



By
John Verity

Cloud computing changing data management

Got data? Will crunch cheap.

posted 02-28-2009

Average Rating: +++++



In 2007, the computer staff at the venerable New York Times faced a daunting task: quickly convert 11 million articles, all previously scanned from newspaper editions published between 1851 and 1980, into Adobe Acrobat's PDF format.

In scanned form, the NYT's article archive comprised a hefty four terabytes (4,000 gigabytes) of TIFF-format image data—a relatively small amount of data to store, but way too much to quickly organize and convert into PDFs. That job would require some serious computer horsepower—high-speed servers and storage systems costing tens of thousands of dollars and requiring months of programming and testing.

Or maybe not. With a deadline looming and the Times hardly eager to invest so much in a one-off task, programmer Derek Gottfrid turned to the cloud. Gottfrid saw that the Times could rent high-speed data storage and computing capacity out on the Internet—in virtually any quantity it needed and for dirt cheap. Up in the Net's sprawling cloud of datacenters and communications links, everything from raw data storage to desktop word processing to sophisticated business software to time on huge supercomputers, is fast becoming available much like 110-volt electricity from a wall socket. Resources are simply there when they're needed—24/7, in any conceivable quantity, managed by professionals, and easily woven into existing IT resources running down on the ground.

Consider: Once Gottfrid had created and stress-tested his PDF-conversion process on a cluster of four servers rented from a service called Elastic Computing Cloud (EC2)—run by Amazon.com, a leading player in cloud computing services—a few clicks of his mouse was all it took to instantly

harness another 100 machines, and, in less than 24 hours, all 11 million PDFs were done and ready for delivery. Total cost: \$240 for EC2's server time, plus a similarly small amount for having Amazon's storage service, called S3, store the initial scans and finished results.

It's not every organization that needs to make 11 million PDFs in one go, but just about every company, from one-person law offices to global corporations, will soon be using some form of cloud computing—if it isn't doing so, already.

Economies of scale

“The cloud is enabling small, medium, and large businesses to outsource non-strategic IT,” says Andi Mann, research director at Enterprise Management Associates, a leading IT consulting firm. “Cloud computing gives businesses the opportunity to make their investments in IT more strategic and deliver a better ROI.” Most companies' strategies don't pivot on email or CRM (Customer Relationship Management), for instance, so why not hand off those tasks to an outside provider?

It goes back to the basic principle of economies of scale. Specialist companies such as Amazon.com, Google, and IBM have figured out how to run “server farms” that are much larger and more efficient than those within most enterprises, and that gives the specialists a big advantage in cost per unit of computing. Google even designs and manufactures its own stripped-down servers and support gear, which it employs by the hundreds of thousands units per year. What's more, the Internet is now fast and reliable enough to make software running in these distant datacenters almost as responsive as if it were executing just down the hall. Furthermore, the Net has spawned a wealth of ultra-low-cost “open source” software that programmers like Gottfrid can use to harness large banks of remote servers and storage systems—and even write new code for—almost as if they were using a single machine.

Slowly but surely, a growing list of computing tasks are migrating up to the cloud, each one more sophisticated than the last. Users of Apple Macintoshes and Microsoft Windows-based PCs already can seamlessly move text, music, photo, and movie files to Web-based data storage and backup services. Just drag a file or folder onto a desktop icon and within seconds, the items are copied to a disk located ... well, located somewhere safe is all you need to know. What really matters is that the files are now secure and retrievable from practically any device on the planet.

That's just the beginning. Google and other companies have turned traditional word processing, spreadsheet, and email into business-ready cloud-based apps. No software to buy or install, everything takes place in a standard Web browser window, and no data's lost if a hard disk crashes or a laptop is lost. Last year, Abilene Christian University, in Abilene, Texas, moved all of its students to the Google Docs suite of online apps, and many other universities are following suit. With teachers and students storing documents centrally, sharing them is a snap and costs are reduced.

Software as a service

Meanwhile, says Jeff Kaplan, chief analyst at THINK Strategies, small and medium-sized businesses are scrambling to reap the benefits of renting “software as a service,” aka SaaS. Companies such as Salesforce.com, NetSuite, Workday, and Inacct, he points out, are providing Web-based services that tackle beefy, strategic tasks like sales force automation, enterprise resource planning (ERP), human resource planning, and financial accounting. A startup called 37Signals has done well with cloud-based groupware—project management, shared calendars, etc.—that helps dispersed teams work together across great distances. Even software giant SAP is in on SaaS, offering medium-sized businesses an online version of its market-leading ERP system.

“Smaller businesses are in a perfect position to capitalize on SaaS,” Kaplan says. “It gives them the opportunity to get top functionality on a pay-as-you-go basis and reduce upfront costs” such as hefty annual software fees and the labor involved in operating datacenters.

Some issues remain. How can disparate cloud apps, run by disparate companies, efficiently share data with each other? To address that problem and help expand its software ecosystem, Salesforce.com, for instance, invites developers to build their code on its SaaS platform.

Still, interest in SaaS is skyrocketing. THINK Strategies has seen adoption grow from 32 percent of the companies it surveyed in late 2007 to 63 percent in 2008. And of those already using SaaS, Kaplan says, more than 90 percent said they were satisfied and would recommend the option to their peers. Advantages? With more than 30 percent of in-house IT projects failing to reach fruition for one reason or another, and with the cost of maintaining apps around 10 times the cost of initial deployment, SaaS appeals as a way to avoid heavy upfront costs and mitigate risks. Says Kaplan: "You can start small and expand, and that raises the likelihood of success."

Clearly, even the largest corporations also stand to benefit from outsourcing certain IT tasks to cloud-based providers. Last fall, GE business units around the world went live with an on-demand service from Aravo Solutions, San Francisco, that helps them to manage information flowing into GE from some 500,000 suppliers. Not only does Aravo's service make for worldwide consistency across GE, it took only 7 months to be deployed in six languages.

A few more caveats and one solution

For now, however, CIOs cite three major risks that make them cautious about moving most or all of their IT operations into the cloud: reliability, security, and compliance. What if the Internet or a SaaS provider fails in some way? That could leave a company twiddling its thumbs, unable to conduct business. And how secure are apps and data residing in a pool of "virtualized" cloud servers and storage systems shared among several customers? Many big companies aren't ready to risk seeing sensitive data leak to competitors, or get stolen. And how can an organization satisfy federal regulations like the Sarbanes-Oxley Act, which calls for documenting precisely where important business data is stored and who has viewed and edited it?

One answer to such concerns is to build one's own cloud. The "internal cloud" may not physically reside on the premises of the organization it serves; it could be located in an IBM datacenter, for example. But its hardware and software resources would be dedicated to a single customer, who would enjoy virtually all of the flexibility and cost advantages of a shared cloud setup, but with added security.

External or internal, the cloud approach also may prove compelling as a way for companies to further "green" their IT departments, too. Computing has emerged as a significant consumer of electricity for corporations, both in powering servers and in cooling them. But if several enterprises pool their IT work in a shared datacenter, with banks of servers set up to automatically take on new tasks whenever they have capacity to spare, each machine can achieve a significantly higher average utilization rate. Today, many servers are idle as much as 80 percent of the time, but a pool of fully "virtualized" servers can drive average idle time below 10 percent. The result: significant energy reductions, because fewer servers will be required.

Even more energy could be saved if PCs and laptops were made simpler by depending on the cloud for the bulk of their processing and data storage needs. The cloud model requires more energy to be consumed by communications links, but most experts calculate that the net result will be significant gains in energy efficiency.

The details of cloud computing are evolving rapidly, but the cloud will surely be an important concept in coming years, especially as enterprises large and small seek shelter from economic storms.