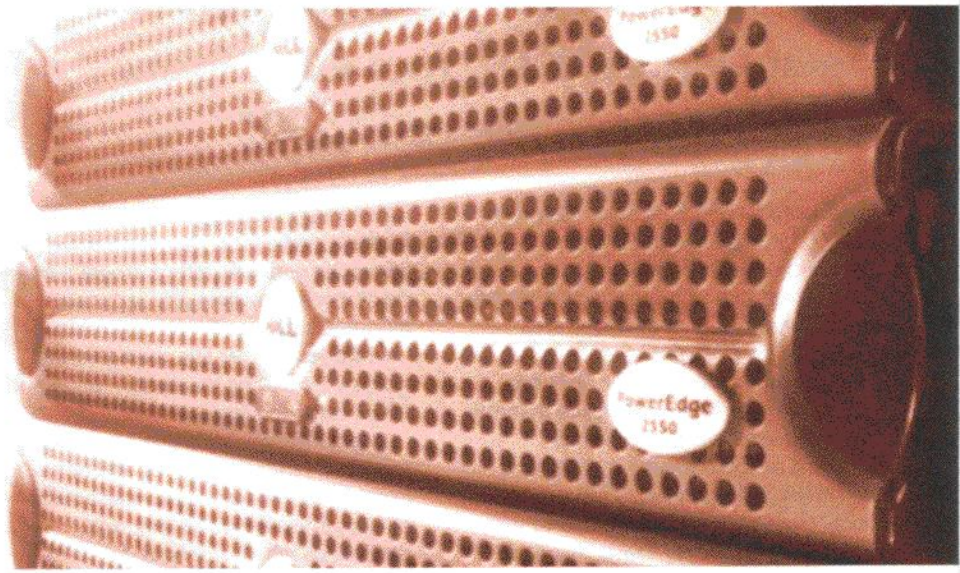
Will IT managers ever find a way to escape the constant round of server upgrades?

Sizing up servers



The scenario is simplistic but far-reaching: Virtually every business with more than a handful of computers has at least one server. It could be a file server or a mail server; it could even be a web server. And as the business grows, the demands placed on the server increase, until the firm has to buy a new one, or upgrade the existing server with new processors, memory, RAID or even a more sophisticated operating system. Then requirements increase again and another upgrade is needed. Then another, and another . . .

Is this the best way to add server capability to a business, though, or is there a better way? The figures would suggest so. Despite the downturn in the world markets, and in the US in particular, servers are still big business in western Europe. Analysts at IDC say that worldwide revenues from servers approached $16.7 billion in the last quarter of 2000, up 14% from the same quarter in 1999, with shipments topping 1.2 million, 16% stronger than Q4 of 1999. In all, that makes $60 billion in revenue on 4.4 million servers shipped in 2000. And while growth is going to slow slightly this year to 7.2%, IDC expects that will be at least sustained through to 2005.

"The western European [server] market has not grown as quickly as in the United States and has, at present, the capacity to take up slack from the US market," according to Martin Hingley, vice president of IDC's European Systems Group. "We expect vendors to look at the fundamentals in Europe – PC penetration is lower than in the United States; Internet build-out still has a way to go; and Europe has not yet achieved US levels of efficiencies from technology use. All these indicate that there is room for expansion in Europe, even ignoring new technologies that require greater computing capacity."

In fact, what led the growth most in 2000 was not so much organic incremental growth, but the purchase of more scalable rack-mounted servers: the small footprint servers that can be stacked with other servers, sharing storage, power supplies, fans and other components. "Rack-mounted systems have been a high point," says Thomas Meyer, research manager in the European Enterprise Server Group of IDC. "The explosion in market demand for these systems, primarily due to reduced space usage and lower administration costs, will ensure their continued growth."

However, according to Tom O'Kill, enterprise marketing manager for Dell UK and Ireland, smaller businesses are more likely to buy traditional tower chassis servers – those like desktop PCs – as are larger enterprises with geographically distributed offices, each with small computing requirements. But as their server needs increase, or if they plan to expand quickly, many look to rack-mounted servers as a way to expand their capacity without expanding their space or maintenance requirements. The small business may not ever need the extra server capacity, so it will not be interested in the initial expense of buying a server rack, which would include heat distribution and power supply for the whole fixture. Instead, it might prefer to buy a single server – or a few, at most that would have these functions built in.

SMALL FOOTPRINTS

Alan Priestley, strategic marketing manager for Intel's European enterprise server group, says that demands for these small footprint servers have been predominately from companies that need a large numbers of servers and for whom extra space is costly. "If you're running a server farm like a data centre or an ISP, you need a colossal amount of space [with traditional tower chassis servers].

But with rack-mounted servers, you can stack them in the bottom of buildings." With racks containing as many as 40 servers, yet only occupying a few square metres, the course for companies needing many servers is clear.

Yet, is anything lost in compacting servers into so small a space? "PCI-card slots are usually the first thing to go. In a tower chassis, you tend to get 10 PCI-card slots, but in the rack servers, you only get one. Some vendors also remove RAID and fast-SCSI options, but most don't." Storage and peripheral connections also come out, but in a rack, the servers can all be connected, via the rack's back plane, to a storage area network (SAN) or peripherals, which can also be rack-mounted. The SAN can contain far more storage space than the tower chassis server ever could.

Most of these smaller servers have relatively limited memory capacities and only two processors, compared to the eight-way or 64-way beasts that have been popular in enterprises until now. They also have far fewer opportunities for expansion and upgradability because of their reduced size. To match even a mid-range server for power therefore requires several of these smaller rivals. Clearly, the break-even point comes when companies begin to run out of space or when enterprises start to need far larger server capacity – the far cheaper, thin servers then begin to measure up against their cousins. It is a matter of simple fractions and addition to determine how many thin servers for a specific vendor, together with their rack, match price and performance with a number of mid or even top-range servers.

O'Kill points out that there are different sizes of server, from 1U (about 1.5 inches) thick upwards, and that these thicknesses affect what vendors can put into them. "In the 1U servers, there tends to be less redundancy. You don't get the hot-swappable power supplies, or the hot-swappable fans. They tend to be built into the rack. As you increase in size to 2U, you can add in those extra levels of redundancy."

So does that mean companies with thin servers are liable to additional maintenance costs, or that thin servers are harder to fix? O'Kill thinks not. "It can be easier to maintain rack-mounted servers [than tower-chassis servers] since they're not spread out among different computer rooms. If a server malfunctions, you can remove it from the rack and replace it with another while you fix it, or, with the higher range servers, pull them out and fix them in the rack." Administration across the network using a program on a client machine is the same for a rack-mounted server as it is for any other 'monitorless' server and the operating systems tend to be the same – Windows 2000, Linux or Netware.

> "If a thin server malfunctions, you can remove it from the rack and replace it with another while you fix it."

Intel's Priestley adds that while rack-mounted servers may bring additional expense at the outset, they can actually be cost-effective in the long run, particularly when problems occur. "It might take four hours to have someone come in and fix a broken server. The question is, is four hours downtime acceptable to a business? If it's not, then simply being able to replace a server is a good thing."

RELIABLE EVIDENCE?

As yet, there is no evidence to suggest that rack-mounted servers are less reliable than their more conventional counterparts, and, as they are made from industry-standard components, they can work out cheaper than a tower chassis, with only the initial outlay for the rack needed. Indeed, many of the reliability issues are more relevant to tower-chassis servers because a single server presents a single point of failure: it is very noticeable if a single server stops working when it is the only one providing a service. But if there are 40, it does not matter so much if one falls over.

So at what point should a company switch from tower-chassis servers to rack-mounted servers? According to Ian Meakin, the UK enterprise server marketing manager for Sun, there is no clear-cut point at which to cross over. "It depends on the organisation's applications and its market. I don't like to think of it as an 'either/or' situation. We usually send a team in to discuss the company's architecture and what its requirements are, rather than say it must have one or the other." Sun's top range Fire server is rack-mounted, using a high-bandwidth (9.6GB/s) backplane to achieve the same speeds as an internal bus for data communication between servers.

Meakin's argument also goes for companies that are looking to buy a mid-frame server that could rival the capabilities of several smaller servers. "A large Oracle database requires 40GB of memory so a smaller server isn't going to be able to cope with that. You need a massively, vertically-scalable server, rather than a series of horizontally-scalable servers as you would for web provision."

While clustering small servers together is an option for load-balancing and failover provision, it cannot support the requirements a large database application places on a single server, Meakin argues. So no matter how good the argument is for smaller, rack-mounted servers, there are some applications that absolutely require a single, highly powerful server.

A company needs to look at what applications it intends to deploy on its servers before deciding how to handle its server needs. It may eventually settle on a hybrid approach – highly-scalable servers with considerable degrees of redundancy and failover for database applications and other programs that need to have single, powerful servers; and racks of thin servers that are far less powerful individually, but together are capable of responding to far larger numbers of transactions, with multiple failover points.

Most server vendors are looking at the dense, rack-mounted server option carefully, as the potential it offers to medium and large enterprises is clear. But the single, highly-powerful server is still invaluable for any enterprise that needs raw power in a single box. ⓘ

CONTACT
Article by Rob Buckley
Email: rbuckley@infoconomy.com