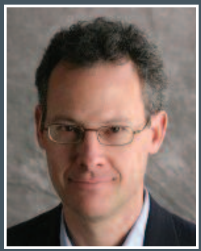
IT in 2018:
From Turing's Machine to the Computing Cloud



By Nicholas Carr

an internet.com IT Management eBook

IT in 2018: From Turing's Machine to the Computing Cloud

By Nicholas Carr

In 1936, as the clouds of war gathered once again over Europe, a 24-year-old Cambridge University mathematician named Alan Turing invented the modern digital computer. At least, he invented the *idea* of the modern digital computer, which, as it turned out, was far more important than constructing any particular physical manifestation of that computer.


Jupiterimages

Turing's theoretical apparatus, which he called a "universal computing machine," was a simple one. In essence, it had the ability to read or write symbols – a one or a zero, say – on an endless roll of paper. It could only take one action at a time, reading or writing a single symbol, but it could remember what it had done, and over an infinite span of time it could take an infinite number of actions.

What Turing had created was, in the words of the historian George Dyson, "a single machine that can exactly duplicate the behavior of any other computing machine." Any calculation, no matter how complex, can be reduced to a series of discrete, simple steps – an algorithm, or a code – and carried out by Turing's machine. What that means, quoting Dyson again, is that "in principle all digital computers are equivalent; any machine that can count, take notes, and follow instructions can compute any computable function." What it also means is this: "Software (coding) can always be substituted for hardware (switching)."

The only real constraints on a universal computing machine are the size of its memory and the speed with which it can carry out its calculations and transmit the results. With enough memory and enough speed, Turing's work implies, a single computer could be programmed, with software code, to do all the work that is today done by all the other physical computers in the world.

And that is why the modern corporate data center, with all its complex and expensive stacks of machinery, is on the path to obsolescence.

But let's not get ahead of ourselves.

> With enough memory and enough speed, Turing's work implies, a single computer could be programmed, with software code, to do all the work that is today done by all the other physical computers in the world.

From Hardware to Software

Randy Mott has been called a "CIO superstar." He originally came to fame during the 1990s, when as head of Wal-Mart's IT organization he played a central role in using computer systems to streamline the giant retailer's supply chain and fine-tune its use of store sales data, giving it a big advantage over K-Mart, Sears, and other rivals. Mott left Wal-Mart in 2000 to join Dell Computer, where he worked his technical magic in expanding the company's IT operations and helping cement Dell's place as the most efficient PC maker.

In 2005, Mott jumped ship again – just as Dell was about to hit some rough waters – and became CIO of Hewlett-Packard. He was lured to the post by the company's new CEO, Mark Hurd, who knew he needed a strong leader to rebuild HP's sprawling and tangled information systems from the ground up. Mott's charge at HP wasn't to redesign and automate processes in other parts of the business – it was to redesign and automate the IT function itself.

"Over the last decade, IBM has replaced 155 of its traditional data centers with just seven modernized facilities."

And that's what he's been doing. Mott and his team are in the process of closing down the 85 data centers that HP has been operating around the world to run its internal systems. They're being replaced by just six giant server farms in the southern United States – four in Texas and two in Georgia. The new centers will be highly automated, able to be managed by wire from a central command post and requiring few on-site employees. Combined with an effort to rationalize the company's application portfolio, the data center consolidation is expected to reduce HP's total IT workforce from 19,000 to 8,000 and trim its IT costs as a percentage of revenues from 4 percent to 2 percent – even while, as Mott put it in a recent speech, "delivering more to the business."

HP is not the only IT company in the midst of remodeling the way it deploys IT. Over the last decade, IBM has replaced 155 of its traditional data centers with just seven modernized facilities. Last summer, Big Blue announced it would go even further, replacing nearly 4,000 of its servers with just 30 Linux-based mainframes. In the process, the company says, it will slash its application licensing costs and save enough electricity "to power a small town."

The consolidation programs of HP and IBM may be particularly aggressive, but they're not unusual. Big companies of all stripes are today moving to save billions of dollars in IT costs by dramatically reducing the number of data centers and servers they run.

And they can all thank Alan Turing.

Turing's discovery that "software can always be substituted for hardware" lies at the heart of "virtualization," which is the technology underpinning the great consolidation wave now reshaping big-company IT. As the cost of computing power and storage capacity has continued its decades-long freefall, it's become possible to turn more and more hardware into software code – to use a single powerful computer to run many virtual machines.

All the pieces of hardware stuffed into corporate data centers – not just servers but storage drives, load balancers, firewalls, switches, and even the cables connecting the gear – exist, after all, to carry out instructions. Virtualization simply turns the hardwired instructions into code and gets rid of the physical machinery. That not only saves a lot of cash, it makes the radical automation of formerly manual IT processes possible. Once IT infrastructure turns into software, it can be programmed, easily and from afar. Code, as always, replaces labor.

The Data Center in the Cloud

If you're of a certain age, you'll probably remember that the first telephone answering machine you used was a bulky, cumbersome device. It recorded voices as analog signals on spools of tape that required frequent rewinding and replacing. But it wasn't long before you replaced that machine with a streamlined digital answering machine that recorded messages as strings of binary code, allowing all sorts of new features to be incorporated into the device through software programming. But the virtualization of telephone messaging didn't end there. Once the device became digital, it didn't have to be a device anymore – it could turn into a service running purely as code out in the telephone company's network.

And so you threw out your answering machine and subscribed to a service. The physical device vaporized into the "cloud" of the network.

Your company's data center, when you boil it down to its essence, is just a much more elaborate version of that telephone answering machine. It, too, can be encapsulated into software and transferred into a network cloud. And if you follow the logic of virtualization and consolidation out to its conclusion, that's where you end up. At some point, the consolidation of an individual company's data centers, servers, and other hardware will hit the point of diminishing returns. You won't be able to squeeze any more capital, labor, or operating expenses out of the process. The next step will require that consolidation takes place not just *within* companies but *across* companies. The private data center will follow the private answering machine into the cloud. Private infrastructure will turn into a shared utility.

I know what you're thinking: "Yeah, right." Well, think again. Think in particular about those IBM mainframes that are sucking up hundreds of physical servers and delivering their applications and data over the network. When computing and communication become fast and cheap enough—and who's going to bet against the continuation of that trend? – the network turns into Turing's "universal computing machine," efficiently running any code that you can throw at it. Or, as Google CEO Eric Schmidt once said, "When the network becomes as fast as the processor, the computer hollows out and spreads across the network."

Goodbye "World Wide Web." Hello "World Wide Computer."

"Once you virtualize the computing infrastructure, you can run any application, including a custom-coded one, on an external computing grid."

Sun Microsystems, which coined its prescient slogan *The network is the computer* 20 years too early, has announced plans to leapfrog the consolidation programs of its archrivals HP and IBM. By 2013, Sun's IT department plans to cut the square footage of its internal data centers by 50 percent. And then, says Brian Cinque, the company's data center architect, it intends to reduce that square footage to zero. As early as 2015, Sun hopes to be running all its applications from shared utility grids – from the cloud. The network will have become not just the computer but the data center.

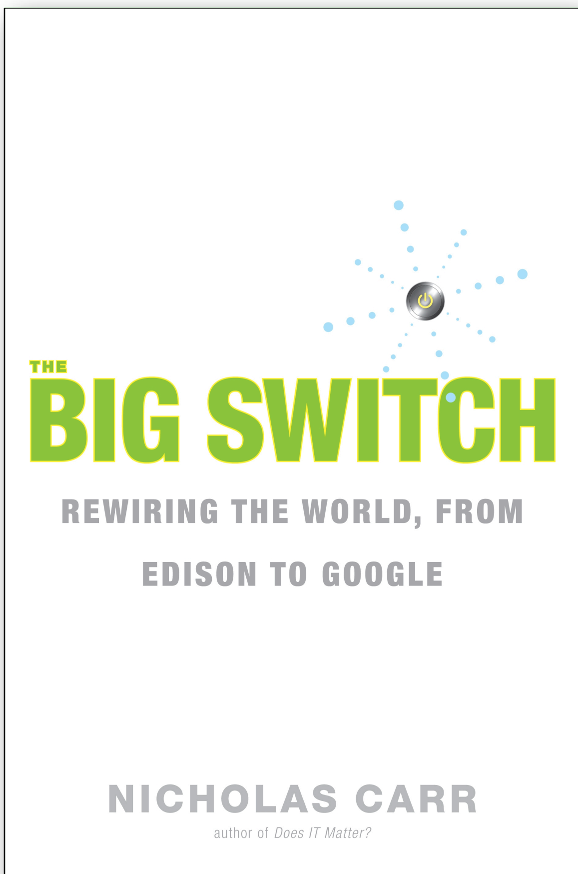
Today, the idea of "software-as-a-service" tends to be tied to fairly generic Web applications – Salesforce.com's customer relationship management service, say, or ADP's Employease service for managing HR functions. But there's no reason that the Web-delivery model needs to be limited to standardized applications. Once you virtualize the computing infrastructure, you can run any application, including a custom-coded one, on an external computing grid. Many small companies are already doing that through the set of utility storage and computing services offered by Amazon Web Services. Other pioneers, like virtualization leader VMware and the startup 3Tera, are developing new tools that make it easy to configure apps and infrastructure to run in the cloud. As those tools advance, and as companies like IBM, Google, and Microsoft plow billions more dollars into the buildout of the computing grid, the shift from local computing to cloud computing will only broaden and accelerate.

This revolution, it's important to emphasize, is not only about saving money by sharing infrastructure. It is also about promoting collaboration by sharing data and software. Up to now, corporate IT systems have been built on the principle of isolation – private hardware, private software, private data stores. But the isolation principle has always been in conflict with the nature of business itself, which is all about shared processes, shared information, and shared ideas. One of the biggest frustrations, and headaches, about IT has been

Are you ready for

THE BIG SWITCH?

The author of **Does IT Matter?** returns with an eye-opening look at the new computer revolution and its consequences for business and society.



"Future Shock for the Web-apps era ... Compulsively readable." -Fast Company

"Destined to influence CEOs and the boards and investors that support them as companies grapple with the constant change of the digital age." -The Wall Street Journal

"Quick, clear read on an important theme." -BusinessWeek

"Persuasive, well-researched, authoritative and convincing ... This is an exceedingly good book." -Techworld

"Mixes history, reportage and a dash of futurism to create a persuasive vision that CIOs may find unsettling—but cannot afford to ignore." -CIO Insight

"The first serious examination of 'Web 2.0' in book form." -The Register

Nicholas Carr's **The Big Switch: Rewiring the World, from Edison to Google** is available now at Amazon.com, Barnes & Noble, and your local bookstore.

For more information, visit **bigswitchbook.com** the never-ending battle to adapt isolated systems to support collaborative work.

The new utility model of IT, which is built from the start on the principle of sharing, helps resolve that tension. If you don't believe it, just look at the millions of people who, at home or at school, are happily using so-called Web 2.0 services to share data, work together on projects, and form online communities. You may love or loathe Facebook, but its social network shows how easy collaboration becomes when software and databases are shared rather than operated in isolation.

IT in 2018

Of course, what's easy for individuals, or even small businesses, is not necessarily easy for big companies. Corporations have loads of capital invested in their internal IT departments and assets. They're not going to rush to throw those investments out the window. And, of course, they have to worry about reliability, security, redundancy, and all the other hobgoblins of corporate computing. The transition from private to shared IT is not, therefore, going to happen overnight. It will proceed slowly and rationally, as the capacity of the cloud expands and as companies move to take advantage of it. For the foreseeable future, companies will rely on hybrid IT operations, blending privately maintained systems with utility services, while at the same time reengineering their internal IT infrastructure to operate more like a virtualized, multi-tenant utility system. What the grid allows is for machinery and functions to be organized in whatever way ensures the greatest efficiency and flexibility for the user.

> "Already today, according to one analyst, Google's parallel-processing grid can carry out a computing job for one-tenth the cost of the systems in a big corporate data center."

Few IT managers have the luxury of looking ahead more than a year or two. They face the distractions of constant operational challenges, relentless annual budgeting cycles, and pressures to hit quarterly targets. But at a moment when so much in IT is in flux – when the future is up for grabs in a way we haven't seen since the arrival of the personal computer a quarter century ago – looking ahead at the likely course of the changes over the next 10 years becomes not just wise but essential. The worst course, at a moment of transformation, is to blindly make big investments in the status quo.

So let's unwrap the crystal ball, give it a squirt of Windex, and take a peek at what the IT landscape may look like in 2018.

On the supply side, we've seen massive investments in the computing grid, with traditional IT suppliers like IBM and Microsoft, upstarts like Google and Amazon, and a variety of other players, from big telcos like Deutsche Telekom to governmental agencies, plowing capital into data centers and the network. Competition for control of the grid is intense, at the global, regional, and national levels, and governments have begun imposing a patchwork of regulations on the new utility, reflecting national concerns and traditions. In 2018, the Net is no longer "neutral." After much public debate, commercial traffic is routinely given priority over personal traffic to ensure the effective and reliable operation of software-based business processes.

A half dozen global vendors, along with some smaller regional players, have emerged as the operators of the computing grids used by most businesses. Over their interconnected networks a vast array of software services flow, in the form of both turnkey applications for routine processes and single-function modules that can be easily combined, or mashed up, into more custom programs. In addition, companies continue to run specialized, bespoke apps, sometimes in their own data centers but more often in the cloud. Although data flows are far from seamless, standards have been adopted that allow data and service modules to be integrated in a much easier way than is possible today. The advance of encryption and other technologies has improved data security dramatically, though the Net remains subject to disruptive criminal and even political attacks.

> Just as the last century's electric utilities spurred the development of thousands of new consumer appliances and services, so the new computing utilities will shake up many markets and open myriad opportunities for innovation.

As more and more hardware and software have shifted out of companies and into the cloud, the IT department has gone through a fundamental transformation (and has likely taken on a new name). The legions of workers once needed to keep the internal machinery running and updated have been replaced by a small squad of architects that use simple management programs, delivered through a browser window, to assemble virtualized IT components into corporate systems, modify them as needed, and monitor use and billing. Other information professionals work more like internal consultants, coordinating information and workflows within and across business units, finding new opportunities for deploying software services to achieve higher levels of automation, and helping product developers incorporate network-supplied data and applications into products and services for delivery to customers.

For employees, the "work desktop" has been entirely virtualized, and is served up, like an application, in a secure window on whatever computing device they happen to be using. (Companies rarely buy PCs or other computing devices for employees anymore, recognizing that the devices serve both personal and business uses, though they sometimes subsidize individual purchases.) Connectivity is ubiquitous and seamless – your data and applications are always at your fingertips, if not fed directly into the circuits of your brain – though many applications use sophisticated caching and backup techniques to take advantage of local data stores and processors. Although everything is controlled centrally; not everything runs centrally.

Nirvana? Well, imagine what it would feel like to live your entire life inside your BlackBerry. If that sounds like nirvana to you, then it will be nirvana. If not, then it will be something else.

But that's another story.

Resistance Is Futile

Crystal balls are notoriously unreliable, and the one we've been looking into is no exception. Ten years from now, utility computing may be more advanced than I've suggested or, more likely, it may be less advanced. But whatever the particular details of the IT world in 2018, it would be foolhardy to bet against the grid. Already today, according to one analyst, Google's parallel-processing grid can carry out a computing job for one-tenth the cost of the systems in a big corporate data center. As the capacity and power of the grid grow, as it becomes the greatest manifestation of Turing's universal computing machine, more functions will flow onto it. As we saw with the shift of factories from private generators to the electric grid in the early years of the last century, companies inevitably move to the cheapest method of supply. It becomes a competitive necessity.

No doubt, utility computing and Web apps pose big challenges to corporate IT departments and workers. They overturn long-held assumptions about IT, and they threaten some traditional jobs. But it's important to see that for the companies that rely on computer systems, this shift is all to the good. After all, it means, at the very least, a proliferation of new choices for how IT needs can be fulfilled. And it means, in the long run, the freeing up of considerable amounts of capital that have long been sunk into privately owned and maintained computing machinery. And in the end it gives workers greater power and flexibility to gather, analyze, manipulate, and share information – which is, of course, why businesses use computers in the first place.

As for computer professionals, the coming of the World Wide Computer means a realignment of the IT workforce, with some jobs disappearing, some shifting from users to suppliers, and others becoming more prominent. On the supplier side, we'll likely see booming demand for the skills required to design and run reliable, large-scale computing plants. Expertise in parallel processing, virtualization, artificial intelligence, energy management and cooling, encryption, high-speed networking, and related fields will be coveted and rewarded. Much software will also need to be written or rewritten to run efficiently on the new infrastructure. In a clear sign of the new labor requirements, Google and IBM have teamed up to spearhead a major education initiative aimed at training university students to write programs for massively parallel systems.

On the user side, as the transition to the utility model accelerates in the years to come, we'll likely see a slow but steady decline in jobs related to building and maintaining in-house computer systems, while skills in information management and process design and automation will remain highly valued. We may also see the rise of a new kind of IT worker – a systems broker who serves as the interface between the utility grid and internal functions, crafting a flexible portfolio of IT services to meet business needs. But the most exciting IT opportunities probably won't lie within traditional IT departments. Rather, they'll involve figuring out how innovative new products and services can be designed to run on top of the grid, incorporating data and software running out on the Net.

Just as the last century's electric utilities spurred the development of thousands of new consumer appliances and services, so the new computing utilities will shake up many markets and open myriad opportunities for innovation. Harnessing the power of the computing grid may be the great enterprise of the twenty-first century. ■

About the author: Nicholas Carr's new book is The Big Switch: Rewiring the World, from Edison to Google. *He is also the author of* Does IT Matter? Information Technology and the Corrosion of Competitive Advantage. *The former executive editor of the* Harvard Business Review, *he writes regularly for the* Financial Times, The Guardian, *and* Strategy & Business *as well as his blog* Rough Type. *More information can be found at nicholasgcarr.com.*