



THE RESOURCE FOR INFORMATION EXECUTIVES
AUGUST 1, 2004

The EXECUTIVE'S GUIDE TO UTILITY COMPUTING

BY CHRISTOPHER LINDQUIST

Pay-as-you-go computing means a lot of things to a lot of people. Use our pass-along guide to explain to your CEO what it is—and is not.

Reader ROI

- ▶ Learn how to cut through the hype
- ▶ Determine which applications are appropriate for utility computing
- ▶ The ROI of utility computing

Utility Computing

IT'S AN APPEALING VISION: You pay for computing power and software only when you need it. No more money wasted on expensive computer servers that are woefully under-used most of the time and software programs that sit on the shelf. Utility computing (UC) is IT efficiency on a grand scale never before experienced—at least in theory.

But what is utility computing, really? Better yet, what does it mean to the CEO? Vendor bombasts would have your CEO believe that the brave, new utility world is just around the corner. And while some businesses are already taking their first steps toward utility (using existing tools and practices), this relatively new form of computing shouldn't be on every company's agenda.

In the interest of making the CIO's life a little easier, we provide this pass-along guide to help CEOs understand whether their company should be on the utility computing path.

WHAT IS UTILITY COMPUTING?

Pinning down a definition of utility computing can be like tying a knot in a snake; just when you think you've done it, the bugger slithers away.

Nearly every major vendor, and a number of minor ones, are promoting their own vision and brand name for utility computing, including Agile Computing, On Demand Computing, N1 and the list goes on. But beyond branding, utility computing is about providing flexible computing resources when and where they're needed. These resources could come from a pool of computing power and software that lives inside a corporate data center, and is metered out and billed to departments as

needed. External service providers also offer such resources, akin to how Salesforce.com currently sells subscription CRM services: You pay for them while you use them or turn them off and don't pay another dime.

"You can provision less and still have the safety margin [to handle unexpected spikes in demand]," says Mike Prince, CIO at Burlington Coat Factory, which is currently moving to a utility computing model built around Oracle databases and clusters of Intel-based hardware. With its prior setup—four separate Unix servers—Burlington had multiple pools of computing resources, with every one of them being bigger than necessary. "None [of the servers] peak or load at the same time of day, day of the month or season of the year. You don't need a 20 percent sludge factor on every single system."

That sludge factor varies dramatically from company to company, but a recent report, "Pay As You Go' IT Services," by analyst firm Saugatuck Technology, states that businesses often have 50 percent or more surplus IT capacity. Utility computing promises to either get rid of that expensive overhead completely or put the surplus to work.

WHY IS UTILITY COMPUTING TAKING SO LONG TO ARRIVE?

Most of the major hardware vendors—including IBM, Sun and Hewlett-Packard—offer some form of physical infrastructure for utility computing already (such as HP's Utility Data Center). But in order to work, utility computing requires a coordinated effort between hardware, applications and management software that can perform tasks such as tracking pools of computing resources. The hardware and management software has

A GLOSSARY OF UTILITY TERMS

Demystifying some terms associated with utility computing

Autonomic computing: The IBM-promoted idea that computer systems should be "self-aware" and able to repair, reconfigure, protect and reallocate themselves with minimal or no human intervention.

Clusters: Groups of computer systems tied together to act as a single computing unit. Clusters are often not geographically distributed, nor can machines be added or removed on an ad hoc basis.

Grid computing: The concept of connecting arrays of smaller computers into a single virtual supercomputer. Grids often involve large numbers of geographically dispersed machines that fall out of or are captured by the grid as they become available. Grid is largely used today as an inexpensive alternative to supercomputers, for research and scientific projects.

Pay as you go: The concept of paying for computing power on a metered basis rather than simply buying a piece of hardware and a software license.

Software as service: Rather than purchasing a software license for a number of users or CPUs and then "owning" that software for some predetermined time, software-as-service agreements allow users to pay for a piece of software on a per-user basis.

Virtualization: The act of pooling servers or storage into a larger "virtual" resource. Rather than knowing exactly which machine or hard drive will be used, users simply request enough space and computing power from the pool to perform a given task. The goal is to increase utilization levels and reduce costs.

—C.L.

UTILITY COMPUTING is about providing flexible computing resources when and where they're needed. You pay for them while you use them or turn them off and don't pay another dime.

arrived, at least in early forms, but it's all still largely in the trial stage for many enterprises. And getting a company ready for utility can take some serious effort.

In today's typical nonutility scenario, a user places a request with an application on a departmental server. The server takes the request and returns the answer, then goes on to other requests. If no other requests are in the queue, the machine sits quietly and acts as a hyperexpensive space heater.

With utility computing (at least in its end-game form), that departmental server may go away entirely, usurped by a collection of machines centrally controlled by the IT department. This could be accomplished either by physically replacing remote servers with a centralized larger server, cluster of smaller servers or rack of blade servers, or by binding the remote systems into a "grid." (For definitions, see "A Glossary of Utility Terms," this page.) When the user makes a request, a series of questions must be answered by the utility infrastructure, determining what resources the user gets from the pool while tracking usage for billing purposes.

Such a level of flexibility and tracking requires management tools that are currently in their infancy, which explains why not every company is jumping on the utility bandwagon (basing your company's IT life on a bunch of relatively untried tools is only for the very brave or the foolhardy). But the real holdup for utility computing is that application providers have yet to move en masse toward UC-ready licensing models. "The software licensing models in particular are currently the barrier to utility pricing models," says Corey Ferengul, senior vice president at Meta Group. Ideally, utility computing pricing models would allow customers to pay "by the sip," much as we do with electricity and water. But software vendors are still predominantly selling their products on a per-seat or per-CPU basis, regardless of how much or how little an individual seat or CPU is utilized.

Vendors are reluctant to move to utility pricing for several reasons, with fear being the biggest factor. A large migration to utility pricing would likely change the Wall Street valuation of software companies dramatically—almost certainly on the downside as vendors would no longer be able to book projected future revenue, but instead could claim only what they collect each month. Utility computing also promises to make it easier for companies to switch software vendors, encouraging healthy competition and reducing margins.

But Ferengul notes that smart software vendors could even-

tually take advantage of what may look like a bad situation, moving to a utility pricing model and locking in customers despite utility's touted flexibility and ease of vendor switching. "[With utility computing] my goal would be to deploy quickly, in hours or days, not weeks, months or years," Ferengul says. "You're not going to stop and evaluate vendors every time."

And like it or not, utility will arrive, he adds, with hardware vendors pushing it as the new model for delivery, and users pushing for software models that match their hardware.

HOW DO I CUT THROUGH THE HYPE TO REALIZE THE PROMISE?

Ask any IT vendor if they have products for utility computing, and they'll likely say yes. But believe it or not, they won't always be telling the truth. Here are two common misconceptions:

It's more than outsourcing. Handing over your entire data center to an outsourcer and paying it to run your applications on its servers is not utility computing. The key to utility is per-use pricing. Few outsourcing deals offer this option; IBM's much ballyhooed, 2002 multibillion-dollar deal with American Express is a notable exception. If your users don't utilize any computing resources for a month, and you still get a bill from your outsourcer, it's not really utility.

It's more than virtualization. Virtualization is the act of consolidating server power and storage space into shared pools. In a virtualized system, you may not know exactly what machine is running your database query or on what disk your quarterly report is stored. Those details are handled by software. This is a big step toward utility, but it doesn't go all the way. In an ideal world, UC allows IT departments to bill users for exact usage—per minute, per transaction or the like. Determining exactly what those units should be is one of the big questions that remains to be answered.

That's not to say that some companies won't opt for the completely external model. When American Express outsourced \$4 billion worth of IT operations to IBM, the transaction was hailed as possibly the first large-scale utility computing arrangement. In reality, the deal was remarkably similar in scope to any other multibillion-dollar outsourcing arrangement. What differed wasn't the technology, but the pricing model. Instead of paying a flat rate, AmEx would in some cases pay only for usage, providing significant cost savings. AmEx Executive Vice President and CIO Glen Salow has claimed that there's potential for "hundreds of millions of dollars" in savings from implementing the utility model.

IS THERE ANYTHING REAL ABOUT UTILITY COMPUTING THAT CAN BE USED TODAY?

Yes, but you need to choose wisely. For example, utility is likely to prove most beneficial in situations where certain high-cost applications get used relatively infrequently. Two companies may run the same engineering package, for instance, with each needing five seats. But one company might use the package once a month, while the other runs it 24/7. For the low-use company, utility pricing would probably appeal because the company wouldn't have to pay full license prices for a rarely used product. But the high-use company would likely do better with standard, per-user licenses.

In addition, certain applications don't lend themselves well to utility. "We're not trying to replace everything out there," says Joe

ERP systems—in Mexico, the United States and Canada—into a single instance of Oracle ERP at headquarters.

For many companies, utility computing will arrive in the form it did at Gates—disguised under names like "virtualization" and "consolidation." As servers and storage become consolidated, IT will need tools to better manage workloads and increase efficiency: the mantra of utility.

Hosted services, meanwhile, present another approach—one that takes the servers out of your hands entirely. Four years ago, Royal Caribbean Cruise Lines began using Akamai Edge computing services for distributing online advertisements because it made more economic sense than trying to build a geographically dispersed server network. Then, two years ago,

SENIOR MANAGEMENT'S JOB will be to smooth ruffled feathers as departmental servers evaporate into a pool of resources that are controlled by the IT department or outsourced to service providers.

Heasley, CIO at industrial products manufacturer Gates. He notes that some applications—such as CAD software—simply perform better on local servers than over a wire.

But in cases where a single application can be shared among multiple groups or where applications face regular peaks and slow periods, utility could be a cost-efficient solution. From there, you need to select the proper approach for your company's situation.

Consolidation is one way to go. Last November, Gates began a move from a distributed-server environment to a centralized HP Superdome system that includes HP's Instant Capacity on Demand (iCOD) feature. With iCOD, Gates can turn on extra processors as needed (and turn them off when the load drops), at a reduced cost. "We decided that consolidation—moving to an environment that is partitionable, virtual and has workload management capabilities—gave us the ability to take that combined horsepower and wield it where we need it, instead of being shoehorned into individual machines," says Heasley. In other words, if a certain application suddenly needs more resources, the new system enables Gates to floor the accelerator and have the business respond.

The Gates project was anything but small. It involved identifying some 42 applications—and their associated hardware—that were prime for consolidation, including custom ERP applications at its Mexico-based manufacturing facilities. In fact, Gates decided to use the Superdome systems to host all of its Mexican computing resources remotely. "We were able to leverage our infrastructure here without adding infrastructure or data centers or outsourcing in Mexico," Heasley says. And under the new model, Heasley's team can redistribute computing power as necessary from a central location. It's worked so well, Heasley says, that the company is currently working on consolidating all of its

RCCL launched a successful advertising campaign that increased its normal load by four to five times as customers logged on to investigate cruises. "I had 18 servers and needed to scale up to 30," says Mike Suttin, vice president of IT at the company. "I have pictures of some of our technicians with servers in their hands running down the hallway." Techs were even forced to cannibalize other projects to keep the website running.

As a result, RCCL shifted more of its Web operations to Akamai's servers, culminating in the piloting of some J2EE applications to Akamai's globally distributed network of IBM WebSphere servers—for which RCCL pays for usage only, not software licenses. Those applications currently access data that resides on RCCL's in-house servers. But that could change, as RCCL gains experience with Akamai and as Akamai rolls out new offerings, Suttin says.

HOW CAN THE CEO HELP US IN ADOPTING A UTILITY COMPUTING MODEL?

If utility is to take hold, cultural change will be a key challenge. By its very nature, utility requires centralization and the sharing of resources in a manner that many companies won't recognize. Senior management's job will be to smooth ruffled feathers as departmental servers evaporate into a pool of resources that are controlled by the IT department or outsourced to service providers.

And the reverse can also be true, with IT managers getting peeved when line-of-business leaders request utility services instead of IT assistance, says Mike West, senior program director at Saugatuck. "In general, the person who is most enthusiastic about [pay as you go] is the line-of-business person who has control of his P&L but not his IT," West says. "On the other hand,

Utility Computing

CIOs are highly skeptical of [the model].” The reasons, West says, are that CIOs are often averse to the additional risk that new technologies can introduce and because they see themselves as technology gatekeepers.

Given that, the CEO and other C-level executives must step in to act as mediators between IT and the line-of-business managers. “If you’re going to do things to get shared resources [such as utility], you’ve got to bring those resources together,” says Jonathan Eunice, an analyst at Illuminata. “That’s why the CEO or COO or CFO at a minimum have to have a say.”

WHAT IS THE ROI FOR UTILITY COMPUTING?

The bottom line, of course, is money. Utility computing promises lower licensing costs by consolidating multiple instances of applications into fewer licenses, and—once pricing models are in place—allowing companies to pay only for the active hours, transactions, megabytes or bits (the exact unit remains to be determined and will probably vary by application). It also

ness processes to be rebuilt on the fly to meet new opportunities.


But that may not matter so much at companies that already run a tight IT group. “If you’ve got an organization that has a good centralized policy base, that has strong governance on change management and how they grow their IT infrastructure, if there’s communication between business units and IT, then the utility model is probably not going to be as attractive to them,” says Gartner Principal Analyst Eric Goodness. Loosely regulated IT organizations, however, could probably benefit from utility’s simplification and control. Unfortunately, however, those organizations are going to be the least prepared for a move to utility. “If you’ve got anarchy, you don’t know what your unit costs are now,” Goodness says. “It is going to take some up-front elbow grease to get [these companies] where they want to be.” Without such preparation, he notes, utility computing may actually cost more than current operations.

Still, certain niche utility markets, such as telecom management services, are also proving cost-efficient for their users. “I’m talking

THE ULTIMATE GOAL of utility computing will be to create a technology framework for a truly agile enterprise, allowing business processes to be rebuilt on the fly to meet new opportunities.

promises reduced maintenance costs, with outsourced UC services handling some operations and internal IT groups freed from laborious tasks such as upgrading countless remote servers one by one. And, as in Royal Caribbean’s case, smaller hardware budgets are possible, thanks to fewer servers and fewer people required to manage them. The ultimate goal will be to create a technology framework for a truly agile enterprise, allowing busi-

ness processes to be rebuilt on the fly to meet new opportunities.

with customers that are saving as much as 30 percent off their corporate telecom spending [using such services],” says Goodness. “It does work, and it’s available right now,” says Burlington’s Prince. “It just takes more engineering than it ought to.” 

Technology Editor Christopher Lindquist can be reached via e-mail at clindquist@cio.com.

