Integrating Windows™ Clients and the SAS® System into the Enterprise

Gary Mehler, SAS Institute Inc., Cary, NC

Abstract

To utilize the full potential of SAS System on Microsoft Window, your PC should be fully integrated into your enterprise. SAS System products for the PC provide many client/server architectures that Information System designers can use to access, update and analyze the entire spectrum of enterprise data. SAS System client/server products are introduced in different PC based scenarios and several PC, network and SAS System configuration suggestions are presented to help you get better returns from your hardware and software investment.

Introduction

Organizations are extending the effective reach of their centralized systems by interfacing workstations to large PC, mainframe, and UNIX server systems. Many of these organizations are planning on using Windows, Windows NT, Windows 95, and UNIX platforms for the development of new applications.

Version 6.12 of the SAS System delivers a synchronized release of advanced features to all of these workstation platforms. This release provides a seamless computing environment on these platforms, allowing UNIX and Windows NT server based servers to be used with Windows 95, Windows NT, and Win32S client machines.

One way to integrate these platforms is through client/server computing, which is described below. Data warehousing is a common implementation, and one for which the SAS System is uniquely qualified. In addition, PC-based systems have a wide range of networking configurations which can help meet information sharing needs.

Other important enterprise considerations include performance optimization and software installation/maintenance issues.

SAS System Client/Server Support

The SAS System for Personal Computers provides a window to your enterprise data and offers cooperative processing configurations that can access the diverse and scattered data formats while making optimal use of the wide range of enterprise hardware. The following SAS System features and products form the basis for these client/server computing solutions.

MultiVendor Architecture™

Version 6 of the SAS System is available for many types of computer hardware and operating system combinations including: MVS, VM/CMS, VSE, OpenVMS VAX, OpenVMS AXP, UNIX, OS/2®, Windows 3.1/Win32s, Windows NT, Windows 95 and Macintosh. The compatibility in terms of SAS programs, user interface, and data transportability make inter-system use of the SAS System on various platforms an attractive foundation for enterprise architecture. Your SAS applications can communicate with each other to exchange data no matter where they are located. You can develop and test applications in one environment and move them to another with minimal modifications. And you can cooperatively run applications across multiple platforms.

The central technology in the SAS System's data access strategy is the *Multiple Engine Architecture* modular design. This architecture allows the SAS System to transparently access all data, regardless of vendor, format or location, though loadable *engines*. Engines provide the framework for translating read, write and search calls to native SAS, OS file formats or external database management systems. Engines can surface data sets as *views* (data about how to access the data) or actual data files.

ODBC Support

Microsoft Corporation's Open Database Connectivity (ODBC) standard, part of the Microsoft Windows Open System Architecture (WOSA) umbrella, is used as a call level API to numerous DBMS formats from the Windows operating system. ODBC has significant DBMS market penetration and ISV support and is being ported by various vendors to OS/2, UNIX and the Macintosh.

SAS System client/server products extend the advantage of SAS Multiple Engine Architecture design to third party applications that support the Microsoft ODBC database standard. The SAS ODBC Driver allows ODBC applications to connect to a local SAS ODBC server and to remote SAS/SHARE servers. SAS/ACCESS to ODBC allows the SAS System for Windows to access third party DBMS data sources.

Implementing a Data Warehouse

Customers are moving from the mainframe batch and OLTP environments down to client/server and desktop analysis. SAS Institute Marketing has estimated that 95% of IT budgets are allocated to collecting data and only 5% to analyzing the data. Data warehousing empowers end users and gets the IT center out of the business of providing custom reports as well as getting end users off the mainframe. Offloading provides better on-line transaction processing performance where response time is crucial.

A data warehouse contains data physically separated from an organization's day to day operational data. The data warehouse includes decision support data that has been collected, cleaned and summarized from isolated departmental data across the enterprise in order to provide more meaningful, accessible information for modeling, forecasting and analysis by the organization decision makers. The data warehouse is designed to be read from and not written to because the end users are making decisions on existing data not entering data. The three most critical elements of a data warehouse are: easily accessing the data, transforming and managing the data and analyzing the data in a friendly manner.

Using the SAS System for Data Warehousing

The SAS System has strengths in all areas of data warehousing and various SAS components can be intermixed with other vendor products that are already a part of the enterprise. One of the primary activities of the IT center should be time spent transforming OLTP data into sensible and meaningful formats. The SAS System provides many tools for data collection, transformation, reduction, modeling and visualization.

SAS/ACCESS and it's various engines can be used to extract data from the different OLTP and archive structures and computing environments which include SQL accessible relational databases as well as many non relational legacy systems where it is estimated near 70% of corporate data still resides. SAS/ACCESS can take advantage of Stored Procedures to process and subset data at the OLTP server to help minimize network traffic. Popular base SAS routines can be used to clean, summarize and otherwise transform

data into more user friendly formats used to stock the data warehouse locations after collecting transaction level data from diverse data sources.

SAS/CONNECT remote submit offers application partitioning which can be used to remotely process queries against the warehouse and can keep users from flooding the network with data. SAS/CONNECT can be used to transfer data to more localized servers or local SAS data sets from different architecture repositories. This allows better performance when large numbers of users around the enterprise want ad-hoc query access to the same information. SAS/CONNECT can be driven from the desktop or from the central location when transferring the data. The SAS ODBC Driver can be used to give other third party desktop analysis tools access to the SAS data on the local or server platforms. SAS/SHARE servers can be positioned at strategic points in the enterprise and can provide SAS and ODBC clients shared read/write access to SAS data.

The SAS Domain Server can bridge protocol domains and simplify and reduce costs for configuring and administering PC clients.

SAS/AF® and other SAS System analysis products can be used to build custom front ends and user interfaces to the data warehouse and which present the data using the latest object oriented analysis techniques.

Using Windows NT for Data Warehousing

Release 6.11 of the SAS System operating under Windows NT can be a good server platform for a data warehouse application. This platform is not inhibited by a two gigabyte file size (and therefore data set) limitation. Instead, maximum data set size is limited only by the available disk space on the Windows NT machine.

Since Windows NT also supports software disk striping, multiple hard drives (currently available in sizes up to nine gigabytes) can be configured into a very large, parallel access disk array, usually limited only by the number of drives that can be accommodated in the PC cabinet. Using nine gigabyte drives, a system with seven available full-height drive bays could be used as a 60 gigabyte data store.

In addition to NT's ability to handle very large files, it also supports symmetric multiprocessing (SMP) to provide an additional performance boost. A system with four CPU's can support nearly four times the computing load as a comparable single CPU system. This can be helpful in configurations in which SAS/SHARE servers and Remote Computing services are used.

PC-based Data Sharing

In most cases, the SAS System is used as what can be called a "*fat client*". The SAS System software acts as the client interface, which provides great power on client systems, but at a cost of disk space and processing power requirements.

PC networks support various ways to allow access to SAS data stored in remote locations. Keep in mind that most of these options allow multiple read/only sessions or a single read/write session when accessing a centrally stored catalog or data set of the SAS System. A SAS/SHARE server is **always** needed in order to provide access for shared read/write sessions to native SAS data for multiple clients whether on the same or different machines.

Defining indexes on SAS data sets can significantly improve data access performance and is especially recommended for large files maintained on file servers and for SAS data sets accessed via Remote Library Services. When used in conjunction with SAS WHERE clause processing, the SAS System can use an index to directly locate the requested records greatly reducing CPU time, disk I/O and network traffic from file or application servers. Indexes are not needed for data sets always accessed sequentially.

Various portions of the computing network can provide services for desktop personal computer client systems running Windows and the SAS System. Although a wide range of configurations are possible, some do not make efficient use of computing and network hardware. Operation of SAS System products in some typical enterprise scenarios including certain advantages and disadvantages follow.

PC File Server Access

Running the SAS System from the local hard drive and accessing only local data sets is the simplest computing model. A local configuration normally allows for the best performance of the SAS System but requires the most disk space on each workstation. Usually, SAS System data is processed locally in conjunction with data sets stored remotely or downloaded from or uploaded to other data sources in the enterprise.

The simplest client/server scenario is desktop PC using a PC-based file server to store SAS data sets. It is also reasonable to access the SAS System itself as well as SAS data sets from network file servers rather than from the local hard drive when local hard drive resources are small or when shared read/only access is desired for easy installation and maintenance.

There are various options concerning the actual distribution of programs and data between the local machine and the remote server. The SAS System Companion manual for Windows contain configuration suggestions for installing onto a file server. Some of these configurations include:

- Running the SAS System from the local hard drive and accessing data sets from a network file server. For large amounts of data, this can be an attractive solution as long as the SAS data is indexed. However, if you make multiple random passes of non-indexed data residing on a file server you will flood the network and degrade everyone's network performance.

- Also common is installing from or running the SAS System from a network file server and accessing data sets from the local hard drive. If quick access to data is required, this may work best. Since the SAS System is a large application, accessing it entirely from a file server saves local disk space but can be time consuming if SAS System sessions are frequently started as would be the case for a large script of SAS batch programs.

- Running both the SAS System and accessing SAS data sets from a network file server minimizes the desktop disk space requirement, but can place a high load on the PC network.

- You can determine which files are required during an often executed SAS session by using the RTRACE and RTRACELOC system options and the COPYSAS autocall macro as documented in the Release 6.12 "Changes and Enhancements to the SAS System". Performance is increased but usage of local disk space is still minimized by not having all the SAS products installed locally.

- Available in Release 6.12 of the SAS System is a network install that allows client machines to access SAS Software already installed on the network. Network setup adds OLE and SAS file type registrations to the client PC's registry, and folder and icon additions are made to the desktop. This type of install is also useful to run from a CD or a previously installed local image when running other Windows OS systems on the same machine. At the Institute, we often run Windows 3.1/Win32s, Windows NT and Windows 95 on the same machine.

Using Another Desktop PC as File Server

Peer-to-peer networking packages, such as Windows for Workgroups, Windows NT and Windows 95 enable any PC to act

as a file server to other PCs which become clients to that PC. While an extremely cost-effective solution for data sharing, this approach does not handle a large number of client PCs. Since the PC acting as a file server is typically an ordinary desktop PC, even a small number of clients can significantly degrade the performance of the desktop PC acting as the file server.

PC-based Thin Clients

In this type of configuration, the SAS System, your application code, and all data reside on the NT server. The client machine runs an NT emulation program, and all processing occurs on the NT server. Note that no SAS software needs to be licensed on the client workstations in this configuration. Using commands like **rcmd**, or remote console programs like **pcAnywhere32** or **WinFrame**, remote sessions can be viewed on the client PCs while all processing actually occurs on the server system.

This approach can be cost-effective when the cost of SAS System licenses for all client systems becomes an issue, and high-powered centralized Windows NT servers are available. When the NT servers are SMP (multi-processor) machines, several users can be supported by each NT server. However, since all users are actually sharing a limited pool of resources on the server, the type of application being run by each user needs to be considered. Another important factor is network bandwidth requirements. Since all input and output to the screen is being transmitted across the network, performance of highly-interactive applications may suffer, particularly if the network is heavily utilized.

SAS System Support for Client/Server Applications

Various products and features of the SAS System enable the development of distributed applications. These configurations could be called **cooperative client** configurations, since both client and server PCs can process the data. These are described below.

SAS Application Servers

SAS/CONNECT and SAS/SHARE software extend the file server model and allow an ODBC or SAS System client on one PC remote access to another SAS System running on a different Windows PC, or on a UNIX or mainframe computer. Remote computers acting as SAS application servers are normally more powerful than PC based systems or have unique and desired hardware or software capabilities. The SAS System client/server components communicate using communications protocols that we refer to as SAS Communication Access Methods or just SAS Access Methods. Appendix A lists the peer to peer SAS Access Methods supported by each SAS System product.

Remote Library Services

The Remote Library Services (RLS) Remote engine provided by SAS/CONNECT or SAS/SHARE software allows SAS System client applications on one hardware/OS platform access to remote SAS data libraries on any other hardware/OS platform even when the machines use different internal data representations. RLS can also provide gateway access to external DBMS systems that the server can access using SAS/ACCESS engines. You can also use RLS with all three types of SAS data views.

When using RLS, one way to reduce the amount of data that moves over the network is to use the WHERE clause on your local data steps and procedures. Another is to use an index on variables that are likely to be used in WHERE clauses to help the data set engine directly locate data.

A SAS System application for data entry or retrieval that allows PCs to access or update data stored on a central computer system works well if relatively small amounts of data are accessed or updated infrequently. An application that accesses a large portion of the data multiple times would generate a large amount of network traffic that would slow completion of the task. Excessive or unneeded network traffic also slows the operation of other network users. SAS/CONNECT Data Transfer Services or Compute Services would make better use of computing and network resources when processing large amounts of data or making multiple passes of the data.

Either the SAS/SHARE or SAS/CONNECT product will supply the Remote engine on the SAS System client machine to enable access to a single user SAS/CONNECT server or to a multi-user SAS/SHARE server.

SAS/SHARE

The SAS/SHARE product includes the client side Remote engine as well as a multi-user server that can be accessed from any SAS System client running the Remote engine. Multiple SAS RLS clients can access a SAS/SHARE server which manages libraries of SAS data sets and catalogs in shared read/write mode. The SAS/SHARE server can provide read/only access to external DBMS systems that are accessed using SAS/ACCESS engines. A multi-user SAS/SHARE server allows you to develop basic on-line transaction processing systems.

A separate SAS process running PROC OPERATE on the same or another machine is used to control a SAS/SHARE server including user access to the SAS/SHARE server libraries. ODBC and SAS clients on the same machine as a SAS/SHARE server can still access the server though not actually remote. Both client and server sides of SAS/SHARE software are available for all computers supported by the Release 6.12 SAS System.

SAS/CONNECT

SAS/CONNECT software provides several client/server features. These include the single user server, the RLS Remote engine, Data Transfer Services and Compute Services.

All features of SAS/CONNECT can be used concurrently to one or more remote platforms. Both client and server sides of SAS/CONNECT software are available for all computers supported by the SAS System. The single user server is not available for Windows 3.1/Win32s or for the Macintosh due to OS restrictions.

SAS/CONNECT Single User Server

The SAS/CONNECT product provides a single user server that can be accessed from any SAS System client running the Remote engine. Each single user server provides one SAS RLS client with shared read/only or exclusive read/write access to SAS data on the server platform.

The SAS/CONNECT single user server can provide read/write access to external DBMS systems that are accessed using SAS/ACCESS engines. Multiple SAS/CONNECT single user servers can be run on all SAS platforms except Windows 3.1/Win32s where only a single instance of SAS is allowed to run.

SAS/CONNECT Data Transfer Services

If batch mode processing is to be performed on a remote data set, or if data will be accessed multiple times, the data can be downloaded from the remote server to the client machine for local processing to minimize redundant network usage. SAS data sets, catalogs, graphics and system files can be transported in either direction between any SAS System platform using commands available in the SAS/CONNECT product, which takes care of all data set and catalog translation between platforms. Data can be selected and subsetted during transfer using WHERE clauses. These transfers are performed through PROC UPLOAD and PROC DOWNLOAD. The Query Window component of SAS/ASSIST® provides an

interactive graphical interface for moving SAS files using Data Transfer Services.

You can also address access to large amounts of data using various data warehousing techniques. For example, SAS/CONNECT could download local summarized copies of the data to individual machines or network servers nightly from a mainframe for next day processing by local SAS Systems. It would be inefficient, however, if the actual use was on a small subset of the overall data transferred. The time required to transfer the entire data set would be large relative to the amount of useful data really transferred. In order to minimize the amount of network traffic generated by large transfers, you may want to download a subset of the data for processing using WHERE clauses or summarizing the data prior to downloading. If the data needed to be later updated to the central location, a combination of SAS/CONNECT Compute and Data Transfer Services can be used to upload and merge the subset data. Or you could directly merge the changes back into the original data set with Remote Library Services.

SAS/CONNECT Compute Services

SAS statements and programs can be forwarded to a SAS System running on another computer using SAS/CONNECT software. As remote processing progresses, the program listing and any graphical output are presented on the local system as if the program had executed locally. This allows compute or data intensive processing to occur on the most capable computing resource or at the location where large volumes of data reside. Unlike SQL Stored Procedures which are coded to a specific RDBMS and so are not portable, SAS programs can be relocated to a wide variety of server platforms due to the MultiVendor Architecture design.

Compute Services are especially useful where large amounts or frequently updated data reside on the remote server. In these cases it would not be practical to download that data or to remotely analyze it via Remote Library Services due to the tremendous network traffic that would be required to get data records to the local client machine.

SAS/ACCESS

SAS/ACCESS engines provide the ability to access many different DBMS and system file formats including text, VSAM, Access, dBase, DB2, SYBASE, SQL Server and ORACLE. In addition, when SAS/ACCESS engines are used by a local ODBC server, a SAS/CONNECT single user server or a SAS/SHARE multi-user server, the ODBC and SAS clients on one platform can use the server as a gateway to access external database systems. The database systems can be local or remote to the SAS application or SAS server that is using the SAS/ACCESS engine.

SAS/ACCESS provides the middleware for SQL Pass-Through connectivity to remote RDBMS systems which can process a SQL request and return result sets back to the SAS System. This architecture is best for activities which return a relatively small amount of data such as running SQL summaries or Stored Procedures that subset the data to be returned. If large amounts of data are requested, too much data is passed over the network impacting network performance as well as moving processing to the workstation impacting it's performance.

SAS/ACCESS to ODBC

SAS/ACCESS to ODBC bears special mention since it allows the SAS System on Windows open access to numerous DBMS formats from a wide range of vendors. For example, the SQL Pass-Through facility and the ODBC engine allow SQL queries and other transactions to be executed on databases such as dBase, Paradox, Access, Excel, ORACLE and SYBASE.

SAS ODBC Driver

The SAS ODBC Driver for the SAS System allows third party ODBC compliant Windows applications, such as Microsoft's Excel or Visual Basic, open access to SAS data sets. The SAS ODBC Driver is a multi-tiered driver meaning that it does not directly access SAS data sets but rather communicates with either a local SAS ODBC server or a remote SAS/SHARE server. The SAS ODBC Driver does not connect to a SAS/CONNECT single user server.

SAS System servers transparently handle all cross architecture issues such as floating point data translation between different hardware/OS platforms. The same SAS ODBC Driver comes included with either the base SAS System or with SAS/SHARE*NET software. The local SAS ODBC server also comes with the base SAS System and allows access to local SAS data. To access SAS data remote from the local machine, you use the SAS ODBC Driver to connect to a SAS/SHARE server that has the SAS/SHARE*NET software installed.

When used in conjunction with SAS/ACCESS software on the local or remote server machines, the SAS ODBC Driver serves as a read/only gateway to the wide variety of DBMS data supported by the various SAS/ACCESS engines. The SAS ODBC Driver supports both local and network DDE and TCP/IP for connection to local ODBC or remote SAS/SHARE servers. SAS Institute will expand the available access methods for the SAS ODBC Driver in the future.

SAS/SHARE*NET

SAS/SHARE*NET software is added to a SAS/SHARE server to allow third party ODBC compliant Windows applications, such as Microsoft's Excel or Visual Basic, open access to remote SAS System data sets in shared read/write mode. When used in conjunction with SAS/ACCESS software on the server machine, the SAS/SHARE server becomes a read/only gateway for the ODBC client applications to the wide variety of data supported by the various SAS/ACCESS engines.

SAS Domain Server

The SAS DOMAIN Server protocol gateway service allows message exchange between SAS sessions that run in different network environments and use different SAS access methods. The goal of this technology is to deliver full, enterprise-wide, protocol independent messaging between all of the SAS client/server products.

The SAS Domain Server can currently be used by either SAS/SHARE or SAS/CONNECT clients to send requests from a network serviced by one protocol to a remote SAS/SHARE or SAS/CONNECT server running on a platform serviced by a different protocol. An example of using the SAS Domain Server would be a Windows SAS/CONNECT client using the TCP/IP protocol to access a SAS/SHARE server on MVS using the APPC protocol. The SAS Domain Server allows the Windows machines that already have the TCP/IP protocol installed to use Remote Library Services on the MVS server without licensing or maintaining APPC for each client machine. A cost-savings advantage can be realized with this configuration because it eliminates the requirement to license, configure, and support both TCP/IP and APPC protocol stacks for each Windows desktop in the enterprise.

SAS System OLE Interoperability

The 6.12 release of the SAS System for Windows fully exploits Microsoft OLE 2.0 functionality which is actually a very sophisticated form of client/server computing. The SAS System for Windows supports the following OLE 2.0 features: drag and drop, visual editing, OLE automation, OLE custom controls, and an OLE automation server.

For a more complete discussion of the OLE support available in the SAS System, please refer to the paper, "OLE and the SAS System for Windows Release 6.12," also in these proceedings.

Other SAS System Interoperability Support

Several new methods are available in Release 6.12 of the SAS System to allow enhanced inter-program operation and data sharing.

The SAS System's support **Lotus Notes** enables users to populate Notes with SAS output including attachments such as data sets, graphs, and reports.

Electronic mail can be sent from within the SAS System either by using the Send Mail dialog (available from the main SAS System **File** pull-down menu), or by using the **EMAIL access method** from a DATA step or SCL code. The SAS System supports MAPI and VIM e-mail client packages and supports SMTP mail delivery directly to central mail servers.

Three access methods available in Release 6.12 provide increased cross-platform data interoperability. The communications protocols can replace normal external file processing, so they can be used anywhere an external file would be accessed from the SAS System. The FILENAME statement can be used to indicate that you want to use an alternate external data source access method. These access methods include:

- The **FTP access method** can be used to read and write remote data files using the industry-standard FTP (file transfer protocol). This method can be used to access directory listings, access anonymous FTP sites, retrieving and running SAS programs from a remote system, as well as transport SAS data libraries.

- The **TCP/IP socket access method** can be used to establish peer-to-peer connections with remote applications. A SAS application can act as a SOCKET client (and attach to a remote application acting as a server and waiting for a connection) or as a SOCKET server, in a wait state waiting for a remote application to connect to it.

- The **URL access method** allows a SAS application to access remote files using a URL (universal resource locator) and a WWW (world wide web) server. An option is used to specify the URL when the FILENAME statement is entered.

SAS System integration with the Windows 95 Explorer allows Windows 95 to understand different types of SAS System files (including SAS data sets, catalogs, and SAS System source code), which are represented by unique icons. Each file type is also associated with default actions that can be accessed through the Explorer's pop-up menus.

The SAS System also supports long filenames, which are supported by many other applications under Windows 95 and Windows NT. These filenames, which can be up to 256 characters and can include embedded space and separator characters, make file system interaction more intuitive and inter-application data sharing transparent.

Finally, Release 6.12's ability to access external dynamic link libraries (DLLs) from within the SAS System can extend your processing power. Functions contained in custom DLLs developed in-house, or system or third-party DLLs can be called from within the SAS System by experienced programmers. The external DLL can be loaded, routines invoked, and the DLL unloaded upon completion.

For a more complete discussion of the new features available in the SAS System, please refer to the paper, "Introducing Release 6.12 of the SAS System for Personal Computers," also in these proceedings.

Performance Considerations

When you select new machines, there are certain features that can have a major impact on both desktop and server machine performance. Some of these are briefly described below:

Multiple processors are a must when multiple applications or users will be supported by the same hardware. Depending on your application, two users running two applications can be serviced as quickly as one, when running on a dual-processor NT system. A four processor SMP system should handle up to four users without any performance degradation, although other factors such as disk channel saturation can become an issue for some types of applications.
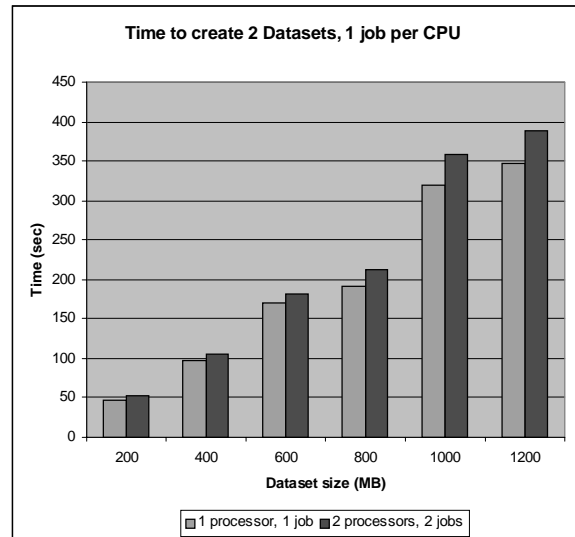

Figure 1. When running one job per CPU, performance varies little until large amounts of data are processed and disk write caches become too busy to remain effective.
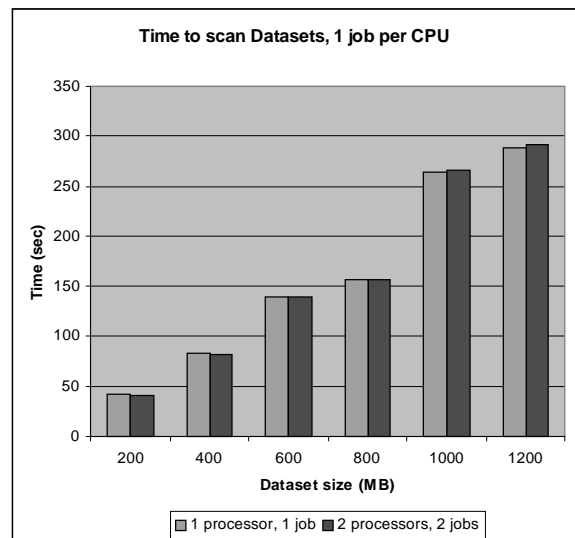

Figure 2. Read performance is consistent for one job per processor.

CPU type and speed are very important. The SAS System is optimized for Pentium (and later) processors, so a performance gain can be realized when using these processors. Keep in mind that internal chip optimizations can make a "newer" CPU faster than an older CPU with the same speed rating. A 200MHz Pentium Pro system should outperform a 200MHz Pentium system. Of course, you should always have enough cache memory as possible to keep your CPU running as quickly as possible. A Pentium Pro system with a 512KB L2 cache should be a minimum for heavy application requirements, and 1MB L2 cache systems will also help performance when they become generally available.

Disk Striping (RAID 0) can dramatically improve performance. Hardware or software approaches to paralleling disk access by decreasing effective disk wait periods. Depending on your hardware configuration, a dual drive configuration can cut the time required to write data by 50%, up to a maximum benefit of 75% with 3-4 drives.
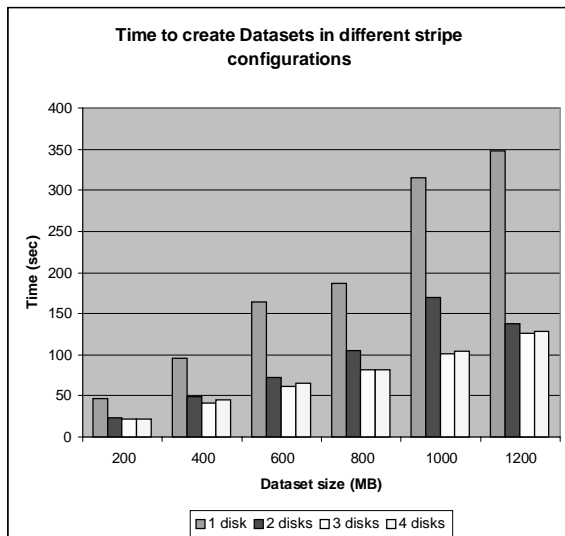

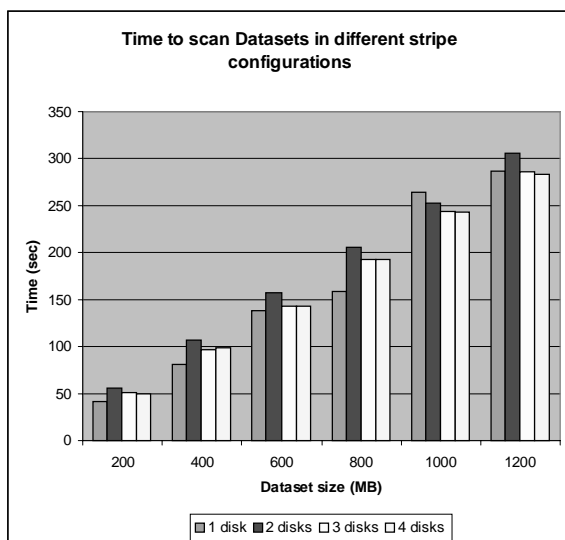Figure 3. Any disk striping dramatically improves disk write performance

Figure 4. Disk striping doesn't have a great effect on read performance.

Disk Format can effect performance. Although NTFS is the most reliable and expandable type of disk format, its features come at a cost. For smaller drives, old-style FAT formatted disks can provide better performance. For some types of applications, operations like writing to a dataset can be improved by 20% by using FAT instead of NTFS. This benefit also occurs on FAT32-formatted disks available beginning with Windows 95 OEM Service Pack 2.
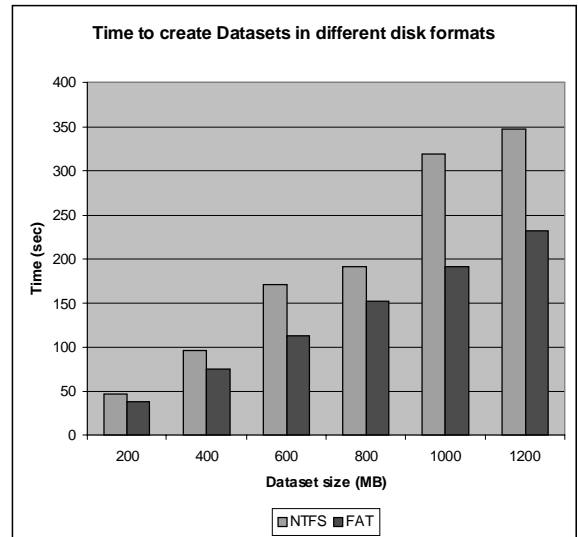

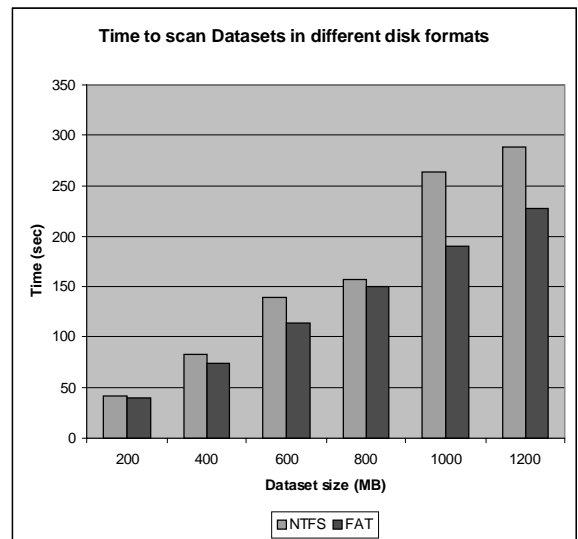Figure 5. FAT performance exceeds that of NTFS when writing files to disk

Figure 6. Read performance of FAT and NTFS are similar for smaller datasets.

SAS Library configuration can effect performance as well. If a large number of libraries and/or users are being serviced by a single disk system, the disk may spend far more time "seeking" to locations requested by the applications than serving data back to the applications. Particularly for large datasets, consider segregating data to separate disks. The closer your applications can come to sequential (as opposed to random) access to data, the closer you will come to the maximum performance of the disk. Depending on your application, run time can easily triple if two applications are heavily accessing data on the same disk in differing access patterns.
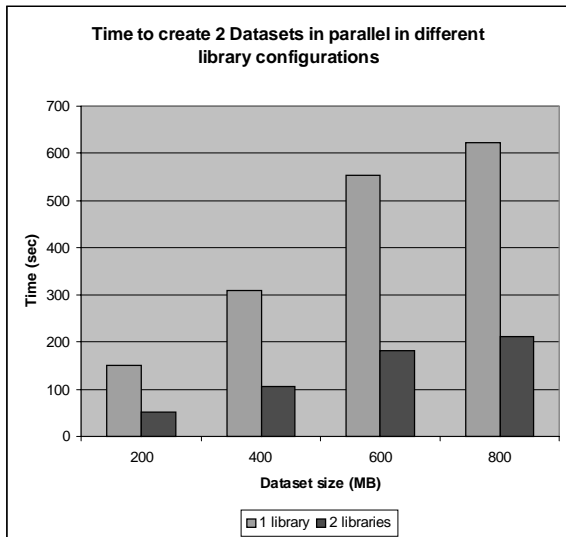
Time to create 2 Datasets in parallel in different library configurations

Figure 7. Multiple jobs running against the same library can have a dramatic performance difference when creating datasets.

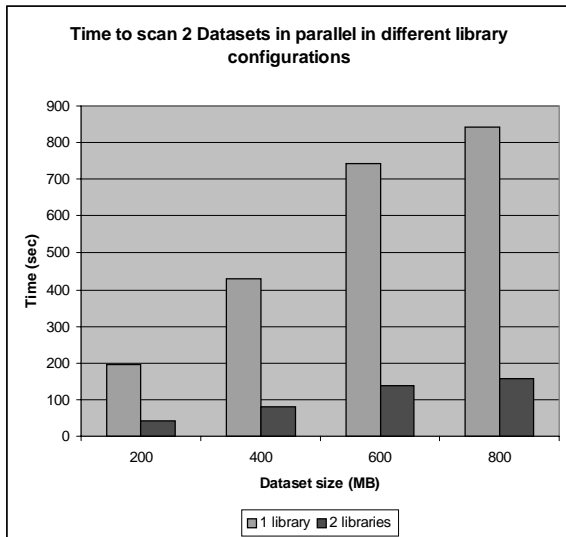**Time to scan 2 Datasets in parallel in different library configurations**

Figure 8. Parallel read processed against the same library can drastically effect performance.

Network Adapters are important in heavily networked configurations. Fast network cards can help speed network-based applications along. As the number of heavy network users increases, the need for 100Mb/sec Fast Ethernet or FDDI network will become apparent

Disk Adapters should be 32 bit PCI bus compatible SCSI or Enhanced IDE to offer the best throughput on new machines. In addition, disk systems that use fast and wide or ultra SCSI can realize additional performance gains due to increased throughput. Also, server class machines can benefit from multiple adapter configurations, especially on Pentium Pro systems where multiple PCI buses can be used.

Disk type (such as fast/wide) should always match the capabilities of the controller, and fast/wide ultra SCSI should provide the best currently-available performance. If your disk is equipped with cache memory (as many new drives are), you should be sure to enable both read and write cache operation to achieve peak performance.

Operating System versions are a factor for determining compatibility, capability and performance Microsoft Windows 95 and Windows NT 4.0 offer much improved integrated client support in conjunction with 32 bit system performance than their 16-bit predecessors.
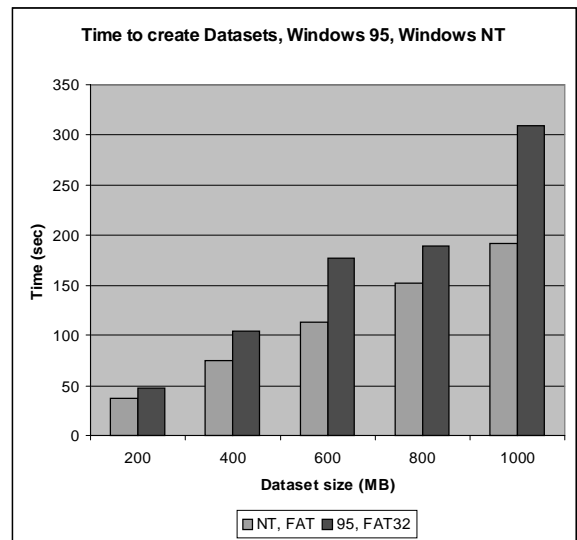
Time to create Datasets, Windows 95, Windows NT

Figure 9. Windows NT allows faster dataset creation, especially for larger files.

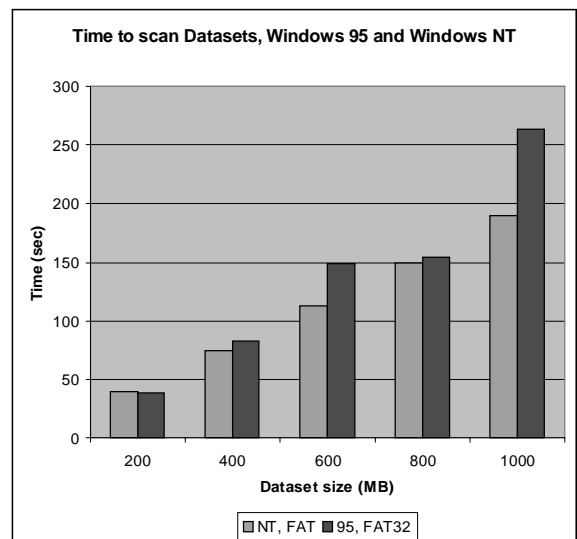**Time to scan Datasets, Windows 95 and Windows NT**

Figure 10. When reading datasets, equivalent performance between Windows NT and Windows 95 should be expected only for small datasets.

Video Adapter technology can make a great difference in apparent CPU performance due to video accelerator chips and 32 bit PCI bus throughputs. Some of this performance gain may be lost, however, by the use of very high display resolution (such as 1600x1200 displays) and pixel depth (such as True-Color, which requires at least 24-bits of data for each pixel displayed on the screen).

Bus Architecture makes a big difference. Most new PC's are equipped with a PCI (Personal computer interconnect) bus. This type of bus provides a very fast backbone for data transfer between major system components such as hard drive, network, CPU, and display adapter. The PCI bus is far faster

than older bus systems (such as ISA) which are common on older machines. Although server class machines tend to be much more expensive than desktop systems, the higher technology bus systems can benefit heavy application loads.

IMPORTANT INFORMATION ABOUT PERFORMANCE AND BENCHMARKS

Although a benchmark program was used in the preceding section, this does not constitute any guarantee about absolute performance on any platform for a given configuration. It is extremely important to remember that *your mileage will vary*, due to application, hardware, and user differences. The sample program used in the preceding charts is included below, and was run on a 200MHz Pentium Pro with dual processors, 64MB RAM, 512kb L2 cache per processor, a Fast/Wide Ultra SCSI controller driving up to four 4 gigabyte fast/wide ultra SCSI hard drives. Operating systems used include Windows NT 4.0 workstation, and Windows 95 OEM Service Pack 2.

More important than the actual numbers represented in this paper is the concept of looking for performance differences in a systematic way and applying the results to your particular enterprise and application situation. That being said, the sample program is included below.

```
libname l 'g:\bench';

data l.a;

/* Loop as many times as we want x 100MB */
do a = 1 to x;
  /* Create about 100MB of data */
  do i = 1 to 1000000;
    x1 = ranuni (0);
    x2 = x1 * 2;
    s1 = 'ABCDEFGHIJKLMNOPQRSTUVWXYZ';
    s2 = '0123456789.0123456789';
    s3 = 'qwertyuiopasdfghjklm';
    output;
  end;
end;

run;

proc means; run;

proc delete data=l.a; run;
```

Hardware Sizing Issues

Before one can answer the question "What size NT server do I need?", you need to understand which model the application fits by selecting the client/server model, as described above. In most Data Warehouse scenarios SAS System users generally have Windows 3.1 or Windows 95 PC clients that need to access data on a Windows NT server or Unix server class machine.

In every case, the size of the server depends on the amount of data, the number of users accessing the data, and the type of processing anticipated. (For example, sorting data requires an amount of space that is three times the size of the data.) Also whether the customer generally wants to have more horsepower on the desktop or in their more centrally administered servers.

It would be a good idea to run the NT Server Performance Monitor on your server system to determine if you have adequate memory to sustain a reasonable page rate for your peak load working set on the server machine. The NT Performance Monitor can also help you determine if you have adequate CPU power and help you determine

your disk I/O and network access characteristics and various other system characteristics and bottlenecks.

Amount of memory needed on the Windows client

An additional 16 megabytes for the SAS System after the operating system and any other concurrent application requirements have been met is recommended for most application types.

Amount of memory needed on the NT Server

An additional 16 megabytes for the SAS System after the operating system and any other concurrent application requirements have been met and an additional 12-16 megabytes per user, for typical processing. If the NT server is being used as a file server to share the SAS System code image then the server should have additional memory to be able to cache the entire code image of the SAS System as determined by the NT performance monitor. For complicated modeling and/or large sorting, add the appropriate amount needed for each application, per user.

Amount of disk space required on the client and the NT server

This depends highly on the type of processing that each user will be doing. Unless running a thin client in which SAS is not running on the client machine, work (temporary) files should typically be located on the local client system for optimal performance. Work files for the SAS System running on a server machine should be on disks local to that machine. There will be work files on both the client and server if SAS is run on both sides. The main point being the work files should certainly never be on a remote file server but should be on a disk local to the machine running SAS. The values for NT are comparable to those required on other platforms, so any information already available for procedure specific requirements should pertain to NT as well.

Installation Issues

The SAS System SETUP program, available with version 6.12 of the SAS System for Windows and adds features that facilitate enterprise use of the SAS System.

Typical and Compact Installation

Easily-selectable default installation types are available to facilitate easy use of the SAS System SETUP program for typical users. A **Typical** configuration will install the most commonly-used parts of all products for which you are licensed, and a **Compact** configuration will install commonly-used parts of only the Base SAS System.

CD-ROM or Network Installation

A capability of the installation program for the SAS System is called a CDROM or Network installation. This capability allows a PC to utilize SAS System software located on either a CD-ROM or a network drive. In this mode of operation, no SAS System files need to reside on the PC's hard drive or in a user's network directory. Instead, all SAS System files are accessed from the remote fileserver or the SAS System CD-ROM. Only files and data generated by the user need to be stored on their own disk space.

Selective Custom Installation

The 6.12 SETUP program also supports a normal standalone selective installation mode. The user can select products and components to copy to their user disk space based on descriptions of the products and components. This is different from a client installation in that only the selected (and copied) products or components can be run by the user: no products on the distribution CD-ROM or a networked location can be accessed in this configuration. This configuration should provide superior performance to a client installation, but will require more local hard disk space to be used.

CD35 Media and floppy diskette generation

A distribution media which is available is called *CD35*, which is a CD-ROM containing diskette images of 3.5" floppy diskettes. While this CD-ROM doesn't support client installation modes, it does provide the new functionality of being able to generate floppy diskettes from the SAS System CD-ROM. An administrator with CD35 media can select a list of products and generate SAS System product floppy diskettes identical to those shipped by SAS Institute. These floppies can then be used for intra-site distribution.

Conclusion

Various approaches to SAS System PC client/server computing and data access are possible and have been explored. Each can be useful in certain configurations and situations. The flexibility of the SAS System allows Information System designers to design an enterprise computing system that utilizes data and resources efficiently. There is a wide range of variables in areas such as application design, remote data access, remote computing access, and desktop PC speed and disk size. No single approach is best for all configurations. Application needs, data quantity and data locations will direct how a system should be organized. Factors to consider include:

- The location of SAS System software. Network installations are possible to save disk space but can effect performance in some cases. If the SAS System is run from a file server, hybrid local and network installations are suggested to minimize local disk space requirements while keeping network delays small.

- The processing power and configuration of the desktop and server computers are of critical importance and PCI bus machines with 32 bit video and disk adapters should be selected. Increasing remote server performance and availability is an easy way to increase performance for users. Plan for eventual migration to newer Windows platforms with integrated client support.

- Offloading jobs to PC platforms from overloaded central resources can get the job done quicker in many cases and will usually lower costs compared to using central resources.

- If you must process all records of a large data set multiple times, use SAS/CONNECT Compute Services to process the data at the server or Data Transfer Services to download the data set for local processing. Either approach will help reduce network traffic.

- Use WHERE clauses on your procedures and data steps and indexes on SAS data set variables that are likely to be used to subset your data by the WHERE clause. These two features will significantly increase data access performance and reduce network traffic.

- Rely on central services when they are plentiful and when data is centrally located. Mainframe and other centralized computing resources typically have access to unique hardware and software products as well as high quality system support and backup facilities that may be required in many cases. SAS/CONNECT and SAS/SHARE software allow critical data to remain in the safest location and be analyzed on the most capable platform.

- Take advantage of interoperability with other applications. Third party ODBC access to SAS data, OLE interoperability, and SAS/ACCESS connections or flat file access to external sources add flexibility for system designers. This means that data can be accessed or processed in the most efficient manner by the most appropriate software tool set.

- Lotus Notes and e-mail access through the SAS System can help meet organizational needs. You can also extend the power of the SAS System through external DLL access and full integration with operating systems like Windows 95.

- Use the SAS Domain Server gateway if you do not want to bear the cost and maintenance of multiple protocol stacks when you must contact diverse protocol domains.

- Network administrators and system designers as well as power users should determine the optimal uses of the enterprise hardware and software which will lead to more optimal use of personnel time.

- A properly configured installation can help performance as well as the maintainability of the installed software.

In summary, the SAS System provides a wide range of computing and data access options that meet just about any system requirement from desktop analysis to data warehousing. The desktop PC running the SAS System can play an important role in your enterprise computing strategy and will increase both your productivity and your capability.

Trademarks

SAS, SAS/ACCESS, SAS/AF, SAS/ASSIST, SAS/CONNECT, SAS/GRAPH, SAS/SHARE, and SAS/SHARE*NET are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. OS/2 is a registered trademark or trademark of International Business Machines Corporation. ® indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their respective companies.

References

Betancourt, Randy (1995) "Data Warehousing Strategies for Using the SAS System," *Proceedings of the Twentieth Annual SAS Users Group International Conference*

Cates. Mark (1997) "Taking Advantage of the SAS System on Windows NT," *Proceedings of the Twenty-second Annual SAS Users Group International Conference*

Clegg, Jennifer and Rigsbee, Carol (1996) "OLE and the SAS System for Windows Release 6.12," *Proceedings of the Twenty-second Annual SAS Users Group International Conference*

Mehler, Gary, (1996) "Integrating the SAS System for Personal Computers into Your Enterprise," *Proceedings of the Twentieth Annual SAS Users Group International Conference*

Rigsbee, Carol (1997), "Using OLE with the SAS System", *Proceedings of the Twenty-second Annual SAS Users Group International Conference*

SAS Institute Inc. (1995) "Client/Server Computing with the SAS System: Tips and Techniques," SAS Institute Inc.

SAS Institute Inc. (1996) "Microsoft Windows Environment: Changes and Enhancements to the SAS System, Release 6.12," SAS Institute Inc.

Author Information

Gary Mehler
SAS Institute, Inc.
SAS Campus Drive
Cary NC 27513
Phone: (919) 677-8000
e-mail: gjm@unx.sas.com