# Remotely Hosted Services and "Cloud Computing"

Dave Cliff

## About the author

Dave Cliff is a Professor of Computer Science at the University of Bristol. He has previously held academic faculty jobs at the Universities of Sussex and Southampton in the UK, and at the MIT Artificial Intelligence Lab in the USA. He also spent seven years working in industry at Hewlett-Packard's European Research Labs in Bristol, and at Deutsche Bank's London Foreign-Exchange Complex Risk Group. He is a founder and co-owner of Syritta Algorithmics Ltd, a technology development and consultancy company. In October 2005 Cliff was appointed Director of a £14m 250 person-year UK research consortium, addressing issues in the science and engineering of Large-Scale Complex IT Systems (see www.lscits.org). He is author or co-author on over 70 academic publications, inventor or co-inventor on 15 patents; and he has undertaken advisory and consultancy work for a number of major companies and for the UK Government. He's given more than 150 invited keynote lectures and seminars; and he and his work have frequently been featured both in the press and on TV and radio.

Cliff has worked as a futures consultant for the UK Government's *Foresight* Team on several of their projects since 2003. In 2008-09 he led the Science & Technology sub-group on the FutureLab *Beyond Current Horizons* project funded by the UK Government Department of Children Schools and Families, and Department of Innovation Universities and Skills, co-authoring a report on major socio-technical trends likely to influence UK education practice over the next twenty years.[1] He also serves as a member of Becta's Research Advisory Group.

[1] D. Cliff, C. O'Malley, & J. Taylor (2008) *Future Issues in Socio-Technical Change for UK Education.* Futurelab briefing paper for DCSF/DIUS *Beyond Current Horizons* project. http://www.beyondcurrenthorizons.org.uk/wp-content/uploads/socio_technical_change_revised_disclaimerv2.pdf.

## Introduction

This report discusses remotely hosted computing services, their recent explosive growth, and how they may affect the educational landscape in the next five years or so. To get the discussion going, let's start with a short story.

Imagine that you and I arrange to meet at a café. You get there on time, but I am running late and so you switch on your laptop computer to do some work while you're waiting for me to turn up. When I eventually arrive, noticing your laptop, I ask if I can borrow it for a few minutes to check something on the web – my Facebook account, say – because I need to send a message to someone. I use your computer for seven minutes and twelve seconds, and then I return it to you. We chat for a while over coffee, and then I pay the café bill. As we're saying goodbye, I reach into my pocket, pull out some loose change, and ask how much I owe you for the use of your laptop. What would you say?

If you're angry with me for turning up late and then immediately logging into Facebook, maybe you'd be tempted to punish me by asking for five or ten pounds. Or maybe not. More likely, you'd say something like "don't be silly, it's nothing". But is it really *nothing,* as in actually zero? Of course not.

At the very least, I've used up some electrical charge in the laptop battery that will need to be replaced by you when you get home. That has to cost something. And my typing has added some wear and tear to your laptop, which will contribute (albeit in a very small way) to its eventual demise, at which point you'll need to replace it. Let's say the laptop cost £1000, and lasts for three years, after which its value is zero. Let's also say that over its lifetime you spend £95 on electricity to keep its battery charged. So that's £1095 over 3 years, and each of those years is 365 days, giving your laptop a total lifetime of 1095 days, so the actual cost of your laptop to you is (conveniently) exactly £1 per day. And I used that laptop for 7.2 minutes which is (conveniently, again) exactly 0.5% of one day. So one way of calculating the cost of my session on your laptop is that it cost you half a penny. There's no longer a halfpenny coin in the UK, so I can't pay you from my loose change, but you could keep a note of my debt and the next time I borrow your computer for 7.2 minutes you could reasonably ask me for a penny to compensate your costs. Or you could charge me a penny for the 7.2 minutes, in which case you're making a profit of 100%. Now, 100% is really a very nice profit-margin. This sets you thinking.

Noticing the opportunity to make some money, you go home, connect your laptop to the internet, and advertise it as a computer for rent. Using standard technologies, other people (your customers) can log in to your computer and have it do work for them – your customers have their own computers, with keyboards and screens, so they don't need access to your laptop's keyboard or screen. A customer can type something on her keyboard, in her home or office, and the characters can be sent across the internet to your laptop's processor chip and memory and disk drive. When your laptop goes to display some of the customer's data, that data can be sent across the internet to the customer's screen. So the customer has *remote access* to the primary computing resources of your distant laptop: the central processor unit (CPU); the memory; and the hard disk – and can interact with them via her own computer's keyboard and screen for input and output. Your laptop is acting as a *server* to the customer's *client* computer. So far, so good. If you can attract enough customers to rent out every second of your laptop's life, at a rate of 1p per 7.2 minutes (or, equivalently, 8.3p per hour), you will have spent £1079 but will have generated income of £2158; that's a profit of £1079; the 100% margin that was predicted. Things are looking up.

But there's a catch. Let's say that your laptop is running Windows 7 as its operating system, with Microsoft Office running on top of that. What if some of your users want to run Lotus Notes on Windows XP instead? What if yet other customers want Linux running Open Office instead? Your business plan doesn't look so smart if you have to provide a different server loaded with different software for every possible combination of operating system and application software that your customers might demand. Luckily, a solution is at hand: you load your laptop server with some *virtualisation software.* This allows your one physical server-computer to run a program that makes it appear as more than one independently simulated but fully-functioning computers or 'virtual machines' (VMs). Each VM is, from the remote user's perspective, a fully functioning computer– and each of the VMs on any one server can be running different operating systems, supporting different application programs for different users. If a user crashes a VM, your one physical server can just start simulating another one (the other VMs that the server is running are independent, and so are unaffected by the crash/reboot). You've only got one server, but with virtualisation you can now make it pretend to be more than one computer, and each of those pretend computers (the VMs) can be independently tailored to a user's need. You're back in business.

Emboldened by the prospect of earning 100% profit, you borrow a large amount of money, quit your job, and prepare to buy a very large number of computers to make some serious money with: let's say you raise enough funds to buy 10,000 computers. Now you certainly don't need 10,000 keyboards and 10,000 screens, so it would be a waste of money to buy 10,000 laptops. But the attractive thing about laptops is that they generally take up much less space than standard desktop PCs. What you really need for your new business is a laptop computer with no screen and no keyboard: just the computer circuit board with its disk and memory and processor, in a conveniently thin plastic or metal case. That would be quite thin – roughly half the thickness of a standard laptop. Luckily, computer manufacturers make exactly this type of computer. Because they're thin, they're known as *blade-servers*, or just 'blades'. So, you buy 10,000 blades and arrange them on shelves running floor-to-ceiling on all the walls of your garden shed (it's a very big shed), connect them to the internet, and advertise them for hire. Now the chances are that when you buy 10,000 blades you will get a handsome discount, and remember also that you're not having to pay for any screens or keyboards either, so let's say those savings combine to mean that your actual costs are 5p per hour per server, but in pursuit of a 100% margin you charge your customers 10p per hour.

What's in it for the customers? There's two good ways of answering that. The first answer is this: if you offer to rent out remote access to your computers, and you take on the responsibility of keeping your software up-to-date, installing the latest bug-fixes and updates and service-packs and so on as they become available, then the customer no longer has to worry about those kinds of tasks, and indeed doesn't even have to worry about buying or installing software; what you're really offering to the customer is metered by-the-hour access to the software you have bought and installed, or '*software as a service'*, and because the service is provided by computers that are some way distant from the users' personal computers (or other *access devices*, such as mobile phones), your service is, from the users' perspective, *remotely hosted*. If there are a bunch of customers that each want occasional use of some expensive software product X, but who are each unable or unwilling to splash out the money to buy their own copy of X, then they can gain access to your remotely hosted copy at 10p or 11p per hour (the extra penny on the rent is to cover the cost of buying the software).

The second answer is this: if you have lots of computers, you can offer customers access to tens or hundreds or thousands of them for just as long as the customer needs access to that many computers. For example, say that I am one of your customers, and that I'm interested in computer animation. You have some computer-animation software installed on your computers and I could rent it at 11p per hour. I want to make a 10-minute movie, 600 seconds long, using 25 frames per second. That's a total of 15000 frames to be rendered. If rendering each one takes 10 seconds, then generating the entire movie will take a little over 41.5 hours on one computer, and at 11p per hour it'll cost me £4.57. Trouble is, 41.5 hours is a long time to wait, and I'm impatient. If instead I pay to hire 5,000 of your computers, the entire 15,000-frame movie will be created in just 30 seconds. And the bill will still be only £4.57 because the number of computer-hours I've used has not altered. What has changed is how those computer-hours are distributed over your collection of blade-servers (rather than tie up one blade for many hours, I'm instead tying up many blades for half a minute). Now it's plausible that I may be able to buy one computer and one copy of the animation software for £1000 say, but I'd have spent £995 more than if I had used your service; and it's much less plausible that I would go out and buy 5,000 computers and 5,000 copies of the software just to make my movie – that would only make sense if I was a computer-animated-movie-production studio like Pixar (makers of *Toy Story*, etc). Companies like Pixar do indeed maintain their own huge banks of computers for making their animations on (Pixar refer to theirs as their "render-farm"), and historically the huge costs of setting up a render-farm have acted as a *barrier to entry*: I can't become a movie-producer because I can't afford a render-farm. But your service changes all that: you've lowered the barrier to entry. Now I don't need tens of thousands of pounds to make a movie, I just need five quid.

Let's end the story here.

Just how realistic is that story? As you've probably guessed, it glosses over a lot of important details. If you put 10,000 blade-servers in your shed and switch them all on, they will generate an awful lot of heat: managing that heat is going to be a major problem: you will need to buy air-conditioning units and they will soak up a lot of electricity too. You can expect to spend roughly as much on electricity for the air-conditioners as you do on electricity for the servers. In reality you also have to pay for the shed that the blades are occupying (properly referred to as a *data-centre*), which in practice would be warehouse-sized, as large as an edge-of-town supermarket, perhaps larger. And you'd have to pay for people to look after the building and the servers inside it, 24 hours a day, seven days a week, every day of the year. Also, the mathematics of what you charge the customers is not nearly so straightforward either – setting prices gets very complicated as soon as you have to take into account the facts that some blades will be idle some of the time, that others will fail occasionally, that some customers will want to make advance bookings, that some customers will quit their usage early, that yet other customers will overrun their bookings and urgently need to rent blades 'on the spot', and so on.

Nevertheless, this story has served to illustrate some of the economics behind a major shift in the commercial provisioning of computer and software resources, one that is happening right now and is being driven by some of the biggest players in the IT industry, along with some surprise new entrants. That shift is the shift to remotely hosted services, where computer software (and computer hardware too) can be remotely accessed by users, often with those users paying small hourly fees, or perhaps not paying anything at all. Where they do pay by the hour, computing becomes one more metered utility, like electricity or water or gas, which users can expect to pay for via monthly bills. The service providers can exploit economies of scale to build larger and larger data-centres to accommodate more and more blade-servers; as they do so, the potential cost per hour to the user falls lower and lower; and significant barriers to entry are lowered in a variety of markets that are relevant to education. In this way, the move to remotely hosted services offers new opportunities in education, opening up the possibility of novel applications and offering significant potential benefits, but it is nevertheless not without unresolved issues and attendant risks.

Before we move on, it's worth pointing out that, within the IT industry, few people use the multi-syllable phrase 'remotely hosted services' in casual conversation. Because the precise geographical location of the remotely hosted services is largely irrelevant to many[2] users, they may as well be located in the clouds – an observation that has given rise to the popular phrase ‘cloud computing’ as a descriptor for remotely hosted computing services. If you store your data at a remote data centre, or use software provided as a remotely hosted service from such a data centre, you might reasonably say that your data is ‘in the cloud’ or that the software you use is ‘a cloud service’.[3]

## The current shift to remotely hosted services

In the early 2000s major IT companies such as Hewlett-Packard, IBM, and Sun Microsystems, each announced their visions of remotely hosted computing services. The different companies used different terminology: HP referred to ‘utility computing’; IBM spoke of ‘computing on demand’, and Sun called theirs ‘N-1’ (referring to some arbitrary number *N* of computers that could be accessed and used as if they were one) but the underlying aims and visions were largely identical. That these companies were proactive in the move to remotely hosted services is probably no surprise as they were already focused on the sale of computer hardware, software, and related services.

Somewhat less expected (and not well anticipated by some of those long-established computer companies) is the fact that the two companies now widely regarded as the leaders in remotely hosted computing are not long-established IT vendors but are instead an online advertising company, and a bookshop: Google, and Amazon. For both these companies, their core businesses required them to build ever-larger data-centres, filled with racks of blade-servers and the associated networking and cooling hardware.

For Google, the servers in their data-centres process users’ internet searches, and insert adverts into the search-result web-pages that are then served back to the users: from the users’ perspective, Google is a global remotely hosted search service, free at the point of use. Google’s customers are the companies and individuals who pay for the placement of their adverts, providing Google with enough revenue to operate this global service and also generate enviable profits.

For Amazon, the servers in their data-centres process customers’ requests for product information and purchases. Customer search requests trigger searches through the database representing the stock of Amazon and its affiliate suppliers, the stock data then needs to be integrated with customer reviews and automatically-

[2] The geographic location of remotely hosted data and services may be irrelevant to many casual users of those services, but for many institutions and corporations it can be of critical importance: a point returned to later in this report, when we discuss jurisdiction.
[3] When this report was commissioned, Becta expressed a clear preference for using “remotely hosted services” rather than “cloud computing”, but the latter phrase is now commonplace in the IT industry

generated recommendations for alternative purchases; any actual customer orders need then to go through the payment fulfilment process.

For both Google and Amazon, the need to serve very large numbers of simultaneous users with near-instant results, 24-7, required global networks of large data-centres to support their core business. Having already built these data-centres, they then moved to renting out spare capacity in them, but the two companies approach this in different ways.

Amazon launched *Amazon Web Services* (AWS)[4] in 2002. AWS is a collection of component services that can be combined by customers to build their own remotely hosted services. A core element of AWS is the Amazon *Elastic Compute Cloud* (EC2) which offers users access to virtual machines running on servers in Amazon's data-centres – these virtual machines are known as *Amazon Machine Instances* (AMIs). AMIs can be configured as Windows or Unix/Linux, and come in various types:

- *High-memory AMIs*, for applications with high data-throughput, such as databases;
- *High-CPU AMIs*, for processor-intensive applications;
- *Standard AMIs*, which are generally the cheapest option.

Each type of AMI is available in a range of sizes: for example, standard AMIs are available as *small* (1.7Gb of memory, one virtual CPU, for roughly 8p per hour – Amazon's actual prices are quoted in US dollars), *large* (7.5Gb of memory, two virtual CPUs, roughly 23p per hour), and *extra large* (15Gb of memory, eight virtual CPUs, roughly 63p). While EC2 AMIs offer remote equivalents of a computer's CPU and memory, the AWS equivalent of hard-disk storage is offered as Amazon's *Simple Storage Service* (S3). EC2 and S3 form the backbone of most AWS applications, and Amazon offer a number of additional services that can be combined with EC2 and S3 to complete applications, such as database servers, payment fulfilment systems, and so on, the full list of which has been growing steadily in recent years, some of which are only available as semi-experimental ('beta') versions.

Amazon Web Services have been used in a number of notable applications, often cited as case-study illustrations of the transformational effect that the shift to remotely hosted services can offer. We'll look at two of those briefly here: one is in newspaper archiving; another is in creating a new remotely hosted web service.

---

[4] See http://aws.amazon.com.

Our first AWS case-study took place in 2007. The *New York Times* (NYT) newspaper was in the process of making digital copies of all 11 million of its past articles from the period 1851-1980. The digital copies are made available to customers as PDF-format documents, but the digital scans of the archived hard-copy newspapers were in a different format: TIFF. To make one article available as PDF, it was frequently necessary to process multiple TIFF images (e.g. if an article ran over a number of pages, possibly non-consecutively numbered) and stitch them together to form a single PDF file. The conversion from TIFF to PDF had been handled dynamically: a computer processed the TIFF files to create the PDF afresh, every time someone requested a copy of that article. As the number of article requests increased, the amount of computer power needed to do this processing was clearly set to rise. Derek Gottfrid, a NYT staff-member, decided to use AWS to pre-generate all 11 million articles, and store them on Amazon's servers so that each was instantly available via static recall anytime a user requested it: this required 4Tb (4,000Gb) of S3 storage. Describing the process in a NYT article,[5] Gottfrid writes:

*'…if I used only four machines, it could take some time to generate all 11 million article PDFs. But thanks to the swell people at Amazon, I got access to a few more machines and churned through all 11 million articles in just under 24 hours using 100 EC2 instances, and generated another 1.5TB of data to store in S3. (In fact, it works so well that we ran it twice, since after we were done we noticed an error in the PDFs.)'*

Gottfrid's article does not go into details of the cost, but using 100 EC2 instances for 'just under 24 hours' is going to incur no more than 2400 times the hourly rate. Say, for the sake of argument, that the cost per instance-hour was 31.3p (that's the average of 8p, 23p, and 63p; the prices quoted above for the three sizes of standard AMIs). Then the cost of 100 AMIs each for 24 hours to collectively process the 4Tb of data that 11 million NYT articles add up to will have been around £750. At today's prices, you can't buy a bottom-of-the-range Apple MacBook for £750.

Our second AWS case-study involves a small American company called Animoto. Animoto run a web-based business[6] that produces videos for its customers. A customer logs into Animoto's web-site, and uploads a set of digital photos, plus a digital audio file of some music. Animoto's service analyses the music, and then creates a 'music video' cutting between the different photos in a sequence that is synchronised with the music: if you upload some relaxing music you can expect a video with slow, smooth transitions between the successive images, whereas if you upload some pumping disco/dance music, with its characteristic repetitive beats, you can expect more abrupt, jumpy transitions between images. The algorithms that drive Animoto's system involve some random choices, so if you don't like the first video it produces, you can request that the algorithm is re-run, creating a different

[5] http://open.blogs.nytimes.com/2007/11/01/self-service-prorated-super-computing-fun/
[6] See http://www.animoto.com.

video, and you can keep on re-requesting the service until it produces a video that you are happy with. Although the user uploads only two-dimensional images to the service, the videos that it produces may have three-dimensional effects in the transition from one to the next: for example the current photo might be rendered as rotating in three-dimensional space, flipping over to reveal that on its rear-side is pasted the next photo in the sequence. Animoto used AWS for their remotely hosted provisioning of compute-power and data-storage. That is, when a user logs into the Animoto website and uploads photos and music, that upload goes straight through the Animoto site and onto AWS servers, but to keep things simple this back-end arrangement is hidden from the user.

Jeff Bezos, the founder and CEO of Amazon, gave a public lecture[7] on April 19th, 2008, where he spoke about what had happened to Animoto in the preceding few days. Animoto had been operating for some time and was routinely renting around 50 AMIs from AWS. Then, over the course of a weekend, the number of people using Animoto exploded as news of the (free) service 'went viral' via Animoto's deployment of an 'app' on Facebook. That is, some Animoto users told their Facebook friends about Animoto, those friends told their friends, they in turn told their friends, and so on. Animoto was new, cool, and free. Over the course of just one weekend, Animoto signed up 750,000 new users, and at the peak sign-up rate they took on 25,000 users in one hour.[8] Very many of these new users immediately used the free service, uploading megabytes of images and audio to Animoto, and receiving in return many more megabytes of digital video. As Jeff Bezos relates, describing a graph of the number of AMIs that Animoto were simultaneously renting from AWS, in that one weekend:

'*You can see they've gone from 50 instances of EC2 usage up to 3,500 instances of EC2 usage. It's completely impractical in your own data-centre over the course of three days to scale from 50 servers to 3,500 servers. Don't try this at home.*

*"And by the way, the other alternative might be to raise enough capital to deploy 3,500 servers. That's sort of equally insane because it's just way too big of a gamble. You don't know whether you're going to get this app to take off in that way and you shouldn't be deploying that kind of capital.'*

Bezos went on to note that in the details of the same graph there were a few large drops in the number of AMIs that Animoto were hiring over the course of the three days: these corresponded to the middle of the night in the US: when the number of Animoto users who were awake had fallen off sharply. In those periods, Animoto no longer needed as many AMIs as at peak time, and so AWS automatically scaled

[7] See http://blog.animoto.com/2008/04/21/amazon-ceo-jeff-bezos-on-animoto/ for a transcript of the section of Bezos' lecture where he discusses Animoto's recent history; for a video of that section of the lecture, see http://www.omnisio.com/v/9ceYTUGdjh9/jeff-bezos-on-animoto.

[8] http://www.nytimes.com/2008/05/25/business/worldbusiness/25iht-proto.3.13191841.html?scp=4&sq=animoto&st=cse

their number of AMI allocations downward – this is the ‘elasticity’ in Amazon’s *Elastic* Compute Cloud (EC2): only as many AMIs as are actually needed by the AWS customer’s application are allocated to it, and the allocation of AMIs varies smoothly as the customer’s computing needs rise and fall.

These two AWS case-studies demonstrate the kind of applications that Amazon’s rental machine-instances can be put to: applications that would be ruinously expensive, or very risky, if attempted via the conventional route of buying and maintaining your own IT systems. Amazon offers users access to remote servers, the users choose the size of the virtual machines (the AMIs) and the operating system, and then run their own programs remotely on those AMIs. The approach that Google has taken to remote hosting is different in a number of respects, and we’ll turn to them next.

Google’s remote-hosted services are referred to collectively as *Google Apps*, offering a combination of email, calendar, instant messaging, internet telephony, office productivity (word processor, spreadsheet, presentation, form-based database), and wiki/web-page creation. As such, Google Apps can reasonably be seen as a remotely hosted service that offers the same functionality as well-established commercial software titles such as Microsoft Office, Lotus Notes, Microsoft Exchange, and Microsoft FrontPage. Traditionally, such software products are sold in the form of CD-ROM/DVD disks, packaged in boxes or wallets and stocked on the shelves of computer stores or shipped direct to users via mail order, where each copy of the software is installed and run on the user’s computer. Instead, Google Apps all run remotely, in Google’s data-centres, and are accessed simply by logging in to the service via a web-browser. This is not the only difference between Google Apps and the comparable commercial products: Google Apps are available for free. Enterprise users requiring additional levels of service in Google Apps, such as extra storage, can pay a fee to Google, but the basic services are all provided at zero cost to routine users, and free education-specific versions of these services are also available.[9] Currently services that have a high level of standardisation, such as email, are most suited to cloud computing models and in education they are free. Free cloud email services aimed at education have matured over the last couple of years and providers have worked with people in education to overcome issues such as data protection regulations, single sign-on, and integration with directory management tools.

Google also offers a service called *Google App Engine*, where software developers can write their own extensions to Google Apps; that is, new programs that run on the same underlying software and hardware foundations as used by the other Google Apps application programs. The provision of the existing software and hardware foundations (or ‘platform’) means that developers have less work to do (in

[9] See http://www.google.com/a/help/intl/en/edu/index.html

comparison to working on Amazon web services) to create a new working Google app, but at the same time they are limited to using only that platform – which may be an unwelcome constraint in some instances.

At this stage, it's worth introducing three acronyms that are commonly used to distinguish between different types of remotely hosted services: SaaS, PaaS, and IaaS. SaaS refers to *Software as a Service.* This is where the remote-hosting provider offers the functionality of complete software packages: Google Apps is an obvious example of SaaS. PaaS refers to *Platform as a Service*, and is well exemplified by Google App Engine: there the remote-hosting service offers a fixed software and hardware foundation (the 'platform') on which to build new applications (new SaaS) but the user has little or no choice in the nature of the platform, and can't alter it. Finally, IaaS refers to *Infrastructure as a Service*, of which Amazon Web Services (AWS) is the leading example – here the users can specify the memory size, processing power, and operating system of each virtual machine (VM) that they hire, and the VMs are then rented out as 'bare' machines, with the customer responsible for adding all further software to the VM. In fact, users of AWS can elect to take the PaaS route, as the range of AMI options is actually much wider than was indicated above: it is possible, for instance, to rent AMIs that come pre-loaded with **L**inux, **A**pache (an open-source web-server), **M**ySQL (an open-source database server), and an open-source programming language such as **P**HP, Perl, or Python. This particular configuration of software components (or 'solution stack') into an application development platform is used extensively in a wide range of applications, and is so familiar that it is commonly referred to as the 'LAMP Stack'.

While Amazon and Google are widely regarded as leaders in the field, they are nevertheless being closely pursued by a number of other companies, some of which are household names and others less so. Microsoft came to the table relatively late,[10] but as a famously cash-rich company they have been able to catch up by investing very heavily in recent years, both in the construction of major data-centres and in the development of new software products tailored to the remotely hosted future, such as their *Azur*e cloud-computing software platform, and their *live@edu* SaaS offering that is aimed specifically students and educational practitioners. Apple Computer, whose market capitalisation recently surpassed that of Microsoft, thereby making Apple the world's most valuable technology stock, run a remote hosting service called *MobileMe* which allows their customers to store their email, contacts, and calendars remotely, and hence to automatically synchronise these data across multiple devices; and Apple's *AppStore*[11] sells application programs (apps) for their

[10] For example, on the 25th February 2008, almost a decade after HP, IBM, and Sun had each started the move toward remotely hosted computing, planning the construction of huge data-centres, Microsoft's CEO Steve Ballmer was interviewed in the *Financial Times* (p.27) under a headline "Microsoft predicts rise of the datacenter".

[11] Apple's Appstore is accessed over the internet via Apple's iTunes desktop application rather than directly from a web-browser.

popular iPod, iPhone, and iPad devices, and AppStore can be thought of as a remotely hosted sales and distribution services: to reach millions of customers an application programmer for Apple devices no longer needs to be concerned with arranging the manufacture, packaging, and distribution of CD-ROMs in cardboard boxes: once the application is on the AppStore, it is globally distributed, instantly.

Two companies that are less well known to the general public, but that are nevertheless widely regarded as members of the ‘big seven’ group[12] of cloud vendors, are Salesforce.com and VMware Inc. Salesforce.com are widely credited as the company that led the way in offering software as a remotely hosted service (SaaS), paid for by subscription; founded in 1999, the company’s current market capitalization is now more than $11bn. Subsequent to the success of their SaaS business, Salesforce launched a commercial PaaS offering, called force.com. VMware’s rise has been similarly meteoric (founded in 1998, current value $27bn); their strengths lie in providing the core software that enables remote hosting – they are the leaders in virtualisation software that allows one physical server-computer to appear as a number of independently fully-functioning virtual machines (VMs – hence the company’s name) to the users accessing that server remotely.

The shift to remotely hosted services is currently a major concern in the IT industry, and it is now the case that practically all major IT systems and software suppliers have a stated strategy and products for cloud computing. Hardware manufacturers such as Dell, HP, IBM, and Lenovo have all announced plans for new products aimed at customers who intend to construct their own data-centres, and/or products aimed at customers who intend to capitalise on the growth in remotely hosted services by replacing their desktop PCs with cheap ‘cloud terminals’. These are computers that have good screens and keyboards to ensure a strong user experience, but do not contain large amounts of memory or hard-disk space, nor super-powerful processors, and which are hence very cheap. In the remotely hosted future, there may well be no need for the machine on your desk (or on your lap, or in your palm) to have any more compute power than is necessary to run a decent web-browser: the necessary processing power and data-storage to do everything, and anything, you want will all be elsewhere, in the cloud.

For these reasons, many corporations are starting to invest in their own ‘private cloud’ data-centres that give remotely hosted functionality to the corporation, and which are constructed and managed in such a way that if excess capacity is needed at peak times, the corporation’s private system can smoothly integrate extra resources that are bought from “public cloud” providers such as Google or Amazon, for just as long as they are needed: an approach known as *cloudbursting*. Mindful of the attractive economies of scale that accrue when bigger data-centres are built, there is talk of some groups of enterprises or institutions clubbing together to

[12] See http://www.technologyreview.com/briefings/cloud.

construct private cloud facilities that are shared equitably among the set of co-owners or co-stakeholders, an approach referred to as *community clouds*. In January 2010 the UK government announced ambitious plans for the construction of what promises to be one of the largest community clouds on the planet: "*G-Cloud*", a network of twelve secure data-centres devoted to serving the entire IT needs of the UK government and public sector, costing £3bn to build but thereafter projected to save £3.2bn per year from the government's annual £16bn expenditure on IT.

The shift to remotely hosted services is not without its issues, challenges, and problems, some of which we briefly review in the next section, before moving on to discussion of the potential impact of the technology on UK education over the next five years.

## Issues and Challenges

The shift to remotely hosted services is well underway, but as with many new technologies, the path of progress is a bumpy one: every now and again, something goes wrong. Perhaps the most newsworthy problems are the major outages, i.e. when a problem in a data-centre, or in more than one data-centre operated by one service provider, deprives many users of the remotely hosted service(s). In the past three years, major providers of remotely hosted services such as Amazon, Apple, and Google have each suffered incidents in which they have offered severely reduced levels of service, or perhaps no service at all, to large number of their customers.[13]

Even if the remote service provider has fully functional data-centres, users of remotely hosted services are critically reliant on their internet connection. With a traditional desktop PC running its application software locally, from the executables stored on its hard-drive and resident in its memory, if the network connection is unplugged by a toddler or chewed through by a mouse or drilled into by a builder, you lose internet connectivity but you still have a workable autonomous computer on which you can run those applications. In the world of remotely hosted services, you don't. If you're using a 'cloud terminal' you have a screen, a keyboard, and very little else.

In the absence of failures of the data-centres that host the remote services, or failures of a user's network connection to those remote facilities, there are still open issues. Building warehouse-sized data-centres requires attention to details, such as power electronics and thermofluidics (the dynamics of hot and cold air circulation within the data-centre) that are not usually the concern of computer systems engineers. In 2009 Luiz André Barroso and Urs Hölzle, two engineers working for

[13] For further details of cloud service failures in 2008 and 2009, see the *MIT Technology Review* article at http://www.technologyreview.com/briefings/cloud.

Google, published a groundbreaking book with the title "*The Datacenter as a Computer: An Introduction to the Design of Warehouse-Scale Machines*",[14] where they argue that the correct design approach is not to treat the data-centre as a large collection of co-located servers; but rather to employ a holistic approach in which the entire data-centre is conceived of as a single, massive, warehouse-scale computer. Barroso and Hölzle's book reveals many aspects of Google's approach to the design and management of such large-scale data-centres, but there are still a number of unresolved research issues remaining.

In recent years a growing number of international conferences[15] have been organised for researchers and practitioners of remotely hosted computing services to share their findings. Two recent survey papers have been well received as concise summaries of the existing academic literature and the significant challenges that remain to be solved if remotely hosted services are to achieve more mainstream acceptance. The first of these is Ilango Sriram and Ali Khajeh-Hosseini's 'Research Agenda in Cloud Technologies',[16] which discusses the challenges facing developers of technologies for providing remotely hosted computing services; and the second is 'Research Challenges for Enterprise Cloud Computing'[17] by Ali Khajeh-Hosseini, Ian Sommerville, and Ilango Sriram, which takes a perspective more centred on the issues for organisations – be they corporate enterprises or non-profit institutions – looking to move to remotely hosted service provision.

Not all of the challenges are technical in nature. One aspect of remotely hosted computing services that is a major concern to many organisations is the security and integrity of the data: assurances are needed that a customer's data (or, indeed, their application programs) won't be leaked to other users and/or corrupted once they are entrusted to the remote-hosting service provider. Furthermore, the prospect of an organisation's data and program-code being in the cloud can strike fear in the hearts of certain organisations, unless those organisations can be guaranteed that the data will not be hosted (or securely backed up) at locations outside the organisation's preferred geographical region. This is primarily for reasons of jurisdiction: if a UK organisation's data is remotely hosted and finds its way onto servers physically housed in data-centres in the USA, then the British data becomes subject to American data-protection and data-interrogation laws, which may not be to the customer-organisation's liking – and may in some cases place the customer-organisation in breach of UK and EU laws or contracts. Most remote-hosting service providers now offer contractual assurances concerning geographic and jurisdictional limits on the data that they host. Providers such as Amazon, Google, and Microsoft

---

[14] Available to download from http://www.morganclaypool.com/doi/pdfplus/10.2200/S00193ED1V01Y200905CAC006.

[15] Notable conference series include LADIS, the Workshop on Large Scale Distributed Systems and Middleware (http://ladisworkshop.org/), first organised in 2007; and the IEEE International Conference on Cloud Computing (http://www.thecloudcomputing.org/2010/), running since 2008.

[16] Available to download from http://arxiv.org/pdf/1001.3259.

[17] Available to download from http://arxiv.org/pdf/1001.3257.

can all offer services contracts where the customer is assured that the data will be constrained to reside within specific geographical regions (e.g. solely within the EU). In addition to concerns over geography and jurisdiction, some potential customers of cloud computing have concerns over issues of auditability and the responsibility not only to ensure data security, but also to ensure secure disposal of data: how do you know for sure that any data you delete really is deleted? To counter such concerns, various service providers may give assurances about “scrub-down” (secure deletion) of virtual machines, “power-cycling” (switching off and then restarting) of physical servers between jobs, and military-grade destruction of servers at the end of their serviceable lifetimes. Nevertheless, some potential and actual customers of remote-hosted services remain nervous, because these modes of operation are so new that the contracts governing the services have not yet been tested in court, and in some cases new skills will be required to specify, evaluate, and manage the service contracts to ensure that appropriately robust service-level agreements (SLAs) are in place.

Another concern in this new and rapidly developing business environment is the prospect of *vendor lock-in*: if you commit to Company X for provision of remotely hosted services, you may find it very difficult or costly to then switch to a different vendor if, for example, Company X goes out of business or sharply increases its prices. Recognising this as a potential stumbling block, a large number of companies active in the provision of remote hosting services (or in the provision of component technologies for such services) have signed up to support the *Open Cloud Manifesto*,[18] a statement of principles for how cloud-computing and remote-hosting services should be organised and provided to enable easy switching between cloud providers, among other things. In the list of approximately 300 companies named on the supporters page[19] of the Open Cloud Manifesto website are major corporations such as IBM, SAP, and VMware, but several other companies mentioned previously in this report are not yet signed up.

Finally, other non-technical challenges that may need to be overcome include formal cost-benefit analysis to check of the reality of the perceived cost-savings. Such analyses should include evaluation of the costs of integrating the remotely hosted services with existing services already in use, in staff re-training, and in overcoming resistance to change.

[18] Available to download from http://www.opencloudmanifesto.org.
[19] See http://www.opencloudmanifesto.org/supporters.htm.

## Potential Applications and Benefits in a Learning Context

There are a number of benefits claimed for remotely hosted computing services that are as applicable to educational contexts as they are to other institutions and corporations. These include:

- IT provisioning ceases to be a capital expenditure item in the accounts, with the costs of purchasing, managing, supporting and replacing IT instead all being accounted for as operating expenses, in the form of whatever fees are due to the service provider. This should result in simplified management and lower costs.

- Money is used more efficiently, as organisations benefit either from free services, or pay only for those services that they actually use, with usage scaling up or down smoothly as required. The savings from economies of scale available to large-scale service providers can be shared with the providers' customers.

- Shared remote services allow customer organisations to benefit from easy access to reliably up-to-date standardised services, maintained by experts, and from access to new services that can be introduced quickly and smoothly. The risk of investing in unsuitable hardware or software is similarly reduced.

- Services are platform independent and can be accessed from most web-enabled devices. New services can be developed to run on cloud platforms, and can be distributed to large numbers of users/customers over wide geographic areas at very low cost.

Together, these benefits can help education establishments and organisations to reduce the time spent on the relative distractions of dealing with information technology, and instead concentrate on their primary purpose (and core competence), education.

In addition to these general benefits, there are two major ways in which the shift to remotely hosted services seems likely to specifically affect education and learning over the next five years: in the administration of educational establishments, and in the delivery of educational materials and experiences.

### Administration

Educational establishments such as schools, colleges, and universities all managed their administrative affairs using paper-based systems prior to the switch to IT-based administration that has taken place over the past two decades. It is now commonplace for all the primary records of an establishment to be held in digital form. Because historically each establishment ran their own local paper-based administration, the IT-enabled administrations are also run locally in each institution. But there is no reason for matters to stay that way.

In the past decade, in the private sector, many multi-national companies, and many multi-site national companies, have moved to centralised IT-based administrative services, where a central 'hub' server or data-centre provides remote hosting to the 'spokes' in the network – the satellite sites for whom the IT services are now remotely provided. In some cases, the hub might be at the corporation's head office; in others, it may be at a remote service centre in a country where labour is cheap. Nevertheless, running such centralised IT services is complicated, expensive, and in many cases is something of a costly distraction, in the sense that it is not part of the core competence of the business. In recent years, the providers of remotely hosted computing services have started to pursue such corporations as customers, offering to take on the hosting and administration duties, in exchange for a commercial fee. That long-established IT service companies such as EDS and IBM are making such offers is probably no surprise. Once again though, the new entrant making the news is Google, in the form of *Google Enterprise*. Google Enterprise has secured a number of high-profile corporate clients, including newspaper publisher Telegraph Media Group and construction company Taylor Woodrow, who have switched to using Google Apps for all their email and office productivity functions. But Google Enterprise are not only securing private-sector clients: Leeds Metropolitan University and the University of Westminster have both now similarly switched to using the remotely hosted services of Google Enterprise. For Google Enterprise customers such as these, there is no need to retain their own centralised IT servers and the staff to maintain them, freeing up space and reducing personnel expenditure.[20]

On this basis, it seems reasonable to expect that in the coming five years we may see more educational establishments switching to remotely hosted services for their administrative IT provision. In some cases this may be through their own volition; in others it may be because they have been required to do so, by their local education authority or perhaps by the government department responsible for education.

The switch to digital record-keeping offers the prospect of smoother hand-over of data as learners move from one institution to another, and in a remotely hosted record-keeping service the data-storage responsibility no longer lies locally with each institution but rather with the shared remote service provider.

Somewhat more fancifully, but certainly within the bounds of possibility, is the prospect of taking advantage of remotely hosted data-storage to create a digital

[20] The costs and benefits of moving to remotely hosted computing services can be explored at greater length via toolkits such as the one described by Ali Khajeh-Hosseini, David Greenwood, James Smith, and Ian Sommerville in "The Cloud Adoption Toolkit: Addressing the Challenges of Cloud Adoption in the Enterprise", available for download from: http://arxiv.org/abs/1003.3866.

record of vastly more than is currently the practice of a learner’s outputs and activities over the course of their education. In the past, for a school to store every school exercise book completed by a child, and every piece of artwork they produce, would have been physically impracticable and so was not done. But now, so long as the digitisation process could be reasonably handled, storing scanned images of every item of paper written upon by a child, every keystroke of theirs on a computer, and perhaps also digital audio and video recordings of every practice and every performance of music, drama, and sports, does not present a major challenge: a 1Tb (1,000Gb) desktop hard-disk drive retails in high-street PC stores for around £60, and it seems reasonable to conjecture that all the outputs produced by one child over the period from Reception to Year 12 could be stored in one terabyte of data. Assuming that economies of scale apply, the likely cost for storing all output data produced by a child over the years R-12 would be probably be much less than £50 per child. Assuming that this data is searchable, it may open up new methods for monitoring progress and identifying opportunities for tailoring the educational experience to the individual learner. Of course, ensuring the integrity and security of that data would be major priorities, and there is likely to be significant and perhaps heated debate over the ethics and privacy implications of such an approach.[21]

### Educational Materials and Experiences

Currently, locally-hosted educational software running on desktop PCs is limited by the processing power and storage capacity of those PCs. But, as the two AWS case-studies described above (*The New York Times* and *Animoto*) make clear, it is now possible to remotely access supercomputer-scale processing power and storage capacities but only pay small amounts for doing so. This opens up the possibility of offering new modes of provisioning educational content and experiences. Here, as illustration, are two speculations on possible mid-term future developments, each enabled by remotely hosted computing services. The first involves access to massive remotely hosted stores of digital data; the second involves new possibilities that are opened up by access to large numbers of remotely hosted computer processors.

***Every Book in Every Schoolbag****:* the libraries of educational establishments have historically been limited by their budgets for buying new stock, and by the amount of storage space they have for that stock – both on the library shelves, and in any back-office storage. As digital-media e-book reader devices, either specialised ones such as Amazon’s *Kindle*, or more general-purpose media devices such as Apple’s *iPad*, increase in prevalence, the possibility arises of totally bookless libraries. The library of the future could be nothing more than a room with pleasant seating and a large number of Kindle and/or iPad reader-devices, possibly with some larger wall-

[21] This is certainly not something that I am personally advocating as a positive future development. I am merely pointing out the possibility. It’s beyond the scope of this report to explore the positive and negative aspects of such an application, but that is an important debate that should not be avoided.

mounted or tabletop colour touch-screens for viewing large-format content such maps and atlases. Users of the library could browse through virtual bookcases, or perhaps view graphic visualisations of what publications are currently most read among their friends or classmates. In principle, every book that is available digitally anywhere in the country would be available simultaneously to every library user in the country. The primary obstacle to this happening is commercial rather than technological: while there are large online libraries of copyright-free material, it would be ruinously expensive to pay for a license for a full copy of every single commercial (in-copyright) e-book, for every single potential reader. But if the reader-devices could be programmed to monitor how many pages of the book have been accessed, and charged on a per-page-viewed basis, the economics could be workable. If the library's reader devices are capable not only of displaying books, but also of playing back digital audio and digital video, then the library service could also provide access to videos of tutorials and lectures made available on YouTube or via Apple's iTunes service: in June 2010, the UK's Open University reported that it was the first content-provider to have served 20 million downloads on iTunes. It did this via the successful dedicated educational area on iTunes, called *iTunesU*, where several world-leading universities (including top American universities Berkeley, Harvard, MIT, Stanford, and Yale) provide educational audio, video, and text content for free. For further discussion of these issues, see Charles Leadbetter's report for Counterpoint, "*Cloud Culture: the Future of Global Cultural Relations*".[22]

***Serious Games and 'Real Virtuality'.*** One class of applications consistently pushes at the boundaries of what is possible with a standalone desktop or laptop computer: games. Computer-game designers habitually write programs that consume all of the available processing power on a PC. As PCs become more powerful, the games designers design more power-hungry games to soak up the available power. Many games involve simulations of real-world scenarios (driving cars, flying aeroplanes, armed combat) visualised in first-person perspective, and in the past ten years a number of network-enabled games, known as massively multiplayer online role-playing games (MMORPGs) have become enormously successful, attracting millions of players from around the world who can meet and interact in the virtual, simulated worlds provided by the games. MMORPGs are remotely hosted computer services too: the player's computer is responsible for monitoring the player's actions and sending them to the central MMORPG servers, and the MMORPGs server sends back data that determines what is rendered on the player's screen (although a lot of the details of the graphics rendering will be handled locally, by the graphics circuitry in the player's computer). Currently the most successful MMORPG is *World of Warcraft* by Blizzard Entertainment, with more than 11 million subscribers worldwide, each paying US$13 per month or more, generating recurring revenues of more than $1.7bn per year. While one significant aspect in an MMORPG is the quality of the interaction with other players, another significant factor for many game-players is the

[22] http://www.counterpoint-online.org/cloud-culture-a-new-counterpoint-publication-and-ica-debate/.

quality of the simulation of a real or imagined world. In recent years, a number of academic and industrial researchers have started to explore the possibility of using computer games technology for educational and instructive purposes – and this approach has become known as ‘serious games’.[23] Remote-hosted MMORPG techniques could be used to join together students at geographically disparate sites in virtual worlds that have been constructed for pedagogic purposes. Technology is also advancing towards making virtual experiences immersively realistic for sight, sound and smell: an approach that Alan Chalmers and Eva Zanyi refer to as ‘real virtuality’.[24]

While the examples given here have been described with reference to conventional primary and secondary schools, colleges and universities, there is no reason why these examples could not also apply to individuals interested in lifelong and post-retirement learning, and to institutions that cater to the requirements of such individuals, such as the UK University of the Third Age. Serious games and ‘real virtuality’ may find just as much use in adult continuing professional development (CPD) contexts as in schools; and the argument that institutional libraries may become book-free could just as easily be applied to municipal public libraries, so long as appropriate per-page-view licensing and rights-management arrangements could be agreed.

## Conclusion

The switch to remotely hosted computing services is well underway. Although the phrase ‘cloud computing’ has certainly become, in some quarters, a bandwagon buzzword, the economics of remotely hosted services are compelling for many situations where local standalone IT services are currently deployed: running server computers in data-centres at close to 100% capacity for close to 100% of their lifetime means that the per-hour cost of renting each of those servers to remote users is only a few pennies, yet it offers the users the possibility of accessing vast amounts of compute resources that would otherwise cost the user hundreds of thousands of pounds. The move to data-centre-driven computing is unlikely to completely wipe out the need for reasonably-powered computers on people’s desks or laps, but those computers, and other devices such as mobile phones and ‘cloud terminals’, will increasingly be able to call upon extra hardware and software

---

[23] See, for example, the Serious Games Institute website at http://www.seriousgamesinstitute.co.uk, the Serious Games Initiative website at http://www.seriousgames.org/, and the various research outputs from FutureLab on Games and Learning, Serious Games in Education, Game-Based Experience in Learning, and Teaching with Games, available at http://www.futurelab.org.uk/projects/. Also a recent extensive report on the use of serious games in military education and training was produced by Caspian Learning for the UK Ministry of Defence, and is available from here: http://www.caspianlearning.co.uk/MoD_Defence_Academy_Serious_games_Report_04.11.08.pdf.

[24] For further discussion of Real Virtuality in educational contexts, see the Becta report available at http://emergingtechnologies.becta.org.uk/index.php?section=etr&rid=15005.

resources, remotely hosted, as those resources are needed, and paying only small amounts of money for those services. For education, the biggest impact in the next five years is most likely to be witnessed in the provision of remotely hosted administrative services – the ‘back office’ record keeping functions of most schools, colleges, universities, and education authorities could all be outsourced to remotely hosted service providers using currently available technology: the only obstacles to doing so are social and political ones. In classrooms, the IT suite of 2015 is likely to look superficially very similar to current installations, but the applications that are running on the PCs in the IT suite may be very different, tapping into vast amounts of computer power and data storage for just so long as they need to do so, running advanced applications that historically would have required a private supercomputer, and incurring costs of a few pennies each time they do so.

## Further Reading

In addition to the articles and papers cited in the footnotes of this report, the following sources may also be of interest.

- Probably the best book for a non-technical audience on the switch to remotely hosted computing services and the likely effects this will have on a variety of industries is Nicholas Carr's *The Big Switch: Rewiring the World from Edison to Google*, first published by W. W. Norton & Co. in 2008 and available as paperback since 2009. The paperback version has an additional postscript from the author, reflecting on the rapid explosion of interest and activity in cloud computing in the year after the book was first published. More detailed overviews can be found in the *Cloud Computing Use Cases Whitepaper*[25] available from the Open Cloud Manifesto organization's "resources" web-page; and in the widely-cited technical report *Above the Clouds: A Berkeley View of Cloud Computing* by Michael Armbrust and ten co-authors.[26]

- *The Economist*, the weekly newspaper, occasionally runs well-researched reports on developments in remotely hosted computing services. In the issue dated 23rd October 2008 they ran a special report "Let IT Rise" consisting of seven articles explaining and discussing various aspects of cloud computing. The issue dated 15th October 2009 ran a cover-story briefing titled "The Clash of the Clouds" exploring the growing levels of competition among leading multi-billion-dollar computing companies involved in providing remotely hosted computing services.

- Readers interested in learning how to write programs to run on Amazon's remotely hosted computers are referred to George Rees's *Cloud Application Architectures* (O'Reilly, 2009) and James Murty's *Programming Amazon Web Services* (O'Reilly, 2008), or to the developers' tutorials and articles on the Amazon Web Services website at http://aws.amazon.com/.

- 451 Group, an independent research and consultancy company, have a team that specialises in publishing authoritative reports on cloud computing that are based on lengthy conversations with large numbers of industry providers and end-users. As these reports are the primary output of 451's consultants, access to them is usually sold for prices substantially higher than for regular books. Despite this, 451's reports are treated as required reading by many people in the growing cloud computing industry. See www.451group.com.

---

[25] http://opencloudmanifesto.org/Cloud_Computing_Use_Cases_Whitepaper-3_0.pdf
[26] http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.html, recently published in revised form as an article in the *Communications of the Association for Computing Machinery*: http://cacm.acm.org/magazines/2010/4/81493-a-view-of-cloud-computing/