CUTTING THROUGH THE FOG: UNDERSTANDING THE COMPETITIVE DYNAMICS IN CLOUD COMPUTING

BRIE Working Paper 190 (*Beta*)
May 1, 2010

Kenji E. Kushida[*]
Dan Breznitz
John Zysman

The Berkeley Roundtable on the
International Economy (BRIE)

[*] Kenji E. Kushida is a PhD Candidate in Political Science at the University of California Berkeley Graduate Researcher at the Berkeley Roundtable on the International Economy. kkushida@berkeley.edu
Dan Breznitz is Assistant Professor at the Sam Nunn School of International Affairs and the School of Public Policy, Georgia Institute of Technology tbvb@gatech.edu
John Zysman is Professor of Political Science, University of California Berkeley, and Co-Director, Berkeley Roundtable on the International Economy zysman@berkeley.edu

The authors with to thank Patrick Keating of Cisco Systems for his help in our research, Cisco System for financial support, and the many industry participants who generously gave their time to help us understand the issues from a variety of competitive vantages. All errors are our own.

INTRODUCTION

Cloud Computing is growing rapidly, and it is likely to become part of the dominant computing infrastructure for individuals, start-up firms, small-medium businesses, and large enterprises. However, as it is still an emerging set of technologies and business models, discussions of Cloud Computing have not reached the level of clarity or shared conceptions of more mature areas of computing. The purpose of this document is threefold.

Part I can be used as a standalone introduction to Cloud Computing for general audiences. It provides an operating definition of Cloud Computing, with an overview of the logic behind competing definitions.

Part II introduces a set of conceptual tools that helps map US competitors and understand their strategies. The National Institute of Standards and Technology (NIST) definition of Cloud Computing distinguishes between *service* and *delivery* models. To the discussion of service models, we contribute a conception of "stacks" that enhances the NIST definitions to better understand the realities of the market as it unfolds. Part II also presents Cisco's conceptions of delivery models, which are enhancements of the NIST definitions.

Part III examines the market and regulatory issues facing different sets of users and providers. It concludes with a brief overview of emerging regulatory issues.

PART I: WHAT IS CLOUD COMPUTING?

OPERATING DEFINITION OF CLOUD COMPUTING

First we provide an operating definition of Cloud Computing, then unpack it into its components.

Cloud Computing provides on-demand network access to a computing environment and computing resources delivered as services. There is elasticity in the resource provision for users, which is allocated dynamically within providers' datacenters. Payment schemes are typically pay-as-you-go models.

Cloud computing is a combination of technical architecture and business model. The "computing environment and computing resources delivered as services" is most usefully disaggregated into applications, platforms, and infrastructure, all delivered as services.

Put simply, applications are the software for users, platforms are the programming language-level environment for developing applications, and infrastructure can include processing power, storage, and user-configurable "virtual machines." Delivered as services, these elements are accessed via networks, with users typically charged by the amount of usage—infrastructure elements are therefore virtualized.

With elastic provisioning, users can rapidly scale up or scale down their usage of the computing resources. With the dynamic provisioning of resources by providers, the providers' resources are allocated "on the fly" when new users are added or existing users expand their usage.

Our operating definition draws upon formulations by the National Institute of Standards and Technology (NIST)[1], an influential study from the University of California Berkeley, "Above the Clouds: A Berkeley View of Cloud Computing,"[2] and Jonathan Murray of Technology Policy Research.[3]

---

[1] The NIST definition begins with Cloud Computing providing "on-demand network access to a shared pool of configurable computing resources (for example, networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction." See Appendix 1 for the full NIST definition.

[2] The definition in "Above the Clouds" offers three main characteristics of Cloud Computing that are new:

1. The illusion of infinite computing resources available on demand
2. Elimination of upfront commitment by users
3. Ability to pay for use of computing resources on short-term basis as needed

[3] Murray defines Cloud Computing concisely as "An operating environment consisting of infrastructure, platform and software service capabilities, connected via internet and web standard protocols which provides elastic scaling, dynamic provisioning and a 'pay-for-use' consumption model." http://www.tpsag.com/archives/19

DIFFERENT VANTAGES ON CLOUD COMPUTING

Cloud Computing is a term used to refer to a wide variety of services and business models. There are currently many definitions floating around, some more loosely defined than others, and some more misleading than others. This has led some observers to even be skeptical whether there is anything new, or if Cloud is simply a marketing term.

We contend that the reality of Cloud Computing as services offering *are* new, although they are built on top of mostly previously existing ideas and technologies. The multiple vantages on Cloud Computing are a result of different sets of users experiencing very different types of Cloud services. Moreover, the commercial logic of Cloud service providers is very different from that of users.

To the general public, "Cloud" is often synonymous with "The Internet." As it is commonly used in the media and by the marketing campaigns of some firms, exposure by the general public to the term "Cloud" is confusing. It is often used loosely as a synonym for "the Internet." In this view, any time applications or data reside outside the users' PC or access device such as a smart phone, it is "in *the* Cloud." Google's Gmail, Google Docs, and other data stored in the Calendar, for example, or social networking sites such as Facebook or MySpace are all in the Cloud. We do not agree with this catch-all usage of the term. The problem is that distinct types of Cloud service offerings are blurred into "the Cloud," and the term loses its significance. Part of the purpose of this document is to dispel this confusion.

Having noted what Cloud Computing *is not*, we now turn to what it *is*, as seen by the different logic between users and providers.

For users, Cloud Computing is what we call an "enhanced utility." Like public utilities including electricity, gas, and water, with Cloud Computing, computing resources can be accessed in user-demanded amounts once the "pipe" is connected. The "pipes," in this case, are broadband or high speed mobile connections. Upfront user infrastructure investments into storage and computing capacity are not required, and costs are incurred according to usage.

Cloud Computing therefore dramatically lowers the entry costs for new players. Users face a radically increased capacity to innovate, experiment, and quickly scale up (or down) their computing operations. Initial startup costs for small-medium enterprises (SME) or startup firms are also lowered considerably. In Silicon Valley, it appears that venture capitalists are increasingly mandating that new startup firms use Cloud Computing for their initial computing needs rather than building their own data centers – a stark contrast to the dot-com bubble era of the late 1990s.[ii]

For large enterprises, Cloud services can provide immediate extra capacity for experimentation or the testing of new services. If they face peak demand that exceeds their own infrastructure, they can purchase extra capacity rapidly and flexibly. Moreover, Cloud Computing architecture can be applied *within the firm*, increasingly the flexibility of allocation of computing power within the firm and fundamentally altering how computing resources are managed. (We will discuss this further in Part III.)

For providers, **Cloud Computing is not a utility, but a competitive service offering.** Massive investments and large scale are required to offer the illusion of infinite

computing resources with the ability for users to scale up and scale down rapidly. We examine different types of providers in the next section. There are many points of entry—some providers offer "virtual machines" (Amazon's EC2), others offer applications (Google Apps), and others provide "platforms," on top of which software developers can create their own software (Force.com by Salesforce.com).

Some providers market themselves as providing "Cloud Solutions," but they do not actually operate Cloud services, in the sense that they do not have their own data centers with Cloud architectures. Instead, they specialize in the *implementation* of Cloud Services offered by a variety of providers. Their role is close to that of system integrators or IT service outsourcers. They may take various Cloud offerings from different vendors, combining them to offer an integrated solution to users. Analytically, we need to keep Cloud integrators from actual Cloud providers, since the scale of investment into infrastructure, and therefore the competitive vantage, differs.

THE DISTINCTION BETWEEN CLOUD AND OTHER FORMS OF COMPUTING

Many of the technologies and concepts underlying Cloud Computing are not new. What is new, however, is the combination of these underlying technologies in a way that offer commercial services with the characteristics we identified as Cloud. (Not every Cloud service offering contains every technology or operating model listed here.)

Virtualization decouples applications and software platforms from the underlying physical hardware.[iii] Software (known as a hypervisor) mimics hardware, "tricking" applications into thinking that they are interfacing with physical servers when they are in fact interfacing with software-created "virtual machines." Attributes such as processing speed and memory, which applications assume belong to the physical servers are actually created by software. With virtualization, the physical hardware is decoupled from hardware specification-dependant application, enabling greater flexibility in how workloads are managed, and how datacenters are constructed.[iv] There may be several virtual machines residing on a particular physical server, or there may be multiple physical servers running one particular "virtual machine." Virtualization opens the possibility for workloads to be moved around, distributed, and scaled to new degrees.[4v]

Grid computing typically refers to a computing environment in which software allows large number of servers to work in tandem as one large system. Unlike most Cloud services, software and applications for grid computing are usually written for the entire grid arrangement, and are unable to scale up or scale down rapidly. Virtualization is typically not used. Overall, manner of resource allocation and usage patterns for grid computing differ from those of Cloud Computing, and while grid computing is primarily a technology,[vi] Cloud Computing refers to both the technology and business models.

Software as a Service (SaaS) refers to software that is delivered or accessed over networks. Software is managed centrally by the SaaS provider, relieving the IT managers of users from dealing with incremental upgrades and security patches to end-user PCs. While many Cloud offerings use SaaS business models, not all SaaS offerings share the characteristics of Cloud Computing.

Utility Computing is essentially outsourced Computing resources offered in a metered, or pay-as-you-go scheme. While many Cloud services adopt utility payment schemes, not all outsourced computing servers are Cloud.

[4] Google, famously, did not utilize virtualization in its datacenters although virtualization is usually considered a key component of Cloud Computing.

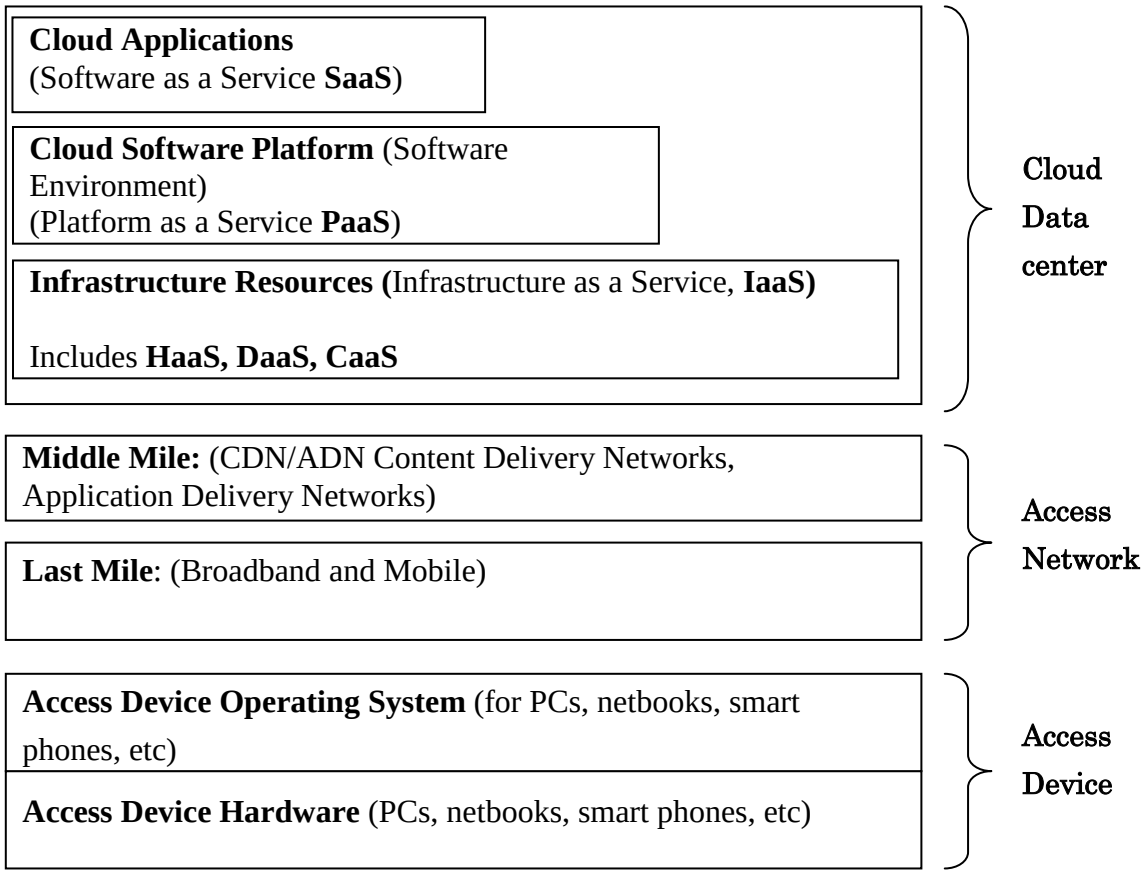
PART II: HOW IS COMPETITION IN CLOUD COMPUTING UNFOLDING? CONCEPTUAL TOOLS

Part II introduces the conceptual tool of services "stacks" to map competition as it unfolds. After introducing the services stacks, Part II takes Cisco's vocabulary and conceptions of Cloud Computing, which captures a separate dimension of competition. Combined, Part II provides analytical leverage into the strategies of Cloud Computing.

The NIST definition of Cloud Computing distinguishes between Cloud Computing service and delivery models. *Service* models refer to the types of services offered, and the level of configurability by end users. Our "stacks" conception incorporates the different service models, but adds other critical aspects of Cloud Computing over which competition is unfolding. *Delivery* refers to the location of infrastructure and management/control of Cloud services. As we will see in the next section, Cisco has gone beyond the NIST definitions to add extra delivery models, and to provide and extra dimension for analytical clarity.

First we introduce our conception of services "stacks." (See Figure 1)

Figure 1: The Cloud Computing Services Stack



The three major layers, labeled on the right, from top to bottom consist of: 1) functions within the Cloud data center, 2) the Access Networks to which they are connected to users, and 3) the Access Devices through which users access Cloud services. We examine each in turn.

CLOUD DATA CENTER

The three categories inside the Cloud data center represent different types of service models. Each offering has a different level of **user configurability** and access to the underlying Cloud infrastructure (graphically represented by the width of the box), which includes network servers, the operating system, storage, and applications capabilities. The target users are therefore different for each layer, and the level of abstraction from the underlying hardware differs. The higher layers are targeted more directly to end users, with lower layers aimed more at developers and service aggregators. Moreover, services at the higher layers can be constructed from services at the lower layers. We will examine each layer in turn, from top to bottom.[vii]

Cloud Applications (SaaS) are applications offerings, usually in the form of Software as a Service (SaaS). Their primary target is end-users – individuals, SMEs, or enterprise. They are available on demand, and users do not have to worry about scale issues or the underlying Cloud datacenter infrastructure. At this top layer, users have the least control and access, typically with no access beyond user-specific application configuration settings. Examples of Cloud Applications include Google Apps and the Salesforce Customer Relationships Management (CRM) from Salesforce.com. They give users powerful, scalable tools to use, but the underlying layers cannot be accessed.

Cloud Software Environment (PaaS) targets developers, supplying them with a programming-language-level environment and a set of well-defined Application Protocol Interfaces (API). Commonly referred to as Platform as a Service (PaaS), this layer essentially provides software development frameworks and components delivered over networks. Users have control over the deployed applications, and sometimes the configuration of the hosting environment. Developers can take advantage of the scalability of the underlying infrastructure, and in some cases, PaaS providers provide services such as authentication and user interfaces that developers can incorporate. Examples include Google's App Engine and Force.com from Salesforce.com.

Infrastructure Resources (IaaS) provides computing resources, storage, and communications resources delivered over networks. They are, in essence, virtual replacements for physical infrastructure – processors, servers, and network equipment. Collectively labeled Infrastructure as a Service (IaaS), the various components can be offered independently, or in combinations.[viii] Typically, IaaS offerings enable consumers to deploy software of their choosing, including operating systems applications, and sometimes, select networking components such as host firewalls.[ix]

Computing resources and storage are the two largest components of IaaS, with a smaller segment for communications resources. *Computing resources* offered as a service commonly take the form of Virtual Machines (also known as Hardware as a Service, HaaS).

Amazon's EC2 service is the paradigmatic example of HaaS. *Storage* offered as a service (also known as Data Storage as a Service, DaaS) can be offered in tandem with computing resources, but others offer it independently. Amazon offers such a service, S3, and firms such as Rackspace.com's Cloud Hosting service falls into this category. Traditional communications firms such as Verizon and AT&T also offer DaaS services.

Communications resources delivered as services (Communications-as-a-Service, CaaS) is less visible among Cloud Services, referring to virtualized communications and network equipment, such as telephony switches and security-enhancing equipment. Network equipment providers such as Cisco Systems, Juniper Networks, and incumbent communications firms such as Verizon offer CaaS services of various types.[x]

ACCESS NETWORKS

Access networks are the pipes that connect Cloud providers' data centers to users. For enterprise users, Cloud providers often connect via private networks. For SMEs and the general public, networks traditionally consisted of the public telecommunications network, owned and operated by telecommunications carriers such as AT&T and Verizon. The ownership of access networks is beginning to change significantly.

The **Middle Mile** in our conception consists of the space between the Cloud providers' datacenters and the "last mile" of network infrastructure connected to users. The middle mile consists of two types of related service markets – primarily Content Delivery Networks (CDN), and Application Delivery Networks (ADN). Since low latency and reliability is a critical requirement for most Cloud applications, CDN and some types of ADN services developed to alleviate transfer speed bottlenecks in the networks and with the provider's servers. Akamai, the largest CDN provider, had over 56,000 cache servers in 70 countries by the end of 2009.[xi]

The **Last Mile** refers to the physical network infrastructure connected to end users. For large enterprises and some SMEs, this means private lines, and for individuals, broadband. Wireless technologies such as 3G, and emerging services such as LTE and WiMax, with data transfer-optimized high bandwidth are also included. The incumbent telecommunications carriers and cable companies operate in this market.

ACCESS DEVICES

Access devices are the devices through which users access the Access Network and Cloud services. Traditionally, this referred to PCs, but netbooks, smart phones, and other networked devices are now included in this category. As the next section shows, control of access devices, once considered unprofitable commodities, are increasingly viewed as valuable portals to firms' Cloud offerings. Two areas that different firms identify as points of commercial leverage are the access device operating system, and the devices themselves.

Access Device Operating System is the operating system used by access devices. While Microsoft dominates PC operating systems, as the variety of access devices expands

rapidly, control of the operating systems is increasingly seen as a portal to their Cloud offerings.

Access Devices – access devices include most any device that can connect to Cloud services via broadband or high speed mobile networks. They include conventional PCs, laptops, netbooks, smartphones, and other portable devices.

USING THE SERVICES STACKS CONCEPTION TO MAP COMPETITIVE STRATEGIES

Our Services "stacks" conception is useful to map the competitive strategies of Cloud providers, not only because it can identify which service provider and Cloud offering fits,[5] but because it can map the *trajectories* of how competitors are moving. Different Cloud providers started at different positions in the stacks. This fact influences their vantage on where they can add value by offering Cloud services, and how they view their competitive strengths.

The first key point to note is that **no one player today provides Cloud Computing that includes all layers of the stack.** Firms started in different positions, and are moving in different directions on the stack. The following are a few of the major strategies.

Moving down the stack. Key firms that began with offerings at the top levels of the stack are moving down the stack. In doing so, they are becoming players in the Access Networks layers, and are *repositioning the role of access devices, once considered commodities.*

Google is both a key example and a major player. It started at the very top of the stack, offering applications. To better deliver its services, it is moving aggressively to acquire Access Network infrastructure. In a significant shift in the pattern of Internet traffic, in 2009 Google ranked third worldwide in the total volume of Internet traffic carried over its networks, displacing AT&T and Sprint.[xii] Google was also involved in a number of well-publicized undersea fiber optic cables linking Asia to North America.[xiii] It also moved into the Access Device layer – first with its Android operating system for mobile handsets, and then with a handset offering of its own. Both the operating system and handset offerings enable users' easy access into their Cloud.

Amazon, which began with a Cloud architecture for its online retail operations, expanded into offering its own access device, the Kindle. While primarily for reading books, the Kindle connects wirelessly to Amazon's online store, and can access the web, with a recent push to make it an applications platform.

Moving up the Stack. Many large firms that began at the lower stacks are moving upwards. Often experiencing the threat of commoditization at the lower layers, they seek to offer higher value-added services by moving up the Cloud Services stacks.

Telecommunications carriers, starting in the Access Networks layers, have been aggressively moving upwards. (In most countries, they are restricted by government policy from moving down the stack into Access Devices, bundling product offerings of their own

[5] *See* Appendix 2 for a table of some major competitors in each layer of the stack

with their Access Network offerings.) Their ongoing competitive challenge since the advent of the Internet has been to offer value-added services to avoid becoming commoditized – merely providers of "dumb pipes." Verizon and AT&T, for example, have begun offering storage and computing power on demand.

Cisco, which began with offerings for infrastructure in both the Access Networks layer as well as inside firms' data centers, is moving up the stack by offering applications for enterprise users, and partnering with other firms that have SaaS and PaaS offerings.

Firms that traditionally manufactured Access Devices, such as Dell and Nokia, have also been moving up the stack. Nokia, which started by making cellular handsets, moved up within the Access Devices layer to purchase Symbian, which made operating systems. It then announced an alliance with Microsoft to integrate its smartphone offerings with Microsoft's Cloud-based Office suite. Dell, originally a PC assembler, owns the domain name cloudcomputing.com, and tried unsuccessfully to trademark the phrase "Cloud Computing." It has been focusing on enabling enterprises to implement Cloud architectures for their own datacenters, and has acquired several software firms, such as enterprise email service MessageOne, remote services firm Everdream, and others.[xiv]

Federating combines Cloud offerings from multiple vendors by connecting their management infrastructures and enabling them to exchange resources and aggregated functions such as billing. Since no one firm offers all pieces of the services stacks,[6] this has opened market opportunities to offer services linking several Cloud offerings, freeing users from the hassle and management resources necessary to manage multiple Cloud interfaces.

Traditional systems integrators, such as IBM and HP are offering these services in conjunction with their other Cloud offerings and servers. New firms, exemplified by Salesforce.com have also been offering software tools and services to enable firms to mix and match Cloud offerings.

LOCATION ON STACKS INFLUENCES VANTAGES ON COMPETITIVE ADVANTAGES

The location and trajectory of firms on the Cloud service stacks also influences their views on several critical aspects of competition as it unfolds. Key areas of contention include Security, Quality of Service (QoS), Location of Intelligence, and Creating Developer Ecosystems.

Security: Firms that begin from the Access Networks layer and move upwards, tend to argue that the highest levels of security demanded by enterprise customers can only be delivered via the network itself. Firms moving down from the top layers, in contrast, contend that the level of security they offer within their datacenters is greater than that of everybody

[6] Recently, Microsoft has begun offering the entire stack of Cloud services within the datacenter layer. Until recently, its offerings were fragmented and lacking a central focus. Cisco's alliance with VMWare and EMC gives provides the capability to offer the full range of services along datacenter portion of the stack.

else. They contend that QoS and security is not about the pipe, but about the standards at each endpoint.

Quality of Service (QoS): the availability of Cloud services without interruption is crucial for enterprise users, making Service Level Agreements (SLA) a key attribute of competition. Again, firms starting from the infrastructure layers contend that high SLA is best delivered by the network itself. Firms in the top layers, on the other hand, are taking a dual approach. While working to expand into the Access Networks layers to enhance SLA for enterprise customers, they are also optimizing other Cloud service offerings for users and usage patterns that do not demand the highest level of SLA.

Location of Intelligence: The previous two issues – security and QoS – are part of a larger discussion about the location of intelligence with respect to Access Devices. Over the past twenty years, the evolution of computing entailed the movement of intelligence from the center outwards; the original server-client model of computing entailed "dumb" terminals with intelligence in the server and networks. The advent of PCs linked by the Internet placed increasing intelligence in the "edges" of the networks, with little intelligence in the network itself.[xv]

With Cloud Computing, the most intensive processing can be done in the datacenter, and management and control of content and applications can be centralized. On the one hand, this allows Access Devices less powerful than PCs to take advantage of Cloud services. However, it is not obvious that the evolution over the past twenty years will be reversed, and that processing power and capacity on the edges will immediately become irrelevant – a return to "dumb" clients. There is a logic for why power at the edge may be expected to persist.

Different visions are currently competing in the marketplace. One of Fujitsu's enterprise Cloud offerings, for example, does entail simple client machines with intelligence concentrated in the enterprise's Cloud-architecture datacenter. Other major Cloud providers strongly contend that latency issues, different Cloud resource utilization patterns, and the requirement for rich user experiences will necessitate significant processing power on the edges. Most Cloud providers' enterprise offerings do not entail simplified clients, though management and control of the applications and data become centralized.

Creating Developer Ecosystems: Firms competing in the Cloud Software Platform (PaaS) layers are interested in having large developer communities write applications for their platforms. The more developers write applications for their platform, the greater its value. Since interoperability standards for applications and data exchange have yet to be solidified or standardized, firms that already have large developer communities see it as an advantage. Microsoft, for example, has long been a company offering platforms, and has a large installed base of developers.

DELIVERY MODELS: PUBLIC/PRIVATE, INTERNAL/EXTERNAL

While our services "stacks" competition was centered on types of service models, we now turn to a second vantage into competition, that of *delivery* models. Variations in delivery models hinge on two axes: 1) the location of control and *management functions* and 2) the ownership and location of the physical Cloud *resources* (the datacenter). The NIST definition does not make this distinction, so here we adopt Cisco's enhancement of the NIST definition.

We first run through the basic definitions before introducing the variety of combinations.

Figure 2: Typology of Private/Public + Internal/External Cloud Deployments

Resources location / Control location	**Internal**	**External**
Private	Idealized "Private Cloud"	Virtual Private Cloud
Public	Managed by Cloud provider, housed locally	Classic "Public Cloud"

Location of Management Control

A Public Cloud refers to deployment models in which the control and management of Cloud resources reside with the providers, and are outside any one customer's firewall.

Private Cloud offerings are those in which control of the Cloud resources is located within a particular firm – inside its firewall.

Location of Cloud Resources

With **External** Cloud offerings, the physical Cloud Computing resources are located outside the customer's premises.

Internal Cloud offerings entail the management structure and infrastructure residing within the customers' physical buildings.

Deployment Combinations

Public/External deployments entail control functions as well as the physical infrastructure residing outside any one company. It refers to Cloud services offered to more than one customer using the same datacenter infrastructure. Services offered to the general public, as well as those offered to enterprises without special arrangements, all fall within this category. From the user's perspective, since they are one of any number of users, the resources they face have the illusion of infinite scalability, and providers can utilize scale attained by providing for a large number of users. Google Apps is a typical example.

Private/Internal deployments, in their idealized form, are within-enterprise Cloud deployments; a Cloud service provided by the enterprise for the enterprise itself. Deployment scale would be large enough that different parts of the company experience something similar to an illusion of infinite scalability, and billing can be metered, with divisions charged according to usage. The advantage of a private internal cloud deployment for enterprises is that they can take advantage of the flexible resource provisioning, but ensure that everything takes place within the company's firewall and within the company's physical buildings. Few enterprises other than Cloud providers are themselves large enough to have truly private/internal Cloud deployments with the scale needed to provide the illusion of infinite scalability. This leads to commercial deployments of virtual private and hybrid forms, which we will review below.

Private/External deployments give management control to the consumer enterprise, but with the infrastructure physically located off-site, in the third party Cloud provider's datacenters. In its pure form, the customer's infrastructure is physically separated from the rest of the provider's cloud deployment within the provider's datacenters. This may make strategic sense for the provider if the customer is extremely large, or customers such as major governments with stringent security requirements. While providers providing true Private/External Cloud deployments will not be able to use the rest of the infrastructure to attain scale, the importance of their relationships with large customers or governments may tilt their cost-benefit perspective towards this type of deployment.

Public/Internal deployments entail the offering configuration in which the Cloud provider manages the Cloud offering, but the datacenter is located within with the datacenter of the client. This is likely to be form taken by many government Cloud service offerings managed by Cloud provider corporations. In this example, the deployment is internal to the government, in that they are located within government premises and facilities. However, the management control resides with the firm providing the Cloud capability rather than the government.

Varieties of Cloud Deployment

Virtual Private Cloud deployments are the common commercial substitute for internal/private Cloud deployments in their idealized form. They are External deployments, in that the infrastructure is owned by the provider, but there are firewalls and software between the resources used by different customers (known as multi-tenancy). In short, it is a virtualized Private/Internal deployment. As long as adequate security is maintained, this gives customers Cloud resources within its firewall while allowing providers to efficiently utilize their datacenters to attain scale.

Virtual private deployment is arguably the most common form of deployment for enterprise users. For government users, there can be a variety of security requirements which vary according to country and agency.

Hybrid Private Internal/External deployments allow firms to take resources from both internal and external resources, but manage it themselves. This arrangement is optimal for firms that use different external Cloud services and want to manage the resources themselves, as though it were their own internal deployment.

Alternatively, users may want to use primarily internal resources, relying upon external resources for bursts of demand for loads that exceeds their peak capacity. This is also

an attractive option for enterprises that do not want to completely reorganize their own datacenters.

Deployment Modes and User Types

The correlation between deployment modes and user types is still being explored. It should be noted, however, that some Cloud providers have pointed out that rather than user types – for example, enterprises, government, SMEs, and individuals – it is the workloads, applications, and business models that are more important in determining the deployment architectures. This will become clearer as the market unfolds.

PART III: REGULATORY AND MARKET DYNAMICS

Part III examines several emerging market and regulatory issues facing different sets of users, and different types of providers. The issues are still developing, and the nature of the debates is still open; it is still very early in the game, especially from a regulatory perspective. We will therefore start with the most obvious but also the most major issues.

LARGE ENTERPRISES

Large enterprises, major consumers of computing resources with massive investments in internal ICT systems, represent a large potential market for Cloud providers. The adoption of Cloud architecture, however, is still not at a stage where enterprises drive the demand for a major overhaul of their IT systems in favor of adopting Cloud Computing architecture. Concerns over security, the fit between existing computing needs and Cloud offerings, and internal corporate organizational aspects in adopting Cloud are among the hurdles.

So far, there are indications that, among the largest early enterprise adopters, those that facilitate the operators of others, such as System Integrators and Service Providers, are the first to deploy Cloud architectures for their internal systems. Cloud services to user enterprises might be considered at this point part of a portfolio of IT offerings. Cost savings, and increased efficiency and flexibility are among the major selling points to large enterprises.

Incentives for Adoption: Currently Not Necessarily Demand-Driven

The first hurdle seen by several Cloud providers in getting enterprises to adopt Cloud architectures for their internal datacenters is that there does not appear to be a "market-pulling" magnitude of demand from enterprises at this point. They are not generally saying "we need Cloud services to solve our problems." It has been argued that a replicable pattern for enterprises to adopt Cloud services has not emerged yet—a replicable pattern analogous to past deployments of successive computing platforms, such as the shift from mainframes and "thin client" terminals to more decentralized but powerful PCs, or the gradual re-concentration of power with the advent of networked PC platforms.

For several types of business models, Cloud Computing presents clear benefits, for example, augmenting internal capacity with external capacity from Cloud services to meet burst demand. Therefore, the emerging first step for one group of large enterprises may be to adopt Cloud architecture as a supplement, retaining their own datacenters for core applications and sensitive data. For firms with internal applications that require rapid or massive scalability, internal Cloud architecture enables an efficient utilization of computing resources. Migration to external clouds of some form may be easier once the internal organizational and computing architectural changes have been undertaken. Virtualization of datacenter operations may be another entry point into adopting Cloud architectures.

It must be remembered, however, that the applications must have the capability for horizontal scaling and dynamic migration (databases recognizing that the source of data may shift from one virtual machine to another, for example) to fully utilize the benefits of Cloud Computing. Therefore, points at which enterprises are rewriting applications offer another potential entry point for Cloud service providers.

Security and Privacy

It is useful to distinguish between the concepts of security from privacy. *Security* is about protecting data from unauthorized access, while *privacy* refers to who is allowed to get access to data. Both are critical concerns for large enterprises, but for slightly different reasons.

As is often pointed out, large enterprises are highly sensitive to the security of their data – in many cases, once there is a leakage, the damage has been done. In general, therefore, enterprises are less willing than SMEs to have their data leave their premises, a barrier to outsourcing the management of data in general.

A challenge for Cloud providers is to provide credible guarantees that large enterprises' data will be secure from unauthorized access. This credibility is harder to sell to enterprises when it entails enterprises giving up direct management of their data, especially when that data no longer resides on their premises and particularly in an environment of multi-tenancy in the providers' datacenter. As noted above, vantages on how to best provide security differ among Cloud providers coming from different layers in the services stacks.

The challenges of privacy —who is allowed to access the data—is arguably more a regulatory than strictly technical matter. If regulations in particular countries legally allow governments to access information physically stored in providers' datacenters, this can be a serious deterrent to enterprises adopting any Cloud architecture that is External. A clear tiering, with the most sensitive data kept internal, with less sensitive data allowed to move to external Cloud deployments, seems to be the emerging pattern, but it is still early. As discussed in the regulatory section below, regulations in many countries over privacy are still unclear, and largely untested in courts.

Service Level Agreements (SLA)

The performance reliability requirements for large enterprises varies according to task and division, but overall, large enterprises are regarded as requiring higher Service Level Agreements than some Cloud providers can easily provide. Especially for service providers that do not provide infrastructure within customers' datacenters, or who do not control Access Networks, the possibility of service disruptions due to network or traffic problems outside their control can be serious. This in turn creates incentives for system integrators or PaaS providers to provide the ability to federate multiple cloud services to provide redundancy in the event that particular providers fail to provide sufficient SLA.

Internal Corporate Politics in Reorganizing IT

Implementing Cloud architectures internally can entail power struggles as the IT management structures change. Traditionally, enterprise IT departments consisted of server, storage, and network groups. However, with widespread virtualization, the role of hardware administrators decreases, replaced by "virtualization administrators" or "operations specialists." In short, many previous functions will be collapsed into monitoring the performance of Cloud services.[xvi] Internal fiefdoms of resistance, which can mobilize risk averseness over sensitive data, are a potential challenge facing Cloud providers. Systems integrators, which already have strong links to the CIO's office, contend that they have an

advantage over infrastructure layer players in their existing capacities to foster the necessary organizational shifts.

A similar point can be made from the vantage of the existing organization of IT, which varies across firms. One proposition is that organizations with autonomous, decentralized IT management will have greater difficulties in adopting Cloud, following the logic above. This may open up opportunities for firms that are able to provide solutions that federate multiple, autonomously deployed Cloud architectures by each group. Alternatively, it may open the room for Cloud providers to aid the central leadership in consolidating IT management by pointing to a technical need for integrating autonomous, decentralized IT operations to drive an organizational integration via Cloud implementation.

In any case, the ability of Cloud providers to work closely with varying pieces of firms to identify the most advantageous implementation of Cloud services, which can include internal political battles, is likely to be of key significance. This leads to a second organization point, that of access to the CIO as it can change with the adoption of Cloud architecture.

SME AND STARTUPS

The benefits of Cloud Computing to SMEs and startup firms are obvious, since costly upfront capital investments can be shifted into scalable service expenditures. It is generally useful to differentiate between three types of IT usage for SMEs and startups. First, they can be traditional businesses using IT as a tool. Second, they may be building tools for others, whether corporate or individuals, to use (for example, Zumodrive, Flickr or online backup services). Third, they can be providing components for Cloud services (third parties offering pre-configured Amazon EC2 "virtual machines" for example).

SLA and Security

SMEs and startups firms, in most cases, do not share the same degree of SLA and security concerns of large enterprises.[7] There is a general consensus that this advantages Cloud providers that are working their way from the top down. For providers working from the lower layers upwards, medium-sized firms that are willing to undertake a complete overhaul of their IT infrastructure provide a large opportunity for entry, but the challenge is how to present their infrastructure-based enhancements in SLA and security as value-added propositions.

Scaling Up SMEs/Startups

For startups with rapidly growing computing demands, Cloud services avoid the initial IT infrastructure investment costs. In the medium-term, however, some firms are finding that external Cloud services are not necessarily cheaper than internal datacenters, depending on the pattern and intensity of computing resources usage (such as large sustained demand for computing resources rather than fluctuating demand with high peaks, for example). The

[7] There are exceptions, of course, such as biotech startups which are often famous for their data security consciousness.

switching point, where startups or SMEs decide to move from an entirely external cloud to some internal capacity, offers the opportunity for Cloud providers to help create hybrid internal/external Clouds or virtual private clouds.

Community Clouds

For traditional businesses using IT, we are beginning to see trade associations offering Cloud services to members, or Cloud services targeting particular professions. For example, a set of Cloud services is offering lawyers and law practices a federated set of Cloud providers to offer applications, platform capabilities, and subscriptions to services used in the legal profession. Security is set up to isolate traffic.[xvii] This type of deployment is an opportunity for Cloud providers that can provide the platform for federating Cloud services, offering the ability of using a single common management interface for multiple Cloud offerings.

GOVERNMENTS

Governments' involvement in Cloud Computing should be differentiated between government as a *user* and government as a *provider*. It is still too early to observe major examples of government as a provider.

Moves by Cloud providers to offer Cloud services to Government users have received recent attention. Offering Cloud services to government users presents its own set of opportunities and challenges. Government systems, especially at the local levels, have been traditionally decentralized (having been built piecemeal, at different times for different purposes), and there is strong demand for higher security, tighter central management, and lower investment and operating costs. Government adoption of Cloud services is also a significant legitimizing tool for Cloud providers, who can point to potential clients that their services are trustworthy enough to have government clients.

Varying Security Requirements

Different governments, and different levels of government, have different security requirements. The most stringent require physical separation of the physical datacenter infrastructure from other users, and from other parts of the datacenter. At its most extreme, the physical servers and network infrastructure may require dedicated protection. In such cases, Cloud providers cannot utilize the benefits of scale that virtualization enables, and deployment decisions are guided by broader cost-benefit calculations (see next section). Some providers contend that the key for such stringent requirements is the *level of modularity at the datacenter level*. Some providers have a greater ability to build datacenters cheaply, and in a modular fashion, with the equivalent of ship containers with water and electricity hookups filled with servers as one unit.[xviii]

REGULATORY ISSUES

Technologically, Cloud Computing may be considered the ultimate global technology, since the potential physical location of datacenters could be anywhere, with geography-blind distributions of applications and data. However, as a practical commercial matter, national regulations do and will influence the actual deployment of Cloud services in countries around the globe.

In general, it is difficult to codify regulations and incorporate them into the service offerings themselves, as not only the regulations themselves, but their interpretations can change over time. Here we provide a brief overview of some of the emerging issues and potential dynamics in their nascent form.

Information Privacy Policy

It is worth reiterating that while information security concerns protecting data from unauthorized access, privacy refers to who has access to what data. Arguably the biggest emerging policy debates that will affect Cloud Computing deployments around the world are those surrounding information privacy. Without concrete guarantees on the privacy of data held by Cloud providers, the diffusion of Cloud services may be bounded by perceived risk in entrusting sensitive data to external Cloud services.

The differences between US and European requirements for keeping personal information or copyrighted data within national borders are well known. The regulations require some Cloud offerings to allow users to stipulate the country in which their data will be stored, to the detriment of scalability for the Cloud providers.[xix] In the US, auditing and compliance requirements for policies such as for the US Sarbanes-Oxley Act (SOX) and Health and Human Services Health Insurance Portability and Accountability Act (HIPAA) are stringent for large enterprises. In some cases, this hinders the casual use of external Cloud services by employees for experimentation "on the side" without going through the enterprise's central IT system.[xx]

National security policies are also a potential threat to information privacy. The application of US Patriot Act and Homeland Security Act, which has not been tested extensively in courts yet, raises concern for non-US firms, whose information can be accessed by the US government if stored on servers physically located in the US.

In the US, some court cases are beginning to show the limitations of expectations of privacy, but major questions have yet to be addressed. For example, the FBI used the Stored Communications Act to access a person's email without a warrant or his consent, and the court was unable to determine whether they were subject to any expectation of privacy. In another event, the FBI raided and seized servers from two Texas datacenters with search-and-seizure warrants for a wide range of equipment. However, included in the equipment seized were servers that had data of a large number of clients unrelated to the investigation. The district court sided with the FBI. A major question is whether the Fourth Amendment, which protects against unreasonable searches and seizures, applied to data held by Cloud providers.[xxi] This has yet to be directly tested in courts.[xxii]

Telecommunications Policy Debates

The position of Cloud services and Cloud service providers to traditional telecommunications policy will become a focal point as policy reforms attempt to catch up to technological advances. Debates are likely to emerge across advanced industrial countries, as each policy framework is driven by a different policy and political logic. There are several issues.

First, how are Cloud services categorized? This has implications for QoS requirements – conventional telephony, and in some countries, Voice over IP (VoIP). VoIP deployments are considered by some states part of the "social infrastructure" and hence are required to meet certain performance obligations. If Cloud services are considered an "enhanced utility," the debate will be over whether that entails performance obligations.

Second, how will Cloud providers be categorized? Currently, these categories matter in the strategies of Cloud providers because most incumbent communications carriers are prevented from moving downwards from the Access Network layer to offer Access Devices directly. However, firms that began in the upper layers are allowed to do so. If the Cloud offerings begin to converge significantly, incumbent carriers can argue that this is a competitively unfair arrangement – especially if firms from the upper layers purchase Access Network infrastructure.

This leads to further fuel for the network neutrality debates. Cloud providers that provide services over the public Internet prefer Access Networks to be commodity "pipes" that do not differentiate between users. Access Network providers (incumbent communications firms), on the other hand, in their continuing attempts to avoid becoming commodity pipes, are likely to continue fighting for the ability to leverage their network ownership to advantages in offering value-added applications – in this case Cloud services.

In particular the rise of Google to surpass AT&T and Verizon in the Internet traffic volume it carries, a product of aggressive investments into its own networks to bypass that of the public telecom carriers, is likely to raise this debate soon. Should third parties get access to Google's network, just as the telecom carriers are forced to do? Indications that Microsoft and other players that are investing massively into Access Networks will likely raise political concern as soon as their positions become clear in the statistical data.

CONCLUSION: WHERE TO GO FROM HERE

This document consisted of three parts, each of which can be put to different uses. Part I was directed at a general audience, "cutting through the fog" of terminology and laying out Cloud Computing as a combination of technologies and business models. It can be adapted to become a standalone document to provide a baseline for discussions about Cloud Computing as the market unfolds.

Part II introduced conceptual tools and frameworks to analyze the unfolding market. The services stacks can be useful in mapping the competition – one can take any particular service or Cloud provider, map it into the stack, and derive strategic implications. Moreover, the strategic vantages of the firms can be understood by mapping it onto an understanding of what other major providers are doing according to the map—for example by moving up or moving down, expanding into Access Networks or making a play in Access Devices. The inclusion of Access Networks and Access Devices into the stacks model makes it easier to see a more complete set of strategies, which can affect how the market unfolds, compared to stack conceptions that are limited to the Cloud offerings themselves.

Part II then introduced the different varieties of Cloud deployment models, Private/Public and Internal/External. These, taken as a separate dimension from the services stacks, are particularly useful in sorting out the issues facing different types of users, covered in Part III.

Part III examined the key issues facing different types of users. The hurdles for adoption by large enterprise, a large potential market, are still high. SMEs face a proportionally greater set of immediate upside benefits with lower hurdles for adoption. Regulatory issues, especially over privacy, loom large on the horizon, which could greatly shape the cost-benefit calculations for the adoption of Cloud services, especially by large enterprise.

Areas for Future Inquiry

This report has opened up several areas for future inquiry, in both the development of markets, and in emerging regulatory debates. First, we pose the proposition that the nature of how Cloud services are being used is likely to strongly shape its future trajectory of deployment. The role of Cloud services as a new business ecosystem for Silicon Valley startups, often mandated by venture capitalists, was alluded to in several of our interviews. It is worth following up to get a clearer picture of patterns of adoption and usage. Case studies of large enterprises in a variety of sectors, with a diverse mix of computing needs and business models, could provide a set of mechanisms useful in identifying patterns of adoption.

Second, an inquiry into understanding how the strategies of Cloud providers are unfolding should be sustained, as the terms over which they compete and positions they stake out in markets become clearer. In particular, the implications of Cloud providers moving down from high levels of the stacks, such as Google and Microsoft, moving into purchasing massive amounts of Access Networks should be continually analyzed for potentially shifting how security and SLA are delivered.

Third, given the concerns over privacy, especially for large enterprises in adopting External Cloud deployments, regulatory developments around the world should be closely

monitored. A map of the actors and debates as they emerge, which differ across advanced industrial countries and regions (court rulings in the US, European Union debates in Europe, government ministry strategies in countries such as Japan and Korea, for example), as well as emerging markets such as China and India, could be useful in following these debates. Broader policy debates over ICT networks and access will also require monitoring, as concepts such as network neutrality may take on a new light as firms in the higher levels of services stack such as Google and Microsoft move down to become major network operators.

Fourth, emerging talks over standards, though currently still in its very early stages, will be critical to follow. Industry standards or government mandates on interconnectivity, or requirements for particular types of information to be open and transferable, are likely to be one area focused upon. Another area may be in agreements over standards of management or SLA, to differentiate firms with good practices versus those with serious problems that could tarnish the entire genre of Cloud services. As talks begin to form, different arenas, such as the ITU or IETF, must be monitored; we are still so early in the game that the dominant arena is still unclear.

Finally, policy is often driven by crises. This raises the importance for Cloud providers to engage in sustained monitoring of the issues outlined here. They also need to monitor potential issues, and keep abreast of new issues as they arise. Only by doing so can they react quickly and take preemptive action if possible.

Thus, this document has attempted to cut through the fog surrounding Cloud Computing. It is our hope that the tools, conceptions, issues, and areas for sustained future monitoring put forth here will provide a clear set of expectations—expectations that will minimize the few surprises in the alignment of actors and markets as Cloud Computing markets and regulations develop and the fog lifts.

APPENDIX 1: THE NIST DEFINITION OF CLOUD COMPUTING

The NIST Definition of Cloud Computing[8]
Authors: Peter Mell and Tim Grance
Version 15, 10-7-09

National Institute of Standards and Technology, Information Technology Laboratory

Note 1: Cloud Computing is still an evolving paradigm. Its definitions, use cases, underlying technologies, issues, risks, and benefits will be refined in a spirited debate by the public and private sectors. These definitions, attributes, and characteristics will evolve and change over time.

Note 2: The Cloud Computing industry represents a large ecosystem of many models, vendors, and market niches. This definition attempts to encompass all of the various cloud approaches.

Definition of Cloud Computing:

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

Essential Characteristics:

On-demand self-service. A consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service's provider.

Broad network access. Capabilities are available over the network and accessed through standard mechanisms that promote use by heterogeneous thin or thick client platforms (e.g., mobile phones, laptops, and PDAs).

Resource pooling. The provider's computing resources are pooled to serve multiple consumers using a multi-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand. There is a sense of location

[8] http://csrc.nist.gov/groups/SNS/cloud-computing/cloud-def-v15.doc

independence in that the customer generally has no control or knowledge over the exact location of the provided resources but may be able to specify location at a higher level of abstraction (e.g., country, state, or datacenter). Examples of resources include storage, processing, memory, network bandwidth, and virtual machines.

Rapid elasticity. Capabilities can be rapidly and elastically provisioned, in some cases automatically, to quickly scale out and rapidly released to quickly scale in. To the consumer, the capabilities available for provisioning often appear to be unlimited and can be purchased in any quantity at any time.

Measured Service. Cloud systems automatically control and optimize resource use by leveraging a metering capability at some level of abstraction appropriate to the type of service (e.g., storage, processing, bandwidth, and active user accounts). Resource usage can be monitored, controlled, and reported providing transparency for both the provider and consumer of the utilized service.

Service Models:

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

Cloud Platform as a Service (PaaS). The capability provided to the consumer is to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages and tools supported by the provider. The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage, but has control over the deployed applications and possibly application hosting environment configurations.

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

Deployment Models:

Private cloud. The cloud infrastructure is operated solely for an organization. It may be managed by the organization or a third party and may exist on premise or off premise.

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

Public cloud. The cloud infrastructure is made available to the general public or a large industry group and is owned by an organization selling cloud services.

Hybrid cloud. The cloud infrastructure is a composition of two or more clouds (private, community, or public) that remain unique entities but are bound together by standardized or proprietary technology that enables data and application portability (e.g., cloud bursting for load-balancing between clouds).

Note: Cloud software takes full advantage of the cloud paradigm by being service oriented with a focus on statelessness, low coupling, modularity, and semantic interoperability.

APPENDIX 2: SOME MAJOR CLOUD SERVICE PROVIDERS

SaaS	Google Apps Microsoft Online Services IBM LotusLive Salesforce.com (enterprise software) SAP	
Paas	Google App Engine Force.com by Salesforce.com Microsoft Azure Services Platform	
IaaS	Amazon EC2 Rackspace	
DaaS	Amazon S3 EMC Storage Managed Service Sun, IBM, HP, AT&T Business, Verizon (virtual datacenters as part of system integration services)	
CaaS	Cisco, Junper, Brocade Communications System, Citrix Systems	
HaaS	Sun Microsystems (Open Cloud), Dell (Cloud Computing Solutions)	
CDN/ADN (Middle Mile)	Akamai Technologies Amazon Web Services	
Last Mile:	Telecom firms (Verizon, AT&T)	
Device Operating Systems	Microsoft Google Android, Chrome OS Symbian	
Access Devices	Nokia (from mobile to netbooks), Apple, HP + Dell (moving from netbooks into mobile), etc	

REFERENCES

Armbrust, Michael, Armando Fox, Rean Griffith, Anthony D. Joseph, Randy H. Katz, Andrew Konwinski, Gunho Lee, David A. Patterson, Ariel Rabkin, Ion Stoica, and Matei Zaharia. "Above the Clouds: A Berkeley View on Cloud Computing." http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.pdf.

Barroso, Luiz André, and Urs Hölzle. "The Datacenter as a Computer an Introduction to the Design of Warehouse-Scale Machines." [San Rafael, Calif.]: Morgan & Claypool, 2009.

Couillard, David A. "Defogging the Cloud: Applying Fourth Amendment Principles to Evolving Privacy Expectations in Cloud Computing." *Minnesota Law Review* 93, no. 6 (2009): 2205-39.

"Japan's Technology Champions: Invisible but Indispensable." *The Economist*, 2009.

KDDI. "Kddi to Join Google in Laying Japan-Singapore-Undersea Cable." http://www.kddi.com/english/corporate/news_release/2009/1102/index.html.

"Kddi to Join Google in Laying Japan-Singapore-Undersea Cable." Asia Pulse, http://www.antara.co.id/en/news/1260528035/kddi-to-join-google-in-laying-japan-singapore-undersea-cable.

Labovitz, C., S lekel-Johnson, D. McPherson, J. Oberheide, F. Jahanian, and M. Karir. "Atlas Internet Observatory 2009 Annual Report." http://www.nanog.org/meetings/nanog47/presentations/Monday/Labovitz_ObserveReport_N47_Mon.pdf.

MIC. "Smaato Kuraudo Senryaku, Chukan Matome [Preliminary Report on the Smart Cloud Strategy]." http://www.soumu.go.jp/menu_news/s-news/02ryutsu02_000023.html.

NIST. "Nist Definition of Cloud Computing V15." http://csrc.nist.gov/groups/SNS/cloud-computing/cloud-def-v15.doc.

Sridhar, T. "Cloud Computing - a Primer." *The Internet Protocol Journal* 12, no. 3 (2009).

Urquhart, James. "Does the Fourth Amendment Cover 'the Cloud'?" In *Wisdom of the Clouds*, 2009.

Youseff, Lamia, Maria Butrico, and Dilma Da Silva. "Toward a Unified Ontology of Cloud Computing." In *Grid Computing Environments Workshop*, 2008.

[i] Michael Armbrust et al., "Above the Clouds: A Berkeley View on Cloud Computing," http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.pdf.

[ii] The exact situations in which venture capitalists mandate startup firms to use Cloud Computing are still under investigation.

[iii] Virtualization technology is not new. IBM in the 1960s introduced virtualization to optimize the usage of its servers, enabling multiple software contexts on its mainframes.

[iv] Scalability, workload migration, and resiliency are critical attributes T. Sridhar, "Cloud Computing - a Primer," *The Internet Protocol Journal* 12, no. 3 (2009). For example, workloads or Virtual Machines can be shifted around physical servers with the operating systems and applications still running, avoiding downtime if some physical servers need to be shut down for maintenance or rebooting.

[v] The full advantages of virtualization must be captured by software. For example, databases experiencing load shifts from one virtual machine to another must be capable of dynamically adjusting to changes in the underlying virtual servers.

[vi] See http://blog.rightscale.com/2008/07/07/cloud-computing-vs-grid-computing/

[vii] Part of this stacks conception was drawn from Lamia Youseff, Maria Butrico, and Dilma Da Silva, "Toward a Unified Ontology of Cloud Computing," in *Grid Computing Environments Workshop* (2008).

[viii] Our definition of IaaS differs from some, which refer to IaaS as virtual machines. Ibid. In our conception, it is the overall category label for the constituent elements of computing power, storage, and communications offered as a service. We label computing power in the form of virtual machines as HaaS (Hardware as a Service).

[ix] NIST, "Nist Definition of Cloud Computing V15."

[x] Some CaaS offerings substitute PBXs and VoIP equipment for services. Others enhance Quality of Service (QoS) and security for networks, including "…dynamic provisioning of virtual overlays for traffic isolation or dedicated bandwidth, guaranteed message delay, communication encryption, and network monitoring. Youseff, Butrico, and Silva, "Toward a Unified Ontology of Cloud Computing."

[xi] The biggest difference between CDN and ADN services is that ADN is able to "see into" the data more to manage traffic loads and offer more security options. ADN services such as those offered by Cisco include installations in users' datacenters to provide enhanced performance and security through hardware network equipment.

[xii] C. Labovitz et al., "Atlas Internet Observatory 2009 Annual Report," http://www.nanog.org/meetings/nanog47/presentations/Monday/Labovitz_ObserveReport_N47_Mon.pdf.

[xiii] KDDI, "Kddi to Join Google in Laying Japan-Singapore-Undersea Cable," http://www.kddi.com/english/corporate/news_release/2009/1102/index.html. "Kddi to Join Google in Laying Japan-Singapore-Undersea Cable," Asia Pulse, http://www.antara.co.id/en/news/1260528035/kddi-to-join-google-in-laying-japan-singapore-undersea-cable.

[xiv] Some of Dell's other recent acquisitions include IT monitoring software vendor SilverBack Technologies and ASAP Software offering other IT solutions.

[xv] The competitive question for Access Network firms became how to add intelligence in the network to prevent themselves from becoming commoditized. Some players moved into CDNs and ADNs, but these were not the traditional telecom operators. Debates over "network neutrality" hinged on whether network providers could leverage the fact that they owned and operated networks to exert control over applications and services providers. Having been blocked by policy from doing so, they have been searching for ways to add intelligence in the networks as value-added propositions.

[xvi] Point from James Urquhart

[xvii] Example from James Urquhart

[xviii] Microsoft, and to a lesser degree Google, have been public about this architecture. Luiz André Barroso and Urs Hölzle, "The Datacenter as a Computer an Introduction to the Design of Warehouse-Scale Machines." ([San Rafael, Calif.]: Morgan & Claypool, 2009), http://dx.doi.org/10.2200/S00193ED1V01Y200905CAC006.

[xix] Even in the European Union, which has EU-level privacy policies, data privacy and data sovereignty are determined at the national level.

[xx] Koike, Ryoji. "Dark Side of the Cloud." http://www.blwisdom.com/trend/51/2.html

[xxi] See James Urquhart, "Does the Fourth Amendment Cover 'the Cloud'?," in *Wisdom of the Clouds* (2009).

[xxii] A recent issues raised is whether digital assets are treated as transactions, in the same manner that phone numbers dialed is seen as a transaction with a third party (so there is no privacy for the number dialed), or as physical assets locked in briefcases or lockers. David A. Couillard, "Defogging the Cloud: Applying Fourth Amendment Principles to Evolving Privacy Expectations in Cloud Computing," *Minnesota Law Review* 93, no. 6 (2009).