Data Warehouse
+  
Microsoft Windows NT Server
=  
A Perfect Solution for Client/Server Computing

Joe Wang, Phamacia & Upjohn. Inc., Kalamazoo, MI

Abstract

SAS Institute Inc. claims that it has all the tools for users to extract data for, organize data in, and explore data from their Data Warehouse. But how will users build the Data Warehouse for their companies? What are the network, hardware, software, and SAS products configurations needed to build Data Warehouse in various size companies?

Microsoft's NT Server is a promising star as the operating system for an application server. Bundled with the BackOffice offer: SQL server, SNA server, Systems Management Server, and Mail server, this is an ideal client/server platform for a company's downsizing project.

For small to midsize companies, this paper will show you how to design an integrated network, hardware, software, and SAS products architecture for users to build their Data Warehouse on Windows NT Servers.

Business Process Reengineering

Thomas H. Davenport, in his book Process Innovation - Reengineering Work through Information Technology, proposed the following five key steps for a business process reengineering effort:

1). Identify process for innovation

Identify and rank processes according to priority as well as analyze the organization's ability to support the new processes. The priority is based on information such as which processes have the greatest effect on customers, which processes have the greatest potential for radical improvement, which processes can be improved simply and quickly, and which processes seem to have the most problems.
2). Identify "levers" of change

Search for new technology enablers which can be applied in the redesign process to help companies do things that they are not already doing.

3). Develop process visions

Based on an organization's strengths and weaknesses as well as market structures and opportunities, develop a business strategy and process visions. Conduct benchmarking based on best competitors' innovations and activities is essential.

4). Understand existing processes

Use process flow diagram, control chart, and distribution analysis to understand current process, identify levers for change, and collect information on time, cost, and other resources consumed by the process.

5). Design and prototype new processes

The iterative design of new technology-supported business processes using quick prototyping is essential to a successful business reengineering effort.

Data Warehouse & SAS's strategies

The advanced client/server networks and a whole range of new middleware technologies are accelerating the move to data warehousing.

The data in the data warehouse is selectively retrieved or summarized from an organization's operational systems. The data warehouse contains only the information necessary for decision support processing. The data is collected over time and used for comparisons, trends, and forecasting. This data is not updated in real-time, but is refreshed from operational systems on a regular basis, when the data transfer will not adversely affect the performance of operational systems.

In addition, the warehouse repository is created to be read from, not written to altered. End-users of the data warehouse are not data entry personnel entering transactions. Instead, they are business managers making decisions based on the information in the data warehouse.

A data warehouse has four major elements: data access, data transformation, data organization, and data exploitation.

1) Data access

Presently, the SAS System provides more than 50 different access methods for a variety of file types found in different hardware environments. These access methods are a part of the SAS/ACCESS family of ACCESS and VIEW technology and include access to:

- relational database management systems
- hierarchical databases
- network database management systems
- data gateways and standard API's such as ODBC
- external file formats such as VSAM
- SAS data sets

The SAS System also offers users
complete control over the SQL passed to the database with the passthru SQL facility in PROC SQL. SAS/ACCESS software can reside on the same system as the RDBMS, or they can reside on a client platform and communicate through SQL-based middleware.

2) Data transformation

The SAS System provides a large number of tools for data transformation. They include:

- the ability to open multiple input files from different sources residing on different platforms simultaneously
- the ability to open multiple output files simultaneously
- the ability to perform look-ahead reads and table look-up logic
- sorts that can use a variety of character sets and collating sequences
- SQL for Groupby, Orderby, and summary functions

3) Data organization

SAS/CONNECT software supports facilities for moving data and forwarding SQL requests to a server platform. It also provides a unique capability for dynamically distributing application logic to remote systems. SAS/CONNECT software allows the SAS System on one machine to communicate with a SAS process on another machine. The software provides three capabilities for distributed processing: Data Transfer Services, Remote Library Services, and Remote Computing Services.

Data Transfer Services enable the transfer of data between two machines. Data Transfer Services supports the movement of SAS data sets; SAS catalogs, SAS graphics output, and external files between local and remote systems. Uploading and downloading of data can be scheduled to occur at off-peak times when network traffic and costs are lower. The Query Window component of SAS/ASSIST software, which provides a graphical interface for retrieving remote data, uses the Data Transfer Services of SAS/CONNECT software.

Remote Library Services allows a user to assign a libname in a local SAS session for data which resides on a remote system. With Remote Library Services the returned data is processed in memory on the local system. A physical copy of the data is not created on the local system. Remote Library Services are ideal for low-volume transaction-oriented applications involving small amounts of data.

Remote Computing Services allow the SAS System to forward statements to a remote machine for execution. Because the SAS System's MultiVendor Architecture insures the same functionality on all platforms, any programs which can operate in a client environment can also be forwarded for execution on the server.

The physical data distribution can be organized through using SAS/CONNECT to achieve an acceptable retrieving time and less data transfer on the network.

4) Data exploitation

The SAS System for data exploitation supports diverse levels of user experience through a range of
custom-tailored interfaces, from an EIS interface for enterprise decision makers to a task-oriented menu-driven interface for business analysts to a full-function programming interface for applications developers.

As more companies right-size their hardware, data and applications must be portable as well as connected in order to take maximum advantage of available processing power. Through SAS System’s MultiVendor Architecture, applications are portable from one platform to another.

97% of the Fortune 100 companies and 95% of the Computerworld Premier 100 are using the SAS System for Enterprise Information System, Decision Support System, Report Writing System, Financial Analysis System, Market Research System, Project Management System, Data Visualization, Network and System management, Clinical Trials Analysis, and Quality Improvement.

Windows NT & Backoffice

Microsoft Windows NT Server provides the connectivity, reliability, base services, and administrative tools necessary to deliver critical business information across a distributed network of computers. Its scale ranges from the small network to the enterprise, providing basic file and print services while supporting mission-critical databases, electronic messaging, host connectivity, and distributed systems management.

You can add a fast application server to a network of NetWare file and print servers, add an easy-to-operate file and print server to a UNIX network, or bring mainframe data to users quickly, simply, and consistently. This makes it easy to take advantage of the technology available with Windows NT Server, while integrating it into your existing computing environment. The combination of performance improvements and reductions in hardware pricing makes Windows NT Server even more cost-effective to deploy as a network operating system.

Workstations connected to the network via Remote Access Service of Windows NT appear as if they are directly attached to the network. They can participate fully in the network, sharing files and printers, accessing databases such as SQL Server, connecting to SNA host via SNA Server, and communicating with colleagues via e-mail.

Microsoft SQL Server relational database features advanced transaction processing, server-enforced data integrity, remote stored procedures, high availability, and distributed transactions.

Microsoft SNA Server connects desktops running on Windows, MS-DOS, Windows NT, Macintosh, UNIX and OS/2 with AS/400 and IBM mainframes. SNA Server is integrated in a client/server manner, so users’ AS/400 and IBM mainframes can view up to 2,000 personal computers as a single node on the SNA network. This improves the performance of the SNA network by reducing the number of nodes that need to be managed. It also increases the performance of customers’ desktops by minimizing the amount of memory devoted to managing host connectivity from each desktop.

Microsoft Systems Management
Server helps reduce support costs by centralizing common network administrative tasks that typically cost organizations significant amounts of time and money. It has four primary functions: hardware and software inventory, automated software distribution and installation, remote systems troubleshooting and control, and network application management.

Microsoft Mail Server allows you to communicate with anyone, anywhere. Administrators will appreciate the powerful administration tools. One administration program can manage an entire system. Mail Server automatically synchronizes directories and creates a global address list. And with the Mail Multitasking MTA (Message Transfer Agent), you can install more postoffices and can service more remote users without needing additional hardware.

Microsoft Exchange Server delivers interoperability with currently available messaging systems such as Microsoft Mail, IBM PROFS, and Lotus cc:Mail. It also works seamlessly in NetWare environments, and the Microsoft Exchange client is supported on the Microsoft-DOS, Windows, Windows for Workgroups, Windows 95, and Windows NT Workstation. Future client support for the Apple Macintosh, OS/2 and UNIX.

Network & Remote Access

With more and more companies decentralizing their operations, the demand of putting employees on the road or sending them home to telecommute is rapidly rising. Designing a remote-access solution with a corporate network can be a complicated, detailed process, dealing with a number of factors, such as:

- the number of remote users who need access
- where users will be when they need access
- the applications and protocols for users
- the volume of data traffic they’ll generate
- how users will connect into the LAN headquarters - that is, dial-up lines or leased circuits.

A good remote-access solution requires an understanding of the entire scope of an organization’s network. This involves considering what applications will be used, discovering which protocols are involved and determining which type of communications services will be required.

One of the biggest challenges in developing remote-access solutions is making them easy to install, upgrade and maintain. This is especially important when considering products for use out in the field. Small local offices typically have limited technical expertise. This means working with products that are leading edge and offering features like dynamic reconfiguration, dial-back and priority queuing, and point-and-click installation and configuration is a plus.
Small Company

After you identify processes for innovation for your company, the next step is to search for Information Delivery Technology enablers. The technology enablers fall into four categories: hardware, LAN, system software, and application software. The powerful PCs and RISC servers connected within a LAN is easy to install, maintain, upgrade and expand. Windows NT is a cost-effective network operating system which co-exist with Novell network and UNIX servers well. SAS products fit all end-users' needs of your company.

For a small company, Figure 1 is a typical network layout. You can start with one PC; then, add more PCs with a peer-to-peer connection. When business grows, connect PCs with a hub and a server. If your company has users that want to dial-in, a remote access server needs to be set up with modems. Windows NT server and Backoffice provides file, print, DBMS, e-mail, remote access, and system management functions even when the LAN expands. SAS products can be either installed on PCs or the server or both. The data warehouse can reside in Microsoft SQL database, Oracle database, or SAS data sets. You can also run SAS applications from home and access company's servers through a modem connection.
Midsize Company

For a midsize company, a headquarter and one or several branch offices located in different buildings or cites is quite typical. The branch office layout is very similar to a small company except the branch office uses a router and CSU to connect to the company headquarter through T1, or frame relay. (Figure 2.) Users can either access the branch office’s server or headquarter’s servers and IBM mainframes from offices or homes.

The company-wide data warehouse which resides on the headquarter server retrieves information from each branch office and provides access to users from each branch office. Users on the road or at home dial-in the local branch office and invoke SAS applications on the branch office’s server.

Users at each branch office can access headquarter’s Oracle database, SAS data sets, or DB2 database from SAS Frame applications through SAS/ACCESS, SAS/CONNECT, SAS/EIS, etc.

With the SAS system’s Business Intelligence Software, The three phases of decision making can be accomplished: reaching and examining data to identify areas of interest, analyzing and confirming data, and presenting data.
The headquarter layout is shown in Figure 3. All branch offices are connected to the headquarter through CSUs and routers. The remote access server handles all dial-in users through modems. For full-time telecommuters, the connection through ISDN bridges provides about four folds of the modem speed. There are dedicated servers for file, print, e-mail, DBMS, and SNA gateway. Microsoft NT server and Backoffice will play an important role in this heterogeneous environment.

**Conclusion**

For right-sizing, down-sizing, client/server projects or whatever you called it, the five key steps of the business process reengineering suggested by Davenport are practical guidelines. First, identify processes needed to be changed by priority. Second, identify technology enablers: LAN, Windows NT, SAS products, etc. Third, develop processes visions with key customers. Fourth, understand existing processes in detail. Finally, Design and prototype new processes with the technology enablers mentioned above. These enablers have to be easy to use, maintain, upgrade, and expand.