

Interface activity by node	Obs	Pct Network Utilization MEAN	Pct Network Utilization MAX	Total bytes received-/sec	Total octets sent/sec	Input error pct MEAN	Output error pct MEAN
Machine							
Chevis.net	882	2%	9%	11592930	8110872	0.0%	0.0%
b-decker.net	5292	.4%	16%	11106475	13286959	0.0%	0.0%
craftsman.net	5292	1%	22%	41678169	56309381	0.0%	0.0%
dewalt.net	5292	2%	50%	59390004	65154840	0.0%	0.0%
hitachi.net	5292	2%	54%	50365539	52184515	0.7%	0.0%
leo.net	5292	.4%	34%	12881531	16679121	0.0%	0.0%
makita.net	5292	2%	29%	42948357	58628339	0.0%	0.0%
milwaukee.net	5292	.3%	24%	8783585	10455124	0.0%	0.0%
pancho.net	5292	1%	31%	54461982	35994337	0.0%	622%
porter.net	5292	2%	37%	43626289	55339702	0.0%	14%
ryobi.net	5292	1%	31%	38396997	58173692	0.0%	0.0%
sears.net	5292	2%	79%	442893351	102781475	0.0%	0.0%
trigger.net	2646	.2%	18%	51738131	26411251	0.0%	0.0%
ville.net	5292	.6%	20%	12838695	26557016	0.0%	0.0%
ALL	67032	1%	79%	4375417021	5623044951	0.1%	60%

Table 3: Summary Report

Exception Report

EXCEPTION STUDY:	Total Observ	Total Excepts	TEST1 Util>.30	TEST2 Iner>.05	TEST3 Oter>.05
Machine					
Chevis.net	882	.	.	.	.
b-decker.net	5292	.	.	.	.
craftsman.net	5292	.	.	.	.
dewalt.net	5292	3	3	.	.
hitachi.net	5292	44	22	21	1
leo.net	5292	3	3	.	.
makita.net	5292	.	.	.	.
milwaukee.net	5292	.	.	.	.
pancho.net	5292	876	2	.	874
porter.net	5292	453	11	.	442
ryobi.net	5292	3	1	.	2
sears.net	5292	61	61	.	.
trigger.net	2646	.	.	.	.
ville.net	5292	.	.	.	.
ALL	67032	1443	103	21	1319

Table 4: Exception Report

Of the three tests, Test 3 shows a very interesting measure. Both pancho and porter fail the 5% threshold of output errors to output packets 1316 times. These failures indicate a possible hardware problem on a network interface and further analysis is in order.

SAS/CPE Software for OpenSystems provides the Network administrators with the data reporting and management tools necessary to analyze network performance and plan for future network expansion.

SAS/CPE and the MVS Department

The MVS Department in ISD is responsible for management of an IBM 3090 processor along with over 150 3380 disk volumes. The user community consists of more than 2000 users working with data on these volumes, including 60,000 or more data sets. This department has found that understanding how and who is using this huge volume to data is made easier using SAS/CPE Software for MVS.

Information about the volumes and data sets is collected daily using the IBM utility program called DCOLLECT. The collected data are then added to the detail reduction level of the SAS/CPE Performance Data Base (PDB) using the Generic collector interface. Appendix 1 of "SAS/CPE Software for MVS: Usage and Reference" describes tow to construct an interface utilizing the Generic collector.

Figure 1 is an example of the type of report in use at the Institute. This graph displays the types of storage that have been allocated by data set type. It can be seen that a significant amount is allocated for Direct Access.

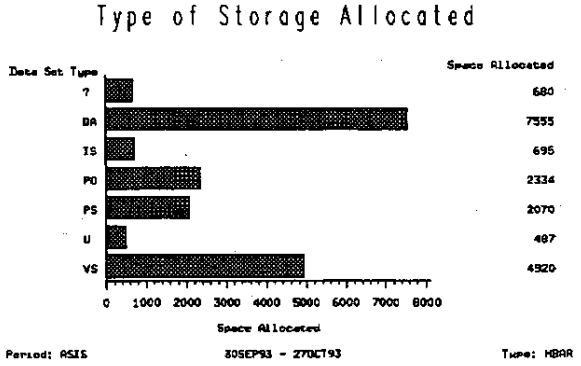

Figure 1: XUTYP Report

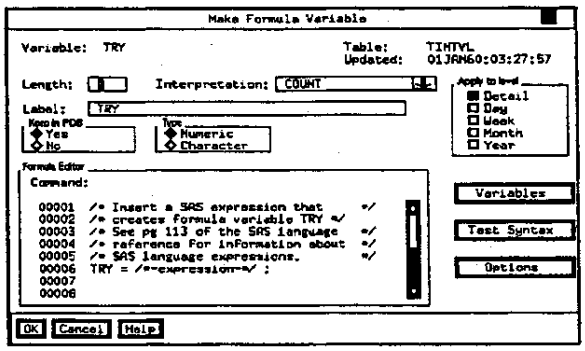

Figure 2: Make a Formula Variable

The MVS administrators were able to use the "Make a Formula Variable" window, Figure 2, to define three formulated variables; DEPT, XCOST and XWCOST. These variables are used to generate a report, Figure 3, showing storage costs by data set type for each department.

Both Figures 1 and 3 are used to summarize the costs of disk storage to both users and management.

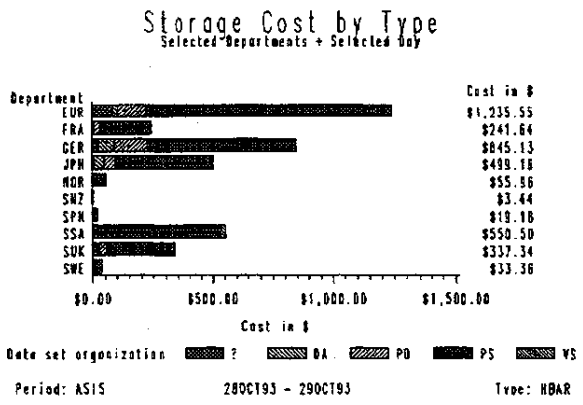

Figure 3: Storage Costs per Department

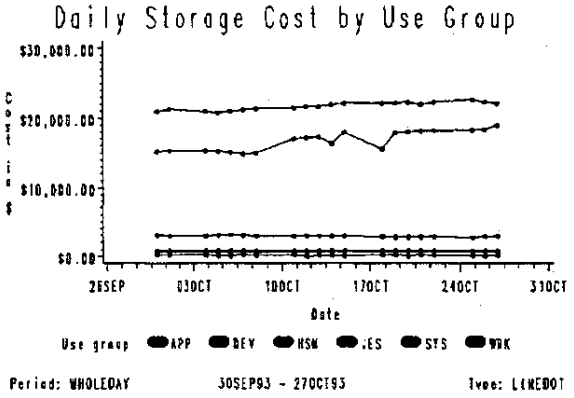

Figure 4: Daily Storage Costs

Formulated variables can also be used to plot costs of allocated storage over a month. Two formulated variables, XUSEGRP (user group) and XPSCOST (calculated rate for each user group) are used to produce the graph shown in Figure 4. This plot shows a constant rise in costs of allocated storage over the month. The same information is presented in a bar chart format in Figure 5 and uses a third formulated variable, XWEEK. This variable is created to allow grouping of data stored by days into a week format as shown.

After several weeks of data are collected, reports showing the percentage of free space and the average amount of fragmentation on the system are produced, as shown in Figure 6. In this case, the graph shows a noticeable decline in the amount of free space available on the system.

Once disturbing trends like these have been found the MVS administrators can use a SAS/CPE tabular exception report to identify users and departments that were using huge quantities

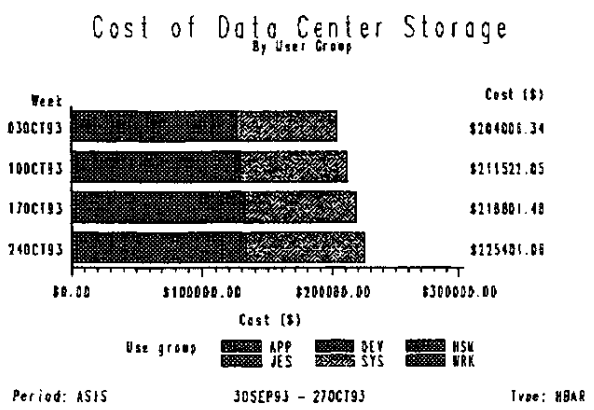

Figure 5: Cost of Data Storage

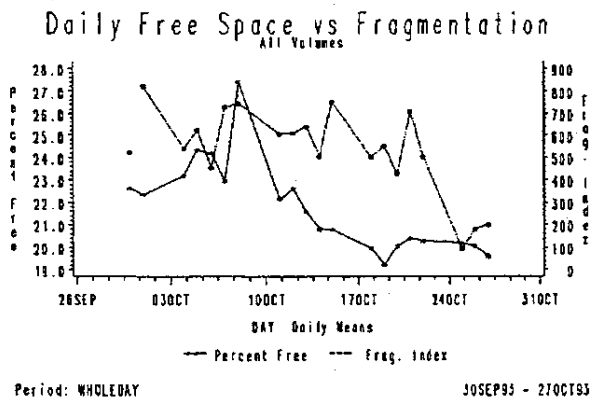

Figure 6: Free Space vs. Fragmentation

of storage, as shown in Table 5.

Next, they examined exceptions at the data set level. Table 6 shows datasets with a large number of extents. If these can be reduced performance would be improved.

SAS/CPE Software for MVS also provides easy access to data that have been collected by MXG Software. This data, combined with the data management and reporting tools in SAS/CPE Software, are used by the MVS administrators for daily and long-term performance analysis, as well as capacity planning.

EXCEPTION STUDY: High level	Total Observ	Total Excepts	TEST1 EXTENTS > 5	TEST2 SPALLOC > 5000
High level qualifier				
A608	20	20	.	20
CPD	20	20	.	20
DBI	20	20	.	20
DFHSM	20	20	.	20
ETS	20	20	.	20
EURCDW	20	9	.	9
EURHXM	20	20	.	20
EURKPH	20	15	.	15
EURLGR	20	19	.	19
EURMIS	20	3	.	3
EURMJW	20	20	.	20
EURPJW2	20	18	.	18
EURSXM	20	7	.	7
EURWEB	20	8	.	8
FRARAV	20	5	.	5
SASMKT	20	37	17	20
SDKSTV	20	17	17	.
ALL	66582	1999	34	1965

Table 5: Exception Report

	Extents	Secondary	Primary	Used
EURHAK.LOG.SEA	7	139	5378	5378
GEREUR.WEEKLY.LIST	16	46	788	788
GERRUS.MSPF2.SPFLOG2.LIST	8	232	1808	1669
GERSUV.MSPF2.SPFLOG1.LIST	6	232	1437	1205
GERTHD.LOG.HELP-NET	6	139	974	974
GERTHD.LOG.W3L	13	464	2040	2040
SASADM2.EDSALES.MASSDB	16	1391	127545	90965
SASADM3.AP.LIBRARY	7	6954	180817	179936
SASADM3.EQUIP.LIBRARY	6	6954	124485	120869
.				
.				
.				
SASTSD.TRACK.REPVSAM.DATA	8	1391	23645	0
SASTV.TVLIB.DATA	9	1391	17966	16042
SASTXD.JOBS.MONTH	13	695	15300	15207
SASTXD.MYSTEST.PDBWORK	9	695	6954	6908
SUKKDJ.PK	7	46	325	325
SUKKDJ.ZOOM2	11	46	510	510
SUKKDJ.ZOOM3	11	46	510	510
SUKKDJ.ZOOM4	8	46	371	371
SUKKDJ.ZOOM5	7	46	325	325
SUKMAF.ADDRESS.INFO	9	139	4126	4126
SUKMJH.WIN125	9	46	417	417
SUKSTR.SAS6.SASUSER	8	232	6398	6305

Table 6: Datasets with Excessive Extents

SAS/CPE for OpenVMS

SAS/CPE Software for OpenVMS provides data management and extensive reporting capabilities for collectors developed and distributed by Digital Equipment Corporation:

- OpenVMS Monitor facility
- OpenVMS Accounting utility
- VAX Software Performance Monitor (SPM)
- Polycenter Performance Solution (PPS) Software

In addition, SAS/CPE Software for OpenVMS includes facilities to:

- Collect and analyze diskusage information (DISKUSAGE)
- Collect and analyze Ethernet activity (ETHERNET)

The OpenVMS Management Department in ISD is responsible for managing a mixed AXP and VAX cluster consisting of the following:

- 13 VAX 3100
- 2 VAX 4090
- 3 VAX 4XXX
- 1 VAX 6550
- 10 AXP 4xxx
- 1 AXP 7000
- 195 disk drives of various types

This cluster is used by over 1000 users, who have been divided into 48 groups for management purposes.

For performance analysis of the cluster and network, and analysis of disk utilization, the OpenVMS administrators make use of the SAS/CPE product to collect and analyze several different types of data, including data that have been collected by the Data Collector portion of Polycenter Performance Solution (PPS), the SAS/CPE Diskusage Facility, and the SAS/CPE Ethernet Facility.

PPS data is collected daily at two minute intervals. At midnight, these data are read by SAS/CPE Software and collapsed to 15 minute intervals. The collapsed data are used to generate daily reports. These reports are used by OpenVMS administrators to quickly review performance of the system. Over 100 supplied reports are provided to quickly analyze memory and file system performance, as well as disk and page file utilization.

Analysis of the responsiveness of the memory management subsystem begins with the graph shown in Figure 7. This graph shows mean page faults per second by type. From this graph, the OpenVMS administrators can determine if there has been excessive faulting. An excessive rate of free list faults and modified list faults can indicate that working sets are too small and should be adjusted.

A high level of demand zero faults can indicate a high number of image activations. This can be confirmed by reviewing a graph of image activations over time, as shown in Figure 8.

After daily reports are generated the collection is collapsed to 30 minute intervals and added to a group. Typically a group contains data collected for a month. This data can be combined with other groups to predict long-term trends and assist in capacity planning.

The OpenVMS administrators are responsible for managing 195 disk devices of various types. It is their responsibility to ensure a sufficient amount of disk space is available to the different groups within the user community at all times. This responsibility requires a knowledge of disk utilization by each user group over time combined with a knowledge of the

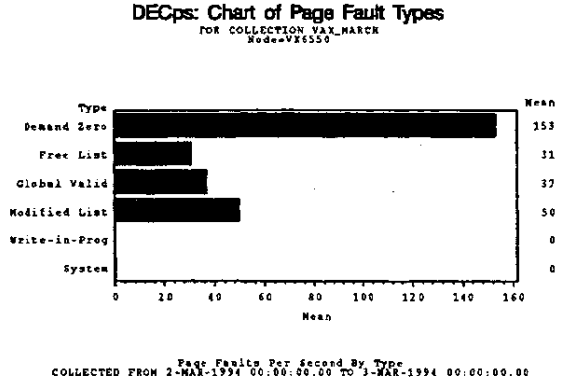

Figure 7: Mean Page Faults by Type

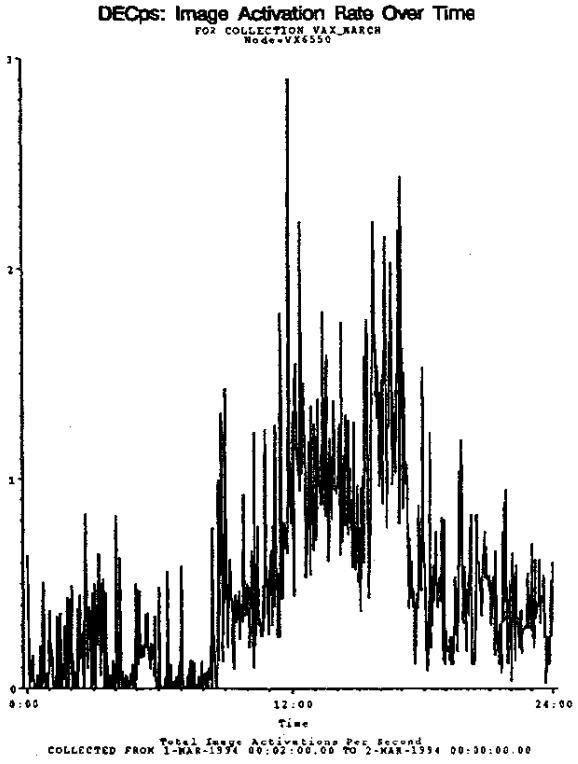

Figure 8: Image Activations over Time

expected increase or decrease of utilization as projects begin or near completion.

The SAS/CPE Diskusage Facility is used to collect disk space utilization statistics over time. Reports on disk utilization by UIC, UIC group, volume and directory can then be produced.

Several graphs have been developed to analyze current usage and predict future utilization. These graphs show the percentage of free and allocated space over time per volume. Figure 9 shows the percentage of free space for volume CMGT1. This volume is used by the Code Management group at SAS Institute. Currently free space on this volume is over 30%, but the graph shows that in January the percentage of free space dropped to 15%. This corresponds to an increas in the number of SAS source and image levels that were contained on this disk at that time. Knowing that the number of levels was due to increase as the Code Managers geared up for another maintenance level and Version 7 development, the OpenVMS administrators reorganized the current levels and moved some of the existing levels to volume CMGT2. This enabled the OpenVMS administrators to ensure that the Code Management group will not be short of disk space at a critical time in the development cycle.

The SAS/CPE Ethernet facility is used by OpenVMS administrators to monitor network performance. The Ethernet facility data collector gathers information on network traffic by node, send/receive pair, and protocol activity. The Ethernet facility is used to gather data daily during peak hours. Supplied reports are generated showing traffic loads, top nodes by activity, network traffic rates over time, etc.

Figure 10 shows a graph of traffic byte rate over time. This shows peak network usage time. It is obvious by looking at this graph that the network is heavily used right before lunch. Usage drops considerably during lunch hours and picks up again right after lunch. Activity slows again around 2:00 p.m. This is probably due to the preponderance of meetings scheduled for the late afternoon. Typically, the data can be correlated with network performance fluctuations.

The daily Ethernet data is then combined with previously collected daily data and graphs are generated to show general trends. If it is detected that the traffic byte rate consistently spikes at 11:00 a.m., then the administrators would further investigate the types activity occurring at or around 11:00. It is possible that this activity is caused by batch jobs which can be run at off-peak hours to reduce workload.

Another graph the OpenVMS administrators find useful is a chart of message rates by protocol. An example of this chart, shown in Figure 11, shows that DECnet protocol has the highest message rate. This indicates that the traffic byte rate shown in Figure 10 is generated by DECnet traffic. DECnet traffic is typically caused by intra-OpenVMS communication.

The analysis capabilities provided with SAS/CPE Software enable the OpenVMS administrators to not only determine the cause of problems that occur today, but prevent problems from occurring in the future.

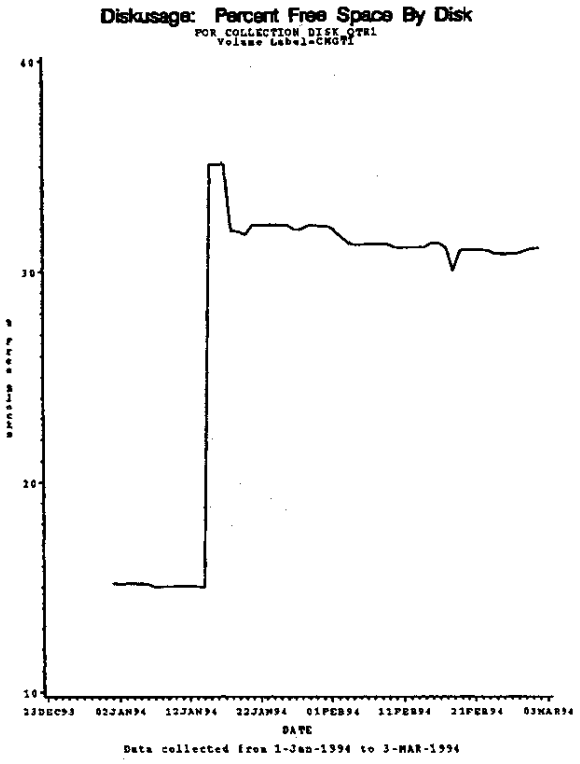

Figure 9: Percent Free Space by Disk

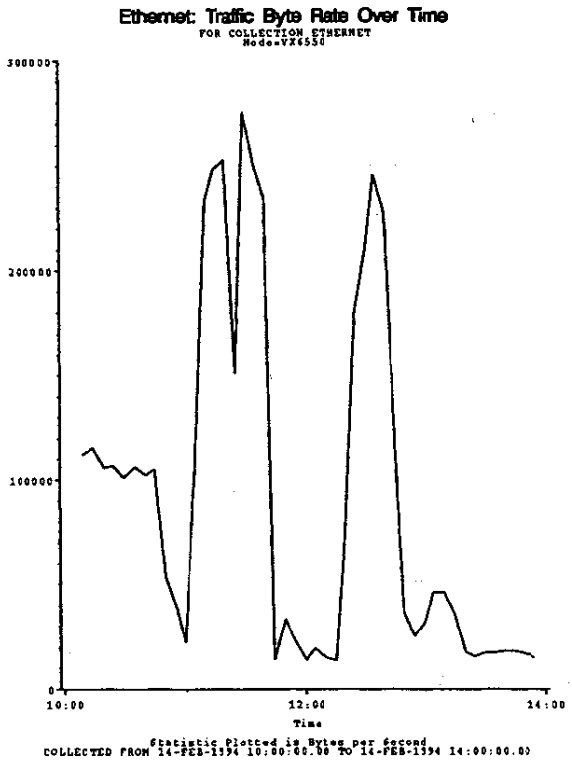

Figure 10: Traffic Byte Rate

Using a variety of examples drawn from our complex multi-vendor networked environment, we have demonstrated ways that SAS/CPE software can meet the needs of today's complex information systems organizations.

Trademarks

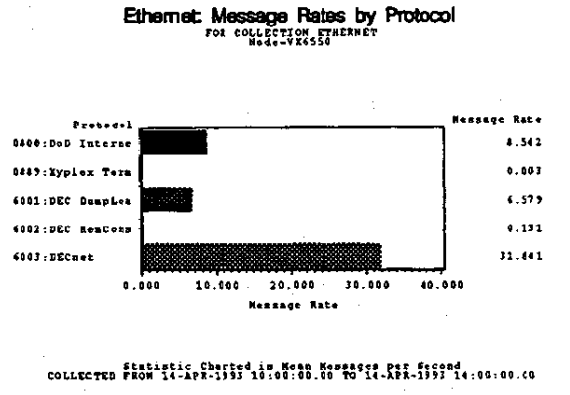

Figure 11: Protocol Traffic by Type