SCIENTIFIC AMERICAN™

Online 24/7: "Life Logging" Pioneer Clarifies the Future of Cloud Computing

Microsoft researcher Gordon Bell, paperless for more than a decade, envisions data centers saturated with information and services readily available via the Internet

By Larry Greenemeier | Wednesday, May 4, 2011 | 3

The idea of cloud computing is to make all the information and services run in data centers around the world available via the Web. The reality of this is daunting. Data centers built by different businesses, government entities and research institutions are not inherently designed for sharing, and not all information can (or should) be available to anyone with access to a Web browser.

Many people think that as mobile, connected devices proliferate and broadband access expands, the cloud-computing model will prevail because it enables access to data and apps without the need for a lot of storage or processing power on the devices themselves.

Perhaps the foremost of those people is Gordon Bell, a principal researcher with Microsoft Research and a veteran "life logger," who spoke with *Scientific American* about what distinguishes cloud computing from the other types of Web services, why scientists need to get on board with the cloud model, and why someone would want to store a lifetime of memories in the digital abyss.

Bell considers cloud computing to be a new chapter in the eponymous Bell's Law, which he formulated in 1972 to describe how different approaches to computing arrive, evolve and eventually die out (or at least fade into the background). These new approaches come along roughly every decade and promise to make computers cheaper and more accessible. In the 1960s the mainframe introduced distributed computing and dumb terminals into the workplace. This was followed by minicomputers that essentially made mainframe capabilities available to smaller businesses. PCs followed, extending the reach of computing into the home and eventually allowing the Internet to grow in prominence. Most recently, wireless gadgets have allowed us to take computing with us wherever we go. The "cloud" is poised to take computing to the next level, according to Bell.

Although Bell has been working for Microsoft Research for nearly two decades, he makes a point that the views expressed in this interview are his alone and not those of his employer.

[*An edited transcript of the interview follows.*]

How do you define "cloud computing"?
It is the next computing platform as described by Bell's Law of computer class formation. Like all new platforms, we can look at it in terms of the four functional components: storage, computational ability, network and user interface. With cloud computing the emphasis is on storage and networking to enable wide-scale, 24-by-seven access to data needed for transactions—scientific, financial or otherwise.

What distinguishes it from earlier hardware, network, application and data-hosting services?
In some ways nothing. The cloud has evolved from the large number of distributed servers that hosted Web content. What is different is the scale of these servers—tens of thousands of computers consuming 50 megawatts of power and hosting thousands of customers. Instead of each customer maintaining their own isolated servers, a hosting company is selling access to their servers as a service. The customers share computer systems, power, data-center space and maintenance services.

At what stage are we in the evolution of cloud services?
Amazon was first to use a cloud-computing model for their business and now is the leader in providing cloud services to other

businesses. Entrepreneurs are exploiting Amazon's Web services, Microsoft's Windows Azure hosting platform and other cloud services in order to start up companies because of the zero capital equipment requirement. Payment is by credit card, and you pay as you go.

Are most people using the cloud in some way today?
Sure. Consider *iTunes*, *Dropbox*, *Salesforce.com* and *HealthVault*—[the latter of] which lets you store and share your medical information, as opposed to, say, the Epic software used by your local hospital. Start-up companies offering online games, project management tools and other services are other ways people are using the cloud.

How are cloud services impacting science?
For science, cloud services haven't really started to any measurable degree. However, for science the cloud is inevitable, driven by several factors. Universities and other research organizations maintaining their own high-performance compute clusters will start to see the cost benefit of having someone else manage these systems. The life of data in a high-performance cluster is suspect and probably only as long-lived as the student is running the experiment. And there may or may not be any redundancy or backup for the data they produce.

Cloud computing offers scientists access to data across a number of research organizations. As science grows beyond a single lab, the administrative details and network costs to support a scientific community requires standards and overhead that are beyond a single lab or university computation center's mentality. Homegrown, grad student–managed computer systems positioned as a type of mini cloud providing 24-by-seven access to data will give way to commercial cloud services that have geographic redundancy and higher reliability. Scientists are also facing pressure to make data available forever, particularly when their experiments are publicly funded. Even more relevant, larger scale experiments driven by the competitive research market and fed data by ubiquitous sensors are producing terabytes of information that are too expensive to manage on in-lab servers. Then there's the skill required to maintain these systems. Is the goal to train computer operators or have the graduate students work on science?

What is "life logging" and how does Microsoft's MyLifeBits relate to this?
MyLifeBits, which I wrote about with Jim Gemmell in the March 2007 issue of *Scientific American*, is a Microsoft Research project to provide people with the tools needed to compile a lifelong digital archive, or life-logging. It is the fulfillment of Vannevar Bush's 1945 memex [hypertext] vision—a digital repository of information accumulated throughout one's life to supplement one's own memory—including full-text search, text and audio annotations, and hyperlinks.

Since 2001 Jim Gemmell and I have demonstrated many aspects of complete life-logging—storing letters, papers, photos, videos and voice recordings associated with my life in an annotated and searchable database. The advent of digital cameras, biosensors and GPS means we can now log everything about an individual in real time, from location to aspects of their physical state such as energy expenditure, heart rate and stress levels.

Utopian vision or dystopian nightmare? The extent of future life-logging will depend not least on the laws and norms we establish on privacy. For example, what right do we have to record our interactions with others? But life-logging's potential to benefit individuals' lives and society generally is immense. In 2009 researchers in the U.K. showed how life-logging with a time-lapse camera can aid those suffering from memory loss to regain control of their lives. For social scientists extensive life-logging will mean an unprecedented flood of data to further our understanding of human behavior. And for each of us it could mean a chance for a little, limited immortality.

For all of the information collected throughout a person's life to be useful, it must be searchable. To what extent can different types of data (documents, audio, video, etcetera) be easily searched?
It is getting better every year. All photos are evolving to be tagged based on geography and time, which helps systems identify them. Printed documents have been searchable for a decade, although I don't believe that handwritten documents are being worked on yet. Video is related to pictures and is being addressed.

You've been paperless for over a decade. How easy is it for you to sift through all of this information to find what you are looking for?
Rarely do I ever give up on finding an item, whether it is a photo, e-mail or document. Those files are always with me and I can work anywhere.

Data breaches have become commonplace, with businesses exposing personal information as a result of cyber

attacks or lost laptops. What do you say to people who are concerned about security and privacy as the world increasingly goes digital?

Well, we have two cases: data is on a local computer or held in the cloud. Anyone who has corporate or institutional data on a local computer really must encrypt their disks in case someone steals a system. This capability is built into Windows. People can protect their personal computers to varying degrees by physical isolation including local data servers and external hard drives. We continually work on making PCs connected to the Web harder to penetrate, especially to access by attacks. Staying off the Web is ideal, basically the idea behind firewalls that we all use. For those who store everything in the cloud, I know of no examples of a major attack where everyone's data is exposed, but it may come.

Amazon recently experienced extended downtime to servers that enable many of the company's customers to engage in cloud computing. Should this serve as a note of caution to anyone considering the outsourcing of all their data and software to a service provider?

I don't think the outage will affect the adoption of cloud computing. Many of Amazon's customers may have been affected, but this is probably the same amount as it would have been if you summed up the downtime incurred across all the separate systems that had operated independently. The outage will undoubtedly affect how [future] applications are designed.

What is next for cloud computing?

It will just continue to grow with more capabilities and scope. I would like to think that science will eventually get with the program that we tried to outline in the 2005 book *The Fourth Paradigm.* How about for science we start to work on what the industrial world has started, building on the massive investment and inevitable cost declines that will accompany economy of scale? The next stage will see sensor data being fed continuously to the cloud, pretty much in the same fashion mobile users interact today.