WRITTEN BY ROB BUCKLEY

From theory to practice

How should a company that wants to consolidate its servers go about the process?

It sounds simple enough. A company has a large 'farm' of servers that are typically only running at 20% of their capacity. To save costs, the company would like to move as much of its processing work as it can onto a few high-end servers, and use the remaining ones for other purposes.

Yet how easy is it to re-deploy applications and servers operating in a production environment? And how can an organisation ensure that 'a few' is not so few that its systems cannot cope when things go wrong or when data traffic increases?

"The typical server only has 17% to 22% utilisation," says Mark Lewis, product marketing manager at Sun Microsystems. "If you have 1,000 servers with fairly low utilisation, that's 800 servers not being fully used at a particular time." Many companies looking at those statistics, particularly those with data centres or web farms, begin to consider consolidation. But what should their first steps be, and how many servers can they get rid of and replace with bigger versions when, as Lewis warns, "the flip side is that some servers go through peak loads of 100%. Next to one server running at 20%, there may be another running itself to pieces."

Ian Benn, marketing director at systems and services company Unisys, specialises in server consolidation and says that even the first small steps in consolidation can pay large dividends. "For the most part, step one is for an organisation to standardise servers on one version of an operating system, applications and systems management tools, just as they have done on the desktop. It's low risk, it's medium payoff: you can get a 15% to 20% return on investment. Physical consolidation of servers into one place and redesigning the network can give a 20% return."

But, says Benn, the way to successful server consolidation is to look at server loads throughout the day and see which servers are running which applications. "The secret is to put all the spare capacity together. There's usually a surge in email activity at 9.30am and far less activity during the rest of the day, so you can share the server with an application whose load is greater at night, say."

WINDOWS DATACENTRE

Benn cites an example: "The first Windows Data Center site in the UK belongs to [UK financial services provider] Abbey National; it runs an application that downloads information to local offices last thing at night and handles email during the day." He suggests various combinations and rules for running multiple applications on a single server. If an enterprise resource planning application shares a server with other applications, it should always have priority over them, except first thing in the morning, for instance. "You get big savings, big performance improvements," says Benn,

---

HANDS ON: BANK LABOUCHERE

Server consolidation becomes necessary when companies expand quickly – but not so quickly that they can invest in a large server instantly. The Netherlands-based investment bank Bank Labouchere launched an electronic share trading service, Alex, in 1999 and quickly grew to become Holland's market leader in 2000. As growth increased, so did the number of servers Alex used – to 250.

"The exponential business growth we experienced since creating Alex triggered us to buy several servers a months – sometimes up to 15," recalls Bart van der Vlis, executive director of information computing technology at Labouchere. "With the number of servers we acquired each month, less and less time could be allocated to maintaining each machine, resulting in unequal data storage levels and increasingly difficult disaster recovery."

The company decided to consolidate 90% of its internal processing and e-brokerage servers. It first consolidated over 30 Windows NT-based applications onto a single, 32-processor ES7000 from Unisys running Windows 2000 Advanced Server. However, the company needed to run both Unix and Windows 2000 Advanced Server on one machine, so chose to partition it.

After the first consolidation, it bought seven more to continue the task and replace the remaining servers. The result of the consolidation, van der Vlis estimates, has been a reduction in total cost of ownership of 35%.

with such application consolidation.

But as more applications are deployed on fewer servers, so the number of possible points of failure decreases – and the chances of one fault knocking out a large number of services increase. Mark Lewis confirms that many businesses thinking of going down this path are "alarmed" at the thought that "they're consolidating multiple platforms onto one single point of failure".

"Obviously, overall reliability and security are an issue as you put more and more processes onto a single server," agrees Colin Grocock, new business director, IBM eServer. Along with Sun and Unisys, IBM has been focusing on bringing the fault

cables and two systems clocks in case one of either develops a fault.

All the servers offer considerable amounts of data throughput since the large number of processors in servers of this size requires a large amount of data to be supplied at any one moment to maintain full utilisation. Lewis boasts that the Sunfire 15k can transfer 43GB per second of data to the processors. Tikiri Wandarugala, senior server consultant of IBM Europe, counters that the Enterprise X-Architecture has its data interconnection built into the memory controller rather than the bus and so is capable of scaling to higher levels.

Indeed, IBM claims the unique selling point of its X-Architecture is

"You can set the machine to be completely self-reliant," he says.

Nevertheless, Lewis says that no one would necessarily want to consolidate onto just one machine. "I don't think one server can do everything anyway," he confesses. Alan Priestley, strategic marketing manager for Intel's enterprise marketing group, points out there may be problems of scalability for applications, for instance. "An application might run great on a 4-way platform, but what happens if I put it onto a 32-way platform? What will my back-up strategy and recovery strategy be when I have to back everything up in one go from one server? Do I have the network bandwidth to deliver data out to users from one site if I used to have servers closer to the user?"

Priestley advises clustering groups of larger servers to consolidate at least some of the small servers, but to avoid the issues he highlights.

"Failover clusters are pretty common. The Oracle 9i database system supports failover clusters. And the clusters can offer greater availability than a single server – six 9s possibly, whereas the individual servers may offer only five 9s," referring to systems that support 99.999% up-time. But Unisys' Ian Benn points out that some applications cannot be partitioned or broken into chunks so need to be on a single server.

For companies intending to consolidate servers, Priestley advocates involving a systems integrator or hardware supplier very closely. But Benn suggests that the project should be treated as a small mainframe project rather than a large Windows project if the company wants to succeed. "Most 'Windows projects' fail. Mainframe thinking doesn't go wrong." **B**

> *"Obviously, overall reliability and security are an issue as you put more and more processes onto a single server."*

Mark Lewis, Sun Microsystems:
"The typical server only has 17% to 22% utilisation."

tolerance of mainframe products to its midrange server lines. Its Enterprise X-Architecture and Project Eliza initiatives are part of its concept of 'autonomic computing' – computers that can manage themselves.

"We have self-healing systems via Eliza. Systems can detect errors before critical memory fails, can even cope if areas of memory fail," says Grocock. In addition to the hot-swappable power supplies, fans and drives of low-range servers, Eliza's features include internal sensors to detect faults in components.

Eliza is currently available on IBM's pSeries servers, but the Enterprise X-Architecture that IBM is building into its latest servers will be the backbone of future Eliza developments, offering memory mirroring, hot-addable, hot-replaceable memory and diagnostics software that can run at the same time as the operating system.

Sun, in turn, offers similar capabilities on its Sunfire servers, having fault-tolerant clock boards, hot-swappable clocks and memory boards. Unisys's ES7000, meanwhile, can "lose memory, lose processor cores, lose complete IO modules," says Ian Benn. It even has two mains

that customers can increase the capabilities of their server at any time. For example, they can add a second server to the first and create a box with double the processors instead of just two servers. They can even switch tasks, converting an 8-way into two 4-ways and back again with a reboot of the operating system.

Partitioning is another critical technique for consolidating multiple low-end servers into one or two high-end machines. Using partitioning, a single server can be made to look like multiple servers with different operating systems and resources. "Partitioning is absolutely vital," argues Lewis. Applications can be placed on different virtual servers according to their requirements and mission critical apps can be distributed so that if one operating system collapses, the server and the other operating systems continue unaffected. Lewis claims Sun is the only company to offer dynamic partitioning, enabling servers to reallocate resources depending on load or according to a schedule. An application that runs its server at 80% utilisation can automatically be allocated more processors to match demand.

CONTACT:

Article by Rob Buckley
Email: rbuckley@infoconomy.com