InformationWeek
BUSINESS INNOVATION POWERED BY TECHNOLOGY

The Red Shift Theory

It's Sun's thesis about the explosive growth in demand for raw computing power--but is it more than the utility computing model warmed over?

By Richard Martin, InformationWeek
Aug. 18, 2007
URL: http://www.informationweek.com/story/showArticle.jhtml?articleID=201800873

For nearly a decade as chief technology officer at Sun Microsystems, Greg Papadopoulos has mulled the best way to build and sell high-performance computers, and more recently how to pull the company out of its financial slump. Earlier this year, after multiple conversations with Sun customers and other CIOs, Papadopoulos says he had a lightning-bolt insight--that an elite group of companies are consuming inordinate amounts of IT infrastructure, well beyond most other businesses, and that their demand is growing exponentially. This trend, Papadopoulos maintains, has implications not just for IT's most insatiable consumers, but for the structure of the computing industry itself--and, naturally, for Sun.

Papadopoulos calls this theory the "red shift." In astronomy, the term refers to what happens to the wavelength of light emitted from an object traveling away from the observer: It lengthens, moving to the redder end of the visible spectrum. Predicted by Christian Andreas Doppler, who first observed the phenomenon in sound waves, the optical red shift was used by Edwin Hubble to demonstrate that the universe is expanding.

Red-shifting companies will experience explosive growth, predicts Sun's Papadopoulos

Photo by Gabriela Hasbun

In IT, Papadopoulos uses red shift to describe the rapidly expanding universe of computing demand as data processing requirements--not only from Web companies like Google, YouTube, MySpace, and Salesforce.com, but also from large conventional users of high-performance computing like pharmaceutical, financial, and energy companies--exceed the ability of Moore's Law to keep up.

It's not just about how many CPU cycles a company uses. Papadopoulos argues that red-shift companies will enjoy exponential business growth in the coming years. Blue-shift companies--those whose processing needs aren't exploding--will grow at about the same rate as GDP, he says.

There's an apparent contradiction in this theory. It will be hard for red-shifting companies to grow at exponential rates if they must spend massively to expand their infrastructures. Papadopoulos' answer to that quandary is, naturally, Sun-centric: A shift to lower-cost, lower-risk utility computing, mostly on sophisticated "big iron" servers, will allow these businesses to overcome the inherent limitations of Moore's Law, he maintains.

To be sure, the red-shift theory is part techno-economic insight and part hype. In talks on the subject by Papadopoulos, a former MIT professor, you can detect a hint of the happy pedagogue expounding on his pet thesis. But there's hard evidence as well: mostly in the growth of cloud-computing platform providers, from giants such as Amazon.com, whose on-demand storage service, called Amazon S3, contains 5 billion objects, up from zero less than a year ago at launch, to lesser-known players like 3tera, which is seeing 100% quarterly growth for its utility computing service. The business of selling software as a subscription service is also explosive, growing 43% annually, according to a recent report by RBC Capital Markets.

At the same time, companies faced with the rising costs of powering, cooling, and maintaining racks of servers in conventional data centers are stretching their IT resources beyond capacity--and looking for alternatives. Papadopoulos uses an energy-utility metaphor to capture this shift: "Why build your own generator in your back yard when you can plug into the energy grid?"

YOUTUBE WANNABES

An example of a red-shift company is Twitter, an instant-update social networking service that lets users post brief messages to the Web from cell phones or PCs. The number of Twitter users, now at more than 50,000, posting 30,000 telegraphic updates a day, is doubling every two to three weeks, according to Obvious Corp., the startup that launched the service in July 2006.

Hoping to grow into that neighborhood is GigaVox Media, a company that provides an Internet-based production and distribution platform for podcasters and video bloggers. "Like all startups, we envision ourselves as the next YouTube," founder Doug Kaye says with a laugh. What does that mean in terms of processing requirements? "I don't know," admits Kaye. "But we have to have that level of scalability."

To ensure that GigaVox does, Kaye has chosen Amazon Web Services, an on-tap application, storage, and computing infrastructure. That's one way red-shift companies are adapting--some, like Amazon and Google, are building supersized IT infrastructures, while others, like GigaVox, are tapping into them.

Red shift presents more complex quandaries for large companies. Papadopoulos began to understand the phenomenon, he says, in conversations over the last 18 months. "I would hear this conflict directly from customers, who would first say, 'I've got this really poor utilization--only 10% or 15% of our data center capability is really being used,'" he recalls. "And then I'd hear, 'I'm running out of data center space and power. I don't know how I'm going to continue to manage the growth,' and so on. On the one hand, the infrastructure is being poorly utilized, yet you're bursting at the seams."

The aha! moment came when Papadopoulos realized that there are really two different application sets driving computing demand: one consisting mostly of newer Web-facing applications driving exponential growth in both user demand and computing requirements; the other comprising back-end systems that are growing at more historical rates. "All this is really about which side of Moore's Law you're on," Papadopoulos says. "If your applications are growing faster than Moore's Law, you've got a fundamental set of issues about scale and power. If they're growing slower than Moore's Law, you've got all kinds of opportunities around consolidation."

It's important to emphasize that these rates of growth have seldom been seen in the computing world, even in the early days of the Internet boom. In August, for example, a little-known startup called Yoomba, which offers a peer-to-peer application that lets users place voice-over-IP calls or send instant messages to e-mail addresses automatically, announced that it had signed up 500,000 people in less than a month.

And it's not just user numbers that have IT infrastructures at Web 2.0 companies bursting at the seams. Research firm ComScore says Americans watched more than 8.3 billion video streams online in the month of May--nearly a video per day for every man, woman, and child in the United States. The largest provider, of course, was YouTube, which served up some 1.7 billion videos during that month--a crushing burden for any storage and networking system, even one as expansive as that of Google, YouTube's owner.

Mainstream companies also are feeling the processing pinch. Arizona Federal Credit Union, a midsize financial institution that services 244,000 consumers and small-business employees, is growing about 25% a year. In itself, that doesn't outpace Moore's Law. (In its original form, Moore's Law said that the number of transistors on a chip would double every two years, which is an annual growth rate of about 42%.) Expanding faster, though, is Arizona Federal's credit card division, a data-processing-intensive business that's growing at "an exponential rate" in terms of cards outstanding and transactions per card, says CIO James Phillips. Putting a strategy in place for that kind of growth has been Phillips' primary task since he joined the company in 2005.

Virtualization software let Phillips reduce the number of servers in Arizona Federal's data center to 61 from 87 over the last year, but that sort of consolidation won't keep up with the requirements he foresees in the next few years. "We will continue to expand our data center, but as it reaches maximum capacity we've got to make some long-range decisions," Phillips says. Over time, that could mean moving up to 30% of Arizona Federal's IT infrastructure to hosted services accessed over the Internet, he says.

What if you're much, much bigger? What if you're Visa, one of the world's largest transaction processors?

Visa's overall transactions are increasing at around 20% a year, which doesn't sound like a red-shift rate until you consider the base from which it's growing. The company processes some 6,800 transactions per second during the peak holiday season, and the nature of those transactions is growing more complex. Visa has spent five years and hundreds of millions of dollars annually modernizing and upgrading its transaction system, known as Visa Integrated Payments. Last year, the company opened a data center in the central United States that will process upwards of $1 trillion in transactions this year.

The rapid increase in computing demand is coming not only in Visa's real-time transaction authorization and clearing system but also in what Peter Ciurea, senior VP at Inovant, Visa's IT division, calls open systems. Those include post-transaction analyses, such as monthly aggregate reports for small businesses, in the form of Web services for Visa customers.

Those systems, Ciurea says, entail a "much more database-intensive workload," with a different processing model. And they're driving a shift in Visa's two main U.S. data centers to server virtualization and an internally developed distributed, grid-computing architecture. "In our core transactions business, we see one type of growth, but when you go to these value-added services, now you have to store that data," Ciurea says. "Storage costs are increasing faster than the actual processing costs." The result is predictable: "We need to invest in more and more processing power. My expectation is that we'll see a higher acceleration to that growth once we move to more of a grid model," he says.

That acceleration is being felt across many industries, including pharmaceuticals, where the Food and Drug Administration is pushing new quality-assurance tools known as Process Analytical Technology that have dramatically increased the need for data collection and processing, and energy, where the search for oil in ever deeper and more remote locations, particularly under the ocean, is generating terabytes of geological data that require minute analysis. It's also easy to spot in the software-as-a-service business, where industry pioneer Salesforce has seen the demands on its systems soar. According to Steve Fisher, senior VP for Salesforce's platform division, its "transactions" (i.e., API calls to its database) increased from about 500 million per quarter three years ago to 5.4 billion per quarter this year.

"Over the past four or five years, people have been talking about the commoditization of computing, as if all the innovation were over," Papadopoulos says. "Sure, general-purpose computing is a commodity. But designing really efficient systems to handle these kinds of workloads, and getting productive with the data center software and its management--those are anything but solved problems."

PRESCRIPTION: SHARING

Red shift, in itself, is a diagnosis. The prescription, as Papadopoulos sees it, is a return to utility computing and shared infrastructure.

The notion of utility computing, of course, has been around since the IBM mainframe. What's new is not just an explosion in demand from companies with mountains of data to climb, but a set of technological advances, including the near-ubiquity of high-speed networks; the advent of high-density, high-powered, highly scalable server and storage systems that pack more and more power into a given space; and the maturation of data center virtualization technologies and the software to stitch utility computing grids together more seamlessly.

"The development of the tools and expertise to do this level of orchestration is well understood at a couple of dozen companies," says Forrest Norrod, VP and general manager of Dell's data center solutions division. "We're seeing that percolate across the industry."

It's not just Sun betting on a large-scale shift of much of today's in-house computing loads into the Internet cloud. Big computer companies such as IBM and Hewlett-Packard have offered some form of utility computing for a while. New to the game are Google, which is building a series of massive data centers (the exact number is a closely guarded secret but is thought to be at least two dozen) from which it reportedly plans to offer in-the-cloud computing services to businesses (Google declined multiple interview requests for this article), and Amazon, which debuted its Elastic Compute Cloud service (known as EC2), part of its Amazon Web Services unit, last year.

Database calls have gone

"In our history, we've built up a large, cost-effective, highly scalable, high-performance compute infrastructure," says Adam Selipsky, Amazon's VP of product management and developer relations. "So what we're doing is essentially exposing the ability to tap into that infrastructure to all developers, regardless of their company size or their economics."

Under Amazon's pay-as-you-go model for general-purpose computing, companies pay 10 cents per CPU per hour--a huge savings on typical in-house server costs. Sun offers a similar service at $1 per CPU-hour, which, according to Sun, is actually less expensive because Sun measures only actual processing time rather than the total time period that the customer subscribes to the cloud infrastructure.

The notion of a utility computing boom has plenty of skeptics. "There's no question the concept of utility computing is growing," says Dell's Norrod. "But it's ridiculous to assert, as some have, that it's 1950 all over again and we're going to have seven computers in the whole world to which everything is outsourced. That prediction is exactly as accurate as the first time it was made."

There's a balance between the economics of utility computing and the need to control your own destiny, says Arizona Federal's Phillips. "CIOs want to control the things they have to manage and change quickly, and it's hard to let go of that." Nevertheless, Phillips thinks utility computing "is the wave of future, and we'll see more and more applications moving into that environ just simply for the business proposition of the cost."

That leaves two questions: No. 1, to what degree will utility computing replace conventional in-house data centers? No. 2, which is even more germane to the future of Sun Microsystems: What will those big service-oriented data centers look like?

BLACKBOX COMPUTING

Earlier this summer, a truck pulled up to Building 50 on the grounds of the Stanford Linear Accelerator Center at Stanford University and delivered a jet-black shipping container. Prefabricated, equipped with a built-in cooling system, and crammed with 252 Sun servers, it's one of the first of Sun's so-called Blackboxes. The unit will sift data for physicists examining the effects of subatomic particles colliding at nearly the speed of light. The containerized data center let Stanford add nearly a third more computing power without the cost of new construction, says Randy Melen, the head of high-performance storage and computing for Stanford's department of scientific computing and computing services.

Red Shift Redefined

- Red shift refers to companies experiencing exponential growth in demand for raw computing power

- Red-shift companies tend to be Web 2.0-focused like YouTube and MySpace, or big

financial, energy, and pharmaceutical companies

• Those companies, Sun CTO Greg Papadopoulos says, will experience similarly high levels of growth in users, revenue, etc., while blue-shift companies will grow relative to GDP

• Along with the cost of powering and cooling in-house data centers, the red shift is driving a surge in utility computing and software as a serviceDatabase calls have gone from millions to billions, says Salesforce's Fisher Red-shifting companies will experience explosive growth, predicts Sun's Papadopoulos

Designed for customers that face massive computing requirements but aren't ready to shift all their infrastructure to the Internet cloud, the Blackbox epitomizes Sun's approach to the red-shift challenge. In the future, as Papadopoulos sees it, "the Google model"--massive server farms running hundreds of low-end machines--will give way to high-end, extremely dense units like Blackbox and Sun's Constellation, a highly efficient, low-power system based on interconnected Sun Blade 6000 servers. Designed to run complex applications, such as climate, weather, and ocean modeling, the first Constellation will go online at the Texas Advanced Computing Center at the University of Texas, in Austin, later this summer.

Referring again to the power grid, Papadopoulos uses the analogy of a massive, very hot power plant that produces multiple megawatts of energy versus an array of portable Honda generators.

But the jury is still out on whether the big iron from Sun will win out over arrays of dozens or hundreds of commodity boxes. "We'll see [an exponential] increase in the number of servers sold, yes," says technology pundit Mark Anderson, author of the Strategic News Service newsletter. "Will they be Sun servers, running Solaris? I'm not so sure. Everything is headed toward open systems, mostly Linux on commodity servers."

So far, though, Papadopoulos' bet on big over small seems to be paying off. After five years of losses following the pop of the dot-com bubble, Sun's fortunes have changed for the better since the appointment of Jonathan Schwartz, Sun's former chief software architect and a big red-shift proponent, as CEO in April 2006. Last month, the company reported its first full-year profit in half a decade. In its latest earnings report, Sun said its gross margins also rose, to more than 47%, though revenue was essentially flat. (A week later, Sun said it would go through another round of restructuring, cutting an unspecified number of jobs.)

So, after years of contraction, Sun's universe is expanding once more. The question is the same one facing theoretical physicists: Will the red shift go on forever?