

MARK BURGESS

the cloud minders



Mark Burgess is professor of network and system administration at Oslo University College, Norway. He is the author of Cfengine, co-author of the SAGE Short Topics Booklet *A System Engineer's Guide to Host Configuration and Maintenance Using Cfengine*, and author of many books and research papers on system administration.

Mark.Burgess@iu.hio.no

EVERYTHING OLD IS NEW AGAIN. SO goes the famous adage, and never more so than in computing.

Distributed computing has conjured many mirages on its broad horizons over the years: from distributed computing environments to fully distributed operating systems, service oriented architecture, the grid, and now, lately, Cloud Computing. Should we be impressed? As a researcher and technologist, I am not. As a consumer, there is more to be said, but first let's look at the technology.

Each time one of these new manifestations of "wishful innovation" comes up at a conference I attend, I wince a little and wonder whether it will be worth committing to memory before the next one takes over. With so few ideas in IT management and so much palpable desperation to come up with something innovative in both research and industry, even the research community seems to have become blindly complacent about these magical phrases and may even see them as a godsend to fund one more round of paper recycling. But perhaps you are thinking, "The gentleman doth protest too much": We are all stuck in the same mess, cheered on by broken funding politics and commercial exuberance; after all, this is only a sign that information technology has truly entered the marketplace. And there is something to be said for the hype that provokes us into thinking about the upsides and downsides of computing economics, especially given current events in money markets.

This column is a comment on what is currently being called Cloud Computing. Normally, I would not bat much of an eyelid to anything so plainly construed, but on this occasion the name Cloud Computing itself is only a distraction. Of greater importance is what is being offered: the idea that computing as a rentable service is preferable to owning your own—and this in itself warrants some remarks.

The name Cloud Computing seems appropriate for something so vaporous, in which people see the shapes they want to. What is it really? Last year it seemed as insubstantial as a wisp of cirrus, often mentioned in connection with Web 2.0. Is it a web? Is it a cloud? No, it's really a kind of utility computing, and the Web is just its presumed application of choice.

Sale of online computing resources is not new, of course. It has been going on for some time, from shared tenancy computing in the 1960s [1], to email accounts with Hotmail, and then Gmail, Yahoo!, etc. Then came the rise of social network-

ing sites such as Facebook, online photography, and YouTube. The list goes on. At several junctures, there were briefly held notions of Application Service Providers (ASPs) changing the face of computing by running all of our software for us in centralized factories on the Internet, freeing ordinary companies from the burdens of coping with ever-changing technology. However, this only ever met with limited success. It still exists in a few forms, and indeed it has now brought on the idea of Cloud Computing, but it did not eradicate the stand-alone PC in favor of lighter, smaller “thin clients” as originally suggested, and this alone should be an omen to prevent us crying Hallelujah!

The Web-related hype has been snowballing into a silly idea: that not only will all computational resources be consolidated into mass production sweatshops of on-demand servers, but that the current flora of technologies would all be collapsed into a single kind of technological packaging—namely, that ever halting darling of ad hoc innovation, Web services. Supposedly, all of our applications might one day be provided by giant providers like Amazon, Google, and Sun, all willing to sell us storage or unlimited mileage accounts for such things as email, and all via the browser.

Where will our data be then? They will be all “out there,” as Captain Kirk might have said expansively with a twinkle in his eye, in the unknown folds of the global Internet, in no known location—just in “cyberspace.” Hmm.

The idea is of course inevitable, just as everything else about the globalization of the economy has been inevitable. But what is disturbing is the lack of thought in presenting this as “The Big Step Forward.” It makes me think of the alacrity with which people threw themselves into the economic bubbles of the past decade. I suggest that “Cloud Computing” is far from “The Future of Distributed Computing,” powering us on to the next generation, but likely only a footnote to a broader view of global services that will find a moderate market share in the commercial future of IT systems alongside a variety of other models. Why? Because of risk.

What Is Cloud Computing?

You might imagine from the sheer size of its Wikipedia entry that Cloud Computing was really something quite innovative and special. It is of course no different from any other kind of computing; it offers no new functions and no special features, and it is not necessarily any cheaper to provide on an hourly basis than any previous model for computing service, despite what is often claimed. In fact, on balance, it might even result in more carbon emissions, given the kinds of customers who are likely to use it. It certainly does nothing to improve the security of users, who still need a PC and a Web browser to access it, with all of those attendant flaws. It exists chiefly for convenience to a certain segment of the market for computer services.

Price and ease are the main driving forces for the online services being offered—large amounts of “cheap” storage, “cheap” applications to replace existing commercial ones, etc. For startups and hobbyists who have neither the expertise nor the resources to run their own servers, the idea of rentable virtual computers is an amazing convenience. Suddenly resources are available without searching for rackspace, network providers, or hosting companies, and without any investment in infrastructure. Fantastic!

But this could be misleading. We’ll look at the economics of this in a minute, but for now consider why someone would want to use a cloud service rather than running software on a PC as everyone has done in the past.

Let us imagine two computer users, Alice and Bob (as they are often cryptographically named), who want to send email to one another, or perhaps simply use an application such as a spreadsheet. Alice, having practiced in front of her looking glass, is an expert computer user and owns her own Custom Classic Computer, complete with air-brushed bodywork and a V8 processor and custom grown software built from spare parts she finds on the Net. Every day she tunes the engine a little to maximize performance or even just for enjoyment, and there are few problems that a little amateur tweaking cannot fix.

Bob, by way of contrast, is just floating along, going with the flow. His heart is not really in the mechanics of computing, but he enjoys a comfortable ride through the spreadsheets and word processors from time to time when he needs them—his needs vary so he doesn't want to tie himself to just one thing. His attitude is that he “just wants it to work” and he is willing to pay a rental company to fix this computing for him because it is cheaper than owning his own. That way he knows he will always have the latest and it will be pumped up with the latest hot air and checked by experts. So he normally rents a cheap service from MegaHertz or CloudAvis and they even throw in a built-in MP3 system and air conditioning for the servers (at the data center, naturally) for their frequent flyers.

Bob can never get the kind of souped-up custom experience that Alice enjoys from her personal computer system, but he wants neither the hassle of her infatuation nor the responsibility of owning a depreciating pile of capital expenditure. Alice, for her part, would be mortally offended by the mere suggestion that she should plump for anything as degradingly generic as a cloud service. She saves a bundle by doing it herself; after all, how hard could it be to add a hard disk, remember to do a backup, or install a new program—things that would make Bob shudder?

You get the idea. Cloud computing is much like the idea of car rental, ski rental, or any other kind of pay-as-you-go service. A kind of online Internet cafe for application services. You will pay more than you strictly need to get something quite generic, possibly with selectable levels of service quality (basic, super, or ultra) or one of a limited number of special-needs solutions (hatchback, van, or snowboard). It would be cheaper on an hour-by-hour basis to have your own, but you haven't; moreover, you don't have to pay for the thing, store it, and maintain it when you don't need it, so in the end it could be a lot cheaper for the occasional fair-weather user. When we compare this to renting a car, it is not so mysterious.

So cloud computing is about providing computing service (a lot of the hardware and all of the software) as a commodity without the need for a large and risky personal investment. The cloud provider will take that risk and investment, which of course is a lower risk if you know you will have sufficient customers—and in the case of the chief providers today, you actually know they can use the machines for their primary business if no one is buying the cloud service, so they are not losing anything. Some authors have likened this to making information technology run like water from a tap or flow like electricity from a wall socket, but, as we shall see, this analogy is not the right one. In fact it is more like a bank account with a credit card, with all of the risks that entails.

Why Is Cloud Computing Not Like Electricity?

The argument for cloud computing is an economic one. The argument is not that it is cheaper for everyone, only that it is a service that some will find

useful and that can be cheaply provided by some giants who do it as a kind of sideline, using their spare capacity to subsidize the sale. Cloud computing is not going to replace other forms of computing any more than car rentals have taken over the transport sector, because the model does not fit everyone's needs, but it could be quite useful to occasional computer users. Certainly the idea that companies might want to set up their own "local cloud" to make effective use of resource virtualization seems faintly ridiculous—you mean set up their own computer infrastructure, the way they've been doing for . . . how many years?

The real issues lurking for inexperienced consumers are the risks.

True, the cloud companies bear the risk of initial investment and they carry the cost of maintenance. But what shall we make of the subsidies they provide? If this is a sideline propped up by the cloud providers' core business model, then we should look rather carefully at whether that model is rock-solid and is likely to survive. Worse, cloud services are not like electricity or car rentals, because those services are "disposable" transactions. You consume these services once and then they are gone; nothing is stored or saved for the future. If they go under one day, you might be inconvenienced but you will not lose any savings. The economics are also easy to understand. There is a big pool of resources that can be shared by a lot of customers. With many customers a single provider can own an efficient fleet of cars or a flexible farm of servers and pay for them with a profit because there is always a sufficiency of customers coming back for more.

A bank is a more comparable service. Banks aspire to make money flow like electricity when needed, but with an important difference: Clients own their savings. A bank provides various services (perhaps for a fee, though these are mostly gone in Scandinavia) and they do it for the privilege of having your money for their use while you don't want it back. As long as there are many customers with enough money, the pool allows the bank to smooth over the inequities of individuals' financial details. The bank even pays you a nominal interest rate to cover the depreciation of the money due to inflation. There is no reason why people could not stuff money into mattresses or have their own private vault for storing money, but banks are successful because they provide certain conveniences. The key difference between banks and car rental is that banks provide safekeeping for something that matters to you: your money. But this is a risk we are usually willing to take. After all, banks don't go under, now do they? If you are in the black, the risk is yours. If you are in the red, the risk is theirs. For them, this evens out, but for you it doesn't, as thousands of pensioners and savers around the world learned over the past decade.

The key to pooling and sharing resources is that the fee for occasional use does not necessarily have to directly cover the actual cost of making the service available, as long as there are enough individuals to balance the incoming payments somehow. Alternatively, the whole thing can be subsidized by external funding. This makes the rental look cheap and stable to casual individuals. However, if the supply of money ever starts to get too low to smooth over the inequities, the transactions will grind to a halt, confidence in the model can be lost, and people will take their money elsewhere, causing a collapse. The model goes into a "recession" and the ones who remain could lose everything.

Cloud computing is much like a bank, because it will contain people's personal data and valuables. But for how long? If one of these services should suddenly stop working, that data would all be gone forever and you would never be able to rescue it, because it would still be "out there," lost in space.

The risk lies in the stability of the collective. If it does not attract enough individuals to maintain its services, or if it grows too many to service and maintain, or if there is not enough money to smooth over the imbalances in the pricing, then confidence in the system can be lost and it can collapse, meaning that all of its users will potentially lose everything stored there. Now, this doesn't matter for electricity or car rentals; you use them and then they are gone anyway. But with a bank you do care. And there is no central bank to bail out cloud computing.

The Stability of the Commons

No one seriously looks at Amazon or Google and thinks that these companies, the very knights of modern marketing savviness, will go under—but take care. Cloud companies do not have cloud computing as their primary business, at least for the moment, so they can effectively subsidize these collectives, making them seem artificially cheap. What happens when too many providers enter the market and prices rise? This could also cause a mass exodus from the providers.

Unlikely? No one expected the present banking crisis to emerge, either. All the funds were guaranteed by someone, weren't they? Unfortunately, when you are playing with margins there is not always an outside source that can come to the rescue in a sudden shortage market fluctuation. The problem with all stochastic systems (systems with fluctuations) is that there is pretty much always a freak wave out there that can wreak mass destruction in the system, one against which it simply doesn't pay to try to protect oneself, because it is so unlikely.

There is actually precedent for this kind of precipitated collapse of a collective commons in the Internet world already. The Internet has its own exchanges for trading spare capacity and pooling its resources (i.e., the Internet Exchanges). There major and minor network providers can trade their capacity either for money or, more often, for “Brownie points” or goodwill. Indeed, studies of these exchanges show that exchange agreements are based more on visibility than on material profit [2]. This means that the larger providers often do more than their fair share of giving away their spare capacity, and on occasion this has led to a major provider withdrawing from the exchange, leading to a crisis for the others, forcing them to pay real money for those Brownie points. This happened in Norway only a few years ago, placing confidence in the system in jeopardy [3].

But even if the possibility of collapse seems small, there are several causes for concern in cloud computing. One is security and privacy (who can see my data, and how do I verify the claim?); another is the question of geography. How about backup? If you need to have a backup for your data locally, then you either need some local infrastructure or you have to diversify your data investment over multiple providers that are not likely to go away all at the same time. What if there is a take-over? Will one copy go away? Will you know the physical location of the data and avoid the next big earthquake or flood?

What are the terms and conditions for the services? Does the (remote) provider retain the right to mine your data for marketing buzzwords? Will it be forced to reveal your private data to someone else under duress? Will it adequately destroy sensitive data when you ask it to, including all of the backup copies? Will its backups be properly secured? What about geography? Where precisely is your data stored? Is the data illegal in the country of storage? Will you always be able to access it? Is there political (or tectonic) stability in

the location of the data? The potential problem is that there is practically no way to assess these risks. It's all just "out there."

For these reasons, cloud computing is not going to be for everyone. The Alices of the world are never going to find Wonderland in outsourcing. They live off the ability to customize nonstandard systems, and they have a heavy weapon against it: competence. Competence and technology actually make it cheap for individuals to manage their own concerns. There is no single recipe for solving the problem of scale, as we have discovered in our research into systems in Oslo. Centralization is but one approach to resource management [4].

Self-Reliance: The Counterpoint of Cloud Computing

There is another weapon in the computing arsenal that could play a role. It has come increasingly to the fore of late. It began in the 1990s with artificial immune systems or computer immunology, and today it is often called "self-healing" technology. By contrast with cloud computing, which is mainly a brute-force cost reduction, self-healing is a set of more subtle technologies. The idea is (as with smart modern cars) to get experts to program the requirements and safe working conditions for computer systems in advance and then equip the units with smart technology that allows them to maintain this condition for the greatest possible time, ultimately eliminating the need for human intervention until an unexpected decision has to be made. Automation is a technology that can level the playing field again, removing some of the benefits of cloud convenience.

Whereas brute-force cost-cutting would try to make everything absolutely identical in order to keep down costs, the self-healing configuration approach actually tries to improve the technology itself to make a more effective system manage itself cheaply. Futurist Alvin Toffler wrote about this phenomenon in manufacturing at the end of the 1960s and concluded, "As technology becomes more sophisticated, the cost of introducing variations declines."

The differentiating self-healing technologies such as Cfengine, and to some extent IBM's autonomic initiative and HP's work in the area, are taking a different path to the idea of the cloud (and the Cloud Minders too can benefit from it), namely, bringing computer expert systems back to support complexity cheaply rather than offering only vanilla and strawberry flavored services to potential buyers (i.e., any color as long as it's black).

Self-healing, then, could be the thorn in the side of naive cloud computing, making resource flexibility easy at home. Do Hertz and Avis outsource cars to specialist companies, or do companies buy their own car pools? Of course both models exist, just as "cloud computing" is likely to coexist with in-house expertise, enabled with powerful self-healing systems in the years to come. Consolidation did not capture the market and change the world before, so why should it now? Pretty much every development in personal technology, starting with the motor car, has been about the opposite of pooling resources: mobile phones, microwave ovens, PDAs, Blackberries, iPods are all about personal enablement, making oneself independent of ties.

Consolidation is a strategy for the non-resource-wealthy that pokes its head up and dives down again like the Loch Ness Monster, at reasonably regular intervals in computing. When a resource becomes scarce, it encourages pooling of those resources through consolidation. Sometimes it was the need for processor capacity, sometimes it was memory, sometimes fast communication. The scarce resource today is *competence*, specifically in the areas of

management and maintenance, but self-healing will take away much of the need for this too.

The technological phenomenon is the growth of computer virtualization for effective resource management. This is a healthy reality check, as power requirements force us to reconsider vulgar excesses. The role of the Internet will only come into play if resources can be moved dynamically around the globe to optimize time zones and traffic burst in a dynamic and secure fashion. That would be a true technology to propel us into the stratosphere.

Unlike some, I am not bowled over by cloud computing, any more than I was impressed by grid services or any other special packaging for distributed computing. Yes, of course there are arguments for it. It has its place where expertise is lacking or temporary, throw-away resources are required on short notice, but this is not a fundamental shift, only a commercial opportunity, and it does not free users from their responsibilities for thinking about backup and security. More convincing are the benefits of renting software as a service: paying a smaller regular fee for continuous updates, in which we keep our own data privately and safely. This is a model for nearly all regular computer users.

Marketing is a powerful force that is sometimes genuinely creative. I only wish that as much effort could be expended in educating competent specialists to solve the technical challenges of resource management as is put into the manufacturing of media hype to merely suggest overcoming them. Brute-force mediocrity is almost a standard for computing today, even in research. Personally, I am holding out for the next level: self-healing computers, with self-scaling automation, that can be deployed anywhere, not just in vast datacenters. This has nothing to do with the Web or the Internet but has everything to do with intelligent configuration.

As Droxine, the lovely daughter of the Cloud Minder, said in the memorable Star Trek story, “I shall go to the mines; I no longer wish to be limited to the clouds” [5].

ACKNOWLEDGMENTS

I would like to acknowledge helpful discussions and input from Khanan Grauer, Rik Farrow, and Helge Skrivervik. None of them necessarily shares any of the opinions expressed here.

REFERENCES

- [1] For example, http://en.wikipedia.org/wiki/Automatic_Data_Processing.
- [2] W.B. Norton, “The Art of Peering: The Peering Playbook,” Equinix.com, 2001.
- [3] For example, <http://www.nettavisen.no/it/article1245673.ece> and <http://www.dagbladet.no/kultur/2007/06/20/504054.html>. (Search Google for “Oslo telenor trekker seg internet exchange” for references in Norwegian.)
- [4] For example, see research at <http://research.iu.hio.no/promises.php>.
- [5] Paramount Pictures, *Star Trek* (Original Series 3), “The Cloud Minders.”