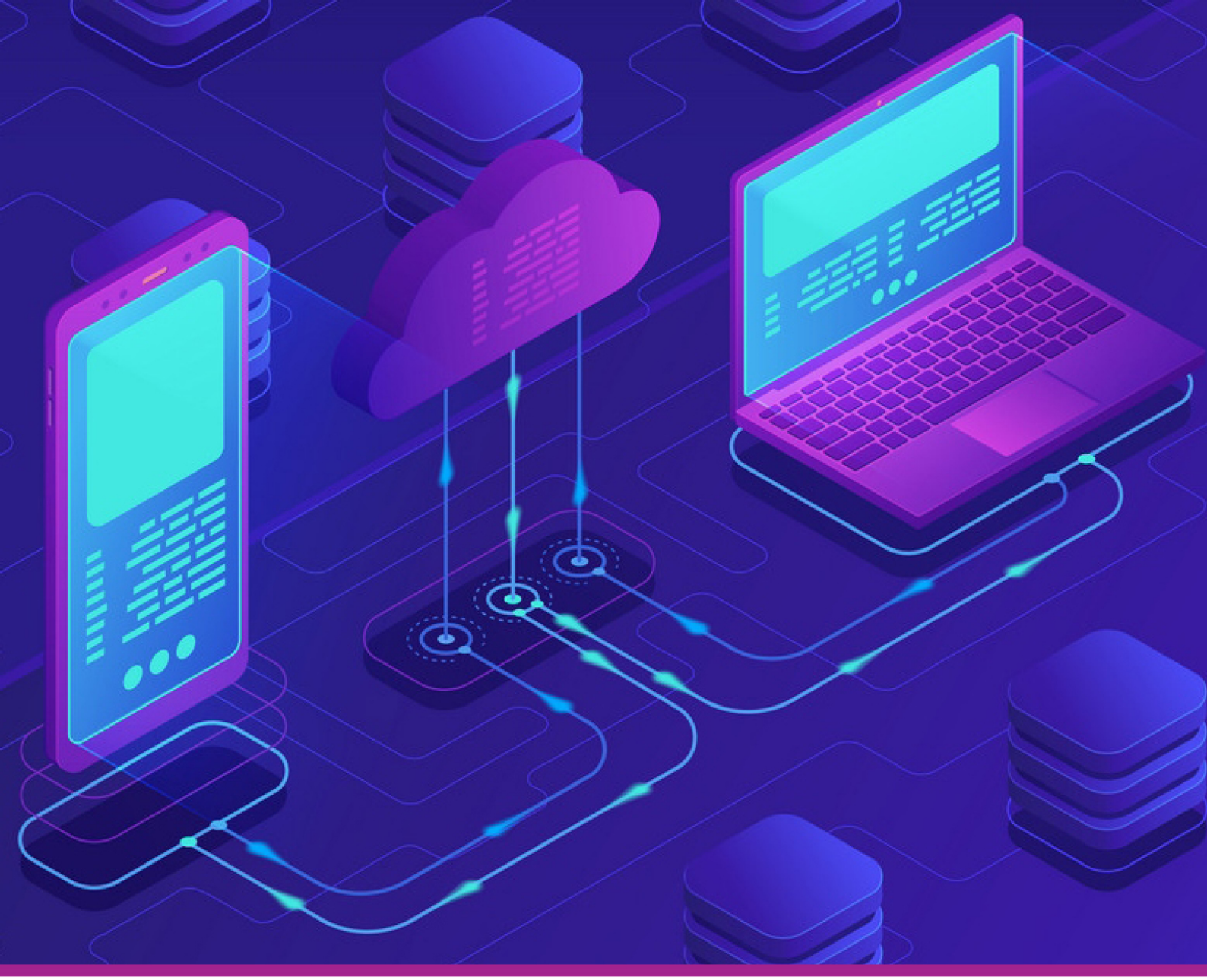


Cloud Computing Elements

Oliver Walls

CLOUD COMPUTING ELEMENTS

CLOUD COMPUTING ELEMENTS

Oliver Walls

Cloud Computing Elements
by Oliver Walls

www.bibliotex.com

Ebook ISBN: 9781984663993

Published by:
Bibliotex
Canada
Website: www.bibliotex.com

Contents

1

Introduction

A hybrid computing model enables an organization to leverage both public and private computing services to create a more flexible and cost-effective computing utility:

- The public cloud is a set of hardware, networking, storage, service, and interfaces owned and operated by a third party for use by other companies or individuals.
- A private cloud is a set of hardware, networking, storage, service, and interfaces owned and operated by an organization for the use of its employees, partners, and customers.
- In a hybrid cloud environment, an organization combines services and data from a variety of models to create a unified, automated, and well-managed computing environment.

Whether your cloud is public, private, or hybrid, you'll need a cloud provider that provides elasticity, scalability, provisioning, standardization, and billed usage. Elasticity is important because it means that you are able to use a service for a long or short period of time based on need. You can add more services from a self-service portal rather than wait for IT to do the heavy lifting for you. Increasingly, as companies begin to understand that they will use a combination of different platforms to meet different business needs, the hybrid cloud will become the foundation for computing. The advent of the hybrid cloud will also help redefine the purpose and use of the traditional data centre as well.

One of the fundamental differences between cloud computing and traditional computing is the way a cloud is designed to manage resources. Whereas the data centre is designed to manage applications, the cloud is intended to manage a pool of resources. A pool of resources is precisely what it sounds like — a set of shared, configured services that are independent of physical location.

For example, suppose you are a cloud provider. You do not want customers to have to select one server or one storage system; rather, the customer is abstracted from that idea. Instead, the customer simply says I need some more storage, and those storage resources are pooled together from various physical systems to create a set of resources. Customers never know which storage system they are accessing. To make

resource pooling work, it's important that each element that is pooled be written with service-oriented constructs in mind. This means that each resource is written as an independent service without dependencies and with well-defined interfaces.

A HISTORY OF CLOUD COMPUTING

Cloud computing has evolved through a number of phases which include grid and utility computing, application service provision (ASP), and Software as a Service (SaaS). But the overarching concept of delivering computing resources through a global network is rooted in the sixties.

The idea of an "intergalactic computer network" was introduced in the sixties by J.C.R. Licklider, who was responsible for enabling the development of ARPANET (Advanced Research Projects Agency Network) in 1969.

His vision was for everyone on the globe to be interconnected and accessing programmes and data at any site, from anywhere, explained Margaret Lewis, product marketing director at AMD. "It is a vision that sounds a lot like what we are calling cloud computing."

Other experts attribute the cloud concept to computer scientist John McCarthy who proposed the idea of computation being delivered as a public utility, similar to the service bureaus which date back to the sixties.

Since the sixties, cloud computing has developed along a number of lines, with Web 2.0 being the most recent evolution. However, since the internet only started to offer significant bandwidth in the nineties, cloud computing for the masses has been something of a late developer.

One of the first milestones for cloud computing was the arrival of Salesforce.com in 1999, which pioneered the concept of delivering enterprise applications via a simple web site. The services firm paved the way for both specialist and mainstream software firms to deliver applications over the internet.

The next development was Amazon Web Services in 2002, which provided a suite of cloud-based services including storage, computation and even human intelligence through the Amazon Mechanical Turk.

Then in 2006, Amazon launched its Elastic Compute cloud (EC2) as a commercial web service that allows small companies and individuals to rent computers on which to run their own computer applications.

"Amazon EC2/S3 was the first widely accessible cloud computing infrastructure service," said Jeremy Allaire, CEO of Brightcove, which provides its SaaS online video platform to UK TV stations and newspapers. Another big milestone came in 2009, as Web 2.0 hit its stride, and Google and others started to offer browser-based enterprise applications, though services such as Google Apps.

"The most important contribution to cloud computing has been the emergence of "killer apps" from leading technology giants such as Microsoft and Google.

When these companies deliver services in a way that is reliable and easy to consume, the knock-on effect to the industry as a whole is a wider general acceptance of online services," said Dan Germain, chief technology officer at IT service provider Cobweb Solutions.

Other key factors that have enabled cloud computing to evolve include the maturing of virtualisation technology, the development of universal high-speed bandwidth, and universal software interoperability standards, said UK cloud computing pioneer Jamie Turner. Turner added, "As cloud computing extends its reach beyond a handful of early-adopter Google Docs users, we can only begin to imagine its scope and reach. Pretty much anything can be delivered from the cloud."

FOLLOWING THE CLOUD

"Many IT professionals recognise the benefits cloud computing offers in terms of increased storage, flexibility and cost reduction," said Songnian Zhou, chief executive officer of Platform Computing. But he added that IT directors still have concerns about the security of their corporate data in the cloud. This means that it will be 2010 at the earliest before cloud adoption sees increased growth. Julian Friedman, a specialist in emerging technologies, said that security and other concerns will soon be resolved.

"Considerations such as security, data privacy, network performance and economics are likely to lead to a mix of cloud computing centres both within the company firewall and outside of it." He added that today's applications will naturally move towards a cloud model as they become more pervasively available through the web, require more data processing, and span the boundaries of multiple devices.

Experts seem to agree that cloud computing will ultimately transform today's computing landscape. Andreas Asander, vice-principal of product management at virtualisation security specialist Clavister, said that once the security issues are resolved, cloud computing services "can enable an enterprise to expand its infrastructure, add capacity on demand, or outsource the whole infrastructure, resulting in greater flexibility, a wider choice of computing resources and significant cost savings." It is clear that cloud computing can bring enormous benefits for IT users. However, the bottom line for IT directors is that they will need to continue to manage their internal computing environments, whilst learning how to secure, manage and monitor the growing range of external resources residing in the cloud.

A BRIEF HISTORY OF CLOUD COMPUTING

Believe it or not, "cloud computing" concepts date back to the 1950s when large-scale mainframes were made available to schools and corporations.

The mainframe's colossal hardware infrastructure was installed in what could literally be called a "server room" (since the room would generally only be able to hold a single mainframe), and multiple users were able to access the mainframe via "dumb terminals" – stations whose sole function was to facilitate access to the mainframes.

Due to the cost of buying and maintaining mainframes, an organization wouldn't be able to afford a mainframe for each user, so it became practice to allow multiple users to share access to the same data storage layer and CPU power from any station. By enabling shared mainframe access, an organization would get a better return on its investment in this sophisticated piece of technology.

A couple decades later in the 1970s, IBM released an operating system called VM that allowed admins on their System/370 mainframe systems to have multiple virtual systems, or "Virtual Machines" (VMs) on a single physical node. The VM operating system took the 1950s application

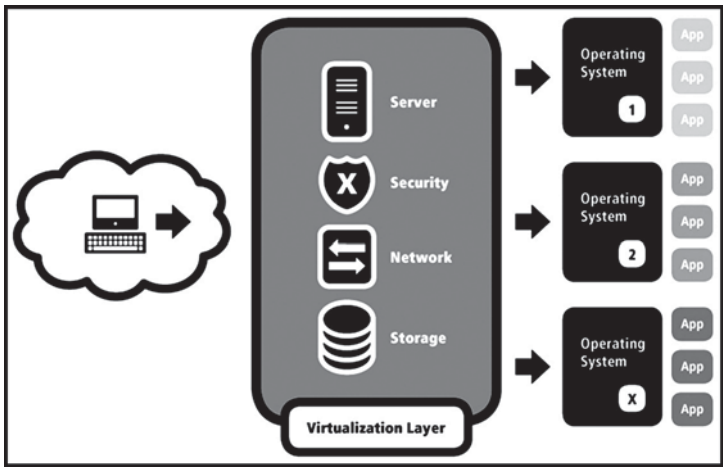
of shared access of a mainframe to the next level by allowing multiple distinct compute environments to live in the same physical environment.

Most of the basic functions of any virtualization software that you see nowadays can be traced back to this early VM OS: Every VM could run custom operating systems or guest operating systems that had their "own" memory, CPU, and hard drives along with CD-ROMs, keyboards and networking, despite the fact that all of those resources would be shared. "Virtualization" became a technology driver, and it became a huge catalyst for some of the biggest evolutions in communications and computing.

In the 1990s, telecommunications companies that had historically only offered single dedicated point–to-point data connections started offering virtualized private network connections with the same service quality as their dedicated services at a reduced cost. Rather than building out physical infrastructure to allow for more users to have their own connections, telco companies were able to provide users with

shared access to the same physical infrastructure. This change allowed the telcos to shift traffic as necessary to allow for better network balance and more control over bandwidth usage. Meanwhile, virtualization for PC-based systems started in earnest, and as the Internet became more accessible, the next logical step was to take virtualization online.

If you were in the market to buy servers ten or twenty years ago, you know that the costs of physical hardware, while not at the same level as the mainframes of the 1950s, were pretty outrageous. As more and more people expressed demand to get online, the costs had to come out of the stratosphere, and one of the ways that was made possible was by … you guessed it … virtualization. Servers were virtualized into shared hosting environments, Virtual Private Servers, and Virtual Dedicated Servers using the same types of functionality provided by the VM OS in the 1950s.



As an example of what that looked like in practice, let's say your company required 13 physical systems to run your sites and applications. With virtualization, you can take those 13 distinct systems and split them up between two physical nodes. Obviously, this kind of environment saves on infrastructure costs and minimizes the amount of actual hardware you would need to meet your company's needs.

As the costs of server hardware slowly came down, more users were able to purchase their own dedicated servers, and they started running into a different kind of problem: One server isn't enough to provide the resources I need. The market shifted from a belief that "these servers are expensive, let's split them up" to "these servers are cheap, let's figure out how to combine them." Because of that shift, the most basic understanding of "cloud computing" was born online. By installing and configuring a piece of software called a hypervisor across multiple physical nodes, a system would present all of the environment's resources as though those resources were in a single physical node. To help visualize that environment, technologists used terms like "utility computing" and "cloud computing" since the sum of the parts seemed to become a nebulous blob of computing resources that you could then segment out as needed (like telcos did in the 90s).

In these cloud computing environments, it became easy add resources to the "cloud": Just add another server to the rack and configure it to become part of the bigger system.

As technologies and hypervisors got better at reliably sharing and delivering resources, many enterprising companies decided to start carving up the bigger environment to make the cloud's benefits to users who don't happen to have an abundance of physical servers available to create their own cloud computing infrastructure.

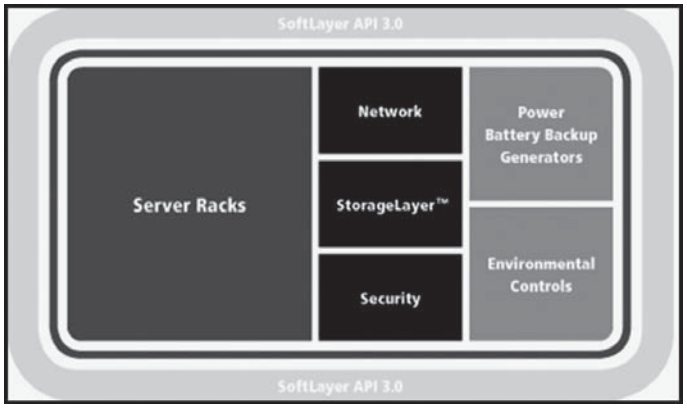
Those users could order "cloud computing instances" (also known as "cloud servers") by ordering the resources they need from the larger pool of available cloud resources, and because the servers are already online, the process of "powering up" a new instance or server is almost instantaneous. Because little overhead is involved for the owner of the cloud computing environment when a new instance is ordered or cancelled (since it's all handled by the cloud's software), management of the environment is much easier. Most companies today operate with this idea of "the cloud" as the current definition, but SoftLayer isn't "most companies."

SoftLayer took the idea of a cloud computing environment and pulled it back one more step: Instead of installing software on a cluster of machines to allow for users to grab

pieces, we built a platform that could automate all of the manual aspects of bringing a server online without a hypervisor on the server. We call this platform "IMS."

What hypervisors and virtualization do for a group of servers, IMS does for an entire data centre. As a result, you can order a bare metal server with all of the resources you need and without any unnecessary software installed, and that server will be delivered to you in a matter of hours. Without a hypervisor layer between your operating system and the bare metal hardware, your servers perform better.

Because we automate almost everything in our data centres, you're able to spin up load balancers and firewalls and storage devices on demand and turn them off when you're done with them. Other providers have cloud-enabled servers. We have cloud-enabled data centres.



IBM and SoftLayer are leading the drive towards wider adoption of innovative cloud services, and we have ambitious goals for the future. If you think we've come a long way from the mainframes of the 1950s, you ain't seen nothin' yet.

BENEFIT OF CLOUD COMPUTING

There is a lot of benefit for the business looking for the service from the cloud service provider. Apart from the bundle of suits they have to offer, it focus all an escape from huge investment into IT infrastructure and operating cost.

REDUCE RUNTIME AND RESPONSE TIME

For applications that use the cloud essentially for running batch jobs, cloud computing makes it straightforward to use 1000 servers to accomplish a task in 1/1000 the time that a single server would require.

The New York Times example cited previously is the perfect example of what is essentially a batch job whose run time was shortened considerably using the cloud.

For applications that need to offer good response time to their customers, refactoring applications so that any CPU-intensive tasks are farmed out to 'worker' virtual machines can help to optimize response time while scaling on demand to meet customer demands.

The Animoto application cited previously is a good example of how the cloud can be used to scale applications and maintain quality of service levels.

Minimise Infrastructure Risk

IT organizations can use the cloud to reduce the risk inherent in purchasing physical servers. Will a new application be successful? If so, how many servers are needed

and can they be deployed as quickly as the workload increases?

If not, will a large investment in servers go to waste? If the application's success is short-lived, will the IT organization invest in a large amount of infrastructure that is idle most of the time? When pushing an application out to the cloud, scalability and the risk of purchasing too much or too little infrastructure becomes the cloud provider's issue.

In a growing number of cases, the cloud provider has such a massive amount of infrastructure that it can absorb the growth and workload spikes of individual customers, reducing the financial risk they face.

Another way in which cloud computing minimizes infrastructure risk is by enabling surge computing, where an enterprise data centre (perhaps one that implements a private cloud) augments its ability to handle workload spikes by a design that allows it to send overflow work to a public cloud.

Application lifecycle management can be handled better in an environment where resources are no longer scarce, and where resources can be better matched to immediate needs, and at lower cost.

Lower Cost of Entry

Since the infrastructure is rented, not purchased, the cost is controlled, and the capital investment can be zero. In addition to the lower costs of purchasing compute cycles

and storage "by the sip," the massive scale of cloud providers helps to minimize cost, helping to further reduce the cost of entry.

Applications are developed more by assembly than programming. This rapid application development is the norm, helping to reduce the time to market, potentially giving organizations deploying applications in a cloud environment a head start against the competition.

Increased Pace of Modernism

Cloud computing can help to increase the pace of innovation. The low cost of entry to new markets helps to level the playing field, allowing start-up companies to deploy new products quickly and at low cost. This allows small companies to compete more effectively with traditional organizations whose deployment process in enterprise data centres can be significantly longer. Increased competition helps to increase the pace of innovation — and with many innovations being realized through the use of open source software, the entire industry serves to benefit from the increased pace of innovation that cloud computing promotes.

Free from Software Licensing/up Gradation/ preservation

Cloud computing frees up user from any further licensing of the software or from up gradation and maintenance. All the services are provided by the service providers. No longer having to worry about constant server updates and other

computing issues, government organizations will be free to concentrate on innovation.

A Mobile Outline

Since all is accessible through internet, it will be accessible globally. It will be too much beneficial for a small and medium sized enterprise that is not willing to invest a lot in network setup and wish to free from maintenance.

AN INTERIM EVALUATION FOR THE COMMERCE

In cloud computing models, customers do not own the infrastructure they are using; they basically rent it, or pay as they use it. The loss of control is seen as a negative, but it is generally out-weighed by several positives. One of the major selling points of cloud computing is lower costs. Companies will have lower technology-based capital expenditures, which should enable companies to focus their money on delivering the goods and services that they specialize in. Still there are key features for consideration before one talk for the need of the business. Since entire gamut of services is available in the market one has to be very choosey and do lots of self evaluation before drawing a final plan for the business.

- In which stage of your business life cycle you are planning to scale for the service of cloud computing?
- What business line you need to support and how much is the requirement os for your business.

- How much cost effective it can be when you rent the services?
- Which type of service is going to be beneficial for you?
- What is the organization preferred technology, development platform and business that require for this type of service?
- Is your organization having the capabilities to handle these services, as these services needs lot of competency to handle it as there are lots of mechanism with different layers of service present in them.
- How much risk is associated with the data dependency when it is a kept in others infrastructure?
- How much performance and bandwidth is required to use this type of service with comparison to the current business needs? Is the company able to cope it up with the existing bandwidth to its business needs?

There is no limit for the evaluation, and consideration should be made with respect to the current business in one is, with respect to the multiple factors with responsiveness towards stake holders and business needs, financial goals, investment capabilities, profitability, future planning, industrial growth, service providers offerings etc.

One can only earn the advantage through the new technology only if they are able to do a correct feasibility study to mitigate the business need.

Disadvantage Technology

As any technology is a boon for an evaluation as the history is evidence, there are disadvantages too which cannot be ignored. Despite a fact cloud computing has so many features which can be awaiting a new horizon there are also key factors which cannot be ignored. Few have been summed up below:

- Lack of connectivity causes 100 per cent downtime, whereas with traditional applications, lack of connectivity allows for some local function to continue until connectivity is restored.
- The lack of industry-wide standards means that a usage surge can easily overwhelm capacity without the ability to push that usage to another provider.
- Companies providing computing services will over-sell these services similar to how bandwidth is over-sold based on average or "peak" usage, instead of "maximum" usage. ISP's typically operate at multiples of 5 to 1, where they sell.
- 5 times more than they have in capacity, assuming users will not use more than 20 per cent of their allotted resources. This works, until there is a popular YouTube video that everyone wants to see at the same time.... resulting in outages. Cloud computing is even more vulnerable to the peak-usage problem than internet bandwidth.

- "Denial of service" attacks, currently common, become easier. What's more they become harder to trace, as compromised "cloud resources" can be leveraged to launch the attacks, rather than compromised "individual pc's". Cloud computing is vulnerable to massive security exploits. Currently, when a system is broken into, only the resources of that system are compromised. With cloud computing, the damages caused by a security breach are multiplied exponentially.
- By "centralising" services, cloud computing increases the likelihood that a systems failure becomes "catastrophic", rather than "isolated".
- No political approach has been made till date to control the uncontrolled factors to bring the service under the boundary lines of trust and owner ship, as these services are beyond country lines.

IMPORTANT CLOUD COMPUTING FOR BUSINESS USERS

Like any new IT trend, Cloud Computing gets its fair share of hype, and with it comes a multitude of vendors that use the terms in ways it was never intended for, therefore making it devoid of any sense. When pushed to the extreme, a simple server connected to a network seems to qualify as a cloud, thereby allowing pundits such as Larry Ellison to deride the concept to no end.

Yet Cloud Computing is much more than a passing fad. It is a major step forward in the development of distributed computing, and one that will reshape the IT industry for years to come. But for it to happen, we must agree on a clear definition of the concept, and the less technical it is, the better. Let us introduce one that focuses exclusively on the business benefits of cloud computing.

Defines cloud computing in the following fashion:

- "Cloud computing is the provision of dynamically scalable and often virtualised resources as a service over the Internet. Users need not have knowledge of, expertise in, or control over the technology infrastructure in the "cloud" that supports them. Cloud computing services often provide common business applications online that are accessed from a web browser, while the software and data are stored on the servers."

UTILITY PRICING OF CLOUD COMPUTING

Cloud Computing is first a foremost defined by its utility-based pricing model. Users of the platform consume computing and storage services on demand and pay for them as they go, using an Operating Expenses (OPEX) budget, instead of paying for infrastructure resources up-front using Capital Expenditures (CAPEX).

For example, a Director of Sales can create CRM accounts for 10 of her sales people on Salesforce.com by using her

corporate credit card, without having to ask the CFO for a budget, and without having the IT Department initiate a requisition process for a new server.

Elastic Resource Capacity

Cloud Computing differs from more traditional forms of distributed computing in the way it scales computing and storage resources up and down. Instead of tapping from a fixed set of resources, users can add or remove capacity at will, almost instantaneously, and only pay for what they actually use. While utility pricing let users pay as they go, elastic resource capacity let them pay as they grow (or shrink). Following our previous example, the Director of Sales can add 5 more accounts for the sales people that were recently added to her team following the merger with another company, without having to worry about adding new servers or buying more hard drives.

Virtualized Capital

Cloud Computing would not be possible without virtualization, not for arcane technical reasons, but for one obvious business requirement: the need for multi-tenancy. In order to benefit from economies of scale, cloud computing is predicated upon the sharing of a common infrastructure by multiple groups of users, often referred to as tenants. And multi-tenancy can only be achieved through some kind of virtualization, either at the database level (Salesforce.com),

application server level (Google AppEngine), kernel level (Red Hat), or CPU level (Amazon EC2). Unlike grid computing, which often pooled and aggregated distributed computing resources for the purpose of handling very large computing jobs that could not fit or would take took long to complete on a single server, Cloud Computing creates virtual slices of resources from clusters of servers and storage devices, perfectly sized to fit the specific needs of multiple users. Such virtual resources can be small or large, and scale elastically as user needs evolve over time. In our previous example, virtualization means that the CRM application used by our sales team is served by an infrastructure also used by over 60,000 other tenants, all securely isolated from each other (hopefully).

Organization Automation

Cloud Computing platforms differ from traditional corporate data-centres in one major way: standardization. While your typical data-centre will usually host every versions of every operating systems and databases known to mankind, thereby creating massive management overhead, most Cloud Computing platforms usually standardize on a single kind of CPU (x86-based predominantly), a single hypervisor (VMware, Xen, etc.), a single operating system (some Linux distribution usually), and a single database (MySQL rules). This standardization has an obvious business benefit: dramatic reduction of operating costs through aggressive

management automation. Following our previous example, the sales team's CRM application is served by one of 16 instances, each made of a few dozens servers. Altogether, this infrastructure might require anywhere from 100 to 200 full-time resources to manage. As a point of comparison, if each of Salesforce.com's 60,000 customers were to require a dedicated infrastructure, it would take several thousands full-time resources to manage it all.

Self-service Provisioning

Cloud Computing and Software as a Service is often compared to the Application Service Provider (ASP) model that became popular for a brief period of time ten years ago. One element makes them fundamentally different from each other though: self-service provisioning. With the ASP model, dedicated servers had to be provisioned for each customers, which meant that technical resources had to be involved every time a new customer would be signed. Hefty setup fees would be added to the bill, and the service would become operational within a few days at best. With Cloud Computing, business end users like our Director of Sales can provision applications and user accounts in a few mouse clicks, and these become available instantly.

Third-party Ownership

Cloud Computing is also a new form of outsourcing. Customers trying to focus the allocation of scarce capital

resources to their core businesses soon realize the benefits of moving IT infrastructure off their balance sheet. Furthermore, as technology evolves and leading service providers roll-out ever larger data-centres, the acquisition and operation of state-of-the-art data-centre facilities makes less and less sense from an economic standpoint for most organizations.

Cloud Computing is all about the transfer of ownership for such resources to a third-party that specializes in their deployment. According to our previous example, the company using the CRM application provided by Salesforce.com does not own any infrastructure beyond a few laptop computers. Everything else, from data-centres to servers and storage systems is owned by Salesforce.com, Inc.

Managed Operations

Cloud Computing is finally about allocating human resources to tasks that will directly impact the business, rather than simply managing the infrastructure that supports it. As such, Cloud Computing advocates a model according to which the IT infrastructure is not only owned by a third-party, but managed by the third-party as well. Software upgrades, data backups, and the countless other tasks required to manage mission-critical business applications on a day to day basis become the responsibility of a third-party, according to well-defined Service Level Agreements. Following our example, the Director of Sales discovered this

morning the snowman adorned logo for the Winter 2010 version of Salesforce.com, without having taken any part in the software upgrade process that took place over the weekend. In the cloud, ignorance is bliss.

CLOUD COMPUTING IS A SERVICE

The simplest thing that a computer does is allow us to store and retrieve information. We can store our family photographs, our favourite songs, or even save movies on it. This is also the most basic service offered by cloud computing. Flickr is a great example of cloud computing as a service. While Flickr started with an emphasis on sharing photos and images, it has emerged as a great place to store those images. In many ways, it is superior to storing the images on your computer. First, Flickr allows you to easily access your images no matter where you are or what type of device you are using. While you might upload the photos of your vacation to Greece from your home computer, you can easily access them from your laptop while on the road or even from your iPhone while sitting in your local coffee house.

Second, Flickr lets you share the images. There's no need to burn them to a compact disc or save them on a flash drive. You can just send someone your Flickr address. Third, Flickr provides data security. If you keep your photos on your local computer, what happens if your hard drive crashes? You'd better hope you backed them up to a CD or a

flash drive! By uploading the images to Flickr, you are providing yourself with data security by creating a backup on the web. And while it is always best to keep a local copy — either on your computer, a compact disc or a flash drive — the truth is that you are far more likely to lose the images you store locally than Flickr is of losing your images.

This is also where grid computing comes into play. Beyond just being used as a place to store and share information, cloud computing can be used to manipulate information. For example, instead of using a local database, businesses could rent CPU time on a web-based database.

The downside? It is not all clear skies and violin music. The major drawback to using cloud computing as a service is that it requires an Internet connection. So, while there are many benefits, you'll lose them off if you are cut off from the Web.

Cloud Computing is a Platform

The web is the operating system of the future. While not exactly true — we'll always need a local operating system — this popular saying really means that the web is the next great platform.

What's a platform? It is the basic structure on which applications stand. In other words, it is what runs our apps. Windows is a platform. The Mac OS is a platform. But a platform doesn't have to be an operating system. Java is a platform even though it is not an operating system.

Through cloud computing, the web is becoming a platform. With trends such as Office 2.0, we are seeing more and more applications that were once the province of desktop computers being converted into web applications. Word processors like Buzzword and office suites likeGoogle Docs are slowly becoming as functional as their desktop counterparts and could easily replace software such as Microsoft Office in many homes or small offices. But cloud computing transcends Office 2.0 to deliver applications of all shapes and sizes from web mashups to Facebook applications to web-based massively multiplayer online role-playing games. With new technologies that help web applications store some information locally — which allows an online word processor to be used offline as well — and a new browser called Chrome to push the envelope, Google is a major player in turning cloud computing into a platform.

Cloud Computing and Interoperability

A major barrier to cloud computing is the interoperability of applications. While it is possible to insert an Adobe Acrobat file into a Microsoft Word document, things get a little bit stickier when we talk about web-based applications.

This is where some of the most attractive elements to cloud computing — storing the information on the web and allowing the web to do most of the 'computing' — becomes a barrier to getting things done. While we might one day be able to insert our Google Docs word processor document into our

Google Docs spreadsheet, things are a little stickier when it comes to inserting a Buzzword document into our Google Docs spreadsheet.

Ignoring for a moment that Google probably doesn't want you to have the ability to insert a competitor's document into their spreadsheet, this creates a ton of data security issues. So not only would we need a standard for web 'documents' to become web 'objects' capable of being generically inserted into any other web document, we'll also need a system to maintain a certain level of security when it comes to this type of data sharing. Possible? Certainly, but it isn't anything that will happen overnight.

2

Characteristics of Cloud Computing

So what, you may reasonably ask, is the cloud? Well, for years the Internet has been represented on network diagrams by a cloud symbol. When, around 2008, a variety of new services started to emerge that permitted computing resources to be accessed over the Internet, the label "cloud computing" therefore emerged as an umbrella term. Does this mean that we really ought to be talking about "Internet computing"? Well, perhaps. However, in the strictest sense, the "cloud" is a label for online computing resources rather than the entire Internet. The term "cloud computing" is also useful to separate the kinds of things we have been doing online for a couple of decades from a totally new age of online software and processing power.

We are already said that cloud computing is where software applications, processing power, data and potentially even artificial intelligence are accessed over the Internet. Building on this basic definition, it can also be stated that cloud computing is where dynamically scalable, device-independent and task-centric computing resources are provided online, with all charges being on a usage basis. Cloud computing is dynamically scalable because users only have to consume the amount of online computing resources they actually want. Just as we are used to drawing as much or as little electricity as we need from the power grid, so anybody can now obtain as many or as few computing resources from the cloud as they require at any particular point in time.

Cloud vendors including Amazon Web Services (AWS) now quite literally sell computer processing power by the hour. For example, anybody can now rent "virtual server instances" from Amazon's Elastic Compute Cloud or "EC2" service for as little as $0.02 an hour (or indeed you even sign up for a one-year trial of the AWS Free Usage Tier for nothing). As Amazon explain, "EC2 reduces the time required to obtain and boot new server instances to minutes, allowing [customers] to quickly scale capacity, both up and down, as [their] computing requirements change".

Cloud computing is device-independent because cloud computing resources can be accessed not just from any computer on the Internet, but also any type of computer.

Provided that it has an Internet connection and a web browser, it really does not matter if the computer being used is a traditional desktop or laptop PC, or a netbook, tablet, smartphone, e-book reader, or any other kind of cloud access device. Such device independency is also a killer feature of cloud computing because it means that users can move between computing devices — such as their work PC, home PC, laptop and netbook — without having to worry that they will always have access to the latest versions of their files.

Cloud computing is task centric because the usage model is based entirely around what users want to achieve, rather than any particular software, hardware or network infrastructure. Users do not have to purchase or install anything before using a cloud computing resource. Nor do they have to maintain or pay for anything during periods in which no resources are being used.

The above means that cloud computing empowers its users to just get on with what they want to do. Today, nobody sits down to use a pencil. However, lots of people do still consciously sit down to use a computer. Cloud developments may, however, start to catalyse a mentality shift from tool-in-hand to task-at-hand computer application.

Due to the fact that cloud computing is charged on a usage basis, it has no fixed costs. A fixed cost is something that has to be paid regardless of the number of people who use something or a company's level of production. This compares

to a variable cost that will change according to output levels. For example, the annual cost of renting a factory is likely to be fixed. However, the cost of staffing a factory and of the raw materials it consumes will vary according to how much it produces.

Traditionally computing has involved substantial fixed costs, such as those costs incurred in the building and equipping of a data centre. However, because cloud computing is dynamically scalable and task-centric, for users such fixed costs disappear. All of the costs of cloud computing are therefore on a per-usage or variable basis. As demonstrated by the earlier example of AmazonEC2, processing power can already be purchased from the cloud by the hour.

The fact that cloud computing has only variable costs is of extreme importance for small companies. This is because small businesses have traditionally not had access to the sophisticated, customised types of business application available to larger organizations.

However, because they do not charge an initial fixed-cost outlay, cloud computing suppliers including Clarizen, Employease, Netsuite, Salesforce and Zoho are now levelling the software-access playing field by allowing companies of all sizes access to the latest types of business application. You can read more about cloud computing and small businesses here.

CHARACTERISTICS

In general, cloud computing customers do not own the physical infrastructure, instead avoiding capital expenditure by renting usage from a third-party provider. They consume resources as a service and pay only for resources that they use. Many cloud-computing offerings employ the utility computing model, which is analogous to how traditional utility services (such as electricity) are consumed, whereas others bill on a subscription basis. Sharing "perishable and intangible" computing power among multiple tenants can improve utilization rates, as servers are not unnecessarily left idle (which can reduce costs significantly while increasing the speed of application development). A side-effect of this approach is that overall computer usage rises dramatically, as customers do not have to engineer for peak load limits. In addition, "increased high-speed bandwidth" makes it possible to receive the same response times from centralized infrastructure at other sites.

ECONOMICS

Diagram showing economics of cloud computing versus traditional IT, including capital expenditure (CapEx) and operational expenditure (OpEx)

Cloud computing users can avoid capital expenditure (CapEx) on hardware, software, and services when they pay a provider only for what they use. Consumption is usually

billed on a utility (e.g., resources consumed, like electricity) or subscription (e.g., time-based, like a newspaper) basis with little or no upfront cost. A few cloud providers are now beginning to offer the service for a flat monthly fee as opposed to on a utility billing basis. Other benefits of this time sharing-style approach are low barriers to entry, shared infrastructure and costs, low management overhead, and immediate access to a broad range of applications. In general, users can terminate the contract at any time (thereby avoiding return on investment risk and uncertainty), and the services are often covered by service level agreements (SLAs) with financial penalties.

According to Nicholas Carr, the strategic importance of information technology is diminishing as it becomes standardized and less expensive. He argues that the cloud computing paradigm shift is similar to the displacement of electricity generators by electricity grids early in the 20th century.

Although companies might be able to save on upfront capital expenditures, they might not save much and might actually pay more for operating expenses. In situations where the capital expense would be relatively small, or where the organization has more flexibility in their capital budget than their operating budget, the cloud model might not make great fiscal sense. Other factors impacting the scale of any potential cost savings include the efficiency of a company's data centre

as compared to the cloud vendor's, the company's existing operating costs, the level of adoption of cloud computing, and the type of functionality being hosted in the cloud.

ARCHITECTURE

The majority of cloud computing infrastructure, as of 2009, consists of reliable services delivered through data centres and built on servers with different levels of virtualization technologies. The services are accessible anywhere that provides access to networking infrastructure. Clouds often appear as single points of access for all consumers' computing needs. Commercial offerings are generally expected to meet quality of service (QoS) requirements of customers and typically offer SLAs. Open standards are critical to the growth of cloud computing, and open source software has provided the foundation for many cloud computing implementations.

HISTORY

The Cloud *is a term that borrows from telephony. Up to the 1990s, data circuits (including those that carried Internet traffic) were hard-wired between destinations. Then, long-haul telephone companies began offering Virtual Private Network (VPN) service for data communications. Telephone companies were able to offer VPN-based services with the same guaranteed bandwidth as fixed circuits at a lower cost because they could switch traffic to balance utilization as they saw fit, thus utilizing their overall network bandwidth more effectively.*

As a result of this arrangement, it was impossible to determine in advance precisely which paths the traffic would be routed over. The term "telecom cloud" was used to describe this type of networking, and cloud computing is in concept somewhat similar.

The underlying concept of cloud computing dates back to 1960, when John McCarthy opined that "computation may someday be organized as a public utility"; indeed it shares characteristics with service bureaus that date back to the 1960s. In 1997, the first academic definition was provided by Ramnath K. Chellappa who called it a computing paradigm where the boundaries of computing will be determined by economic rationale rather than technical limits*. The term* cloud *had already come into commercial use in the early 1990s to refer to large Asynchronous Transfer Mode (ATM) networks.*

Loudcloud, founded in 1999 by Marc Andreessen, was one of the first to attempt to commercialize cloud computing with an Infrastructure as a Service model. By the turn of the 21st century, the term "cloud computing" began to appear more widely, although most of the focus at that time was limited to SaaS, called "ASP's" or Application Service Providers, under the terminology of the day.

In the early 2000s, Microsoft extended the concept of SaaS through the development of web services. IBM detailed these concepts in 2001 in the Autonomic Computing Manifesto, which described advanced automation techniques such as

self-monitoring, self-healing, self-configuring, and self-optimizing in the management of complex IT systems with heterogeneous storage, servers, applications, networks, security mechanisms, and other system elements that can be virtualized across an enterprise.

Amazon played a key role in the development of cloud computing by modernizing their data centres after the dot-com bubble, which, like most computer networks, were using as little as 10% of their capacity at any one time just to leave room for occasional spikes. Having found that the new cloud architecture resulted in significant internal efficiency improvements whereby small, fast-moving "two-pizza teams" could add new features faster and easier, Amazon started providing access to their systems through Amazon Web Services on a utility computing basis in 2005. This characterization of the genesis of Amazon Web Services has been characterized as an extreme over-simplification by a technical contributor to the Amazon Web Services project.

In 2007, Google, IBM, and a number of universities embarked on a large scale cloud computing research project. By mid-2008, Gartner saw an opportunity for cloud computing "to shape the relationship among consumers of IT services, those who use IT services and those who sell them", and observed that "[o]rganisations are switching from company-owned hardware and software assets to per-use service-based models" so that the "projected shift to cloud

computing… will result in dramatic growth in IT products in some areas and in significant reductions in other areas."

POLITICAL ISSUES

The Cloud *spans many borders and "may be the ultimate form of globalization." As such, it becomes subject to complex geopolitical issues, and providers are pressed to satisfy myriad regulatory environments in order to deliver service to a global market. This dates back to the early days of the Internet, when libertarian thinkers felt that "cyberspace was a distinct place calling for laws and legal institutions of its own".*

Despite efforts (such as US-EU Safe Harbor) to harmonize the legal environment, as of 2009, providers such as Amazon cater to major markets (typically the United States and the European Union) by deploying local infrastructure and allowing customers to select "availability zones." Nonetheless, concerns persist about security and privacy from individual through governmental levels (e.g., the USA PATRIOT Act, the use of national security letters, and the Electronic Communications Privacy Act's Stored Communications Act*).*

LEGAL ISSUES

In March 2007, Dell applied to trademark the term "cloud computing" (U.S. Trademark 77,139,082) in the United States. The "Notice of Allowance" the company received in July 2008 was cancelled in August, resulting in a formal rejection of the trademark application less than a week later.

In November 2007, the Free Software Foundation released the Affero General Public License, a version of GPLv3 intended to close a perceived legal loophole associated with free software designed to be run over a network. Founder and president, Richard Stallman has also warned that cloud computing "will force people to buy into locked, proprietary systems that will cost more and more over time".

KEY CHARACTERISTICS

- Agility improves with users able to rapidly and inexpensively re-provision technological infrastructure resources..
- Cost is claimed to be greatly reduced and capital expenditure is converted to operational expenditure. This ostensibly lowers barriers to entry, as infrastructure is typically provided by a third-party and does not need to be purchased for one-time or infrequent intensive computing tasks. Pricing on a utility computing basis is fine-grained with usage-based options and fewer IT skills are required for implementation (in-house).
- Device and location independence enable users to access systems using a web browser regardless of their location or what device they are using (e.g., PC, mobile). As infrastructure is off-site (typically provided by a third-party) and accessed via the Internet, users can connect from anywhere.

- Multi-tenancy enables sharing of resources and costs across a large pool of users thus allowing for:
- Centralization of infrastructure in locations with lower costs (such as real estate, electricity, etc.)
- Peak-load capacity increases (users need not engineer for highest possible load-levels)
- Utilization and efficiency improvements for systems that are often only 10–20% utilized.
- Reliability improves through the use of multiple redundant sites, which makes cloud computing suitable for business continuity and disaster recovery. Nonetheless, many major cloud computing services have suffered outages, and IT and business managers can at times do little when they are affected.
- Scalability via dynamic ("on-demand") provisioning of resources on a fine-grained, self-service basis near real-time, without users having to engineer for peak loads. Performance is monitored, and consistent and loosely-coupled architectures are constructed using web services as the system interface.
- Security typically improves due to centralization of data, increased security-focused resources, etc., but concerns can persist about loss of control over certain sensitive data, and the lack of security for stored kernels. Security is often as good as or better than under traditional systems, in part because providers are able

to devote resources to solving security issues that many customers cannot afford. Providers typically log accesses, but accessing the audit logs themselves can be difficult or impossible. Furthermore, the complexity of security is greatly increased when data is distributed over a wider area and/or number of devices.

- Sustainability comes about through improved resource utilization, more efficient systems, and carbon neutrality. Nonetheless, computers and associated infrastructure are major consumers of energy.

THE THREE REASONS TO CLOUD COMPUTE

Cloud vendors offering SaaS, PaaS and IaaS services now provide most individuals and organizations with a real alternative to running local computing resources in-house. There are also now three reasons why, within a decade, cloud computing is likely to be the only mainstream computing show in town.

Specifically, the key reasons that cloud computing will become inevitable are that:

- Cloud computing will be essential to remain competitive
- Cloud computing will be essential to be green, and
- Cloud computing will be essential for next-generation applications

The above drivers of the Cloud Computing Revolution are explained below and in the following video:

The Competitive Cloud

Cloud computing will become essential for firms that wish to remain competitive simply because it will cost them less to cloud compute than to run an in-house data centre. Already many cloud computing vendors claim that their customers can reap cost savings of up to eighty per cent. Because cloud computing is dynamically-scalable, task-centric and does not require substantial fixed-cost investments, firms that cloud compute are also likely to be more competitively agile than their competitors.

In his cloud computing manifesto *The Big Switch*, Nicholas Carr compares the growth of cloud computing to the development of the electricity network around a century ago. Before that time, businesses had to make their own power using internal generators. However, when a reliable electricity grid became available, companies were increasingly freed from having to generate their own energy in-house. The opportunity to plug-in to cheaper electricity on-tap from a national power grid also offered two competitive benefits. Firstly, it saved companies money. And secondly, it allowed companies to focus their resources on other aspects of their business.

In a striking parallel to the Big Switch from local to central electricity generation a century ago, today we are entering a

Cloud Computing Age in which businesses will be able to dispense with an internal data centre, and instead plug-in to cloud computing resources that will fuel their information processing requirements. For example, ThinDesk is now offering small and medium-sized companies in Canada the opportunity to move to a thin client/ cloud infrastructure which it claims will provide them with up to forty per cent cost savings, coupled with increased reliability and productivity. Carbon footprint and energy reduction savings of up to eighty per cent can potentially also be achieved with a ThinDesk solution.

The Green Cloud

Talking of energy reduction and carbon footprint savings, cloud computing is also inevitable because it is more green. Today, most internal company data centres run their servers at around thirty per cent capacity. In contrast, the servers in a large cloud data centre typically run at eighty per cent capacity or more. This means that less energy is wasted, with the carbon footprint of each unit of computing power being reduced. Cloud computing is also more environmentally friendly than traditional computing because it removes the need for most users to have high-power PCs and laptops. As discussed on the green computing page, lower-power computers based around processors such as Intel's Atom are perfectly sufficient to run cloud applications. Their use can also cut end-user energy bills and carbon footprints by

as much as eighty per cent. You can see the construction of such a green PC in this video.

The use of cloud collaboration and virtual meeting tools may also allow some people to work from home a little more and to make fewer business trips. As more and more people and companies adopt cloud computing, it may therefore help to take some cars off the road and some planes out of the sky.

The Next Generation Cloud

As well as being more cost-effective and more green, cloud computing will be increasingly essential for many next-generation computing developments, such as Big Data. In other words, cloud computing will be driven not only by a desire and necessity to do existing things in more effective ways, but by a demand to do entirely new things.

One of the defining characteristics of cloud computing is that it enables value to created via collaboration and data sharing. Local software and data inevitably constrain collaboration and the anytime, anyplace, anywhere use of information resources. As a consequence, we will be prevented from obtaining the benefits of new developments such as “crowdsouring” unless many of us cloud compute.

Crowdsourcing uses the Internet to help generate value from the activities of a great many people. Today, crowdsourcing mainly involves lots of people working together to tackle a problem that in the past would have been left to just one individual or a small team.

This had already lead to many open source developments where all of the involved intellectual property is created and shared online for mutual benefit. Already crowdsourced, open source products and services in use or under development include computer software, robots, 3D printers, prosthetic limbs and electric cars. You can find links to many of these developments in the Cloud Computing Directory. The rise of cloud computing will make it easier for individuals to consciously work together on crowdsourcing projects. However, the Cloud Computing Revolution will also be driven by the significant potential to crowdsource data from many of the things that people consume and the objects that they manipulate.

Increasingly many everyday objects — ranging from fridges to clothing — are being given their own Internet connection and are going online. As examined in my Explaining Web Squared video, using data from cameras and other sensors smart cloud computing applications are also starting to recognise and monitor objects that do not have their own Internet connection. Increasingly, this is allowing the development of some quite innovative cloud-centric applications, such as the Google Goggles visual search application. This allows items viewed by a smartphone camera to be identified and used as the basis for a web search.

Perhaps most notably at present when it comes to next generation applications, cloud computing developments are

enabling the rise of augmented reality. This is where real-time cloud data is overlaid on the camera feed of a smartphone or other mobile device. For example, users of the Layar augmented reality browser can now see other the photos and tweets of Twitter users in their area, or floating burgers showing them where they might like to eat. Links to some augmented reality pioneers are also included in the Cloud Computing Directory. Soon there will be so many cameras, microphones and other sensors online that a great many objects will start to cast a constant data shadow in the cloud. Whilst this may raise concerns, it will also allow us to reap crowdsourcing benefits similar to those of online social networking. For example, satellite navigation systems will be able to advise on routes based not only on internal maps, but also the position and predicted intent of all other vehicles on the road. However, this will only happen if a great deal of data is pooled and shared in the cloud rather than being held and processed on local computing devices.

Developments in artificial intelligence will also depend on crowdsourced cloud data. Programming a mobile phone or a robot to recognise everything in view is likely to remain very difficult if internal data and processing power have to be relied on. However, a phone or robot with access to cloud resources including video feeds from other nearby cameras will be in a far better position to usefully make sense of the world around it. What this means is that for computers to

be usefully smart they will require access to information from our immediate environment that can only be crowdsourced from the cloud. In turn, along with augmented reality, the development of artificial intelligence is likely to be a very strong driver for the mass adoption of cloud computing.

CLOUD AS A SERVICE TO CUSTOMER

The cloud computing that are evolving as a service in the cloud are being provided by big enterprises with a heavy investment with resource and technology which are accessed by others via the internet. The resources are accessed in this manner as a service – often on a subscription basis. The users of the services being offered often have very little knowledge of the technology being used. The users also have no control over the infrastructure that supports the technology they are using.

There are six different forms that have been consolidated so far to understand how the services are being provided to the customers:

SaaS and Types of Cloud Computing

This types of cloud computing delivers a single application through the browser to thousands of customers using a multitenant architecture. On the customer side, it means no upfront investment in servers or software licensing; on the provider side, with just one app to maintain, costs are low compared to conventional hosting. SaaS is also common for

HR apps and has even worked its way up the food chain to ERP, with players such as Workday. And some who could have predicted the sudden rise of SaaS desktop applications, such as Google Apps and Zoho Office.

Utility Computing

The idea is not new, but this form of cloud computing is getting new life from Amazon.com, Sun, IBM, and others who now offer storage and virtual servers that IT can access on demand. Early enterprise adopters mainly use utility computing for supplemental, non-mission-critical needs, but one day, they may replace parts of the datacenter. Other providers offer solutions that help IT create virtual datacenters from commodity servers, such as 3Tera's AppLogic and Cohesive Flexible Technologies Elastic Server on Demand. Liquid Computing's LiquidQ offers similar capabilities, enabling IT to stitch together memory, I/O, storage, and computational capacity as a virtualized resource pool available over the network.

Web Services in the Cloud Intimately Related to SaaS

Web service providers offer APIs that enable developers to exploit functionality over the Internet, rather than delivering full-blown applications. They range from providers offering discrete business services -- such as Strike Iron and Xignite — to the full range of APIs offered by Google Maps, ADP

payroll processing, the U.S. Postal Service, Bloomberg, and even conventional credit card processing services.

Platform as a Service 'One more SaaS Variation'

This type of cloud computing deliver development environments as a service. You build your own applications that run on the provider's infrastructure and are delivered to your users via the Internet from the provider's servers.

Like Legos, these services are constrained by the vendor's design and capabilities, so you don't get complete freedom, but you do get predictability and pre-integration. Prime examples include Coghead and the new Google App Engine. For extremely lightweight development, cloud-based abound, such as Yahoo Pipes or Dapper.net.

MSP (Managed Service Providers)

One of the oldest forms of cloud computing, a managed service is basically an application exposed to IT rather than to end-users, such as a virus scanning service for e-mail or an application monitoring service (which Mercury, among others, provides). Managed security services delivered by SecureWorks, IBM, and Verizon fall into this category, as do such cloud-based anti-spam services as Postini, recently acquired by Google. Other offerings include desktop management services, such as those offered by CenterBeam or Everdream.

Service Commerce Platforms

A hybrid of SaaS and MSP, this cloud computing service offers a service hub that users interact with. They're most common in trading environments, such as expense management systems that allow users to order travel or secretarial services from a common platform that then coordinates the service delivery and pricing within the specifications set by the user. Think of it as an automated service bureau. Well- known examples include Rearden Commerce and Ariba.

CLOUD SERVICE ARCHITECTURE

The new way of the world for most web software development is the assembly of applications from cloud-based APIs. Developers are saving loads of time by pulling in various cloud services and focusing their attention on the novel business logic of their solutions.

Hundreds of new APIs are sprouting up monthly, as tracked on ProgrammableWeb. And as the very existence of a web site dedicated to tracking APIs implies, application assembly has fundamentally changed software development. Yet as monumental as that is, it's the tip of the iceberg when it comes to realizing the potential of a cloud service-based architecture. Distributing the logic within an application across numerous centralized cloud services will enable more automated, intelligent applications. And the timing couldn't be better.

Interestingly, this separation is similar to how software defined networking (SDN) splits the components of a network into a "control" plane and "data" plane to enable more automated, intelligent networks. Just as an SDN controller can analyse data from the various nodes in a network and automatically change the behaviour of a network to improve performance or security, a cloud service can analyse data across all the applications it powers and make changes to the applications to improve their behaviour or performance. In fact, a deeper look at SDN offers important clues about the benefits a decoupled architecture can provide for cloud applications. The control plane in SDN contains the intelligence responsible for defining the behaviour of the network (the "rules") while the data plane moves packets within the network according to these rules (the "processing"). Separating the "rules" of how network packets should be processed from the actual processing allows the feedback loop of measure, analyse, and modify that is critical for enabling automated, intelligent networks.

Cloud application architecture does bear some resemblance to SDN, where a centralized intelligent element (a cloud service) often plays the role of analysing data from across numerous end nodes (in this case application instances) and modifying the behaviour of those end nodes. Most startups haven't exploited this because they have been able to deliver so much value to customers by simply offering

a service or application with a modern web-based delivery model. That in and of itself adds enough of a value proposition to get off the ground.

However, as they mature, more and more of these companies will deploy intelligence and automation in their service. This will enable them to use the SDN qualities of their architecture to analyse data from all the applications powered by these services and then modify the application to improve performance or change the behaviour of these applications.

One area already leveraging the intelligence enabled by a disaggregated architecture to provide a leap forward in value delivered to customers is security, where large-scale correlation analyses, machine learning, and other big-data techniques are utilized to determine if a threat is present. If a threat is detected, the data is shared across the network, alerting people of the threat. This is an intelligent cloud service at its best.

When I meet with startups offering cloud-based services, the discussion often leads to the tremendous value of the data they are gathering from all the applications that implement their service. But in order to benefit from this data, machine learning heuristics and other advanced analysis techniques will need to be applied so that action can be taken in real time to modify these applications. Centralized, intelligent cloud services will improve the

functionality or performance of applications automatically based on the data they are seeing — not unlike Amazon making recommendations for you based on past purchases or ad-tech companies optimize retargeting. While a majority of cloud services startups today are creating a lot of value simply by delivering their service as an easy-to-use API, intelligence and automation is where the next large opportunity lies.

CLOUD COMPUTING SERVICES

Cloud computing is among the leading disruptive trends and strategic technologies of this decade that offers a new IT delivery model.

Several recent new developments have made security, risk management and governance in the Cloud more manageable, and hence opened up new options for enterprises that want to leverage the Cloud.

Enterprises stand to derive a host of benefits from a smart and consistent Cloud strategy. iGATE can help you along your transformation journey to Cloud enabled IT for addressing the challenges faced by you today, namely:

- Reducing Capital Expenditure (CapEx)
- Adjusting to fluctuations in demand for computing resources based on dynamic business conditions
- Rapidly setting up technology stacks for IT projects to be delivered in critical timelines

- Maximizing utilization of computing resources, reducing their management complexity, flexibly reusing computing resources across widely ranging needs of multiple projects
- Catering to high-end expensive IT infrastructure needs for which investment in CapEx is not a practical approach for you. These include:
  - Large scale compute resources for short-term experimentation or testing
  - High performance computing resources for compute-intensive processing
  - Distributed fault-tolerant environment for high availability and business continuity.

Cloud provides several features that help to overcome these challenges:

- It offers a variety of pre-installed, dynamically scalable computing and storage infrastructure, platform and application software as a service
- It is accessed on-demand from distributed locations using Open standards based automated, self-service interfaces
- It is charged on subscription or usage basis at low rates that benefit from economies of scale
- It is hosted on an optimized, fault-tolerant, highly scalable, secure infrastructure
- Its management is automated and its complexity is hidden from users.

How iGATE Can Help You

We deliver end-to-end Cloud services from consulting, architecture, design, implementation to management-monitoring to help you throughout the lifecycle of your Cloud adoption initiative. We can extend your IT to a Cloud in public domain to satisfy your needs of Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS), Application Software-as-a-Service (SaaS), or help to transform your data centre into a private cloud for internal use, or implement and manage a hybrid cloud for you. We also deliver industry facing Cloud solutions.

Our services enable you to gain several benefits:

- CapEx Reduction and Dynamic Scalability: We take a consulting-led approach. Through our Cloud Acceleration Programme (CAP), we help you to make complex decisions such as which of your applications to transition to Cloud, Cloud type and vendor platform to be chosen, and provide a smooth roadmap for transitioning to Cloud infrastructure that flexibly scales up or down as your demand varies and reduces CapEx and procurement cycle time.
- Business Focus and Speed to remain Competitive: Our Development/Testing on Cloud (DToC) and Testing-as-a-Service (TaaS) offerings allow you to flexibly use required pre-configured development/ testing

technology stacks on demand from Cloud, and get you quickly started, reducing time-to-market, allowing focus on business problems enabling business innovation in order to remain competitive.

- Efficiency, Agility, Availability and Accountability: Our Private Cloud offering enables transforming your data centre or server rooms into Cloud under your control for enterprise-wide use. It also enables optimal and flexible utilization of fault-tolerant IT resources providing efficiency, agility, availability; and allows intelligence and accountability to be provided for shared IT resource consumption by business users.
- Risk Reduction: Our Unified Cloud Management and Monitoring offering enables visibility, control of IT operation seamlessly across Cloud based and on-premise IT to reduce risk.

iGATE Advantage

Leveraging our deep expertise, proven methods, frameworks, and multi-vendor alliances, we accelerate, and lower the risk of, transitioning your IT to the best suited Cloud to derive the above benefits.

- *Expert Team for Superior Solution:* We have a team of trained-certified, experienced Cloud Computing experts who continuously explore evolving Cloud technology, develop deep skills through hands-on work in our Cloud

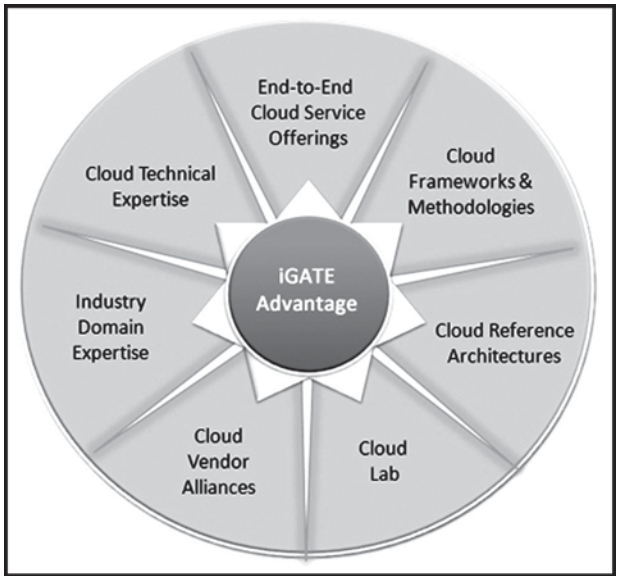
Lab and by networking with allied major Cloud ecosystem vendors. Our internal training programmes and Cloud Computing Community of Practice enable rapid generation of Cloud Computing skills.

- *Proven Methods for Accelerated Delivery and Lowered Risk:* We perform structured delivery of services using accelerators tested in our Lab that capture our expertise and best practices that combined with our global delivery model reduce your risks and time in transitioning to Cloud. Our integrated technology and operations (iTOPS) led delivery model ensures assured business outcomes.
- *Multi-vendor Approach for Best-of-Breed Technology-based Solution:* We take a multi-vendor approach to recommend a solution based on best-of-breed technologies most suited to your problem and environment.

Track Record for Client Success:

- Messaging and collaboration Proof-of-Concept, Private cloud on Microsoft Platforms for a global foods manufacturer serving 28000 restaurants worldwide
- Proof-of-Concept on AWS public cloud for HPC design automation with multiple tool compatibility for one of the global top three electronic design automation companies
- Migration of operations management and reporting system to Azure platform, SSO system for 1.6 million+

transient user population for one of the world's largest quick service restaurant chains



- Cloud-readiness assessment on six core applications for one of the largest North American power utilities company
- Consulting and migration to hybrid environment including private and public cloud platforms for a leading American real-estate owner-operator with over $1 billion revenues.

3

Security Concerns of Cloud Computing

Cloud computing frees both individuals and organizations from the cost and hassle of installing, maintaining and constantly upgrading software applications on their desktops and in their data centres. It also allows companies to focus on their core competencies, rather than investing in centralized computing facilities that have to be maintained and upgraded and that may not be utilized at an optimum capacity.

This said, the many critics of cloud computing point to the fact that users become totally reliant on a high quality Internet connection. Cloud computing also creates a reliance on external suppliers which may also raise potential business continuity, data protection and security risks.

The aforementioned concerns do have to be carefully considered. This said, our reliance on the Internet is now so great that even if most of our applications are locally installed then the disruption caused by an Internet outage is already highly significant. In a sense, a reliable Internet connection has now become as requisite a utility service for business and personal activities as a constantly available phone network and electricity supply. All that cloud computing is therefore doing is making us even more explicitly aware of this. Like it or not, we have already mortgaged our souls to the Internet.

Many of the fears associated with cloud computing security and data protection are also largely perceptual. Granted, using cloud computing makes individuals and organizations dependent on both their cloud vendors and the integrity of their Internet connection.

However, anybody or any company that makes web searches or sends e-mail — let alone who makes online purchases or does their banking online — is already trusting both the quality of available online security, as well as the security of those they purchase from, not to mention the integrity of their Internet Service Provider. We should also not forget that storing data locally can also create data protection problems of its own. As events in the United Kingdom public sector have demonstrated time and time again in recent years, it is apparently very easy to loose

thousands and even millions of highly sensitive personal records on lost or stolen laptops, USB keys or CDs. One of the key things highlighted by cloud computing developments is the need for secure personal computing devices. In January 2010 the Google e-mail accounts of Chinese dissidents were allegedly hacked by the Chinese Government. What most reports of this global news story suggested was that the security of Google's cloud data centres was compromised. However, this was not the case. Rather, a security flaw in the Internet Explorer web browser was used to plant spyware on the PCs of individual Chinese dissidents. This malicious software then e-mailed the Google usernames and passwords of the dissidents back to the hackers, who used it to "legitimately" access their e-mail accounts. The end result may have been the same as hacking a Google data centre. However, it is very important to appreciate that — as is usually the case when problems occur — security was compromised at the user and not the vendor end of the cloud computing chain.

Anybody using cloud computing services needs to take appropriate measures for ensuring safe web access. These include setting a strong password, ensuring antivirus, antispyware and firewall software are installed, and ensuring that their operating system and web browser(s) are always updated with the latest security patches. All users also need to be educated not to open suspicious e-mails that may

contain and install malware. More information on appropriate online security measures can be found on the security page or from GetSafeOnline.org.

In time, cloud computing developments themselves may significantly increase the security of individual PCs and other computing devices. For example, Google's recently released Chrome OSoperating system relies entirely on Google Docs and other cloud based applications and storage. In turn, as Google advertise, this will make Chrome-based computers far more secure as it is simply not be possible for applications to be locally installed. For years we have heralded computers as programmable devices on which we can install local software. However, in security terms, the best cloud access devices will be non-programmable computers on which neither users or hackers can install any type of local programme code.

CLOUD COMPUTING: A MINDSET SHIFT

Somebody recently informed me that cloud computing sounds like a return to the age of centralized computing when dumb terminals were totally dependent on a centralized mainframe. To some extent there is more than a little truth in this. However, there are also important differences to the previous mainframe era. For a start, cloud computing is levelling the playing field by bringing the potential benefits of remote and highly professional computing resources to all sizes of business. Any company and indeed anybody can

now connect to software or hardware as an online utility, with fewer companies having to invest in a large-scale computing infrastructure.

As the above hopefully highlights, cloud computing significantly differs from the previous dumb-terminal/ mainframe era in that users do not become reliant on a single, specific, centralized computing resource that their organization has to invest in and maintain.

Rather, they become reliant on a far looser external web of resources which they will always to an extent be free to "mash" as their needs dictate. In other words, a competent use of the cloud by either an individual or an organization ought always to involve a loose rather than a tight-coupling of computing resources. Reliance on individual SaaS, PaaS or IaaS vendors ought therefore to be minimised in the same way that nobody obtaining power from a national electricity grid is dependent on the continuous functioning of an single, specific power plant.

In comparison to previous computing eras and paradigms, cloud computing potentially also offers many advantages. Not least, over the coming decade fewer individual users will be tied to a particular device when they want to access their data and applications. Collaborative working will also become far more common, with the benefits of using an online word processor or spreadsheet far outweighing the potential drawbacks in most circumstances. Cloud computing is also

likely to be far more environmentally friendly than many currently mainstream computing practices.

It will also enable the next generation of computing developments like augmented reality, visual search and artificial intelligence In tandem with Web 2.0, cloud computing is already changing the landscape of the computing industry. Google's CEO Eric Schmidt has stated that cloud computing is bigger than the PC revolution and with that sentiment I tend to agree. Following the launch of new its cloud web site at microsoft.com/cloud in February 2010, even Microsoft has "gone cloud". As I said at the beginning of this article, the only real choice for most of us now is whether we want to be part of the cloud computing steamroller or the traditional computing road. And I would suggest that plumping for the option that will result in getting flattened has to be second best!

To embrace cloud computing requires a new mindset. For too long computing has been about hoarding and putting up barriers — both technologically and culturally — rather than sharing and opening up communication and collaboration. For business, the computing industry, and the planet more generally, cloud computing ought therefore to be welcomed as a breath of fresh air that will become part of the solution rather than part of the problem.

The world faces too many global challenges for us all to continue to compute by ourselves.

CLOUD COMPUTING SECURITY

Cloud computing security (sometimes referred to simply as "cloud security") is an evolving sub-domain of computer security, network security, and, more broadly, information security. It refers to a broad set of policies, technologies, and controls deployed to protect data, applications, and the associated infrastructure of cloud computing. Cloud security is not to be confused with security software offerings that are "cloud-based" (a.k.a. security-as-a-service).

SECURITY ISSUES ASSOCIATED WITH THE CLOUD

Organizations use the Cloud in a variety of different service models (SaaS, PaaS, IaaS) and deployment models (Private, Public, Hybrid). There are a number of security issues/ concerns associated with cloud computing but these issues fall into two broad categories: Security issues faced by cloud providers (organizations providing software-,platform-, or infrastructure-as-a-service via the cloud) and security issues faced by their customers. In most cases, the provider must ensure that their infrastructure is secure and that their clients' data and applications are protected while the customer must ensure that the provider has taken the proper security measures to protect their information.

The extensive use of virtualization in implementing cloud infrastructure brings unique security concerns for

customers or tenants of a public cloud service. Virtualization alters the relationship between the OS and underlying hardware - be it computing, storage or even networking. This introduces an additional layer - virtualization - that itself must be properly configured, managed and secured. Specific concerns include the potential to compromise the virtualization software, or "hypervisor". While these concerns are largely theoretical, they do exist. For example, a breach in the administrator workstation with the management software of the virtualization software can cause the whole datacenter to go down or be reconfigured to an attacker's liking.

CLOUD SECURITY CONTROLS

Cloud security architecture is effective only if the correct defensive implementations are in place. An efficient cloud security architecture should recognize the issues that will arise with security management. The security management addresses these issues with security controls. These controls are put in place to safeguard any weaknesses in the system and reduce the effect of an attack. While there are many types of controls behind a cloud security architecture, they can usually be found in one of the following categories:

Deterrent Controls

These controls are set in place to prevent any purposeful attack on a cloud system. Much like a warning sign on a

fence or a property, these controls do not reduce the actual vulnerability of a system.

Preventative Controls

These controls upgrade the strength of the system by managing the vulnerabilities. The preventative control will safeguard vulnerabilities of the system. If an attack were to occur, the preventative controls are in place to cover the attack and reduce the damage and violation to the system's security.

Corrective Controls

Corrective controls are used to reduce the effect of an attack. Unlike the preventative controls, the corrective controls take action as an attack is occurring.

Detective Controls

Detective controls are used to detect any attacks that may be occurring to the system. In the event of an attack, the detective control will signal the preventative or corrective controls to address the issue.

DIMENSIONS OF CLOUD SECURITY

Correct security controls should be implemented according to asset, threat, and vulnerability risk assessment matrices.

While cloud security concerns can be grouped into any number of dimensions (Gartner names seven while the Cloud Security Alliance identifies fourteen areas of concern) these

dimensions have been aggregated into three general areas: Security and Privacy, Compliance, and Legal or Contractual Issues.

SECURITY AND PRIVACY

Identity Management

Every enterprise will have its own identity management system to control access to information and computing resources. Cloud providers either integrate the customer's identity management system into their own infrastructure, using federation or SSO technology, or provide an identity management solution of their own.

Physical and Personnel Security

Providers ensure that physical machines are adequately secure and that access to these machines as well as all relevant customer data is not only restricted but that access is documented.

Availability

Cloud providers assure customers that they will have regular and predictable access to their data and applications.

Application Security

Cloud providers ensure that applications available as a service via the cloud are secure by implementing testing and acceptance procedures for outsourced or packaged

application code. It also requires application security measures be in place in the production environment.

Privacy

Finally, providers ensure that all critical data (credit card numbers, for example) are masked and that only authorized users have access to data in its entirety. Moreover, digital identities and credentials must be protected as should any data that the provider collects or produces about customer activity in the cloud.

Legal Issues

In addition, providers and customers must consider legal issues, such as Contracts and E-Discovery, and the related laws, which may vary by country.

Compliance

Numerous regulations pertain to the storage and use of data, including Payment Card Industry Data Security Standard (PCI DSS), the Health Insurance Portability and Accountability Act (HIPAA), the Sarbanes-Oxley Act, among others.

Many of these regulations require regular reporting and audit trails. Cloud providers must enable their customers to comply appropriately with these regulations.

Business Continuity and Data Recovery

Cloud providers have business continuity and data

recovery plans in place to ensure that service can be maintained in case of a disaster or an emergency and that any data loss will be recovered. These plans are shared with and reviewed by their customers.

Logs and Audit Trails

In addition to producing logs and audit trails, cloud providers work with their customers to ensure that these logs and audit trails are properly secured, maintained for as long as the customer requires, and are accessible for the purposes of forensic investigation.

Unique Compliance Requirements

In addition to the requirements to which customers are subject, the data centres maintained by cloud providers may also be subject to compliance requirements. Using a cloud service provider (CSP) can lead to additional security concerns around data jurisdiction since customer or tenant data may not remain on the same system, or in the same data centre or even within the same provider's cloud.

LEGAL AND CONTRACTUAL ISSUES

Aside from the security and compliance issues enumerated above, cloud providers and their customers will negotiate terms around liability (stipulating how incidents involving data loss or compromise will be resolved, for example), intellectual property, and end-of-service (when data and

applications are ultimately returned to the customer).

Public Records

Legal issues may also include records-keeping requirements in the public sector, where many agencies are required by law to retain and make available electronic records in a specific fashion. This may be determined by legislation, or law may require agencies to conform to the rules and practices set by a records-keeping agency. Public agencies using cloud computing and storage must take these concerns into the account.

SECURING CLOUD COMPUTING

Cloud computing is attractive, seductive and perhaps irresistible. The benefits are compelling, particularly the pay-as-you-go model that has been likened to buying electricity (or, if you, prefer, buying your drinks by the glass rather than the bottle).

There's a powerful business case for buying computational power, disk storage, collaboration, application development resources, CRM, on demand. Rather than buying more servers and disks or expanding or deploying expensive infrastructure and programmes, cloud computing is flexible and scalable. It can meet short-term initiatives and requirements and deal with peaks and valleys in business cycles. But where does security fit into all this? Security

analysts and practitioners generally say proceed, but proceed with caution. All the risks to sensitive corporate data associated with outsourcing apply to cloud computing, and then some. Enforcing security policy and meeting compliance requirements are tough enough when you deal with third parties and their known or unknown subcontractors, especially on a global scale. Add the blurry characteristics of the cloud and the entry of non-traditional vendors into the technology market, and some red flags go up. In an IDC survey of 244 IT executives/CIOs published last fall, 75 percent of the respondents cited security as a significant or very significant challenge with cloud computing. Compare that with 63 per cent cited for the next two concerns—performance and availability.

THE BIGGEST CLOUD COMPUTING SECURITY RISK

The past couple of years have been tough for those defending the security of cloud computing and those trying to establish secure cloud infrastructures for themselves. For the most part, there have been DDOS attacks or defacements designed to embarrass or punish site owners.

However, even considering only web sites or services from which hackers actually took over accounts, stole data or money, or planted malware to help steal data or money from

others, the list of security failures is long and distinguished: Google, LinkedIn, Twitter, Hotmail, Global Payments (credit-card clearinghouse for Visa, MasterCard and others), Federal Express, Zappos, a host of local bank and police agencies, and the China Software Developer Network (which, all by itself, lost personal information on 6 million users to a single hacker named Zeng).

MORE INSIGHTS

True, some of those victims offer services with access too restricted and services too limited to be considered "cloud." Except for the potential booty (money, data or notoriety), cloud and non-cloud services look pretty much the same to criminals trying to crack them open.

During 2010, only 4 million user accounts were compromised by hackers; in 2011 hackers penetrated 174 million accounts (thanks, Anonymous), according to the Data Breach Investigations Report published by Verizon in March. If anything, 2012 is going to be even worse, according to anti-virus/security vendor Kaspersky Labs.

All that hackery did generate tons of publicity, plenty of vocal outrage from victims, diluvian volumes of frettage from pundits and solemn warnings from security companies thrilled to find themselves centre stage in a drama they've been narrating all but unheard for years. Despite all that noise, it's odd to realize that the incident with the greatest

potential to cause a change in attitude among users and IT is the hack of a single reporter's backup account on Apple's consumer-oriented iCloud storage service.

It wasn't even a cool hack. It barely qualified as social engineering.

Due only to the weak identity verification of Apple and Amazon, *Wired* reporter Mat Honan suffered what he described as his own digital death and dissolution. Not only did hackers get into his e-mail and iCloud accounts, but they also used the data-sync connections among Honan's all-Apple suite of personal devices to wipe out all his backup data in the cloud. They deleted everything from his MacBook and iPhone, leaving him with no easily recoverable copy of any of his data, most of which was valuable to only him.

"In the space of one hour, my entire digital life was destroyed. First my Google account was taken over, then deleted," he explained in a piece for *Wired* explaining "How Apple and Amazon Security Flaws Led to My Epic Hacking." "Next my Twitter account was compromised, and used as a platform to broadcast racist and homophobic messages. And worst of all, my AppleID account was broken into, and my hackers used it to remotely erase all of the data on my iPhone, iPad, and MacBook." Details of the attack are telling, but have been left out to annoy the readership into even more outrage at the poor security/customer service of Apple and Amazon. (They're actually in "Anatomy of a Successful

Personal Hack," if you haven't read them elsewhere, follow the link to be appalled by them.)

Honan acknowledges that some random hacker couldn't have rolled up and eaten his whole digital life without help from the victim himself.

"Had I been regularly backing up the data on my MacBook, I wouldn't have had to worry about losing more than a year's worth of photos, covering the entire lifespan of my daughter, or documents and e-mails that I had stored in no other location," he wrote in his *Wired* piece. "Those security lapses are my fault, and I deeply, deeply regret them. But what happened to me exposes vital security flaws in several customer service systems, most notably Apple's and Amazon's."

MORE INSIGHTS

Unsurprisingly, Honan recommends against daisy chaining all your data-heavy devices to the same control account. More usefully, publishing his story prompted both Apple and Amazon to revamp security, at least to the extent of eliminating the specific gaps Honan's hackers exploited. Neither company required frequent password changes, secure passwords or two-factor authentication for anything. It wouldn't do any good, anyway. Most end users—and most IT people, for that matter—aren't interested in going to the amount of trouble it would take to keep from being digitally

gutted by the same hack that eviscerated Honan. No matter how many warnings they get, an astonishing number still use simplistic passwords (*123456* is a favourite) and the same passwords for everything (easier to remember), and link as many accounts as possible (to avoid multiple logins).

In fact, single sign-on—the secure version of the same practice—has been the goal of dozens of major enterprise networking products. No one likes having to remember passwords or log in separately to every application or web site. Apple, Google, Facebook, Twitter and most other consumer-oriented services count on that to get customers to agree to link their social networking accounts—a major marketing benefit to the vendors that offers users almost nothing good.

Last week, Apple co-founder Steve Wozniak got roasted for suggesting that relying on the cloud too heavily would result in "horrendous" consequences for end users. "I really worry about everything going to the cloud. I think it's going to be horrendous. I think there are going to be a lot of horrible problems in the next five years," he said. "With the cloud, you don't own anything. You already signed it away." Woz did get plenty of support from cloud haters, digital paranoids and from experts who realize the cloud is just as dangerous and filled with security flaws as any other Web service, data centre or other computerized structure invented by and configured for the use of demonstrably imperfect humans.

CLOUD SERVICE PROVIDER

IT organizations must address a set of obvious technology issues to support this model, including integration, identity management and security. But to compete as a service provider, they're finding they also need unfamiliar business skills and capabilities.

Among them:

- *Offer Design:* It's not enough to just build a private cloud and let it loose. Just as Amazon and Rackspace do, enterprise IT must identify the starting points, bundles and configurations it will offer internally, including CPU, memory, storage, network and other services and components. Post- launch, the IT organization needs to respond to user needs and usage behaviour and modify its offers accordingly, just as a third-party service provider would do. This is easier said than done.
- *Pricing:* With cloud service catalogues, many CIOs are implementing chargeback models, whereby business users receive a bill at the end of the month for all IT services, whether they're internal or external. To compete with vendors, IT organization must evolve the pricing of their private cloud services beyond "cost-plus" models. That pricing must reflect market dynamics and provide incentives to keep volumes in house.

- *PR And Marketing:* Bear with me. This isn't a stretch. While CIOs don't need to retain PR firms for internal communications, they do need to market and evangelize their services internally. It's not just about promotion. It's about understanding the needs and pains of your customers, whether they're business executives or developers, and effectively communicating your value prop.
- *Demand Management:* If your organization offers users and developers a true choice, it might initially be difficult to predict demand for internal services. This is particularly true of private cloud IaaS and PaaS services. A variety of factors will drive usage, including features, capabilities and chargeback pricing. Demand forecasting and management are critical to avoiding bad capacity decisions. Although capacity planning has always been a core IT skill, effectively forecasting demand in a competitive environment is a capability few marketing organizations even possess.

If this all sounds suspiciously close to product marketing and management, it is. IT departments must effectively become cloud service providers themselves if they're going to compete as such. While many organizations have considered the technology implications of this choice, few have fully considered the required business skills.

COMPUTING POWER IN THE NETWORK

Take advantage of cloud technology to keep your business one step ahead.

In an age where innovation and technology play a vital role in both the strategic direction and financial performance of any global organisation, it is critical to keep up to date with new technologies to maintain competitive advantage. We can provide your organisation with reliable, high-end cloud solutions which support a wide range of service requirements and are delivered in the most relevant way for your business (public, private or hybrid). Our global model means you have a consistent solution in all your operational locations and our local focus means you will receive a personalised service that delivers the next wave of cloud solutions around smart devices for your business.

Cloud computing marks a new era in the evolution and delivery of solutions and enables your business to take advantage of the benefits of new technology immediately. Our solutions reside in our network, and are accessible seamlessly, utilising verification protocols, through a simple internet connection (Wi-Fi, 3G, Cable) from any device. This means you have instant access, wherever and whenever you want, and are always in control.

Our solutions are modular, scalable and flexible. Importantly they are physically secure and redundant, with contingency solutions operated and maintained

24x7, so your information is safely saved in the network.

Our global Cloud Computing solutions provide a unique integrated environment designed to meet your business challenges globally, with local customer care and support delivering end-to-end levels of service. We manage the complete solution (the platform, the network and the infrastructure) while you maintain control and management of your cloud services. Furthermore, we provide access to on-demand experts so you don't need to hire or train new support teams within your business.

How do our Global Cloud Computing Services Help?

Removing entry barriers to technology:

- Ease of use and access - you can access and use the service simply by logging onto the application on the internet.
- End-to-end (e2e) support - we provide e2e support for cloud computing solutions, network communications and management.
- OPEX model - you only pay for what you use, enabling you to replace CAPEX with a lower OPEX (capital expenditure and operating expense).
- Self-service - you're in control and can request what you want, whenever you want, wherever you want, 24 hours a day, 365 days a year.

- Multi-device - you can access your applications from any device, including mobile, PC, lap-top and iPad.
- Multi-location - you can access from wherever you are and the service is the same irrespective of location.
- Automatic updates - service updates are transparent and automatic. Collaborative tools in real time - every employee in the organisation can update their information immediately and every other worker can have real time access to that information.

What are our Cloud Computing Services?

Expert technology combined with excellent support. We have global coverage for cloud services supported by local integration and local customer support.

In addition we have an infrastructure for global support and maintenance which is redundant and secure. This operates 24x7 using a ¨Follow-the-Sun¨ model and has management centres in both Europe and the Americas.

The core elements of our global Cloud Computing solution include a global e2e management model based on:

- A supervision platform (Network Operation Centres) - allows advanced diagnostics, real time alerts and monitoring, simple set up and reporting.
- Service Level Agreements and premium support - specially designed for the unique demands of business-critical cloud computing solutions to ensure availability and security.

- 24x7 centralised support - ensures the highest quality service that is always available. Our professional staff are committed to ensuring the fastest issue detection and resolution.

Our Virtual Data Centre provides Infrastructure as a Service (IaaS) as an integral solution for storage, security, communications, monitoring, administration and backup.

This solution delivers the following benefits:

- Online self-provisioning and management - You will have full control of your Virtual Data Centre. You make a request via the user portal which will be available in minutes with monitoring tools provided anytime and anywhere.
- Choose a service consumption model - You only pay for what you use. You can choose a flat rate to control your expenses or simply contract some virtual machines to support your business needs to improve efficiency and avoid unnecessary expenditure.
- Commercial flexibility - The modular format of our solutions means you can adapt to the changing needs of your business at any time.

Other services available soon will include Virtual PC, online storage, a collaboration portal to share content and an application marketplace.

4

Cloud Computing and Information Security

As cloud computing options proliferate for individuals and large organizations, it is increasingly important for both to make informed choices about appropriate use of cloud services, taking into consideration both benefits and risks. To assist in making this assessment, faculty and staff can see at a glance whether or not it is permissible to maintain a specific data type in a U-M or external vendor cloud service by viewing the Sensitive Data Guide to IT Services.

CLOUD COMPUTING

Cloud computing has several distinct characteristics that distinguish it from a traditionally-hosted computing environment:

- Users often have on-demand access to scalable information technology capabilities and services that are provided through internet-based technologies.
- These resources run on an external or third-party service provider's system. This is in contrast to traditional systems, which run on locally-hosted servers. Unlike traditional systems which are under the user's personal control or institutional control, cloud computing services are fully managed by the provider.
- Typically, many unaffiliated and unconnected users share the service provider's infrastructure.
- Using cloud services reduces the need to carry data on removable media because of network access anywhere, anytime.

Cloud services, sometimes called "software as a service" (SaaS), "infrastructure as a service" (IaaS), or "platform as a service" (PaaS), facilitate rapid deployment of applications and infrastructure without the cost and complexity of purchasing, managing, and maintaining the underlying hardware and software.

Organizations and institutions are increasingly driven to cloud computing as a way to increase functionality, lower cost, and enhance convenience to users by making the services and resources available anywhere there is an internet connection. With cloud computing, users have readily

available a suite of applications, features, and infrastructure that would normally require significant investment if provided in the traditional in-house computing environment.

U-M AND THE CLOUD

There are different ways in which cloud computing is being introduced to U-M students, faculty, staff, and researchers. Individuals across campus routinely access cloud applications or services on their smartphone or laptop. Faculty are increasingly using cloud computing applications as class or laboratory tools to supplement or even replace campus-provided resources. U-M researchers work frequently with other researchers across the globe and share data in the cloud.

As part of the NextGen Michigan initiatives, the university is implementing a full service environment and shared internal cloud by migrating from current servers to new virtual servers. The most significant of these new services are M+Box, M+Google, MiDatabase, and MiServer:

- M+Box provides a storage solution for U-M students, faculty, and staff to store and share files online. It's part of a two-year agreement between Internet2, U-M, and several other peer institutions.
- M+Google provides a platform for collaboration, including shared documents, as well as e-mail and calendaring.

- MiDatabase is a U-M hosted cloud service, managed by ITS, consisting of a virtual server and a managed database.
- MiServer is a virtual server environment managed by ITS which allows users to focus on managing applications instead of the operating system (physical servers).

Proper Use of Cloud Computing Services at U-M

Cloud computing should not be used for information that is private, personal, or sensitive, unless there is a contractual agreement between U-M and the service provider that protects the confidentiality of the information and data. A contractual agreement is a formal contract that would typically be reviewed by the Office of General Counsel.

U-M engages in research, teaching, and business activities that encompass a variety of regulated sensitive data. There are important institutional and individual responsibilities for compliance to ensure that such data are properly protected. Faculty, researchers, and staff (including student employees and students conducting research) need to assess whether federal and state laws, contractual obligations, and/or grant restrictions limit the ability to store institutional or research data in cloud computing services.

Sensitive and Regulated Data: Permitted and Restricted Uses establishes mandatory expectations for complying with

statutory and regulatory requirements related to protecting sensitive regulated data.

The standard references the following Standard Practice Guide Policies:

- SPG 601.12 – Institutional Data Resource Management Policy, which includes the definition of sensitive data.
- SPG 601.07 – Proper Use of Information Resources, Information Technology, and Networks at U-M

Please refer to the Sensitive Data Guide to IT Services to determine where storage of sensitive data is permitted in the U-Mcomputing environment and among current U-M cloud computing service providers.

Security and Privacy

The integrity, availability, and maintenance of appropriate confidentiality of institutional data is critical to U-M's reputation and to minimizing institutional exposure to legal and compliance risks. Much of the challenge in deciding whether cloud computing is desirable and appropriate for an institution like U-M is determining whether a prospective cloud computing vendor has adequate physical, technical, and administrative safeguards as good as or better than the local on-campus systems.

While cloud computing services have numerous potential benefits, there are also potentially significant privacy and security considerations that should be accounted for before

collecting, processing, sharing, or storing institutional or personal data in the cloud.

Consequently, institutions should conduct careful risk assessment prior to adoption of any cloud computing service.

Specific risks and challenges to consider include:

- Vendor transparency and inadequate or unclear service level agreement
- Privacy and confidentiality of personal, sensitive, or regulated data and information
- Legal and regulatory compliance
- Cyber security and support for incident forensics
- Records preservation, access, and management
- Service availability and reliability

UNDERSTAND THE RISKS OF CLOUD COMPUTING

Cloud service users need to be vigilant in understanding the risks of data breaches in this new environment.

"At the heart of cloud infrastructure is this idea of multi-tenancy and decoupling between specific hardware resources and applications," explains Datamonitor senior analyst Vuk Trifkoviæ. "In the jungle of multi-tenant data, you need to trust the cloud provider that your information will not be exposed."

For their part, companies need to be vigilant, for instance about how passwords are assigned, protected and changed.

Cloud service providers typically work with numbers of third parties, and customers are advised to gain information about those companies which could potentially access their data.

IDC's Bradshaw says an important measure of security often overlooked by companies is how much downtime a cloud service provider experiences. He recommends that companies ask to see service providers' reliability reports to determine whether these meet the requirements of the business. Exception monitoring systems is another important area which companies should ask their service providers about, he adds.

London-based financial transaction specialists SmartStream Technologies made its foray into the cloud services space last month with a new SaaS product aimed at providing smaller banks and other financial institutions with a cheap means of reconciling transactions. Product manager Darryl Twiggs says that the service has attracted a good deal of interest amongst small to mid-tier banks, but that some top tier players are also being attracted by the potential cost savings. An important consideration for cloud service customers, especially those responsible for highly sensitive data, Twiggs says, is to find out about the hosting company used by the provider and if possible seek an independent audit of their security status.

"Customers we engage with haven't been as stringent as we thought they would have been with this".

How Cloud Hosting Companies have Approached Security

As with most SaaS offerings, the applications forming SmartClear's offering are constantly being tweaked and revised, a fact which raises more security issues for customers. Companies need to know, for instance, whether a software change might actually alter its security settings. "For every update we review the security requirements for every user in the system," Twiggs says.

One of the world's largest technology companies, Google, has invested a lot of money into the cloud space, where it recognises that having a reputation for security is a key determinant of success. "Security is built into the DNA of our products," says a company spokesperson. "Google practices a defence-in-depth security strategy, by architecting security into our people, process and technologies". However, according to Datamonitor's Trifkoviæ, the cloud is still very much a new frontier with very little in the way of specific standards for security or data privacy. In many ways he says that cloud computing is in a similar position to where the recording industry found itself when it was trying to combat peer-to-peer file sharing with copyright laws created in the age of analogue.

"In terms of legislation, at the moment there's nothing that grabs my attention that is specifically built for cloud computing," he says. "As is frequently the case with disruptive technologies, the law lags behind the technology development

for cloud computing." What's more, many are concerned that cloud computing remains at such an embryonic stage that the imposition of strict standards could do more harm than good. IBM, Cisco, SAP, EMC and several other leading technology companies announced in late March that they had created an 'Open Cloud Manifesto' calling for more consistent security and monitoring of cloud services. But the fact that neither Amazon.com, Google nor Salesforce.com agreed to take part suggests that broad industry consensus may be some way off. Microsoft also abstained, charging that IBM was forcing its agenda.

"Standards by definition are restrictive. Consequently, people are questioning whether cloud computing can benefit from standardisation at this stage of market development." says Trifkoviæ. "There is a slight reluctance on the part of cloud providers to create standards before the market landscape is fully formed." Until it is there are nevertheless a handful of existing web standards which companies in the cloud should know about. Chief among these is ISO27001, which is designed to provide the foundations for third party audit, and implements OECD principles governing security of information and network systems. The SAS70 auditing standard is also used by cloud service providers.

Local Law and Jurisdiction where Data is Held

Possibly even more pressing an issue than standards in this new frontier is the emerging question of jurisdiction.

Data that might be secure in one country may not be secure in another. In many cases though, users of cloud services don't know where their information is held. Currently in the process of trying to harmonise the data laws of its member states, the EU favours very strict protection of privacy, while in America laws such as the US Patriot Act invest government and other agencies with virtually limitless powers to access information including that belonging to companies.

UK-based electronics distributor ACAL is using NetSuite OneWorld for its CRM. Simon Rush, IT manager at ACAL, has needed to ensure that ACAL had immediate access to all of its data should its contract with NetSuite be terminated for any reason, so that the information could be quickly relocated. Part of this included knowing in which jurisdiction the data is held. "We had to make sure that, as a company, our data was correctly and legally held."

European concerns about US privacy laws led to creation of the US Safe Harbour Privacy Principles, which are intended to provide European companies with a degree of insulation from US laws. James Blake from e-mail management SaaS provider Mimecast suspects that these powers are being abused. "Counter terrorism legislation is increasingly being used to gain access to data for other reasons," he warns.

Mimecast provides a comprehensive e-mail management service in the cloud for over 25,000 customers, including 40 per cent of the top legal firms in the UK. Customers benefit

from advanced encryption that only they are able to decode, ensuring that Mimecast acts only as the custodian, rather than the controller of the data, offering companies concerned about privacy another layer of protection. Mimecast also gives customers the option of having their data stored in different jurisdictions.

For John Tyreman, IT manager for outsourced business services provider Liberata, flexibility over jurisdiction was a key factor in his choosing Mimecast to help the company meet its obligations to store and manage e-mails from 2500 or so staff spread across 20 countries. The company is one of the UK's leading outsourcing providers for the Public Sector, Life Pensions and Investments and Corporate Pensions leading. "Storing our data in the US would have been a major concern," Tyreman says.

5

Practice in Cloud Computing

As already mentioned, cloud computing can encompass activities such as the use of social networking sites and other forms of interpersonal computing as examined on the Web 2.0 page. However, most of the time cloud computing is concerned with accessing online software applications, data storage and processing power. The fundamental, practical building blocks of cloud computing are therefore what are known as software as a service (SaaS), platform as a service (PaaS) and infrastructure as a service (IaaS). As discussed in my Three Ways to Cloud Compute video (embedded below), anybody wanting to understand cloud computing really has to get to grips with what SaaS, PaaS and IaaS are all about.

SaaS, PaaS and IaaS all involve a cloud vendor supplying servers on which their customers can store data and run

applications. However, there are differences in the level of control provided to the customer, as well as the type of cloud hardware on which a customer's cloud applications are run.

In very simple terms, when businesses opt for SaaS they can only run those applications that their cloud supplier has on offer. When they opt for PaaS they can create their own applications but only in a manner determined by their cloud supplier. And when they opt for IaaS they can run any applications they please on cloud hardware of their own choice. OK, that may at this stage be as clear as well stirred pond water! So let's now work through it again in more detail.

Software as a Service (SaaS)

Software as a service is where computer applications are accessed over the Internet rather than being installed on a local computing device or in a local data centre. So, for example, people may use an online word processor like Google Docs, an online database application like Zoho Creator, an online photo editor like Pixlr, or an online invoicing application such as Zoho Invoice. Many SaaS applications are free to use, at least initially. You can find links to a great many in the Cloud Computing Directory. And you can find a video introduction to some great free SaaS applications here.

SaaS can provide its users with many benefits. These include the general cloud computing advantages of dynamic scalability and any device independence, as well as the benefit of being able to use an application without incurring fixed

costs. Many SaaS applications are also collaborative. This allows multiple users to share documents and even to work on them at the same time. For example, in the Google Docs spreadsheet different users can work on different cells simultaneously. The cells different users are working on are locked-off and highlighted in different colours. A real-time chat window can also be opened up alongside the spreadsheet to further enhance collaboration. For more information on collaborative working using Google Docs, you can watch the now classic video Google Docs in Plain English.

Taking collaboration further still, the outputs of some SaaS applications can be embedded in other web pages as web service gadgets. For example, a Google Docs or Zoho Sheet chart can be mashed into another web site. There it will automatically update when the data in the online spreadsheet that is generating it is changed. SaaS applications are also constantly updated, which can free users from the "upgrade hell" of a major traditional software package revision.

The disadvantage of SaaS is that it is basically a take-it-or-leave-it form of cloud computing. This means that businesses and individuals who require direct access to cloud computing hardware on which they can run their own applications cannot use SaaS. Rather, they need to cloud compute at the platform or infrastructure level using either platform as a service (PaaS) or infrastructure as a service (IaaS).

Platform as a Service (PaaS)

A platform is a software environment used to develop and run applications. For example, Microsoft Word is an application that runs on the Microsoft Windows platform. When people choose to cloud compute using platform as a service or 'PaaS', they obtain access to an online platform provided by a cloud computing vendor. They can then use this platform to develop and deliver their own online (SaaS) applications.

Applications developed using PaaS may be used privately by just one or a few users within a particular company. However, they can also be offered free or for-a-fee to anybody on the web. This means that if you have a great idea for a new online application then you can use PaaS to turn it into a reality.

Several cloud suppliers now offer PaaS tools. Most notably these include Google App Engine, Microsoft Windows Azure, and Force.com. All such offerings effectively provide their customers with a box of cloud computing Lego. New applications are then constructed from the plastic bricks on offer. With Force.com, some applications can even be built using a simple drag-and-drop interface. Relatively non-technical people can therefore create new online applications very quickly. Indeed, Force.com claim that their "simplified programming model and cloud-based environment mean [customers] can build and run applications five times faster,

at about half the cost of traditional software platforms". Google App Engine and Force.com also allow an initial application to be created for free!

Whilst PaaS is great in many situations, its users do need to be mindful of the involved flexibility verses power trade-off. What this means is that whilst PaaS makes it relatively easy to create new online applications, users are nevertheless constrained by the particular programming languages and tools provided by their PaaS supplier.

In other words, PaaS vendors have total control over which Lego bricks they allow their customers to build with. Whilst this ensures that applications built using the tools on offer will always function correctly, it is nevertheless restrictive. It is for this reason that many companies and some individuals choose to cloud compute at the infrastructure level.

Infrastructure as a Service (IaaS)

Infrastructure as a service or "IaaS" is where a cloud supplier provides online infrastructure on which their customers can store data and develop and run whatever applications they please. IaaS therefore allows companies to move their existing programmes and data into the cloud and to close down their own local servers and data centres. Whilst computing applications run on platforms, platforms in turn run on computing infrastructure. So, for example, whilst the Microsoft Word application runs on the Microsoft Windows platform, in

turn the Microsoft Windows platform runs on the infrastructure of an IBM-compatible PC.

The fundamental building block of cloud computing infrastructure is the server. Cloud computing servers are basically computers on which online applications can be run and data can be stored. When provided by an IaaS vendor, cloud servers can also be real or virtual.

Real or "dedicated" servers are individual circuit boards – known as blades – mounted within equipment racks in a data centre. In contrast virtual servers – also known as "virtual server instances" – are software-controlled slices of real, physical servers. Virtual servers are created by a process called virtualization that allows many users to share the processing power of one physical server.

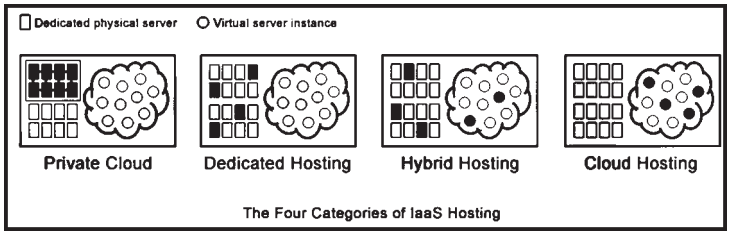
Dedicated physical servers and virtual server instances can perform exactly the same functions.

However, there are some differences between them. For a start, virtual server instances are cheaper to supply as each does not require its own piece of physical hardware in a cloud data centre. On the other hand, virtual server instances are sometimes seen as less secure by those who do not want to share server hardware with other customers.

For this reason, four categories of IaaS are available. These are most commonly known as "private clouds", "dedicated hosting", "hybrid hosting" and "cloud hosting".

IaaS Categories

The four categories of IaaS are represented in the figure below. In each case a large rectangle outlines a cloud data centre. In this data centre there are then a number of dedicated physical servers (shown as small rectangles), together with a number of virtual server instances (shown as circles within a cloud). Dedicated physical servers or virtual server instances in the figure are shown in solid rather than outline when they are part of a particular IaaS category.



Under the first IaaS category of a private cloud (or more fully a vendor managed private cloud), a customer rents a number of co-located servers in part of a data centre. This means that their cloud hardware is as separate as possible from that of other users. Private clouds are therefore considered the most secure form of IaaS.

However, a private cloud cannot be dynamically scaled and is the most expensive form of IaaS as a block of servers is permanently dedicated to one customer. (As an aside, it should be noted that an increasing number of IT companies are also starting to use the term "private cloud" to refer to the building of a cloud computing infrastructure — or

"internal cloud" — within a company's own data centre. Such a development is not really cloud computing at all, and may be regarded as a last-ditch attempt to maintain the status quo. Under any sensible definition, a "private cloud" has to be a cloud computing arrangement where the hardware concerned is owned and housed in a vendor's shared data centre. OK, rant over, and back to the three remaining IaaS categories!).

In the second IaaS category of dedicated hosting, a customer rents dedicated physical servers on demand from anywhere within a data centre. Whilst this means that the hardware they use is mixed-in with that of other customers, in this IaaS category once again customers do not share the particular servers they use with anybody else. As well as being less costly than a private cloud, dedicated hosting can therefore be dynamically scaled. This means that the customer is able to increase or decrease the number of servers they are both using and paying for on a daily or even hourly basis. Under the third IaaS option of hybrid hosting, a customer rents on demand a mix of dedicated physical servers and as well as some less expensive virtual server instances. For example, a company may run all of its applications on dedicated physical servers, but store its data on virtual server instances. Or a business may rent virtual service instances by the hour to cope with occasional peak processing demands. Once again, the whole offering is dynamically

scalable, with both dedicated and virtual servers able to be added or taken away as required.

Finally, in the last IaaS category of cloud hosting, a customer rents as many or as few virtual server instances as they require on demand. This means that customers share all of the servers they use with other customers. Some companies subsequently see this as too risky. However, cloud hosting is also the lowest-cost and by far the most technically and environmentally efficient form of IaaS. This is because cloud hosting allows an IaaS provider to run all of their physical servers in use to capacity and to close down those not required.

IaaS Providers

Many companies now offer IaaS services. For example, as already noted, Amazon has a product range called Amazon Web Services or "AWS". This falls under the fourth IaaS category of cloud hosting, with Amazon offering the rental of virtual server instances.

At the heart of AWS is Amazon Elastic Compute Cloud or "EC2". This allows customers to run either new or existing applications in Amazon's data centres. EC2 is described as "elastic" because customers can increase or decrease the infrastructure capacity they are using within minutes.

EC2 users can purchase and activate one, hundreds or even thousands of virtual server instances simultaneously. They do this by setting up Amazon Machine Images or "AMIs"

that contain all of the applications, data and configuration settings that their virtual servers will need. AMIs can be created from scratch, or chosen from a range of pre-configured templates. AMIs can even be pre-loaded with licensed software from vendors including IBM. Another key component of AWS is the Amazon Simple Storage Service or "S3". This enables customers to store data online in so-termed "buckets". As Amazon explain "S3 provides a simple web interface that can be used to store and retrieve any amount of data, at any time, from anywhere on the web. It gives any developer access to the same highly scalable, reliable, fast, inexpensive data storage infrastructure that Amazon uses to run its own global network of web sites". Another IaaS provider is Rackspace, which provides private clouds, dedicated hosting, and cloud hosting services. The latter include Rackspace Cloud Servers (as a competitor to Amazon EC2) and Rackspace Cloud Files (as a competitor to Amazon's S3). IaaS services can also be purchased fromGoGrid, which offers cloud hosting, hybrid hosting and dedicated hosting solutions. A longer list of IaaS providers is included in my Cloud Computing Directory.

MOTIVATION: OUR DATA-DRIVEN WORLD

Advances in digital sensors, communications, computation, and storage have created huge collections of data, capturing information of value to business, science,

government, and society. For example, search engine companies such as Google, Yahoo!, and Microsoft have created an entirely new business by capturing the information freely available on the World Wide Web and providing it to people in useful ways.

These companies collect trillions of bytes of data every day and continually add new services such as satellite images, driving directions, and image retrieval. The societal benefits of these services are immeasurable, having transformed how people find and make use of information on a daily basis.

Just as search engines have transformed how we access information, other forms of *big-data computing* can and will transform the activities of companies, scientific researchers, medical practitioners, and our nation's defence and intelligence operations.

Some examples include:

- Wal-Mart recently contracted with Hewlett Packard to construct a data warehouse capable of storing 4 *petabytes* (4000 trillion bytes) of data, representing every single purchase recorded by their point-of-sale terminals (around 267 million transactions per day) at their 6000 stores worldwide. By applying *machine learning* to this data, they can detect patterns indicating the effectiveness of their pricing strategies and advertising campaigns, and better manage their inventory and supply chains.

- Many scientific disciplines have become data-driven. For example, a modern telescope is really just a very large digital camera. The proposed Large Synoptic Survey Telescope (LSST) will scan the sky from a mountaintop in Chile, recording 30 trillion bytes of image data every day – a data volume equal to *two entire Sloan Digital Sky Surveys daily*! Astronomers will apply massive computing power to this data to probe the origins of our universe. The Large Hadron Collider (LHC), a particle accelerator that will revolutionize our understanding of the workings of the Universe, will generate 60 terabytes of data per day – 15 petabytes (15 million gigabytes) annually. Similar *eScience* projects are proposed or underway in a wide variety of other disciplines, from biology to environmental science to oceanography. These projects generate such enormous data sets that automated analysis is required. Additionally, it becomes impractical to replicate copies at the sites of individual research groups, so investigators pool their resources to construct a large data centre that can run the analysis programmes for all of the affiliated scientists.
- Modern medicine collects huge amounts of information about patients through imaging technology (CAT scans, MRI), genetic analysis (DNA microarrays), and other forms of diagnostic

equipment. By applying *data mining* to data sets for large numbers of patients, medical researchers are gaining fundamental insights into the genetic and environmental causes of diseases, and creating more effective means of diagnosis.

- Understanding the environment requires collecting and analysing data from thousands of sensors monitoring air and water quality and meteorological conditions, another example of eScience. These measurements can then be used to guide simulations of climate and groundwater models to create reliable methods to predict the effects of long-term trends, such as increased CO_2 emissions and the use of chemical fertilizers.
- Our intelligence agencies are being overwhelmed by the vast amounts of data being collected through satellite imagery, signal intercepts, and even from publicly available sources such as the Internet and news media. Finding and evaluating possible threats from this data requires "connecting the dots" between multiple sources, *e.g.*, to automatically match the voice in an intercepted cell phone call with one in a video posted on a terrorist web site.
- The collection of all documents on the World Wide Web (several hundred trillion bytes of text) is proving to be a corpus that can be mined and processed in

many different ways. For example, language translation programmes can be guided by statistical language models generated by analysing billions of documents in the source and target languages, as well as multilingual documents, such as the minutes of the United Nations. Specialized web crawlers scan for documents at different reading levels to aid English-language education for first graders to adults. A conceptual network of noun-verb associations has been constructed based on word combinations found in web documents to guide a research project at Carnegie Mellon University in which fMRI images are used to detect how human brains store information.

These are but a small sample of the ways that all facets of commerce, science, society, and national security are being transformed by the availability of large amounts of data and the means to extract new forms of understanding from this data.

BIG-DATA TECHNOLOGY: SENSE, COLLECT, STORE, AND ANALYSE

The rising importance of big-data computing stems from advances in many different technologies:

- *Sensors:* Digital data are being generated by many different sources, including digital imagers (telescopes, video cameras, MRI machines), chemical and biological sensors (microarrays, environmental monitors), and

even the millions of individuals and organizations generating web pages.

- *Computer networks:* Data from the many different sources can be collected into massive data sets via localized sensor networks, as well as the Internet.
- *Data storage:* Advances in magnetic disk technology have dramatically decreased the cost of storing data. For example, a one-terabyte disk drive, holding one trillion bytes of data, costs around $100. As a reference, it is estimated that if all of the text in all of the books in the Library of Congress could be converted to digital form, it would add up to only around 20 terabytes.
- *Cluster computer systems:* A new form of computer systems, consisting of thousands of "nodes," each having several processors and disks, connected by high-speed local-area networks, has become the chosen hardware configuration for data-intensive computing systems. These clusters provide both the storage capacity for large data sets, and the computing power to organize the data, to analyse it, and to respond to queries about the data from remote users. Compared with traditional high-performance computing (*e.g.*, supercomputers), where the focus is on maximizing the raw computing power of a system, cluster computers are designed to maximize the

reliability and efficiency with which they can manage and analyse very large data sets. The "trick" is in the software algorithms – cluster computer systems are composed of huge numbers of cheap commodity hardware parts, with scalability, reliability, and programmability achieved by new software paradigms.

- *Cloud computing facilities:* The rise of large data centres and cluster computers has created a new business model, where businesses and individuals can *rent* storage and computing capacity, rather than making the large capital investments needed to construct and provision large-scale computer installations. For example, Amazon Web Services (AWS) provides both network-accessible storage priced by the gigabyte-month and computing cycles priced by the CPU-hour. Just as few organizations operate their own power plants, we can foresee an era where data storage and computing become utilities that are ubiquitously available.
- *Data analysis algorithms:* The enormous volumes of data require automated or semi-automated analysis – techniques to detect patterns, identify anomalies, and extract knowledge. Again, the "trick" is in the software algorithms - new forms of computation, combining statistical analysis, optimization, and artificial intelligence, are able to construct statistical

models from large collections of data and to infer how the system should respond to new data. For example, Netflix uses machine learning in its recommendation system, predicting the interests of a customer by comparing her movie viewing history to a statistical model generated from the collective viewing habits of millions of other customers.

Technology and Submission Challenges

Much of the technology required for big-data computing is developing at a satisfactory rate due to market forces and technological evolution. For example, disk drive capacity is increasing and prices are dropping due to the ongoing progress of magnetic storage technology and the large economies of scale provided by both personal computers and large data centres. Other aspects require more focused attention, including:

- *High-speed networking:* Although one terabyte can be stored on disk for just $100, transferring that much data requires an hour or more within a cluster and roughly a day over a typical "high-speed" Internet connection. (Curiously, the most practical method for transferring bulk data from one site to another is to ship a disk drive via Federal Express.) These bandwidth limitations increase the challenge of making efficient use of the computing and storage resources in a cluster. They also limit the ability to

link geographically dispersed clusters and to transfer data between a cluster and an end user. This disparity between the amount of data that is practical to store, vs. the amount that is practical to communicate will continue to increase. We need a "Moore's Law" technology for networking, where declining costs for networking infrastructure combine with increasing bandwidth.

- *Cluster computer programming:* Programming large-scale, distributed computer systems is a longstanding challenge that becomes essential to process very large data sets in reasonable amounts of time. The software must distribute the data and computation across the nodes in a cluster, and detect and remediate the inevitable hardware and software errors that occur in systems of this scale. Major innovations have been made in methods to organize and programme such systems, including the MapReduce programming framework introduced by Google. Much more powerful and general techniques must be developed to fully realize the power of big-data computing across multiple domains.
- *Extending the reach of cloud computing:* Although Amazon is making good money with AWS, technological limitations, especially communication bandwidth, make AWS unsuitable for tasks that

require extensive computation over large amounts of data. In addition, the bandwidth limitations of getting data in and out of a cloud facility incur considerable time and expense. In an ideal world, the cloud systems should be geographically dispersed to reduce their vulnerability due to earthquakes and other catastrophes. But, this requires much greater levels of interoperability and data mobility. The OpenCirrus project is pointed in this direction, setting up an international testbed to allow experiments on interlinked cluster systems. On the administrative side, organizations must adjust to a new costing model. For example, government contracts to universities do not charge overhead for capital costs (*e.g.*, buying a large machine) but they do for operating costs (*e.g.*, renting from AWS). Over time, we can envision an entire ecology of cloud facilities, some providing generic computing capabilities and others targeted towards specific services or holding specialized data sets.

- *Machine learning and other data analysis techniques:* As a scientific discipline, machine learning is still in its early stages of development. Many algorithms do not scale beyond data sets of a few million elements or cannot tolerate the statistical noise and gaps found in real-world data. Further research is required to

develop algorithms that apply in real-world situations and on data sets of trillions of elements. The automated or semi-automated analysis of enormous volumes of data lies at the heart of big-data computing for all application domains.

- *Widespread deployment:* Until recently, the main innovators in this domain have been companies with Internet-enabled businesses, such as search engines, online retailers, and social networking sites. Only now are technologists in other organizations (including universities) becoming familiar with the capabilities and tools. Although many organizations are collecting large amounts of data, only a handful are making full use of the insights that this data can provide. We expect "big-data science" – often referred to as eScience – to be pervasive, with far broader reach and impact even than previous-generation computational science.
- *Security and privacy:* Data sets consisting of so much, possibly sensitive data, and the tools to extract and make use of this information give rise to many possibilities for unauthorized access and use. Much of our preservation of privacy in society relies on current inefficiencies. For example, people are monitored by video cameras in many locations – ATMs, convenience stores, airport security lines, and urban

intersections. Once these sources are networked together, and sophisticated computing technology makes it possible to correlate and analyse these data streams, the prospect for abuse becomes significant. In addition, cloud facilities become a cost-effective platform for malicious agents, *e.g.*, to launch a botnet or to apply massive parallelism to break a cryptosystem. Along with developing this technology to enable useful capabilities, we must create safeguards to prevent abuse.

Leadership

This is an area where industry has been in the lead, especially the Internet-enabled service companies. These companies are investing billions of dollars in computing infrastructure that dwarf the world's largest traditional supercomputing installations. The main innovators in the configuration and programming of cluster computing systems have been at Google, Yahoo!, and Amazon. Other companies, from retailers to financial services, are taking notice of the business advantages and the operating efficiencies these companies are finding.

University researchers have been relatively late to this game, due to a combination of lack of access to large-scale cluster computing facilities and to a lack of appreciation for the new insights that can be gained by scaling up to terabyte-scale data sets. This situation is rapidly changing through

access to facilities and training, and due to the successes of their research counterparts in industry. Google, IBM, Yahoo!, and Amazon have provided access to some of their computing resources for students and researchers, using a cloud computing model. This has been enough to whet appetites but not nearly enough to satisfy the potential needs for widespread application of data-intensive computing. Many large-scale scientific projects are formulating plans for how they will manage and provide computing capacity for their collected data. Spirited debate is taking place between proponents of this new approach to data management and computing, with those pursuing more traditional approaches such as database technology and supercomputing. A lack of sufficient research funding is the major obstacle to getting greater involvement among university researchers.

Unfortunately, leadership from government agencies in this area has been mixed at best. The NSF has embraced data-intensive computing enthusiastically, with several new funding initiatives both within the CISE Directorate and with ties to other fields of science. However, their recent budgets have been so constrained that these programmes have not been able to scale up to their full potential. The Cyber-enabled Discovery and Innovation programme is scheduled to receive $26M for FY09. This is a step in the right direction, but further funding growth is required. Both the DoD and the DoE *should* be deeply involved in the development and deployment of

big-data computing, since it will be of direct benefit to many of their missions. Sadly, neither of these agencies has been a driving force for innovations in computing technology in recent times. Both are making heavy investments in traditional high-performance computing infrastructure and approaches, but very little in new eScience facilities and technologies.

CLOUD COMPUTING ENVIRONMENT

The National Institute of Standards and Technology (NIST), "Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (*e.g.*, networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model promotes availability and is composed of five essential characteristics, three service models, and four deployment models."

Recently the Federal Government has released the Federal Risk and Authorization Management Programme (FedRAMP) to account for the unique security requirements surrounding cloud computing. FedRAMP consists of a subset of NIST 800-53 security controls targeted towards cloud provider and customer security requirements.

As agencies look to reduce costs and improve reliability of business operations, cloud computing may offer promise as

an alternative to traditional data centre models. By utilizing the following cloud service models, agencies may be able to reduce hardware and personnel costs by eliminating redundant operations and consolidating resources. Cloud services offered by third party providers are often tailored to provide agencies with very precise environments to meet their operating needs. An agency's cloud implementation is a combination of a service model and a deployment model. NIST SP 800-145 outlines the possible service models that may be employed during a cloud implementation:

- Cloud Software as a Service (SaaS). The capability provided to the consumer is to use the provider's applications running on a cloud infrastructure. The applications are accessible from various client devices through a thin client interface such as a web browser (*e.g.*, web-based e-mail). The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited user-specific application configuration settings.
- Cloud Platform as a Service (PaaS). The capability provided to the consumer is to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages and tools supported by the provider. The consumer

does not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage, but has control over the deployed applications and possibly application hosting environment configurations.

- Cloud Infrastructure as a Service (IaaS). The capability provided to the consumer is to provision processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications. The consumer does not manage or control the underlying cloud infrastructure but has control over operating systems; storage, deployed applications, and possibly limited control of select networking components (*e.g.*, host firewalls).

Organizations have several choices for deploying a cloud computing model, as defined by NIST in SP 800-145:

- *Private cloud:* The cloud infrastructure is operated solely for an organization. It may be managed by the organization or a third party and may exist on premise or off premise.
- *Community cloud:* The cloud infrastructure is shared by several organizations and supports a specific community that has shared concerns (*e.g.*, mission, security requirements, policy, and compliance considerations). It may be managed by the

organizations or a third party and may exist on premise or off premise.

- *Public cloud:* The cloud infrastructure is made available to the general public or a large industry group and is owned by an organization selling cloud services.
- *Hybrid cloud:* The cloud infrastructure is a composition of two or more clouds (private, community, or public) that remain unique entities but are bound together by standardized or proprietary technology that enables data and application portability (*e.g.*, cloud bursting for load balancing between clouds).

Based on NIST guidance, industry best practices, and the Internal Revenue Service (IRS) Publication 1075, this memo provides agencies guidance for securing FTI in a cloud environment. These preliminary requirements are subject to change, based on updated standards or guidance. Agencies and their cloud providers should also review the requirements of FedRAMP and ensure overall compliance with these guidelines. While cloud computing offers many potential benefits, it is not without risk. The primary security concerns with cloud computing are:

- Data is not stored in an agency-managed data centre,
- The agency must rely on the vendor's security controls for protection, and

- Data from multiple customers are potentially commingled in the cloud environment.

Limiting access to authorized individuals becomes a much greater challenge with the increased availability of data in the cloud, and agencies may have greater difficulties to identify FTI when segregated or commingled in the cloud environment. Agencies that utilize a public cloud model should have increased oversight and governance over the security controls implemented by their cloud vendor. Monitoring and addressing security issues that arise with FTI in a cloud environment remain in the purview of the agency.

MANDATORY REQUIREMENTS FOR FTI IN A CLOUD COMPUTING ENVIRONMENT (CCE)

To utilize a cloud computing model to receive, transmit, store, or process FTI, the agency must be in compliance with all Publication 1075 requirements. The following mandatory requirements are in effect for introducing FTI to a CCE:

- *Notification Requirement:* The agency must notify the IRS Office of Safeguards at least 45 days prior to transmitting FTI into a cloud environment.
- *Data Isolation:* Software, data, and services that receive, transmit, process, or store FTI must be isolated within the cloud environment so that other cloud customers sharing physical or virtual space cannot access other customer data or applications.

- *Service Level Agreements (SLA):* The agency must establish security policies and procedures based on IRS Publication 1075 for how FTI is stored, handled, and accessed inside the cloud through a legally binding contract or Service Level Agreement (SLA) with their third party cloud provider.
- *Data Encryption in Transit:* FTI must be encrypted in transit within the cloud environment. All mechanisms used to encrypt FTI must be FIPS 140-2 compliant, and operate utilizing the FIPS 140-2 compliant module. This requirement must be included in the SLA.
- *Data Encryption at Rest:* FTI must be encrypted while at rest in the cloud. All mechanisms used to encrypt FTI must be FIPS 140-2 compliant, and operate utilizing the FIPS 140-2 compliant module. This requirement must be included in the SLA.
- *Persistence of Data in Relieved Assets:* Storage devices where FTI has resided must be securely sanitized and/or destroyed using methods acceptable by National Security Agency/Central Security Service (NSA/CSS). This requirement must be included in the SLA.
- *Risk Assessment:* The agency must conduct an annual assessment of the security controls in place on all information systems used for receiving, processing, storing and transmitting FTI. For the annual assessment

immediately prior to implementation of the cloud environment and each annual risk assessment (or update to an existing risk assessment) thereafter, the agency must include the cloud environment. The IRS Office of Safeguards will evaluate the risk assessment as part of the notification requirement in #1.

- *Security Control Implementation:* Customer defined security controls must be identified, documented and implemented. The customer defined security controls, as implemented, must comply with Publication 1075 requirements. These requirements are explained in detail in the sections below.

Notification

To utilize a cloud environment that receives, processes, stores or transmits FTI, the agency must meet the following mandatory notification requirements:

- If the agency's approved Safeguard Procedures Report (SPR) is less than six years old and reflects the agency's current process, procedures and systems, the agency must submit the Cloud Computing Notification, which will serve as an addendum to their SPR.
- If the agency's SPR is more than six years old or does not reflect the agency's current process, procedures and systems, the agency must submit a new SPR and the Cloud Computing Notification.

Before the SPR has been updated with the information from the Cloud Computing Notification Requirements, the IRS strongly recommends that a state agency planning on implementing a virtual environment contact the Office of Safeguards at SafeguardReports@irs.gov to schedule a conference call to discuss the details of the planned cloud computing implementation.

Data Isolation

One of the most common compliance issues with FTI is data location. Use of an agency-owned computing centre allows the agency to structure its computing environment and to know in detail where FTI is stored and what safeguards are used to protect the data. In contrast, a characteristic of many cloud computing services is that detailed information about the location of an organization's data is unavailable or not disclosed to the service subscriber. This makes it difficult to ascertain whether sufficient safeguards are in place and whether legal and regulatory compliance requirements are being met. IRS Publication 1075, section 5.3 recommends separating FTI from other information to the maximum extent possible. Organizing data in this manner will reduce the likelihood of unauthorized data access and disclosure. If complete separation is not possible, the agency must label FTI down to the data element level. Labelling must occur prior to introducing the data to the cloud and the data must be tracked accordingly through audit trails captured for

operating systems, databases and applications that receive, store, process or transmit FTI. The agency must be able to verify with the cloud provider, at all times, where the FTI has travelled in the cloud and where it currently resides.

IRS Publication 1075, section 9.3, *Audit and Accountability,* states audit logs must enable tracking activities taking place on the system. IRS Publication 1075 Exhibit 9, *System Audit Management Guidelines*, contains requirements for creating audit-related processes at both the application and system levels. Within the application, auditing must be enabled to the extent necessary to capture access, modification, deletion and movement of FTI by each unique user. This auditing requirement also applies to data tables or databases embedded in or residing outside of the application.

SERVICE LEVEL AGREEMENTS AND CONTRACTS

While the agency may not have direct control over FTI at all times, they ultimately maintain accountability while it is in the cloud, and the ownership rights over the data must be firmly established in the service contract to enable a basis for trust. A Service Level Agreement (SLA) is a mechanism to mitigate security risk that comes with the agency's lack of visibility and control in a cloud environment. It is important that agencies enter into SLAs with cloud providers that clearly identify Publication 1075 security control requirements and determine who has responsibility (provider, customer) for

their implementation. At a minimum, SLAs with cloud providers must include:

- IRS Publication 1075, Exhibit 7 contract language
- Identification of computer security requirements the cloud provider must meet. IRS Publication 1075, section 9, *Computer System Security* provides the security control requirements to include in agreements with third party cloud providers.
- Identification of requirements for cloud provider personnel who have access to FTI. All cloud provider personnel with FTI access must have a justifiable need for that access and submit to a background investigation.
- Identification of requirement that FTI may not be accessed by contractors located "offshore", outside of the United States or its territories. Further, FTI may not be received, stored, processed or disposed via information technology systems located off-shore.
- Identification of requirements for incident response to ensure cloud providers follow the incident notification procedures required by IRS Publication 1075. In the event of an unauthorized disclosure or data breach, the cloud provider and agency must report incident information to the appropriate Agent-in-charge, TIGTA, and the IRS Office of Safeguards within 24 hours according to Publication 1075, section 10.

- Agreement on the scope of the security boundary for the section of the cloud where FTI is accessible and systems with FTI reside. The agency must ensure that boundary details are included in the SLA between the two parties.
- Clearly state that agencies have the right to require changes to their section of the cloud environment, conduct inspections and Safeguard reviews, and cloud providers will comply with IT policies and procedures provided by the agency.
- IRS Publication 1075, Exhibit 12 45-day notification requirement for notifying the IRS prior to executing any agreement to disclose FTI to a contractor the cloud vendor may utilize, or at least 45 days prior to the disclosure of FTI, to ensure appropriate contractual language is included and that contractors are held to safeguarding requirements
- Identification of cloud provider employee awareness and training requirements for access to FTI. IRS Publication 1075, 6.2, Employee Awareness states employees must be certified to understand the agency's security policy and procedures for safeguarding IRS information prior to being granted access to FTI, and must maintain their authorization to access FTI through annual recertification.

Data Encryption in Transit

IRS Publication 1075 requires encryption of FTI in transit. The agency must ensure that encryption requirements are included in contracts with third party providers. The IRS does not advocate specific mechanisms to accomplish encryption as long as they are FIPS 140-2 compliant and configured securely. Additionally, agencies must retain control of the encryption keys used to encrypt and decrypt the FTI at all times and be able to provide information as to who has access to and knows information regarding the key passphrase.

Data Encryption at Rest

In a cloud environment, protection of data and data isolation are a primary concern. Encryption of data a rest provides the agency with assurance that FTI is being properly protected in the cloud. NIST's Draft Special Publication 800-144 recommends "Data must be secured while at rest, in transit, and in use, and access to the data must be controlled." The IRS does not advocate specific mechanisms to accomplish encryption as long as they are FIPS 140-2 compliant and configured securely. Additionally, agencies must retain control of the encryption keys used to encrypt and decrypt the FTI at all times and be able to provide information as to who has access to and knows information regarding the key passphrase.

Persistence of Data in Relieved Assets

If a storage device fails, or in situations where the data is moved within or removed from a cloud environment, actions must be taken to ensure residual FTI is no longer accessible. The destruction or sanitization methods apply to both individual devices that have failed as well as in situations where the agency removes data from the CCE or relocates FTI to another environment.

The technique for clearing, purging, and destroying media depends on the type of media being sanitized. Acceptable physical destruction methods would include disintegration, incineration, pulverizing, shredding, or melting. Repurposed media must be purged to ensure no residual FTI remains on the device.

As there are varied approaches towards secure sanitization based on vendor specifications, cloud providers should consult their data storage vendor to determine the best method to sanitize the asset. If the storage device will no longer be in service, the residual data must be purged using Secure Erase or through degaussing using a NSA/CSS approved degausser. The cloud provider is required to notify the agency upon destroying or repurposing storage media. The agency must verify that FTI has been removed or destroyed and notify the IRS Office of Safeguards of the destruction of storage media in the agency's annual Safeguard Activity Report (SAR).

Risk Assessment. Agencies are required to conduct a risk assessment (or update an existing risk assessment, if one exists) when migrating FTI to a cloud environment. Subsequently, the risk assessment must be reviewed annually to account for changes to the environment. This implementation and an evaluation of the associated risks should be part of the risk assessment. The IRS Office of Safeguards will evaluate the risk assessment as part of the notification requirement in #1.

Security Control Implementation. Cloud providers may designate selected controls as customer defined. For customer defined security controls, the agency must identify, document and implement the customer defined controls, in accordance with Publication 1075. Implementation of some controls may need to be done in partnership with the agency's cloud provider, however the agency has primary responsibility for ensuring it is completed. The agency's capability to test the functionality and security control implementation of a subsystem within a CCE is more limited than the ability to perform testing within the agency's own infrastructure. However, other mechanisms such as third-party assessments may be used to establish a level of trust with the cloud provider.

CLOUD COMPUTING AND THE LAW

The legal implications of using cloud computing solutions are broadly similar to those for any outsourcing arrangement

with a third party. One major difference in using a cloud provider arises from the flexibility and movement of data between servers that may be located in various parts of the world. This makes it difficult to identify which law applies at any given time to the data, particularly as the data may also have been fragmented to suit particular cloud availability or capability. However, the key point is to consider the legal implications of using the cloud at the outset when planning your cloud provision to ensure that you have considered the risks and their management and mitigation to a satisfactory level for your institution.

Key legal areas to consider with cloud computing are:

- Data security and data protection compliance
- Jurisdiction
- Confidentiality
- Freedom of information
- Copyright
- Equality legislation
- Law of contract (including contract enforcement)

Information Security and Compliance

There is ongoing legal research and argument that data protection law is outdated and is currently not a good fit with cloud computing. In particular, it is argued that unreasonable demands are being placed on data controllers wishing to transfer data overseas. Furthermore, due to the nature of the cloud and the amount of control the cloud

provider may exert over the data and its movement, there is a view that a cloud provider is also the data controller (and that therefore an institution is effectively passing its data to a third party). However, despite the legal debate, the current position in UK law is that an institution will usually be considered the data controller with regard to its personal data being processed using cloud computing facilities. As such, your institution will need to comply with the Data Protection Act 1998.

The Data Protection Act 1998 (DPA) applies to the 'processing' of personal data. The definition of processing is broad and will include transfer, storage, alteration, and deletion *i.e.* it covers all interaction with the data. The DPA applies to personal data only. This is defined as data relating to a living individual from which you can identify the individual or which, if combined with other data, may identify the individual.

In using a cloud service, an institution will usually be the data controller responsible for compliance with the DPA when processing personal data and the cloud provider will be the data processor. The cloud provider as data processor should act in accordance with the agreed terms under the contract with your institution in order to ensure compliance with the DPA.

Institutions will also have other confidential data which is not personal data, for example, sensitive financial planning

data which it will consider as confidential or highly sensitive and requiring adequate protection from unauthorised access or release. This will be discussed in more detail under confidentiality below.

Whether outsourcing data processing to a processor (*i.e.* the cloud provider) or processing the information within the institution, a data controller has eight data protection principles to adhere to in order to comply with the DPA. The principles are intended to provide a technology-neutral framework for balancing an organisation's need to make the best use of personal data, while safeguarding that information and respecting individuals' private lives.

The eight data protection principles state that personal data must be:

- Fairly and lawfully processed.
- Processed for limited, stated purposes.
- Adequate, relevant and not excessive.
- Accurate and up-to-date.
- Kept no longer than necessary.
- Processed in accordance with the individual's rights.
- Secure.
- Not transferred to a country outside the European Economic Area unless that country has adequate data protection itself.

There are also additional conditions to meet to ensure compliance and these depend on whether the data is personal

data or sensitive personal data. As stated above, 'Personal data' is any information, including photographs or other images, about an identifiable living individual regardless of the format of information. The overriding test is whether the information in question on its own or when combined with other information, is significant biographical information that would identify the individual. 'Sensitive personal data' includes information regarding an individual's race or ethnic origin, and physical or mental health.

Security of Personal Data

An institution as data controller has an obligation to ensure that the cloud provider has adequate measures in place to protect personal data securely against unauthorised or unlawful processing, and against accidental loss, destruction, and damage. There is no set definition of what would constitute 'adequate security' and the Information Commissioner's Office (ICO), which is responsible for enforcing compliance with the DPA, suggests a risk based approach.

When choosing a cloud provider, your institution should enquire as to how that provider handles personal data. It should also investigate assurances offered, responses to breaches, reactions to UK DPA requirements, and use of security measures such as encryption. It is important to ascertain which other third parties have access to the data, for example, those to whom elements of the cloud service provided is subcontracted *e.g.* an IAAS or PAAS.

Your institution will need to consider whether the security level offered meets both the institution's requirements and that of the DPA. It will also need to ensure that the terms of the contract with the cloud provider reflect these requirements.

Transfer of Information to a Country Outside the EEA

The DPA states that personal data is not to be transferred outside the EEA (European Economic Area) unless there is an adequate level of protection for the data subjects regarding the processing of personal data.

Cloud providers intrinsically store and move data around multiple servers potentially situated in a number of jurisdictions which may very likely be outside the EEA. This activity will breach the DPA unless these jurisdictions have adequate security measures in place.

Compliance may be achieved through using EU approved contract terms with your cloud provider, or a cloud provider in the US who has signed up to the Safe Harbour provisions, or by getting informed consent from the data subjects to transfer it to an 'unsafe' location (which is not a recommended solution).

Jurisdiction

Broadly speaking, a UK court can only rule on a dispute if it has jurisdiction and similarly, law enforcement agencies

such as the police may only operate where they have agreement or jurisdiction. To complicate matters, there are a number of international agreements relating to jurisdiction. Varying rules as to which jurisdiction applies have been agreed upon depending on the area of law and nature of the dispute. In addition, there may be layering of provision, for example, a cloud provider may outsource some of his service provision *e.g.* data storage or infrastructure. This will make it more difficult to ascertain where the information is at any given time. Also, local laws may apply which permit wider access than you anticipated. A well publicised example of this is the US Patriot Act. The Patriot Act is intended to assist terrorism prevention in the US and permits access to data by the US Federal Government in certain circumstances, mainly in the interest of national security. The assumption is that in using a cloud provider, data will be moved. However, without knowing the jurisdiction to which it is moved, it will be difficult to assess the jurisdiction and its suitability. If this issue is not discussed and agreed at the outset, the result may be protracted disputes. There are examples of such disputes involving Google and Yahoo. Law enforcement agencies may also experience difficulty in (a) tracking down information and (b) finding information in jurisdictions where they have no authority.

Confidentiality

The flow and movement integral to cloud computing may make it difficult to locate the data at any given time and

difficult to ascertain whether it is in a secure location. It may often be difficult to assess which third parties have access or access capability to the data as this may change with new or additional providers, services and server locations. Information handled in a cloud environment, although not personal data, may be confidential in nature. An example would be an institution's dealings with a commercial research collaborator (*e.g.* a pharmaceutical company), where your institution may be subject to contract confidentiality clauses which constitute a legal obligation of confidence. If confidentiality of information is crucial, then risk decisions will need to be made as to its security. Although it might be implied that the cloud provider will maintain confidentiality, it may be desirable to state clearly in the contract terms what obligations of confidentiality are owed between the parties.

Freedom of Information

Institutions in the UK, as public authorities, have a legal duty to comply with freedom of information (Freedom of Information Act 2000 and Freedom of Information (Scotland) Act 2002) and other related legislation such as the environmental regulations. If a request is made to an institution for information and your institution holds the information, it is required to release it to the requester within 20 days, unless an exemption or an exception applies. It is likely that even although the information may be stored in

the cloud, an institution will still be deemed to be holding it for the purpose of FOI. It is therefore necessary to ensure that access to information is timely - outage, failure, and back-up details should be assessed when choosing your cloud provider.

Intellectual Property Rights

Intellectual property rights (IPR) are, broadly, rights granted to creators and owners of works that are the result of human intellectual creativity. These works can be in the industrial, scientific, literary or artistic domains. The types of IPR considered here are copyright, the database right, and patents.

Essentially, copyright protects original works, including films or broadcasts, and the typographical layout of published editions. This will include works such as teaching and research materials and blogs. Software (computer programmes) and databases may be protected as literary works, in addition to other possible rights such as database right.

A college or university will usually be the owner of copyright works created by its staff, unless there is an agreement otherwise. A copyright owner has the right to control the copying, adaptation, publishing, performance and broadcast of the work, and under what conditions this may be done. In addition to creating materials to which copyright will apply, staff and students of colleges and universities are likely to

use work that belongs to others extensively. Compliance with copyright law remains necessary in migration to the cloud.

In addition to any copyright protection, a database may be protected by the database right. The database right applies in the EU and is intended to protect and reward investment in the creation and arrangement of databases.

A patent protects the features and processes that make things work, allowing inventors to profit from their inventions. It gives the patent owner the right to prevent others from making, using, importing or selling the invention without permission.

Using a cloud provider for IT service provision raises particular IPR issues for institutions to consider prior to agreeing the terms of their cloud computing provision.

Two main issues arise:

1. The cloud provider (*i.e.* a third party) may have access to data belonging to an institution
2. The location of the data is not fixed

This has implications for an institution's IPR compliance.

Licence Restrictions

Software licences may be location specific and these will require review to ensure continued compliance when considering a cloud infrastructure service. An institution will have contractually agreed with publishers via current educational resource licences (*e.g.* Copyright Licensing Agency Limited (CLA) licence) to safeguard resources.

The licence agreement, for example, may state that only authorised persons *e.g.* staff and students may view the digital resource or storage of digital material may be restricted under the licence to local servers. There is a possibility of third party (*i.e.* cloud provider and their sub-contractors) access and the cloud is intrinsically not location specific. Contractual agreements with your resource suppliers, on access and location, need to be reflected in your contract with your cloud provider via warranties. The cloud provider should provide assurances that best efforts will be made to prevent access by unlicensed users and to prevent any unauthorised usage of the licensed resources.

Creation of Content in the Cloud

Where content is created in the cloud then whilst it may usually be possible to identify the creator and therefore the first copyright owner, it may be more difficult to identify where the material was created. This will not affect copyright protection *per se*, but may affect whether correct formalities have been followed in a particular jurisdiction, which in turn may affect ability to take court action if necessary.

Database Right

If a database is recorded on a server in an EU member state then it is clear that a valid database right may apply, provided of course that the database meets the criteria outlined above for protection. However, research has raised

the question of whether it is where the database is made or where it is recorded that is key and whether these are different places according to the legislation. This may potentially affect whether database right applies or not as there is no database right, for example, in the US. As there is no court decision on the interpretation, some uncertainty exists as to whether a database recorded on a non EU server will be protected by the database right. It is important to ensure that no residual database rights should be created for the cloud provider.

THE INTERFACE OF AN EMAIL CLIENT, THUNDERBIRD

Messages are exchanged between hosts using the Simple Mail Transfer Protocol with software programs called mail transfer agents. Users can retrieve their messages from servers using standard protocols such as POP or IMAP, or, as is more likely in a large corporate environment, with a proprietary protocol specific to Lotus Notes or Microsoft Exchange Servers. Webmail interfaces allow users to access their mail with any standard web browser, from any computer, rather than relying on an email client.

Mail can be stored on the client, on the server side, or in both places. Standard formats for mailboxes include Maildir and mbox. Several prominent email clients use their own proprietary format and require conversion software to transfer email between them. Accepting a message obliges an MTA to

deliver it, and when a message cannot be delivered, that MTA must send a bounce message back to the sender, indicating the problem.

FILENAME EXTENSIONS

Upon reception of email messages, email client applications save message in operating system files in the file system. Some clients save individual messages as separate files, while others use various database formats, often proprietary, for collective storage. A historical standard of storage is the *mbox* format. The specific format used is often indicated by special filename extensions:

- eml : Used by many email clients including Microsoft Outlook Express, Windows Mail and Mozilla Thunderbird. The files are plain text in MIME format, containing the email header as well as the message contents and attachments in one or more of several formats.
- emlx

  Used by Apple Mail.
- msg

  Used by Microsoft Office Outlook.
- mbx : Used by Opera Mail, KMail, and Apple Mail based on the mbox format.

Some applications (like Apple Mail) also encode attachments into messages for searching while also producing a physical copy of the files on a disk. Others

separate attachments from messages by depositing them into designated folders on disk.

URI SCHEME MAILTO

The URI scheme, as registered with the IANA, defines the mailto: scheme for SMTP email addresses. Though its use is not strictly defined, URLs of this form are intended to be used to open the new message window of the user's mail client when the URL is activated, with the address as defined by the URL in the "To:" field.

USE

This section needs additional citations for verification. Please help improve this article by adding reliable references. Unsourced material may be challenged and removed. *(November 2007)*

IN SOCIETY

There are numerous ways in which people have changed the way they communicate in the last 50 years; email is certainly one of them. Traditionally, social interaction in the local community was the basis for communication – face to face. Yet, today face-to-face meetings are no longer the primary way to communicate as one can use a landline telephone, mobile phones or any number of the computer mediated communications such as email. Research has shown that people actively use email to maintain core social networks, particularly when others live at a distance.

However, contradictory to previous research, the results suggest that increases in Internet usage are associated with decreases in other modes of communication, with proficiency of Internet and email use serving as a mediating factor in this relationship. With the introduction of chat messengers and video conference there are more ways to communicate.

FLAMING

Flaming occurs when a person sends a message with angry or antagonistic content. Flaming is assumed to be more common today because of the ease and impersonality of email communications: confrontations in person or via telephone require direct interaction, where social norms encourage civility, whereas typing a message to another person is an indirect interaction, so civility may be forgotten. Flaming is generally looked down upon by Internet communities as it is considered rude and non-productive.

EMAIL BANKRUPTCY

Also known as "email fatigue", email bankruptcy is when a user ignores a large number of email messages after falling behind in reading and answering them. The reason for falling behind is often due to information overload and a general sense there is so much information that it is not possible to read it all. As a solution, people occasionally send a boilerplate message explaining that the email inbox is being cleared out.

Stanford University law professor Lawrence Lessig is credited with coining this term, but he may only have popularized it.

IN BUSINESS

Email was widely accepted by the business community as the first broad electronic communication medium and was the first 'e-revolution' in business communication. Email is very simple to understand and like postal mail, email solves two basic problems of communication: logistics and synchronization.

LAN based email is also an emerging form of usage for business. It not only allows the business user to download mail when *offline*, it also provides the small business user to have multiple users email ID's with just *one email connection*.

PROS

- The problem of logistics : Much of the business world relies upon communications between people who are not physically in the same building, area or even country; setting up and attending an in-person meeting, telephone call, or conference call can be inconvenient, time-consuming, and costly. Email provides a way to exchange information between two or more people with no setup costs and that is generally far less expensive than physical meetings or phone calls.

- The problem of synchronization : With real time communication by meetings or phone calls, participants have to work on the same schedule, and each participant must spend the same amount of time in the meeting or call. Email allows asynchrony: each participant may control their schedule independently.

CONS

This section may contain original research or unverified claims. Please improve the article by adding references. Most business workers today spend from one to two hours of their working day on email: reading, ordering, sorting, 're-contextualizing' fragmented information, and writing email. The use of email is increasing due to increasing levels of globalization—labour division and outsourcing amongst other things. Email can lead to some well-known problems:

- Loss of Context: which means that the context is lost forever; there is no way to get the text back.

Information in context (as in a newspaper) is much easier and faster to understand than unedited and sometimes unrelated fragments of information. Communicating in context can only be achieved when both parties have a full understanding of the context and issue in question.

- Information overload: Email is a push technology—the sender controls who receives the information. Convenient availability of mailing lists and use of "copy

all" can lead to people receiving unwanted or irrelevant information of no use to them.

- Inconsistency: Email can duplicate information. This can be a problem when a large team is working on documents and information while not in constant contact with the other members of their team.

Despite these disadvantages, email has become the most widely used medium of communication within the business world.

PROBLEMS

This section needs additional citations for verification. Please help improve this article by adding reliable references. Unsourced material may be challenged and removed. *(November 2007)*

INFORMATION OVERLOAD

A December 2007 New York Times blog post described Email as "a $650 Billion Drag on the Economy", and the New York Times reported in April 2008 that "Email has become the bane of some people's professional lives" due to information overload, yet "none of the current wave of high-profile Internet start-ups focused on email really eliminates the problem of email overload because none helps us prepare replies".

6

Key Elements of Cloud Computing

Cloud computing involves three key elements, namely, resource pooling, capability supply, and the service model. As a process, the cloud provider concentrates a mass of resources and seamlessly provides them for users. These three elements are described below:

Resource Pooling

Various kinds of resources are converged to form a cloud. Though a revolutionary concept, this already exists in practice in various forms. Search engines construct a "search cloud" by prearranging and converging all available information on the Internet so that the user can quickly get the search result. The highly popular taobao.com centralizes and converges online stores to form a "cloud" market much in the same way that a traditional shopping mall converges physical

stores. Moreover, if we converge video data collected from myriad video cameras installed on city streets to provide VOD, we can establish a "cloud view".

Capability Supply

What does a cloud bring us? To be called a cloud, it must supply computing and processing capabilities and share resources. In the early 1990s, we researched computing capability supply with a focus on presenting storage space and databases and interfaces for delivering their storage and search functions. Converged resources and computing capabilities are useless without a supply channel in the same way that taobao.com would be useless without a shopping platform to enable transactions.

Service Model

The "as a service" (aaS) model is gaining popularity in the computing world. It is about services, not technologies, for users usually care about the services they are getting, not the technologies or resources involved. For example, we initially bought full-priced software and prepared necessary resources for it, no matter whether we would use it or not. Later, we shared software on the Internet at a lower cost without needing to prepare resources on the client though a fixed cost was still incurred for the usage over a certain period of time. With cloud computing, we can directly use the software function (the service) provided by the cloud even

without knowing the software, on a pay-per-use basis without an awareness of the software involved.

Many similar applications exist to convert available capabilities and resources into the services required by users. Users care about software functions but not where the software is installed, which has spawned "Software as a Service" (SaaS). Equally, users are interested in storage space size, but not storage mechanics. This has given us "Software testing as a Service" (StaaS). Users want their requests to be satisfied, but are not concerned with the process, creating "Platform as a Service" (PaaS). Finally, users are concerned with the availability of computing systems, not the infrastructure through which they are implemented. This has brought about "Infrastructure as a Service" (IaaS).

CLOUD COMPUTING – IS IT REALLY ALL THAT BENEFICIAL?

Cloud computing is now evolving like never before, with companies of all shapes and sizes adapting to this new technology. Industry experts believe that this trend will only continue to grow and develop even further in the coming few years. While cloud computing is undoubtedly beneficial for mid-size to large companies, it is not without its downsides, especially for smaller businesses. We now bring you a list of advantages of disadvantages of cloud computing, with a view to helping

such establishments fully understand the concept of cloud computing.

ADVANTAGES OF CLOUD COMPUTING

If used properly and to the extent necessary, working with data in the cloud can vastly benefit all types of businesses. Mentioned below are some of the advantages of this technology:

Why Cloud Computing Is Ideal for Small Businesses

- Cost Efficient: Cloud computing is probably the most cost efficient method to use, maintain and upgrade. Traditional desktop software costs companies a lot in terms of finance. Adding up the licensing fees for multiple users can prove to be very expensive for the establishment concerned. The cloud, on the other hand, is available at much cheaper rates and hence, can significantly lower the company's IT expenses. Besides, there are many one-time-payment, pay-as-you-go and other scalable options available, which makes it very reasonable for the company in question.
- Almost Unlimited Storage: Storing information in the cloud gives you almost unlimited storage capacity. Hence, you no more need to worry about running out of storage space or increasing your current storage space availability.
- Backup and Recovery: Since all your data is stored in the cloud, backing it up and restoring the same is

relatively much easier than storing the same on a physical device. Furthermore, most cloud service providers are usually competent enough to handle recovery of information. Hence, this makes the entire process of backup and recovery much simpler than other traditional methods of data storage.

5 of the Best Third-Party Mobile Cloud Sync Providers

- Automatic Software Integration: In the cloud, software integration is usually something that occurs automatically. This means that you do not need to take additional efforts to customize and integrate your applications as per your preferences. This aspect usually takes care of itself. Not only that, cloud computing allows you to customize your options with great ease. Hence, you can handpick just those services and software applications that you think will best suit your particular enterprise.
- Easy Access to Information: Once you register yourself in the cloud, you can access the information from anywhere, where there is an Internet connection. This convenient feature lets you move beyond time zone and geographic location issues.

  Cloud Computing – Is it Possible to Assign a Standard?
- Quick Deployment: Lastly and most importantly, cloud computing gives you the advantage of quick

deployment. Once you opt for this method of functioning, your entire system can be fully functional in a matter of a few minutes. Of course, the amount of time taken here will depend on the exact kind of technology that you need for your business.

DISADVANTAGES OF CLOUD COMPUTING

In spite of its many benefits, as mentioned above, cloud computing also has its disadvantages. Businesses, especially smaller ones, need to be aware of these cons before going in for this technology.

- Technical Issues: Though it is true that information and data on the cloud can be accessed anytime and from anywhere at all, there are times when this system can have some serious dysfunction. You should be aware of the fact that this technology is always prone to outages and other technical issues. Even the best cloud service providers run into this kind of trouble, in spite of keeping up high standards of maintenance. Besides, you will need a very good Internet connection to be logged onto the server at all times. You will invariably be stuck in case of network and connectivity problems.
- Security in the Cloud: The other major issue while in the cloud is that of security issues. Before adopting this technology, you should know that you will be surrendering all your company's sensitive information

to a third-party cloud service provider. This could potentially put your company to great risk. Hence, you need to make absolutely sure that you choose the most reliable service provider, who will keep your information totally secure.

What Strategies Should an Enterprise Adopt in Order to Ensure Data Protection?

- Prone to Attack: Storing information in the cloud could make your company vulnerable to external hack attacks and threats. As you are well aware, nothing on the Internet is completely secure and hence, there is always the lurking possibility of stealth of sensitive data.

CLOUD COMPUTING SECURITY

Cloud computing changes personal and enterprise computing models in a way that makes information security as relevant as it is for online banking services. Though cloud computing has been applied to network security, it remains to be seen whether its architecture is an information trap that is ripe for misuse or exploitation. To provide secure services, cloud computing must address this issue at the following three levels:

How does Cloud Security Work?

In the current networking environment, client-based Trojan checks are increasingly discredited as a solution. To

check malicious codes, security vendors need a cloud computing platform where an inbuilt cloud security system pre-scans web pages and immediately informs the user of a page's safety. The advantage of cloud security is its ability to scan all web pages using large-scale computing capabilities. For end users, the web is only one danger source; others include e-mails and USBs, though the cloud security system does not apply to users who do not wish to publish their personal information.

However, the cloud security system has a fatal weakness– its over-reliance on transmission channels. User information and resources are handed over to the cloud for processing and transmission and security depends on the internal transmission channels. In order for cloud computing to unleash its potential, cloud service providers must work with broadband service providers to build a broadband transmission system appropriate for cloud services. A recommended solution involves the integration of the cloud security system into telecom networks so that the former scans web pages and the latter sends risk alerts.

Is the Cloud itself Safe?

The cloud must be open to provide service. Openness usually leads to vulnerability, though. So it is a problem how to protect the cloud against attacks and ensure that the cloud provides services continuously. The recent system crash at Amazon web services, the cloud service provider,

caused Twitter and other prominent web sites to fail. If this type of crash causes loss of user data, users will doubt the security of cloud computing. Solutions may include backups and additional monitoring, both of which should provide vital areas for future research.

How does the Cloud Ensure User Security?

In the cloud, the security level of user routines is not analysed, nor is data copied, in order to protect business secrets and personal data of users. Cloud services can be widely used only when they are reliable. However, unlike investigating a retailer before buying products from it, cloud service users cannot check the reliability of a cloud because they do not know which part of the cloud is serving them. If users transmit encrypted data, the user routines will be inefficient.

The openness of a cloud may render it a malicious tool. Currently, harmful Internet activities require the control of the terminal. For example, phishing requires fake sites that look and feel almost identical to legitimate ones, and Trojans require network controllers. The cloud model opens up new possibilities for criminal and malicious behaviour.

ECONOMICS OF CLOUD COMPUTING

The rationale behind the cloud model and the idea behind resource provision is flexibility. For example, a user requires 10,000 computers as nodes to work at a full load for a couple

of months but for the rest of the year requires only 5 per cent to 20 per cent of these nodes. Such a user can apply for 200 to 1,000 nodes for normal operation and 10,000 nodes in the peak period.

To ensure security, some existing cloud systems provide exclusive private resources for users at certain costs that basically equal those required to own these resources. This solution obviously goes against the economics of cloud computing, making it unnecessary for users to apply for resources in the cloud. It remains a subject of debate and research whether this is in fact a cloud model.

CLOUD MANAGEMENT TOOLS GUIDE FOR BEGINNERS

Cloud management is a hot topic, so hot that every startup and established vendor has some form of tool for managing cloud computing environments. There are tools that monitor, tools that provision, and tools that cross the divide between both. Then there's just vaporware, and sorting through that can be a challenge.

If your cloud deployment is fairly static or not mission-critical, then you may not need a dynamic provisioning system. In that case, the standard tools for resource adds/changes/removals included with the product may suffice. Several providers have products designed for cloud computing management (VMware, OpenQRM,

CloudKick, and Managed Methods), along with the big players like BMC, HP, IBM Tivoli and CA. Each uses a variety of methods to warn of impending problems or send up the red flag when a sudden problem occurs. Each also tracks performance trends.

While they all have features that differentiate them from each other, they're also focused on one key concept: providing information about cloud computing systems. If your needs run into provisioning, the choices become more distinct than choosing "agent vs. agentless" or "SNMP vs. WBEM."

The main cloud infrastructure management products offer similar core features:

- Most support different cloud types (often referred to as hybrid clouds).
- Most support the on-the-fly creation and provisioning of new objects and the destruction of unnecessary objects, like servers, storage, and/or apps.
- Most provide the usual suite of reports on status (uptime, response time, quota use, etc.) and have a dashboard that can be drilled into.

When it comes to meeting those three criteria, there are a few vendors that offer pervasive approaches in handling provisioning and managing metrics in hybrid environments: RightScale, Kaavo, Zeus, Scalr and Morph. There are also options offered by cloud vendors themselves that meet the second and third criteria, such as CloudWatch from Amazon

Web Services. The large companies known for their traditional data centre monitoring applications have been slow to hit the cloud market, and what products they do have are rehashes of existing applications that do little in the way of providing more than reporting and alerting tools. CA is on an acquisition spree to fix this and just acquired 3Tera, a cloud provisioning player.

An example of the confusion in the industry is IBM's Tivoli product page for cloud computing. You'll notice that clicking the Getting Started tab results in a 404 error. Nice work, IBM.

Meanwhile, HP's OpenView (now called Operations Manager) can manage cloud-based servers, but only insofar as it can manage any other server. BMC is working on a cloud management tool, but doesn't have anything beyond its normal products out at the moment. In place of these behemoths, secondary players making splashes on the market are offering monitoring-focused applications from companies like Scout, UpTime Systems, Cloudkick, NetIQ and ScienceLogic. There is also the "app formerly known as" Hyperic, now owned by VMware through the acquisition of SpringSource.

In truth, we could rival John Steinbeck and Robert Jordan in word count when it comes to writing about all the products in this field, though within a year or two it should be a much smaller space as acquisitions occur, companies fail and the

market sorts itself out. There's a lot on the way in cloud computing, not the least of which is specifications. Right now the cloud is the Wild West: vast, underpopulated, and lacking order except for a few spots of light.

These are the best infrastructure management and provisioning options available today:

RightScale

RightScale is the big boy on the block right now. Like many vendors in the nascent market, they offer a free edition with limitations on features and capacity, designed to introduce you to the product (and maybe get you hooked, ala K.C. Gillette's famous business model at the turn of the 20th century).

RightScale's product is broken down into four components:

1. Cloud Management Environment
2. Cloud-Ready ServerTemplate and Best Practice Deployment Library
3. Adaptable Automation Engine
4. Multi-Cloud Engine

A fifth feature states that the "Readily Extensible Platform supports programmatic access to the functionality of the RightScale Platform." In looking at the product, these features aren't really separate from one another, but make a nice, integrated offering.

RightScale's management environment is the main interface users will have with the software. It is designed to

walk a user through the initial process of migrating to the cloud using their templates and library. The management environment is then used for (surprise!) managing that environment, namely continuing builds and ensuring resource availability. This is where the automation engine comes into play: being able to quickly provision and put into operation additional capacity, or remove that excess capacity, as needed. Lastly, there is the Multi-Cloud Engine, supporting Amazon, GoGrid, Eucalyptus and Rackspace. RightScale is also working on supporting the Chef open-source systems integration specifications, as well. Chef is designed from the ground up for the cloud.

Kaavo

Kaavo plays in a very similar space to RightScale.

The product is typically used for:

- Single-click deployment of complex multi-tier applications in the cloud (Dev, QA, Prod)
- Handling demand bursts/variations by automatically adding/removing resources
- Run-time management of application infrastructure in the cloud
- Encryption of persisted data in the cloud
- Automation of workflows to handle run-time production exceptions without human intervention

The core of Kaavo's product is called IMOD. IMOD handles configuration, provisioning and changes (adjustments in their

terminology) to the cloud environment, and across multiple vendors in a hybrid model. Like all major CIM players, Kaavo's IMOD sits at the "top" of the stack, managing the infrastructure and application layers. One great feature in IMOD is its multi-cloud, single system tool. For instance, you can create a database backend in Rackspace while putting your presentation servers on Amazon. Supporting Amazon and Rackspace in the public space and Eucalyptus in the private space is a strong selling point, though it should be noted that most cloud management can support Eucalyptus if it can also support Amazon, as Eucalyptus mimics Amazon EC2 very closely.

Both Kaavo and RightScale offer scheduled "ramp-ups" or "ramp-downs" (dynamic allocation based on demand) and monitoring tools to ensure that information and internal metrics (like SLAs) are transparently available. The dynamic allocation even helps meet the demands of those SLAs. Both offer the ability to keep templates as well to ease the deployment of multi-tier systems.

Zeus

Zeus was famous for its rock-solid Web server, one that didn't have a lot of market share but *did* have a lot of fanatical fans and top-tier customers. With Apache, and to a lesser-extent, IIS, dominating that market, not to mention the glut of load balancers out there, Zeus took its expertise in the application server space and came up with the Application

Delivery Controller piece of the Zeus Traffic Controller. It uses traditional load balancing tools to test availability and then spontaneously generate or destroy additional instances in the cloud, providing on-the-fly provisioning. Zeus currently supports this on the Rackspace and, to a lesser extent, Amazon platforms.

Scalr

Scalr is a young project hosted on Google Code and Scalr.net that creates dynamic clusters, similar to Kaavo and RightScale, on the Amazon platform. It supports triggered upsizing and downsizing based on traffic demands, snapshots (which can be shared, incidentally, a very cool feature), and the custom building of images for each server or server-type, also similar to RightScale. Being a new release, Scalr does not support the wide number of platforms, operating systems, applications, and databases that the largest competitors do, sticking to the traditional expanded-LAMP architecture (LAMP plus Ruby, Tomcat, etc.) that comprises many content systems.

Morph

While not a true management platform, the MSP-minded Morph products offers similar functionality in its own private space. Morph CloudServer is a newer product on the market, filling the management and provisioning space as an appliance. It is aimed at the enterprise seeking to deploy a private cloud.

Its top-tier product, the Morph CloudServer is based on the IBM BladeCenter, and supports hundreds of virtual machines.

Under the core is an Ubuntu Linux operating system and the Eucalyptus cloud computing platform. Aimed at the managed service provider market, Morph allows for the creation of private clouds and the dynamic provisioning within those closed clouds. While still up-and-coming, Morph has made quite a splash and bears watching, particularly because of its open-source roots and participation in open-cloud organizations.

CloudWatch

Amazon's CloudWatch works on Amazon's platform only, which limits its overall usefulness as it cannot be a hybrid cloud management tool. Since Amazon's Elastic Compute Cloud (EC2) is the biggest platform out there (though Rackspace claims it is closing that gap quickly), it still bears mentioning. CloudWatch for EC2 supports dynamic provisioning (called auto-scaling), monitoring, and load-balancing, all managed through a central management console — the same central management console used by Amazon Web Services. Its biggest advantage is that it requires no additional software to install and no additional web site to access applications through. While the product is clearly not for enterprises that need hybrid support, those that exclusively use Amazon should know that it is as robust and functional as the other market players.

CLOUD SERVICE MANAGEMENT SOLUTION

Cloud-computing is one of the hottest topics in the information technology industry today. It simplifies the way information technology resources are consumed and managed, with promises of improved cost efficiencies, faster time-to-market and the ability to scale applications on demand. Infrastructure as a Service (IaaS) is a cloud-computing service model, which delivers virtualized computing, storage and network infrastructures as service. By being able to add services without adding physical servers, end-users gain the benefits of flexible server provisioning while saving cost on power and cooling as well as space since a lot more can be done on the same physical footprint. The key to success of cloud-computing technologies is automated management of cloud resources. It provides the automation needed to address the operational complexities in managing virtual resources. With Teamup Technology's cloud service management solution, called the Cloud Service Manager, you get a full set of functionalities to meet your business needs, including:

- Wizards for simplified service provisioning and deployment.
- Automated service request, approval and fulfilment.
- Full life-cycle management of virtual machines.
- Continuous monitoring of virtual machines and triggering of user alerts when resource usage goes beyond predefined thresholds.

- Automatic reconfiguration of virtual machines in accordance with service level agreements.
- Support of multi-tenancy for today's organization hierarchies.
- Role-based access control to prevent unauthorized access.
- Logging of all user interactions to facilitate auditing.
- Comprehensive usage reports in tabular and graphical formats.
- Flexible chargeback with fixed and variable charging models based on resource usage.
- Integration with external billing systems through periodic generation of resource usage detail records.
- Web portal for easy access to cloud service from the internet.
- Use of mobile computing devices to access cloud services, *e.g.* iPhone and iPad.
- Internationalization with support for multiple languages.

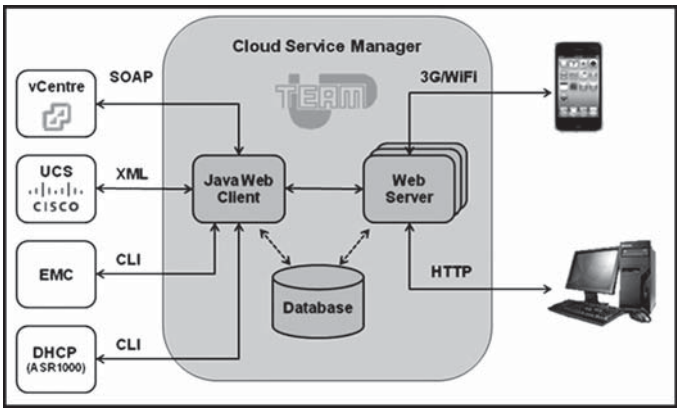


Fig. System Block Diagram

Figure below shows a block diagram of Cloud Service Manager:

In broad terms, Cloud Service Manager consists of two major subsystems:

- Frontend System for providing users with access to cloud services everywhere from various fixed and mobile computing devices, including web browsers, iPhone and iPad.
- Backend System for managing computing, storage and network resources in the backend datacenter infrastructure.

The frontend and backend systems communicate through an open XML API which can be used for the development of customized user interfaces.

Besides, the frontend web portal is developed with pluggable portlets (JSR168, JSR286) which make it easy to customize the look-and-feel of the web portal.

Target Industry

Cloud-computing is the future for datacenters. Telecom service providers are in a unique position to build virtualized data centres to address enterprise needs across the realm of public cloud services. The key ingredients for success are already in place for service providers, including established IP network infrastructures and trusted billing.

Service providers offer infrastructure as a service (IaaS) to consumers and businesses at zero startup cost and

charge their customers on a pay-as-you-go model. Rather than purchasing hardware and software, consumers instead buy those resources as a fully outsourced service.

Yet, enterprises may choose to adopt the private cloud, whereas they build their own cloud infrastructures instead of turning to the internet-based public cloud services and still enjoy the benefits of cloud-computing, including scalability and time-to-market.

Features and Benefits

Key features and benefits of Cisco Unified Computing System (UCS) include:

- Scalability - One Cisco UCS Manager instance can manage two Cisco UCS 6100 Series Fabric Interconnects, multiple Cisco UCS 5100 Series Chassis, 80 Cisco UCS 2100 Series Fabric Extenders, and hundreds of Cisco UCS-B Series Blade Servers.
- Service Profiles - The service profile allows the servers in the Cisco UCS to be treated as raw computing capacity that can be allocated and reallocated among application workloads, enabling a much more dynamic and efficient use of the server capacity than exists in today's data centres. Server deployment with service profile takes minutes instead of the many days or weeks that server deployment takes in many existing data centres.

- Cisco VN-Link - The Cisco UCS implements Cisco VN-Link technology which enables policy-based virtual machine connectivity, mobility of network and security properties during VMware VMotion migration, and a non-disruptive operating model in which network administrators perform network tasks and server administrators perform server tasks.
- Auto-discovery and configuration - Cisco UCS Manager automatically discovers devices that are added, moved, or removed from the system; adds them to its inventory; and applies service profile configurations as appropriate.
- Unified fabric - Cisco UCS provides the foundation for the operation of the internal unified fabric created by the Cisco UCS 6200 Series Fabric Interconnects, the Cisco UCS 2100 Series Fabric Extenders, and the network adapters that are present on the Cisco UCS B-Series Blade Servers. This foundation enables the use of Fibre Channel over Ethernet (FCoE) in the internal fabric, while preserving traditional Ethernet and Fibre Channel connectivity to the core LAN and SAN.
- XML-based API - A full-featured XML API exposing over 9000 objects provides powerful new opportunities for service providers, independent software vendors (ISVs), and users interested in customizing the behaviour of the Cisco UCS to enhance its value in their own unique environments.

- Unified Management - Cisco UCS Manager provides unified, centralized, embedded management of all software and hardware components of the Cisco UCS. By enabling better automation of processes, Cisco UCS Manager allows data centre managers to achieve greater agility and scale in their server operations while reducing complexity and risk.

Key features and benefits of Teamup Technology's Cloud Service Manager (CSM) include:

- Designed for both private and public clouds.
- Aim for both administrators and end-users.
- Right-sized "just-for-clouds" architecture with no legacy and small footprint.
- Cloud services accessible from various mobile internet devices, *e.g.* iPhone and iPad.
- XML-based open API for development of customized user interfaces.
- Internationalization.
- Multi-vendor support with minimal dependency on specific hypervisor; begins with VMware-Cisco-EMC and Cisco devices.
- Support for network devices.
- Policy-driven framework for virtual machine life-cycle management and automation for on-demand resource provisioning.
- Chargeback with flexible schemes and UDRs.

- Multi-tenancy for maximizing resource utilization.
- Designed for scalability and high availability.

System Components

- Cisco Unified Computing System (UCS) is a next-generation data centre platform that unites computing, network, storage access and virtualization into a cohesive system designed to reduce total cost of ownership (TCO) and increase business agility. It is built from the following components:
- Cisco UCS 6100 Series Fabric Interconnects: line-rate, low-latency, lossless interconnect switches
- Cisco UCS 5100 Series Blade Server Chassis: support for up to eight blade servers and up to two fabric extenders in a 6 rack unit (RU) enclosure
- Cisco UCS 2100 Series Fabric Extenders: unified fabric in the blade-server chassis, up to four 10-Gbps connections each
- Cisco UCS B-Series Blade Servers: enhanced support for application demands, energy use, and virtualization
- Cisco UCS B-Series Network Adapters: a range of adapters optimized for virtualization, compatibility with existing driver stacks, or efficient, high-performance Ethernet
- Cisco UCS C-Series Rack-Mount Servers: the benefits of the Unified Computing System in a rack-mount form factor

- Cisco UCS C-Series Network Adapters: a choice of four types of PCI Express (PCIe) adapters
- Cisco UCS Manager: centralized management capabilities

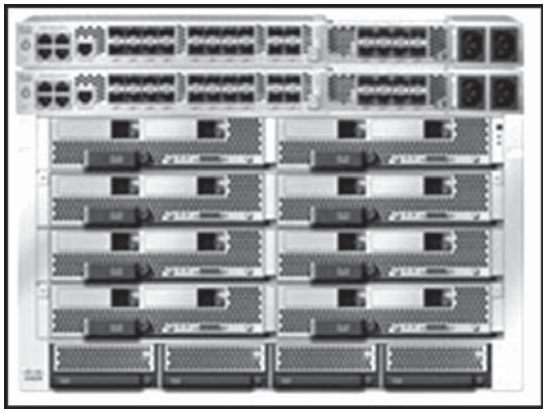


Fig. Below shows the functional diagram of Cloud Service Manager:

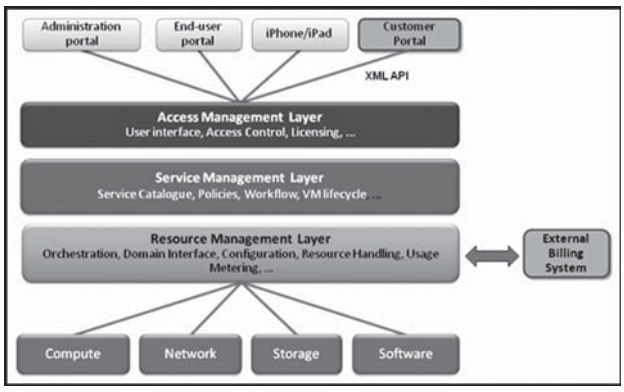


Fig. Functional Diagram

Cloud Service Manager is composed of the following layered structure:

- User Interfaces consisting of web portals, iPhone and iPad.
- Access Management Layer for role-based access control (RBAC), logging of user interactions for auditing purpose and software licensing control.

- Service Management Layer for full life-cycle management of virtual machines, workflow automation, policy-driven service management, etc.
- Resource Management Layer for managing computing, storage, network and software resources, monitoring of server performance and generating user alerts, periodic generation of usage detail records (UDRs), charging, etc.
- Device Drivers for interfacing with the backend datacenter infrastructure, including VMware vSphere, Cisco UCS servers and EMC storage systems.

Figure below shows the detailed system architecture of Cloud Service Manager:

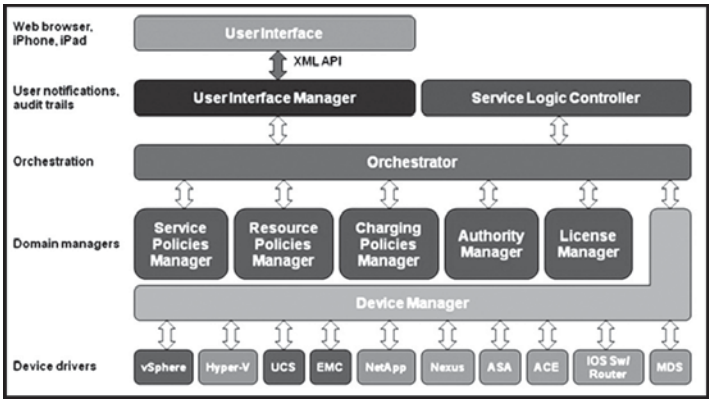


Fig. System Architecture

The Service Logic Controller is the brain of the system, routing the user requests via the Orchestrator to the appropriate software modules for execution.

Services are offered to end-users in the form of service catalogues which are constructed from a bunch of policies

consisting of Service Policies, Resource Policies and Charging Policies. Service Policies specify what the services are and how the services are offered. An example of service policies is the SLA-Aware Auto-Reconfiguration Policy which will automatically add or remove resources from a virtual machine to keep resource usage within predefined bounds. Resource policies are used to manage the computing, storage, network and software resources for service provisioning. Charging policies provide flexible usage-based charging models.

Company Info

Teamup Technology Limited was established in Hong Kong in February 2004. It was founded by a group of technology professionals with rich experience in the field of telecommunication.

Our initial focus was on telecom technology and solutions for fixed and wireless network operators and now we are expanding into the cloud-computing business. Our vision is to be the leading provider of cloud engineering services in designing the systems necessary to leverage the power and economics of cloud resources to maximize profitability and competitive differentiation.

Our strong local development team is committed to ensure that software applications are delivered with high quality and reliability to meet the business needs of our customers.

CLOUD APPLICATION MANAGEMENT COMPLEXITY A GROWING CHALLENGE

IT organizations are already using cloud in highly sophisticated ways in spite of the added level of complexity it introduces in managing applications. That was one of the surprise findings from Enterprise Management Associates' "Radar for Application Performance Management for Cloud Services: Q1 2012," according to the report's author, Julie Craig, research director, application management, at EMA.

"As application management becomes much more complex as they turn to cloud, [IT organizations are] using cloud in ways that surprised me in terms of sophistication," Craig says. For example, almost half of the companies surveyed are running tiered transactions/services spanning both cloud and on-premise, according to the report, while 35 per cent have either integrated or are in the process of integrating multiple software-as-a-service (SaaS) applications.

"The research showed quite a large number of companies that are actually already running transactions that span multiple SaaS services, and I didn't expect them to be this far along," Craig notes. "They are very sophisticated in terms of deployment but struggling as an industry with finding APM products that can deal with this kind of complexity."

Her research found, for example, that many public cloud providers do not yet offer monitoring agents or APIs, which is hampering vendors' ability to build capabilities management into APM products.

Additionally, the majority of midsize to large businesses have already embraced private cloud as a viable delivery option for business-critical applications, the report finds, as some 66 per cent of companies are either already using infrastructure as a service (IaaS) or planning to do so within the next year. Among other findings are that 44 per cent of companies are already utilizing at least one SaaS service, with another 33 per cent planning to do so in the next year.

There are a couple of key challenges in terms of managing cloud applications, says Craig. "One is that you have to build an end-to-end picture of the transaction or application. Unless you have that [visibility], you don't know where to start in terms of actually solving application performance problems." But the challenge is being able to drill down and understand as many as 30 to 40 components supporting the application to determine the origin of a slowdown, for example. "You have to be able to see the application from start to finish in context with the technology that's supporting it," she says. "IT organizations are starting to experience this problem. In many cases they're trying to manage these very complex environments with manual kinds of activities, and they see the need to get away from that because it's taking a

tremendous amount of time." It's also very expensive to manage these complex environments with people alone, she adds.

In spite of the struggles APM vendors are having, Craig was also surprised by "how fast vendors are evolving products to address" the complexity during the timeframe she conducted her research. "There are constant updates and enhancements to cloud APM solutions," she says, "so it's a very rapidly evolving market." APM product vendors covered in the report include OpTier, AppDynamics, AppFirst, Aternity, CA, Compuware, Correlsense, eG Innovations, HP, IBM, INETCO, Nastel, Netuitive, New Relic, OPNET, Quest, SolarWinds and Splunk. More than 40 users of the products were interviewed for the report. But even with the upgrades and enhancements, Craig doesn't see the level of APM complexity diminishing anytime soon. "It's still an evolving market," she says. "The task of managing applications is probably the most challenging of any of the enterprise management disciplines. Application management relies on assimilation of metrics from across the entire application execution ecosystem." Her vision is to see APM become "increasingly automated," although full automation capabilities will likely take five to 10 years to evolve. Flow-based network analytics will likely be a powerful force in enabling full automation of APM systems, Craig believes. Network

analytics leveraging such information to track and model end-to-end application execution will be the answer to enabling greater levels of automation in the future, she maintains. "I believe the network will ultimately provide the unifying information necessary to automate the process of application performance management," Craig says, adding that within the next two to four years, "APM vendors are going to turn to the network in very creative ways to build this end-to-end view of application execution."

CLOUD COMPUTING STANDARDS TAKE SHAPE

In a surprising move to improve interoperability and portability between public and private clouds, seven cloud services providers took off their competitive gloves and formed Cloud Application Management for Platforms (CAMP) last week, the first Platform as a Service (PaaS) management API.

More on Cloud Computing Standards

Collaboration for the standard between Oracle Corp., CloudBees Inc., Cloudsoft Corp., Red Hat Inc., Rackspace Inc., Huawei Technologies Co. and Software AG began in late 2010. CAMP was submitted to OASIS to create a common foundation for deploying and managing applications across multiple cloud environments; the API will also increase interoperability and foster innovation, according to founding companies.

"We offer the opportunity for vendors and consumers to join," said Carol Geyer, senior director of communications and development at OASIS.

"[CAMP] allows us to compete [on] a more common ground," said Steven G. Harris, CloudBees' senior vice president of products. The standard can help guide the industry into an ecosystem of interoperable and portable cloud systems. And that's where OASIS comes into play; it ensures the standards developed respond to the marketplace. The group felt OASIS was the best option to help ensure broad industry participation, "providing an open, collaborative and productive setting," said Jeff Mischkinsky, a senior director at Oracle.

When developing CAMP, OASIS will use a non-assertion mode, an open mode that fosters vendor and consumer adoption as well as community input. Open mode also means CAMP implementers do not need a license, which is appealing to many cloud consumers.

United the Cloud Market Stands, but it's Still Divided

Not all cloud services providers are on board with adopting CAMP specifications, said Richard Pharro, CEO of APM Group Ltd., who believes small vendor groups will form and create their own standards, especially because the largest vendors such as Google and Amazon are not on board. "There is really no compelling reason to join CAMP, unless suddenly there is

a demand from those consuming PaaS to drive towards CAMP, which is highly unlikely," said Linthicum.

Nonetheless, Pharro said he agrees that cloud standardization has value. "It opens up opportunity by providing innovation," he said. "Transportability of data makes it easier for users to manage data across multiple suppliers." The truth is staring the cloud industry in the face: Cloud standards are a necessity. The question of adoption by cloud services providers still lingers unanswered.

MONTAGE APPLICATION

So far, we focused on the technology-side of the equation. In this section, we examine a single application, which is a very important and popular astronomy application.

We use the application as a basis of evaluating the cost/performance tradeoffs of running applications on the Cloud. It also allows us to compare the cost of the Cloud for generating science products as compared to the cost of using your own compute infrastructure.

What Is Montage and Why Is It Useful?

Montage is a toolkit for aggregating astronomical images into mosaics.

Its scientific value derives from three features of its design:

1. It preserves the calibration and astrometric fidelity of the input images to deliver mosaics that meet user-specified parameters of projection, coordinates, and

spatial scale. It supports all projections and coordinate systems in use in astronomy.

2. It contains independent modules for analysing the geometry of images on the sky, and for creating and managing mosaics; these modules are powerful tools in their own right and have applicability outside mosaic production, in areas such as data validation.
3. It is written in *American National Standards Institute* (ANSI)-compliant C, and is portable and scalable – the same engine runs on desktop, cluster, supercomputer or cloud environments running common Unix-based operating systems such as Linux, Solaris, Mac OS X and AIX.

The code is available for download for non-commercial. The current distribution, version 3.0, includes the image mosaic processing modules and executives for running them, utilities for managing and manipulating images, and all third-party libraries, including standard astronomy libraries for reading images. The distribution also includes modules for installation of Montage on computational grids. A web-based Help Desk is available to support users, and documentation is available on-line, including the specification of the *Applications Programming Interface* (API).

Montage is highly scalable. It uses the same set of modules to support two instances of parallelization: MPI a library specification for message passing, and *Planning and Execution*

for Grids (Pegasus), a toolkit that maps workflows on to distributed processing environments. Parallelization and performance.

Montage is in active use in generating science data products, in underpinning quality assurance and validation of data, in analysing scientific data and in creating Education and Public Outreach products.

MONTAGE ARCHITECTURE AND ALGORITHMS

Supported File Formats

Montage supports two-dimensional images that adhere to the definition of the *Flexible Image Transport System* (FITS) standard, the international standard file format in astronomy. The relationship between the pixel coordinates in the image and physical units is defined by the *World Coordinate System* (WCS). Included in the WCS is a definition of how celestial coordinates and projections are represented in the FITS format as *keyword=value* pairs in the file headers.

Montage analyzes these pairs of values to discover the footprints of the images on the sky and calculates the footprint of the image mosaic that encloses the input footprints. Montage supports all projections supported by WCS, and all common astronomical coordinate systems. The output mosaic is FITS-compliant, with the specification of the image parameters written as keywords in the FITS header.

AN ON-DEMAND IMAGE MOSAIC SERVICE

The NASA/IPAC Infrared Science Archive has deployed an on-request image mosaic service. It uses low cost, commodity hardware with portable, Open Source software, and yet is fault-tolerant, scalable, extensible and distributable. Users request a mosaic on a simple web form at http://hachi.ipac.caltech.edu:8080/montage.

The service returns mosaics from three wide-area survey data sets: the 2-Micron All-Sky Survey (2MASS), housed at the *NASA IPAC Infrared Science Archive* (IRSA), the *Sloan Digital Sky Survey* (SDSS), housed at *FermiLab*, and the *Digital Sky Survey* (DSS), housed at the *Space Telescope Science Institute* (STScI).

The first release of the service restricts the size of the mosaics to 1 degree on a side in the native projections of the three datasets.

Users may submit any number of jobs, but only ten may run simultaneously and the mosaics will be kept for only 72 hours after creation.

These restrictions will be eased once the operational load on the service is better understood. The return page shows a JPEG of the mosaic, and provides download links for the mosaic and an associated weighting file. Users may monitor the status of all their jobs on a web page that is refreshed every 15 seconds, and may request e-mail notification of the completion of their jobs.

SCIENTIFIC APPLICATIONS OF CLOUD

Science applications today are becoming ever more complex. They are composed of a number of different application components, often written by different individuals and targeting a heterogeneous set of resources. The applications often involve many computational steps that may require custom execution environments. These applications also often process large amounts of data and generate large results. As the complexity of the scientific questions grows so does the complexity of the applications being developed to answer these questions.

Getting a result is only part of the scientific process. There are three other critical components of scientific endeavours: reproducibility, provenance, and knowledge sharing. We describe them in turn in the context of the scientific applications and revisit them towards the end of the chapter, evaluating how Clouds can meet these three challenges.

As the complexity of the applications increases, reproducibility, the cornerstone of the scientific method, is becoming ever harder to achieve. Scientists often differentiate between scientific and engineering reproducibility. The former implies that another researcher can follow the same analytical steps, possibly on different data, and reach the same conclusions. Engineering reproducibility implies that one can reproduce the same result (on the same data with the same software) bit-by-bit. Reproducibility is hard to achieve

because applications rely on a number of different software and different software versions (some at the system level and some at the application level) and access a number of data that can be distributed in the environment and can change over time (for example raw data may be calibrated in different ways as the understanding of the instrument behaviour improves).

Reproducibility is only one of the critical components of the scientific method. As the complexity of the analysis grows, it is becoming very difficult to determine how the data were created. This is especially complex when the analysis consists of a large-scale computation with thousands of tasks accessing hundred of data files. Thus the "capture and generation of provenance information is a critical part of the <...> generated data".

Sharing of knowledge, of how to obtain particular results, of how to go about approaching a particular problem, of how to calibrate the raw data, etc. are fundamental elements of educating new generations of scientists and of accelerating knowledge dissemination. When a new student joins a lab, it is important to quickly bring them up to speed, to teach him or her how to run a complex analysis on data being collected. When sharing results with a colleague, it is important to be able to describe exactly the steps that took place, which parameters were chosen, which software was used, etc. Today sharing is difficult because of the complexity

of the software and of how it needs to be used, of what parameters need to set, of what are the acceptable data to use, and of the complexity of the execution environment and its configuration (what systems support given codes, what message passing libraries to use, etc.).

Besides, these over-reaching goals, applications also face computational challenges. Applications need to be able to take advantage of smaller, fully encapsulated components. They need to execute the computations reliably and efficiently while taking advantage of any number and type of resources including a local cluster, a shared cyberinfrastructure, or the Cloud. In all these environments there is a tradeoff between cost, availability, reliability, and ease of use and access.

One possible solution to the management of applications in heterogeneous execution environments is to structure the application as a workflow and let the workflow management system manage the execution of the application in different environments. Workflows enable the stitching of different computational tasks together and formalize the order in which the tasks need to execute. In astronomy, scientists are using workflows to generate science-grade mosaics of the sky, to examine the structure of galaxies and in general to understand the structure of the universe. In bioinformatics, they are using workflows to understand the underpinnings of complex diseases. In earthquake science,

workflows are used to predict the magnitude of earthquakes within a geographic area over a period of time. In physics workflows are used to try to measure gravitational waves.

In our work, we have developed the Pegasus Workflow Management System (Pegasus-WMS) to map and executed complex scientific workflows on a number of different resources. In this context, the application is described in terms of logical components and logical data (independent of the actual execution environment) and the dependencies between the components. Since the application description is independent of the execution environment, mappings can be developed that can pick the right type of resources in an number of different execution environments, that can optimize workflow execution, and that can recover from execution failures. In this chapter we examine the issues of running workflow-based applications on the Cloud focusing on the costs incurred by an application when using the Cloud for computing and/or data storage. With the use of simulations, we evaluate the cost of running an astronomy application Montage on the Cloud such as Amazon EC2/S3.

THE OPPORTUNITY OF THE CLOUD

Clouds have recently appeared as an option for on-demand computing. Originating in the business sector, Clouds can provide computational and storage capacity when needed, which can result in infrastructure savings for a business.

For example, when a business invests in a given amount of computational capacity, buying servers, etc., they often need to plan for enough capacity to meet peak demands.

This leaves the resources underutilized most of the time. The idea behind the Cloud is that businesses can plan only for a sustained level of capacity while reaching out to the Cloud resources in times of peak demand. When using the Cloud, applications pay only for what they use in terms of computational resources, storage, and data transfer in and out of the Cloud. In the extreme, a business can outsource all of its computing to the Cloud. Clouds are delivered by data centres strategically located in various energy-rich locations in the US and abroad. Because of the advances in network technologies, accessing data and computing across the wide area network is efficient from the point of view of performance. At the same time locating large-computing capabilities close to energy sources such as rivers, etc. is efficient from the point of energy usage.

Today Clouds are also emerging in the academic arena, providing a limited number of computational platforms on demand: Nimbus, Eucalyptus, Cumulus, etc. These Science Clouds provide a great opportunity for researchers to test out their ideas and harden codes before investing more significant resources and money into the potentially larger-scale commercial infrastructure. In order to support the needs of a large number of different users with different

demands on the software environment, Clouds are primarily built using resource virtualization technologies that enable the hosting of a number of different operating systems and associated software and configurations on a single hardware host.

Clouds that provide computational capacities (Amazon EC2, Nimbus, Cumulus, etc.) are often referred as an Infrastructure as a Service (IaaS) because they provide the basic computing capabilities needed to deploy service. Other forms of Clouds include Platform as a Service (PaaS) that provide an entire application development environment and deployment container such as Google App Engine. Finally, Clouds also provide complete services such as photo sharing, instant messaging, and many others (termed as Software as a Service (SaaS).

As already mentioned, commercial Clouds were built with business users in mind, however, scientific applications often have different requirements than enterprise customers. In particular, scientific codes often have parallel components and use MPI or shared memory to manage the message-based communication between processors. More coarse-grained parallel applications often rely on a shared file system to pass data between processes. Additionally, as mentioned before, scientific applications are often composed of many inter-dependent tasks and consume and produce large amounts of data (often in the TeraByte range). Today, these

applications are running on the national and international cyberinfrastructure such as the Open Science Grid, the TeraGrid, EGEE, and others. However, scientists are interested in exploring the capabilities of the Cloud for their work.

Clouds can provide benefits to today's science applications. They are similar to the Grid, as they can be configured (with additional work and tools) to look like a remote cluster, presenting interfaces for remote job submission and data stage-in. As such scientists can use their existing grid software and tools to get their work done. Another interesting aspect of the Cloud is that by default it includes resource provisioning as part of the usage mode. Unlike the Grid, where jobs are often executed on a best-effort basis, when running on the Cloud, a user requests a certain amount of resources and has them dedicated for a given duration of time. (An open question in today's Clouds is how many resources and how fast can anyone request at any given time.) Resource provisioning is particularly useful for workflow-based applications, where overheads of scheduling individual, inter-dependent tasks in isolation (as it is done by Grid clusters) can be very costly. For example, if there are two dependent jobs in the workflow, the second job will not be released to a local resource manager on the cluster until the first job successfully completes. Thus the second job will incur additional queuing time delays. In the provisioned case, as

soon as the first job finishes, the second job is released to the local resource manager and since the resource is dedicated, it can be scheduled right away. Thus the overall workflow can be executed much more efficiently.

Virtualization also opens up a greater number of resources to legacy applications. These applications are often very brittle and require a very specific software environment to execute successfully. Today, scientists struggle to make the codes that they rely on for weather prediction, ocean modelling, and many other computations to work on different execution sites. Noone wants to touch the codes that have been designed and validated many years ago in fear of breaking their scientific quality. Clouds and their use of virtualization technologies may make these legacy codes much easier to run. Now, the environment can be customized with a given OS, libraries, software packages, etc. The needed directory structure can be created to anchor the application in its preferred location without interfering with other users of the system. The downside is obviously that the environment needs to be created and this may require more knowledge and effort on the part of the scientist then they are willing or able to spend.

In this chapter, we focus on a particular Cloud, Amazon EC2. On Amazon, a user requests a certain number of a certain class of machines to host the computations. One also can request storage on Amazon S3 storage system. This is a fairly basic environment in which virtual images need to

deployed and configured. Virtual images are critical to making Clouds such as Amazon EC2 work.

One needs to build an image with the right operating system, software packages etc. and then store them in S3 for deployment. The images can also contain the basic grid tools such as Condor, Globus, higher-level software tools such as workflow management systems (for example Pegasus-WMS), application codes, and even application data (although this is not always practical for data-intensive science applications). Science applications often deal with large amounts of data. Although EC2-like Clouds provide 100-300GB of local storage that is often not enough, especially since it also needs to hosts the OS and all other software. Amazon S3 can provide additional long-term storage with simple put/get/delete operations. The drawback to S3 for current grid applications is that it does not provide any grid-like data access such as GridFTP. Once an image is built it can be easily deployed at any number of locations. Since the environment is dynamic and network IPs are not known beforehand, dynamic configuration of the environment is key. In the next section we describe a technology that can manage multiple virtual machines and configure them as a Personal Cluster.

CLOUD COMPUTING: THREAT OR OPPORTUNITY?

Many solution providers have heard the phrase *cloud computing* thrown around as if it were the end-all, be-all

future of information technology. But cloud computing can mean many different things to different people, and the channel is no exception. For many system integrators and VARs, cloud computing seems to be more of a threat than an opportunity.

The perceived threat comes from the idea that a customer's computing resources can be converted into a virtual offering, where applications are hosted and storage is sold based upon need and capacity. What's more, that virtual offering can be located in a data centre hundreds or thousands of miles away, far from the reach of the local solution provider.

That creates a major dilemma for most small integrators and VARs: How can you compete with something that is not quite tangible and requires minimal up-front expense? The answer comes from demystifying the IT entity called the "cloud." To fully understand how cloud computing can affect IT buying decisions means, you have to take a closer look at how cloud technologies affect customer business processes. First off, cloud computing is a catch-all term that can cover many technologies — such as managed service providers, application service providers, hosted servers, hosted storage technology, hosted security solutions and other services and technologies. A typical business can combine a few of those services to build a complete hosted solution that eliminates most of the traditional networking hardware found in a business.

That may sound like a major negative to a system integrator trying to make a living off hardware sales. But today's system integrators make most of their profits from integration services and not off hardware margins. Although the cloud may eliminate or curtail the sale of hardware, it does still require integration chores, as well as training, support and maintenance — all of which can be translated into profitable revenue streams. It comes down to a shift in focus, providing services to customers and defining those services as packages that are purchased over time. Solution providers looking to profit from cloud services will have to make several decisions before venturing down the path of cloud computing. Those decisions include:

- What services to offer
- How to offer those services
- What to charge for those services
- How to bill for those services
- How to support those services

Some solution providers may choose to blaze a new trail and build their own data centres to offer specialized services. Others may want to create hosted applications and act as an application service provider, or ASP, for customers needing customized or vertical-market applications. Another opportunity comes in the form of hosted file servers, where the traditional onsite file server is replaced with a hosted file server and kept in a data centre. Each of those approaches

come with their own challenges as well as varying levels of profitability.

A Deeper Dive

Solution providers have several options for hopping on the cloud computing bandwagon. The hard part is deciding where to begin. For many, following the path of a managed service provider, or MSP, may prove to be the best starting point.

Interestingly, with managed services, a solution provider can become both a customer of cloud services and a purveyor of cloud services. The typical MSP solution consists of a channel partner providing system management services to an end customer. But, to make that happen, the channel partner must leverage an existing managed services platform, set up management applications on endpoints at the customer's site and then access a portal to support the customer via the management platform.

Ideally, the provider of the management platform will charge the solution provider for each device monitored, and the solution provider will add margins to those charges and then bill the customer accordingly. By becoming an MSP, a solution provider is still providing services to its customers without having to create a management platform from scratch to make it happen. Another path into building cloud services comes from the ASP model. Solution providers looking to transform into an application service provider can either develop a custom application and deliver it via the cloud —

often referred to as Software as a Service — or partner with another existing ASP to deliver an off-the-shelf solution.

Of course, those developing custom applications and delivering access to the application via the cloud will have more of a captive audience, especially when it comes to vertical-market solutions. But developing the hosting mechanism, application delivery mechanism and the application itself can be a daunting and expensive endeavour.

Another option is for a solution provider to sign on as a partner for an established application — such as SalesForce or WebEx — and resell those services. Perhaps the biggest potential for ASPs comes in the form of hosted e-mail systems, which incorporate archiving, indexing and searching. The need for those capabilities is driven by compliance and discovery, both of which are legal requirements for many businesses.

For solution providers that support smaller businesses, hosted file servers may make the most sense. That model consists of delivering a virtual file server via the cloud, which incorporates file sharing, collaboration, remote access and many other traditional services that would normally be present on a local file server.

The hosted file server model brings several opportunities with it — solution providers can leverage virtualization technology, quickly build a small data centre and deliver a hosted file server completely on their own. The virtual servers

can run Linux to help keep costs even lower. What's more, solution providers can bundle in additional options, such as backup, continuity services and security.

For most solution providers, cloud computing spells opportunity — it all comes down to adapting the technology to meet customer needs and building business practices around the cloud so that it can deliver profits while increasing service opportunities. Cloud computing exemplifies that, sometimes, change is good.

Potential Growth Areas for Cloud Solution Providers

With the growth in available bandwidth, the rapidly falling prices of PCs and the mobilization of the workforce, cloud computing now makes more sense than ever before. The following new services have come about that may be a good fit for mobile workers as well as the channel.

- *Security as a Service:* Many security applications have been moved into the cloud, with the latest being security suites. Some notable vendors are offering hosted security solutions for desktops and servers — the desktop solutions work with a host system to deliver anti-malware, firewall, antiphishing and several other security capabilities — fully integrated with the desktop. The advantages offered by hosted security include up-to-date protection — no signature updates needed — deep packet inspection taking place at the host and comprehensive reporting.

- *Elastic Clouds:* Companies such as Amazon are now offering services where virtual servers can be rented to run customer applications. The idea is that a customer can instantly scale as needed, have complete control over a virtual instance of a server OS and pay based upon demand. At the same time, a method like Amazon's EC2 is a way for customers to leverage cloud computing. That service can also be used by solution providers to build out virtual data centres, which then can be provisioned to create cloud services to customers. In other words, services such as EC2 can be used to create an instant cloud-related business.
- *Virtual Desktop Infrastructures:* Many businesses are looking to desktop virtualization as a method to centralize and control PC desktops. A VDI works by delivering a virtual PC down to a client device or end point. Users then have full access to that virtual PC either as a remote client or using synchronization technology. VDI is poised to become a cloud computing service, where custom virtual PCs can be created and then delivered down to an endpoint. Customers pay for access to that virtual PC, which is also constantly backed up and part of a business continuity solution.

Those are just a few examples of emerging technologies that are delivered via the cloud and that could prove to be

very channel friendly if managed and delivered properly. Solution providers may want to explore those technologies and others to find a ground floor opportunity that they can transform into a new business opportunity.