UNIX Server Consolidation Strategies

Exceed onDemand and Exceed Freedom
Remote X Window Application Delivery Solutions

Published: February 2009
While every attempt has been made to ensure the accuracy and completeness of the information in this document, some typographical or technical errors may exist. The Open Text Connectivity Solutions Group cannot accept responsibility for customers' losses resulting from the use of this document. The information contained in this document is subject to change without notice.

This document contains proprietary information that is protected by copyright. This document, in whole or in part, may not be photocopied, reproduced, or translated into another language without prior written consent from the Open Text Connectivity Solutions Group.

This edition published February 2009.
Contents

Overview ..............................................................................................................................................1
Server Sprawl ......................................................................................................................................2
   Skyrocketing IT Cost ..................................................................................................................2
   Runaway System Management .................................................................................................2
   Applications and Services Reliability ......................................................................................3
Server Consolidation is the Antidote ..........................................................................................4
UNIX Server Consolidation ............................................................................................................5
Recipe for a Successful UNIX Server Consolidation ..................................................................6
Conclusion .........................................................................................................................................8
Overview

Server sprawl is a term that describes the IT phenomenon in which multiple over-distributed and under-utilized servers are incurring a higher cost of operation than their return on investment. It is an offspring of the marriage of two other influential IT phenomena of the past decades – the highly popular distributed computing strategy and the increasingly affordable server.

In the current economic climate, server sprawl is synonymous with irresponsible spending and unplanned growth, neither of which is tolerated when every company in every industry is looking for ways to cut costs to weather the financial storm of the century.

This whitepaper examines the potential issues with uncontrolled server expansion, methods to rein in those servers, and the associated costs. Lastly, it offers a simple guideline to help manage a server consolidation project.

This document is part of the Open Text Remote X Window Application Delivery Solutions Whitepaper series. For other technical topics, please visit www.opentext.com/connectivity.
Server Sprawl

Server Sprawl is a term that describes the IT phenomenon in which multiple over-distributed and under-utilized servers are incurring a higher cost of operation than their return on investment. It is an offspring of the marriage of two other influential IT phenomena of the past decades – the highly popular distributed computing strategy and the increasingly affordable server.

In the 80’s and 90’s, the IT world was blessed with the flexibility of adding as many low-cost and inexpensive servers as required. Typically, an organization would deploy new applications on new servers and distribute them widely to quickly satisfy business demands. Also, thanks to the increasing affordability of UNIX servers, server sprawl was not limited to Microsoft Windows-based servers. On the surface, small-scale server deployment seemed to be an inexpensive solution that improved business agility; however, in reality this was a rather shortsighted strategy with many trade-offs. The constant influx of new servers finally reached a critical mass in which those trade-offs manifested into a list of challenges that neither the executives nor the administrative members could afford to ignore, as those challenges ultimately affect the business bottom line.

Skyrocketing IT Costs

The initial cost of acquiring server hardware is just a small part of the overall cost of operation; there are other downstream costs such as hardware and software maintenance that must be included in the TCO. Moreover, most organizations have policies that define a 3 to 5 year hardware replacement cycle. At which time, not only do organizations have to shell out valuable IT budgets to acquire new systems, but they may have to enlist professional services to aid in the application and data migration tasks, not to mention manage business disruptions and risks generated by the replacement project.

Server sprawl has also had a deep impact on the infrastructure cost – the networking requirements, power consumption, heat management, and floor and rack space management, which are equally costly.

Runaway System Management

It requires tremendous IT resources to keep track of the server population, maintain operating systems and applications, deploy new servers with new applications, and other general administrative tasks related to system management. In addition, IT organizations are further taxed by tasks related to ensuring the security of the corporate assets – both the physical hardware and the business data. The larger the server population, the harder it is to protect them from theft, viruses, phishing attacks, and other forms of hacking.
Application and Service Reliability

High availability and disaster recovery can effectively protect corporate assets that reside on application servers and data warehouses from natural disasters and a wide variety of other failures. On the other hand, they also add more servers to the already over-populated IT ecosystem and increase the workload of the stretched IT organization. The multiplication of complexity for implementing high availability and ensuring service reliability further magnifies the cost of managing and controlling the systems.
Server Consolidation is the Antidote

Server consolidation is an approach to increase the effective usage of server resources by reducing the number of physical servers that an organization has to deploy and the number of locations affected. It is a compelling solution to server sprawl, and it helps companies to take the control of the cost of management, service reliability, and security. Depending on the size of the organization, the type of existing servers and services, and the nature of business applications, IT policy setters may choose to employ one or more consolidation strategies to achieve desirable infrastructure improvements.

- **Geographical Consolidation**: Physically relocate servers from various local offices to a centralized facility can help centralize server management.
- **Server Consolidation**: Reduce the total number of servers required to provide the same business functions.
- **Application Consolidation**: Reduce the total number of applications required through standardization.
- **Data Integration**: Combine multiple data warehouses into fewer data repositories.

Different consolidation strategies may target various layers of the IT fabric, but the common goal is to ultimately help organizations bring the server population under control by pooling under-utilized resources and logically reducing the number of physical servers. The benefits of reducing the number of servers trickle down through the rest of the problems – the utilization rate of servers becomes higher, and the cost of managing those servers decreases. In addition, streamlined server management can simultaneously benefit the implementation of security management and improve the reliability and availability of servers and applications.
UNIX Server Consolidation

Regardless of the consolidation strategy – where application servers are physically relocated from remote offices to an application center or multiple applications are being grouped on fewer physical servers – the impact to end users is most likely the same: consolidation creates distance between users and the servers where applications reside. Common by-products of the growing distance are an increase in network latency and a decrease in network bandwidth simply because connections between users and applications are now made over the Internet and are therefore subject to network traffic and conditions. Business applications respond to the change in network conditions differently. For instance, consolidating web servers may have a minor impact to end users whereas moving storage servers off-site may make accessing large files over the Internet ineffective or even impossible for end users.

Large organizations are heavily populated with UNIX servers. In almost all cases, they are used to house some of the most complex applications that many businesses depend upon. Semiconductor companies rely on UNIX to run their Electronic Design Automation tools to create circuitry in computer chips that we use in computers, PDAs, MP3 players, and cell phones. The Oil and Gas industry has its own unique set of seismic analysis applications that geologists use to discover oil patches and plan for drilling. The Medical industry uses UNIX-based applications to interface with MRI and CAT SCAN machines. The underlying technology of these UNIX applications is called the X Window System, and these applications are collectively referred to as X Window applications. At the time of development, high speed LAN was the only way users could connect to UNIX servers hosting X applications. The fundamental design of the X Window System is based on such assumptions; therefore, the system was unprepared for the changing network conditions. As a result, X Window applications can be inefficient or even inoperable if servers are relocated outside of a LAN environment. In order to maintain a level of usability of these X applications that is acceptable by users, and guarantee a satisfactory degree of productivity, a successful UNIX server consolidation project must also consider implementing alternative means of accessing business critical X applications over remote connections in order to overcome the productivity, performance, and mobility challenges. That is the key difference between Windows and UNIX server consolidation projects.
Recipe for a Successful UNIX Server Consolidation

Like other major IT initiatives, a UNIX Server Consolidation project has extensive impact to the IT infrastructure, the business processes, and the productivity of the end users. The degree of preparation, the thoroughness of research, the precision of implementation, and the follow-up can greatly affect the outcome and the success of the project. The following guidelines should help most organizations to manage a UNIX server consolidation project successfully.

Defining Goals

Why are you interested in consolidating your server resources? Is it because you are mandated to reduce operational costs and you have identified server sprawl as the area where you are losing the most money? Is the issue more specific in that the islands of servers in your organization are hampering your effort to implement a strong security framework? The demand for consolidation may even come from the user community – segregated application servers and databases make the coordination of global product development ineffective. Whether it is reducing cost, strengthening security, enhancing collaboration, or improving performance, the goals must be identified and then substantiated by business benefits and agreed upon by all stakeholders.

Understanding Applications and Usage

Server Consolidation reduces the number of servers handling applications and users, so before you start slashing machines from your server population, you should understand the impact to each X application’s user community if these UNIX servers are removed or relocated to a remote office. You need to involve your users in an exploratory step to help you gain the understanding of the various types of X applications used by business users, their functions, and how they are being accessed. You may find some X applications used by your organization are text-based, while others rely on a mixture of text and graphical interfaces. There may be some highly complex 2D design tools, or 3D modeling applications, such as CAD/CAM and seismic analysis applications, which top the complexity chart. Generally speaking, the more complex the X applications the higher the impact on users.

The usage pattern of these X applications is also a decisive factor affecting how the server consolidation strategy is planned out: what is the number of users accessing those X applications? Is there a peak hour when most users access the applications? These questions are good starting points towards better understanding your users and their environments.

Setting Tangible Performance Targets

Your study of the X applications and their usage pattern may provide you with information that leads to a new set of constraints. One constraint should be the
minimum performance benchmark that the applications must be able to achieve over a remote connection. It may be measured in the time it takes for an application to launch, or frames per second, or some other quantifiable value. In your discussion with your users, you need to identify this performance benchmark and use it throughout your project to measure your success.

Another type of tangible target is the server utilization rate. It is one of the parameters that organizations often use to quickly assess the ROI. A recent study shows that most organizations have their servers running at less than 20% of their capacities; the remaining 80% is largely wasted. Defining an attainable utilization rate can help you achieve your goals.

**Evaluating Options**

Once you have defined viable goals, gained a thorough understanding of your users and their business processes, and set tangible targets to help you measure your success, it is time to enter the evaluation phase.

One of the first things you need to decide on is the type of consolidation – geographical, physical, or application. Geographical consolidation involves adding distance between users and servers, but it can cut down the number of facilities where server resource reside. Physical consolidation can reduce the total number of individual servers required for businesses to operate. Depending on the goals that you set and the performance targets that you are trying to attain, you may even have to deploy a combination of consolidation strategies.

You also need to evaluate remote X application access solutions for your users to access remote UNIX servers. There are many remote X application access solutions on the market, and your decision should not be based on finding the one with the lowest price sticker. You want to identify the solution that best fits your IT strategy, that can reliably connect your users to X applications over any network connection, and that aptly satisfies the performance targets that your users have set. Also, keep in mind that network connections can be fragile and unstable at times; therefore the solution of choice must be able to retain user productivity by shielding them and their sessions from network hiccups and safe guarding their work against many common system or network failures. Other factors you should consider when choosing your remote X application access solution include evaluating the solution’s ease of deployment, scalability, manageability, and security. Lastly, you should consider the quality of the vendor’s pre-sales and post-sales support and professional services.

**Measuring Success**

We all know too well that what is planned in a meeting room usually does not come to fruition in reality. What you have envisioned as the perfect execution of a server consolidation project may have many obstacles and may not yield the optimal outcomes that you hoped for. Thus, it is important to track and report results, use metrics to calculate checks and balances, and be prepared to make mid-course adjustments to the assumptions and the plan in order to reach your goals.
Conclusion

Server consolidation is definitely one of the hottest and most discussed IT initiatives in the industry right now. The latest research reported by ChangeWave: Corporate Virtualization Software Purchasing Trends (May 19, 2008) shows that out of 1,956 respondents to the survey, only 12% say their company will spend more money on software, whereas 25% say they will spend less. And yet, though companies are slowing down their software acquisition, the same survey shows that spending for virtualization software is one of the only two software categories that are on the rise!

Although there are countless numbers of documents, best practices and whitepapers on the topic, server consolidation is still a daunting task for any IT organization to undertake, and no organization should tackle this IT initiative without a dedicated team of professionals with the necessary technical and project management skills. It is vitally important to get the right people from the right departments involved in the early planning stages and to keep them abreast of ongoing developments. This ensures a higher level of support and cooperation throughout the project span.

Exceed onDemand and Exceed Freedom, the Open Text remote X window delivery platforms are solutions designed for the modern business world and IT infrastructure, where technology resources are being centralized by the likes of virtualization technology, and labor resources are dispersed throughout the world. With Exceed onDemand or Exceed Freedom, the global work force is sheltered from the effect of the server consolidation project. Users are empowered with access to business applications without the burden of the physical distance that separates them.

For more information on the power of Exceed onDemand and Exceed Freedom, please visit us at www.opentext.com/connectivity.