# *The next generation of computing – A walk through the clouds Cloud Computing*

**Dr. Joe.Lopez**
*Associate Professor*
*Matrix School of Management Studies Pune*
**&**
**Mr.Sameer Kshirsagar**
*Project Manager*
*SAS eSourcing*

## Objective

This paper endeavors to dwell, understand and highlight the advantage of Cloud Computing in the current era of Globalization, in the highly challenging and competitive IT industry.

## Introduction to Cloud Computing

Cloud Computing, the long-held dream of computing as a utility, has the potential to transform a large part of the IT industry, making software even more attractive as a service and shaping the way IT hardware is designed and purchased. Developers with innovative ideas for new Internet services no longer require the large capital outlays in hardware to deploy their service or the human expense to operate it. They need not be concerned about over-provisioning for a service whose popularity does not meet their predictions, thus wasting costly resources, or under-provisioning for one that becomes wildly popular, thus missing potential customers and revenue. Moreover, companies with large batch-oriented tasks can get results as quickly as their programs can scale, since using 1000 servers for one hour costs no more than using one server for 1000 hours. This elasticity of resources, without paying a premium for large scale, is unprecedented in the history of IT. Cloud Computing refers to both the applications delivered as services over the Internet and the hardware and systems software in the datacenters that provide those services. The services themselves have long been referred to as Software as a Service (SaaS). The datacenter hardware and software is what we will call a Cloud. When a Cloud is made available in a pay-as-you-go manner to the general public, we call it a Public Cloud; the service being sold is Utility Computing. We use the term Private Cloud to refer to internal datacenters of a business or other organization, not made available to the general public. Thus, Cloud Computing is the sum of SaaS and Utility Computing, but does not include Private Clouds. People can be users or providers of SaaS, or users or providers of Utility Computing.

## What is Cloud Computing?

Cloud Computing refers to both the applications delivered as services over the Internet and the hardware and systems software in the datacenters that provide those services. The datacenter hardware and software is what we will call a Cloud. When a Cloud is made available in a pay-as-you-go manner to the public, we call it a Public Cloud; the service being sold is Utility Computing. Current examples of public Utility Computing include AmazonWeb Services, Google AppEngine, and Microsoft Azure. We use the term Private Cloud to refer to internal datacenters of a business or other organization that are not made available to the public. Thus, Cloud Computing is the sum of SaaS and Utility Computing, but does not normally include Private Clouds. We'll generally use Cloud Computing, replacing it with one of the other terms only when clarity demands it. Figure 1 shows the roles of the people as users or providers of these layers of Cloud Computing, and we'll use those terms to help make our arguments clear.

The advantages of SaaS to both end users and service providers are well understood. Service providers enjoy greatly simplified software installation and maintenance and centralized control over versioning; end users can access the service "anytime, anywhere", share data and collaborate more easily, and keep their data stored safely in the infrastructure. Cloud Computing does not change these arguments, but it does give more application providers the choice of deploying their product as SaaS without provisioning a datacenter: just as the emergence of semiconductor foundries gave chip companies the opportunity to design and sell chips without owning a fab, Cloud Computing allows deploying SaaS—and scaling on demand—without building or provisioning a datacenter. Analogously to how SaaS allows the user to offload some problems to the SaaS provider, the SaaS provider can now offload some of his problems to the Cloud Computing provider. From now on, we will focus on issues related to the potential SaaS Provider (Cloud User) and to the Cloud Providers, which have received less attention.

We will eschew terminology such as "X as a service (XaaS)"; values of X we have seen in print include Infrastructure, Hardware, and Platform, but we were unable to agree even among ourselves what the precise differences among them might be.1 (We are using Endnotes instead of footnotes. Go to page 20 at the end of paper to read the notes, which have more details.) Instead, we present a simple classification of Utility Computing services in Section 5 that focuses on the tradeoffs among programmer convenience, flexibility, and portability, from both the cloud provider's and the cloud user's point of view.

## From a hardware point of view, three aspects are new in Cloud Computing

1. The illusion of infinite computing resources available on demand, thereby eliminating the need for Cloud Computing users to plan far ahead for provisioning
2. The elimination of an up-front commitment by Cloud users, thereby allowing companies to start small and increase hardware resources only when there is an increase in their needs
3. The ability to pay for use of computing resources on a short-term basis as needed (e.g., processors by the hour and storage by the day) and release them as needed, thereby rewarding conservation by letting machines and storage go when they are no longer useful.

## What does Cloud Computing comprise of ?

Cloud computing can be visualized as a pyramid consisting of three sections:

## 1. Cloud Application

This is the apex of the cloud pyramid, where applications are run and interacted with via a web browser, hosted desktop or remote client. A hallmark of commercial cloud computing applications is that users never need to purchase expensive software licenses themselves. Instead, the cost is incorporated into the subscription fee. A cloud application eliminates the need to install and run the application on the customer's own computer, thus removing the burden of software maintenance, ongoing operation and support.

## 2. Cloud Platform

The middle layer of the cloud pyramid, which provides a computing platform or framework as a service. A cloud computing platform dynamically provisions, configures, reconfigures and de-provisions servers as needed to cope with increases or decreases in demand. This in reality is a distributed computing model, where many services pull together to deliver an application or infrastructure request.

## 3. Cloud Infrastructure

The foundation of the cloud pyramid is the delivery of IT infrastructure through virtualization. Virtualization allows the splitting of a single physical piece of hardware into independent, self governed environments, which can be scaled in terms of CPU, RAM, Disk and other elements. The infrastructure includes servers, networks and other hardware appliances delivered as either Infrastructure "Web Services", "farms" or "cloud centers". These are then interlinked with others for resilience and additional capacity.

## Types of Cloud Computing

## Public Cloud

Public cloud (also referred to as 'external' cloud) describes the conventional meaning of cloud computing: scalable, dynamically provisioned, often virtualized resources available over the Internet from an off-site third-party provider, which divides up resources and bills its customers on a 'utility' basis. An example is ThinkGrid, a company that provides a multi-tenant architecture for supplying services such as Hosted Desktops, Software as a Service and Platform as a Service. Other popular cloud vendors include Salesforce.com, Amazon EC2 and Flexiscale.

## Private Cloud

Private cloud (also referred to as 'corporate' or 'internal' cloud) is a term used to denote a proprietary computing architecture providing hosted services on private networks. This type of cloud computing is generally used by large companies, and allows their corporate network and data centre administrators to effectively become in-house 'service providers' catering to 'customers' within the corporation. However, it negates many of the benefits of cloud computing, as organizations still need to purchase, set up and manage their own clouds.

## Hybrid Cloud

It has been suggested that a hybrid cloud environment combining resources from both internal and external providers will become the most popular choice for enterprises. For example, a company could choose to use a public cloud service for general computing, but store its business-critical data within its own data centre. This may be because larger organizations are likely to have already invested heavily in the infrastructure required to provide resources in-house – or they may be concerned about the security of public clouds (see page 9 for more on this subject).

## Characteristics of a Cloud: A Cloud should have at least the following characteristics:

- Clouds should be uniquely identifiable so that they can be individually managed even when combined with other clouds. This will be necessary to distinguish and harmonize cloud business and infrastructure policies in- force.

- The cloud should be dynamically configurable: configuration should be automatable in varying and unpredictable, possibly even event-driven conditions.
- Systems management technologies for clouds must integrate constraints on business with constraints on infrastructure to make them manageable in aggregate.
- The cloud should be able to dynamically provision itself and optimize its own construction and resource consumption over time.
- The cloud must be able to recover from routine and extraordinary events that might cause some or all of its parts to malfunction.
- The cloud must be aware of the contexts in which it is used so that cloud contents can behave accordingly.
    - *For example,* if clouds are composted, policy will have to be harmonized across cloud boundaries; when in multi-tenant mode, service level agreements may be used to determine priority access to physical resources. Application platforms today are unaware of their usage context, but business functionality in next-generation platforms will have to be managed with context in mind.
- A cloud must be secure, and it must be able to secure itself.

## What services can be used in the cloud ?

There are numerous services that can be delivered through cloud computing, taking advantage of the distributed cloud model. Here are some brief descriptions of a few of the most popular cloud-based IT solutions:

## 1. Hosted Desktops

Hosted desktops remove the need for traditional desktop PCs in the office environment, and reduce the cost of providing the services that you need. A hosted desktop looks and behaves like a regular desktop PC, but the software and data customers use are housed in remote, highly secure data centers, rather than on their own machines. Users can simply access their hosted desktops via an internet connection from anywhere in the world, using either an existing PC or laptop or, for maximum cost efficiency, a specialized device called a thin client.

## 2. Hosted Email

As more organizations look for a secure, reliable email solution that will not cost the earth, they are increasingly turning to hosted Microsoft Exchange® email plans. Using the world's premier email platform, this service lets organizations both large and small reap the benefits of using MS Exchange® accounts without having to invest in the costly infrastructure themselves. Email is stored centrally on managed servers, providing redundancy and fast connectivity from any location. This allows users to access their email, calendar, contacts and shared files by a variety of means, including Outlook®, Outlook Mobile Access (OMA) and Outlook Web Access (OWA).

## 3. Hosted Telephony (VOIP)

VOIP (Voice Over IP) is a means of carrying phone calls and services across digital internet networks. In terms of basic usage and functionality, VOIP is no different to traditional telephony, and a VOIP-enabled telephone works exactly like a 'normal' one, but it has distinct cost advantages. A hosted VOIP system replaces expensive phone systems, installation, handsets, BT lines and numbers with a simple, cost-efficient alternative that is available to use on a monthly subscription basis. Typically, a pre-configured handset just needs to be plugged into your broadband or office network to allow you to access features such as voicemail, IVR and more.

## 4. Cloud Storage

Cloud storage is growing in popularity due to the benefits it provides, such as simple, CapEx-free costs, anywhere access and the removal of the burden of in-house maintenance and management. It is basically the delivery of data storage as a service, from a third party provider, with access via the internet and billing calculated on capacity used in a certain period (e.g. per month).

## 5. Dynamic Servers

Dynamic servers are the next generation of server environment, replacing the conventional concept of the dedicated server. A provider like ThinkGrid gives its customers access to resources that look and feel exactly like a dedicated server, but that are fully scalable. You can directly control the amount of processing power and space you use, meaning you don't have to pay for hardware you don't need. Typically, you can make changes to your dynamic server at any time, on the fly, without the costs associated with moving from one server to another.

## Preview of Top 10 Obstacles and Opportunities in Cloud Computing:

| Sr.No | Obstacles | Opportunities |
|---|---|---|
| 1 | Availability of Service | Use Multiple Cloud Providers; Use Elasticity to Prevent DDOS |
| 2 | Data Lock-In | Standardize APIs; Compatible SW to enable Surge Computing |
| 3 | Data Confidentiality and Auditability | Deploy Encryption, VLANs, Firewalls; Geographical Data Storage |
| 4 | Data Transfer Bottlenecks | Fed-Exing Disks; Data Backup/Archival; Higher BW Switches |
| 5 | Performance Unpredictability | Improved VM Support; Flash Memory; Gang Schedule VMs |
| 6 | Scalable Storage | Invent Scalable Store |
| 7 | Bugs in Large Distributed Systems | Invent Debugger that relies on Distributed VMs |
| 8 | Scaling Quickly | Invent Auto-Scaler that relies on ML; Snapshots for Conservation |
| 9 | Reputation Fate Sharing | Offer reputation-guarding services like those for email |
| 10 | Software Licensing | Pay-for-use licenses; Bulk use sales |

## Cloud Value Preposition:

Cloud computing value prepositions is reduction in total cost of ownership, translation of the fixed to variable cost, improvement of business agility and the ability to build systems of a global class. The cost model allows the business to free up budgets on infrastructure and the platform allows using them for delivering innovation services quickly. Cloud computing can bring about strategic, transformational, and even revolutionary benefits fundamental to future enterprise computing, but it also offers immediate and pragmatic opportunities to improve efficiencies today while cost effectively and systematically setting the stage for strategic change.

And in many cases, the technology supporting cloud offerings serves to facilitate the design of a migration path that reduces initial investment and accelerates returns from each stage. For organizations with significant investment in traditional software and hardware infrastructure, migration to the cloud will occur systematically and over time, as with any other significant technology transition. For other less-constrained organizations or those with infrastructure nearing end-of-life, technology re-architectures and adoption may be more.

## Cloud Computing as a part of IT Strategy:

| | |
|---|---|
| **Reduce Cost** | Reduction in the total cost of ownership by optimally using the software and hardware licenses<br>Ability to scale with the demand for peak loads and seasonal variations thus optimizing the cost model |
| **De-Risk** | Investment are translated form the upfront Capital Expense (CAPEX) to Operational Expense (OpEx) for consuming IT services. Further investments can be put based on the success of the initiative . Alternate sourcing strategy for IT services provides primary and fallbacks options |
| **Agility** | The infrastructure can be provisioned quickly therefore improves the time to market |
| **Global Scale** | Massively scalable engines allows building highly scalable services for customers and partners. Infrastructure scale with demand for peak loads and seasonal variations |

## Cloud Computing Maturity Model

The establishment of a cloud computing maturity model (CCMM) provides a framework for successful implementation. A phased approach to the CCMM, encompassing five key

Components.

• Consolidation

• Virtualization

• Automation

• Utility

• Cloud

## Consolidation

An organization's migration towards cloud computing begins with the consolidation of server, storage, and network resources, which works to reduce redundancy, decrease wasted space,

and increase equipment usage, all through the measured planning of both architecture and process. Consolidation is achieved primarily through virtualization but can also be approached by the use of denser computing hardware or even high performance computing. By boosting the speed of critical processes and enabling greater flexibility, the consolidation of data centers and desktops allows agencies to do more with fewer resources a significant concern in today's economic environment. Moreover, the shift to a unified fabric provides both physical and virtual access to the storage area network (SAN), creating greater efficiency and cost savings by allowing more storage to be consolidated in the SAN Network and application modernization is also an important initial step in enabling the transition to a cloud computing environment. A viable alternative to replacing infrastructure components or rewriting critical applications, modernization promotes communication between older systems and newer solutions, all while preserving the value in existing IT systems. Freed from the bonds of a mainframe environment,

critical applications modernized through a service-oriented architecture provide agencies with the increased ability to leverage newer technologies. As for security concerns surrounding cloud computing, modernization actually works to enhance the security of sensitive information stored on critical applications. When established properly, the cloud platform provides security of all data in motion, traveling between the cloud and the desktop, and all data at rest in cloud storage.

## Virtualization

Virtualization forms a solid foundation for all cloud architectures. It enables the abstraction and aggregation of all data center resources, thereby creating a unified resource that can be shared

by all application loads. Hardware such as servers, storage devices, and other components are treated as a pool of resources rather than a discrete system, thereby allowing the allocation of

resources on demand. By decoupling the physical IT infrastructure from the applications and services being hosted, virtualization allows greater efficiency and flexibility, without any effect on system administration productivity or tools and processes. By separating the workload from the underlying OS and hardware, virtualization allows extreme portability. When extended to every system component, desktop, network, storage, and servers – it enables the mobility of applications and data, not only across servers and storage arrays, but also across data centers and networks. Moreover, through consolidation – one of the critical applications of virtualization – agencies can regain control of their distributed resources by creating shared pools of

standardized resources that enable centralized management, speeding up service provisioning and reducing unplanned down time. Ultimately, the result is increased use of assets and simplified lifecycle management through the mobility of applications and data.

Although many agencies turn to virtualization to improve resource usage and decrease both capital and operating costs, the ultimate goal in cloud computing is the use of the abstraction between applications and infrastructure to manage IT as a Service (IaaS) in a true cloud environment.

## Automation

In this stage, automation optimizes an agency's virtualized IT resources. Through a transformative procedure, the infrastructure is automated, and critical IT processes become more dynamic and greater control is achieved by trusted policies. With automation, data centers can systematically removed manual labor requirements for run-time operations. Among the various forms of automation in practice today, provisioning automation is perhaps the best known and most often implemented .Rather than managing underlying infrastructure, agencies in pursuit of cloud computing need to move toward managing service levels based on what is appropriate for the application users, whether it's minimum tolerable application latency or the availability level of an application – whatever are deemed critical factors. In this regard, automation becomes an essential element. With centralized IT and self-service for end users, automation helps agencies to disentangle themselves from the burden

of repetitive management procedures, all while enabling end users to quickly access what they require. Ultimately, automation can help agencies to reduce their operating expenses by:

• Reallocating computing resources on-demand

• Establishing run-time responses to capacity demands

• Automating trouble-ticket responses (or eliminating trouble tickets for most automated response scenarios)

• Integrating system management and measurement

## Utility

In addition to automation, both self-service and metering -- feedback about the cost of the resources allocated are necessary requirements in creating a cloud service. With breakthrough capabilities for end users and agencies, self-service and metering facilitate not only better IT management but the further extension of the user experience. In the cloud, there is no intermediary between the user of a resource and the processes for acquiring and allocating resources for critical mission needs and initiatives. Since the user initiates the service requests, IT becomes an on-demand service and the costs of operation drop significantly, because costs are incurred only when the service is used and fewer dollars are spent attending to the needs of the infrastructure. Essential to IT administration is the question of how to maintain service delivery in a fully virtualized, multi-tenancy environment while at the same time providing the highest levels of security.

## Cloud

Through cloud internetworking federation, disparate cloud systems can be linked in such a way as to accommodate both the particular nature of cloud computing and the running of IT workloads. This federation allows the sharing of a range of IT resources and capabilities – including capacity, monitoring, and management – and the movement of application loads between clouds. Moreover, since federation can occur across data center and agency boundaries, it enables such processes as unified metering and billing and one-stop self-service provisioning. With cloud computing, communication increases significantly, as data sharing between previously separate systems is fully enabled – and collaboration within and between government agencies grows exponentially. Ultimately, rather than each agency operating in isolation, constricted by the boundaries of its own data center, not only can services be shared among groups, but also costs can be shared and lessened.

## Why switch from traditional IT to the cloud?

There are many reasons why organizations of all sizes and types are adopting this model of IT. It provides a way to increase capacity or add capabilities on the fly without investing in new infrastructure, training new personnel, or licensing new software. Ultimately, it can save companies a considerable amount of money.

## Removal / reduction of capital expenditure

Customers can avoid spending large amounts of capital on purchasing and installing their IT infrastructure or applications by moving to the cloud model. Capital expenditure on IT reduces available working capital for other critical operations and business investments. Cloud computing offers a simple operational expense that is easier to budget for month-by-month, and prevents money being wasted on depreciating assets. Additionally, customers do not need to pay for excess resource capacity in-house to meet fluctuating demand.

## Reduced administration costs

IT solutions can be deployed extremely quickly and managed, maintained, patched and upgraded remotely by your service provider. Technical support is provided round the clock by Cloud service providers at no extra charge, reducing the burden on IT staff. This means that they are free to focus on business-critical tasks, and businesses can avoid incurring additional manpower and training costs. IT giant IBM has pointed out that cloud computing allows organizations to streamline procurement processes, and eliminates the need to duplicate certain computer administrative skills related to setup, configuration, and support.

## Improved resource utilization

Combining resources into large clouds reduces costs and maximizes utilization by delivering resources only when they are needed. Businesses needn't worry about over-provisioning for a service whose use does not meet their predictions, or under-provisioning for one that becomes unexpectedly popular. Moving more and more applications, infrastructure, and even support into the cloud can free up precious time, effort and budgets to concentrate on the real job of exploiting technology to improve the mission of the company. It really comes down to making better use of your time – focusing on your business and allowing cloud providers to manage the resources to get you to where you need to go. Sharing computing power among multiple tenants can improve utilization rates, as servers are not left idle, which can reduce costs significantly while increasing the speed of application development. A side effect of this approach is that computer capacity rises dramatically, as customers do not have to engineer for peak loads.

## Economies of scale

Cloud computing customers can benefit from the economies of scale enjoyed by providers, who typically use very large-scale data centres operating at much higher efficiency levels, and multi-tenant architecture to share resources between many different customers. This model of IT provision allows them to pass on savings to their customers.

## Scalability on demand

Scalability and flexibility are highly valuable advantages offered by cloud computing, allowing customers to react quickly to changing IT needs, adding or subtracting capacity and users as and when required and responding to real rather than projected requirements. Even better, because cloud-computing follows a utility model in which service costs are based on actual consumption, you only pay for what you use. Customers benefit from greater elasticity of resources, without paying a premium for large scale.

## Quick and easy implementation

Without the need to purchase hardware, software licenses or implementation services, a company can get its cloud-computing arrangement off the ground in minutes.

## Helps smaller businesses compete

Historically, there has been a huge disparity between the IT resources available to small businesses and to enterprises. Cloud computing has made it possible for smaller companies to compete on an even playing field with much bigger competitors. ‘Renting’ IT services instead of investing in hardware and software makes them much more affordable, and means that capital can instead be used for other vital projects. Cloud Service providers take enterprise technology and offer SMBs services that would otherwise cost hundreds of thousands of pounds for a low monthly fee.

## Quality of service

Your selected vendor should offer 24/7 customer support and an immediate response to emergency situations.

## Guaranteed uptime, SLAs.

Always ask a prospective provider about reliability and guaranteed service levels – ensure your applications and/or services are always online and accessible.

## Anywhere Access

Cloud-based IT services let you access your applications and data securely from any location via an internet connection. It’s easier to collaborate too; with both the application and the data stored in the cloud, multiple users can work together on the same project, share calendars and contacts etc. It has been pointed out that if your internet connection fails, you will not be able to access your data. However, due to the 'anywhere access' nature of the cloud, users can simply connect from a different location – so if your office connection fails and you have no redundancy, you can access your data from home or the nearest Wi-Fi enabled point. Because of this, flexible / remote working is easily enabled, allowing you to cut overheads, meet new working regulations and keep your staff happy!

## Technical Support

A good cloud computing provider will offer round the clock technical support. ThinkGrid customers, for instance, are assigned one of our support pods, and all subsequent contact is then handled by the same small group of skilled engineers, who are available 24/7. This type of support model allows a provider to build a better understanding of your business requirements, effectively becoming an extension of your team.

## Disaster recovery / backup

Recent research has indicated that around 90% of businesses do not have adequate disaster recovery or business continuity plans, leaving them vulnerable to any disruptions that might occur. Cloud computing providers can provide an array of disaster recovery services, from cloud backup (allowing you to store important files from your desktop or office network within their data centres) to having ready-to-go desktops and services in case your business is hit by problems. Hosted Desktops (or Hosted VDI) for example, mean you don't have to worry about worry about data backup or disaster recovery, as this is taken care of as part of the service. Files are stored twice at different remote locations to ensure that there's always a copy available 24 hours a day, 7 days per week.

## Security Concerns on the Cloud

Many companies that are considering adopting cloud computing raise concerns over the security of data being stored and accessed via the internet. Cloud computing providers adhere to strict privacy policies and sophisticated security measures, with data encryption being one example of this. Companies can choose to encrypt data before even storing it on a third-party provider's servers. As a result, many cloud-computing vendors offer greater data security and confidentiality than companies that choose to store their data in-house. However, not all vendors will offer the same level of security. It is recommended that anyone with concerns over security and access should research vendors' policies before using their services.

Technology analyst and consulting firm Gartner lists some security issues to bear in mind when considering a particular vendor's services:

- Privileged user access—enquire about who has access to data and about the hiring and management of such administrators
- Regulatory compliance—make sure a vendor is willing to undergo external audits and/or security certifications
- Data location—ask if a provider allows for any control over the location of data
- Data segregation—make sure that encryption is available at all stages and that these "encryption schemes were designed and tested by experienced professionals"
- Recovery—find out what will happen to data in the case of a disaster; do they offer complete restoration and, if so, how long that would take
- Investigative Support—inquire whether a vendor has the ability to investigate any inappropriate or illegal activity

- Long-term viability—ask what will happen to data if the company goes out of business; how will data be returned and in what format

## Best practice for companies in the cloud

- Inquire about exception monitoring systems
- Be vigilant around updates and making sure that staff don't suddenly gain access privileges they're not supposed to.
- Ask where the data is kept and inquire as to the details of data protection laws in the relevant jurisdictions.
- Seek an independent security audit of the host
- Find out which third parties the company deals with and whether they are able to access your data
- Be careful to develop good policies around passwords; how they are created, protected and changed.
- Look into availability guarantees and penalties.
- Find out whether the cloud provider will accommodate your own security policies

Generally speaking, however, security is usually improved by keeping data in one centralized location. In high security data centers security is typically as good as or better than traditional systems, in part because providers are able to devote resources to solving security issues that many customers cannot afford.

## Moving from Traditional IT to Cloud Computing

***Architectural planning, simplification, and transformation –***

Moving IT platforms to the clouds represents the next logical step in a service oriented world, and Build v. Buy v. Lease is the new decision framework in service selection within this context. Understanding the level of cloud and internal company maturity will guide decisions such as How and When to leverage cloud services to support core as well as non-core business capabilities, and how software assets should interoperate to provision business functionality. It is also critical in this step to give explicit focus to policy-based architectures that support agility and innovation.

***Policy-oriented business and risk management –***

Policy within and across organization boundaries has traditionally been embedded within enterprise IT platforms and applications. However scaling businesses globally will require implementing new ways to combine and harmonize policies within and across external process networks and value chains.

It will become increasingly critical for companies to establish clear and explicit definitions of governance, policy (regulatory, security, privacy, etc) and SLAs if they are to operate effectively with diverse entities in the cloud.

***Cloud composition -***

The ability of one cloud to participate in managing another will become critical to scaling a cloud. It will provide a means for a private cloud to temporarily use the resources of a public cloud as part of an elastic resource capacity strategy. It also will make it possible to more immediately share functionality, information, and computing resources. One real-life example of a composite cloud is Skype. While Skype may be considered to be just a p2p application, it actually is a global mesh network of managed network elements (servers, routers, switches, encoders/decoders, etc.) that provisions a global VoIP network with voice endpoints that are laptop/desktop computers or handheld devices that run Skype's client application at the edge of the Skype cloud. When the Skype application is not running on a laptop/desktop/handheld device, VoIP calls are not conducted through it. But when the application is running and calls can be conducted, the Skype cloud expands to use the laptop/desktop/handheld device to route traffic and manage network exceptions if needed.

***Cloud management –***

To conduct business within a cloud (recognizing what is available today), it is important for cloud consumers and Providers to align on graduated SLAs and corresponding pricing models. Maturing cloud capabilities into more advanced offerings, such as virtual supply chains, requires support for fully abstracted, policy-driven interactions across clouds. This is a big jump, and it will become a major challenge for the cloud providers to adequately model, expose and extend policies in order to provide integrated services across distributed and heterogeneous business processes and infrastructure. The data associated with these business processes and infrastructure will need to be managed appropriately to address and mitigate various risks from a security, privacy, and regulatory compliance perspective. This is

particularly important as intellectual property, customer, employee, and business partner data flows across clouds and along virtual supply chains.

***Cloud Computing: Applicability***

Not everyone agrees, but McKinsey has concluded1 as follows. “Clouds already make sense for many small and medium-size businesses, but technical, operational and financial hurdles will need to be overcome before clouds will be used extensively by large public and private enterprises. Rather than create unrealizable expectations for “internal clouds”, CIOs should focus now on the immediate benefits of virtualizing server storage, network operations, and other critical building blocks”. They recommend that users should develop an overall strategy based on solid business cases not “cloud for the sake of cloud”; use modular design in all new software to minimize costs when it comes time to migrate to the cloud; and set up a Cloud CIO Council to advise industry.

***Example of Cloud Applicability: Pharmaceutical Industry***

In the pharmaceutical sector, where large amounts of sensitive data are currently kept behind protective firewalls, security is a real concern, as is policing individual researchers’ access to the cloud. Nevertheless, chemin-formatics vendors are starting to look at cloud options, especially in terms of Software as a Service (SaaS) and hosted informatics. In bioinformatics and number-crunching, the cloud has distinct advantages. EC2 billing is typically hours times number of cpus, so, as an over-generalization, the cost for 1 CPU for 1000 hours is the same as the cost of 1000 cpus for 1 hour. This makes cloud computing appealing for speedy answers to complex calculations. Over the past two years, new DNA sequencing technology has emerged allowing a much more comprehensive view of biological systems at the genetic level. This so-called next-generation sequencing has increased by orders of magnitude the already daunting deluge of laboratory data, resulting in an immense IT challenge. Could the cloud provide a solution

## Cloud Computing: New Application Opportunities

While we have yet to see fundamentally new types of applications enabled by Cloud Computing, we believe that several important classes of existing applications will become even more compelling with Cloud Computing and contribute further to its momentum. When Jim Gray examined technological trends in 2003 , he concluded that economic necessity mandates putting the data near the application, since the cost of wide-area networking has fallen more slowly (and remains relatively higher) than all other IT hardware costs. Although hardware costs have changed since Gray’s analysis, his idea of this “breakeven point” has not. Although we defer a more thorough discussion of Cloud Computing economics to Section 6, we use Gray’s insight in examining what kinds of applications represent particularly good opportunities and drivers for Cloud Computing.

***Mobile interactive application:.***

Tim O’Reilly believes that “the future belongs to services that respond in real time to information provided either by their users or by nonhuman sensors. Such services will be attracted to the cloud not only because they must be highly available, but also because these services generally rely on large datasets that are most conveniently hosted in large datacenters. This is especially the case for services that combine two or more data sources or other services, e.g., mash-ups. While not all mobile devices enjoy connectivity to the cloud 100% of the time, the challenge of disconnected operation has been addressed successfully in specific application domains,2 so we do not see this as a significant obstacle to the appeal of mobile applications.

***Parallel batch processing:***

Although thus far we have concentrated on using Cloud Computing for interactive SaaS, Cloud Computing presents a unique opportunity for batch-processing and analytics jobs that analyze terabytes of data and can take hours to finish. If there is enough data parallelism in the application, users can take advantage of the cloud's new "cost associatively": using hundreds of computers for a short time costs the same as using a few computers for a long time. For example, Peter Harkins, a Senior Engineer at The Washington Post, used 200 EC2 instances (1,407 server hours) to convert 17,481 pages of Hillary Clinton's travel documents into a form more friendly to use on the WWW within nine hours after they were released. Programming abstractions such as Google's MapReduce and its open-source counterpart Hadoop allow programmers to express such tasks while hiding the operational complexity of choreographing parallel execution across hundreds of Cloud Computing servers. Indeed, Cloudera is pursuing commercial opportunities in this space. Again, using Gray's insight, the cost/benefit analysis must weigh the cost of moving large datasets into the cloud against the benefit of potential speedup in the data analysis. When we return to economic models later, we speculate that part of Amazon's motivation to host large public datasets for free may be to mitigate the cost side of this analysis and thereby attract users to purchase Cloud Computing cycles near this data.

***The rise of analytics:***

A special case of compute-intensive batch processing is business analytics. While the large database industry was originally dominated by transaction processing, that demand is leveling off. A growing share of computing resources is now spent on understanding customers, supply chains, buying habits, ranking, and so on.Hence, while online transaction volumes will continue to grow slowly, decision support is growing rapidly, shifting the resource balance in database processing from transactions to business analytics.

***Extension of compute-intensive desktop applications:***

The latest versions of the mathematics software packages Matlab and Mathematica are capable of using Cloud Computing to perform expensive evaluations. Other desktop applications might similarly benet from seamless extension into the cloud. Again, a reasonable test is comparing the cost of computing in the Cloud plus the cost of moving data in and out of the Cloud to the time savings from using the Cloud. Symbolic mathematics involves a great deal of computing per unit of data, making it a domain worth investigating. An interesting alternative model might be to keep the data in the cloud and rely on having sufficient bandwidth to enable suitable visualization and a responsive GUI back to the human user. Offline image rendering or 3Danimation might be a similar example: given a compact description of the objects in a 3D scene and the characteristics of the lighting sources, rendering the image is an embarrassingly parallel task with a high computation-to-bytes ratio.

***"Earthbound" applications:***

Some applications that would otherwise be good candidates for the cloud's elasticity and parallelism may be thwarted by data movement costs, the fundamental latency limits of getting into and out of the cloud, or both. For example, while the analytics associated with making long-term financial decisions are appropriate for the Cloud, stock trading that requires microsecond precision is not. Until the cost (and possibly latency) of wide area data transfer decrease, such applications may be less obvious candidates for the cloud.

## Cloud Computing: Economics

In this section we make some observations about Cloud Computing economic models:

- In deciding whether hosting a service in the cloud makes sense over the long term, we argue that the fine grained economic models enabled by Cloud Computing make tradeoff decisions more fluid, and in particular the elasticity offered by clouds serves to transfer risk.

- As well, although hardware resource costs continue to decline, they do so at variable rates; for example, computing and storage costs are falling faster than WAN costs. Cloud Computing can track these changes and potentially pass them through to the customer—more effectively than building one's own datacenter, resulting in a closer match of expenditure to actual resource usage.
- In making the decision about whether to move an existing service to the cloud, one must additionally examine the expected average and peak resource utilization, especially if the application may have highly variable spikes in resource demand; the practical limits on real-world utilization of purchased equipment; and various operational costs that vary depending on the type of cloud environment being considered.

## Cap-Ex vs. Op-Ex

Although the economic appeal of Cloud Computing is often described as "converting capital expenses to operating expenses" (CapEx to OpEx), we believe the phrase "pay as you go" more directly captures the economic benefit to the buyer. Hours purchased via Cloud Computing can be distributed non-uniformly in time (e.g., use 100 server-hours today and no server-hours tomorrow, and still pay only for what you use); in the networking community, this way of selling bandwidth is already known as usage-based pricing. 3 In addition, the absence of up-front capital expense allows capital to be redirected to core business investment. Therefore, even though Amazon's pay-as-you-go pricing (for example) could be more expensive than buying and depreciating a comparable server over the same period, we argue that the cost is outweighed by the extremely important Cloud Computing economic benefits of elasticity and transference of risk, especially the risks of over provisioning (underutilization) and under provisioning (saturation).

## Benefits of moving to the Cloud

Whereas the previous section tried to quantify the economic value of specific Cloud Computing benefits such as elasticity, this section tackles an equally important but larger question: Is it more economical to move my existing datacenter-hosted service to the cloud, or to keep it in a datacenter. While computing costs have improved the most in 5 years, the ability to use the extra computing power is based on the assumption that programs can utilize all the cores on both sockets in the computer. This assumption is likely more true for Utility Computing, with many Virtual Machines serving thousands to millions of customers, than it is for programs inside the datacenter of a single company.

***Pay separately per resource:***

Most applications do not make equal use of computation, storage, and network bandwidth; some are CPU-bound, others network-bound, and so on, and may saturate one resource while under utilizing others. Pay-as-you-go Cloud Computing can charge the application separately for each type of resource, reducing the waste of underutilization. While the exact savings depends on the application, suppose the CPU is only 50% utilized while the network is at capacity; then in a datacenter you are effectively paying for double the number of CPU cycles actually being used. So rather than saying it costs $2.56 to rent only $1 worth of CPU, it would be more accurate to say it costs $2.56 to rent $2 worth of CPU. As a side note, AWS's prices for wide-area networking are actually more competitive than what a medium-sized company would pay for the same bandwidth.

***Power, cooling and physical plant costs:***

The costs of power, cooling, and the amortized cost of the building are missing from our simple analyses so far. Hamilton estimates that the costs of CPU, storage and bandwidth roughly double when those costs are amortized over the building's lifetime [23, 26]. Using this estimate, buying 128 hours of CPU in 2008 really costs $2 rather than $1, compared to $2.56 on EC2. Similarly, 10 GB of disk space costs $2 rather than $1, compared to $1.20–$1.50 per month on S3. Lastly, S3 actually replicates the data at least 3 times for durability and performance, ensure durability, and will replicate it further for performance is there is high demand for the data. That means the costs are $6.00 when purchasing vs. $1.20 to $1.50 per month on S3.

***Operations costs:***

Today, hardware operations costs are very low—rebooting servers is easy (e.g., IP addressable power strips, separate out of band controllers, and so on) and minimally trained staff can replace broken components at the rack or server level. On one hand, since Utility Computing uses virtual machines instead of physical machines, from the cloud user's point of view these tasks are shifted to the cloud provider. On the other hand, depending on the level of virtualization, much of the software management costs may remain—upgrades, applying patches, and so on. Returning to the "managed vs. unmanaged" discussion of Section 5, we believe these costs will be lower for managed environments (e.g. Microsoft Azure, Google AppEngine, Force.com) than for hardware-level utility computing (e.g. Amazon EC2).

## Conclusion

With its convenient, on-demand model for network access to a shared pool of configurable computing resources, cloud computing is rapidly emerging as a viable alternative to traditional approaches and is carrying a host of proven benefits to government agencies. Costs are being significantly reduced, along with personnel time spent on computing issues. Storage availability increases, high automation eliminates worries about keeping applications up to date, and flexibility and mobility are heightened, allowing workers to access information anytime, anywhere. Cloud computing can be rapidly provisioned and released with minimal management effort or service provider interaction.

Ultimately, with its offering of scalable, real-time, internet-based information technology services and resources, the cloud can satisfy the computing needs of a universe of users, without the users incurring the costs of maintaining the underlying infrastructure.

Cloud computing as we see it emerging today is somewhat amorphously defined, making it difficult to form a point of view about the capabilities of currently available cloud computing instances to manage next century platforms. While it is clear that they can manage today's common platforms, we see architectural challenges for the future that we believe will be difficult to address using current cloud architectures and architecture styles. We identify technical challenges including architecture style, user and access control management, the need to have externally managed business and infrastructure policies through interaction containers, and the need for Utility Computing capabilities that must be addressed to meet future architecture requirements. Aiming at implementation of an ecosystem platform will take us beyond the management capabilities of current cloud offerings. Adding architecture components like the interaction container and externalized policy engine will improve cloud capabilities, but until these become fundamental components in cloud architecture, it is unlikely that a cloud will be able to manage the concerns of a service grid. It is interesting to note, however, that the construct of a service grid enables it to manage the concerns of a cloud. A service grid, as an autonomic architecture We see, in this paper, that cloud computing can be used to address tactical problems with which IT continually deals, like resource availability and reliability, data center costs, and operational process standardization. These near-term objectives represent sufficient justification for companies to use cloud computing technologies even when they have no need to improve their platforms or practices. But there are longer-term business imperatives as well, like the need for a company to be agile in combining their capabilities with those of their partners by creating a distributed platform that will drive aggressively toward cloud and service grid computing. We believe that clouds, service grids, and service-oriented architectures are technologies that will be fundamental to twenty-first-century corporations' successfully navigating the changes that they now face.

## References

1. Cloudera, Hadoop training and support [online]. Available from: http://www.cloudera.com/
2. TC3 Health Case Study: Amazon Web Services [online]. Available from: http://aws.amazon.com/solutions/case-studies/tc3-health/.
3. Amazon.com CEO Jeff Bezos on Animoto [online]. April 2008. Available from: http://blog.animoto.com/2008/04/21/amazon-ceo-jeff-bezos-on-animoto/.

4. Black Friday traffic takes down Sears.com. Associated Press (November 2008).

5. ABRAMSON, D., BUYYA, R., AND GIDDY, J. A computational economy for grid computing and its implementation in the Nimrod-G resource broker. Future Generation Computer Systems 18, 8 (2002), 1061–1074.

6. AMAZON AWS. Public Data Sets on AWS [online]. 2008. Available from: http://aws.amazon.com/ publicdatasets/.

7. BARROSO, L. A., AND HOLZLE, U. The Case for Energy-Proportional Computing. IEEE Computer 40, 12 (December 2007).

8. BECHTOLSHEIM, A. Cloud Computing and Cloud Networking. talk at UC Berkeley, December 2008.

9. BIALECKI, A., CAFARELLA, M., CUTTING, D., AND O'MALLEY, O. Hadoop: a framework for running applications on large clusters built of commodity hardware. Wiki at http://lucene. apache. org/hadoop.

10. BRODKIN, J. Loss of customer data spurs closure of online storage service 'The Linkup'. Network World (August 2008).

11. CARR, N. Rough Type [online]. 2008. Available from: http://www.roughtype.com.