Consolidation in the Data Center

David Hornby, Global Sales Organization

Ken Pepple, Enterprise Services

Sun BluePrints™ OnLine—September 2002

http://www.sun.com/blueprints

Sun Microsystems, Inc.
4150 Network Circle
Santa Clara, CA 95045 U.S.A.
650 960-1300

Part No. 817-0181-10
Edition: September 2002

Please
Recycle

Types of Consolidation

When we started working on consolidation four years ago, the focus was on server consolidation; now, the emphasis has changed. When you walk into a data center and look around, everything you see is a candidate for consolidation. Today, consolidation is about reducing the number of devices you have to manage and reducing the number of ways you use to manage them.

With that in mind, this Sun BluePrints™ OnLine article takes a look at some of the consolidation opportunities that exist in today's Information Technology (IT) environment. These include:

Consolidating Servers

Servers are still the primary focal point for consolidation because they are so obvious. Whether you have 100 servers or 5000 servers, you probably have too many to manage effectively. Today's distributed computing environment lends itself to a proliferation of servers. Reducing and controlling the number of devices to manage and simplifying ways to manage them is the goal of most IT groups.

Applying Vertical and Horizontal Scalability

When we talk about consolidating servers, we generally refer to scaling them vertically or horizontally.

- Vertical scalability enables you to reduce the number of servers by consolidating multiple applications onto a single server.
- Horizontal scalability enables you to deal with increased workloads through the replication of servers and the distribution of workloads across those servers.

By thinking of consolidation in these terms, you begin to define the approach required by your particular consolidation project. Once you decide whether your consolidation project requires horizontal scaling, vertical scaling, or a combination of the two, you can further refine your approach by identifying patterns in your server population. Examples of vertical and horizontal scaling are presented in the Sun BluePrints book *Consolidation in the Data Center* (ISBN #0-13-045495-8).

Identifying Patterns in an End-to-End Architecture

In the end-to-end architectures that are prevalent today, tiers of servers are specialized for particular tasks. When you look at consolidating servers, you need to look for patterns in your server population. When you identify these patterns within tiers, you can start to devise a consolidation strategy. Scalability is the key, here. Because you are expected to deliver predictable service levels in response to unpredictable workloads, it is important that you use the right type of scalability for each part of a consolidated architecture. The following sections describe common patterns in an end-to-end architecture.

For consolidation discussions, we generally assume that there are three server types, or tiers:

- The presentation tier is the closest tier to the end user.
- The business, or middleware, tier is where applications or middleware run in conjunction with the other tiers.
- The resource tier is where large, scalable servers run mission-critical applications and databases.

Although architectures with these characteristics have been around for awhile, most corporations still have many servers running monolithic applications. In many cases, these are older servers running mature applications. These servers are generally excellent candidates for server and application consolidation.

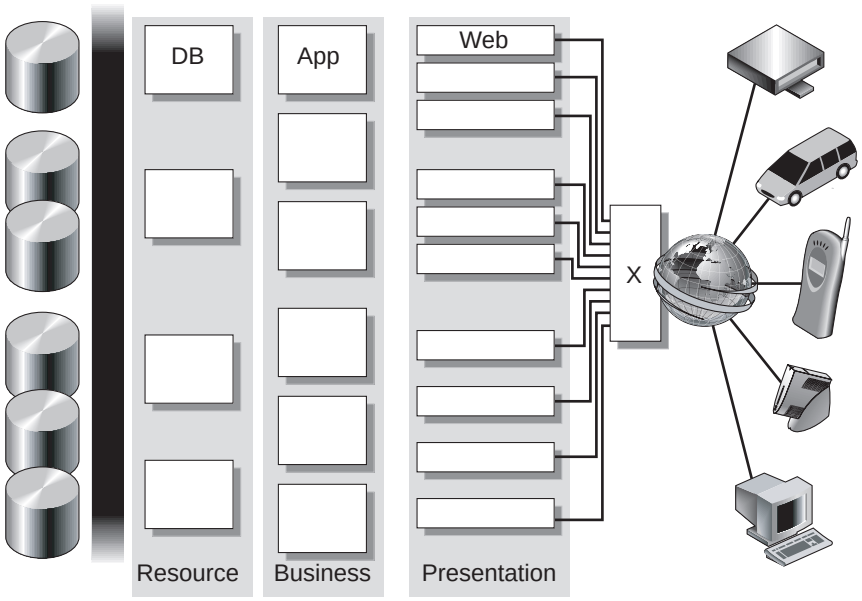

FIGURE 0-1 End-to-End Architecture

Presentation Tier

As you get closer to the Internet and the delivery devices for your applications, the most common techniques for scalability are replication and load balancing across an array of generally small servers. This is usually referred to as horizontal scalability.

On the presentation tier, you usually find several small, one or two processor servers running UNIX™, Linux, or Microsoft Windows. Because current architectures call for many small servers in this tier, the opportunities for consolidating to fewer servers are limited. Despite these limitations, there is still ample opportunity to implement standards and best practices to simplify the management of these servers and to reduce total cost of ownership (TCO).

Business Tier

After the presentation tier, you find business tier servers. These generally run applications or middleware in conjunction with the other tiers. While we used to see primarily horizontal scaling in this tier, there is growing interest in reducing the number of applications using vertical scalability, as well.

Resource Tier

The resource tier is usually thought of as the heart of the data center. It is characterized by large, scalable servers that run mission-critical applications and databases. We usually find the most interest in vertical scalability and server consolidation on this tier.

Identifying Types of Consolidation

Some major consulting groups identify different categories of server consolidation. For example, Gartner refers to physical, logical, and rational consolidation groups—categories we define in the following sections. The techniques used to achieve these specific types of consolidations are discussed in later chapters.

Physical Consolidation

Physical consolidation involves consolidating data centers and moving servers to fewer physical locations. The theory behind physical consolidation is that by having servers in fewer physical locations, you can achieve management consistencies and economies of scale more easily than you can when your servers are dispersed. Physical consolidations may also enable you to reduce data center real estate costs. It is generally felt that physical consolidation has the lowest risk, but that it also has the lowest payback.

Logical Consolidation

Logical consolidation involves implementing standards and best practices across your server population. By doing this, you can realize substantial benefits in the productivity of your IT staff. They can manage the environment more efficiently and more effectively. This can often result in lower systems management costs and in lower TCO. Logical consolidation is often implemented with physical consolidation and rationalization.

Rationalization

Rationalization involves the deployment of multiple applications on fewer, larger servers and in fewer instances of the operating system (OS). Because TCO reduction is closely tied to the number of instances of an OS you manage, reducing the number is the best way to reduce TCO. While rationalization is the riskiest form of server consolidation, it offers the biggest TCO reduction and return on investment (ROI).

Because of its potential to reduce TCO and increase ROI, we find that most of our customers find rationalization to be the most attractive method for consolidation, despite its higher risks. During a recent web seminar on consolidation, we polled over 300 attendees on whether their main interest was in physical consolidation, logical consolidation, or rationalization. Over 85 percent said their primary interest was in rationalization.

Consolidating Applications

Understanding which server resources applications use and understanding usage patterns is essential to the process of consolidating servers. This process is explained in detail in the chapters about assessment and architecture.

Everywhere we go to talk about consolidation, people always ask which applications they should or should not consolidate. There is no easy answer to this question, and you need to evaluate each application individually. The people who know the applications best are the developers, database administrators, and system administrators who take care of them on a daily basis. Talking with these people usually identifies good and bad consolidation candidates rather quickly.

Eliminating Consolidation Candidates

While it is not easy to make specific recommendations about which applications are good consolidation candidates, there are some general guidelines you can follow to eliminate candidates from consolidation.

First and foremost, consider eliminating all servers and storage that are deliberately isolated for other servers. This includes:

- Firewalls
- Intrusions detection servers
- Sensitive databases
- Geographically separated servers

Isolation and separation are vital attributes of many security-related servers. Unless you are specifically targeting these machines, they are probably too complex to consolidate without a complete reevaluation of the security architecture. After you have consolidated the rest of your environment, you can come back and look for opportunities for security consolidations. The same is true for servers or storage that are geographically separated. For instance, if you have separated servers for redundancy or wide area network (WAN) reasons, it hardly makes sense to try to consolidate these together.

Third-party applications developed and sold by independent software vendors (ISVs) may be problematic as consolidation candidates. In many cases, ISVs specify that their applications run in standalone mode on their own servers. If this is not done, they usually do not provide support for the application. Make sure you check with your vendors for their support requirements.

Other applications that do not qualify for consolidation include those that lock up system resources such as physical memory. If you aren't sure about an application's suitability for consolidation in this respect, make sure you analyze it with a performance tool such as TeamQuest or BMC's Patrol-Perform and Predict.

Applying Backward and Forward Consolidation Strategies

When most people begin their consolidation efforts, they often focus on existing applications. This process is known as a backward consolidation. In many cases, these consolidations are hugely successful. In others, organizational issues prevent them from achieving their goals. There is no doubt that some of the largest and most successful consolidations we have seen have been backward consolidations.

To avoid future server sprawl, it is also important to begin to develop new applications in a consolidated environment, referred to as forward consolidation. It's interesting to note that applications that are developed in a consolidated environment usually behave with each other, and can be moved into production in a consolidated environment.

Consolidating Application Sets

When you analyze a group of servers for consolidation, it is important to identify the applications that run on them. It is equally important to identify the servers and applications with which the consolidation candidates interact. It's very common to find groups of applications that receive data from other applications or that supply data to other applications. Because these applications process information sequentially, they are often excellent candidates for consolidation. These application sets don't usually compete for the same resources at the same time, and when you group them within a single instance of the OS, you often find that they perform better because of intra-domain communications. A decrease in network traffic is often a side benefit of consolidating application sets.

When deciding which application sets to consolidate, we generally categorize them as being tightly coupled or loosely coupled. In tightly coupled application sets, applications actively interact with each other. In loosely coupled sets, a group of applications acquire and process data and then pass them on to the next server.

As an example of a successful consolidation of tightly coupled applications, consider a situation where a customer gathered home-grown Enterprise Resource Planning applications, Tuxedo messaging, and an Oracle database server and moved them into a single instance of the OS. The results were extremely successful. Application throughput increased, and network usage decreased. From a scalability viewpoint, the transaction volume of the application set increased from roughly 40,000 sales order lines per day two years ago to over 400,000 sales order lines per day. This was obviously a successful consolidation and a successful application of vertical scalability.

As an example of a loosely coupled application set, consider a situation where a server runs an online transaction processing (OLTP) application during the day, then at night, passes the transaction information to a batch server for further processing and reporting. The batch server would then update a data warehouse server, which would, in turn, update multiple data mart servers. Each server depends on another, but not simultaneously.

Consolidating Multiple Database Instances

A very common example of server and application consolidation is to put multiple instances of a database in a single instance of the OS. This is a very common tactic with most major databases. It's important to note that you don't want to mix these two databases on the same instance of the OS; doing so may cause fatal application conflicts.

The success of this consolidation strategy depends heavily on development standards and practices. Where strict naming conventions are followed and each instance of the database is unique, consolidations are successful. However, on an occasion when we worked with a customer who wrote multiple Oracle applications and chose to use default naming conventions, multiple conflicts resulted between database instances. Obviously, this consolidation was not successful.

Consolidating Similar Workloads

Another consolidation technique that some consultants recommend is to consolidate similar workloads with like resource usage and time scheduling. If you do this, you need to understand the application’s resource usage very clearly. When using this consolidation technique, the key to success is that you must offset the timing of resource usage for each workload. Otherwise, you simply end up with overlapping peaks and valleys of resource usage.

Consolidating Storage

Today, there is as much interest in storage consolidation as there is in server consolidation. As we said earlier, every new server you deploy results in more storage. In many cases, the cost of storage for a server exceeds the cost of the server, and although the server may not grow very much, the amount of storage required for an application will grow indefinitely.

When you look at application sets, you find a lot of data replication because multiple applications look at the same data. For example, if you have an OLTP server, a batch server, a data warehouse server, and four data mart servers, you may have seven copies of a particular dataset such as a customer master file. While, in theory, all seven copies are identical, because different developers probably created their applications, it is likely that there will be differences in each copy of the data. This situation has the potential to create a situation where reports run on the “same” data, yet yield different results.

As with servers, the goals in storage consolidation are to reduce complexity, increase utilization, and reduce TCO. These goals can be achieved through a variety of consolidation techniques. The ultimate goal is data sharing among applications. Unfortunately, this often requires redesigning and redeveloping applications, so it is a long-term goal, at least for backward consolidation. In a forward consolidation, data sharing should absolutely be a goal.

Other benefits of storage consolidation include:

- Easier backup and recovery
- Increased availability
- Improved scalability
- Storage resource pooling

When undertaking a storage consolidation effort, it is critical that you understand how disk space is utilized. While most of our customers can't tell us exactly what their current disk space utilization rates are, many companies we have surveyed estimate a rate of 40 percent, or less. For companies that have tried to accurately assess disk utilization, complexity often hinders their efforts. For example, one company we visited started counting storage, but stopped at 110 terabytes after realizing that they just couldn't count it all. When we evaluated their disk utilization, we found they were utilizing only 20 percent of their available storage.

There are several types of storage consolidation available today. The following sections describe the three most common types we see.

Consolidating Servers and Their Associated Storage

With every server consolidation, there is an accompanying storage consolidation. As you move multiple applications to a single instance of the OS, you must also move their storage, as shown in the following graphic. In theory, once you have moved the storage, data will be available to any of the applications. This is the most primitive form of storage consolidation.

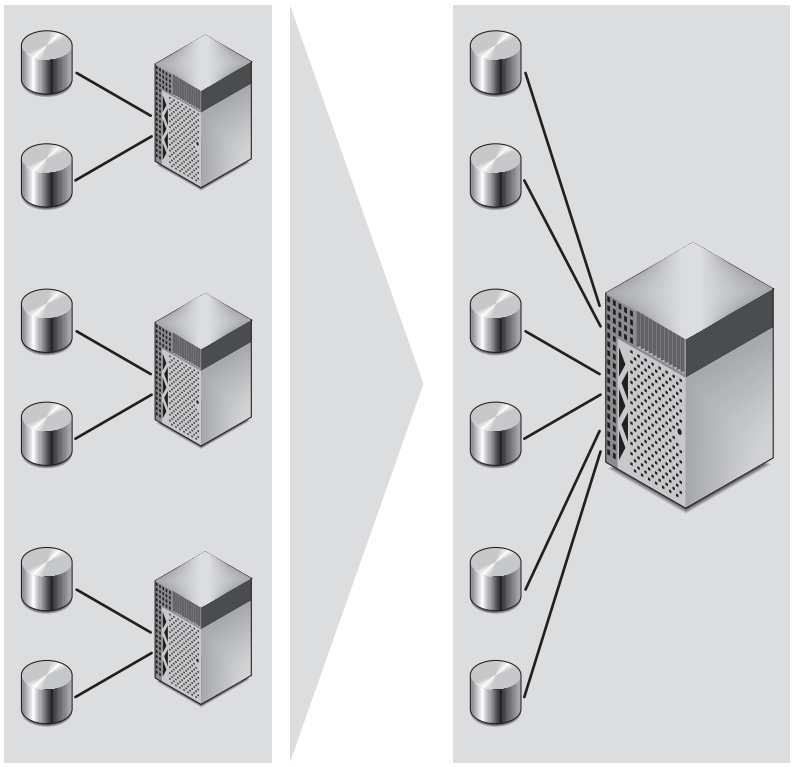

FIGURE 0-2 Consolidating Direct-Attached Storage

Connecting Heterogeneous Environments to a Single Storage Component

Many of our customers have heterogeneous IT environments. They run servers and OSes from many vendors, and they want to access data from a variety of different servers. With direct-attached storage, this was difficult to do. Now, with products like the Sun StorEdge™ 9900 storage array, it is possible to connect the Solaris™ Operating Environment servers, other UNIX servers, Microsoft Windows NT servers, and mainframe servers to the same storage array. The following graphic demonstrates this capability.

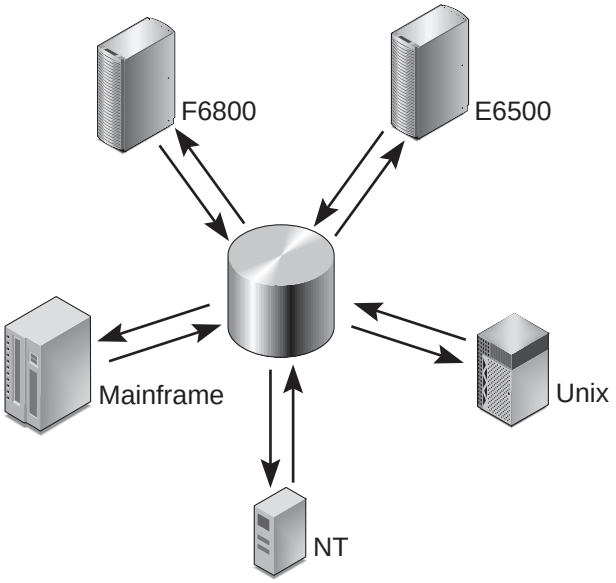

FIGURE 0-3 Heterogeneous Environments to Single Storage

This is a very popular type of storage consolidation, especially where there are multiple mission-critical applications running on servers from a variety of vendors.

Consolidating With Storage Area Networks

Storage area networks (SAN) have been the hottest trend in storage architectures for the last few years. As a technology, the SAN is now mature enough that it can be implemented using standards and standard configurations. As shown in the following graphic, SAN technology inserts a network, or fabric, of switching devices

between servers and storage that enable any server or application to access any storage connected to it. The fabric can then be configured to allow various servers to access various storage.

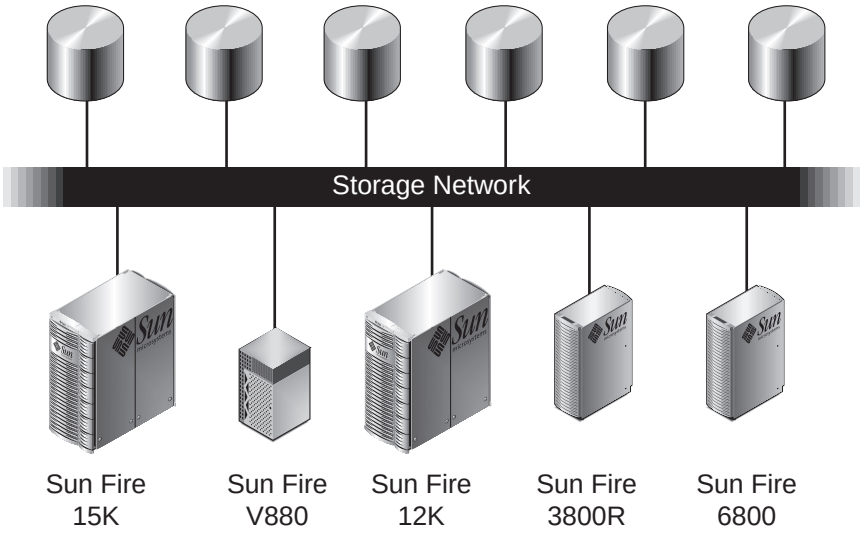

FIGURE 0-4 Storage Area Network (SAN) Configuration

Another hot storage technology is network attached storage (NAS). This technology allows servers and clients to utilize storage directly over the network using common networking protocols like network file system (NFS) and Server Message Block (SMB). Although not used greatly in many server and application consolidations within the data center, it is used extensively in file and print consolidations at the department and workgroup level. As the technology matures, expect it to work itself into data center consolidations.

Consolidating Shared Services

Another type of consolidation that is rapidly gaining popularity is middleware or shared services consolidation. Over the last few years, companies have implemented shared services (such as file, print, authentication, and e-mail) in a variety of ways. Business units typically implement their own versions of these services using software from a variety of vendors; however, when they want to share some of this information, they find that the inconsistencies and incompatibilities among these various architectures increase complexity, thereby increasing cost and making it

difficult to share. As a result, we often find that companies want to rebuild their shared services as web services using a standards-based, unified architecture such as Sun™ Open Net Environment (Sun ONE).

A common example of this is the use of directory services. Directory services have been implemented over the years using a variety of architectures. Now that standards based architectures are available, products like the Sun ONE Directory Server, which is based on the Lightweight Directory Access Protocol (LDAP), are being used to design and implement corporate-wide directory services. Condensing multiple, disparate directory architectures into a single, corporate architecture allows corporations to simplify directory maintenance and easily replicate directory services, as needed.

We expect to see more of this type of consolidation over the next few years.

Consolidating Networks

When you consolidate servers and applications, network consolidation is usually either a big issue or not an issue at all. When you collapse data centers and concentrate servers and applications into fewer physical locations, there may be a severe impact on the network. This impact must be evaluated as part of the consolidation project. You certainly need to ensure that there is sufficient network bandwidth to handle network traffic to and from the consolidated location.

Conversely, when you do a rationalization within a single data center, you often find that there are no major network changes that need to take place. The overall traffic does not increase, since new servers and applications are not being introduced. Further, rationalization may actually decrease network traffic, because applications in a single instance of the OS don't need the network to communicate with each other.

Consolidating Data Centers

Many organizations are looking to consolidate multiple data centers into one site. These consolidations range from simple city-wide consolidations to complex region-wide consolidations. Most companies are being driven toward data center consolidation because of the dramatic drop in the cost of telecommunication wide area network lines, the huge disparity of IT wages between certain regions of the

world, and the high real estate costs of large cities (especially in New York, London, and Tokyo). For those considering simple local-site consolidations, consolidation offers cost savings and enables disaster recovery initiatives.

If your organization is seriously looking to consolidate a data center, carefully consider the goals for consolidation. Shutting a data center is a huge task, and before you even start down the path, it is vital that you can articulate and defend your reasons for doing it. Further, once a data center is shut down, the costs of reopening it can be enormous. From there, data center consolidations are similar to other types of consolidation, except that assessment (especially application, networking, and physical planning) and implementation become much more complex.

Consolidating People Resources and Processes

In any consolidation project, you must not neglect the people and processes you use to manage your environment. Time after time, when we work with clients who have both mainframe and distributed-computing environments, we find that the mainframe side of the house runs smoothly and problem free, while the distributed-computing side of the house is often chaotic, with few developed standards and procedures. The problems that result from this situation demonstrate the importance of resolving these people and process issues. While mainframes consistently run with high availability and high service levels, distributed computing systems often suffer from low service levels and low availability.

Some consulting groups estimate that only 20 percent of data center availability is hardware or technology related; the other 80 percent is estimated to be directly related to people and process issues. The lesson is that successful consolidations must address standards, people, and processes. Without these, availability and service levels may be compromised.

Another benefit of implementing standards and best practices is that you frequently see a 10–20 percent reduction in TCO. You may be able to realize these types of savings by following the recommendations outlined in the Sun BluePrints book *Consolidation in the Data Center* (ISBN #0-13-045495-8).

Related Resources

This article is an excerpt from the Sun BluePrints book *Consolidation in the Data Center* (ISBN #0-13-045495-8). Refer to the book for more information about the topics presented in this article.

About the Authors

David Hornby has been with Sun for five years in Sun Professional Services (Sun PS) and in Sun's Global Sales Organization. While with Sun PS, he specialized in server consolidation and data center consulting, helped develop the Sun PS consolidation methodology, and participated in the sale and implementation of multiple server consolidation projects. David currently focuses on consolidation and TCO strategies. Prior to joining Sun, he spent 25 years in IT with an emphasis on providing IT solutions to business problems. He has over 10 years experience managing IT organizations at the CIO level.

Ken Pepple is an IT Architect in the Sun PS Asia Pacific practice. In this role, Ken assists clients with enterprise computing architecture solutions, concentrating on advanced data center projects. Before this, Ken managed the Sun PS high-end platform-services program and focused on complex performance issues for the IT Consulting and Operations practice. While there, he co-wrote and taught Sun Education's “Solaris Operating Environment Performance and Tuning Secrets” seminar.

Ordering Sun Documents

The SunDocs^{SM} program provides more than 250 manuals from Sun Microsystems, Inc. If you live in the United States, Canada, Europe, or Japan, you can purchase documentation sets or individual manuals through this program.

Accessing Sun Documentation Online

The `docs.sun.com` web site enables you to access Sun technical documentation online. You can browse the `docs.sun.com` archive or search for a specific book title or subject. The URL is `http://docs.sun.com/`

To reference Sun BluePrints OnLine articles, visit the Sun BluePrints OnLine Web site at: `http://www.sun.com/blueprints/online.html`