

Innovation for Innovators

an occasional column exploring principles, models, and theories of innovation in business and management

Roger Smith

Computing in the Cloud

Inside a New York City loft in 2006, two entrepreneurs are working on the software to launch a new dot.com business. They hope to offer a service that will allow people to turn a series of photographs into a simple movie with a nice sound track in the background. Though their business offers an online version of something found in most photo viewing applications, they hope that the ability to send the finished product to friends and relatives will make their service a hit with millions of amateur photographers. They have finished the basic functionality of the software, but their next major hurdle is attracting enough venture capital to purchase the server farm and network bandwidth necessary to go live on the Internet. Unfortunately, this will require that they give up significant ownership and control of their company to venture capitalists. But this is how all dot.com companies get started, hoping to make it big like Yahoo!, YouTube, or MySpace.

But, in the 21st century, these entrepreneurs have a new alternative that was not available to start-ups of the past. They can rent the necessary computing, storage, and communication capacity from a large service provider who already has all of these assets connected to the Internet. They can pay only for the volume of these services that they use, they can quickly add or subtract resources from their order, and they never have to take possession of the hardware and all of the technical support headaches associated with it. This will allow them to retain more ownership in their fledgling company, hopefully keeping millions of dollars into their pockets instead of the venture capitalists' pockets. They can launch their company "in the cloud". They can tap into the "cloud computing" services offered by a big vendor like Amazon.com, relying on Amazon's ability to purchase hardware cheaply, maintain it reliably, and staff it competently.

Animoto has become one of the poster children for the cloud computing concept. This producer of online slide-shows was able to launch their company without purchasing millions of dollars worth of computer equipment, hiring an IT department, or selling majority ownership to venture capitalists to fund their company. As a start-up, they really had no idea how big a customer base they could attract, and therefore, did not know how much IT equipment they would need to purchase. But Amazon.com's cloud computing service, known as Amazon Web Services, offered a means to rent this capacity by the compute hour, storage gigabyte, and network gigabit. Amazon also offered to scale up and down as the customer demand increased or decreased. This was perfect for Animoto since they could not predict demand – and since the demand they actually experience varied widely as their service was featured in various press stories. Over one three day period they rocketed from 25,000 registered users to 250,000 as the result of a posting on the popular web site Slash.dot. In a one week period they ramped up their usage of

Amazon computers from a couple of dozen machines to nearly 5,000 machines (Figure 1).

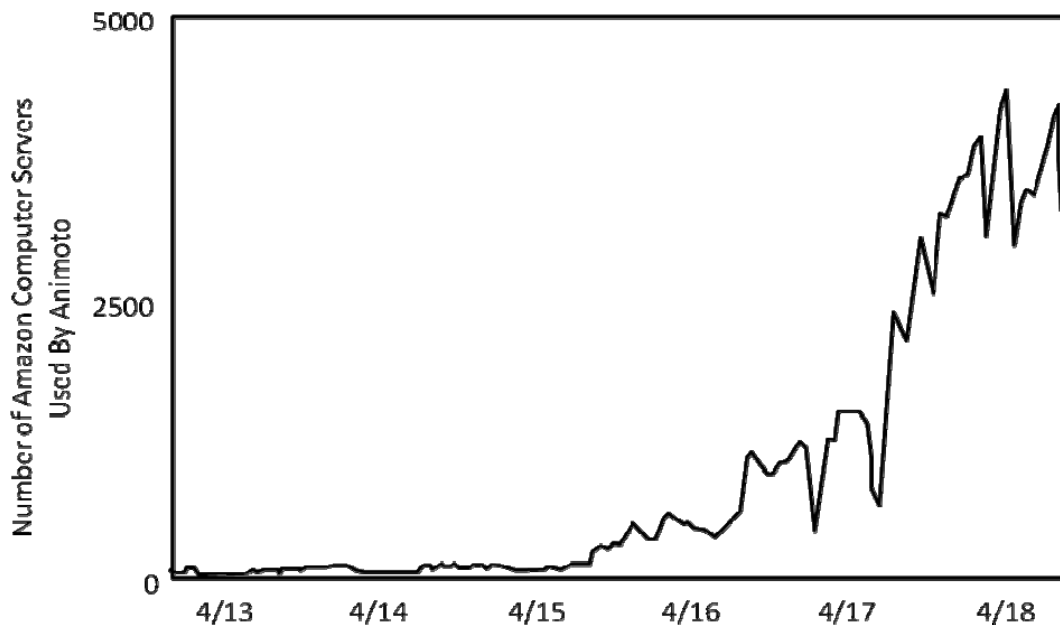


Figure 1. Animoto demand for computer servers from Amazon Web Services during April 2008.

Source: Information Week, January 23, 2009

What is it?

Cloud computing has become a very hot term in the last few years, but a clear description of what it is, what it can do, and why companies might use it is often difficult to find. The concept and a number of predecessor technologies have been around for decades. Gartner, Forrester, and major computer companies offer their own definitions of the term [see Side Bar at the end]. In essence it is a means of renting computers, storage, and network capacity on an hourly basis from some company that already has these resources in their own data center and can make them available to you and your customers via the Internet.

Nicholas Carr has suggested that computing will follow electricity generation in the pattern of business use [Carr, 2007]. A century ago, most companies had to build their own dedicated power generation capabilities. Today, that is performed by an electric utility and companies only purchase the amount of electricity that they need each day, each hour, and each minute. A century ago, power generation was a barrier to market entry because of the huge up-front cost to build dedicated factory generation capacity. In the late 20th century, the cost of IT was a similar barrier to new companies. This was initially overcome through the use of venture capitalist financing for the equipment. Cloud computing offers to eliminate this barrier by turning computing into a utility that companies can purchase based on specific demand.

Telecommunications went through a similar shift in its structure in the 1990's. Traditionally the telecom providers had installed and sold a hard-wired capacity between fixed destinations. But in the 1990's they sub-divided the capacity, packaged it into Virtual Private Networks (VPN), and sold volume between destinations that was really composed of many linked segments that had been grouped into a VPN. This was known as the "telecom cloud", the term that evolved into the current cloud computing moniker. [Wikipedia]

Amazon.com has become the leader in providing cloud services to thousands of businesses in the last few years. IT companies like IBM, HP, Sun, Microsoft, and many smaller companies are getting into the business and base their own cloud offerings on the model pioneered by Amazon. Each offer a unique package of services, security, accessibility, reliability, and support that they feel will appeal to the customers they are interested in.

Advantages

Prior to the popularity of cloud computing, there were a number of related service offerings that attracted only minor attention – i.e. grid computing, utility computing, elastic computing, and software as a service. The basic technologies from each of these has been incorporated into cloud computing and appears to be attracting more interest than these predecessors. This attraction may be a function of the maturity of the technology and the services offered, or it may be driven by a marketing blitz that has occurred as Amazon, Google, Apple, and other big names have gotten behind it.

There are a number of business benefits to this technology.

First, cloud computing is dynamically scalable. Businesses can draw as much computing power as is necessary on an hourly basis. As demand from internal users or external customers grows and shrinks, the necessary computer, storage, and network capacity can be added or subtracted on an hourly basis. Most service providers leave this provisioning up to the customer; though automating this is a valuable advantage that is being pursued by those following in Amazon's wake. Second, the resources can be purchased with operational funds, rather than as a capital expenditure. Many IT departments face a long approval process for capital funding, in addition to the wait for equipment delivery and installation. Cloud computing allows them to bring capacity online within a day and to do so using their operational budgets. Third, the equipment does not reside in the company facility. It does not require upgrades to the electrical system, the allocation of floor space, modifications to the air conditioning, or expanding the IT staff. Computers at Amazon.com consume space, power, and staffing support at Amazon instead of within the customer's company. Fourth, there are competing providers for this service. If the first cloud provider does not deliver acceptable performance, a company can always shift their business to another company offering better service or lower prices.

Concerns

Of course, each of these advantages has within it a corresponding disadvantage or concern. First among these is security. Many companies are hesitant to host their internal data on a computer that is external to their own company and that is potentially co-hosted with another company's applications. So far, there has been no client-to-client penetration of software or data hosted in the cloud. That may be due to sufficient security provisions, or it may be because there has been no value in this kind of attack in the past. The second concern is location. Companies may be concerned about the physical location of the data that is being stored in the cloud. The laws of the host country of the equipment apply to the data on the machines. European and Asian companies have expressed concerns about having their data stored on computers in the USA which fall under the jurisdiction of the US PATRIOT Act, allowing the U.S. government to access that data very easily. Third, experienced users of cloud computing services have noticed a big variation in the performance of their applications running in the cloud. When you start an application in the cloud you never know who your neighbors will be on the allocated computer and network. Since many companies are all sharing the resources, it is possible to arrive in a neighborhood that is extremely busy and very noisy, leaving little room for your applications to run and communicate. Fourth, interesting bugs in this large system have yet to be worked out. There have been instances in which entire cloud services have crashed and been unavailable for hours or days. When this happens, your application will be offline until the larger problem is fixed. Fifth, each cloud vendor offers unique services and unique ways to communicate with the computer resources. It is possible for your company to get so deeply embedded into these unique and proprietary services that you cannot move your applications without some major changes to both your software and your data. Sixth, it appears that a cloud provider has an infinite number of computers and storage disks to meet your needs. But, there are a finite number of these resources available and your provider is multiplexing these between the thousands of applications that are starting and stopping every hour. If all customers called for services at the same time, the provider could run out of available resources. This is the cloud computing equivalent of a busy signal on Mother's Day or an insurance claim following a major hurricane.

These concerns, and others that are much more technical in nature, are well known to the major cloud computing providers and the intermediary support companies. All are working on solutions to eliminate them. The real question is, "Can your company increase profits or decrease costs by using cloud computing as it exists now?" and "How will that change as the technology matures?"

Where Will It Go?

If Nicholas Carr is correct, and computing follows the path of electricity in becoming a utility, then it is just a matter of time before all companies have added cloud computing to the arsenal of technologies that support their internal operations and their external customers. But it is not clear how long this might take. Paul Saffo has said, "Silicon Valley is littered with the corpses of companies who mistook a clear view for a short distance. One of the secrets in my business is that everything changes slower than people imagine. Change only seems fast because people overlook the antecedents. Most ideas

take 20 years to become overnight successes.” [Saffo, 2007] Even if some technological change “has to happen”, it is never clear *when* it will happen. Customers did not jump on utility computing and it is still too early to know how many will really adopt cloud computing, though the numbers appear to be much higher than its predecessors. In support of the adoption of their experience with cloud computing, Amazon offers Figure 2 which compares the bandwidth of merchandise buying customers with the bandwidth of thousands of cloud computing applications in their data centers. In mid-2007, they began using more bandwidth for cloud computing than for their traditional retail business.

Cloud computing service providers have made their systems so inexpensive use and easy to access, that there is little reason that companies should not be exploring this option for providing data and services that are not proprietary to their business. There are many forms of internal data exchange, online courseware, and regulatory information that can be hosted and distributed using cloud services. These can give the internal IT department firsthand experience with the ease of use, cost effectiveness, and available functionality of clouds without disrupting or exposing the company's core operations.

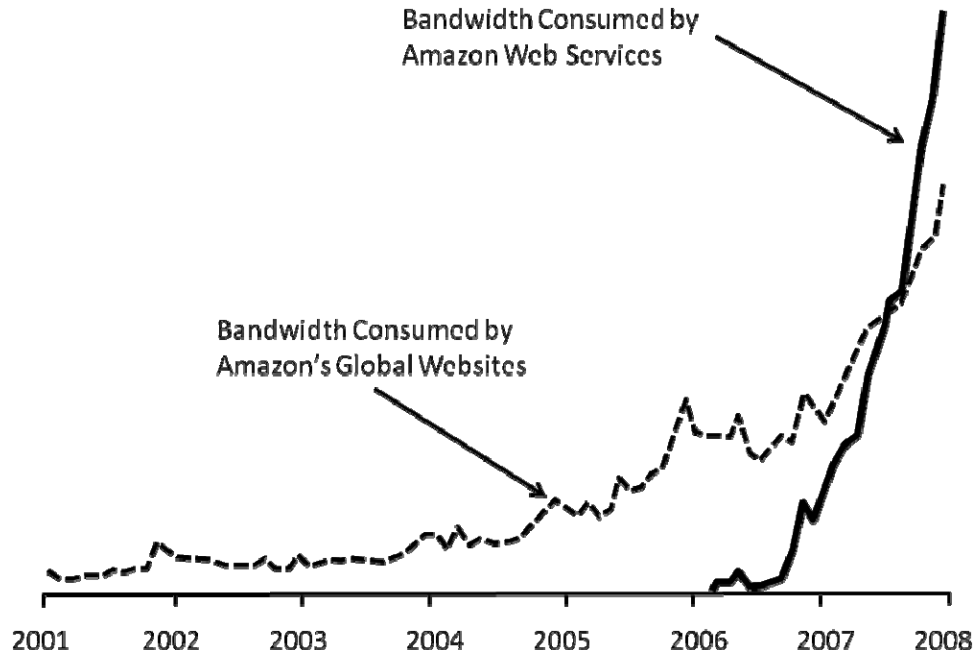


Figure 2. Network bandwidth used by Amazon.com retail versus Amazon Web Services
Source: Vogels, 2008

Cloud Computing Resources

1. Armbrust, M. et. al. (February 10, 2009). Above the Cloud: A Berkeley View of Cloud Computing. <http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.html>
2. Baker, S. (Dec 13, 2007). Google and the wisdom of crowds. Business Week Magazine, http://www.businessweek.com/magazine/content/07_52/b4064048925836.htm

3. Boss, G. et. al. (2007). Cloud Computing. IBM White Paper. http://download.boulder.ibm.com/ibmdl/pub/software/dw/wes/hipods/Cloud_computing_wp_final_8Oct.pdf
4. Carr, N. (2008). *The Big Switch: Rewiring the world from Edison to Google*. W.W. Norton & Co. <http://www.nicholasgarr.com/bigswitch/>
5. Lucky, R. (May 2009). Reflections on Cloud Computing. *IEEE Spectrum*. <http://www.spectrum.ieee.org/may09/8773>
6. Open Cloud Manifesto (2009). <http://www.opencloudmanifesto.org/>
7. Paul Saffo. (July 16, 2007). The Future Really is Now. *ComputerWorld*. http://www.accessmylibrary.com/coms2/summary_0286-33434502_ITM
8. Vogel, W. (April 15, 2008). Ahead in the Cloud: The Power of Infrastructure as a Service. Amazon presentation at MySQL Conference. <http://mysqlconf.blip.tv/file/981534/>
9. Wikipedia article on "Cloud Computing". http://en.wikipedia.org/wiki/Cloud_computing

Side-Bar: Cloud Computing Definitions

Gartner: Cloud computing is a style of computing where massively scalable IT-related capabilities are provided as a service across the Internet to multiple external customers.

Forrester: A pool of abstracted, highly scalable, and managed infrastructure capable of hosting end-customer applications and billed by consumption.

The 451 Group: The cloud is IT as a service, delivered by IT resources that are independent of location.

Wikipedia: A style of computing in which dynamically scalable and often virtualized resources are provided as a service over the Internet.

IBM: A cloud computing platform dynamically provisions, configures, reconfigures, and deprovisions servers as needed. Cloud applications use large data centers and powerful servers that host web applications and web services.

UC Berkeley: The illusion of infinite computing resources available on demand, the elimination of up-front commitments by cloud users, and the ability to pay for use of computing resources on a short-term basis as needed.

Roger Smith is the chief scientist and chief technology officer for U.S. Army Simulation, Training, and Instrumentation, in Orlando, Florida. He has also served as a group-level CTO for Titan Corporation and as a vice president of technology for BTG Inc. A member of RTM's Board of Editors, Smith has led technology innovation for software and computer systems for military training and command systems. He holds a Ph.D. in computer science and Doctorate in business administration.

