

Cloud Computing: From Scarcity to Abundance

Kenji E. Kushida · Jonathan Murray · John Zysman

Received: 15 April 2014 / Revised: 29 September 2014/
Accepted: 3 December 2014 / Published online: 15 February 2015
© The Author(s) 2015. This article is published with open access at Springerlink.com

Abstract Cloud computing is a revolution in computing architecture, transforming not only the “where” (location) of computing, but also the “how” (the manner in which software is produced and the tools available for the automation of business processes). Cloud computing emerged as we transitioned from an era in which underlying computing resources were both scarce and expensive to an era in which the same resources were cheap and abundant. There are many ways to implement cloud architectures, and most people are familiar with public cloud services such as Gmail or Facebook. However, much of the impact of cloud computing on the economy will be driven by how large enterprises implement cloud architectures. Cloud is also poised to disrupt the Information Technology (IT) industry, broadly conceived, with a new wave of commoditization. Offerings optimized for high performance in an era of computing resource scarcity are giving way to loosely coupled, elastically managed architectures making use of cheap, abundant computing resources today.

Keywords Cloud computing · IT management · Disruptive technology · Commoditization

JEL Classification L86 · M15 · O14

1 Introduction

Throughout history, computing power has been a scarce, expensive resource. Now, with the advent of cloud computing, it is becoming abundant and cheap, driving a fundamental paradigm shift—a transformation from the *computing of scarcity to the computing of*

K. E. Kushida (✉)
Shorenstein Asia-Pacific Research Center, Stanford University,
Encina Hall E301, Stanford, CA 94305, USA
e-mail: kkushida@stanfordalumni.org

J. Murray
Innovia Ventures, New York, NY, USA
e-mail: jm@innoviaventures.com

J. Zysman
University of California Berkeley, Berkeley Roundtable on the International Economy (BRIE), 2234
Piedmont Ave, Berkeley, CA 94720, USA
e-mail: zysman@berkeley.edu

abundance. This revolution in computing is accelerating the *commoditization* of products, services, and business models and it is disrupting the existing Information and Communications Technology (ICT) industry. Moreover, the spread of cloud computing raises significant policy issues that have yet to be resolved, which will affect how it is adopted and develops around the world.

Cloud computing, in its simplest sense, is a computing resources management model. It is a method for pooling and sharing hardware infrastructure resources on a massive scale. Finite hardware resources are shared between competing demands, giving each user the illusion of exclusive control over the underlying environment and without the user needing to know anything about how the physical resources are configured. Cloud computing is uniquely new, simultaneously serving as an innovation ecosystem, production environment, and a global marketplace (Kushida et al. 2011). It affects the very underlying activities of production, influencing not only *where* work takes place, but also *how* it is done.

Cloud computing lowers the bar for entrepreneurship and innovation. Large, cutting-edge enterprises are also entering a new era of how they can organize themselves and compete. We view this as an integral part of the ICT-enabled services transformation that creates ICT-enabled services systems (Zysman et al. 2013). For general consumers, tasks involving computer processing and storage become ubiquitous—already, some estimate that a third of all websites visited by the general public are built on Amazon’s cloud service (AWS—Amazon Web Services) (Labovitz 2012). Email services such as Gmail and storage services such as Dropbox are offered without upfront fees, and video streaming service Netflix depends on Amazon’s AWS.

Much of the impact of cloud computing on the economy will be through how it affects enterprise computing, where lead users are moving to deploy cloud computing architectures for their computing to provide vastly greater flexibility and speed at far lower cost than ever before. Historically, large corporations have been the lead users of technology, responsible for driving new disruptions in production and economic activity by deploying new technologies in new ways *at scale*, shaping how the technology is adopted and used throughout economic activity.

The shift in computing resources from scarcity to abundance, combined with how lead enterprise users are reconfiguring their computing, is also driving a new wave of disruption for the Information and Communications Technology (ICT) industry, broadly conceived. The advent of cloud *platform* services is radically accelerating the *commoditization* of existing services; users are unlikely to pay a premium for solutions designed to optimize scarce resources—the previously dominant paradigm in successful business models. It is this commoditization of computing resources and services that gives rise to new possibilities for enterprise users of the technology to reorganize themselves and position IT as a strategic competitive weapon.

This paper asks the following questions: how did this revolution in computing come about; how is it unfolding; and what the implications are for industries and policies? It answers these questions in three parts. Part one delves deeper into understanding cloud computing itself since there is still confusion about what it is—to some degree intentionally exacerbated by IT vendors relabeling existing offerings. Part two examines the drivers of the disruption: how computing power transformed from a scarce to an abundant resource as the computer industry developed, and the role of lead users in driving new uses. Part three shows how the dynamics of the disruption will play out in various parts of the ICT industry, with rapid commoditization of previously high-end offerings. This paper concludes by pointing to critical policy issues raised by cloud, which will shape how cloud technologies, services, business models, and lead user adoption will unfold around the world.

2 What Is Cloud Computing?

First we must define and clarify cloud computing itself. While the term is increasingly commonplace, it is still often used synonymously and often for marketing purposes—in our view wrongly—for “the Internet” or “anything online”. These uses obscure the true transformative nature of cloud computing. The set of characteristics used by the U.S. National Institute of Standards and Technology (NIST) is the definition most commonly cited by academic work (Mell and Grance 2011). However, ours is more precise and discriminating, capturing the essence while excluding others. Our definition, derived from our previous work (Kushida et al. 2011), has withstood the rapid industry developments over the past few years.

A key point is that we distinguish between cloud computing itself and cloud computing *architectures*. Cloud architectures can be broken down into distinct layers, making it easier to trace how cloud computing is deployed, how the industry develops, and how it drives commoditization. The discussion of architectures comes after our definition of cloud computing itself.

We define cloud computing as follows: *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.*

Next we unpack the definition by adding characteristics and examples:

- Users procure the “amount of computing” they want without investing in their own infrastructure. Only an internet connection is required.
- Cloud services provide the illusion of infinite resources on demand available to users, regardless of their size and number.
- Physical infrastructure is decoupled from applications and platforms, which allocate compute, memory, and storage resources without reference to the underlying physical infrastructures. This is known as *virtualization*. Note also that the physical location is decoupled between the physical location of users and cloud datacenters.
- Cloud services transform computing from a capital expense to an operating expense; this changes the role of IT expenditures within the firm.
- Providers can dynamically add, remove or modify hardware resources without reconfiguring the services that depend on them; this differs significantly from traditional datacenter outsourcing.
- Cloud computing changes the location of data processing in the network. Processing moves from the “edge” of the network, in PCs and private datacenters, towards the center of the network, in shared cloud datacenters.
- Only a few firms are able to offer truly global-scale cloud infrastructure (eg., Amazon, Google, Microsoft), with each firm requiring numerous datacenters costing almost a billion USD each, worldwide. For example, according to Google’s financial reports, Google spent over \$7 billion USD in 2013 on capital expenditures, more than doubling the amount from 2012.

Since much of the confusion surrounding cloud computing is the inclusion of characteristics that we do not consider cloud, we need to note how some offerings often labeled “cloud” for marketing purposes are not actually cloud.

First, cloud computing is not simply all datacenter outsourcing, and a large enterprise with a single datacenter is not a cloud service provider. The real power is in the dynamics allocation of resources and the “illusion” of infinite scale.

Second, hardware that combines storage, processing, networking, and databases—known to the IT industry as “appliances”—that are leased to enterprises for their datacenters, usually through service contracts, are not cloud computing services per se.

2.1 Three Architecture Layers

Cloud computing must be delineated into three *architecture* layers: Infrastructure as a service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS). A service that does not delineate between the architecture layers cannot be considered cloud. Here we differ from some definitions, which consider IaaS, PaaS, and SaaS as different types of clouds.

IaaS can be thought of as a “management” model (i.e. how computing, storage and memory resources are allocated to applications based on demand from users over the Internet).

PaaS is a “development” model, which defines how software developers design and build applications that access and make use of the computing resources managed by the underlying IaaS layer. PaaS typically offers software developers common services such as user authentication and database access designed to enable written applications to take advantage of the scalability and resilience of the underlying infrastructure.

SaaS is a “delivery” model that defines how software written by developers using the PaaS layers is made available to users over the Internet. Users typically only require an Internet connection and web browser to access SaaS-based applications and SaaS is often associated with a subscription-based economic model (Fig. 1).

As we will note later, the transformation of computing resources from scarcity to abundance is rapidly commoditizing the lower layer of Infrastructure.

2.2 Cloud Computing as a Dynamic Utility, Critical Social Infrastructure

Beyond a definition and series of characteristics, there are a series of conceptions that capture a broader vantage of the function of cloud computing from both users’ and providers’ vantages.

From users’ vantage, cloud computing is a *dynamic utility*. From the users’ vantage, as with a traditional utility, cloud resources are always available, paid according to the amount consumed, and can be consumed in any quantity. Users are offered contractual levels of availability and reliability. With services provided over Internet connections, providers do not care about the device used to consume the service, and users do not care how providers

Fig. 1 Cloud computing architecture layers



technical configure or operate the service backend as long as quality and price are acceptable. Users are free to use the resources as they see fit.

Cloud providers, like utility providers, are large companies operating at significant scale, serving small users as well as giant corporations. Aggregate demand is amortized over this highly scalable infrastructure and resources are sold back to the user at a much lower—per unit resource—cost than users could provide themselves.

Cloud computing is poised to become part of societies' critical infrastructure, as an increasingly dominant means through which the world's computational demands are met. It will approach the level of economic critical dependency as electricity, gas, water and telephony.

These utility-like characteristics create incentives for national government to search for regulatory frameworks that approach cloud services as critical national infrastructure. This is particularly important for non-US governments, since the major global service providers are American and hence subject to American regulations and rules—as dramatically revealed by information leaked by former US government contactor Edward Snowden in being both cooperative and victims to the government's information gathering activities.

For providers, however, cloud services are not utilities. They are *competitive propositions* that differ from utilities in several important ways. Providers themselves strongly resist being regulated as utilities. To foster innovation, there are strong arguments to refrain from immediately labeling cloud services as utilities.

First, cloud providers do not want to offer commodities—goods that are differentiated primarily on the basis of price. Cloud providers are competing on value-based differentiation on attributes such as service level and functionality. A race to the bottom for commodity offerings is already underway between major players. Any price move by Amazon is immediately followed by Microsoft, for example. If ever-greater scale by the major players drives prices near zero for the cost of computing, it acts not only as an entry barrier for newcomers, but also increases pressure for differentiation through value-added offerings.

Second, unlike many utility providers such as gas or electricity that are granted local monopolies, cloud providers do not enjoy inherent, geographic lock-in of users. Therefore, cloud providers face pressure to create their own service level lock-in mechanisms, including proprietary software components such as a Platform-as-a-Service layer, or specific characteristics tailored for vertical industry or regulatory requirements.

Third, the actual data bits delivered in cloud services are not interchangeable in the manner of electrons or molecules in traditional utilities. The bits combine into quite different users, some of which are mission-critical and some of which are more casual. Users care a great deal about the whereabouts of the bits carrying sensitive personal or mission-critical corporate data, but far less about the location of the constituent bits of a photo or video.

Finally, arguably the biggest difference between cloud services and traditional utilities lies in the degree to which cloud services are uniquely and *dynamically configured* to the needs of each application and class of user. Cloud services are built from a common set of building blocks, but unlike the electricity provider, cloud providers configure them in unique ways for each specific application. For example, the building block configuration for a global public email system differs from an airline reservation system.

2.3 Cloud Computing as Innovation Ecosystem, Production Environment, and Marketplace

One line of skepticism about cloud computing is that the view that there is not new, but rather an agglomeration of existing computing concepts and technologies such as virtualization and

applications residing on remote servers. However, cloud computing *is* uniquely new compared to previous computing technology platforms by simultaneously being an *innovation ecosystem*, a *production platform* and a *global marketplace* (Kushida et al. 2011).

Cloud computing feeds the innovation ecosystem by lowering the bar for new entrants and facilitating experimentation. Most startups no longer require substantial capital outlays to build ICT capabilities. They can scale up or down operations rapidly as needed, and both startups and large firms can experiment with highly computing-intensive tasks.

Cloud computing is also becoming a production platform, with not only raw storage and processing power, but platform-level tools to provide building blocks for creating systems. As we enter an era in which IT services are best considered part of production—with systems built, then delivering services through IT network—cloud services are increasingly providing the resources and tools upon which others build their service systems. Dropbox’s popular file-synchronization and storage services, and Netflix’s video-streaming service, for example, both use Amazon’s cloud infrastructure. Google and Microsoft’s powerful developer tools enable the ability to automatically generate cloud-based services and applications. Cutting-edge enterprise-scale users are now building capabilities from existing open source “building blocks” to enable internal IT services to be easily built and delivered to meet business demands—previously a costly and time-consuming undertaking.

Cloud also provides marketplaces with global reach. This is accentuated by the spread of apps for smartphones, tablets, and browsers, putting powerful building blocks, tools, and entire ecosystem of third-party tools to anyone anywhere with an internet connection.

3 The Drivers of Disruption

The underlying driver of the disruption delivered by cloud computing is the transformation of computing from a scarce to an abundant resource. The key evolution was a progressive *decoupling* of hardware and software. Lead users in the form of innovative large enterprises, following their historical role of driving the IT revolution by adopting and then discovering new uses for IT tools, are likely to be key driving actors in the next round of cloud-enabled innovation. Enterprises’ ability to use IT as a strategic weapon will be critical in the transformation.

3.1 Computing Resources: Transforming Scarce to Abundant

The fundamental driver of the cloud computing revolution is the transformation of computing from a scarce to an abundant resource—a transformation from the economics of scarcity to the economics of abundance (Murray 2013a). Earlier in the history of the computing industry, hardware resources were extremely scarce. Processors’ computational capacities were limited, and limited computer program memory and disk storage space severely constrained the size and complexity of computer applications. Early networking could only transmit data very slowly, at high cost.

Given the high cost and limited capacity of computation, memory, storage, and network bandwidth—the foundation of computing infrastructure—the complexity of software operating systems and applications were limited. *Software and operating systems were optimized for scarce computing resources*. This optimization entailed software written for specific underlying hardware—whether IBM mainframes, DEC mini-computers, or Sun Microsystems workstations. Put the other way around, high performance used to require tight coupling between hardware and software optimized for that particular hardware.

Pioneering companies of the enterprise computing industry were those that successfully optimized for the scarce nature of computing resources. Companies willingly paid premiums for software and hardware solutions that optimized scarce resources to lower operating costs. Cisco, Oracle, and EMC are good examples; each succeeded in being the best of their kind to help customers optimize and manage network, database scale, and disk storage, respectively. Their best-in-class optimization performance enabled them to charge premium prices.

The transformation of computing resources from scarcity to abundance began with foundations laid in the 1980s with the decoupling of hardware and software. The IBM PC broke with the company's traditional model of integrated proprietary hardware and software by outsourcing the computer processor and operating system. Critically, IBM's decision to not bind Intel (processor) and Microsoft (operating system) from exclusively supplying IBM provided the opportunity for an PC industry to develop around the de facto standard of the IBM PC. Compaq and others emerged to create competitive alternatives to IBM.

The decomposition of the PC as a product into components unleashed a wave of innovation and competition at every layer—software, storage, memory, and the like. Prices decreased and the potential use for PCs rose dramatically, leading to exponential growth in consumer demand, driving further economies of scale in the production of PC components. As improvements in hardware fabrication technology (following Moore's Law) led to a doubling of computational capacity every 12–18 months, the computing resources became ever cheaper and more abundant.

When underlying computational hardware resources became abundant, the paradigm for software development transformed. The development of Linux was a paradigmatic example of this transformation. Unix was an operating system invented at Bell Labs and designed to be a general purpose operating system that could run on various types of underlying hardware. However, in practice, when hardware resources were scarce, Unix had to be optimized for each type of computer hardware. This, in essence, led Unix to become a collection of related, but quite different operating systems. Software applications could not reliably run on all variants of Unix, limiting the potential of Unix as a broad-based operating system.

As computing resources became abundant, however, significant amounts of computational capacity could be “wasted” by layers of software that insulates application developers and users from underlying differences in hardware architecture from machine to machine. Linux emerged as an operating system for the era of abundance, capable of running on almost any hardware architecture available.

As the abundance of computing resources increased even more, the one-to-one relationship between hardware and operating systems began to decouple as well. Traditionally, operating systems such as Windows and Linux were designed so that each copy of the operating system ran on one machine. Now, however, enough processing power is available to allow “virtualization,” in which a layer of software mimics the hardware attributes available to an operating system. The result is the possibility of multiple operating systems running on the same hardware—critically, without significant performance compromises. A single computer can run Windows, Linux, and if it is a Mac, the native Mac operating system as well. Conversely, virtualization also enables an operating system to run on numerous physical hardware deployments. An operating system can utilize the pooled resources of multiple computers. This ability—virtualization—is at the base of the cloud computing revolution (Fig. 2).

Silicon Valley-based VMWare was a major success that helped drive the new interest in virtualization. Others quickly followed, with offerings from Microsoft, Oracle, Citrix and Parallels, and the open source based Xen.

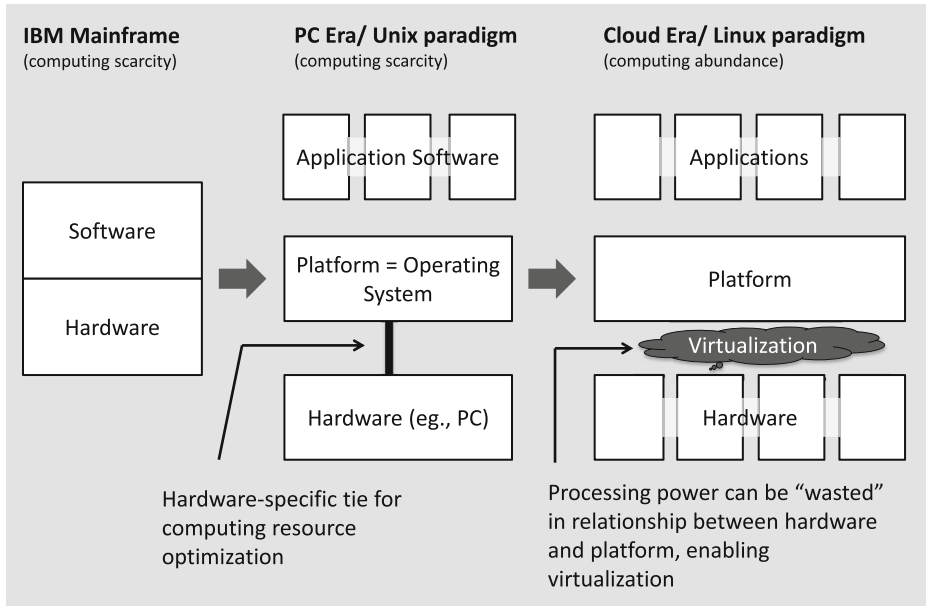


Fig. 2 Progression of software-hardware unbundling

3.2 The Role of Lead Users in IT Innovation

Major industry disruptions are almost never simply about the introduction of new technologies or equipment. Rather, they are about the experimentation by lead users—usually large companies—and the new uses to which they put the new technologies. The initial “IT” revolution that began in the US occurred through this mechanism (Cohen et al. 2000).

For example, the earliest computers were simply replacements for calculations. Only after innovative lead users installed them to solve one set of problems—performing large numbers of calculations quickly—did they discover new uses. Airlines, for example, installed them to handle reservation systems, but then discovered they could manage and adjust routes based on reservation information. With the advent of databases, computers transformed from powerful calculators to “what-if” machines that could calculate probabilities and contingencies (Cohen et al. 2000).

This follows a longer historical trajectory; it took almost 50 years for electrification to yield productivity gain in factor floors. Factories were initially set up according to the logic of steam engines, with machines connected to centrally located steam engines with belts. The first electric motors simply replaced steam engines. Only after the factories were reconfigured, with machines placed according to the logic of production rather than the previous logic of team power, did productivity gains skyrocket (Cohen et al. 2000).

Cloud computing is the next stage in this evolution of production. As enterprise lead users are adopting cloud architectures, they have an opportunity to rethink how IT can be used as a strategic asset.

The initial moves by enterprises to cloud architectures are usually to reduce IT costs. This is, however, only the initial problem that the new technology is implemented to solve. Once implemented, cloud computing architecture can allow firms to solve a variety of other problems, transforming both their strategies and the very nature of how they organize

themselves to compete. We contend that the world is at the cusp of this transformation, with lead users already engaged in new uses of cloud for enterprise-level computing.

3.3 Enterprises IT as a Strategic Weapon

While unforeseen developments in how firms use cloud computing are likely—as has been the case throughout the history of technology and production development—we can point to several specific emerging possibilities, given current efforts by cutting-edge lead users (Murray 2013a).

Throughout the history of companies using IT, IT systems have rarely kept pace with the demands of business at the ground level. Potential business ideas, experimentation, and full strategic implementation of possibilities have been subordinated to limitation of IT systems. Currently, most large corporations subordinate IT to a position lower than that of a core strategic weapon. Legacy IT systems limit the way in which information is used, processed, and stored, and this is often reflected in organizational configurations.

Legacy applications within firms were built to support existing business organizations, supporting functional silos within firms such as manufacturing, finance, sales, supply-chain management, marketing, and human resources. They were not designed to share information horizontally; most large corporate IT maps consist of spaghetti-like mixtures of partially overlapping proprietary legacy systems built over decades.

With the deployment of cloud computing architecture, resources can be deployed dynamically across previously siloed groups. Whenever business needs arise, an IT solution should be deliverable in hours and days rather than months and years. Datasets previously otherwise locked into particular business groups—whether simply due to database incompatibility or from organizational incentives to monopolize the information to gain advantage within the corporation—can be opened up. As corporations are increasingly collections of services that can be purchased on markets, linked with IT systems (Zysman et al. 2013), organizations can become more modular, supported by IT systems to do so.

Put simply, the new *component architecture model* of IT infrastructure, applications, and services, can lead to a new *composable enterprise model* of the firm (Murray 2013b). With a composable enterprise model, business processes and functions can be rapidly and continually re-configured at low cost and operational impact. The imperative to do so is increasing as traditional sectoral boundaries are collapsing with the advent of IT tools (Zysman et al. 2013).

4 The IT Industry Disrupted: Commoditization of Lower Layers

The disruption delivered by cloud computing to the IT industry is essentially a *commoditization* of high-end software and hardware offerings. This commoditization is driven by computing resources transforming from a scarce to an abundance resource.

For providers of IT software and hardware, cloud computing commoditizes a major portion of computing activity offered on a custom basis to deliver high performance, commanding high prices. For users, cloud computing democratizes large-scale computing, enabling even the smallest of users to access global-scale computing capacity. The commoditization delivered by cloud computing is best understood as the latest step in the evolution of computing paradigms (Murray 2013a).

4.1 The Evolution of Computing Paradigms

In the initial era of computing, all layers in the stacks—software, platform, infrastructure—were vertically integrated. An IBM mainframe was integrated with its operating system and software—interlocking components of a system that could not be broken apart.

The advent of the PC era *unbundled* this vertical integration by decoupling software and hardware layers. Each layer became a marketplace with different sets of competitive logic.

The physical PC itself fragmented into its constituent components such as memory, processors, and hard disks, linked by standard interfaces. Competition developed in each of the components, and value moved away from the final assembly, which became increasingly commoditized. Processors, in which Intel dominated, became the areas within physical infrastructure that retained value.

The new logic of competition entailed the emergence of Microsoft Windows as its own platform layer. As a platform, Windows provided a common set of Application Protocol Interfaces (APIs) that freed software providers from writing the code to control specific basic hardware functions, such as accessing memory and storage file systems. The software would work on any machine regardless of who made the components of the underlying PC, as long as it had an Intel processor and could run Windows. Windows quickly became dominant through a positive feedback loop, in which the more users adopted Windows, the more valuable it became as a platform for future users and software developers.

In the vertically unbundled PC era, Microsoft and Intel captured a disproportionate amount of value from the rapidly commoditizing PCs. Microsoft licensed Windows and built an ecosystem of software application that required Windows to function. And within the PC, although almost all other components could be assembled from any number of manufacturers, the Intel processor architecture was necessary to run Windows, enabling Intel to avoid commoditization as well—an era we have called “Wintelism” (Borras and Zysman 1997).

A major consequence of unbundling the PC architecture was the emergence of software as a discrete industry independent from hardware. Microsoft eventually prevailed over independent competitors (such as Lotus 1-2-3 and WordPerfect) through integrating its offerings into suites, and by leveraging its operating system.

4.2 Paths to Escape Commoditization: Customization, High-End, and Embedding Business Logic

As enterprises rapidly embraced commoditized Wintel PCs, large-scale IT providers moved towards high-end, customized, and performance optimized IT “solutions.” IBM, for example, sold its hardware by bundling it with the services it delivered—a shift from its historical selling of hardware servers with built-in functionality. The business model of IT integrators was to find value in vertically integrating the different layers of the stacks (infrastructure, platforms and applications) in custom packages for users firms, charging millions of dollars upfront for an implementation. The complexity and overwhelming plethora of choices facing business users led to the development of the *prima facie* appeal of IT integrators who possessed the specialized insight and expertise necessary to source and implement key software projects. The integrators were therefore responsible for building and operating corporate datacenters, with the corporate applications to provide customized “solutions” tailored to clients’ needs. This approach was highly lucrative for the first generation of Enterprise Resource Management (RP), supply chain and Customer Relationship Management (CRM) applications.

Another path to avoid commoditization was to provide high-end equipment to enterprise users, or to offer solutions that had highly customized built-in functionality that could not be

separated into constituent elements. Servers from Sun Microsystem and Cisco Systems’ routers exemplify the high-end offerings. Oracle’s enterprise solutions, in which Oracle Financials could not be separated out from Oracle Database, for example, represent the latter.

The trend towards emphasis on high-end engineered systems and systems that combined hardware and software were particularly attractive in the face of commoditization. Concerns over network security, for example, let to Cisco’s emphasis on an integrated approach offering a set operating system and high performance optimization and throughput.

Oracle’s Exadata systems combined server, storage, and networking hardware integrated and optimized with database, middleware, and analytics tools in a single chassis. Oracle’s focus on was driving growth from high-end “engineered systems,” as it called them, to offset declining revenue streams from its hardware business (commodity storage and Intel-based servers).

4.3 The Advent of Cloud Computing: Disintegrating the Vertical Stack

With the advent of cloud computing, a more radical type of vertical disintegration and commoditization is occurring. The physical infrastructure itself is becoming unbundled from the platform layer to an entirely new degree. Offerings in the lower layers of the stack, such as storage, network and even databases, are becoming commoditized more than ever before. “Value-added” management controls embedded in storage, network or databases that optimize their resources and allow providers to charge premiums are falling by the wayside as value moves upwards in the stacks (Fig. 3).

To illustrate, let us examine a single-user example. In the PC era, a user would not care who made their hardware as long as the operating system worked to run the software needed. However, the copy of the operating system was tied to a specific computer. With the advent of cloud computing, for applications built on top of cloud computing infrastructure, the user does not care *where* the hardware is, or *what it is running*. Users do not really care how the backend

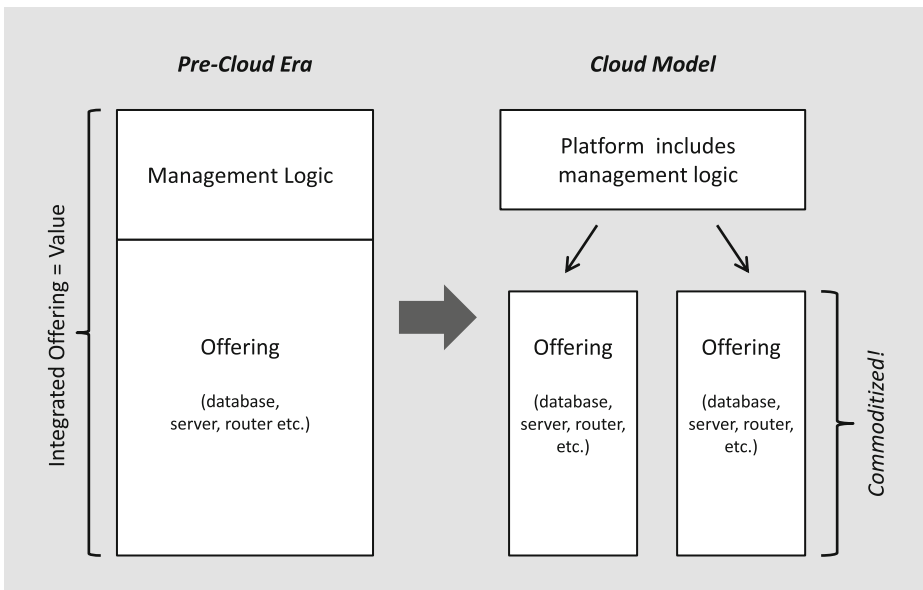


Fig. 3 Pre-cloud era value to cloud model commoditization

servers of Gmail, Google Docs, Microsoft Office 365, Dropbox, or Netflix services are running. To start a service, one can simply rent capacity from Amazon. This logic transferred to the business context is how value in the stack is moving up towards the applications and platform, commoditizing the lower infrastructure layers.

4.4 The Infrastructure Layer Commoditized

Infrastructure is rapidly being commoditized. Take servers, for example. High-end, high-performance servers such as those offered by Sun Microsystems used to give users requiring massive computing power a competitive edge. However, with the advent of cloud computing, the advantage of individual high-performance machines with high price premiums is far less attractive.

Google spearheaded the new paradigm of taking computing resources as abundant as a starting point. It used cheap, off-the-shelf computers rather than high-end servers to build its datacenters. Its approach was that higher hardware failure rates could be built into the parameters of its software design, with the ability for databases and other tasks to be distributed among large numbers of physical hardware, unaffected by failures of particular hardware. Performance came from the algorithms for distributing the tasks rather than the performance of the hardware itself (Levy 2011; Barroso and Hölzle 2009). This approach enabled Google to offer Gmail initially with 1 gigabyte (GB) of storage in 2004 (doubling it to 2GB the following year) at a time when Microsoft's Hotmail offered 2 megabytes (MB) and Yahoo!Mail was 4 MB—500 times and 250 times, respectively, the amount of their established competitors' capacity (though initially to a limited number of users). Put simply, this was the first major step in physical infrastructure servers becoming unbundled and commoditized.

Networking hardware is also rapidly becoming commoditized, threatening firms such as Cisco. As investments to build out the Internet took off in the mid-1990s, Cisco Systems dominated global markets in providing network backbone equipment and was a strong presence in corporate networking solutions. The physical need for routers to connect datacenters, servers and machines to each other and to the Internet gave Cisco and its competitor Juniper Networks opportunities in the networking portion of the Infrastructure layers.

Recently, however, high-end networking is increasingly achievable through software. VMWare's acquisition of Nicira's Software Defined Networking technology raised concerns that networking routers and switches could face disruption analogous to the impact of virtualization on the server market. The parallels are not exact; under-utilization and overcapacity of non-virtualized servers were a prime target for cost reduction. Nonetheless, cheaper commodity hardware can replace both proprietary server and network hardware. For example, major users of IT equipment revamping their entire IT infrastructure are increasingly finding that companies such as Tier 3, a Seattle-based company, was able to match Cisco's high-end, hardware-based networking solution through software—critically, at a far lower price.

4.5 Value Moving to the Platform

Management functions within software are increasingly migrating to the platform rather than being encapsulated in individual components of datacenters, whether it be storage, networking or databases. This drives commoditization of the components.

Take databases, for example. At the simplest level they are comprised of the database itself, a business logic that manages the database, and the user experience. These three used to be integrated—as illustrated earlier with the example of Oracle Financials being unable to be

decoupled from Oracle Database. Oracle was, in essence, embedding management logic into its database offerings.

With the advent of cloud-based platforms, however, it is rapidly becoming possible to procure just a pure database offering without any management logic embedded in the database. The management logic instead resides in the platform level, managed through an open API. From the user's vantage, you can manage multiple resources through the platform, with multiple components open at the API level to control from the platform. Therefore, as a user, you want standalone components that can be mixed and matched, with the possibility of migration from one to another—your financial system will not be coupled with a particular storage engine, and either can be moved to a new provider.

From the perspective of providers, this dramatically increases the pressures for commoditization, since there are far fewer ways to lock in customers to their offering. As a current business reality, businesses in all layers have some business logic in each layer, including the lower layers, such as “stored procedures” in databases. But these stored procedures are what lock users into particular databases.

The initial rationale for such stored procedures was to optimize performance and security. However, in the new paradigm of cloud computing enabled computing abundance, it is simple to allocate more computing resources rather than pursued customized performance databases.

4.6 Lead Users Driving Commoditization

Cutting-edge lead users adopting cloud computing solutions on a major scale are driving commoditization of infrastructure layers. They are beginning to move towards not allowing any stored procedures in the database, instead forcing the database to be controlled only by the platform layer through an open API. The implications are that those providing business functionality in the package and charging premium prices will face a different kind of competition.

This disruption is so pervasive that even some of the frontrunners of recent disruptions are affected. VMware, for example, a company that provided virtualization, purchased software-based network virtualization company Nicira in 2012. VMWare, however, is owned by EMC, which operates many of the virtualized datacenter component offerings such as storage—the areas in which VMWare is driving rapid commoditization. VMWare then went on to spin out many of its services into a joint venture with GM called Pivotal, which offers PaaS for enterprises.

Facing the disruption, leading technology vendors are forced to compete against themselves to mitigate the impact of the transition to cloud. Oracle, for example, has been aggressively acquiring SaaS vendors to stave off customer defections to cloud-based competition. Microsoft launched Office 365 and Windows Azure to provide cloud-based alternatives to productivity suites, database, middleware, and server offerings, while IBM is similarly investing in its own hybrid cloud offerings.

5 Conclusion

In this paper we have argued that the advent of cloud computing was driven by a fundamental shift in computing power, as it transformed from a scarce to an abundant resource. Cloud computing architectures implemented into large enterprises, the lead users, is where the full impact of cloud computing will be felt on the global economy.

This paper has focused primarily on industry issues, but the advent of cloud computing raises numerous critical policy issues that will in turn shape how the cloud technologies, services, business models, and adoption by lead users will progress. The next research agenda should focus on political and regulatory ramifications of *cloud computing becoming the fundamental infrastructure of the global economy*.

Especially as more and more devices are connected to the Internet—commonly characterized as “the Internet of Things”—the architectures of cloud computing will become the underlying fabric of what we call “ICT-enabled services systems.” Political and regulatory debates that had been settled at one point are poised to be reopened. Issues such as antitrust, privacy (who is allowed access to what data), security (protection against unauthorized access and manipulation), jurisdiction, liability, and industrial promotion policies have all developed in a variety of national and regional contexts. We conclude by highlighting several of these issues and drawing implications.

With cloud computing simultaneously being an innovation ecosystem, production platform, and global marketplace, regulations in previously disparate areas all converge on cloud computing. For example, how should industrial promotion policies be conceived, when much of the global, commoditized computing infrastructure is delivered by a small handful of US-based multinational firms? On the one hand, investing massive sums to build national-scale datacenters or supercomputers runs the risk of incurring massive upfront and running costs into something that is commoditized and underpowered even before completion. On the other hand, however, simply deciding to rely on foreign-provided computing infrastructure upon which to build national competencies may be deeply unsettling for the political and bureaucratic leaderships. How should antitrust be conceived and executed, when computing power and platforms are indeed being concentrated in a handful of large firms, but the various “bottleneck” or competitive leverage points shift rapidly (see Kenney and Pon in this issue, for example). How should product/service liability be conceived if services deployed in one particular national context are built on top of building blocks that physically reside elsewhere? The political and regulatory battles in various arenas, across countries, have only just begun, and they can be expected to unfold for the foreseeable near future.

Security and privacy issues—where security is the protection against unauthorized access and manipulation and privacy concerns the rules governing who can view and use what information—are also clearly critical policy areas. Revelations of the US government’s extensive digital espionage activities by subcontractor Edward Snowden have put these issues on the forefront of national concerns, with policies considered in some countries that would explicitly exclude US companies from offering services to government or particular areas such as banking. Since the scale merits of cloud computing implies a centralization of services into a handful of global-scale providers, the question is whether policy considerations will enable nationally or regionally based cloud architecture services that follow a political logic—such as being beyond the reach of the US Patriot Act that enables the US government to access any information passing through the US, or held by US-based firms.

The policy debates raised by cloud computing will unfold at national, regional, and international organizational levels. Issues have yet to be settled, as a variety of policymaking processes interact in multiple regulatory arenas. While there is much uncertainty over the outcomes, this paper provides an understanding of the underlying technological and industry-level foundations that inform the debates.

Many of the current regulations resulted from political settlements among interested parties, mediated by political and/or bureaucratic coordination. With cloud computing as an innovation ecosystem, production platform, and global marketplace causing many of these hitherto distinct policy domains to converge on a single set of actors, technologies, and markets, these

political battles are likely to be reopened, but with a different set of dynamics. Just as fundamental shifts in production paradigms of the global economy—from agrarian to industrial, from steam-powered to electric-powered industry, and from electro-mechanical to digital—unleashed new industry dynamics and transformed political debates at the heart of capitalist societies (Breznitz and Zysman 2013; Gourevitch 1986; Zysman and Huberty 2013; Zysman and Newman 2006), the advent of cloud computing as the new infrastructure underlying the global economy will reopen and transform key issues that will shape the global economy for years to come.

Acknowledgments We wish to thank Ed McGuire of CLSA. Portions of an earlier version of this paper appeared as ‘Clouducopia: Into an Era of Abundance’ CLSA Blue Book, 2013. We also wish to thank two anonymous reviewers for their insightful feedback and comments. This research is a part of the ongoing collaboration of BRIE, the Berkeley Roundtable on the International Economy at the University of California at Berkeley, and ETLA, the Research Institute of the Finnish Economy.

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

References

- Barroso LA, Hölzle U (2009) The datacenter as a computer an introduction to the design of warehouse-scale machines. Morgan & Claypool, San Rafael
- Borras M, Zysman J (1997) Globalization with borders: the rise of wintelism as the future of industrial competition. *Ind Innov* 4(2):141–166
- Breznitz D, Zysman J (eds) (2013) *The third globalization: can wealthy nations stay rich in the twenty-first century?* Oxford University Press, Oxford
- Cohen S, DeLong JB, Zysman J (2000) Tools for thought: what is new and important about the “e-economy”. Berkeley roundtable on the international economy. University of California, Berkeley
- Gourevitch P (1986) *Politics in hard times: comparative responses to international economic crises*. Cornell University Press, Ithaca
- Kushida KE, Murray J, Zysman J (2011) Diffusing the fog: cloud computing and public policy. *J Ind Compet Trade* 11(3):209–237
- Labovitz C (2012) How big is Amazon’s cloud? <http://www.deepfield.com/2012/04/how-big-is-amazons-cloud/>. Accessed 25 May 2014
- Levy S (2011) *In the plex: how Google thinks, works, and shapes our lives*. Simon & Schuster, New York
- Mell P, Grance T (2011) The NIST definition of cloud computing. Recommendations of the National Institute of Standards and Technology. <http://csrc.nist.gov/publications/PubsSPs.html#800-145>. Accessed 25 May 2014
- Murray J (2013a) Economics and platform architecture. Adamalthus, blog post. <http://www.adamalthus.com/blog/2013/02/27/economics-and-platform-architecture-i/>. Accessed 15 January 2013
- Murray J (2013b) The composable enterprise. Adamalthus, blog post. <http://www.adamalthus.com/blog/2013/04/04/the-composable-enterprise/>. Accessed 25 May 2014
- Zysman J, Huberty M (2013) Can green sustain growth? From the religion to the reality of sustainable prosperity. Innovation and technology in the world economy. Stanford University Press, Stanford
- Zysman J, Newman A (eds) (2006) *How revolutionary was the digital revolution? National responses, market transitions, and global technology*. Stanford University Press, Stanford
- Zysman J, Feldman S, Murray J, Nielsen NC, Kushida KE (2013) Services with everything: the ICT-enabled digital transformation of services. In: Breznitz D, Zysman J (eds) *The third globalization: can wealthy nations stay rich in the twenty-first century?* Oxford University Press, New York, pp 99–129