

Diffusing the Cloud: Cloud Computing and Implications for Public Policy

Kenji E. Kushida · Jonathan Murray · John Zysman

Received: 1 May 2010 / Revised: 20 January 2011

Accepted: 18 April 2011

© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract Cloud Computing is rapidly emerging as the new information technology platform. It is, however, much more than simply a new set of technologies and business models. Cloud Computing is transforming how consumers, companies, and governments store information, how they process that information, and how they utilize computing power. It can be an engine of innovation, a platform for entrepreneurship, and driver of corporate efficiency. While an increasingly commonly term, confusion remains over what exactly constitutes Cloud Computing, how the markets are unfolding, and what forces will drive their evolution and diffusion. This paper provides an overview and conceptual tools for business leaders, policymakers, and non-specialist scholars to identify, distill, and easily understand the core aspects of how Cloud Computing service markets are developing, and how an array of policy issues will influence how this new computing platform unfolds across the world.

Keywords cloud computing · IT services · information technology · political economy

K. E. Kushida (✉)

Shorenstein Asia-Pacific Research Center, Stanford University, 616 Serra St E301, Stanford University, Stanford, CA 94305, USA
e-mail: kkushida@stanford.edu

K. E. Kushida · J. Zysman

Berkeley Roundtable on the International Economy (BRIE), University of California Berkeley, 2234 Piedmont Ave, Berkeley, CA 94720, USA

J. Zysman

e-mail: zysman@berkeley.edu

J. Murray

Innovia Ventures, Zurich, Switzerland
e-mail: Jm@innoviaventures.com

J. Murray

School of International Studies, University of California Berkeley, 2234 Piedmont Ave, Berkeley, CA 94720, USA

J. Zysman

Political Science Department, University of California Berkeley, 210 Barrows Hall, Berkeley, CA 94720, USA

JEL Classification L63 · L86 · L96

1 Introduction: cloud computing, the next computing platform

Cloud Computing is rapidly emerging as the new platform for computing. It is, however, much more than simply a new set of technologies and business models. Cloud Computing is transforming how consumers, companies, and governments store information, how they process that information, and how they utilize computing power. It can be an engine of innovation, a platform for entrepreneurship, and driver of corporate efficiency. However, while the term is increasingly commonly used, confusion remains over what exactly constitutes Cloud Computing, how the markets are unfolding, and what forces will drive their evolution and diffusion.

This paper provides an overview and conceptual tools for business leaders, policy-makers, and non-specialist scholars to identify, distill, and easily understand the core aspects of how Cloud Computing service markets are developing, and how an array of policy issues will influence how this new computing platform unfolds across the world. The introduction of any new platform of this potential opens a wide range of opportunities for firms large and small, countries, regions, and the global economy. At the same time, it raises new challenges for firms, industries, economies, and policies.

1.1 Cloud computing driving transformations

Cloud services are already transforming the business models of companies around the world—from small startups and medium-size firms to large multinational enterprises. Cloud Computing offers new capabilities for innovation and entrepreneurship, lowering the bar for new entrants and facilitating experimentation. At the same time, it raises the scale required to be an effective computing infrastructure provider, and customer expectations of service flexibility and costs are undergoing a sea change.

For advanced industrial countries, Cloud Computing provides new opportunities for innovation and entrepreneurship, and promises substantial efficiency gains. For developing countries, Cloud services open up new possibilities to enter international markets and find niches in global value networks. As with the previous computing platforms—mainframes, PCs, and networks of PCs—Cloud computing is becoming a baseline for national and corporate IT infrastructure against which other forms of infrastructure and service delivery must be measured.

Cloud Computing offers unprecedented new levels of configurability for diverse groups of users. Services are *dynamically configured* to the needs of each user with a single unified, usually global-scaled, architecture. Cloud providers' scale and cost merits are available to all types of users, from individuals to multinational corporations.

Critical to Cloud Computing is the abstraction of end-user applications from the underlying hardware. Put simply, application software is not tied to any particular server or physical hardware; instead, it can utilize the massive scalability and resiliency of the underlying global-scale datacenters to deliver the same services to one user or several million users. In general, users do not know or care about where the datacenters are or how they are configured.

The abstraction of user-applications from the underlying datacenter infrastructure, combined with the truly massive scale required to harness the full potential of Cloud Computing makes it susceptible to a wide array of government policies. While the technological logic of Cloud services transcends traditional political and economic boundaries, national and regional-level differences will matter a great deal in its actual patterns of implementation. Policy issues such as information privacy, security, national

network policy, and jurisdiction are among the critical issues, and can be settled differently across countries. Moreover, the policy debates opened by Cloud Computing are beginning to reshape these debates within countries across the globe. Many of these policy debates reach the core of how governments interact with their citizens, and can shape a possible international regulatory architecture over Cloud services.

1.2 Roadmap of the paper

Both the central public policy and the business strategy issues about the Cloud are embedded in the details of this new computing ecosystem. This paper begins to provide a guide to policy and strategy debates, both of which will be profoundly recast by Cloud Computing. The paper unfolds in four parts, following a simple logic of inquiry. In Part I we answer the question: *what exactly are Cloud Computing services?* We provide an operating definition of Cloud Computing as a “dynamic utility,” presenting the major characteristics that define Cloud services and differentiate them from traditional utilities and previous IT platforms.

Part II asks the question: *How can we make sense of the emerging, complex and confusing Cloud services ecosystem?* We offer a simple but powerful typology, the “Cloud Services Framework,” that maps different types of *Cloud providers* (Cloud Services, Access Networks, Access Devices) onto different *Cloud architecture types* (Infrastructure, Platform, Applications). This framework allows us to understand the business models in different parts of the Cloud ecosystem, and how they come together in the marketplace. It shows us how different sets of policy issues bear upon different types of providers and their business models.

Part III answers the question: *How is competition among the actual major Cloud providers unfolding?* Using our “Cloud Services Framework,” we analyze the business strategies of paradigmatic global Cloud providers Amazon, Google, Microsoft, Salesforce.com, other major IT vendors, and Apple.

Part IV asks: *How can we get the big picture for how the Cloud services ecosystem(s) is developing globally?* Here we augment our “Cloud Services Framework” with two additional variables to construct a deployment framework with the acronym PLUMS—Providers, Layers, Users, Modality, and Scope. We apply the framework to major global Cloud providers, comparing with a sample national case of Japan, an example that provides the basis for a comparative discussion of how Cloud ecosystems develop in a variety of places.

2 What is cloud computing?

First, what exactly is Cloud Computing? Firms are marketing a wide variety of services as “Cloud Solutions,” leading to some confusion, some of it deliberate. However, if overused to include all online services, it loses meaning.

Put simply, Cloud Computing delivers computing services—data storage, computation and networking—to users at the time, to the location and in the quantity they wish to consume, with costs based only on the resources used. Users simply procure from providers the “amount of computing” they want without needing to invest in computing infrastructure.

Cloud Computing abstracts the applications and software platforms underlying physical hardware. In most instances, a layer of software mimics hardware, “tricking” applications

into thinking that they are interfacing with physical servers when they are in fact interfacing with software-created “virtual machines.” 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 enables greater flexibility in how workloads are managed, and how datacenters are constructed, since providers can *dynamically* add, remove or modify hardware resources without having to reconfigure the services that depend on them.

Cloud Computing changes the location of data processing. In traditional models of computing, data and applications are local—usually a personal computer (PC) for consumers and private data centers for firms. In both cases, processing takes place at the “edge” of the network, close to users. The new Cloud model, however, moves both the applications and data into shared Cloud Datacenters, away from the edges and towards the center of the network.

The competitiveness of Cloud Computing service provision critically depends on providers’ ability to build out capacity at a scale far greater than any individual user or firm could afford. Aggregate demand can then be amortized over this highly scalable infrastructure and sold back to the user at a much lower—per unit resource—cost than users could provide themselves.¹

Cloud Computing is frequently referred to as a “utility” service, sharing many of the characteristics of other utilities such as contractual levels of availability, reliability, and consumption-based pricing. It is also ‘utility-like’ in the sense that Cloud providers are large companies operating at a significant scale, serving small users as well as giant corporations.

Cloud Computing does, however, differ from traditional utilities in several critical respects. Cloud providers compete aggressively with differentiated service offerings, service levels and technologies. Cloud services can be completely customized to the needs of the largest commercial users. Consequently, we have often referred to Cloud Computing as an “enhanced utility”.

But Cloud computing is much more than just an “enhanced utility”. It is a utility which can be configured and customized depending on user needs. In a very real sense Cloud computing is—as we will call it—a *dynamic utility*.

2.1 Cloud computing as a dynamic utility

Traditional utility services provide the *same* resource to all consumers. The electricity company does not care whether their electrons are used to charge the batteries of an electric car or to run a heater. Neither choice demands that the supplier generate different ‘types’ of electrons.

Traditional definitions of ‘utility’ also refer to the critical nature of these resource provisions for the effective functioning of the economy or society. The regulated nature of traditional utilities—through public rules and standards—recognizes the unique economic role of the resources: that all members of society should enjoy equal access to these

¹ The upper bound for the benefits of scale in Cloud computing are still unclear. For many technologies, the merits of attaining large scale reach a plateau, or begin to taper off after reaching a certain size. In other cases, such as Moore’s Law, Gordon Moore’s 1965 prediction that the number of transistors that could be fit on an integrated circuit would double every 2 years—effectively doubling the processing power of a CPU every 2 years—the maximum limit has not been reached yet. This is a topic to be explored in future research, with significant ramifications how Cloud computing services develop.

resources, (contingent on minimum tariff payments), and that providers tend to be monopolies or limited oligopolies. Does Cloud Computing qualify as a utility under this traditional definition? The answer is both yes and no.

As with a traditional utility, Cloud Computing resources are always available, they are paid according to the amount consumed, and can be consumed in any quantity. Services are delivered through Internet connections, and the provider does not care about the device used to consume the service. Users do not care about how providers technically configure or operate the service on the back end as long as quality and price are acceptable, and users are free to use the resources as they see fit.

Unlike a traditional utility, however, there is little public standardization of Cloud service interfaces. As Cloud services becomes increasingly pervasive, this will raise significant policy issues regarding public rights to interoperability between Cloud services from different providers, and the portability of data and applications from one provider's service to another. An early illustration is the dispute between Google and Facebook over the ability of users to move their contact data between the two (O'Brien 2010).

For *providers*, Cloud Computing services are not utilities. They are competitive propositions that differ from utilities in several important ways.

First, Cloud services are not commodities. Commodities are goods offering little value-added, that are interchangeable with others, and which compete primarily on the basis of price. Cloud providers are competing on value-based differentiation on attributes such as service level and functionality.

Second, Cloud providers do not enjoy inherent—geographic—lock-in of users. Public utility providers such as water or gas are granted local monopolies. By contrast, Cloud services are not geographically bound. Since only an Internet connection is required for users to access the entire market of Cloud services, it is as easy to provide the same service in Lima as it is in London. For this reason Cloud providers face pressure to create their own service level lock-in mechanisms. Once a user has been acquired—often at some expense in terms of marketing or other incentives—it becomes a business imperative to prevent that user from defecting to an alternate provider.

Third, the actual data bits delivered in Cloud services are not interchangeable in the way electrons, atoms, or molecules are substitutes in electric, gas, or water utilities are. Users care a great deal about the whereabouts of the bits carrying sensitive personal or mission-critical corporate data. However, most care far less about the location of the constituent bits of a photo or video, as demonstrated by the success of the Cloud services such as Flickr or YouTube.

Finally, major Cloud service providers make massive capital investments in global datacenter infrastructure—approximately \$500 million per datacenter—to give the illusion of infinite resource availability. Technology *allows* these datacenters to be located anywhere, and factors such as the cost of electricity, cooling, real estate, local tax incentives and network connectivity tend to drive location choices. There is rarely a technological or scale imperative that data centers must to be close to the markets they serve. Importantly, as we show later, this is where national regulations matter a great deal.

Perhaps the biggest difference between Cloud computing service and traditional utility service models lies in the degree to which Cloud services are uniquely and *dynamically configured* for the needs of each application and class of users. Cloud computing services are built from a common set of building blocks—equivalent and electricity providers turbines, transformers and distribution cables. Unlike the electricity provider, the Cloud provider configures those building blocks in unique ways for each specific application. The

building blocks required to deliver a global public email system are configured differently from those required to deliver an airline reservation system.

This ambiguity about the utility nature of Cloud Computing matters, and must be resolved. The lack of definitional clarity is not an abstract issue, since it impacts both business strategy and policy decisions in meaningful ways.²

Cloud Computing is poised to become an increasingly important, and even dominant, means through which the world's computational demands are met. In this respect Cloud Computing infrastructure will approach the same level of economic critical dependency as electricity, gas, water and telephony. The *dynamic* and highly *competitive* nature of Cloud computing service provision will challenge traditional approaches to utility regulation and policy making.

This paper proposes that Cloud Computing should be conceived as a *dynamically configured utility*: A service that combines the scalable, always-available, pay for use and—economically critical—attributes of traditional utility models with the configurable, technologically differentiated and non-geographically bound aspects required by a highly competitive services market.

2.2 Cloud computing as a driver of innovation, experimentation, and efficiency

What are the potential benefits offered by Cloud Computing—at least from the current vantage? (We must remember that few foresee the actual impact of truly transformative technologies. It took decades for the potential of electricity to be harnessed, and transistors have become far more ubiquitous than initially imagined (Cohen et al. 2000).) Cloud Computing is potentially transformative as a driver of innovation, a platform for entrepreneurship, and an enabler of entirely new business models. It also provides the baseline for efficiency and functionality that can become the lowest common denominator for the IT infrastructure of a modern business—something everybody has and cannot do without, much as a PC was a decade ago.

Cloud Computing offers a *platform for innovation and entrepreneurship* by lowering the bar for new entrants and facilitating experimentation. By removing the requirement for substantial capital outlays to build ICT capabilities, startup firms no longer require datacenters full of servers to introduce new services. They can rapidly scale up or scale down operations as needed, and they can experiment with highly computing intensive tasks. Larger firms that already own their own datacenters can also benefit from Cloud services. By procuring Cloud services from outside the firm, they can augment their internal resources with bursts of computing capacity for experimentation—the testing of a new online marketing campaign in a specific market, for example. Implementing Cloud Computing-style architecture within their own datacenters enables a more efficient allocation of internal IT resources, to the extent possible with their corporate organization.

Cloud Computing is becoming the *baseline for efficiency and functionality* for firms' IT infrastructure. Global Cloud providers' scale and ability to aggregate the demands of multiple firms enabled them to offer much lower total operating costs than consumers' own infrastructure. Cloud providers can also upgrade service capability in real-time without requiring costly IT infrastructure upgrades by end users. This enables accelerates the implementation of new technologies and approaches that can create significant competitive advantage.

² See the work of UC Berkeley linguistics professor George Lakoff for how conceptual framing can significantly shape policy debates. For example, Lakoff (2004).

There are, of course, still unresolved questions and potential risks associated with Cloud Computing services. For example, what is the business resumption strategy in the event of a catastrophic failure, though unlikely, of the Cloud providers' services? While many larger firms will choose to retain some on-premise capacity, smaller firms may not have that luxury, becoming entirely dependent on a particular Cloud provider for their business operation infrastructure. Cloud service reliability also critically depends on network service provision; no network, no Cloud. This raises serious implications about different national contexts, with different regulatory regimes governing who can provide network infrastructure and the rules under which they operate.³

3 Cloud providers, cloud users, and the networks that link them

As the business ecosystem surrounding Cloud Computing develops, it is critical to have a conception of how it works, the diverse types of service offerings, and the way that services fit together. We need to understand how the Cloud works to identify where the commercial battles are taking place and what territory firms are competing over. A holistic understanding of the Cloud ecosystem is also critical for policymaking by diverse government agencies across various jurisdictions to avoid undue conflict and confusion for all parties involved.

3.1 The cloud ecosystem in summary

Let us first summarize key components of the evolving ecosystem. As with most technology stories, it is easy to get distracted by the minutiae of the underlying bits and bytes of Cloud Computing, especially given its technical complexity. Our contention, however, is that the competitive arena for these complex technologies and business models can be described in quite simple terms.

As in all markets, commercial battles in Cloud Computing revolve around “ownership”—control of the *end user relationship* and the ability to maximize value-extraction from that relationship. The intimate relationship between scale and cost efficiency in Cloud Computing operations makes this an acute issue. The faster a Cloud provider can grow its customer base and achieve scale, the faster it can lower the per transaction cost of offering the service. This *scale efficiency* effect can provide a significant competitive advantage to early movers who attract a large customer base.⁴

This is a simple story of the digital prairie, starring those who provide Cloud services (Cloud *Providers*), those that purchase and consume the services (Cloud *Users*), and those who connect providers with the consumers (*Connectors*).

Since this paper focuses on the emergent competitive and policy issues facing the Cloud ecosystem, providers and connectors are our primary object of analysis. We leave a full-scale analysis of different types of users, and what they can do with Cloud services, for a separate study.

³ The recent events in Egypt, when the government succeeded in essentially shutting down Internet access by blocking the communications lines carrying Internet traffic to and from the country, serves as a stark reminder of the possible disruptions of services requiring Internet access. There are also questions raised about how China's censorship of information applied through its Internet Service Providers is applied to Cloud services.

⁴ See Footnote 1.

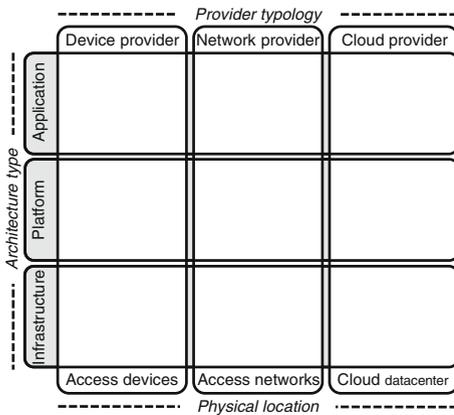
There are three types of providers and connectors. (The top horizontal axis of Fig. 1, Pane A.)

1. The *Cloud Providers* create, configure, run and distribute services from their *Cloud Datacenters*
2. The *Network Providers* offer *Access Networks* that enable the distribution of Cloud services from the *Cloud Providers* to *Users*.
3. The *Device Providers* offer the mobile phones, tablets and PCs through which *users* access Cloud services.

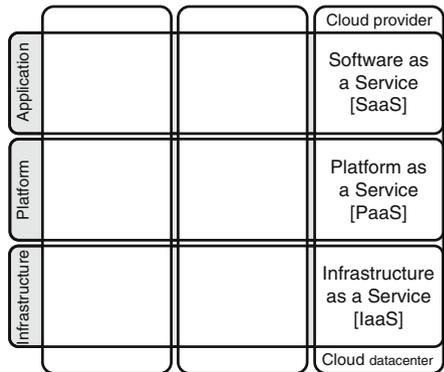
Cloud Computing’s technical, business, and policy issues play out across three layers of *technical architecture*. (The vertical axis of Fig. 1, Panel A.)

1. The *Infrastructure* layer encompasses the hardware, networks and operating systems responsible for managing fundamental resources such as data storage, computation and network bandwidth. A critical element of the Cloud Infrastructure layer is the ability to

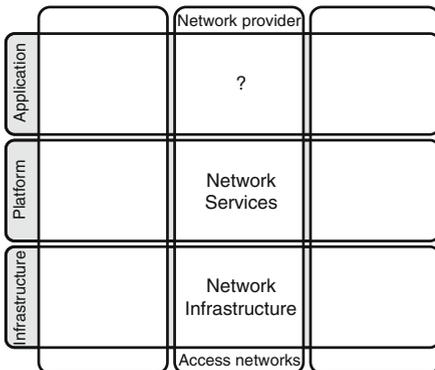
A. The “cloud services stack”



B. Cloud providers in the *cloud services framework*



C. Network providers in the *cloud services framework*



D. Device providers in the *cloud services framework*

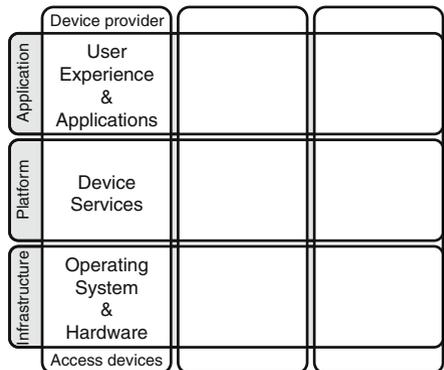


Fig. 1 The cloud services framework

virtualize the connection between physical resources and the services that consume them.⁵ Put simply, virtualization decouples applications and software platforms from the underlying physical hardware with software mimics hardware, “tricking” applications into thinking that they are interfacing with physical servers when they are in fact interfacing with software-created “virtual machines.” 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 enables greater flexibility in how workloads are managed, and how datacenters are constructed, since providers can *dynamically* add, remove or modify hardware resources without having to reconfigure the services that depend on them.

2. The *Platform* layer serves two purposes. It provides a set of common services, such as databases, messaging, and business rules engines, that are shared by applications. It also insulates application developers from the complexity of the underlying infrastructure through a set of higher level Application Programming Interfaces (APIs).
3. The *Application* layer provides the mechanism through which *users* interact with the Cloud applications—often through a web browser. In the *Cloud datacenter* the application layer is where the business logic for the application is run.

These two typologies constitute the two axes of our framework illustrated in Fig. 1, Pane A. Now we examine in detail the specific characteristics of each element. We proceed by moving through the provider typology, examining Cloud Providers, Network Providers, and Device Providers in turn. Each faces a specific set of policy issues that cannot be understood without a grasp of the technical architecture.

3.2 Cloud providers: provisioning cloud services

Cloud Providers, who provide services from their *Cloud Datacenter* fall into three broad types according to their technical architecture type. Figure 1, Panel B, shows how Cloud providers with different architecture types entail different business models. We start with the bottom right corner.

Infrastructure services—commonly referred to as Infrastructure as a Service (IaaS)—are virtual, Cloud-based replacements for physical hardware such as processors and hard drives. For example, Amazon offers virtual servers accessed through web interfaces, for which users pay by the hour per virtual processor. These virtual servers mimic the attributes of physical servers, although in reality the underlying processing power can be distributed over a large number of physical servers. They provide the flexibility of renting truly massive amounts of processing power for short amounts of time.

IaaS offered as standalone services can act as inputs to other Cloud service offerings. For example, a variety of Cloud service providers rent Amazon’s virtual servers and storage as their backend datacenter rather than building their own. For example, Animoto—a startup firm that automatically generates music to match users’ pictures and videos—uses Amazon’s Cloud-based IaaS services as its backend. When the service experienced sudden exponential growth as its popularity skyrocketed, it was able to add capacity almost instantly by increasing its use of Amazon’s virtual servers and storage, avoiding service slowdowns or outages.

⁵ 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. The software that mimics physical hardware is known as the “hypervisor,” and attributes such as processing speed and memory, which applications assume belong to the physical servers are actually created by software.

As inputs to others' Cloud service offerings, IaaS providers can benefit from the growth of Cloud applications and services as a whole, rather than being tied to the fortunes of particular companies. At the same time, they are commoditized more easily, since switching storage or processing power is relatively easy. This leads to pressure to attain massive scale to offer lower costs.

Platform services (Platform as a Service—PaaS) can be thought of as a virtual version of an operating system, such as Windows for PCs. Windows as a platform offers tools and interfaces for third party developers that take advantage of the underlying hardware; each piece of software does not have to include all the rules to control how the processor interacts with the memory and hard drive, how the hard drive stores information, or how the computer interfaces with the screen. Similarly, Cloud platform services provide software developers with tools that take advantage of the massive scalability and flexible resource allocation offered by the underlying Cloud Computing data centers. Microsoft's Windows Azure platform, Salesforce.com's Force.com, and Google's App Engine are examples of Cloud platforms.

PaaS providers' competitive position depends on the breadth and range of applications written to its platform—the attractiveness its ecosystem. Microsoft's dominance of the PC platform with Windows is the paradigmatic example. PaaS providers will compete on the basis of the unique attributes of their platform, the efficiency with which applications can be developed, and the size of the user population they can offer to third party developers.

Most PaaS providers monetize their services by charging developers to use the underlying processing power, storage and network capacity utilization, and other higher level services such as billing handled by the platform provider, optimized content delivery, and service-level guarantees. A few providers, such as Salesforce.com, harness their third party application market to enhance the functionality of their service offering itself—in this case its Customer Relationships Management (CRM) service.

Applications and Content Services (Software as a Service—SaaS) are the actual services, such as office productivity (Google Apps, Microsoft Office365 etc.), email, CRM, Enterprise Resource Planning (ERP) and the like, used by users.

A key advantage of Cloud-based software for users is that users essentially outsource the operation and maintenance of software. Upgrades happen automatically at the back end, eliminating the need for local technical support teams. Users can rapidly scale up the number of subscriptions or usage volume by paying more, without needing to redesign datacenters or undertake costly IT system upgrades. Salesforce.com's CRM, for example, is sold as a Cloud service accessed through a Web interface, with a monthly subscription fee structure.

Not only does the move to Cloud services eliminate the operational complexity and cost of installing, deploying and maintaining complex hardware and software systems in the users own environment, but it also alters the accounting for these services. Users' *Capital Expense* (CapEx) based accounting model for IT provision, in which investments are depreciated over time, can be transformed an *Operational Expense* (OpEx) based model, where expenses can be offset immediately against income.

3.3 The policy issues for cloud providers: global meets national

Cloud providers face a particular set of policy issues. By their very nature Cloud services are borderless since users only required Internet access and an Access Device, and for providers, the location of the datacenter(s) is irrelevant.

Major Cloud service providers such as Google and Microsoft distribute their datacenters across the world. A Hotmail or Gmail user never knows on what server, in which datacenter, and in which country their mailbox is stored. The technological advantages to this approach include significant levels of fault tolerance and disaster protection, a more responsive user experience regardless of location, and the ‘illusion’ of limitless scale provided by these services.⁶

However, every country in which the services are consumed, or in which the physical datacenters reside, has its own set of local policies and regulations bearing directly on electronic service provision and data protection.

Legal issues such as information privacy, security, and legal jurisdiction are highly nation-specific. In the US, for example, the Patriot Act allows the US government to demand disclosure of *any* data stored in *any* datacenter, *anywhere* in the world if that system is operated by a US-based company, broadly defined. That single law places US-based Cloud service providers such as Google, Microsoft, Amazon, and others at a great disadvantage when competing for business in foreign markets. Governments, even close allies, will think twice about using US Cloud providers if their sensitive data can fall under the reach of this act.

Data privacy is another area with no common global policy standards to which Cloud providers can adhere, and to which all governments will agree. In the US, the 1996 *Health Insurance Portability and Accountability Act* (HIPAA) protects personal health information and the 2009 *Health Information Technology for Economic and Clinical Health Act* (HITECH) prohibits disclosing health data to third party providers without specific business agreements and security requirements. The Gramm-Leach-Bliley Act protects personal financial information, and the Sarbanes-Oxley Act contains an array of reporting requirements for companies’ internal controls and reporting procedures.

Europe is a challenging regulatory environment for Cloud providers, with stringent data privacy regulations and substantial differences between individual states. In many European jurisdictions users must actively consent to the collection and storage of their personal information. Providers must disclose on request what information is stored and in general, data about European citizens may not be stored or processed outside EU borders.

Legal issues surrounding international jurisdiction and accountability have yet to be settled. For example, it is still unclear which rule of law applies for the arbitration of contract disputes—the country in which the service is consumed, or the country in which the service originated. In other words, is it Microsoft’s datacenter in Singapore or the US in which Microsoft is headquartered? Debates of this sort will result in much case law over the years ahead. Another US Patriot Act example further illustrates jurisdictional complexities. If the US government seizes some servers from a particular datacenter citing potential terrorist threats under the Patriot Act, those servers may host the data of many more—possibly international—clients than the suspected target.⁷ The issue is what right of protection for search and seizure those clients should expect. Amazon’s removal of Wikileaks

⁶ For a UC Berkeley vantage on Cloud Computing, in which the “illusion of infinite resources” is a key criterion of Cloud services, see Armbrust et al. (2009).

⁷ 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, applies to data held by Cloud providers. This has yet to be directly tested in courts (Urquhart 2009; Couillard 2009).

data from its Cloud storage environment may have been driven by this concern. We will return to these issues in Part III when delving into specific national deployment issues.

3.4 Network providers: connecting the user to the cloud

Network Providers provide the connectivity enabling users to consume the services provisioned and served from Cloud datacenters. These *Access Networks* are strongly shaped by national government policies, particularly telecommunications regulations and information access laws. The outcomes of policy debates in these areas can potentially drive substantive divergences in Cloud deployments across the world.

Following the architecture types again, we begin at the bottom (See Fig. 1, Panel C). The *Infrastructure* layer includes the physical means by which network services are delivered. This includes both landline and wireless. Corporate buildings are typically connected to fiber optic lines, providing extremely high-speed access. For homes and mobile connections, a diverse array of networks is deployed around the world. Broadband networks to homes include Digital Subscriber Line (DSL) that sends high-speed data over conventional copper telephone lines, cable, and Fiber-to-the-Home (FTTH). FTTH is by far the fastest, but the most costly. Wireless technologies range from evolutions of existing WiFi and Mobile WiMax networks to the next (4th) generation cellular technologies, such as Long Term Evolution (LTE).

Government policy strongly shapes the deployment of these technologies in their domestic economies, thereby determining the network environment available for user connection to Cloud services. There are various ways in which governments can facilitate the deployment of landline and wireless technologies. They can increase regulations on incumbent carriers, liberalize markets through deregulation or licensing new carriers, and actively facilitate the deployment of wireless technologies, for example. In Japan and South Korea, regulatory support led to rapid DSL diffusion, and in Japan's case extensive FTTH availability at low prices (Kushida and Oh 2007).

Insufficiently developed connectivity vis-à-vis other countries can hinder users from enjoying the benefits of Cloud services available elsewhere. Having the fastest possible networks, however, do not necessarily confer countries with an automatic advantage. In Japan's case, for example, the platform services taking advantage of those networks were limited to the domestic market. Therefore, services and applications predicated on Japan's high speed broadband and wireless networks were "trapped" in the domestic market.⁸

Telecommunications carriers, which generally own most of the network infrastructure that delivers Cloud services to users, are also highly regulated. In some cases, regulation precludes them from extending their reach and offering into adjacent domains. For example, in most advanced industrialized countries, national state-owned communications monopolies were partially privatized and their hardware manufacturing arms were spun out, preventing them from competing as *Device Providers*. In other cases, incumbent carriers were prevented from entering and dominating software and applications using data networks. In the US, for example, the FCC consistently prevented the Baby Bells from controlling the services and control layers of the telecommunications stack.⁹ This regulatory

⁸ See, for example, Kushida and Zysman (2009).

⁹ The telecommunications stacks consisted of the infrastructure layer, the control layer, and the services layer. The infrastructure layer was the physical network, the control layer included the protocols and standards necessary to use the physical networks for data communications (which mattered a great deal before TCP/IP became the de facto control layer standard with the advent of the Internet), and the services layer was the Value Added Networks (VANs) that businesses used in the early days to connect their computing resources. See Borrus et al. (1993).

stance was a critical enabler for a new class of Internet service providers, opening opportunities for others to use these services as a platform to innovate and experiment (Bar et al. 2000). This policy trajectory was critical to the emergence of today's Cloud service providers, who build upon the operational heritage of those early Internet service providers.

Free from the regulations imposed on *network providers*, Cloud providers such as Google and Microsoft have been free to build out their own substantial network infrastructures. This lowers their costs to carry data, obviating the need to pay other network operators. Control over more of the network also increases opportunities for experimentation. In 2009 Google ranked third worldwide in the total volume of Internet traffic carried over its own private networks, displacing AT&T and Sprint (Labovitz et al. 2009). Google was also involved in a number of well-publicized undersea fiber optic cable investments linking Asia to North America (KDDI 2009). With neither Microsoft nor Google regulated as telecom carriers, they have been free to engage in strategies prohibited to the latter, entering all three domains—*Cloud services*, *Access Networks* and *Access Devices*. As companies such as Google built out new Cloud-based telephony services—in this case Google Voice—allowing users free unlimited calls within the US from Gmail—incumbent network providers may make it an area of policy contention.

The critical business challenge for incumbent network providers is how to participate in the application layer of the Cloud ecosystem. Unlike the lower layers of infrastructure and platform, the application layer is less easily commoditized. Google Voice is at the application layer, and since Google derives its value from elsewhere, users are not paying directly to use the access networks to place their calls. This challenges incumbent network providers' business models of charging for use of their networks.

The ongoing debate about network neutrality—largely in the United States—is really a debate about who gets to extract value at the application layer. Cloud service providers argue that all bits should be treated equally when flowing over the network provider's access networks but this transforms those access networks into truly *dumb pipes*. Network providers argue they should be able to introduce differential tariffs or service levels depending on the type of data and the Cloud provider's willingness to pay for carriage. The introduction of new technologies to inspect the contents of data packets to assess their content (known as Deep Packet Inspection) and prioritize traffic (Anderson 2010), and network providers' attempts to inject their own advertising into data flows (Keane 2008) indicates that this debate will continue.

Incumbent *network providers* are also the firms through which governments pursue wiretapping and surveillance activities—both formal and informal. In the US, for example, carriers such as Verizon shared customer records with the US government after the Patriot Act was passed in the wake of the 2001 terrorist attacks. As these carriers enter Cloud services by offering Infrastructure resources—leasing storage and processing capacity by leveraging their strengths as a provider of reliable access to high speed networks—this intimate relationship with government security agencies can be a major concern for users. While governments need to establish new ties, and in some cases new legislation to obtain information from new entrants or startup firms, existing regulations and relationships are targeted at incumbent carriers.

3.5 Device providers: not your grandmother's telephone

It is a startling fact that a Formula 1 racing car is closer in heritage to a Model T Ford than a mobile Internet *Access Device* is to a plain old telephone from only 20 years ago. Apart

from the common functionality of being able to place voice telephony calls these two devices share almost nothing in common.

Today's Access Devices derive their lineage from computing equipment and like the personal computer forebears they have become one of the most intensely competitive battlegrounds of the technology industry. Winners of this battle will not only define the nature of the end user experience, but may also play a leading role in determining the nature and success of Cloud services.

The direct lineage from PCs to modern network-connected access devices suggests that similar competitive battles are likely to occur, albeit on a smaller and more mobile platform. There will be distinct battles: at the infrastructure layer over operating systems, processors and hardware; at the platform layer for the hearts and minds of the developer community; and at the application layer for the loyalty of end users. (See Fig. 1, Pane D)

The advent of Cloud service access on mobile devices and tablets has reshaped the nature of competition in hardware. *Device hardware* has become an ever-expanding category as manufacturers experiment with new form factors that serve the different uses that Cloud services enable. The present hardware ecosystem includes traditional PCs, netbooks, smart phones and tablet devices (such as the Apple iPad, Samsung Galaxy Tab and Motorola Xoom).

It is important to reflect that both PC hardware and traditional mobile handsets were rapidly becoming commoditized by the mid-2000s. It was the advent of a new generation of smart phones, spearheaded by Apple's iPhone and then the iPad, that reignited interest in—and reimagined—the role of access devices. Cellular handset manufacturers such as Nokia, Samsung, and Motorola rushed to introduce Internet capable smart phones following the success of Apple's iPhone. PC manufacturers such as Dell and HP also raced to offer tablet devices following Apple's successful iPad.

Below that surface there is aggressive competition for the central processing units (CPUs) upon which the software runs. The old battles between Intel and AMD for desktop processor supremacy have been replaced by a battle between Intel and multiple vendors producing processors based on ARM Holding's architecture. In this new incarnation of the processor wars, Intel is the underdog

Apple's control of its products over all three layers of architecture—from device hardware to platform to user experience—for both the iPhone and iPad differentiates it from other players in the market, and clearly factors in its popularity. Apple's profitability across the ecosystem it controls is the envy of others. Apple already provides a number of Cloud services tied to its devices—iTunes, Mobile Me, et cetera. Moving forward, Apple is likely to parlay this advantage into Cloud services to find ever-greater opportunities to extract value from the ecosystem. As it does so, it must tread carefully to avoid the scrutiny of competition authorities around the world, and to avoid the fate of other players dominant in their markets.

Google's initial foray into Access Devices in early 2010 with its NexusOne handset, manufactured by Taiwanese firm HTC, was an interesting and ultimately unsuccessful experiment. Google attempted to alter the dynamics of competition in cellular handset business by decoupling handsets from the traditional *network provider* channel and offering a new direct online retail channel. Google folded its online handset retail store later that year after failing to capture a significant market share.

In operating systems as well, old battles from the PC era are once again playing out in modern Access Devices. Again, the PC era's dominant player, Microsoft, is the underdog. Google's Android and Apple's iOS are battling for dominance of smart phone operating

systems. Nokia's recent endorsement of Microsoft's new Windows Phone 7 operating system is an acknowledgement that it could not compete in operating systems.

The Access Device battleground is one in which multiple business models are vying for market share (Kenney and Pon 2011). Apple's iOS operating system is integrated into its devices, and only available on Apple products. It includes a platform for third-party applications with Apple playing gatekeeper. Microsoft has always offered PC style-licensing for its mobile operating systems, charging manufacturers a license fee for each unit shipped. The sophistication and quality of the user experience Apple was able to deliver with its closed ecosystem model has been difficult for Microsoft to compete with. The company's latest attempt with the Windows Phone 7 operating system introduced in 2010 was favorably received but has yet to carve a significant market share. Google's Android operating system, by contrast, is free and largely open source, with Google extracting value through increased penetration of its Cloud services and advertising revenue streams on Android-based devices.

For PCs, Microsoft's dominance of operating systems became a policy issue in the US and Europe. Microsoft's bundling of its Internet Explorer operating system into Windows PC was the focal point of antitrust action. In the context of Cloud computing and Access Device operating systems, policy issues may develop around linkages between operating systems and search. One mobile operating system is unlikely to dominate in the way that Windows dominated PCs, but the ability for Google, with Android, and Apple, with iOS, leveraging a dominant position in one market to gain traction in another is already causing concern and scrutiny by regulators.

The *platform layer* on mobile devices is becoming increasingly sophisticated and significant in competition between access device and operating system manufacturers to entice developers and users. The breadth and depth of device platform services that can be delivered is directly correlated to the amount of processing capacity available on the device. The capacity available in today's high-end smartphones is already equivalent to the most powerful desktop PCs only a few years ago.

The canonical example of a platform layer service that has become the focal point of competition between device providers is the AppStore service. Each of the major device and operating system providers now offer an "Application Store" service, entailing a Cloud service to browse available applications, the commerce infrastructure to bill for applications and subscriptions, and finally the distribution, license management and update services required to provision the purchased application on the licensed device. This end-to-end capability requires services in each of the *Device*, *Network* and *Cloud* layers. Moving forward it can be expected that investments in *platform layer* common services, bound to backend Cloud services, will become a major competitive point of leverage for device providers.

4 The battle for the cloud

A distinctive characteristic about the emerging Cloud services business ecosystem is that firms from previously distinct sectors are rapidly finding themselves competing in the Cloud, as Cloud providers. These firms range from consumer online services firms such as Google, Microsoft and Amazon, established technology companies such as IBM and HP, Cloud-based startups such as Salesforce.com and Rackspace, incumbent telecom carriers such as AT&T and Verizon, and other major computer-related companies such as Apple. As a result of the diversity of these companies' origins and strengths, there are sharply

contrasting vantages on the best sources of value added activity—which areas are most lucrative, and how can firms carve out a sustainable market positions.

The struggle is for market advantage, margins, and defensible market positions, as we move into an era of Cloud Computing. The questions are similar to those posed at the advent of the open architecture PC in the early 1980s. Then we saw value shift from final assembly to its constituent elements such as processors, memory, hard disks, operating systems, and software. In this stage, Intel processors, Microsoft's Windows operating systems and Office application software captured the majority of value. We called this new competition "Wintelism" (Borrus and Zysman 1997).

In Cloud services, the battles over where the value is best captured, and how it is captured, are still raging. In this section we build a typology of Cloud computing that attempts to capture different activities and market segments, enabling us to see different vantages on competition, as well as how policy affects each layer.

Key competitive games are focused on who owns the *relationship with the user* and who will maximize the value generated from this relationship. This can be sorted according to the Provider types.

The three competitive games can be summarized as:

1. *The Device Wars*—Competition among *Device Providers* is pitting a new generation of mobile connected phones, tablets and PCs, and the operating systems on which they run against each other—Apple/iOS, Google/Android, Microsoft/Windows Phone, HP/WebOS, and more. The dynamics are elegantly described by Kenney and Pon.¹⁰
2. *Winning The User*—The central game of the emerging Cloud Computing marketplace positions major *Cloud Providers*—Microsoft, Google, Amazon, Salesforce, etc.—and a plethora of smaller players vying to win the hearts of minds of both consumers and firms with an ever widening and overlapping set of service offerings. We will explore these dynamics in more detail below.
3. *The Search for Network Value*—The emergence of Cloud Computing places even greater urgency on the *Network Providers'* search for value and relevance. Cloud Computing places ever-increasing demands on network bandwidth and infrastructure, but the lack of a compelling *application layer* offer places network providers at a significant disadvantage.

4.1 The business strategies of global cloud providers

Cloud Computing differs from a traditional utility because competing providers, despite offering common types of service, provide their version of these services in quite distinct ways. Service offerings may be interoperable between providers, meaning that data can be exchanged using common open standard protocols. However, they are not interchangeable, in that users cannot easily swap one provider's service with the same service type offered by another. In short, Cloud Computing services are not substitutable commodities competing on the basis of price.

Cloud Computing was not *invented* by any one architect—person or company. Today's Cloud offerings emerged from the infrastructures created by the major providers for their core businesses: for Amazon, the delivery of an online retailing service; for Google, a search and digital advertising business; for Microsoft, global web email and small business

¹⁰ *Kenney*, Martin & *Pon*, Bryan (10 Feb. 2011): "Structuring the Smartphone Industry: Is the Mobile Internet OS Platform the Key?", ETLA Discussion Papers, No 1238.

services; and for Salesforce, an online Customer Relationship Marketing application. Despite their entirely different business areas these firms share a common set of IT challenges—delivering continuously reliable (24 h a day, 7 days a week, 365 days a year), responsive service at global scale and low cost.

That these providers now generally share a common view and definition of Cloud Computing today is a result of the common business challenges shared by each of them. The difference in technology infrastructure starting point and development trajectory has resulted in a set of common Cloud services with quite distinct technical implementations. Cloud providers compete on the basis of these differences in implementation.

Understanding the broad historical trajectories of providers' development is critical to capturing the advantages and uncertainties associated with Cloud Computing services.

We trace the strategies of four major Cloud providers, Amazon, Google, Salesforce and Microsoft and approach this in two ways. One vantage is the Cloud ecosystem framework introduced in Part II. The second is by considering *target users*, and *providers' core businesses*. Target users are differentiated between the consumer/small- and mid-sized firms, and large corporations and government organizations. The preferences and needs of these two types of users contrast sharply.

Consumers and small businesses are usually more price conscious and may be less concerned about performance guarantees. *Large enterprises and government* users that invest millions of dollars in their IT systems need to balance cost considerations against a wide range of factors including reliability, security and performance. Many enterprise and government organizations are subject to regulations surrounding the handling of information, creating specific legal obligations. Integrating Cloud services with existing IT infrastructure and datacenters is also a major issue, since new Cloud services are usually extensions of existing business systems. Despite the challenges in meeting this myriad of requirements, the substantial IT budgets of large enterprises represent perhaps the largest new market for Cloud service providers.

We now turn to the cases of Amazon, Google, Salesforce, and Microsoft—paradigmatic examples illustrating the different types of services and contrasting strategies, each starting from a different place.

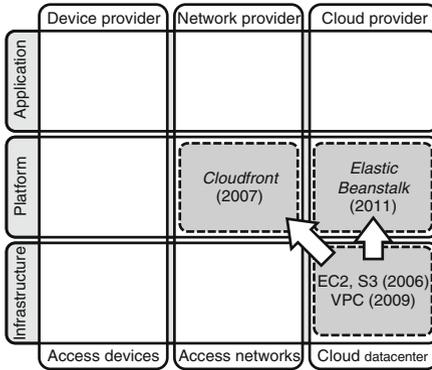
4.2 Amazon

Amazon was the pioneer Cloud service provider. It cultivated this new market out of the demands of its core online retailing business. This business is highly cyclical, with major demand peaks during the December holiday shopping season and serious lulls during the summer. Amazon's IT infrastructure had to handle the peak periods with excess capacity to ensure they could deal with demand increases. For most of the year, therefore, this capacity sat idle. The introduction of Amazon's 'Elastic Compute Cloud' (EC2) service in 2006 was an attempt to monetize this idle capacity.

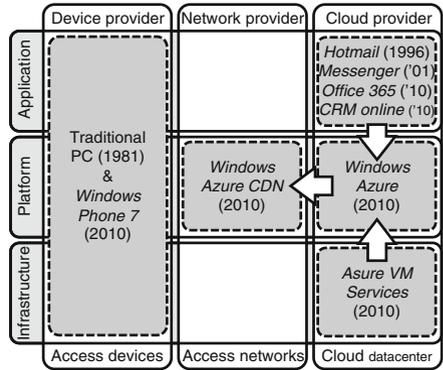
EC2 and subsequent services such as S3, a data storage service, are classical 'Infrastructure as a Service' (IaaS) offerings—basic component services that are combined by developers into higher level services and applications. Amazon offered EC2 and S3 as a dynamic utility, with users renting units of processing power by the hour and storage by the gigabyte. Amazon's rapid success with these initial services opened up an entirely new Cloud Computing business for the company. This new business has moved well beyond utilizing excess capacity from the core retail business, now driving its own investments in datacenters and new services.

Amazon has continued to build out from its IaaS roots both within the *Cloud Datacenter*, and into *Access Networks*. (See Fig. 2, Pane A—the physical location is the

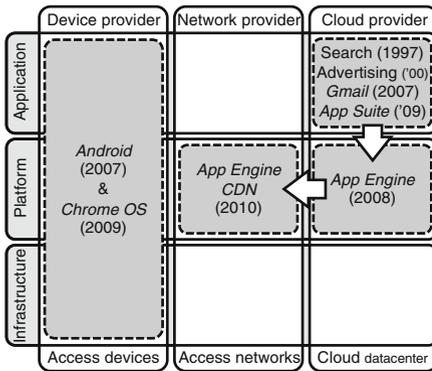
A. Amazon's strategy in the *cloud services framework*



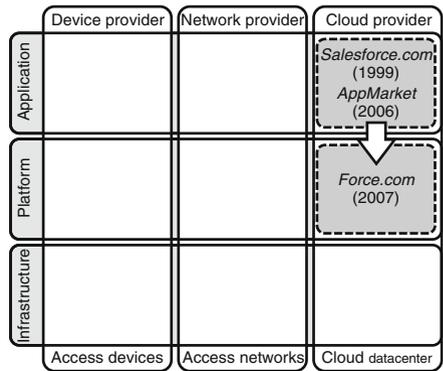
B. Microsoft's strategy in the *cloud services framework*



C. Google's strategy in the *cloud services framework*



D. Salesforce.com strat. in the *cloud services framework*



E. Apple's AppStore cloud service architecture

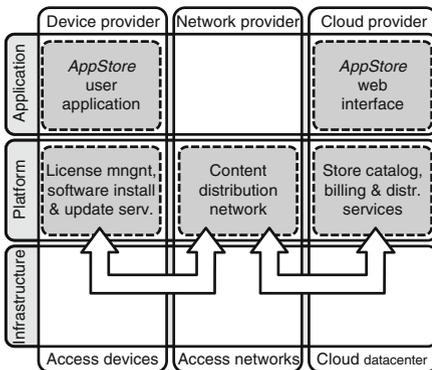


Fig. 2 Corporate strategies in the cloud services framework

bottom horizontal axis.) In 2007 Amazon introduced CloudFront, a Content Distribution Network (CDN) service. A CDN distributes copies of commonly requested information across several geographically dispersed datacenters, accelerating response times to users regardless of location. By 2011, Amazon's major datacenter locations outside the US included the UK, Ireland, Netherlands, and Germany in Europe, and Singapore and Japan in Asia.

Amazon strengthened its offering to enterprises in 2009, introducing its Virtual Private Cloud (VPC) service. This service helps create a seamless link between new Cloud-based services and a user firm's traditional IT systems.

The entry of several new, smaller and perhaps more agile competitors such as RackSpace, CloudSigma, and others, in the IaaS layer increases cost pressures on Amazon's core Cloud service business. In addition, major competitors such as Google and Microsoft have offered a broader range of services across the Cloud Data Center stack. Amazon's announcement of its 'Elastic Beanstalk' Platform as a Service (PaaS) service in 2011 can be seen as a direct response to these pressures. This PaaS offering is Amazon's move up the Cloud value chain, offering a higher level set of platform services to facilitate developers creating complex Cloud applications. This moves Amazon squarely into direct competition with Google and Microsoft for the hearts and minds of the next generation of Cloud application developers.

4.3 Microsoft

Microsoft, a dominant player of the PC era, is moving into almost all areas of Cloud services. In some respects, Microsoft's move into the Cloud is similar to that of Amazon. Both faced pressures from other core businesses which catalyzed their Cloud strategies—excess capacity requirements for Amazon, and for Microsoft, the need to decrease cost and increase scale in its (often free) online services such as Hotmail and MSN Messenger. These services date from the mid 1990s and early 2000s, well before any conception of Cloud computing.

In many respects services such as Hotmail and Messenger are—and always have been—Cloud services. They are massively scalable, with several hundred million global users, and available anywhere in the world via the Internet. However, until recently these services did not share a common infrastructure inside Microsoft. The drive to reduce operating costs and increase service flexibility led Microsoft to develop the underlying platform technologies upon which its Cloud service strategy is based.

Microsoft has always been in the 'Platform' business i.e. creating common interfaces and services that developers can use to build compelling applications. In the early days this started with Windows. As Microsoft's business evolved, it moved to provide platforms for the datacenter allowing firms and their developers to build new line of business applications. (See Fig. 2, Pane B) In the early 2000's Microsoft introduced a new set of platform technologies for web and Internet based developers. Finally in 2008, facing growing competitive pressure from Google and Amazon, the company introduced its Microsoft Azure platform, enabling developers to build Cloud-based applications.

Windows Azure charges for consumption of the underlying computing, storage and network resources on a per unit basis. Microsoft intends Azure to become a major new revenue source. As applications transition from traditional IT infrastructures to a Cloud-based delivery model Microsoft is competing to ensure that Azure becomes the Cloud platform of choice. This would mark a significant change in Microsoft's business model. When Microsoft sells a copy of Windows it makes a onetime gain on that sale. Microsoft

does not share in the revenue stream generated by 3rd party Windows applications. That will change as these applications move to the Cloud—hence the strategic emphasis on Cloud.

In addition to its PaaS offerings, Microsoft continues to build out its own suite of ‘Software as a Service’ Cloud applications such as Office 365 and Dynamics CRM Online. They aim to broaden Microsoft’s appeal to wider user audiences, but they are also defensive strategies for its core traditional application businesses. Microsoft needs to retain existing customers as they migrate to the Cloud.

The company is extending its reach vertically within the *Cloud datacenter*, introducing an IaaS service supporting ‘virtual machines’ on Windows Azure. It is also extending horizontally into *network services* by providing CDN services for Azure developers.

Finally, Microsoft’s heritage is based its strong position the PC operating systems, platform and applications. That position faces increasing pressure in recent years with the introduction of new classes of mobile, Internet connected *Access Devices* that do not run Microsoft’s operating systems—most notably Apple’s iPhone, iPad and the plethora of devices running Google’s Android operating system. Microsoft moved early into many of these categories, such as mobile phones with its Windows Mobile operating system and tablets running traditional Windows, but failed to keep pace with the competition. It remains unclear whether its new attempts—Windows Phone 7 and Windows 8 Tablets—will enable it to succeed in these important and growing categories.

Microsoft has invested billions of dollars to build out its global Cloud services data center infrastructure. The company is tightlipped about the locations in which it operates datacenters—primarily for security reasons—but was thought to be operating in around 18 locations worldwide at the end of 2010.

4.4 Google

Google, dominant in online search and advertising, now has a similar overlap with Microsoft in the breadth of its Cloud services, but with very different starting points and trajectories, its business model contrasts sharply with that of Microsoft. While many consider company a search engine provider, it is now actually a digital advertising company that uses its search engine for lead generation. Google’s success in monetizing search through digital advertising services has provided the capital to experiment and expand into other areas as it seeks new longer term revenue streams and diversification. The company’s massive investments into Datacenters and Access Networks underpinning its search and advertising business provided the launch point for a range of additional Cloud services.

Google began with the introduction of Google Apps providing a Cloud-based alternative to Microsoft’s ubiquitous Office applications. (See Fig. 2, Pane C) Google Apps are true Cloud-based applications, delivered from Google’s datacenters and consumed through a standard web browser. Users’ files are also stored in the Cloud, making them accessible from anywhere through any Internet connected device.

Google went on to extend these services by adding email and other collaboration services. While they started as consumer offerings, Google is quickly moving to add features making them attractive to firms of all sizes, including largest enterprises and governments.

Google shares a similar trajectory to Microsoft’s more recent Cloud strategy. Both companies started by delivering large global scale consumer and business application

services (SaaS), moving down the stack to build capabilities to attract the Cloud developer community. In 2008 Google introduced its App Engine PaaS offering ‘Platform as a Service’ (PaaS) offering that allows developers to leverage some of the powerful underlying services Google had developed to enable its own search and advertising businesses.

Google’s App Engine PaaS offering is limited to a narrower set of technology choices and configurations than alternative services offered by Microsoft and Amazon. It remains to be seen whether competitive pressures drive Google to offer a more generalized service, including IaaS services which compete with Amazon, that might appeal to a wider set of developers and customers.

In 2008, Google released its Android mobile operating system which moved it firmly into the Access Devices stack. Google has continued to maintain Android as an open source platform, supported by a consortium of carriers and software providers advocating for open standards for mobile devices and application platforms. Google does not generate revenue directly from Android but clearly believes that its participation in the Access Devices stack will drive more traffic to its other Cloud services.

4.5 Salesforce.com

Salesforce.com is by far the smallest Cloud provider included here, but it is widely recognized as pioneering of the “Software as a Service” (SaaS) business model. The company was founded in 1999 with a single product focus, CRM. Salesforce’s unique approach was to make its software available over the Internet without requiring customers to take on the complex and expensive burden of installing and maintaining the software on their in-house IT systems. At the time, prevailing options entailed implementing on-premise products such as Siebel Systems’ offering. These cost several tens of millions of dollars for large customers, largely driven by the technical complexity of integrating on-premise CRM software with customers’ existing systems. Salesforce’s delivery of the same functionality *as a service*, avoiding on-premise implementations, was a key driver of its rapid success.

The traditional approach to acquiring CRM software entailed a substantial upfront expenditure based on the number of potential users. This was a significant barrier to adoption by small and mid-sized business customers. Salesforce inverted this sales strategy by using a flexible, pay-as-you-go monthly fee structure attractive to smaller users, significantly lowering upfront costs and adding the ability to quickly scale usage up or down. The company soon expanded to larger users, often by getting individual business units or divisions to adopt its service before making a pitch to the central decision makers (Benioff and Adler 2009). As the product matured and expanded in functionality the financial advantages became increasingly compelling for larger firms.

The process of managing a customer relationship in the automotive industry is entirely different from those of pharmaceuticals or travel; every industry has its own unique approach and processes. Yet there remains a core set of information and processes that are shared by all industries. Salesforce.com’s strategy has been to focus on providing these core CRM services while encouraging the development of a supporting ecosystem of software developers focusing on the requirements of specific industries and other niche markets. In 2006, Salesforce.com formalized this arrangement by introducing its AppExchange marketplace. The AppMarket provided marketing, commerce and distribution service for third-party developers who wrote extensions to Salesforce.com’s core CRM application.

The success of Salesforce.com's AppMarket service became a major competitive strength. Potential customers were now attracted not only to the core functionality of Salesforce.com offerings, but also to the richness of the surrounding third party application portfolio. In essence, Salesforce.com turned its CRM service into a platform, a set of common services upon which other developers could build.

In 2007 Salesforce.com further formalized this *platform strategy* approach when it introduced Force.com, its Platform as a Service (PaaS) offering (see Fig. 2, Pane D). Force.com extended the AppMarket service with a set of tools and lower level services for developers, making it easier to develop applications for the Salesforce.com ecosystem, and for which Salesforce.com would provide the hosting and distribution.

Salesforce.com's development trajectory as a Cloud provider follows a very similar path to both Google and Microsoft. In each case these providers started in at the Software as a Service layer of the stack. The operational pressure to optimize cost, performance and scale led to the development of a shared set of platform services which ultimately were made available to the 3rd party developer community through Platform as a Service offerings.

4.6 The broader cloud provider ecosystem

Cloud Computing has become the buzzword de jour for the marketing department of any self-respecting technology company. The problem for external observers is separating those companies that are true Cloud providers from those who use the terminology to re-brand more traditional and less flexible offerings. With this in mind we consider a number of the other players as both illustration of the dynamic evolution of the broader Cloud ecosystem and some of the definitional problems in this nascent market.

4.6.1 IT Infrastructure Vendors: VMWare, Cisco, EMC

In 2009 three leading infrastructure component vendors—VMWare (Computation virtualization), Cisco (Networking) and EMC (Storage)—announced the formation of the "Virtual Computing Environment (VCE)" Coalition. Cisco, with SaaS offerings such as its WebEx virtual meeting service, is closest to the Cloud provider definition outlined in this paper. However, unlike the four providers above, none of the companies offers a wide range of services in each of the Cloud Datacenter layers

EMC, Cisco and VMware are important providers of the basic components upon which other Cloud providers build their service infrastructures, but each faces significant competition in their core markets: Microsoft's virtualization offerings with VMWare; Juniper Networks and others with Cisco for core networking infrastructure; and HP and others with EMC for storage technology. The VCE Coalition is an attempt to attain scale and to offer a *one-stop-shop* for companies building out their own Cloud service capabilities. The degree to which this moves beyond a marketing and joint sales strategy towards an integrated R&D strategy remains to be seen. However, the combination of VMware, Cisco and EMC will likely remain a formidable force in the provision of Cloud infrastructure components in the years ahead.

4.6.2 Major IT outsourcing vendors: IBM, EDS, CDC

Ten years ago, the buzzword in IT infrastructure circles was *outsourcing*. Major IT vendors such as IBM, EDS, and CSC competed to take over management of the IT infrastructures of

large industrial clients and newswires were filled with stories of multi-million dollar contracts. Unfortunately for the industry, the long-term cost structures of outsourcing proved not to be so attractive. The outsourcing providers quickly discovered that it was far harder to drive down the operational costs as they aggregated the systems of multiple clients than they had anticipated. Clients rapidly discovered that provider's motivations to drive down costs resulted in infrastructures that were far less adaptable to changing business conditions than they needed. The root problem lay with the inflexibility and complexity of traditional vertically integrated IT infrastructure.

These traditional outsourcing providers were quick to rebrand themselves as Cloud providers, but the reality remains questionable. To meet the definition we adopt, these vendors would need to offer IaaS and PaaS services with the dynamic scaling, management and configuration capabilities, and a pay-as-you-go consumption pricing model, as offered by the likes of Google, Microsoft and Amazon. It remains to be seen whether these traditional outsource providers do indeed deliver fully fledged Cloud service infrastructure.

4.6.3 Competition in Infrastructure as a Services (IaaS) and Platform as a Service (PaaS)

The IaaS layer is currently experiencing an explosion of entrants. Companies such as RackSpace, CloudSigma, Joyent and others are entering the market in direct competition with Amazon. It remains to be seen whether these new entrants can build sustainable and defensible business models. IaaS is the layer closest to being truly commoditized in the Cloud Datacenter stack. There is not a great deal of opportunity for competitive differentiation in the provision of basic compute, storage and networking infrastructure services. Ultimately services and service levels will be very similar, and price is likely to become the key point of competition. Since the key to reducing operational cost and therefore pricing flexibility is scale, success in the IaaS layer will depend on each competitor's ability to build scale over time. The current period of rapid expansion in the number of IaaS providers is therefore likely to give way to a period of consolidation as the successful providers gain scale by acquiring smaller, less competitive providers.

A number of the IaaS players are attempting to follow Amazon's development trajectory by moving up the stack, offering Platform as a Service (PaaS) capabilities. PaaS offers wider opportunities for differentiation. Joyent is currently beta testing a new PaaS service called "no.de", based on a new high performance development framework called "node.js". As a first mover in offering "node.js" services, Joyent hopes to capitalize on the growth potential of this development framework. This is an innovative but risky strategy. Success depends on the degree to which developers eventually adopt this new development platform and, should it gain momentum and developer mind share, nothing prevents a larger provider from offering node.js support within their own PaaS service and outcompeting Joyent on price.

4.6.4 Open source

The emergence of Cloud has created an interesting inflexion point for the Open Source community. The Open Source business model revolved around making the software freely available while charging for services such as support and maintenance. Cloud Computing moves in this direction since customers no longer pay for software licenses, but rather consumption-based service fees. Several other aspects of Cloud Computing, however, act very much against the Open Source philosophy. Setting up as a Cloud provider of any

reasonable size requires a substantial capital investment which few organizations in the Open Source ecosystem would be able to sustain. Also, the software systems of the major Cloud providers are only accessible through the interfaces the providers choose to open to the outside world. Unlike software installed on local PCs and servers this prevents the Open Source community from inspecting and reverse engineering these systems. Richard Stallman, president of the Free Software Foundation and founder of GNU, has been very vocal about the “evils” of the SaaS model (Stallman 2010). The ‘OpenStack’ organization is tackling the problems head on by driving an initiative to develop an entire Open Source based IaaS layer, which will then be freely available to anyone wishing to implement an IaaS offering. OpenStack derives from work initially carried out by NASA on its ‘Nebula’ Cloud service and has now attracted substantial backing from major players such as RackSpace and Dell.

4.6.5 Apple

Apple is the outlier in this discussion of the broader Cloud ecosystem, but is clearly a major player. The company has established itself as one of the dominant providers in the Access Device stack with the iPhone and iPad. The question is whether or not it should be considered a Cloud service provider. A qualified yes is our view. When you tap on the ‘Weather’ app on your iPhone, for example, you are initiating a connection with a Cloud SaaS layer service which provides weather information. It is the same for stock quotations, maps and driving directions, and music services such as Pandora or SoundCloud. Each of these iPhone services is the user interface to a Cloud service accessed through the Internet. None of these services are managed or run by Apple; it merely provides the device and the operating system enabling those connections. However, consider how the application arrived on your phone in the first place. That was achieved through the infrastructure provided by Apple’s AppStore service. The AppStore is a very large-scale Cloud service that leverages multiple layers of the Cloud Datacenter, Access Network and Access Device stacks (see Fig. 2, Pane E)

Apple is unique as a Cloud provider in using Cloud to primarily serve the needs of its own vertically integrated ecosystem. The AppStore is not available and does not serve up applications for Android or Windows Phone 7. iTunes is also a very large-scale Cloud service which enables you to browse, buy and download music and videos from your iPhone or iPad over your Internet connection. Finally, the company continues to expand its MobileMe service that provides a range of Cloud based capabilities including, email, storage and device management. In all these respects Apple can be viewed as one of the largest global Cloud providers operating today and the company’s recent billion dollar investment in a North Carolina Cloud Datacenter speaks loudly about its future ambitions.

5 The dynamics of cloud development

Next we turn to the question of how to capture the broader Cloud Computing ecosystem as it unfolds across the globe. Here we introduce a “Cloud deployment framework,” that is useful in capturing entire ecosystems. Any set of Cloud providers can be plugged into this framework, providing a comparative basis at the broader ecosystem level. That will then allow us to see how the policy issues can play out. We provide a brief overview of the framework with a preliminary sample analysis below, but leave an in-depth comparison for subsequent research.

The *globally local* nature of Cloud services, with provision using a global logic of large datacenters coupled with local configurability at the users' end, will lead to Cloud business ecosystems developing differently across countries and regions.

There are several drivers of diversity. *Rules and regulations* over various Cloud provider and architecture layers will differ, as seen in Part II—especially in areas such as privacy, security, and network infrastructure. Different *preexisting industry structures* of both IT providers and users, create different sets of opportunities and constraints for global and smaller Cloud providers. And the *roles that governments play* differ: many are enthusiastically embracing Cloud services for their own use; some are actively fostering diffusion; some are aggressively supporting domestic providers; and others will leave the pattern of diffusion more or less to markets.

The components of this framework combine three elements we have already covered, adding two more. They include the following, with different values denoted by letters in parentheses:

[P]roviders—The “provider typology” axis of our “Cloud services framework” (the bottom horizontal axis in the chart.) Values include:

Cloud Provider (C), Access Networks (N), and Access Devices (D)

[L]ayers—The “architecture type” axis of our “Cloud services framework”. (the vertical axis) Values include:

Infrastructure (I), Platform (P), Applications (A)

[U]sers—Users, both consumers (C) and firms (F)

[M]odality—Deployment models of Cloud services, explained below. Values are: Public (Pu), Private (Pr) and Hybrid (H)

[S]cope—The degree to which Cloud service providers deploy their services across geographic and corporate boundaries. Value include:

Global (G), Local (L) or Community (C)

Providers, Layers, and Users are directly from our discussion above. **Modality** corresponds to the “deployment types” referred in the industry and by the National Institute of Standards and Technology (NIST 2009). With *Public* deployments, the same Cloud services are offered to the general public as well as firms, without particular special arrangements with companies. Different users' data are managed by Cloud providers in the latter's Cloud Datacenters, and users have no control over where that data may be, or whom else is using the same Datacenter.

Large companies and governments, however, often prefer to keep highly sensitive information separated from the physical infrastructure used by the general public as well as other firms; they may insist on servers and datacenters dedicated to their exclusive use, sometimes often physically within their own buildings. These are *Private* deployments, in which Cloud providers make special arrangement with the user to provide dedicate infrastructure.

In practical terms, a purely private Cloud deployment loses crucial benefits of Cloud Computing itself. The benefits of scale—to lower costs for providers and to give the illusion of infinite resource availability for users—are limited. Cost savings from reducing idle capacity are also difficult to realize. Therefore, *Hybrid* deployments that combine Private and Public deployment types are increasingly popular for large corporate Cloud deployments. While the core of sensitive data or operations live on the user's private Cloud

infrastructure, the users procure additional resources from the Cloud provider's Public deployment. For example, a pharmaceutical company engaged in drug trials in which sensitive data must be kept within a private Cloud infrastructure may utilize Public Cloud services for its other activities.

Scope is the degree to which Cloud providers deploy their services across geographic and corporate boundaries. *Global* services are essentially borderless, serving any user who signs up from anywhere in the world, such as Google's Gmail or EBay's PayPal. *Local* services serve the needs of a particular geographic locality, such as a city government's electronic water tax payment service. Finally there is a class of *Community* Cloud services (a term used in the industry but which remains somewhat ill defined), which can be thought of as services that provide for the needs of particular groups of users—for example, applications for doctors or attorneys or members of a professional association.

The PLUMS Cloud service framework provides a valuable first step to understand national deployments. For example, if a particularly company in any country claims to provide "Cloud-based solutions," we can figure out if it is offering Applications, a Platform, or Infrastructure resources. If it is an Application, we can assess its competitiveness; if it is a Platform we can evaluate the reach of its developer community; and if it offers Infrastructure resources, we can ask what differentiates from other easily commoditized offerings. We can also determine whether it is targeted at the general public and small businesses or enterprise-scale companies to see how the national context provides opportunities and limitations. For example, countries with a large number of small-medium firms such as Italy or Taiwan are quite different from tightly knit Korean or Japanese business groups that are likely to have their preferred IT vendors.

6 Applying the PLUMS framework: a sample of future research

We apply the PLUMS framework to notable global Cloud service providers such as Microsoft, Google, and Amazon, as well as IT vendors such as IBM and HP, Cloud providers such as Salesforce and Rackspace, and social networking services Twitter and Facebook. (See Table 1).

Table 1 The PLUMS framework applied to major global IT firms

	Provider	Layer	Users	Modality	Scope
Microsoft	Cloud, Network, Device	Infra, Platform, Apps	Consumers, Firms	Public, Private, Hybrid	Global
Google	Cloud, Network, Device	Infra, Platform, Apps	Consumers, Firms	Public	Global
Amazon	Cloud, Network	Infra, Platform	Consumers, Firms	Public	Global
IBM	Cloud	Infra, Platform, Apps	Firms	Private	Local, Community
HP	Cloud, Device	Infra	Consumers, Firms	Private	Local, Community
Salesforce.com	Cloud	Platform, Apps	Firms	Public	Global
Rackspace	Cloud	Infra, Platform	Firms	Public	Global
Twitter	Cloud	Platform, Apps	Consumers	Public	Global
Facebook	Cloud	Platform, Apps	Consumers	Public	Global

It is clear that there is a clear differentiation between the major providers that are offer services at multiple layers, toward multiple types of users. IT vendors and specialized Cloud firms are focused on firm users, while social network services are focused on consumers. Unsurprisingly, their scope is global—a contrast to the sample case of Japan we present next.

The strength of the PLUMS framework is in capturing a variety of ecosystems. We present a sample case for this exercise by using Japan. Japan is a good example because it is the second largest market for IT investments after the US, and because it has provided major growth opportunities for global Cloud providers. For example, Salesforce.com's largest revenue driver outside the US is Japan, and the government aggressively adopted its platform in areas such as the postal system. Japan also has large IT vendors, but they tend to serve the domestic market. For the past couple years, the Japanese media has been awash with stories about the disruptions, adaptations required, and new opportunities opened up by "Cloud Computing."

So how does what does PLUMS tell us about the emerging ecosystem in Japan? We include major players in Japan's IT sector to provide a broad picture: the major IT vendors catering primarily to large corporations; telecommunications carriers; consumers electronics firms; the online service firms that have grown to become large presences in Japan's corporate landscape; and Japan's sizable and profitable social networking services. (See Table 2).

The picture we get is quite revealing. Few are expanding beyond their stack, and they are mostly staying local to Japan. IT vendors such as NEC, Fujitsu, and NTT Communications, are offering mostly private deployments to large corporations. There is some doubt about whether their "Cloud" solutions employ true Cloud Computing architecture, with the illusion of scalability and dynamic provisioning of resources. They are not moving into Networks or Access Devices, and they are not offering their services abroad in a large scale.

Japanese telecommunications carriers are not directly moving into access devices, but they are working closely with consumer electronics firms to create tablets and mobile handsets that are closely bundled with their online services. Their scale, however, is local to the Japanese market.¹¹ Japanese consumer electronics manufacturers (along with NEC and Fujitsu) long dominated the domestic market for cellular handsets with their sophisticated offerings, but were confined to Japan. Their market shares were assaulted by a tsunami of iPhone, and later Android-based smart phones since 2008. They responses, however, remain focused on the domestic market, with Sharp even naming its tablet "Galapagos"—a common label for Japan's mobile ecosystem in which isolation from global markets led to a sophisticated but distinctive evolutionary trajectory. Japan's major online services firms and social networking sites, which have succeeded more than their global counterparts such as eBay, Google, and Facebook, are also largely local in their deployment, with offerings limited to Japanese.

Overall, a significant observation is that there are few major Japanese firms moving to Platforms. Those that are limit themselves to the Japanese market. None are moving to offer the entire range of services types and architecture layers. Therefore, unfortunately for Japan, unlike automobiles, machine tools, and high tech components, we are unlikely to see anytime soon a broad surge of Japanese Cloud services firms that go beyond specialized niches.

¹¹ For more on Japan's IT ecosystem, see Kushida (2011).

Table 2 The PLUMS Framework Applied to Japan's IT Firms

	Provider	Layer	Audience	Modality	Scope
Major IT Vendors (NEC, Fujitsu, NTT Communications)	Cloud	Apps	Firms	Private	Local
Japan's Telecom Carriers (NTT, KDDI, Softbank)	Network	Network, Apps	Consumers, Firms	Public	Local
Japan's Consumer Electronics firms (Panasonic, Sharp, Sony)	–	–	Consumers, Firms	–	Local
Japan's Online Services (Rakuten, Yahoo! Japan)	Cloud	Apps	Consumers, Firms	Public	Local
Japanese Social Networking Service (Gree, Mixi)	Cloud	Platform, Apps	Consumers	Public	Local

This is a sampling of an in-depth comparative analysis of emerging Cloud Computing ecosystems around the world. In future research, will apply this PLUMS framework to areas such as Europe, and emerging areas such as China.

7 Conclusion

This paper cuts through much of the confusion surrounding Cloud Computing. We answer the most pressing questions in non-technical language, and providing conceptual frameworks to distill and simplify the complexity. These frameworks are a critical first step in understanding the market as it unfolds across the world, at a global scale with local differences. They also enable us to see how different sets of public policies, driven by different political dynamics, can act upon the various areas across provider types and architecture layers. The combination of these market dynamics and policy outcomes will shape the market. The main points can be summed up below.

What is Cloud Computing? Put simply, it is a dynamic utility enabling users to obtain computing as a utility-like resource, offered by providers as a competitive proposition at a large scale, but configurable to users' needs. Why should we care? Cloud computing is emerging as the next computing platform—a driver of innovation, platform for entrepreneurship, and baseline resource available to all firms. How can we make sense of the rapidly growing Cloud Computing market, since everything seems to have the Cloud label? Our “Cloud services framework” distills the activities of firms in the broader Cloud ecosystem into three types of providers, with three architectural layers. The framework shows how different sets of policies affect each type of provider. In particular, policies over information privacy, data security, interoperability and portability, which are likely to differ across countries, will influence how the Cloud markets develop. Our overview of the business strategies of major global carriers illuminates the origins and trajectories of their Cloud computing strategies. Finally, our PLUMS framework is a tool to usefully analyze Cloud deployment ecosystems across countries. With these tools, informed observers and participants have the critical basic information and policy issues that will shape how this next platform diffuses across the globe.

Acknowledgements The authors wish to thank the following people for their knowledge, feedback, and constructive criticism in putting this paper together: Stu Feldman, Patrick Keating, David Hytha, Martin Kenney, Petri Rouvinen, Dan Breznitz, and two anonymous reviewers. They also wish to thank the Alfred P. Sloan Foundation, Cisco Systems, and Microsoft for support to the Berkeley Roundtable on the International Economy (BRIE).

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Anderson N (2010) Deep packet inspection soon to be \$1.5 billion business. *Law & Disorder*
- Armbrust M, Fox A, Griffith R, Joseph AD, Katz R, Konwinski A, Lee G, Patterson D, Rabkin A, Stoica I, Zaharia M (2009) Above the clouds: a Berkeley view of cloud computing. www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.pdf
- Bar F, Cohen S, Cowhey P, DeLong B, Kleeman M, Zysman J (2000) Access and innovation policy for the third-generation internet. *Telecommun Policy* 24:489–518
- Benioff MR, Adler C (2009) Behind the cloud: the untold story of how Salesforce.com went from idea to billion-dollar company—and revolutionized an industry, 1st edn. Jossey-Bass, San Francisco
- Borras M, Zysman J (1997) Globalization with borders: the rise of Wintelism as the future of industrial competition. *Ind Innov* 4(2):141–166
- Borras M, Bar F, Berkeley Roundtable on the International Economy (1993) The future of networking. Berkeley Roundtable on the International Economy, Berkeley
- Cohen S, Bradford DeLong J, Zysman J (2000) Tools for thought: what is new and important about the “e-economy”. Berkeley Roundtable on the International Economy, University of California at Berkeley, Berkeley
- Couillard DA (2009) Defogging the Cloud: applying fourth amendment principles to evolving privacy expectations in cloud computing. *Minn Law Rev* 93(6):2205–2239
- KDDI (2009) KDDI to join Google in laying Japan-Singapore-undersea cable. http://www.kddi.com/english/corporate/news_release/2009/1102/index.html. Accessed January 10 2010
- Keane M (2008) British firm phorm trudges through the deep packet storm. *Wiredcom*
- Kenney M, Pon B (2011) Structuring the smartphone industry: is the mobile internet OS platform the key? *J Ind Compet Trade*. doi:10.1007/s10842-011-0105-6
- Kushida K (2011) Leading without followers: how politics and market dynamics trapped innovations in Japan’s domestic telecommunications sector. *J Ind Compet Trade*. doi:10.1007/s10842-011-0104-7
- Kushida K, Oh S-Y (2007) The political economies of broadband in South Korea and Japan. *Asian Surv* 47(3):481–504
- Kushida KE, Zysman J (2009) The services transformation and network policy: the new logic of value creation. *Review of Policy Research* 26(1–2):173–194
- Labovitz C, Ikel-Johnson S, McPherson D, Oberheide J, Jahanian F, Karir M (2009) ATLAS Internet Observatory 2009 Annual Report. http://www.nanog.org/meetings/nanog47/presentations/Monday/Labovitz_ObserveReport_N47_Mon.pdf. Accessed February 1 2010
- Lakoff G (2004) Don’t think of an elephant!: know your values and frame the debate. Chelsea Green, White River Junction
- NIST (2009) NIST Definition of Cloud Computing v15. <http://csrc.nist.gov/groups/SNS/cloud-computing/cloud-def-v15.doc>
- O’Brien L (2010) Facebook and Google’s Gmail contact conflict deepens. *Silicon Republic* 2011 (March 1)
- Stallman R (2010) Who does that server really serve? *GNU Operating System*
- Urquhart J (2009) Does the fourth amendment cover ‘the cloud’? *Wisdom of the Clouds*, vol 2010