THE BERKELEY ROUNDTABLE ON THE INTERNATIONAL ECONOMY



UNIVERSITY OF CALIFORNIA, BERKELEY

BRIE Working Paper 2014-4

The Implications of Cloud Computing for Integrated Research and Innovation Strategy

9/15/2014

Kenji E. Kushida Stanford University

Jonathan Murray

Innovia Ventures

Patrick Scaglia University of California Berkeley

John Zysman

University of California Berkeley, Berkeley Roundtable on the International Economy

Prepared in contract with the European Commission with John Zysman, URL: <u>https://ec.europa.eu/digital-agenda/en/news/public-consultation-cloud-computing-and-software</u>

ABSTRACT

The advent of Cloud computing as the new underlying global infrastructure of computing presents distinctive new opportunities and challenges for Europe. Cloud computing is transforming computing resources from a scarce to an abundant resource, driving a wave of commoditization in previously high-end software and hardware. For Europe to gain independence from US-based global scale Cloud providers, our view is that it needs to move towards a distributed model of computing with federated governance. Distributed Cloud means the distribution of computation close to the geographic location of the data and the users, as opposed to the centralized model of today. Our research and innovation strategy recommendations reflect this view.

OUTLINE:

Introduction

Part I. What is Distinctive about Cloud?

- 1. Varied meanings of cloud
- 2. The importance of cloud
- 3. Evaluating Europe's position

Part II. Research Strategies in a Cloud Era

- 1. The fate of "cloud" suppliers and implications for research strategy
- 2. Priorities for research in Cloud Computing

Part III. Innovation Strategies in a Cloud Era

- 1. Implementation of commoditized parts
- 2. Testbed projects
- 3. Developing regional capacities for rapid innovation
- 4. Leveraging public provision of services towards innovation

Appendix 1: Distributed Cloud

Appendix 2: Historical Strategies of Global Cloud Providers

EXECUTIVE SUMMARY

The advent of Cloud computing as the new underlying global infrastructure of computing presents distinctive new opportunities and challenges for Europe.

In our view, true 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. It consists of abstracted architectural layers of Software, Platform, and Infrastructure as a single computing system. Cloud computing is transforming computing resources from a scarce to an abundant resource, driving a wave of commoditization in previously high-end software and hardware. Only a handful of multinational firms, rooted in the US, currently deliver global-scale Cloud services.

For Europe to gain independence from US-based global scale Cloud providers, our view is that it needs to move towards a distributed model of computing with federated governance.

A *distributed* model of computing refers to distributed workloads that are transparent and automated. Distributed Cloud means the distribution of computation close to the geographic location of the data and the users, as opposed to the centralized model of today. It goes beyond interconnection in the limited sense of interconnecting legacy architectures. The distribution of workloads within each of the architectures must be visible, adjustable, and automated.

Federation is the governance structure of managing a distributed model of computing. The governance structure must manage the distribution, creation, and closing of workloads, which must be in real time. This differs from federating conducted by telecom carriers for traditional services such as mobile, with a broker dealing with roaming and billing accordingly. Here, the workload must be moved around, coordinating with the location of data. Authentication and authorization become critical issues.

A distributed model of Cloud is driven by fundamental network constraints, combined with the imminent deluge of data that ubiquitous sensors (the so-called "Internet of Things") will bring, far surpassing what we have today.

We recommend research and innovation strategies that explore the challenges and develop possibilities of distributed Cloud with federated governance.

INTRODUCTION

The advent of Cloud computing as the new underlying global infrastructure of computing presents distinctive new opportunities and challenges for Europe.

The formal definitions of Cloud computing vary, but we have a particular vantage, described in detail later, which emphasizes that Cloud computing is in essence a new architecture of how to organize computing. This architecture is the abstraction of traditional layers of infrastructure, platform, and software, with a Cloud consisting of the set of architecture layers as a single computing system. The advent of Cloud is driving a new phenomenon—an abundance, rather than a scarcity, of computing resources.

One immediate consequence of the advent of Cloud computing, already observable, is that the distinctive proprietary advantages of established purveyors of the components of computing are eroded. Components that used to be high value added propositions become commodities. Whenever existing players are dislocated, market space for newcomers is generated. This creates distinctive opportunities for Europe.

A critical area towards which Cloud computing is headed is *distributed* Cloud with *federated* governance, defined more in detail later. Currently, global-scale Cloud computing is provided by only a handful of multinational firms rooted in the US, with centralized processing in each firms' network of datacenters. Others are unlikely to be able to pursue a similar scale as a single firm. However, new approaches and technology developments are moving towards the realization of distributed Clouds with federated governance, which is a deep integration of smaller-scale Clouds. The integration goes beyond simple interconnection, and extends into distributing workloads and data transparently between multiple Clouds in a scalable manner.

The potential of distributed Clouds with federated governance opens major opportunities for Europe. Another driver of potential European advantage is the combination of policy desire for local control of Cloud services, which are poised to become the underlying information technology fabric of society, and the new economic logic of distributed Cloud computing. It is the convergence of security and privacy policy on the one hand, and the economics that enables distributed Cloud that gives Europe a real set of opportunities.

This document provides a set of recommended directions towards research and innovation. We contend that the focus must shift beyond traditionally conceived research and innovation strategies.

Key questions that must be addressed include:

1. What will be the European response to US cloud service providers whose vast – and rapidly growing – scale confers significant economic and market power advantages? Can the aggregation of independent local or regional cloud services provide a competitive response?

2. As service components become rapidly commodifized, where are the areas in which the ability to design, implement and integrate cloud service components can be captured by research and innovation strategies?

PART I. WHAT IS DISTINCTIVE ABOUT CLOUD?

First, a clear definition of Cloud is needed because there still seems to be confusion among various IT sector participants, and therefore government reports, about what Cloud is. Much of this confusion is intentional—an effort by legacy IT solutions providers to rebrand their existing services as "Cloud." This has led to confusion among those who are not providers of global-scale Cloud services. Some of this is a response to large users who feel they should be implementing Cloud, but are not prepared to do so.¹ This confusion can lead to some misleading recommendations on what the research and innovation issues are regarding Cloud.

However, from the vantage of users and providers of global-scale, cutting edge Cloud services, there is very little ambiguity about what Cloud is. In this section we provide a definition of Cloud and show how it differs from other conceptions. We also provide a framework for analyzing the Cloud industry ecosystem in Europe, which is useful for evaluating the position of European players.

1. THE VARIED MEANINGS OF CLOUD.

Our definition of Cloud is derived from the vantage of global-scale Cloud providers, and leading edge users who are deploying global-scale Cloud services in large companies. It is a definition that makes sense to practitioners:

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. 2

We specify that Cloud computing rests on abstracted architecture layers. They are abstracted in the sense that they are not physically demarcated, but are logical, software-based separations. They include: Infrastructure as a service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS). Taken together, they constitute Cloud. The individual layers are not Cloud.³

¹ These positions, and the intentional confusion, was recently well captured and articulated by a Forbes article. <u>http://www.forbes.com/sites/jasonbloomberg/2014/07/15/why-implement-cloud-when-cloudwashing-will-suffice/</u>

² This definition is expanded upon in: Kenji E. Kushida, Jonathan Murray, and John Zysman, "The Gathering Storm: Analyzing the Cloud Computing Ecosystem and Implications for Public Policy," *Communications and Strategies* 85 (2012). And "Cloud Computing: From Scarcity to Abundance," *BRIE Working Paper* 2014, no. 2 (2014), http://brie.berkeley.edu/publications/Cloud2014.pdf.

³ This is a critical difference from definition by the EC included in "Advances in Clouds" which considers SaaS, PaaS, and IaaS as *types* of Clouds. For example, SaaS is not a type of Cloud—it is one of the abstracted layers within Cloud. There are many SaaS services that are not Cloud. http://cordis.europa.eu//fp7/ict/ssai/docs/future-cc-2may-finalreport-experts.pdf

- 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.



Figure 1: Cloud Computing Architecture Layers

Cloud computing can be deployed in several *deployment patterns*.

- A *private* Cloud deployment is entirely behind a corporate firewall.
- A *public* Cloud deployment faces the general public in terms of offering services to anybody with an Internet connection
- A *hybrid* Cloud combines private and public deployments, with some portion behind a corporate firewall.

The general public is most familiar with public-facing deployments, but for enterprises, outsourced datacenters in of themselves are not automatically Cloud. *Only when Cloud architectures are deployed within enterprises* (behind corporate firewalls) can we call Cloud

deployments "private" – a term too often used for marketing conventional datacenter outsourcing.

Large enterprises whose traditional datacenters were behind corporate firewalls are moving to private and hybrid deployments. It is this adoption by large enterprises that will drive the next wave of disruption, both for users and for the IT providers.

How is our definition of Cloud as a set of architecture layers different from other definitions? The main point is that for something to be considered Cloud, it must have the architecture layers. Otherwise it is a traditional datacenter, whether in-house or outsourced.

Commonly used definitions, such as that of the National Institute of Standards and Technology (NIST) in the United States, define cloud computing as a set of essential characteristics, with three services models of SaaS, PaaS, and IaaS, and then three deployment models of Private, Community, Public, and Hybrid.⁴

In our definition, whether any set of deployment models can be considered cloud or not depends on whether they have the architecture layers. Architecture layers are a necessary condition for Cloud. For example, a so-called "private cloud" service that offers Cloud deployment to companies, which comes as an integrated solution that cannot be separated into the three layers, is not a "private Cloud." It is a datacenter outsourcing service. This has serious implications for considering Europe's opportunities, since vendors of traditional datacenter outsourcing services will want to brand their services and solutions as Cloud—but this should not lead to confusion about what it actually is.

2. THE IMPORTANCE OF CLOUD.

In this section we raise 4 main points:

- Cloud computing is commodifizing traditionally high-value added offerings by transforming computing from a scarce to an abundant resource.
- Value is moving up the architecture stack, away from IaaS, with global-scale provider providing all layers of service.
- The role of lead users in driving innovation by adopting and implementing Cloud services.
- Cloud as uniquely new, as an innovation ecosystem, a production platform, and a global marketplace.

i. Commoditization Driver: from Scarcity to Abundance of Computing Resources

Cloud computing services are driving an acceleration of *commoditization*, particularly of previously high-value added offerings. Commoditization, applicable to both products and services, is a situation in which differentiation is primarily on the basis of price, with little potential for firms to capture economic rent (profitability) beyond costs.

The advent of Cloud computing is a driver of commoditization in several ways, but most importantly, it is driving a fundamentally new paradigm in the nature of computing resources. Until now, most value-added computer hardware and software were created with the assumption that computing power (and storage) were scarce resources. Optimization of the scarce resources was therefore what differentiated high-end offerings—the ability to achieve higher processing power, greater storage, or to most efficiently utilize limited resources.

⁴ http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf

In the software layer, Oracle databases, for example, optimized databases by building management functions into the database itself. For platforms, an important attribute of operating system performance was how efficiently it utilized the hardware it was running upon. And for hardware, high performance at a particular time almost always entailed a premium price.

However, now, with cloud computing, computing resources are abundant, with everdecreasing prices and ever more powerful tools.⁵ Processing power can be "wasted" and still provide high performance. Rather than paying premium prices for high-end software and hardware offerings, large numbers of lower-end offerings can be orchestrated to match or exceed premium high-end offerings.

Intel provides high-end server processors, but firms are finding that in many cases, a large number of far lower-cost ARM processors can be deployed to attain greater cost-performance. Cisco provides high end networking equipment, but Software Defined Networking (explored more fully later) is a paradigm that threatens its hardware-embedded solutions. Therefore, business positions optimizing for computing resources as scarce are likely to face this massive wave of disruption.

ii. Value Moving Up the Architecture Stack

For Cloud service providers, an emerging pattern is that *value is moving up the stack*, away from IaaS, to PaaS and SaaS.

The historical strategies of American global providers Amazon, Google, and Microsoft reveal how they started at different architecture layers, and moved towards platform offerings. Google and Salesforce started with SaaS, Amazon with IaaS. Competition moved towards who could provide the best platform. The global-scale providers then moved towards offering services at all layers of architecture, although with different degrees in the ability of users to procure each service independently of the others.⁶

In IaaS in particular, American multinational corporations have built a formidable lead in scale, with billion dollar data centers around the world. Any service competing with them that is not differentiated meaningfully will be annihilated in the cost pressures of the markets. Amazon Web Services (AWS), for example, the largest provider of virtual servers, decreased its prices by approximately 75% from May 2008 to April 2014.⁷ Amazon, Google, and Microsoft, which each have dozens of billion dollar datacenters around the world. Smaller players, such as Rackspace, and other services offering IaaS, cannot compete with global players' scale and cost.

As large corporate enterprises adopt Cloud, and the value shifts away from IaaS, they are less likely to be interested in becoming locked into a service that includes a specific IaaS service. Their preference for not getting locked into a commodity service extends into areas such as databases as well, which may also be considered IaaS for this this purpose. We are beginning to see large lead users preferring have all the management logic in the platform, with the IaaS providing an open API (Application Protocol Interface) that can be controlled from the platform layer. For such lead users, Oracle databases, which have management logic embedded in their

⁵ By April 2014, for example, Amazon had decreased its effective prices for processing power by an estimated 75% since 2008. <u>http://gigaom.com/2014/04/19/moores-law-gives-way-to-bezoss-law/</u>

⁶ For details on the historical evolution of their strategies, see Appendix 2.

⁷ http://gigaom.com/2014/04/19/moores-law-gives-way-to-bezoss-law/

offerings, becomes less attractive than a commodity database with an open API that can deliver sufficient performance if enough processing power is applied.

iii. Dynamic Utility: Lead user-driven innovation

There are several other characteristics of Cloud services that differentiate it from previous paradigms of computing, which merit a brief discussion. First, 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 through Service Level Agreements. 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 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.

Lowering costs is the just the baseline – everybody will have the same tools available. Note that the process of innovation in IT has been as much about *how lead users solve new problems with available IT tools* as the performance attributes of IT tools.⁸

iv. Innovation Ecosystem, Production Platform, and Global Marketplace

Although Cloud services as we know it today are built on top of many existing technologies, such as virtualization and multi-tenancy, global-scale Cloud computing has taken these well beyond their original forms. Cloud computing is unique compared to previous technology platforms by simultaneously being an *innovation ecosystem*, a *production platform* and a *global marketplace*.⁹

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.

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.

⁸ Stephen Cohen, J. Bradford DeLong, and John Zysman, *Tools for Thought: What Is New and Important About the "E-Conomy"* (Berkeley, CA: Berkeley Roundtable on the International Economy, University of California at Berkeley, 2000).

⁹ Kushida, Murray, and Zysman, "The Gathering Storm: Analyzing the Cloud Computing Ecosystem and Implications for Public Policy."

v. Global Scale Cloud Providers Rooted in the US

Only a small number of firms can offer Cloud services at a global scale, and the performance and price gap between the global-scale players and the rest is widening. The global scale providers are Amazon, Google, and Microsoft, which each have dozens of billion dollar datacenters around the world. For example, Google added approximately USD 13 billion worth of datacenter capacity in 2013. The major drivers of commoditization are these global players, which have US origins and are in many ways, deeply rooted in the US. This provides a challenge for the rest of the world, which experience the commoditization drive as exogenous shocks.

On the other hand, as described more in detail below, there is ample reason to believe that distributed Cloud with federated governance is a model that can provide significant opportunity for Europe. There are technical reasons, as outlined below and in the appendix, as well as regulatory factors, including serious concerns about US privacy laws and the Patriot Act, which provides the US government with the authority to access data held by any US-based multinational corporation, or any corporation with business in the US.

3. EVALUATING EUROPE'S POSITION

Europe is in the position of being a follower in Cloud services development and utilization. While this presents serious challenges, it also provides multiple paths for opportunity. Europe should never lose sight of the fact that value will be created in the applications and reorganizations enabled by Cloud, rather than in Cloud infrastructure itself. There are potential risks of being a follower in provision or aspects of the components of the computing systems. But these two issues need to be separated. Furthermore, widespread successful application can develop the foundation for the creation of leadership in provision and component technology.

Here we provide a typology for analyzing players in the Cloud ecosystem, which can be applied to accurately depict the position of various European players.

i. Provider Types: Framework for Analyzing Cloud Ecosystem Frame

Cloud services (Cloud *Providers*), those that purchase and consume the services (Cloud *Users*), and those who connect providers with the consumers (*Connectors*).

There are three types of providers and connectors. (The top horizontal axis of Figure 2.)

- 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 Figure 2.)

- 1. The *Infrastructure* layer encompasses the hardware, networks and operating systems responsible for managing fundamental resources such as data storage, computation and network bandwidth.
- 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 Programing 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.



Figure 2: The Cloud Services Stack

These two typologies constitute the two axes of our framework illustrated in Figure 2.

Cloud Providers, who provide services from their *Cloud Datacenter* fall into three broad types according to their technical architecture type. Figure 3 shows how Cloud providers with the three architecture types of SaaS, PaaS, and IaaS.



Figure 3: Cloud Providers in the "Cloud Services

Today's Access Devices derive their lineage from computing equipment and like their 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.

Figure 4: Device Providers in the "Cloud Services Framework"



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 Figure 4)

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 our 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:

- 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 have emerged largely successful so far.¹⁰
- 2. *Winning The User* The central game of the emerging Cloud Computing marketplace positions major *Cloud Providers* the major players and smaller players vying to win the hearts of minds of both consumers and firms, with large firms offering scalability and low cost, with smaller providers struggling to differentiate.
- 3. *The Search for Network Value* The emergence of Cloud Computing places even greater urgency on the *Network Providers*' search for value and relevance.

ii. PLUMS Framework for Comparison

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. In this paper we provide a brief overview of the framework with a preliminary sample analysis, but leave an in-depth comparison for subsequent research.

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	s –	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:

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.

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" of public, private, and hybrid.

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. 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 useful framework 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 it from other easily commoditized offerings. We can also determine whether it is targeted at the general public and small businesses or enterprisescale companies to determine how the national context provides opportunities and limitations.

Table 1 shows the PLUMs framework applied to the US-based global-scale firms in various areas of the Cloud ecosystem. While it is beyond the scope of this report to make a detailed analysis of the various European providers in various areas of the Cloud ecosystem, it should be clear through even a preliminary sketch of Europe that even the larger telecom carriers do not offer all layers of the stack, and their deployments are more local than global.

	Provider	Layer	Users	Modality	Scope
Microsoft	C, N, D	I, P, A	C, F	Pu,Pr,H	G
Google	C, N, D	I, P, A	C, F	Pu	G
Amazon	C, N	I,P	C, F	Pu	G
IBM	С	I,P,A	F	Pr	L,C
HP	C, D	Ι	C, F	Pr	L, C
Salesforce.com	С	P, A	F	Pu	G
Rackspace	С	I, P	F	Pu	G
Twitter	С	P, A	С	Pu	G
Facebook	С	P, A	С	Pu	G

Table 1: The PLUMS Framework Applied to Major Global IT Firms

Table 2: The PLUMS Framework Applied to Europe's IT Firms-a Template

	Stack	Layer	Audience	Modality	Scope
Major IT Vendors (SAP)					
Telecom Carriers (eg., Deustche Telecom, FT)					
Electronics firms (eg., Siemens, Alcatel)					
Online Services					
Social Networking Service					

iii. Why Europe is a Follower

Why is Europe in the position of being a follower? A major reason is the "critical mass" of the ecosystem that led to the development of IT firms that took advantage of the Internet as it developed into an open global platform. The transition from mainframes to the paradigm of interconnected PCs were driven initially by US newcomers, funded by venture capital and a

business ecosystem surrounding Silicon Valley.¹¹ Sun Microsystems, Apple, Cisco, 3Com, Oracle, Intel, and others are examples of venture capital-funded firms that shaped the interconnected PC era.

In the Internet era, US companies moved first to begin large-scale Cloud operations about a decade ago. Examples include AWS, Google, Yahoo, eBay, MSFT, Apple, Facebook, Twitter, Netflix, Dropbox, Box, SalesForce.com, Rackspace, Palm, Snapfish, Photobucket, and other. Critically, as these firms built out their IT infrastructures, they created many of the supporting technologies along the way, as the scale and complexity of their services grew. As a result, a significant number of engineers and developers created new skills in this area for a decade. These people moved across various companies and now represent highly sophisticated and trained intellectual capital. The current ecosystem therefore enjoys a critical mass of knowledge and skill. This takes time to build, and there is no "quick fix" or substitute.

A similar phenomenon occurred with smartphone operating systems. iOS, Android, Palm (webOS), and Amazon's Android derivative, were all created within a 50 mile radius in the Bay Area. The people involved have moved between those companies. It is, like Cloud, a highly specialized technology space, where new ideas emerged in a short period of time. This has given the region a first-mover advantage of expertise, creating a formidable handicap for others to compete.

PART II: RESEARCH STRATEGIES IN A CLOUD ERA

The first and most important point to note is that in a world of fast-moving Cloud computing-enabled computing resource abundance, there can no longer be a clean separation between research strategies and innovation strategies. Instead, there is a particular relationship between research and innovation strategies that relies on a continuous feedback loop. Therefore, the potential for flexibility and rapid adjustment of research strategies is critical.

Cloud development occurs in a "DevOps" operational model.¹² DevOps is a software development approach which stresses communication, collaboration, and integration between software developers ("Dev") and IT operators ("Ops"), aimed at rapidly producing software products and services. A DevOps approach standardizes development environments and enables faster product delivery, quality testing, features development, and maintenance research to improve reliability and security. DevOps evolved out of the practicalities of running Cloud services; it is needed because Cloud services, unlike traditional non-Cloud architecture datacenters, must run continuously.

As a result, research and innovation strategies must have active feedback cycles. Activities traditionally considered innovation in the EC framework, such as prototyping, testing, demonstrating, and piloting, with the aim of validating technical and economic viability, or supporting initial application and deployment, should feed back into research in the form of discovering new problems to solve, with solutions rapidly fed back into what is being called "innovation" activities.

 ¹¹ Martin Kenney, Understanding Silicon Valley : The Anatomy of an Entrepreneurial Region (Stanford, Calif.: Stanford University Press, 2000).
¹² In the 2012 "Advances in Clouds" report, DevOps is referred to as a future development

¹² In the 2012 "Advances in Clouds" report, DevOps is referred to as a future development paradigm for the development of Cloud services, but DevOps emerged out of the needs of Cloud development in the first place, since there cannot be any down-time.

http://cordis.europa.eu//fp7/ict/ssai/docs/future-cc-2may-finalreport-experts.pdf

The American technical university approach, exemplified by MIT, Caltech, Stanford, Carnegie Mellon, is a science-based engineering paradigm, where people start to undertake new projects, filling in the research problems as they go along. The research programs enabling Moore's law follows this pattern, and firms such as Cadence mobilized the US semiconductor industry association to drive the frontier of science.

1. THE FATE OF THE "CLOUD SUPPLIERS" AND THE IMPLICATIONS FOR RESEARCH STRATEGY

With the advent of global-scale Cloud computing, everything that is aimed at the highend by optimizing for scarce computing resources will face a new wave of commoditization, whether it is a software, hardware, or network offering.

In particular, IaaS offerings will be commodifized rapidly if they attempt to compete head on with the global-scale Cloud providers. This is the area in which the European telecoms are searching to differentiate themselves—offering services that better serve national regulatory and EU-level requirements. However, global-scale, US-based Cloud providers are moving to optimize their services for local regulatory conditions as well.

A key recommendation is to avoid technologies that depend on, or support vertically integrated solutions with an IaaS component that locks in users into that specific IaaS. The problem here is that users are locked into higher cost IaaS even as global-scale IaaS providers rapidly decrease their costs and prices as they increase scale. The performance of solutions providers that can substitute in a variety of IaaS offerings and control them through a platform offering are already showing cost advantages in performance.¹³

An example of an outmoded strategy is exemplified by scientific computing efforts seen around the world, which are aimed at optimizing fixed computing resources rather than more powerfully deploying ever-increasingly abundant resources. For example, the Japanese government invested considerable resources (over USD 1 billion initially with annual maintenance costs of greater than USD 10 million) into building the world's fastest supercomputer, "K". K succeeded in becoming the fastest supercomputer in the world in 2011, but was quickly eclipsed by others—notably one from China. The applications of this supercomputer were not a focus, and K has been deployed for weather forecasting calculations and the like. The argument of the Japanese computer researcher community was that the experience of building K was giving them valuable expertise in building supercomputers, and that it represented a valuable government-industry collaboration with the Japanese firm Fujitsu. However, the view from frontier research in places such as University of California Berkeley is that most of the scientific computing needs can be taken care of using Amazon's AWS.

So what types of computing infrastructure research investments are more useful? We see great potential in the possibility of building shared global infrastructure with some scale to experiment on. For example, even in the US, the National Science Foundation is recognizing that Universities are falling behind when it comes to Cloud; there is currently no shared global infrastructure of a massive scale they can experiment on, and no single university can afford to build a cloud that has massive scale scale. The NSF has therefore launched a call for proposals to develop a nationwide distributed cloud that would support all the academic research in the US. A

¹³ While there are other factors involved as well, in recent years firms such as WiPro, which offers integration with existing IaaS services, have had faster-growing revenues than IBM and HP, which have their IaaS offerings included as in-house offerings integrated with their service.

reasonable analogy for this is physics research; Europe has had success in building very expensive shared infrastructure (i.e, CERN).

While long-term commitment to research can be valuable (as noted in Horizon 2020, "Advances in Clouds" etc), the possibility of radically realigning part way through the process is also important. This point is made at greater length in section 2 (iii) on disjunctures below.

2. PRIORITIES FOR RESEARCH IN CLOUD COMPUTING.

Our core recommendation for Europe's priority for research in Cloud computing is research supporting a *distributed model of computing with federated governance*.

A *distributed* model of computing refers to distributed workloads that are transparent and automated. It goes beyond interconnection in the limited sense of interconnecting legacy architectures. The distribution of workloads within each of the architectures must be visible, adjustable, and automated.

Federation is the governance structure of managing a distributed model of computing. The governance structure must manage the distribution, creation, and closing of workloads, which must be in real time. This differs from federating conducted by telecom carriers for traditional services such as mobile, with a broker dealing with roaming and billing accordingly. Here, the workload must be moved around, coordinating with the location of data. Authentication and authorization become critical issues.

There are several reasons that the integration of smaller Cloud systems into an integrated, scalable system is a valuable research strategy. Many of the following points in this section are derived from McGeer 2014, prepared over the course of developing this report. The paper is summarized in Appendix 1, and can be found in its entirely in the open link.¹⁴

The primary motivation for a distributed architecture at the technical level is the limitation of network capacity as Cloud delivers services via networks. This raises two issues.

First, there is the problem of sheer data volume. The current Cloud architectures increase dependency on the network with data processed primarily in concentrated locations. Distributed architecture decentralizes the processing of data to reduce the amount of data transmitted. The amount of data generated and transmitted has the potential to expand exponentially with the advent of the "internet of things," which radically increases sensor and computation needs.¹⁵

Second, there is the issue of unified availability, replication, and consistency—a challenge given latency between datacenters in different geographic sites. Without these attributes, it is difficult to have high tolerance against disasters hitting one particular geography, or fully adapt to locally based activities—due to factors such as privacy regulations, and for logistical operators reasons such as having different parts of global operations being in charge during daytime hours. Goals to aim for include a high degree of unified user management, user

¹⁴ Rick McGeer, "The Distributed Cloud as the Future of the Internet," (2014), https://us-igniteorg.s3.amazonaws.com/resource/Distributed_Cloud.pdf.

¹⁵ Ibid.

data available everywhere regardless of origin, unified logical networks that can handle distributed processes across geographies, unified user interfaces, and unified API access.¹⁶

i. Distributed Cloud

To manage the vast increase of data collection and to move processing closer to the edge of the network, several research efforts are likely to be fruitful. The key is how to reduce data through processing in local areas to reduce the amount of data transmitted.

First, one focus should be on content-aware technology. Analysis of the vast amount of data collected to prioritize and decide which should be sent along is an area with vast potential. Algorithmic rendering of complex data into simpler forms is another.

Second, next generations of Distributed Real Time Interactive Simulation (DRTIS), originally developed in gaming but allowing some combination of Cloud-hosts applications and local rendering, is receiving serious attention; the current wave is enabled by HTML5. Education, defense, and other areas are currently using DRTIS in their respective fields.

Research communities in the US and EU have begun to focus on distributed Cloud, led by PlanetLab Consortium, which began with Intel HP, and three prominent US universities including UC Berkeley, Princeton, and University of Washington. With over 1000 nodes at over 300 sites, Planet Lab has grown internationally. ONELab became the European, independently managed set of nodes.

In the words of Rick McGreer,

"PlanetLab and OneLab then formed the first federated, distributed Cloud, with independent management, separate user bases and resources, and agreements and tools which permitted users of each federate to acquire and use resources across the federation." (Rick McGreer¹⁷)

ii. Federation of Governance

To manage a distributed Cloud, a federated governance structure is needed. The critical issues are authentication and authorization. Since the current global-scale clouds have a unified administration, global-scale Cloud providers are not likely to be the drivers behind innovating standards and implementing them to allow ad-hoc federations of multiple independent Clouds. Therefore this is an attractive area for research.

Currently, there is no solution for attaining all three goals of *consistency* (all nodes see the same data at the same time), *availability* (guarantee that every request receives response about whether it succeeded or failed), and *partition tolerance/replication* (allowing the system to operate despite stochastic failures of part of the system)—as laid out in Brewer's CAP Theorem.¹⁸ Two out of the three must be chosen at the expense of the other. For example, the current option facing most users at the moment for geographically segregating Cloud services is to use Openstack with completely separate Openstack installations across different regions,

¹⁶ Many of these insights are articulated in: Florian Haas, "Greetings from Havana: A Fresh Perspective on Globally Distributed Openstack,"

https://www.youtube.com/watch?v=28p6Ls6hQJM.

¹⁷ McGeer, "The Distributed Cloud as the Future of the Internet".

¹⁸ See references and definition: http://en.wikipedia.org/wiki/CAP_theorem

unified by a single Openstack identity service (keystone). However, the split between metadata (in a relational database such as MySQL) and data—often service specific—makes replication a challenge. Other solutions sacrifice consistency, enabling both metadata and data to be replicated, but are impossible to synchronize with the metadata. Another current choice is to sacrifice availability, with service that will not be available if the data and metadata are not consistent. While there are nascent solutions such as specific services in Openstack that do not split between the metadata and data, allowing everything to reside in a relational database and allow a replicated keystone, the issue is the continuous volume replication for data synchronization has yet to be realized.¹⁹

iii. Software-defined networking

Software-defined networking abstracts lower level functionality of network services.²⁰ This is a critical area of development as the data flows from Cloud are poised to overwhelm current network capacity, especially with the advent of the "Internet of Things."

With distributed Cloud, networks can be programmed at the same level of application control as computation and storage. Traditionally, networking was not the core concern of applications, but now, in an era of computation and storage abundance, application performance often relies on network performance. To optimize network performance, programming it to recognize, differentiate, and apply different rules for specific applications is critical. The OpenFlow protocol allows this, and is being deployed by firms such as Google, Microsoft, and major telecom carriers. The core issue is to distribute and place virtual machines, and then control the network between them.²¹ This control of the network is through software-defined networking, making it a critical area for research.

iv. Disjunctures that may drive new architectures

While undertaking research in the areas above, it is critical to keep in mind that previous disjunctures, such as the mainframe to PCs, PCs to networked PCs via the Internet, and Internet PCs to Cloud, as well as the rise of mobile, and shift to smartphones, has led to the rise of new firms as dominant players. Few people have successfully predicted the next disjuncture in each shift, particularly the uses to which the new technologies would be put. New uses led to new directions in research.

A critical point in formulating research strategies, therefore, is to react quickly and recognize that *research directions may need to pivot* when it becomes clear that a new disjuncture has arrived, even if it takes an unexpected turn. For example, when it became clear that the Internet was going to become the next open global platform, the underlying protocols of the Internet, TCP/IP, rendered much of the existing research into ATM (Asynchronous Transfer Mode) networking technology obsolete. Those that were pursuing ATM the longest were the

¹⁹ For more elaboration on these points clearly laid out, see Haas, "Greetings from Havana: A Fresh Perspective on Globally Distributed Openstack".

²⁰ For a fuller definition, see http://en.wikipedia.org/wiki/Software-defined_networking

²¹ McGeer, "The Distributed Cloud as the Future of the Internet".

slowest to adjust, such as Japan's NTT, were slow to embrace the Internet and its associated possibilities.

Research into WAP (Wireless Access Protocol) with various layers of middleware to enable applications to interface with traditional mobile handset hardware, became obsolete when the advent of Smartphones, spearheaded by Apple, led to users connecting to the "full Internet" rather than a truncated, mobile-specific version on Internet content. Moreover, although WAP as it developed in the mid-1990s was pushed as a set of protocols, it was the Japanese carriers who did not use WAP—that came up with successful commercial mobile Internet service platforms. The pioneer, NTT DoCoMo, had actually been a member of the WAP forum, but withdrew as it developed its own mobile Internet platform services. The success of Japanese mobile Internet platforms led to growth of not only a multi-billion dollar content ecosystem in Japan, but also a market for middleware that solved some of the coordination and platform compatibility issues across carriers. The Japanese experience was therefore a successful pivot away from the existing trajectory of WAP.²² Yet, they too were disrupted by Apple and Google's "smartphones."

A final point about potential disjunctures stems from the regulatory context. The regulatory environment of the EU, with stronger privacy regulations and legitimate security concerns from entities such as the US government, provide opportunities to become a leader in developing a framework or infrastructure for end-to-end transactions. For business-to-business, business-to-consumer, and consumer-to-consumer transactions, privacy and security will only increase in importance. While the first deployments and widespread use of infrastructure and services for transactions came from the US, the EU, with stronger privacy laws than the US, may be more similar to other parts of the world than the US. Therefore, the EU is positioned to use the stronger privacy and security regulations to become a lead deployer of such transaction services and infrastructure.

PART III. INNOVATION STRATEGIES IN A CLOUD ERA: WHY INNOVATION STRATEGIES WILL BE KEY COMPLEMENTS TO AND COMPONENTS OF RESEARCH STRATEGIES

Innovation strategies are fundamentally not separate from research strategies. In the older production paradigms, first came a research phase, then an innovation phase in which that research was implemented. However, in the current era, the pace has accelerated, and programs that first develop research, then move towards innovation strategies, run the risk of being obsolete by the time they are implemented. The feedback from innovation strategies to research strategies is critical.

1. IMPLEMENTATION OF INCREASINGLY AVAILABLE, COMMODITIZED PARTS.

As Cloud services provide increasingly powerful, commoditized tools and parts, much of the innovation will be in the recombination and implementation of these various parts. The key is

²² Kenji E. Kushida, "Leading without Followers: How Politics and Market Dynamics Trapped Innovations in Japan's Domestic "Galapagos" Telecommunications Sector," *Journal of Industry, Competition and Trade* 11, no. 3 (2011).

http://www.springerlink.com/content/x426717k86034hh8/fulltext.pdf

to "innovate on top" rather than re-invent. The objective should be to leverage, expand, and build on top of existing open source cloud platforms, such as OpenStack.

For Cloud, a potentially successful innovation strategy includes adding sophisticated privacy and security models in these Cloud platforms, and encouraging their adoption. As noted above, given the higher level of privacy and security that will be required in Europe, this could place Europe as leader in the field. Europe does have a good track record with Linux, for example.

Importantly, the experience gained by developers orchestrating commoditized parts can feed back into research strategies. Likewise, the key for research strategies is how to gain experience in the aggregation of Cloud services that provides a basis for innovation in implementation.

2. TESTBED PROJECTS

Testbed projects can play a major role in developing capabilities surrounding the implementation of distributed Cloud with federated governance structures. They can also play a critical role in shaping the next round of research in new and promising areas.

GENIE in the US and FIRE in the EU are the current focal points where distributed Cloud standards are emerging. GENIE, (Global Environment for Network Innovations) is a US National Science Foundation (NSF) testbed that has nodes at universities around the country. Each node, consisting of about 100 cores and several terabytes of storage, can support virtual machines, connected by software-defined (programmable) networks. It uses the OpenFlow protocol, and allows users to allocate virtual machines anywhere on the network, and also specify how they are interconnected with attributes such as prioritized traffic. In Europe, the EU Future Internet Research (FIRE) initiative produced several testbeds. As exemplified Europe's success in building very expensive shared infrastructure such as CERN, large-scale testbeds offer a promising possibility.

3. DEVELOPING REGIONAL CAPACITIES FOR RAPID INNOVATION

Cloud offers significant possibilities for rapid innovation based on regional and local development. Here we raise questions rather than provide answers, but the focus should be on the following issues for innovation strategies.

First, what will be the business models for local and regional cloud based architectures and implementations models? Local and regional regulatory requirements are unlikely to allow the same elasticity and economic scale advantages of the monolithic cloud services provides by Amazon, Microsoft and Google. Therefore, what are the distinctive regional conditions that provide regional deployments with an advantage?

Second, at the level of architectures, what is the most effective way for local and regional Cloud deployments to become part of distributed Cloud services? Until the technical challenges are resolved to allow the simultaneous replication, consistency, and availability—currently not achieved—thereby enabling full-scale distributed Cloud networks, the interaction between local and regional Cloud services and global-scale services provides opportunities for innovation.

Third, what are the local capabilities that can be developed to most effectively take advantage of local conditions? Global Cloud services are growing to reach incredible scale—as of June 2014, for example, Amazon's capacity, excluding those not publically available such as Google, was estimated to exceed by fivefold those of the rest combined. The implication of this is that the number of people with deep expertise on how the global-scale Cloud really works is only a few hundred computer scientists and engineers. This expertise concentration can give rise to national and regional policy strategies that focus on local and regional Cloud deployments to ensure that national and regional IT infrastructure is not fully dependent on a handful of external experts. Finding the sweet spot between the advantages of global-scale Cloud deployments and local/regional scale deployments, from a local expertise standpoint, is another area to focus on in order to build possibilities for innovation.

4. LEVERAGING PUBLIC PROVISION OF SERVICES TOWARDS INNOVATION

Public provision of traditional services can, if crafted in tandem with research strategies, support innovation in the deployment and competitive development of Cloud services. How should Europe leverage this approach towards innovation?

Procurement bidding and contests can spur innovation, broadly, if designed well. If not designed well, however, they can significantly hinder the possibility for innovation. Tender processes, for example, that proceed in multiple steps, with a proposed offer, followed by the actual delivery multiple years later, is almost guaranteed to be several steps obsolete compared to processes that move significantly faster.

A core principle to follow in procurement should be the pervasive recognition that much of the infrastructure and tools will be commodity offerings that can be swapped in and out. In this case, procurement should be designed to spur innovation by creating as modular an architecture as possible, allowing for new generations of infrastructure services and tools to be implemented dynamically in the future, rather than becoming locked into a particular proprietary solution.

Finally, the desire for relatively local control of Cloud services, through privacy and security regulations, now has an economic reality to underpin it, with distributed Cloud architectures. The application of distributed Cloud with federated governance architectures to a variety of privacy and security regulatory contexts can spur innovations in development and deployment. This could give rise to a distinct European advantage—applied first in Europe, and then adopted elsewhere with interest in stronger privacy and security arrangements than those of the US. However, the US-based global-scale firms will continue to innovate and operate under multiple regulatory structures, so fast speed will be critical.

APPENDIX 1: DISTRIBUTED CLOUD THE DISTRIBUTED CLOUD AS THE FUTURE OF THE CLOUD AND THE INTERNET

Rick McGeer, Chief Scientist, US Ignite

Available freely at:

https://us-ignite-org.s3.amazonaws.com/resource/Distributed_Cloud.pdf

SUMMARY:

The key problem of the Cloud is the limitations of network resources, since Cloud makes IT services much more dependent on the network—while dramatically increasing the delay and reducing the capacity of the network to deliver those services. The mismatch between bandwidth supply and demand will grow over time as long as the current network and cloud architecture remains unchanged. The *only* thing which will reduce latency and bandwidth demand is ubiquitous competition. Essentially, the Distributed Cloud permits computation to move close to data and to users, both by physical movement of computation and by using software-defined networking technologies to optimize network transit for specific applications and services. The effect of programs is to reduce data; the essence of a program is to extract information from a large amount of data.

Critical applications of the distributed computing include the "Internet of Things" and other conceptions that share essentially the same idea of using ubiquitous sensors to gather realtime information, using computation to optimize various environmental parameters. If one considers the economic effect of the "computer revolution" to date enabled by widespread adoption of personal computers and instantaneous communication, the transformation wrought by the deployment of cheap, ubiquitous sensing coupled to computation is likely to dwarf the changes we have experienced so far.

In some sense, the Distributed Cloud is nothing complex: it is simply a standard Infrastructure-as-a-Service platform distributed across a wide area, optimally with deeply programmable networking between sites. However, the apparent simplicity masks significant complexities in implementation, most importantly authentication and authorization. Distributed Cloud is arising on an *ad hoc* basis in both public and private spheres, and will continue to do so as pragmatic solutions to enterprise and other needs.

As with most infrastructures, the compelling services and applications enabled by the Distributed Cloud will only become apparent after the Distributed Cloud has emerged. Most of the obstacles to the development of the Distributed Cloud are regulatory and legal in nature, and therefore are amenable to modifications in the regulatory framework. Prototype distributed clouds are already emerging, for both special purpose applications and as testbeds in the US, Europe, Canada, and Japan. Significant progress will be made on the infrastructure, services, and applications front in the next few years, and in particular significant integration with software-defined networking technologies.

APPENDIX 2: HISTORICAL STRATEGIES OF GLOBAL-SCALE, US-BASED CLOUD PROVIDERS (2006-2011)

Figure 1: Amazon



Figure 2: Google



Figure 3: Microsoft



Figures adapted from: Kushida, Kenji E., Jonathan Murray, and John Zysman. "Clouducopia: Into the Era of Abundance." *CLSA Blue Book* (2013).

CITATIONS:

- Cohen, Stephen, J. Bradford DeLong, and John Zysman. *Tools for Thought: What Is New and Important About the "E-Conomy"*. Berkeley, CA: Berkeley Roundtable on the International Economy, University of California at Berkeley, 2000.
- Haas, Florian. "Greetings from Havana: A Fresh Perspective on Globally Distributed Openstack." https://www.youtube.com/watch?v=28p6Ls6hQJM.
- Kenney, Martin. *Understanding Silicon Valley : The Anatomy of an Entrepreneurial Region*. Stanford, Calif.: Stanford University Press, 2000.
- Kushida, Kenji E. "Leading without Followers: How Politics and Market Dynamics Trapped Innovations in Japan's Domestic "Galapagos" Telecommunications Sector." *Journal of Industry, Competition and Trade* 11, no. 3 (2011): 279-307.

http://www.springerlink.com/content/x426717k86034hh8/fulltext.pdf Kushida, Kenji E., Jonathan Murray, and John Zysman. "Cloud Computing: From Scarcity to Abundance." *BRIE Working Paper* 2014, no. 2 (2014).

http://brie.berkeley.edu/publications/Cloud2014.pdf.

- ------. "The Gathering Storm: Analyzing the Cloud Computing Ecosystem and Implications for Public Policy." *Communications and Strategies* 85 (2012): 63-85.
- McGeer, Rick. "The Distributed Cloud as the Future of the Internet." (2014). https://us-igniteorg.s3.amazonaws.com/resource/Distributed_Cloud.pdf.