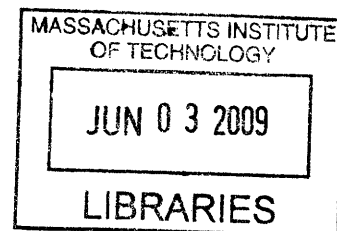# An Analysis of the Cloud Computing Platform

by

Ratnadeep Bhattacharjee

Submitted to the System Design and Management Program
in Partial Fulfillment of the Requirements for the Degree of

Master Of Science In Engineering And Management
Massachusetts Institute Of Technology
January 2009.

Thesis Advisor

Prof. Michael Cusumano
SMR Distinguished Professor of Management & Engineering Systems
Head, Technological Innovation & Entrepreneurship Group
MIT Sloan School of Management

Signature of Author_____

Ratnadeep Bhattacharjee
Systems Design and Management Program, January 2009

Certified By_____

Michael Cusumano
Thesis Supervisor
MIT Sloan School of Management

Accepted By_____

Patrick Hale
Director
Systems Design and Management Program

Abstract

A slew of articles have been written about the fact that computing will eventually go in the direction of electricity. Just as most software users these days also own the hardware that runs the software, electricity users in the days of yore used to generate their own power. However, over time with standardization in voltage and frequency of generated power and better distribution mechanisms the generation of electricity was consolidated amongst fewer utility providers. The same is being forecast for computing infrastructure. Its is being touted that more and more users will rent computing infrastructure from a utility or "cloud" provider instead of maintaining their own hardware. This phenomenon or technology is being referred to Cloud Computing or Utility Computing.

Cloud computing has been in existence in some form or the other since the beginning of computing. However, the advent of vastly improved software, hardware and communication technologies has given special meaning to the term cloud computing and opened up a world of possibilities. It is possible today to start an ecommerce or related company without investing in datacenters. This has turned out to be very beneficial to startups and smaller companies that want to test the efficacy of their idea before making any investment in expensive hardware. Corporations like Amazon, SalesForce.com, Google, IBM, Sun Microsystems, and many more are offering or planning to offer these infrastructure services in one form or another. An ecosystem has already been created and going by the investment and enthusiasm in this space the ecosystem is bound to grow.

This thesis tries to define and explain the fundamentals of cloud computing. It looks at the technical aspects of this industry and the kind of applications where cloud can be used. It also looks at the economic value created by the platform, the network externalities, its effect on traditional software companies and their reaction to this technology. The thesis also tries to apply the principle of multi-homing, coring and tipping to the cloud-computing platform and explain the results. The hurdles for both users and providers of this service are also examined in this thesis.

Acknowledgements

Table Of Contents

Chapter 1: Introduction

A major shift in the area of information technology is beginning to happen. It is being driven by a few key factors:

a) Better software technology

b) Improved broadband access

c) Desire to be energy efficient

d) Desire to dwell on key competencies and outsource the rest to a specialist.

There is a trend towards computing moving out of infrastructure housed within the company to a "cloud" that is hosted by a provider of raw computing infrastructure and in many cases core services around this infrastructure. One might say this has been happening for a long time. However, the nature of the recent change and the usage patterns are very different. So much so that it has a profound effect on how companies do business and make their infrastructure purchasing decisions. This has the impact on hardware companies that have so far thrived on selling their wares to data centers of all sizes. This also has an impact on small and medium software vendors and on startup companies who would otherwise have had to spend a fair amount of their fixed costs on infrastructure. This thesis examines all of these issues in details.

Before delving further into the investigation of the current technologies and their impact I would like to talk a bit about the history of infrastructure as a service.

Evolution

Cloud computing in some form or another can be traced back to the early sixties. Time sharing and multitasking – operating system features of all modern day computers can be

seen as the early enablers of utility or cloud computing. Without these features no form of utility computing would have been possible.

During the early stages of the computer revolution IBM was considered the leader in this field. The IBM mainframe was the first real and commercially viable system where the users desktop machine did not perform much processing besides "dumb" terminal functions. The mainframe in essence was the cloud and in general was hosted in a data center located elsewhere from where the client was. Multiple users could use the mainframe's resources at the same time. The mainframe could process multiple tasks at the same time too. IBM is still a big player in the utility computing market today. There were other corporations and technologies that could also be considered the early innovators in this field. Some of them have changed their service offerings drastically to survive while others have perished.

The seventies – Tymeshare Inc:

Tymeshare was an early mainframe and services company that rented out mainframes space and compute power via telephone lines. The company provided high level computational services to companies like Bank Americard, the IRS, etc. It was located in Cupertino, CA. The company does not exist today. This could have been due to the domination of IBM and DEC in the mainframe and mid range categories. There was no need for the kind of service that Tymeshare provided.

Hosting service providers:

The eighties and nineties saw the birth or numerous companies that hosted large main frames and mini-computer clusters. The hardware was rented out to corporations for their data center processing at fixed monthly or yearly rates. The corporations were connected to these data centers via dedicated ISDN lines. Along with computational power storage services were also provided. Customers' data was backed up in tapes and restored on demand. Usually demand for restoration of data had to be made via phone call to a

service desk. Email support was added later on. This allowed corporate data centers to be hosted outside the premises and corporations did not need to manage the infrastructure themselves. However, they had to pay a negotiated price for the infrastructure as well as the service. The cost of service depended on the Service Level Agreement (SLA) between the corporation and the provider.

Software development within the corporation was usually done on PCs in an environment that simulated the actual datacenter. Once development and local testing was done the software was "uploaded" to the real datacenter machines. A round of testing was done with real data and the then the systems we made "live". The software systems were unavailable during this process – at times the process could last days, depending on the complexity of the operation.

Application Service Providers:

The eighties and nineties also saw the entrance of application service providers or ASPs. As computation became more advanced and a lot more business applications were being implemented using IT outsourcing of these applications also picked up. [1] suggests that application service provision belongs to the third wave of outsourcing behind human resources and hardware .

Application service providers (ASPs) at that time were defined as third-party service firms that deploy, manage and remotely host software applications through centrally located services in a rental or lease agreement. Initially ASPs provided compelling value proposition to SMEs and start-up companies. It allowed small businesses to gain access to technical expertise and "best-of-breed" applications at a lower cost of ownership. Among other advantages for the customers offered by ASPs were scalability of applications over time; access to better IT expertise and state-of-the-art technologies; rapid implementation time; reduced downtime; and free upgrades. The pricing-model of ASPs provided a predictable "cash-flow" because pricing was typically based on per user per month. It also provided a scalable solution in a market place where rapid changes

were occurring in terms of technology as well as within business. The convergence of software and IT infrastructure towards an open network centric environment enabled the ASP concept to emerge. ASPs managed and delivered application capabilities to multiple entities from data centers across a wide area network (WAN). Pricing, customer service, SLAs and RAS (Reliability, Availability, Serviceability) were the parameters on which ASPs were measured.

By the year 2000 they had become the next greatest thing since applications themselves. ASP proponents were able to ride the wave of inflated expectations created by the dot.com mania. Even the big telecommunications firms such as AT&T and Sprint – jumped on the ASP bandwagon. Large IT service providers such as IBM Global Services and EDS were pushing to both host and manage applications. Some software vendors were getting more aggressive, too. Oracle, for example, expanded deployment options for its E-Business Suite Online, a hosting service for ERP, CRM, and E-business applications.

All of this failed to convince too many users. It turned out that ASPs were relatively expensive to set up and maintain. A lot of employees were required to support the customers using these applications as well as for maintaining these applications. The ASP vendors hung their hat on a weak value proposition, namely the marginal cost savings available from hosting someone else's applications. Users did the math and were not impressed with what they saw. Given the marginal economics, the questionable business models of many of the ASPs, and the consequences of failure, the lack of user enthusiasm is understandable. Today the term ASP has assumed negative connotations and is on the way to becoming a historical footnote. Most providers have exited the space. However, some ASPs are hanging on, whether from hope or simply contractual obligations.

A majority of the large ASP vendors have moved their customer support off shore to cut down service related expenses. Prime among them are IBM, EDS, and other software consulting firms. This has enabled them to stay competitive. They have also instituted

SLAs with their customers and instituted sufficient risk management and mitigation strategies.

Web Hosting:

The advent of the web saw the growth of numerous e-commerce and other web applications. In a lot of cases these web applications are hosted on infrastructure that is owned by some other entity. The application or content is hosted by one entity (the provider) and owned exclusively by one other (the user organization). The resources are off-premises, relative to the user organization, and are dedicated to it. Access to the resources is available via the Web (that is, by HTTP). Some Web hosting uses fully fixed resources allocated exclusively to a paying user organization (as in classic hosting). In other cases, Web hosting is optimized, and some underlying resources may be shared with some degree of elasticity. Thus, some web hosting extends into the cloud by virtue of being elastic, although the resources in use remain dedicated to one user organization. There are numerous small and large companies that still offer this hosting service. The service essentially consists of a machine with a choice of wither Windows or Linux operating system and an environment that supports either an application server where the application can be deployed or engines that can handle scripting languages such as Perl, Python, PhP, Ruby, etc. They also offer domain name registration, data backup and similar value added services. The users are charged a fixed rate per month based on the size of the hardware and type of software their web application needs.

Native Web Applications:

Gartner defines these as web-based applications in which the two parties are the provider (for example, Google Search or Orbitz.com) and the individual. The resources (that is, the applications and content) are off-premises, relative to the individual. The resources are dedicated to only one user organization (the provider) and are on-premises relative to it. In some notable cases (such as Google, Yahoo and Amazon), the provider offers

multiple applications and uses elastic, massively scalable methods of dynamically allocating shared resources to its many applications. Thus, these are native cloud applications. In other cases, (such as most e-commerce sites), the resources are fixed and have a limited scale capacity — these are not part of the cloud, although they are certainly Web offerings.

The first web based services started appearing shortly after the internet revolution in 1995-96. Email was one of the first real software available as a service. The Hotmail web-based email service was founded by Jack Smith and Sabeer Bhatia and launched in 1996. Hotmail was one of the first free web-mail services. Hotmail was acquired in 1997 by Microsoft for an estimated $400M, who re-branded it as "MSN Hotmail". The current version, "Live Hotmail" was officially announced on November 1, 2005 as an update to Microsoft's existing MSN Hotmail service. Over time a host of other services became available over the web. These included email, calendar, chat, talk, map services from Yahoo, Google and other players.

A very important feature of these services is that apart from displaying content on browsers these services also expose themselves using well defined Application Programming Interfaces (APIs). This allows others to incorporate these services in their web applications. This has created a slew of new ecommerce business categories. Consider the market for aggregate services like vertical job search engines such as indeed.com and simplyhired.com. These web applications scour other job related web applications for relevant jobs and publish all the results on their web site. Other sites like kayak.com aggregate data from other travel related sites like orbitz.com, travelocity.com, priceline.com, etc. Zillow.com is a real estate web application that embeds maps showing location of properties – this is done using Google Maps.

The revenue model for these web applications or web services tend to be quite simple. Most of these services are available to retail users for free. For businesses that want to use the APIs the web application provider generally charges a fee that varies from one party to another.

These web applications are the precursor to Utility and/or Cloud computing which we define and talk about in details in the next chapter. Chapter 3 talks about the economic value of cloud computing, and the changes in software development methodologies brought about by the cloud. Chapter 4 looks at the issue from a platform, and economic point of view. It evaluates the current providers on various parameters to determine their efficacy. In Chapter 5 we talk about the road ahead for this platform.

Chapter 2: Cloud Computing – The Technology

In this chapter we look at Cloud Computing from a technology point of view. The industry has defined the technology in various ways. Industry analysts and bloggers tend to differentiate between the terms *cloud* and *utility* computing. The definition[1] of utility computing has focused on a business model that provides the computing services. Utility computing service is one in which customers receive computing resources from a service provider (hardware and/or software) and "pay by the drink", much as we normally do for utility services such as electricity. Amazon Web Services is often used as one example of such a service. Industry experts define cloud computing to be a broader concept than utility computing and refer to it as the underlying architecture in which the services are designed.

Sun's CTO of cloud computing, Lew Tucker, sees cloud computing as delivery of applications across the network. He cites the example of document management system like Google Docs that delivers an application over the network. Cloud computing is an "all encompassing" term that includes any service that is delivered over the network, says Tucker.

We have decided to go with a subset of these definitions to help explaining the technology and how it is used. In the next chapter we define Cloud Computing from a business perspective. Before introducing the definitions we talk about the technologies that have been instrumental in enabling Cloud Computing as a viable business model.

The Push Towards Cloud Computing

---

[1] Based on an article in gigaom.com written by Geva Perry, chief marketing officer at GigaSpace Technologies. http://gigaom.com/2008/02/28/how-cloud-utility-computing-are-different/

Improvements in technology, particularly virtualization, provisioning, and the use of Application Programming Interfaces (API) have enabled the business of both cloud and utility computing. We explain these technological terms and then define the term cloud computing.

Virtualization allows the representation of computing horsepower as a logical entity. This virtual computer could be a single physical machine, a series of physical machines connected by a network, or part of a physical machine that has enough horsepower to be shared by multiple applications in need of computing resources. Virtualization has existed in one form or another since the days of the IBM mainframe. More recently companies like VMWare created the new generation virtualization software. Others like Xen provide a platform that companies like Sun, Citrix, Oracle have built their virtualization software. The latest virtualization offering from these companies allows different operating systems to be hosted on the same physical hardware. This has really been a boon for cloud providers. Using virtualization they can meet the needs of a diverse set of Operating System requirements with the similar hardware in their datacenter. This helps reduce their cost of management of these datacenters. For these virtual machines to talk to one another or to a network outside companies like VMWare have written software that represent virtual routers. According to a recent New York Times technology article [10] Cisco has even come up with a virtual switch that it's selling in tandem with VMWare.

Provisioning is a general term used to describe the way machines are made usable by installing an Operating System and other application software on top of it. In the area of provisioning two important things happened over the last few decades that have profoundly affected the way software was being deployed in the hardware: server provisioning and resource provisioning. Server provisioning involves selecting a server from a pool of available servers; loading the appropriate software (operating system, device drivers, middleware, and applications); appropriately customizing and configuring the system. This customization involves creating or changing a boot image for this server and changing other parameters such as IP address, IP Gateway, etc.

Resource provisioning involves associating the provisioned server with an associated network and storage resources. All of this can now be managed from a remote console. There are many software products available to automate the provisioning of servers, from vendors such as BladeLogic, IBM, Sun Microsystems, VMWare, HP among others.

The third important development that has helped the cause of cloud computing is the use of Application Programming Interface (API). According to Lew Tucker, the CTO of Sun's Cloud Computing division, this is an underrated aspect of cloud services. Internet services are accessible to other services and applications because the former expose their features as software APIs. What this means is that not only can we use a browser to interact with a service provided over the internet, we can directly connect to the service using these APIs. Technologies like SOAP, XML, WSDL have become standard formats for data and logic interchange between disparate systems. HTTP protocol is usually used as the transport mechanism. These have been adopted to build internet based services. The term RESTful[2] API is used a lot these days. This refers to a concept called REpresentational STate transfer which postulates that data within an internet service be represented as resources each with a unique URI. Using the HTTP protocol these resources can be created, read, updated, and deleted thereby manipulating resources on the web.

According to Peter Coffee, Director Strategy, at Salesforce.com, many of the reasons for moving to the cloud are not technical at all. The complexity of administering computing is rising. This is partly because governance expectations have risen dramatically, primarily due to regulations such as SOX and HIPPA. He also notes that the processors of today they are not the fastest in single threaded chips. These are multi threaded at the core. This means that tasks don't improve linearly simply by running on a machine having these new multi core chips. shared data centers can continue to exploit the

_____

[2] REST refers to REpresentational State Transfer. The aim is to express every service delivered over the web as a set of resources using universal resource identifiers (URI). The client can perform create/read/update/delete operations on these resources during the process of running the application.

improvements in chip technology that small organizations cannot hope to. The CAPEX per dollar is tilting towards shared data centers.

Defining Cloud Computing

We have decided to define the Cloud Computing on the basis of how the technology is delivered and used. We see three broad definitions:

Infrastructure as a Service (IaaS)

This is generally defined as the delivery of computing infrastructure as a service. The services include raw computing power, databases, networks, load balancers, etc. This obviates the need for a business to have its own datacenter. Rather than purchasing servers, software, data center space or network equipment, business can access these resources from a cloud service provider. The service is typically billed on pay as you go model and amount of resources consumed (and therefore the cost) will typically reflect the level of activity. A typical bill plan may be 30 cents per GB of storage used per month or 40 cents per CPU hour of compute resource used.

The infrastructure services are made available to clients via software APIs. Typically RESTful and SOAP APIs are commonly used to access these infrastructures. IaaS is heavily dependent on platform virtualization technologies for running client specific virtual machines. One user may want to run an instance of the Windows Operating System (OS) while another user may want to run an instance of the Solaris OS.

The implementation of IaaS may also include computer network including firewalls, load balancers that one would need in a typical datacenter to provide security, high performance and reliability to an application. Some cloud providers like Amazon Web Services[3] have created a large community of users who have created language bindings for the RESTful APIs that the web services expose. Other companies like RightScale[4]

---

[3] Amazon Web Services – http://aws.amazon.com
[4] http://www.rightscale.com

provide a nifty browser based user interface that allows application developers to assemble the infrastructure required to run an application by dragging and dropping widgets representing CPUs, databases, storage devices, load balancers, firewalls, etc.

The various providers of this service differentiate their services in the number of OS platforms they support, the software that comes with the OS, the pricing model and the Service Level Agreements they have.

Platform as a Service.

Platform as a Service (PaaS) is a model of computing that makes all of the facilities required to support the complete life cycle of building, testing, and delivering applications and services over the Internet from a single cloud—with no software downloads or installation for developers, IT managers or end-users. According to Wikipedia this is also known as cloudware. PaaS is the next stage of what is already available as mashups. Users can create applications by adding features from Google Maps, Google Calendar, and other web services and make these features as part of their applications. The next stage is to automate the process and that's what we see in the PaaS model. In Platform as a Service the application developer does not have to write code per se as the application can be developed by logic and visual customizations that the platform provides.

PaaS offerings include workflow facilities for application design, application development, testing, deployment and hosting as well as application services such as team collaboration, web service integration and marshalling, database integration, security, scalability, storage, persistence, state management, application versioning, application instrumentation and developer community facilitation. These services are provisioned as an integrated solution over the web. Salesforce.com's "Force.com" and "VisulaForce" platforms are the poster child of the PaaS model. This model seems like the web equivalent of Visual Basic development platform where developers could

visually assemble an application and add bits of code to customize it to their needs. In this model the customization of the application is part of the development process. According to Dave Mitchell, the CEO and founder of Bungee Labs[5], a PaaS Startup out of Utah, the PaaS model will stop the current practice of developing "here" and deploying "there". In the traditional model of software development applications are written in one environment, tested in another, and redeployed to yet another for production. In addition to the costs of building, configuring and maintaining these separate environments, applications almost always need to be maintained, which incurs even more costs along the way. In the conventional model, these costs and attendant risks fall on the application owner, and are considered part of the cost of deploying a web-scale application. In a PaaS environment, the entire software lifecycle is supported on the same computing environment, dramatically reducing costs of development and maintenance, time-to-market and project risk. PaaS should let developers spend their time creating great software, rather than building environments and wrestling with configurations just to make their applications run–let alone testing, tuning and debugging them.

Another characteristic of PaaS is the fact that it provides integration with other Web Services some of which may not be part of its offerings. For instance an application assembled using SalesForce.com's platform allows developers to connect to other native and non-native services residing elsewhere in the cloud.

Given that the entire development environment is available on a cloud that is accessible via the internet collaboration becomes easy. Needless to say the system is access controlled so that a developer can collaborate only with those she chooses to.

Software as a Service (SaaS)

This is defined as the delivery of an application across the internet. The application is hosted as a web application or service and is made available to users either through a browser interface or via well known web service interfaces. Some of the services offered

---

[5] http://www.bungeelabs.com

are free. Prime examples of such services are those provided by Google, Yahoo and other providers. Most client server based applications can be made available using this model. The browser acts as the client. Other services or non browser applications can also act as clients. A service can be hosted on any datacenter as long as it is exposed to the internet. Companies like Google, Microsoft, Yahoo provide common services like mail, calendar, mapping, etc. from their datacenters. There are various other specialized services such as CRM applications that are also offered as a service by companies like SalesForce.com. The former services are usually free unless the user is expecting one or more value added features. The latter type of services is usually offered at a monthly or annual subscription. The SaaS model eliminates the need to install and run the application on the customer's own computer, it also alleviates the customer's burden of software maintenance, and support. On the flip side customers relinquish control over software versions or changing requirements. One major problem with upgrading to newer versions of traditional software is the problem of backward compatibility with the existing data. In the SaaS model the data is stored along with the service in the provider's repository. It is the provider's job to ensure that the new version of the service being rolled out works with the existing data. SaaS reduces the upfront expense of software purchases through monthly on-demand pricing. If software is sold as a service it cannot be pirated. Hence, from the software vendor's standpoint, SaaS has the attraction of providing stronger protection of its intellectual property. The SaaS software vendor may host the application on its own web server. This job may be handled by an infrastructure service provider!

Application areas such as Customer relationship management (CRM), video conferencing (Webex, Citrix), human resources, IT service management, accounting, IT security, web analytics, web content management, e-mail, calendar are some of the initial markets showing SaaS success. A new set of SaaS providers are coming up with newer applications like Billing as a Service, and Monitoring as a Service. The distinction between SaaS and earlier applications delivered over the Internet is that the latter were developed specifically to leverage web technologies such as the browser and provide an alternate user interface to the user. The browser solved the problem of thick clients allowing applications to write a single front end that worked on all browsers. The user did

not need to install any client code. With SaaS the applications are now services that can be accessed from a browser as well as via APIs from other services. According to Pete Koomen, a product manager at Google, advances in SaaS were possible because of the inroads made by rich web applications brought about by advances in web2.0 technologies. This increased the use of web-based applications like mail, calendar, maps etc. that eventually led to these applications being exposed as services.

Common Features

All three of the cloud services defined above have some common characteristics:

Delivery Mechanism: A key feature of cloud based services is that there is no need to install or update the application or software that is being delivered. The service is essentially available "on demand". Application delivery is closer to a one-to-many model than to a one-to-one model, including architecture, pricing, partnering, and management characteristics. There is essentially one instance of the application running in the cloud.

Management: The activities of the cloud are managed from one or more central locations rather than at each customer's site. This also means that the features of the application are updated from a centralized location without the user having to do anything at all. This obviates the need for downloadable patches and upgrades.

Measurable: The usage of cloud services must be measurable. Otherwise, they cannot be billed. The way they are measured is by building telemetry as part of the service offering. At any point in time the resource utilization for a given user can be measured. Also the granularity at which resource utilization is measured is very important. This allows the provider to change bill plans from time to time without changing the underlying metering software. For instance, the service provider may choose to bill compute services at the rate of 30 cents per CPU hour for a few months. Later it may want to change the bill plan to 40 cents per hour during the day (8:00 am to 5: 00 pm, say) and at 20 cents per CPU

hour for the rest of the time. In order to charge the users in this way the service should be measurable at least on an hourly basis.

Self-healing: To the user of cloud services there should be no disruption of service as far as possible. In case of failure, there will be a hot backup instance of the application ready to take over without disruption (known as failover). It also means that when a policy is in place that says everything should always have a backup, when such a failure occurs and the backup becomes the primary, the system launches a new backup, maintaining my reliability policies.

Multitenancy: One key feature of all forms of Cloud Computing is the fact that different clients are accessing the same service or using the same infrastructure without the applications knowing about each other. This is known as multi tenancy. Multitenancy is contrasted with a multi-instance architecture where separate software instances (or hardware systems) are set up for different client organizations. With a multi-tenant architecture, a software application is designed to virtually partition its data and configuration so that each client organization works with a customized virtual application instance. The concept is similar to having more than one tenant living in a house.

When it comes to "Infrastructure as a Service" multitenancy is the direct result of virtualization. With both SaaS and PaaS the design of the data models and the applications are specifically built with a 'multi-tenant' backend, thus enabling multiple customers or users to access a shared data model. This further differentiates SaaS from client/server or 'ASP' (Application Service Provider) solutions in that the former providers are leveraging enormous economies of scale in the deployment, management, support through multitenancy.

Service Level Agreement (SLA)
As cloud services are being used more and more users are asking for some sort of accountability from the service providers. The service providers are responding by

introducing a certain minimum level of service. These guarantees are enshrined in the

Service Level Agreements (SLAs) that the providers agree to when the user signs up for

the service. The system is dynamically managed by service-level agreements that define

policies such as how quickly responses to requests need to be delivered. If the system is

experiencing peaks in load, it will create additional instances of the application on more

servers in order to comply with the committed service levels — even at the expense of a

low-priority application. Sometime the users are made to pay extra for a certain level of

service over and above the "free" SLA. If the cloud provider violates the SLA the user is

credited either financially or by offering free service for a certain period of time.

Google has an SLA for mail, calendar, docs clubbed under the Google Apps SLA[6]. The

agreement starts with the following paragraph:

"Google Apps SLA. During the Term of the applicable Google Apps Agreement, the Google
Apps Covered Services web interface will be operational and available to Customer at least
99.9% of the time in any calendar month (the "Google Apps SLA"). If Google does not meet the
Google Apps SLA, and if Customer meets its obligations under this Google Apps SLA, Customer
will be eligible to receive the Service Credits described below. This Google Apps SLA states
Customer's sole and exclusive remedy for any failure by Google to provide the Service."

Amazon has a service commitment for its Elastic Computing Services (EC2)[7] which

states the following among other things:

"AWS will use commercially reasonable efforts to make Amazon EC2 available with an Annual
Uptime Percentage (defined below) of at least 99.95% during the Service Year. In the event
Amazon EC2 does not meet the Annual Uptime Percentage commitment, you will be eligible to
receive a Service Credit as described below."

Most cloud providers are also making the availability of their systems available via APIs.

This is being used to report the status of their cloud in other websites like

cloudstatus.com:

---

[6] The full SLA can be found at: http://www.google.com/apps/intl/en/terms/sla.html
[7] Amazon EC2 SLA can be found at: http://aws.amazon.com/ec2-sla/
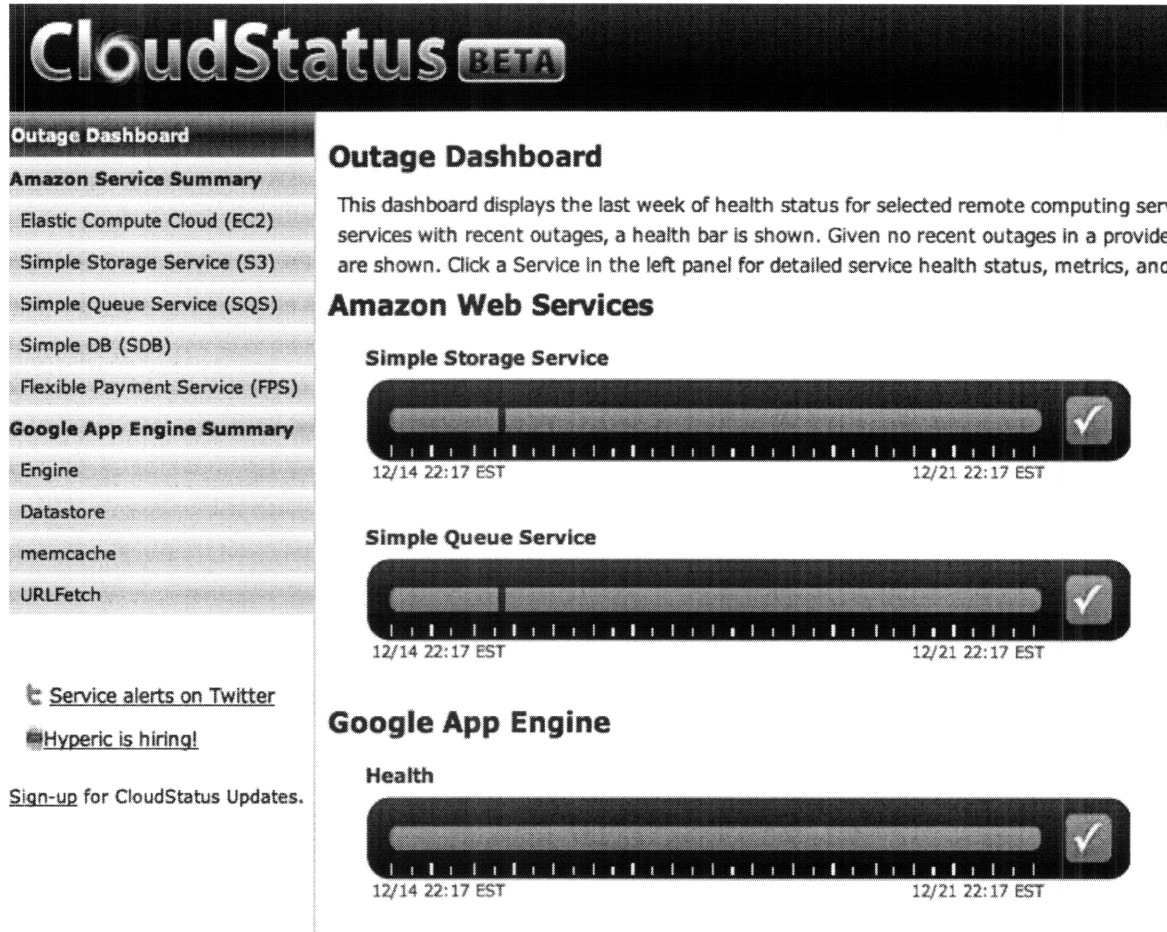
Figure1: Website showing the outage dashboard of Amazon and Google App Engine.

(Source: http://www.cloudstatus.com)

Chapter 3: The Value of Cloud Computing

Cloud computing looms larger on corporate horizon
Sunday December 21, 3:17 pm ET
By Michael Liedtke, AP Technology Writer
Cloud computing: Pie-in-the-sky concept or next big breakthrough on tech horizon?

SAN FRANCISCO (AP) -- Todd Pierce recently put his job on the line.

To meet the computing needs of 16,300 employees and contractors at Genentech Inc., Pierce took a chance and decided not to rely entirely on business software from Microsoft, IBM or another long-established supplier that would have let Genentech own the technology. Instead, Pierce decided to rent these indispensable products from Google Inc.
....

The above clipping is from a recent article published by Associated Press[8]. The case of Todd Pierce and Genentech illustrates one of the ways in which businesses are adopting cloud computing.

In this chapter we define cloud computing as a business and look at the value of cloud computing. We look at the values created by each of the three forms of cloud computing defined in the previous chapter – SaaS, PaaS, and IaaS. We also look the issues that could offset these values.

Defining The Business of Cloud Computing

Cloud computing is a business model where IT services, both infrastructure and software, are provided by external service providers, called cloud providers. These services are either free or users pay for each service based on actual usage. This allows users to concentrate on the actual task they need to accomplish leaving the all other dependent tasks, such as installation, upgrades, maintenance of software, on the cloud provider. This allows users to spend less on non-core expenditures. The service providers tend to have

---

[8] http://biz.yahoo.com/ap/081221/tec_cloud_computing.html?.v=2

many low paying customers as opposed to a few high paying ones. Owing to a large number of customers the service providers are able to amortize their cost over a large set of users bringing down the cost of serving each customer.

Cloud computing has been in the limelight in 2008. Companies like Amazon, Google, and Salesforce.com have led the way in terms of mindshare and revenue. However, revenue from cloud based services for Google and Amazon still do not come close to their other sources of revenue. Amazon Web Services provide infrastructure and core services like content management that commercial websites need. Their income from these services as reflected by the "other income" line on their income statement was 22 Million Dollars during the quarter ending September 2008 as compared to loss of 3 Million for the same period last year[9].

Google doesn't disclose the results of its business applications division, but it's relatively small, too. According to Associate Press the Mountain View-based company's non-advertising operations generated combined revenue of just $540 million during the past four quarters, while Google's advertising sales totaled $20 billion. After studying cloud computing trends, Sanford Bernstein analyst Jeffrey Lindsay predicted Google's cloud based services (Google Apps) will rake in revenue of about $1.5 billion by 2012, a small share next to the estimated $18 billion for Microsoft's desktop office software.

Microsoft recently announced Azure, their vision of cloud computing. This does not seem to be a pure cloud offering as that would preclude usage of their existing revenue generating server and office software business. Azure allows users to run software either locally on their machine or remotely in a cloud offering from Microsoft.

There are numerous smaller players offering a variety of services using different business models – these will be discussed in the next chapter.

---

[9] Source: Amazon's financial statements publicly available on their website.

Value Creation In Cloud Computing

In order to analyze the business value of cloud computing we ask the following question: Why should a business move its computational needs to the cloud? The paragraphs below explain a set of possible reasons to move to the cloud.

Better Utilization of Resources:

As mentioned in the previous chapter Virtualization is a key technology that fueled the move towards cloud computing. Non-virtualized enterprise datacenters are not 100% utilized. According to Gartner in the case of non-virtualized datacenter the utilization rate can be as low as 20% to 30% at times. This is because an application running on a given hardware configuration may not utilize all of the resources of the hardware. Also, the application may not need to be up and running all the time. Different applications may need different hardware because of the way the datacenter is organized. A cloud provider of infrastructure services would typically run a virtualized datacenter where there is better utilization of hardware. Fewer or no physical servers in the data center leads to reduced power and cooling costs, not to mention savings on server hardware and related maintenance over time. Application, backup, and disaster recovery testing is now completed in a fraction of the time it takes with purely physical servers. According to Ian Pratt[10], the chief architect at Xen, a leading provider of open source virtualization software, one of the features of the technology is the ability to transfer the location of virtual machines between different physical machines. The hosting companies love this because it means they can support more users from the same hardware by dynamically moving virtual machines around to balance the load. If there are a few users making heavy use of their virtual machines on one server, the hosting company could arrange to move them to different physical servers and then rebalance that over time. Rather than selling their server capacity to (say) 100 users, this extra "statistical multiplexing game"

---

[10] Gartner Research – publication # G00159774: Gartner Interviews Ian Pratt, Virtualization Visionary.

would mean they could sell it to 200 users. These ideas have been brought to the cloud arena leading to much better data center utilization in cloud computing business.

Economies of Scale and Economies of Scope:

According to Peter Coffee, director of strategy at Salesforce.com, the number of regulations that businesses need to comply with has increased over the years. As more and more services are being offered using software the types of compliance at the enterprise level is increasing as well. There is SOX and HIPPA, for instance. Besides, users' expectations from software have also increased a lot over the years. Users now expect their software to be more secure, responsive, and data oriented. Users of a mail service for instance would want the service to take care of controlling spams as well as preventing Trojan Horses from infecting their PCs. Users of CRM software from Salesforce.com would not like their data to be visible to a competitor who is using the same service from Salesforce.com. Enterprises that offer their services over the web – a search service, for instance, would like to prevent denial of service attacks that could bring their business to a grinding halt. All of this requires extra investment that the business would want to do without. A public cloud provider who serves more than one business has the necessary scale to employ the best of breed solutions and amortize the cost across a large set of users and/or businesses thereby giving their customers more "bang for the buck". Small businesses may simply not be able to afford a security specialist or extra hardware that could take care of securing their infrastructure.

Peter Coffee also mentions that cloud providers would get better chip utilizations than traditional data center customers. This is because the new breed of chips are essentially getting better performance from their predecessors by virtue of chip level multithreading. Applications performance does not scale linearly as these applications may not have been written in a way to take advantage of the new chipset's architecture. However, cloud providers who expose their infrastructure via an API can implement this API in a way that makes better utilization of the chipset's architecture.

New Datacenter Requirements:

For companies to derive more value out of their IT spending the enterprise IT professionals are now expected to provide a richer set of services which require knowledge of not just software and hardware but other subjects like Statistics, governance, etc. Datacenter employees need to engage business units in more of their data-related tasks, to provide strong policy leadership in information security and business process governance, and to ramp up their initiatives in data analytics and search. These are very specialized tasks that require specialized skills that companies may not be able to invest in. Cloud providers, on the other hand, would have the capability and the know how to provide these complex services. Take for example, the analytics[11] service being offered by Google. More and more businesses are using this service to get rich insights into their website traffic. This is not easy to build in house – the business would have to invest in new employees with the right skill set. They would also have to delay their time to market as this software would take some amount of time to be implemented. All of this can be avoided by subscribing to the service offered by Google or some other provider.

Lowering the entry barrier for new businesses:

New businesses that use software would typically have to set aside money for purchase of hardware, software packages that would be installed in the hardware, and human resources that can install and maintain that software. This is a sunk cost that companies can do without, especially since they do not know whether their business will succeed or not. If the business fails the money spent on this infrastructure would be wasted. If the business takes off really well the company may not have the infrastructure to support the all their customers. This may cause dissatisfaction amongst their customers causing the company to lose business. They may decide to spend more money to augment their infrastructure only to realize that the demand was very short lived. The crux of the

---

[11] http://www.google.com/analytics/

problem is that it is hard to forecast how a new business would shape up and if companies have an opportunity to use and pay for computing infrastructure based on the needs of their business it would work out to be more economical for them. This reduces the money needed to start a new business thereby reducing the barrier of entry for many businesses.

Easier to install and test new software.

In the traditional software development model the enterprise needs to set aside hardware and software resources for the purposes of testing. These resources are a sunk cost to the organization in that they cannot derive much value from these once the testing is over and the next round of testing is yet to begin. In a pay as you go model enterprises can use as much infrastructure resources for testing purposes as is needed. However, immediately after the testing is over they can let go of these resources. No additional costs are incurred and there is no idle hardware/software cost involved.

Software upgrades are also easy. The enterprise can provision a new set of servers, load the latest software and "turn it on" once the testing phase is over. While turning on the next set of servers they can turn off and de-provision the existing set of servers. This is much less cumbersome than upgrading a piece of software while it is running.

Dynamic and Granular Sourcing of Infrastructure:

According to Thomas Bittman, a datacenter analyst with the Gartner Group, enterprises can source their infrastructure needs differently now. He predicts that traditional outsourcing and hosting is not dead, but it is dying. Economies of scale are replacing skills, and speed is replacing static contracts. Once enterprises decide to use infrastructure as a service they would base their choice of vendor based on various factors – price, reliability, availability of service, support etc. Even within these constraints it is possible for enterprises to go with one provider for a certain set of infrastructures services and use another provider for a different set of needs. For instance, an enterprise could choose to

use storage services from Amazon, computing services from Google and use Salesforce.com as their CRM software provider. All of this is possible because the services are accessed via standard APIs. Although cloud providers have different APIs for accessing their services there are cloud based middleware companies that are building a uniform API layer fronting these various cloud providers. These APIs allow datacenter applications to stay the same while switching between cloud providers. While it would be hard for an enterprise to move from Salesforce.com to some other CRM service provider they could easily move to a different provider for storage and compute services. IT managers can make sourcing decisions a project at a time, and dynamically based on changing workloads, priorities and costs. This is a huge benefit that enterprises did not have before.

Reduced Risk from Project Failure:

Software projects often face the risk of delay or non-completion. Enterprise managers may realize that the project they started a few quarters back is not relevant in a new business environment. It may also be that the implementation did not go as planned and the enterprise is forced to reevaluate its choice of software, hardware, human resources assigned to this project. Risk of IT project failure is a huge concern for many business buyers that make them not want to buy any software, SaaS or on-premise. SaaS has an advantage in that if a project fails six months into a program, the SaaS buyer has invested very little and can easily switch providers, whereas the on-premise buyer has already spent millions of dollars for licenses and infrastructure and consultants.

Process Improvements:

One of the key characteristics of cloud services, particularly SaaS and PaaS, is that it reduces the amount of time users spend on doing non core activities that traditional software users have to contend with – installation, configuration management, upgrades and other change management. This allows enterprises to revamp their processes to be much leaner as compared to enterprises that use traditional on-premise software. This is a

very profound change based on my experience developing traditional software for the last 15 years. One cannot underestimate the relief of not having to continuously go through testing, debugging, reinstallation cycles. That does not mean there is no learning curve for SaaS based applications but once that knowledge has been acquired things would be much more smooth sailing. Enterprises can have a leaner workforce, as some of the work mentioned above is not there anymore. The extra workforce could be retrained to perform some other task more useful the core business of the company.

Speed and Agility of Development:

As mentioned in the previous chapter one of the key characteristics of Cloud Computing in general and SaaS in particular is that the software is upgraded much more frequently than its traditional counterpart. Oracle, for instance may release newer versions of its CRM software once a year. Salesforce.com, on the other hand deliver updates every quarter. According to John Martin[12], CTO and SVP Strategy for IQNavigator, a SaaS provider for on demand services procurement, the business buyer is faced with a choice of the following:

a) can their implementation team logs in today to a hosted test environment to start the implementation process, or

b) would they rather wait for the IT department to finish making an environment ready with multiple servers, the Oracle database, an application server, a half-dozen ancillary setups, and Internet browser access for their business partners.

The wait for option 2 is usually 3-6 months (assuming the business buyer can prioritize the project high enough for IT's involvement and mid-six-figure investment), so inevitably the business buyer would choose the hosted approach because it delivers value much faster than on-premise.

He also states that Cloud Services have several advantages in evolving with the customer's process requirements. SaaS is designed to be inherently configurable, with switches to alter functionality for a single customer in real-time. Because SaaS operates

---

[12] As stated in his blog:
http://buildingsaas.typepad.com/blog/2006/08/what_businesses.html

on a single code base and releases new capabilities every few months, new software features and innovations are quickly available to every customer without the multi-year product cycles or upgrade delays of on-premise software. In many cases, customizations can be added within a product cycle or two for customers that require functionality specific to them. In my personal experience of working with a Billing as a Service provider we saw the company incorporate new features in their product within a span of just 2 months. This helps users of the SaaS based software to change their applications sooner.

Analysis of Cost Differentiation

Providers of cloud computing services and systems integrators believe that the following cost reductions can be achieved by moving enterprise software to the cloud[13]:

- Costs savings for servers, storage, networking, software licenses, power & cooling, real estate.
- System administration labor – with the enterprise software running on an external cloud.
- Costs savings due to decreased provisioning time.
- Leaner software development workforce due to decrease in amount of software to be written.
- Shortening of application lifecycle and decreased time-to-market of new products and services.
- Increased reliability at lower costs.
- Future costs that one encounters during maintenance, downtime, upgrades or hardware, infrastructure applications, etc.

These can be quantified by the industry or vertical function in question.

---

[13] Based on ongoing discussions on the "cloud computing" Google group.
http://groups.google.ca/group/cloud-computing

Software Development Lifecycle Revisited

The difference between cloud computing and on-premise software development is bound to have some impact on the traditional software development methodologies.

For one, an enterprise's traditional software development lifecycle, agile or waterfall, can become shorter once cloud computing is adopted. This is because some of the functionalities that were performed in house would now be provided by the cloud vendor. This includes installation, patching, upgrades, etc. Most of the issues surrounding performance and scalability will rest with the cloud service provider too, especially if the software was initially being developed in house.

The quality assurance process also becomes shorter in some cases. Particularly if an enterprise which initially had the software written in house decides to use a cloud service provider for the same. The cloud provider usually supplies a sandbox to perform system testing as a result of which no time is wasted in creating and maintaining test beds. Enterprises who are moving their application to a cloud can create an on-demand test environment on the cloud very easily without having to buy any hardware. For example, if the application were to be eventually deployed on the EC2 platform, one can easily deploy the application on as many EC2 images as needed to perform the testing and then let go of that environment once the testing is over. This saves a lot of time.

According to Peter Coffee, Director Platform Strategy, at SalesForce.com, his integration partners now have a 90 day cycle for deploying CRM applications based on SalesForce.com. This is divided into three 30-day phases. In the first phase the application is designed and implemented. In the next 30 days the integrator sits with the enterprise and addresses visual and business issues as they show up – there is no rebuilding etc. to be done. Over the next 30 days his team adds additional features that are "nice to have" from the enterprises perspective. This was unthinkable with the traditional software development model. Platform as a Service, like force.com, by their very nature allow applications to be developed without too many lines of code bring

written. An enterprise simply has to customize the platform to its needs. This obviates the need for some of the major steps in traditional software development.

Between the waterfall and agile models of software development the latter tends to lend itself more to the new technology especially with SaaS. Agile chooses to do things in small increments with minimal planning, rather than long-term planning. Each such iteration involves a full software development cycle, including planning, requirements analysis, design, coding, unit testing, and acceptance testing when a working product is demonstrated to stakeholders. This is well suited to a cloud environment as the infrastructure to test, deploy or validate is already taken care of by the provider. Cloud services, at least on paper, are always available and this makes the agile goal of always having software that's working a distinct possibility. That does not mean the waterfall development model is going away – I am involved in a project where we have chosen to build an application using a third party software as a service and have used the waterfall methodology of requirement analysis, design and development. Needless to say, the process of integration simply goes away with SaaS as any code that is written is integrated with the SaaS application right from the beginning.

For software to be deployed on the cloud an enterprise has to think about the deployment strategies early in its development life cycle – at the design stage. This is because the cloud provider would have exposed its environment via a fixed set of APIs and any software deployed in the cloud has to conform to those APIs. The software should have algorithms in place that adds new virtual machine instances or new storage volumes as more data is generated by the application or more requests are being served by the application. This is in contrast to the traditional software development model where the situation is much more adhoc.

The Stumbling Blocks

According to the Associated Press Cloud Computing has already swelled into an estimated $36 billion market this year, representing roughly 13 percent of global software sales. The big question now is whether it has reached its moment of reckoning, will it

sweep the likes of Microsoft, Oracle, and other software industry staples into obsolescence? Not yet, in our opinion. There are several real and perceived problems that need to be overcome.

Performance Bottlenecks:

Performance of an application residing in the cloud is still a question that is being evaluated. According to Forrester Research 40 percent of companies consider SaaS application performance a key concern. One thing is sure, there is some amount of latency in the transit of data from the cloud to the client accessing the data which does not exist for applications running in in-house datacenters. At issue is the protocol for data interchange. Most cloud delivered applications rely on the traditional get/post mechanism of the HTTP protocol. While great at document delivery, that architecture was never really designed for dynamic user interfaces. We have added various layers on top of the basic delivery mechanism to make things easier for but the underlying delivery mechanism is still the same. Thus, we're really using well-designed, well-delivered Web sites when using SaaS applications, not true, dynamic native interfaces.
A few things can be done to improve the situation - make sure the pipe between the user and the server is as wide and short as possible. This can be achieved if the cloud provider has multiple data centers and clients are served from the data center closest to them. In some cases a geographically disbursed company may end up using several data centers or the services of content delivery networks like Akamai to provide the best performance.

Switching Costs:

The traditional software model has its own set of problems. However, it would be safe to say that enterprises have a set of applications in "steady state" that are running on traditional software packages from companies like Oracle, SAP, Microsoft etc. Enterprises would have to incur a huge cost in moving these applications to the cloud. They would have to undergo organizational changes in order to move to a cloud based computational model – fewer support personal would be required, the application

development lifecycle would change. Most organizations may not be ready for such changes.

Also, what happens to all of the existing data center hardware if applications are moved to the cloud or enterprises start using software services such as CRM from Salesforce.com? These would have to be disposed off at a discount, something companies may not be willing to do.

Business-software powerhouses Microsoft, IBM Corp., Oracle Corp. and SAP are all maneuvering to protect their existing, lucrative software franchises while also setting up their own online services to compete with the industry upstarts. Even Genentech, the biggest U.S. company to buy Google's applications package so far, isn't ready to abandon Microsoft entirely. It's still licensing Microsoft programs like Word for writing documents and Excel for creating spreadsheets.

Single Point of Failure:

Corporate data centers, websites, and ISPs are prone to security threats on a regular basis. These happen in various ways – denial of service attacks, virus infection, machine break-ins etc. If a single corporate data center is attacked the effect is felt only by that organization and the few that do business with it. If a large website like Microsoft.com is attacked the effect is felt by a larger community – all users who want to download patches or updates to existing Microsoft software, for instance. Cloud providers are not immune to such attacks. However, they are playing host to a much larger set of customers and possibly some enterprises whose business systems have completely moved to the cloud. A failure of any sort will have a much larger impact. Cloud providers try to minimize this by having multiple datacenters replicating user data and applications. Hence, total shutdown may not be imminent in the case of failure in one of the data centers. However, application performance will definitely be affected.

Privacy and Security:

The privacy and security of applications is a major issue. Many enterprises would hesitate to move their data outside the firewall due to the nature of the data or the nature of their business. Various countries or groups of countries have their own regulations around where enterprises can store their data. Canada for instance, bars businesses from storing data in the US due to the latter's draconian Patriot Act. EU has norms which mandate that certain businesses store data within EU countries. Cloud providers are getting around these by establishing data centers in more parts of the world. However, there isn't enough distribution of the datacenters to overcome these regulatory constraints. One would imagine that it will be hard to do business in China at all for a cloud provider that does not have a local data center in China.

Lack of proper Service Level Agreements:

Currently cloud providers use uptime as the only measure in their service level agreements (SLA). There are various other factors that cold come into play here. For instance, customers who are willing to pay a premium price may want to get a lower response times for their requests. According to InformationWeek[14], Microsoft, through its Azure cloud platform, is trying to create additional SLAs. These are still in their infancy. The other problem with SLAs is that it is difficult to express and implement these SLAs at a technical level. For instance, if a cloud provider promises a maximum length of time for a certain transaction, the provider cannot guarantee it simply because part of the transaction involves data flowing through the internet over which they really don't have much control.

Lack of variety of services:

One other problem that cloud consumers face is the lack of all types of services. A case in point is that of Northrop Grummen, the defense contractor. They wanted to use

---

[14] http://www.informationweek.com/cloud-computing/blog/archives/2008/10/will_microsoft_2.html

SalesForce.com's CRM application. However, the tool they used for project estimation and management, SEER developed by Galorath, was not available on the cloud. Being a very large customer they were able to convince Galorath to make SEER available as a service. Galorath ported SEER to the SelseForce.com platform thereby making it available to Northrop as a service[15]. The fact is that not all applications that enterprises use on a daily basis are available on the cloud. Until that happens it would be difficult for many enterprises to move to the cloud without considerable disruption to their business.

Vendor Lock-in:

Although the move towards cloud computing has accelerated there are still a handful of large cloud providers – Amazon, Google, Saleforce.com. There are a whole bunch of smaller players like RightScale, Hyperic, Mosso, Elastra – these startups are playing a vital role in furthering the cloud ecosystem. However, this being the inception phase of the technology some of these companies may not survive because of poor execution, or a flawed business model. Others may be acquired by larger companies. What happens to the user's applications and data in such cases? In the case of traditional software companies one always has a tangible piece of software in hand whether the company selling the software perishes or not.

High Initial Investment:

Most of the cloud providers are not able to reap the kind of profits that traditional software companies make. Mark Benioff's Salesforce.com is now the largest cloud computing service for businesses, with a market value of $4 billion, about 52,000 customers and revenue totaling $1 billion in its past four fiscal quarters. Yet according to Associated Press, Salesforce's income of $37 million during that time translates into a measly $3.70 profit on every $100 in sales. That looks anemic alongside Oracle's net margin of about $24.80 for every $100 in sales in the comparable period. San Mateo-

---

[15] http://www.galorath.com/wp/how-galorath-quantified-the-salesforcecom-platform-with-seer-for-software-seer-sem.php

based NetSuite still hasn't eked out its first quarterly profit after a decade in business, despite steady growth that boosted its revenue during the past four quarters to $143 million. The slim profit margins reflect the expenses cloud computing providers must absorb to build big data centers and hire the engineers to run their software applications, while they charge relatively modest fees to use their service. The other problem is that they don't require their customers to pay additional money for product updates and maintenance -- a gold mine for traditional software makers.

In the next chapter we look at how cloud providers can capture value, the various business models that are in play or might come into being in future. We also look at how network externalities may be a factor in this business.

Chapter 4: Cloud Computing Business Models

In this chapter we look at existing cloud computing business models. We also look at cloud computing as a platform and try to apply the levers of platform leadership to a few well-known companies in this space.

Business Models

According to Doug Hauger, General Manager of Marketing and Business Strategy at Microsoft, we are in the very early stages of business models using this technology. One major factor favoring the cloud platform is that end users seem to be fine with the idea of keeping their personal data in datacenters managed by others. Most, if not all, of us have a Yahoo, MSN, GMail, or some other public email account. A lot of us use the internet for banking, bill payment, and shopping and readily provide our credit card number, social security number, PINs, to these web sites in return for convenience. The convenience comes in the form of availability and ubiquity of services, ease of use, etc. On the other hand enterprise users are forced to either be within their firewall or connect from outside via VPN (Virtual Private Network) to access their office email, calendar etc. Individuals are beginning to ask for the same level of ubiquity for their enterprise data and applications as they have for their personal accounts. This may push more enterprises for cloud based solutions. Companies like Google are actively courting enterprises by explaining the aforementioned virtues of the cloud.

Cloud computing can be looked as a stack of services as shown in the figure below – each layer of the stack offers services that are built on layers below it. There are buyers for these services at different levels of the stack depending on the needs of the business, the size, and competency of those running the business. Each layer in the stack has the potential to offer a different business model. There are some core services that span

across the various stacks. For instance, companies that supply metering and billing services are needed for all types of cloud services. So are management services.

Metering and Billing Services (Amazon DevPay, Zuora)			Workflow Engine	Visual Authoring	Management Monitoring, Deployment Services. (RightScale, Cloudstatus)
	SaaS: Google, Yahoo, MSN Live, Salesforce.com				
	PaaS: Google App Engine, Force.com, CogHead				
	IaaS: Amazon Web Services, Sun Microsystems, EMC				
	Mail/ Calendar/ Docs	Analytics			
	Search	Analytics			
	Identity Services, Messaging Services, Content Management Services				
	Compute Services, Storage/Data Services, Network Services/ Load Balancers				

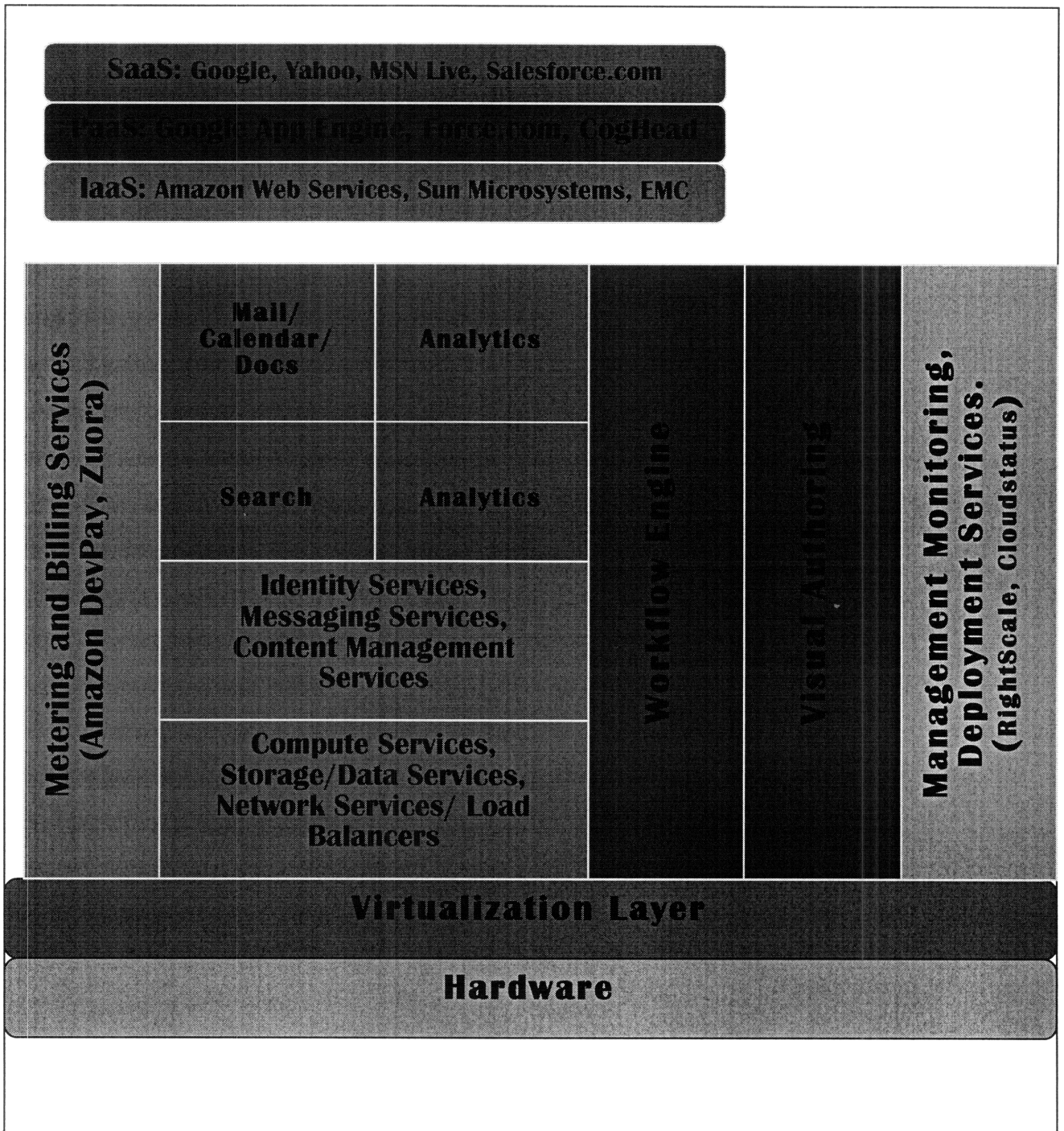**Virtualization Layer**

Hardware

Fig: Stack of cloud Services.

Infrastructure as a Service (IaaS) providers such as Amazon, Google (with their App Engine), Sun Microsystems, EMC, offer their service in ways that differ in terms of what the user gets and how she is able to access the services. Amazon for instance makes a machine instance available to developers. They have to choose the operating system (Linux, Solaris, or Windows). Each such machine is referred to as Amazon Machine Instance (AMI). Once the user gets an AMI she is free to install her own applications. Amazon also makes storage, database, content management and other core services required for someone to build a commercial web application for conducting business online.

Google on the other hand makes core functionalities of its systems available only via a Python based software interface. Google does not provide access to an OS image or any standard database. They provide APIs to access services like mail, a proprietary datastore, images, users etc. Users are expected to write applications using these APIs. That way Google is much more restricted than Amazon's offerings. SalesForce.com's "Force.com" is similar to Google in that they allow the developer to build an application using their workflow engine and visual authoring tools and leave the underlying implementation of the application to the cloud.

Each of the two offerings above sustains a different group of users. Small and medium businesses that do not want to spend money on data centers and yet want to control how they configure their systems would go to cloud providers such as Amazon. There are others who need to create a web site that needs services like mail, user management, a data store and don't really care about the inner working of these services. For them the Google App Engine would suffice.

According to Lydia Leong, an analyst with Gartner, virtualized hardware alone is not enough for companies to realize the true value of the cloud. A managed, automated way to architect, deploy, scale and maintain software which supports applications in virtualized environments is needed, along with infrastructure software that enables organizations to actually use the cloud to its fullest potential. The industry has created

numerous "secondary cloud" providers who rely on the services of infrastructure providers like Amazon. They provide value by supplying management software that makes it easy for applications to provision cloud infrastructure dynamically. Without these services customers would have to write additional software to manage the deployment of their applications to the cloud – software to add or remove compute resources depending on the load on the application. Most of them provide an application stack and/or a graphical user interfaces that developers can use to easily assemble a virtual datacenter to run their application on a cloud. These secondary cloud service companies tend to charge a premium price from their customers. For instance, if Amazon's EC2 charges user 30 cents per "raw" CPU hour of usage, Elastra charges 50 cents for a CPU hour of a virtual machine that comes preinstalled with an application server.

Quite a few of these secondary cloud companies are funded by companies that already have large datacenters and are using the cloud to expand their market. One of them is GoGrid, a provider of cloud servers supporting a variety of Linux and Windows operating systems with preinstalled software, such as IIS, Apache, PHP, Microsoft SQL Server, and MySQL. It is a subsidiary of ServePath a leading provider of server and network hosting services. Similarly Mosso, another IaaS provider is backed by Rackspace, an industry leader that currently manages over 30,000 servers for customers all over the world.

Some of these secondary companies are also funded by companies like Amazon. Elastra, for instance, is partially funded by Amazon. The company helps deploy and manage applications on Amazon's EC2 and S3 platforms. RightScale is another company that creates web solutions that run on Amazon's cloud. This has two benefits for Amazon – it provides them with additional customers who come to Amazon by virtue of using Elastra's management software, it also furthers the cause of Amazon's APIs if more businesses like Elastra build their management software to work on the Amazon's cloud.

Next we look at how the Cloud Computing business model is different from the traditional software model. We also look at how the various cloud providers differentiate themselves in the marketplace.

High volume, low subscription fee:

One of the values of this technology discussed in the previous chapter was that, yes, service providers can spread their costs over a large number of users. However, the payments from each user are likely to be small, so this is a very different and potentially more risky business model than the traditional product business. A common cliché used by industry analysts is that there are likely to be a million customers paying $100 per month for using cloud services like Amazon EC2 and S3 as opposed to a 100 million Dollar customers. There is some truth in this cliché. This is one reason cloud providers are trying to reach out to enterprise customers where they can sign contracts for a specific period and be assured of revenue during that period. Enterprise users are also likely to buy support services which is something that has high margins.

Revenue Source:

The revenue model for cloud providers is quite different from traditional software providers in that the former do not charge an upfront cost for selling the software. There are no additional license renewal fees either. While customers like Genentech are buying enterprise licenses for cloud services like Google Apps, the bulk of the consumers of these services are non-paying individuals, educational institutions, charitable organizations. Companies like Google, Yahoo, Microsoft (with its Live Platform) makes money by displaying relevant advertisement on the same browser window where the mail or the calendar is being viewed. Google for instance, uses technology to parse the text of the emails received and displays advertisements related to the contents of the email. In fact, Google, which does not sell any traditional software, had its revenues from advertisements cross 20 billion Dollars over the last 4 quarters (as reported by Associated Press). In summary, advertisement is a major source of revenue for SaaS providers who

offer their services for free. In most cases these same services are also offered at a price with no advertisement to small and medium businesses. The same holds true for enterprises users who usually have to buy licenses for each user in the enterprise.

Other software services like maps, analytics, etc. are also available to enterprises for a price. The price covers customer support and technical help. This is a small segment of the revenue model. Companies like Kayak.com, Trulia.com, Tripadvisor.com all use Google Map's enterprise addition for a fee[16].

Companies like Salesforce.com provide commercial applications like CRM as a service. They are now the largest cloud computing service for businesses, with a market value of $4 billion, about 52,000 customers and revenue totaling $1 billion in its past four fiscal quarters. Other companies in this space are CogHead.com and Bungee Labs.

Pricing Models:

The way cloud services are priced differs from one provider to another. As mentioned above a large chunk of the software delivered as a service, GMail, Google Calendar, Search, Maps, Yahoo Mail, etc., are available to retail customers for free. There are corporate customers who pay for the services. In fact Google has been actively trying to push these services under a single bundle, Google Apps, among corporate customers for $50 per license per year.

In the case of Platform as a Service (PaaS) Slaesforce.com, the pricing for unlimited application usage of those applications is $50 per user per month. However, the company has faced price pressures from smaller startups like Bungeelabs.com, coghead.com, etc. and has been forced to come up with alternative revenue models. According Ariel Kelman, Salesforce.com senior director of platform product marketing, as quoted on ZDnet, there are a wide variety of apps on the Force.com platform, some that have high

---

[16] http://www.google.com/enterprise/maps/success_stories.html

usage, such as recruiting, and others such as vacation request, employee surveys and expense reports, are widely distributed but less frequently used. To address that segment more equitably, salesforce.com is charging $5.00 per login, with a maximum of 5 logins per user per month. They also set the metered pricing at a promotional price of 99 cents per login throughout 2008.

The predominant pricing structure for Infrastructure as a Service (IaaS) seems to be a "pay as you go" model. Users of the services only pay for the resources and services they use. Amazon EC2 charges for resource usage as shown in the screenshots below. The user pays for the machine on a per CPU-hour basis.

Pricing

Pay only for what you use. There is no minimum fee. Estimate your monthly bill using AWS Simple Monthly Calculator.

Instances

United States   **Europe**

Standard Instances	Linux/UNIX	Windows
Small (Default)	$0.10 per hour	$0.125 per hour
Large	$0.40 per hour	$0.50 per hour
Extra Large	$0.80 per hour	$1.00 per hour
High CPU Instances	**Linux/UNIX**	**Windows**
Medium	$0.20 per hour	$0.30 per hour
Extra Large	$0.80 per hour	$1.20 per hour

Pricing is per instance-hour consumed for each instance type. Partial instance-hours consumed are billed as full hours.

Data Transfer

Internet Data Transfer

The pricing below is based on data transferred "in" and "out" of Amazon EC2.

Data Transfer In	
All Data Transfer	$0.10 per GB

Data Transfer Out	
First 10 TB per Month	$0.17 per GB
Next 40 TB per Month	$0.13 per GB
Next 100TB per Month	$0.11 per GB
Over 150 TB per Month	$0.10 per GB

Data transferred between two Amazon Web Services within the same region (i.e. between Amazon EC2 US and another AWS service in the US, or between Amazon EC2 Europe and another AWS service in Europe) is free of charge (i.e., $0.00 per GB). Data transferred between AWS services in different regions will be charged as Internet Data Transfer on both sides of the transfer.

Usage for other Amazon Web Services is billed separately from Amazon EC2.

Amazon Elastic Block Store

United States   **Europe**

Amazon EBS Volumes

- $0.10 per GB-month of provisioned storage
- $0.10 per 1 million I/O requests

Amazon EBS Snapshots to Amazon S3 (priced the same as Amazon S3)

- $0.15 per GB-month of data stored
- $0.01 per 1,000 PUT requests (when saving a snapshot)
- $0.01 per 10,000 GET requests (when loading a snapshot)

Elastic IP Addresses

No cost for Elastic IP addresses while in use

- $0.01 per non-attached Elastic IP address per complete hour
- $0.00 per Elastic IP address remap – first 100 remaps / month
- $0.10 per Elastic IP address remap – additional remap / month over 100

(Amazon EC2 is sold by Amazon Web Services LLC.)

Figure: Amazon EC2 Pricing Plan. Price as on 01/10/2009.

(Source: http://aws.amazon.com)

As can be seen from the figures above the developer pays for CPU, data transfer, and IP address mappings beyond a certain limit. There is a similar pricing model for storage services from Amazon. Secondary providers of cloud services, companies like Elastra, charge a slightly higher because of the additional value they provide. They charge higher for the virtual computer instances that come with a predefined application stack such as an application server.

The table below provides an example of how pricing for our solution works:

Software Components*	ECU Rating (A)	Cost/ECU/Hour ($) (B)	Component Hourly Rate (AxB=C)	# of Instances Consumed (D)	Total Hourly Cost ($) (CxD=E)
Database Instance	1	$.50	$.50	1	$.50
Application Server Instance	1	$.50	$.50	1	$.50
Software Load Balancer Instance	.25	$.50	$.13	1	$.13
Software Charges Sub-Total					**$1.13**
Server Charges**			Cost/Unit/Hour ($) (A)	# of Instances Consumed/Hr (B)	Total Cost ($) (AxB=C)
AWS Small Server Linux/UNIX Instance**			$.10	1	$.10
AWS Large Server Linux/UNIX Instance**			$.40	2	$.80
Server Charges Sub-Total					**$.90**
Storage & Network Charges**	Unit Quantity (A)	Cost/GB/Month ($) (B)	Total Monthly Cost ($) (AxB=C)	Total Hourly Cost ($) (C/720)	
Elastic Block Store (GB)**	100	$.10	$10.00	$.014	
Data Transfer (GB)**	1000	$0	$0	$0	
Storage & Network Sub-Total				**$0.14**	
Total Hourly Charge**	Software Charges + Server Charges + Storage/Network Charges			**$2.17**	

Elastra provides up to 1TB of Data Transfer per month at NO CHARGE

Fig: Elastra's price list. (Source: www.elastra.com)

Leveraging existing business:

Cloud computing, being at the very early phase of a normal business life cycle, comes with its own set of risks. The cost of setting up a data center is quite high. What happens if you don't get enough customers after you have invested heavily in a data center? As a result most companies who have entered this business either had an existing data center business or had a cloud centric internal data center for their own business as in the case of Amazon. The company had built a vast swath of infrastructure technologies to support their booming internet retailing business. Initially this platform was made available to other vendors who wanted to sell their wares from Amazon.com. With time, sections of the platform were opened up as services like EC2, S3 etc. Salesforce.com seems to be the first exception. Huge investments in data centers might be one reason why their profit margin is so much lower compared to companies like Oracle and Microsoft. Salesforce's income of $37 million during the year 2008 when it had total revenue of $1 billion translates into a measly $3.70 profit on every $100 in sales. That looks anemic alongside Oracle's net margin of about $24.80 for every $100 in sales in the comparable period.

Google already had a number of very large data centers to provide search results. These were replicated for ads, mail, calendar services etc. With the App Engine, Google opened up its datacenter to the outside world using a subset of the APIs that their internal services used. Companies like Sun, HP, EMC, already run data centers for their customers. Second tier cloud service providers are either backed by large server hosting companies or are funded by mainstream cloud providers like Amazon as described earlier in this chapter. As the technology gathers traction we see more of the same trend. Companies like Verizon, AT&T who have a significant amount of data center assets are getting into the cloud business. It is unlikely that someone not connected with datacenters will enter the cloud business anytime soon.

Platform Leadership in Cloud Computing

In this section we look at Cloud Computing from a platform point of view and consider the factors that could contribute to platform leadership in this technology. Platform leadership involves driving industry wide innovation for an evolving system of separately developed pieces of technology. We have seen numerous instances of platform leadership in the recent past – Microsoft with its Windows Operating System, Google with its search platform are some good examples. According to Professor Michael Cusumano and Annabelle Gawer, there are four levers of platform leadership[17]: scope, product technology, relationship with complementors, and internal organization. We analyze a few of the major players in this arena against these 4 levers.

Before that we need to understand if Cloud Computing has the potential to be a platform? Most of the industry leaders we spoke with believe it is a platform or will be one much like software operating systems (like Windows, for example.) Microsoft's Hauger believes the technology is at the initial stages of becoming a platform.

We believe it can be - if we separate the services available on the cloud from the platform on which these services reside. Products and services that bring together groups of users in two-sided networks are platforms. They provide infrastructure and rules that facilitate the two groups transactions and can take many forms. Cloud computing can be envisioned as a two sided market that connects infrastructure providers to application developers. So far though, many of these services are being made available by the platform provider themselves. Amazon, for instance, is providing the compute, storage, database services, content management services, etc. However, for some of these services it has partnered with the vendors of the software. It has partnered with Sun to offer EC2 images of the Solaris Operating System. It also offers Windows images in partnership with Microsoft. There are a slew of value added service providers who provide various EC2 images with custom application stacks – these include MySQL database servers,

---

[17] The Elements of Platform Leadership, Michael A Cusumano, Annabelle Gawer. Sloan Management Review, Spring 2002.

application servers, content management systems, mail servers etc. These are turning out to be quite popular among the developer community looking for specific application stacks. We take these as the beginnings of a platform in this space.

Providers of platforms for two-sided networks are able to draw revenue from both sides. In the case of cloud platform providers, that would be infrastructure providers like those who add their own application stack to Amazon EC2 virtual machines and developers who use these stacks. In this case it makes sense for Amazon to subsidize the supply side of the platform as these providers of value added services bring more developers to Amazon. This in turn gets more providers to deploy specialized stacks on Amazon EC2.

Since cloud computing is at its infancy and different players entered the market using different technologies or domain expertise the current market has multiple platforms. That is not to say we will not have platform providers who would seek market dominance at some point. However, these days the IT community is much more aware of the tricks of the trade in the business of software. Customers for instance understand the value of interoperability, and the need for open APIs. Also, platform providers are trying to woo customers who are already invested in other cloud based products. For instance, businesses that use the CRM service from SalesForce.com are being wooed by Amazon and Google. They are doing so by supporting Salesforce.com's APIs on their platform. This shows that the various platform providers are sharing the platform, at least for now.

According to Bert Armijo, VP of Marketing at 3tera.com, vendors often use APIs as a lock-in mechanism, of-course, and the API wars for cloud are just getting started. However, this strategy can backfire. First, customers see this coming. In cloud, for instance, large users are already starting to push for standard APIs. Unfortunately standards take time and agreement amongst the vendors. Second, introducing new APIs implies that on-boarding software requires writing new code. He sites the example of EC2 which, for the first two years of its availability, did not allow databases to run reliably because its storage system was "unique."

He also believes that new APIs really aren't needed in all cases. Cloud systems that support general purpose computing (as opposed to SaaS and PaaS) use virtualization to allow running existing operating systems that already offer abstractions for memory, storage and network. Although difficult, it is possible to leverage these abstractions. Unfortunately, many cloud vendors chose to "augment" them with their own - hence the APIs. 3tera's AppLogic has no API, so there's absolutely no code writing required to make use of the system thereby reducing the barrier of entry.

Multi-homing costs can be quite high in the case of cloud platforms. For instance, just as it would not make sense to write some of your applications using .Net, some in C++, and some in Java, it would also be quite expensive to write, maintain, and deploy applications on multiple cloud platforms. This is because the APIs are different from one provider to another. A number of secondary cloud service providers are trying to provide solutions that bring down multi-homing costs. It remains to be seen how the landscape changes. The holy grail of Cloud Computing seems to be a desire among consumers to be able to choose whichever cloud they want, whenever they want. Much like the electric grid where your utility companies could choose to buy power from the cheapest source. Cloud users would like to run their computation and store their data in the cheapest possible location and possibly move them around as the price offer by service provides changes.

Even after a cloud provider does a good job of addressing pricing and winner-take-all challenges he may face the danger of envelopment where one provider's services may be "enveloped" by an adjacent provider that enters the market. Cloud platforms have overlapping user bases. Salesforce.com's APIs are now available as part of Google's App Engine API. Developers can deploy their applications on the App Engine and use SalesForce.com's CRM features. There is nothing to prevent Google from making similar features available to these customers in future. The same holds for the Azure platform from Microsoft. Initially they may open up their platform to allow developers to use other service providers but in course of time these could be closed down as Microsoft rolls out its own services. Another envelopment threat is posed by second tier cloud providers – the likes of Elastra who are providing an abstraction layer over Amazon's EC2 and

building their own internal cloud solutions. If these secondary platforms can gain enough traction they have the capability to render the core IaaS providers as mere commodity suppliers.

Levers of Platform Leadership:

As mentioned above the four levers of platform leadership are scope, product technology, relationship with complementors, and internal organization. Let us briefly look at what these mean in the context of cloud computing.

Scope:

Scope refers to the range of services and the level of services that a cloud provider might offer. We defined cloud computing to include SaaS, PaaS, and IaaS. Companies could operate in one, two or all three of these spaces. For instance, Google operates in PaaS by virtue of its App Engine offering. It also has numerous SaaS applications: mail, calendar, map, analytics, etc. Amazon on the other hand is perceived as operating only in the IaaS space. Even within each type of service the scope can vary. For instance one provider may provide a single service while another may provide a host of services. SalesForce.com, for instance, specializes in CRM as a service and has a platform around that. For everything else it depends on other cloud providers. The scope at which an organization operates has a great impact on whether it has the potential to become a platform leader.

Technology:

Cloud providers can establish platform leadership on the basis of their technology. One aspect of this technology is the user interface. User interface to access the cloud can act as a barrier to entry. Cloud services involve writing software slightly differently to take advantage of the cloud. One has to make use of application programming interfaces (APIs) to access the cloud. Usually SOAP or REST based APIs are supplied. Providing a

language binding (in PHP, Java, Ruby, Javascript, etc.) for the REST or SOAP based APIs goes a long way in reducing the barrier to adoption of a cloud platform. In many cases the providers themselves provide all the language bindings. In other cases, the community creates language bindings and makes them available to other developers. Google's App Engine seems to be a lone hold out in that it only allows usage of the Python programming language. According to Microsoft's Doug Hauger, the Azure cloud platform would allow developers to build their applications in most major languages including PHP and Ruby. Coming from Microsoft this seems like a major change in mindset. Yet another advantage can be having a Integrated Development Environment (IDE) that automatically supports a particular cloud platform. Aptana, a software company that offers an IDE for web2.0 development extended its features by allowing users to store their source code in the cloud during development and also deploy their completed application on their newly created cloud platform. Amazon, for instance, has an EC2 plugin for Eclipse IDE.

Providing a set of services that application developers use to write traditional software is another way of luring them to the cloud. For instance, most enterprise applications need access to a database, identity services, messaging services, content management especially if it's a web application. Providing these services gives a big boost to the platform.

According to Lew Tucker, the value of a vendor's cloud depends both on applications and platforms. He believes that platforms occur at multiple levels. One platform could host another software platform. For example RightScale provides a management framework that application developers can then use. Smugmug.com may use that management framework for their application in the cloud. The old Independent Software Vendors are slowly going to become "Independent Service Vendors" or "Internet Service Vendors", says Tucker. One could have a "Ruby and Rails" service provider who provides a platform for hosting Ruby applications on the cloud. This calls for a service provider interface that other providers of services as well as platforms could leverage.

Technology could also be used to induce to a given cloud platform thereby increasing the switching cost to other cloud vendors. One way of doing this is by storing data in a non standard format. Amazon is already doing it by providing services like the simpleDB, which stores data in a proprietary format. On the other hand if data is stored in the cloud using MySQL or some other relational database it can easily be imported to another cloud that also supports relational databases. The same holds true for the APIs. Each provider has its own flavor of APIs to access its cloud. Hence, an application written to live in one cloud cannot easily be ported to another cloud.

Relationship with complementors:

This is a vital part of establishing platform leadership. It would be foolish for cloud providers to build all the services themselves. They should encourage the traditional ISVs to enable their software on the cloud provider's platform. The good things about cloud computing is that multiple implementations of the same service can exist on the cloud. For instance, the same cloud can provide both MySQL and Postgres database services. In fact, the more the merrier, as it gives application developers greater choice.

Internal organization:

The right internal structure for success in the cloud technology seems to be having internal organizations that themselves believe in the cloud model. Specifically, if all internal systems are deployed as cloud services it definitely helps the company develop better understanding of the technology issues first hand. This also helps generate enough knowledge base that can help gain developer mindshare outside the company.

In the next few paragraphs we analyze a few cloud providers against the four levers described above.

Amazon Web Services.

Amazon is by far the most successful provider of Infrastructure as a Service in the world today. The following types of businesses are leveraging Amazon's cloud:

Application Hosting

Backup and Storage

Content Delivery

E-Commerce

High Performance Computing

Media Hosting

On-Demand Workforce

Search Engines

Web Hosting

Clearly their customers cover a wide variety of businesses. However, their scope is limited to providing infrastructure services only. They do not have application services like the ones Google or Salesforce.com does. Amazon has a large number of developers too. Currently there are over 29 billion objects stored in Amazon S3 vs. 22 billion at the end of Q2 2008. That is sequential growth of 32 percent. On Oct. 1 2008, the service peaked at over 70,000 requests per second to store, retrieve, or delete an object. Over 400,000 developers have registered to use Amazon Web Service[18].

From a technology point of view they have created a Firefox plugin, called ElasticFox, allows developers to view details of their resource usage and other account information. Their APIs have various language bindings supplied by members of a vibrant developer community.

In order to get greater developer mindshare Amazon organizes various cloud camps across the US and EU. They have also increased the scope of their offerings by making EC2 and S3 services locally available to European customers.

---

[18] Source: http://gigaom.com/2008/10/09/amazon-cuts-prices-on-s3/

In terms of service offerings, Amazon, being a web retailer, has made available some of the key services that application developers need to build web services. These include storage, compute service, messaging, content management, billing. They could add other services like identity as part of the service.

Amazon has funded a few companies that build value added products leveraging their cloud platform. They have also provided services like Amazon Devpay that allows external vendors to provide value added services on EC2 and S3 without having to spend any time managing metering and billing for these services. This has fueled the availability of various application stacks on the EC2 platform.

Organizationally, the company has been using these services internally for its business for a while. This infrastructure was also available to its partners. This gave them the opportunity to hone their technology for prime time. There is no conflict of interest within the various organizations as having a great cloud infrastructure helps their ecommerce business as well.

Amazon has the potential to be a platform leader by being an early mover in this space and doing the right things to establish platform leadership.

Google:

Google does not sell any of its software using the traditional model. It started off by offering search as a service and over time built various services around it – primarily in the area of search based advertising and analytics. Over time Google offered a few pioneering SaaS applications: Google Maps which provided a very user-friendly web interface to view maps. Then came Google Mail and its USP was the amount of free storage allocated to each user – 1 GB! Google Docs was introduced thereafter as collaborative way to create and share documents. To support all of its software as a

service offerings Google built massive datacenters with proprietary technology. These datacenters helped them drive down the cost of computation andstorage. It also gave them the opportunity to offer this same computational power to developers outside of Google – using the Google App Engine.

As you can see Google started off with a very small set of services but over the last 8 years increased its scope to include both SaaS and PaaS. While Google doesn't disclose the results of its business applications division, it's relatively small, too. The Mountain View-based company's non-advertising operations generated combined revenue of just $540 million during the past four quarters, while Google's advertising sales totaled $20 billion. While Google still makes most of its money from advertising Sanford Bernstein analyst Jeffrey Lindsay believes Google's applications will rake in revenue of about $1.5 billion by 2012, a small share next to the estimated $18 billion for Microsoft's desktop office software. In terms of scope Google has a large set of applications waiting to take off and new ones are constantly being churned out in their labs.

From a technology point of view Google has been a pioneer in making the web richer and more user friendly. From Google Maps, to Google Web Toolkit, to the scalability and efficiency of their data centers they have lead the way in innovation. However, Google offers a rather restrictive view of the cloud than, say, Amazon. This may not be suitable for those enterprises looking for their choice of operating system and application stack. This could come in the way of establishing them as a platform leader. Google has also come out with a browser, Chrome. The browser could be used as a vehicle to showcase their services by making these available as part of the browser's user experience.

Google maintains a very good relationship with the developer community by being very open with its APIs right from the time it started. Its philosophy of "do no evil" has endeared them to the community. The community has flourished and members have contributed by providing various plugins for popular IDEs, tools that further reduces the barrier to entry for developers. The architecture of its data centers is though a zealously guarded secret. One other characteristic of Google is that the company builds its own

services. It does not have the notion of an SDK for service developers as such. Some services were added by acquisitions of smaller companies. So, while on the demand side of the platform they have a very large developer and user community on the supply side its all Google. This may hamper them in the long run if ISVs start offering their services in a competitor's cloud platform.

In terms of organization the company has been built from the ground up to think of Software as a service. The company is relatively new and, hence, is not burdened by traditional software business models.

Salesforce.com:

Salesforce.com was founded with a motto of "No Software" to reflect the alternate delivery model of software as a service. The company started off as a provider of CRM software as a service. In course of time it created a platform around this by establishing AppExchange – a repository of applications that other companies build using Salesforce.com's CRM APIs. Users are charged a flat fee for accessing these applications. Recently it has also come up with it PaaS offering called Force.com that allows developers to build and deploy their own application on Salesforce.com's cloud. The scope of the platform has definitely increased over the last few years - from allowing other vendors to create apps using the platform to allowing ordinary developers to create and deploy applications.

In terms of technology Salesforce.com is fairly limited (as compared to Google or Microsoft, say) due to their focus on CRM applications only. In order to encourage more developers to adopt their platform, Salesforce.com have also announced a $1 million dollar developer challenge as part of their "Tour De Force" visiting 20 cities as a global developer road show. Also, Salesforce.com does not have any pre-existing technologies that could have given them ready access to a set of customers. Other cloud vendors like Google and Amazon Web Services are adding support for Salesforce.com's APIs from their platforms. This may help the company be more ubiquitous.

Salesforce.com charges their customers on both sides of the platform. Just as developers are expected to pay a fee, ISVs writing applications for the AppExchange platform are also expected to pay a fee. This may lead to ISVs going to some other platform if those platforms offer better financial incentives. Two of the major players in this business, Oracle and SAP, have not yet started offering cloud offerings on a large scale. The ERP/CRM SaaS marketplace may change considerable once that happens.

From an organization point of view they are similar to Google in that there is no traditional software technology that the company has to unlearn in order to develop their cloud based platform. This definitely helps.

Microsoft:

On the face of it Microsoft seems like the one company that would benefit the least from pursuing a SaaS strategy. Bulk of their business is based on the traditional software model and provides a large and steady stream of revenue. Their office suite generated 18 billion Dollars of their revenue last year. The only thing going against Microsoft is that the growth in the cloud computing market is going to be huge and most of it will come at the expense of traditional software market. According to a quote in sys-con.com, Merrill Lynch analysts reckon that by 2011 the volume of cloud computing market opportunity will amount to $160BN, including $95N in business and productivity apps (e-mail, office, CRM, etc.) and $65BN in online advertising. All of this has prompted Microsoft to recently launch their Azure platform. According to Doug Hauger of Microsoft, they are not late in the game – it's just that their customers did not ask for a cloud based solution all these years but are asking now.
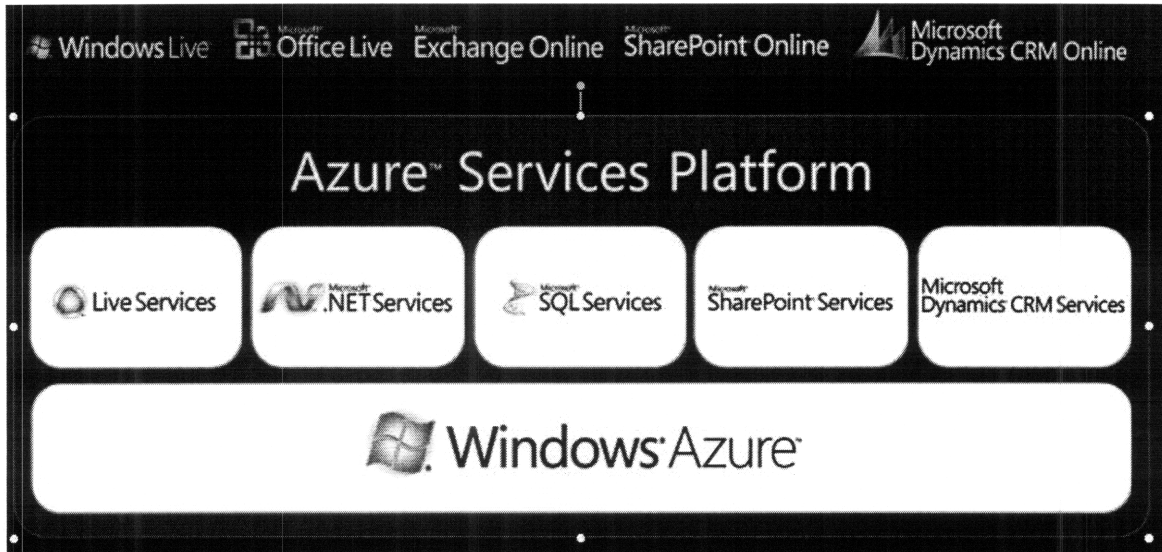
Fig: The Azure platform.

(Source: www.microsoft.com)

In terms of scope most technology analysts believe Azure lies somewhere in between the offerings of Amazon and Google. While the platform does not provide direct access to a virtual computer instance it does provide access to a host of infrastructure services from Microsoft: Live services, SQL services, CRM services, .NET services and Sharepoint Services. Azure hosted applications are expected to use these services. The platform also comes with some "finished services" built on top of these services: Windows Live, Office Live, etc. However, we must note that not all of these have been implemented yet. Sharepoint and Microsoft Dynamics Services have not been implemented at all. The other services are in various stages of preview. The platform, when completely implemented, would be in a position to handle everything that one could do with the traditional version of Microsoft's software. Hence, the scope is pretty broad.

From a technology point of view Microsoft has a great advantage in that being a platform leader in the Operating Systems market it already has a large user base for these technologies. Azure would simply make them available as a service. One of the problems that Microsoft has always had is one of interoperability with other platforms. In fact this was the basis of their dominance for a long time. In the cloud space they do not have a first mover advantage – there are other entrenched providers in the market. The big

question is whether Azure based services can coexist with other non Microsoft cloud services. For instance, can an application hosted on Azure use Microsoft Live Services along with SalesForce.com's CRM application.

In terms of relationship with complementors, Microsoft has a large population of developers and ISVs writing to Window's OS and services. Microsoft has managed to have a great relationship with them in the past. It remains to be seen if they can leverage that relationship to get more ISVs to build services on the platform. The Azure platform has an SDK for ISVs – this is in contrast to Google's App Engine. According to Hauger, providing the right level of expectation about service levels is also an important criteria for a successful platform – the SLA does not have say 99.999% availability, however, it must state the expected response time of the service accurately. At this point we are not sure what those service levels will be.

One major problem that Microsoft is likely to face is how to organize itself to deliver on the promise of the Azure platform without affecting their core businesses – specifically Windows and Office. As long as the major chunk of their revenue is coming from traditional software technologies those departments will have more clout. A strategy for both these organizations to coexist with mutual respect has to be devised at the highest levels of management. One other problem they might face is adjusting to the competitive landscape in the cloud computing arena. Other players, especially Google and Amazon have established themselves in the early movers.

Chapter 5: The Road Ahead

Cloud computing has made many strides in the last few years. As mentioned in the previous chapters the business is slated to grow further in the coming years – in fact if predictions are true cloud business will overtake the traditional software market. However, there are still some impediments to its progress. We look at these issues, and what companies can do solve these.

The Hurdles

Existing software and hardware infrastructure:

One of the greatest hurdles to increased adoption of cloud computing as a viable business model is the capital that has already been spent on acquiring hardware and software licenses for the enterprise. This coupled with the software that has been developed to run on this infrastructure. One cannot change to cloud computing by flipping a switch. Most of the software would have to be rewritten to use cloud APIs. Also, there may not be a one-to-one replacement for all on-premise software running in an enterprise. In addition the cost of hardware has decreased considerably over time – so much so that many of the hardware suppliers are finding it hard to stay afloat.

Additionally, enterprises have generated volumes of data that are hosted on databases like Oracle or in ERP systems from SAP, Oracle and other companies that do not have a cloud product of their own nor do these companies provide a way to port this data to the cloud. Companies would have to sell their existing infrastructure, rewrite a lot of their code to be able to take advantage of the cloud. According to Microsoft's Hauger, any new technology rarely eclipses the install base. This is especially true for cloud computing in the absence of a clear migration path.

Change in software development methodologies:

In order to take advantage of the cloud platform enterprises have to re-evaluate their software development methodologies. In the new scheme of things the choice of programming language, architecture, and functionality, is restricted by the choice of the cloud platform. Also, the build test deploy and maintain cycle is not the same as it used to be. This calls for changes in program and project management that can take advantage of the new paradigm. The software industry has to adopt a cloud based software development model. It also calls on software engineers to change the way they look at development. This change from the ground up cannot be done instantly. Some organizations may not be able to make that transition at all leaving them to always depend on traditional software development practices.

Lack of Frequently Used Tools:

Along with the core software that an enterprise has for running its software, there are other enabling software that is vital for the smooth management of software development and operations. These include among others, project management software, bug tracking systems, customer quality measurement indicators. These have to reside in the cloud along with enterprise applications. A case in point was that of Northrop Grumman's desire to use the CRM solution from Salesforce.com. Initially they could not do so because their project management software, SEER from Galorath, was not cloud enabled. They worked with Galorath to make the software available on Salesforce.com's platform only after which Northrop Grummen start using their CRM service.

Identity, Privacy and Security Concerns:

Lew Tucker, CTO of Sun's cloud services believes Identity Management is a still an issue. Each cloud has its own identity solution. There is no single identity across multiple clouds. While there have been attempts to solve this problem there is no clear solution yet.

Increased demand for security and compliance is considered a good reason to move to the cloud. According to Peter Coffee of Salesforce.com, only a cloud provider may be able to spend that kind of money or amortize the cost of providing high security and auditability across a large number of users. However, privacy is turning out to be a major concern among enterprises. They are still leery of storing data outside their firewalls. What if hackers get access to the data? There might be an SLA in place for handling these scenarios but once the data has been compromised no SLA can undo that. Most companies need to make their enterprise and data available for audit – this becomes an issue if the data center is no longer in-house. Cloud providers are trying to get around that problem by showing their datacenters to customers (after making them sign non-disclosure agreements!). They are also getting their processes certified in order that their customers can breathe easy.

Protocol to Handle Failure:

One of the issues Lew Tucker, CTO of Cloud services at Sun, brought up was around rules of failure. What happens when an application is connected to 5 cloud services and one of the services goes down? What is the protocol to handle failure? How should the other services respond? Do we see the development of a new multi phase commit protocol or can the two phase commit for transactions be applied to the cloud? These issues have not been answered properly yet. While companies talk about high availability they have not yet addressed the technical issues when the system is unavailable.

Bandwidth:

As more applications move to the cloud more bandwidth will be needed to transport data between the cloud providers and the client accessing the application. This coupled with the fact that media companies are also making their bandwidth hogging content available on the net. Is there enough bandwidth to handle all the data movement? This may become a serious concern going forward especially in areas of the world where the current internet infrastructure is lacking. In addition network providers are challenging the Net

Neutrality act as they are not making as much money from the proliferation of broadband usage as they had hoped. Any change in laws may alter the equation adversely affecting cloud computing. Irrespective of that there needs to be more investment in building data centers closer to the customer. There needs to be new networking technologies that can optimize movement of data across networks and provide a stable response time to end-users.

Cisco Systems recently introduced Nexus 7000, a new network switch for companies burdened with rapidly growing Internet data transfers and the increased use of applications that draw on remote data storage, known as cloud computing. The switch is supposed to provide a sharp increase in traffic capacity over the company's current products — to 15 trillion bits of data a second. To illustrate speed, the switch could transfer all 90,000 Netflix movies in 38.4 seconds or send a two-megapixel digital image to every human being on earth in 28 minutes. Better cloud services to serve the types of files that are hosted by web 2.0 sites are also needed.

The HTTP Protocol as a Stumbling Block:

There are many tasks that cannot be accomplished with ease due to the limitations of the HTTP protocol's request/response nature. In order to handle all types of messaging across the cloud as is possible within an enterprise's data center we may need a better protocol than HTTP. This is yet to happen.

Data Portability Issues:

Countries, and regions within countries, have their own rules around where data can reside and how that data should be protected. Since 9/11 the US has introduced the Patriot Act, which allows the US government access to private data. As a result government's of other countries like Canada do not want data generated in their countries to be stored in the US. The EU has similar rules which says data generated in the EU

should be stored within the region. This puts additional constraint on cloud computing. Enterprises cannot move to the cloud if the cloud provider does not have a data center in the region where they need to operate.

Lack of Channel Incentives:

Companies that sell traditional hardware and/or traditional software have various sales channels to sell their products. Each channel has its own set of incentives. For instance, if the company employs its own sales force to sell to an enterprise the sales personnel generally get a share of the revenue. In cloud computing many customers currently buy services from cloud providers by simply going to their website and supplying a credit card number. There may not be a clear relationship yet between service metering and billing and linking that to the sales person responsible for bringing in that business. This has to be addressed wherever it hasn't been. Also cloud computing removes some of the middlemen involved in the sales process. The hope is that as cloud service providers try to acquire more enterprise customers they will involve enterprise sales force and possibly other channel partners and provide the right incentives.

It was resellers that made Microsoft great and so, following in those well-trod footsteps, Google has started recruiting resellers to push its Google Apps to businesses of all sizes everywhere in the world, taking Microsoft on where it lives. According to sys-con.com, so far Google has collected 50 "pilot partners." The authorized resellers are supposed to be able to sell, customize and support Google Apps Premier Edition, creating new revenue opportunities for themselves and easier access to Google's cloud services. And they'd get a 20% discount off the $50-a-user-a-year price.

The Future of Cloud Computing

All the interviewers we spoke with were upbeat about the future of cloud computing. Lew Tucker, CTO of cloud computing at Sun, is a believer in the "big switch" concept described by Nicholas Carr. He thinks fewer and fewer enterprises will build traditional datacenters of their own. There will be virtual datacenters. He also sees enterprises

getting a lot more interested in cloud computing and that too much earlier in the game than one thought they would. This is because they private clouds are going to change how data centers are run. These would be built much more around self-service. Developers in various departments can do their own provisioning or internal requisition. The strategic benefit for them is that through virtualization enterprises can dynamically shift their resources to provide improved levels of service at different times in different geographic locations. Hi sites a futuristic example of an investment bank that deals across continents – it can dynamically shift its resources to offer faster response times to APAC customers for a part of the day, EMEA customers at some other time and US customers for the remaining time, all with the same hardware resources. A cloud within an enterprise would provide increased computational effectiveness at lower cost.

In the near term Lew sees a shift from web 2.0 to cloud. Startups would not buy any hardware; they will deploy their application on a cloud. He also sees issues going forward – we have not figured out quite a few things. As people try writing mashups that draw upon multiple services new problems might show up. What happens when one of these services are not available? We have to solve the issues surrounding dependency on large disparate systems.

Peter Coffee of Salesforce.com, sees financial services, medical systems moving to this model soon. Eventually R&D and government systems will also move to this platform. They already have a customer in Dolby Labs who have built their entire bug tracking system around Saleforce.com. Technicians running the sound systems in cinema halls can go to the web and file a bug by simply logging on to the system and entering the product ID of the system.

Doug Hager of Microsoft strongly believes that most companies will have moved their business to a cloud of some form in 10 to 50 years! The install base will require an evolution. In the shorter term we will have hybrid solutions, where on premise applications written on Windows, Oracle, Sun platforms, for example, will come with a level of abstraction into cloud based services. He sees this happening on 2 to 5 years. The

current economic crisis will accelerate the use of specialized services that are available on the cloud.

We believe in the future of cloud computing as a viable technology and, hence, a viable business too. One major shift in the near future will be towards private data centers. This is because of the current industry landscape. Big players are getting into this space. Sun recently announced the acquisition of Q-Layer, a Belgium-based infrastructure management company that has technology to automate the deployment and management of both public and private clouds[19]. Elastra is building a cloud solution that can be deployed in a private data center. One recent revelation that may further help the adoption of cloud computing in private datacenters is the story of Bechtel. Its CIO was inducted into the CIO hall of fame last year for shifting its IT services towards a more cloud centric model. He incorporated high-bandwidth networking practices from companies such as YouTube, the standardized server approach of Google, extreme virtualization techniques from Amazon, and the multitenant application support strategy of Salesforce.com, among others. This helped provide a secure, ubiquitous, simplified and rapidly deployable access to corporate and customer information for any Bechtel user around the globe. Ramleth calls his approach the "consumerization of the computing environment"—an internal cloud-computing infrastructure serving up in-house applications on demand[20]. This shows the value enterprises can derive by moving to a cloud based platform.

Among the provider of cloud computing services we see a trend towards providing one company's cloud solution on another company's platform. Google's support for SalesForce.com's platform via its Python APIs is a step in that direction. This is a trend that will continue until companies begin to realize the risk of envelopment.

AMD recently announced that it was building a specialty supercomputer to deliver gaming through a computing cloud. This news bolsters the cloud business model —

---

[19] http://gigaom.com/2009/01/08/for-sun-q-layer-is-a-smart-buy/
[20] http://www.cio.com/article/453214/Cloud_Computing_to_the_Max_at_Bechtel

taking it beyond storage and run-of-the mill computing into the realm of specialty clouds. Since gaming may involve mobile devices these cloud services may have to be optimized to deliver services to the mobile platform too. This may be true for all clouds in future with the convergence of PCs and mobile devices.

If the cloud computing industry takes the shape of electric utilities as many are predicting, they may behave like the current utility providers. They may be reluctant to change the architecture of their systems or incorporate greener and better technologies as that would mean a complete overhaul of their data centers. Proper regulation must be put in place so that data centers are forced to reduce per unit energy spend over time.

With the move towards cloud and the decrease in traditional datacenters, hardware providers will have to find innovative ways to stay in business. They have to decide whether they want to start their own clouds or continue to sell to other cloud providers only. Their profits might decrease over time as we achieve a saturation level in data center business. Needless to say this is will take some time to happen.

Over the last 10 years I have attended almost a dozen all-hands by the charismatic Scott McNealy, former CEO of Sun Microsystems. He often talked about Sun being the company that delivers "the big freaking web tone" (BFWT). I liked the vision but always wondered why the world did not buy into that idea. With cloud computing the answer seems quite obvious. He was ahead of his time. The cloud is a BFWT that comes with applications that people need. It's the applications and the API to access this web tone that adds value. The web tone delivers search results, images, blogs, calendars, emails, customer relationship management software.

The promise of cloud computing is one that can revolutionize computing and change the way we go about our business. Services would become more ubiquitous – available on our mobile phones or from our computers. It is also a path towards better data center utilization – this is essential as the need for computation increases in the face of growing environmental challenges.

Appendix A: Interviews

Interview format:

Most of the interviews were conducted over the phone. We asked the following questions to each of the interviewers:

- What are the technical improvements that have enabled cloud computing as a viable business model?
- What in your view are the economic values of cloud computing?
- If cloud computing is a platform what are the platform levers? What strategies could a platform provider adopt to dominate in this technology?
- What are the various business models that are cropping up as a result of this technology?
- What are the impediments to the growth of Cloud Computing?

In some cases we asked a few questions specific to the company they worked for.

Interview with Lew Tucker. Chief Technology Officer, Cloud Services, Sun Microsystems.

Q: Where do you draw the line between what is cloud computing and what is not?

Answer: I see delivering of applications across the network as cloud computing. For example a document management system like Google Docs delivers an application over the network. Cloud computing is an "all encompassing" term that includes any service that is delivered over the network.

Q: In your mind what are the technological breakthroughs that have enabled cloud computing?

Answer: Virtualization is the first thing that comes to mind. The idea that multiple Operating Systems (OS) can be hosted on the same physical hardware has really helped the cloud computing business. This increases server utilization. It also allows for the growth of new kinds of appliances constituting an OS and a software stack. Before virtualization one had to provision different machines, one for Linux, one for windows server, one for an application server running on Ubuntu, for example. The other really overlooked area has been the use of Application Programming Interfaces (APIs) on the web. This has allowed web services to consume other services on the internet. Cloud computing allows these APIs to be used by web applications. These are simple HTTP based APIs. Saleforce.com was one of the players that ensured all their applications always had APIs.

Q: What are the technical hurdles that still remain to be conquered?

Answer: I believe Identity Management is a still an issue. Each cloud has its own identity solution. There is no single identity across multiple clouds. While there have been attempts to solve this problem there is no clear solution yet.

The other issues are related to security. What is the right way to ensure security in the cloud? Govt. regulation of where users data can reside and their desire to have control over such data is another pain point.

Q: Do you see cloud computing as a platform?

Answer: The value of a vendor's cloud depends on applications and platforms. Note that platforms occur at multiple levels. One platform could host another software platform. For example RightScale provides a management framework that application developers can then use. Smugmug may use that management framework for their application in the cloud. The old Independent Software Vendors are slowly going to become "Independent Service Vendors" or "Internet Service Vendors". You could have a "Ruby and Rails" service provider who provides a platform for hosting Ruby applications on the cloud.

Q: What are your thoughts on private clouds?

Answer: Enterprises are a lot more interested and that too much earlier in the game than we thought they would. This is because they private clouds are going to change how data centers are run. These would be built much more around self service. Developers in various departments can do their own provisioning or internal requisition. The strategic benefit for them is that through virtualization enterprises can dynamically shift their resources to provide improved levels of service at different times in different geographic locations. Think of an investment bank that deals across continents – it can dynamically shift its resources to offer faster response times to APAC customers for a part of the day, EMEA customers at some other time and US customers for the remaining time, all with the same hardware resources. A cloud within an enterprise would provide increased computational effectiveness at lower cost.

There are barriers to the adoption of cloud technologies too. For one, this would require a rewrite of existing applications to take the benefit of cloud APIs. It seems like existing applications will only be able to extract the benefits of virtualization. However, any new application can take advantage of the cloud.

Q: Do you see companies like Elastra, Hyperic, and others who are fronting cloud services with their APIs reducing the clouds to commodities?

Answer: There will be big shifts in utility computing. However, as users start using a cloud provider and store more and more of their data in that cloud it would be difficult for them to move their data. It could be the case the case that data stays in the same cloud but computation moves from one to the other depending on who is offering the cheapest rates at a given point in time.

Q: If all cloud providers start providing the same features how you customers differentiate these providers?

Answer: If you look at the US telecom market there are only a handful of providers. Yest cutomers find a way to choose between one or the others. I think there will be factors like customer service, SLAs, connectivity to other clouds on which cloud providers will compete. The container for sure would become a commodity.

Q: Where do you see Cloud Computing going in the near term? In the long term?

Answer: In the near term I see a shift from web 2.0 to cloud. Startups would not buy any hardware, they will deploy their application on a cloud. In the long term I am a big believer in the "big switch" concept as described by Nicholas Carr. Fewer and fewer companies will build their own datacenters. There will be virtual datacenters.

I also see issues going forward – we have not figured out quite a few things. As people try writing mashups that draw upon multiple services new problems might show up. What happens when one of these services are not available? We have to solve the issues surrounding dependency on large disparate systems.

Interview with Peter Coffee, Director of Platform Research at Salesforce.com.

Q: What in your mind are the technical improvements over the last decade that has enabled this new business?

First lets look at the negative aspects of today's data centers. Organizations are unhappy with the state of things are actively looking for alternatives. The complexity of administering computing is rising. This is partly because governance expectations have risen dramatically, primarily due to regulations such as SOX and HIPPA. The industry mindshare of data protection and governance has increased. The cost of implementing and administering security has increased too. Would enterprises like to bear these costs themselves or should it be amortized across thousands of users? Clearly it's the latter. The industry is also revaluating where the tasks of administering security and data protection can be better performed. If an organization has a massive data center that is managing the data and computation resources of more than one enterprise they can spend more money to hire the right talent and enforce the strictest policies. This is something even large enterprises may not want to do.

The reason mentioned above is not technical. If you look at the processors of today they are not the fastest in single threaded chips. These are multi threaded at the core. This means that tasks don't improve linearly simply by running on a machine having these new multi core chips. Shared data centers can continue to exploit the improvements in chip technology that small organizations cannot hope to. The CAPEX per dollar is tilting towards shared data centers.

Coming to the other improvements that have enabled cloud computing has been the increase memory size, increased availability of bandwidth and the very high level of interactivity among applications. Services on the other hand have become unproductive

and undifferentiated. The increased economy of processing and bandwidth is pushing towards an increase of capacity.

Ever since September 2007, I have been asking enterprises to look at the rising cost of capital. Does the cost of capital match the value derived from the capital. Enterprises either over provision their data centers to meet anticipated demand or under provision their datacenters and rush to meet increased demand later at the risk of losing customers.

Improved wireless connectivity has also helped. Users now expect to connect to their applications not just from the desktop but also from virtually anywhere. That calls for an infrastructure that can service up these applications with optimal performance from anywhere.

Q: What according to you is the economic value of cloud computing?

Answer: According to the CIO of Bechtel who studied YouTube, Google and Salesforce.com they found startling differences in the cost savings per unit of bandwidth these companies get. The number of machines managed by a system administrator is far higher in a company like Google as compared to Bechtel. According to the CIO any enterprise that does not have a strategy to move to the cloud will end up on the losing side. Clouds provide economic value by bringing down the cost of bandwidth to $20 per mb/sec of b/w. The number of machines per systems administrators is very high in cloud enables datacenters. Companies like Salesforce.com can push up to three updates to their software each year and our customer's can accommodate that upgrade with ease as they do not have to migrate their data or change their programs. There is a huge opportunity cost. Cloud computing is a major tipping point.

Q: If cloud computing is a platform what are the platform levers?

Answer: For SalesForce.com the system had to scale to meet the computing needs of the planet. To meet the needs of the business you had to have a customization story.

Salesforce.com has code that we run and metadata that customers have specific to their application. It's the metadata that allows applications to customize. We asked the following questions: suppose we were able to provide our customers with customization of the logic – something that allows them to manipulate not only the descriptive stuff but active parts of an application we could get users build the application on our platform without having to leave the web. We also added customization of the visual element using a technology we called Visualforce. This gives us more leverage with our customers and acts as a platform lever. It allows users to develop their application much more rapidly. The cost of project management reduced too.

Q: What are the various business models that are cropping up as a result of this technology?

The main business model is software as a service. For us its CRM and software that handles customer service and support. Collaboration services is also another area. We also have applications that tread both Salesforce.com and other applications that run outside of Salseforce.com.

We also allow businesses to go to market on their own by building their application or infrastructure on SalesForce.com. We do the billing for them. Their customer does not have anything to do with Salesforce.com. Its like these applications become an OEM on our platform.

In Force.com our customers use their own URL and render their application completely out of our own infrastructure.

Q: How can cloud providers create value and how can they capture value?

Cloud computing creates value by allowing developers to innovate quickly. The cost of adoption is low. There is scope for rapid improvement with cost reductions. I believe the one can get superior Ease of Use in the cloud. One can get superior adaptability for a

global workforce. Multiple end user language cab be supported. The platform in general provides more stability, security, governability and auditability. The platform is also very predictable in my opinion. We just signed an enterprise license with Dell – they have decided to build all their IT apps on Salesforce.com.

Q: I also want to get your thoughts on some specific questions around Platform as a Service (PaaS).

a) What are the strategies one might adopt to beat companies like SAP and Oracle.

The major issue is industry's perception that customizations in a platform like Salesforce.com is hard. We need to drive home the point that user experience delivery is very quick.

b) Where do you see the PaaS model going from here?

I see the platform as a service being more widely adopted. System integrators can implement projects faster and customize better to meet the requirements of their users in a shorter period of time.

c) What in your opinion are the stumbling blocks to users moving to a PaaS model?

As I mentioned above the perception in the community is a major stumbling block. The industry thinks that having the enterprise data center in their basement gives them more control. This is not true. There are some latency issues that need to be resolved but this is not a major issue. There is of course the issue of government regulations around data protection and storage. We are working with various governmental authorities to simplify these rules.

d) Besides CRM what other types of software do you see shifting a PaaS model?

I see financial services, medical systems moving to this model soon. Eventually R&D and government systems will also move to this platform. We already have a customer in Dolby Labs who have built their entire bug tracking system around Saleforce.com. Technicians running the sound systems in cinema halls can go to the web and file a bug by simply logging on to the system and entering the product ID of the system.

Interview with Bert Armijo, VP of sales, Marketing and Product Management at 3Tera.

Q: What are the technical improvements over the last decade that has enabled this new business.?

Answer: There are a few key developments that had to happen before cloud computing could be realized - virtualization, plentiful bandwidth, powerful microprocessors, distributed computing, and the web all play a role. The web provided a standard user interface experience and broke the tight coupling between the user and the application. We're now completely comfortable that applications run remotely. More powerful x86 processors and distributed computing ended our dependence on big-iron systems like mainframes and SMP and their proprietary operating systems. Virtualization created a separation between software and the hardware that runs it. And lastly, bandwidth is now available to move large workloads around.

Q: What in your opinion are the economic values of cloud computing?

IT operations have gotten unbelievably complex over the past decade or so. One analyst report published a couple years ago showed that while labor comprised 20% of IT spending in the 90's it's now 80%. Just getting a new server in an enterprise data center can take months and require numerous approvals. If you start a small web company setting up servers, networking and storage in a traditional colocation facility will cost more than your developers salaries. Yet, servers are all the same. Cloud computing cuts through this morass, making resources available on-demand, over the internet.

This has several affects. For small companies, it puts world class data center resources within their budgets for the first time. For enterprises it puts a price tag on server operations; a benchmark that most found they simply couldn't come close to matching.

Plus, it puts resources mere minutes away, which makes many projects viable that couldn't be implemented before because of the delay in deployment.

Q: If cloud computing is a platform what are the platform levers? What I meant to ask was what would it take for cloud providers to retain their customers. Each cloud has its own set of APIs to connecting to the cloud as well as proprietary data formats like Amazon's simpleDB. Seems to me one can move computation around from one cloud to another. However, data seems to be hard to move around. How do you look at this aspect of the cloud business?

Answer: Keep in mind that barriers work both ways . . .

Vendors often use APIs as a lock-in mechanism, of-course, and the API wars for cloud are just getting started. However, this strategy can backfire. First, customers see this coming. In cloud, for instance, large users are already starting to push for standard APIs. Unfortunately standards take time and agreement amongst the vendors. In the first two gatherings to discuss standards Amazon, Google and Microsoft didn't bother to send a participant. Second, introducing new APIs implies that on-boarding software requires writing new code. For the first two years EC2 was available, it was nearly impossible to run a database reliably because the storage system was "unique."

Moreover, new APIs really aren't needed. Cloud systems supporting general purpose computing (as opposed to SaaS and PaaS) use virtualization to allow running existing operating systems which of course already offer abstractions for memory, storage and network. Although difficult, it is possible to leverage these abstractions. Unfortunately, many cloud vendors chose to "augment" them with their own - hence the APIs.

3tera's AppLogic on the other hand has no API, so there's absolutely no code writing required to make use of the system. We've made this choice specifically to allow as much existing software to be used in the cloud as possible. This required a tremendous amount

of extra engineering, but as a result you can literally install software like Apache straight from the rpm off their site. Even the Wordpress 5 minute install works unaltered.

So how do you keep your customers? Innovate. Produce a good product. Create a better business model. Support your customers.

As revolutionary as AppLogic is, our approach to selling and our pricing model are as much a strategic advantage as the product itself. But that's a subject for a different email.

Q: What are the various business models that are cropping up as a result of this technology?

The one most people are familiar with are service providers who set up the data centers, servers, etc and make the resources available to subscribers.

3tera is almost alone at the moment selling a platform for cloud computing but others will likely follow. Our customers include service providers as well as enterprises.

There are a couple companies focusing on software packaging for cloud computing and we're also seeing a few system integrators starting to offer cloud readiness services for enterprises.

Q: How can cloud providers create value and how can the capture value?

Done right, cloud computing creates immense value for users.

Service providers create value by offering "resources on demand." Their value creation is in efficient reliable operations. 3tera licenses our AppLogic system to service providers or enterprises who want to create clouds. Our value to service providers is in enabling them to enter the space and generate revenue from new customers. For enterprises our