Who's in Control? Factors Influencing the Emergence of Industry Architectures in Cloud Computing.

Abstract

We provide an explanation of cloud computing's polarization around two emergent industry architectures. A comparative case study of two dominant cloud deployment models is presented. Archival data is collected and analyzed to characterize industry architectures and explain how they attract an installed base. We find one emergent industry architecture is based around a single actor who owns and controls an entire closed vertically integrated cloud architecture. The other is based around an actor who owns a key asset within layered modular cloud architecture and through architectural advantage attracts partners with valuable secondary assets. This second arrangement draws an installed base seeking flexible and transparent arrangements so that they can maintain control of their virtualized assets. We conclude the desire to control assets is key to the polarization within the industry. One or both sides must relinquish some control over their assets to reduce polarization. Our research generates insight into factors affecting the creation and capture of value in emerging industries in the digital economy.

1. Introduction

The internet is spawning digital infrastructures, which are defined as “the constitutive information technologies and organizational structures, along with the related services and facilities necessary for an enterprise or industry to function." [1, P748]. Some of these, notably mobile platforms such as Apple's iOS and Google's Android, enable applications to be executed based on native code that resides on the hardware employed [2, 3]. Others, such as cloud computing platforms like Amazon's AWS, enable applications to be executed remotely based on code that resides on virtualized computing resources and that are connected with users' devices via the internet [4]. Digital infrastructures, such as cloud computing deployments, share the characteristic of being delivered by networks of commercially driven actors, who individually seek to create and capture value and are organized in industry architectures [5]. In this context, organizations seek to find commercial opportunities by entering nascent industries and by contributing to new and evolving digital infrastructures. When they do so, they may seek to understand the structure of emerging industry architectures to identify attractive roles [5].

Emerging industries are often characterized as being fluid [6], as their structure has yet to become defined, stabilized and rigid. Under these conditions supply side technology has yet to mature, there exist numerous competing solutions in a phase of experimentation, and a dominant design [7] has yet to emerge. Similarly the demand side is characterized by great uncertainty, as needs may be unrecognized and unknown. In this way an industry's hierarchy of design [8] goes through an "architectural phase", such that the roles and relationships of actors in an industry go through a period of flux. In these circumstances, positions in industry are "up for grabs" as roles have not yet fully stabilized. In times of uncertainty, and with competing architectural solutions, industries may be prone to wars of architecture [9] and standards [10].

Cloud computing is an example of an emerging industry. Some claim that is "only 3 to 4 years old in reality" [11], although others [4] trace its origins to grid computing, application service provision and older technologies. cloud computing is defined by the National Institute of Standards and Technology (NIST), s US based Governmental standards body, as:

"a model for enabling ubiquitous, convenient, on–demand network access to a shared pool of configurable computing resources (e.g. networks, servers, storage, applications and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction" [12, P2]

Cloud computing is a means of delivering virtualized computing resources as services to the user in the form of: 1) virtualized infrastructure, such as computing, storage or network capabilities in a capacity known as infrastructure as a service (IaaS); 2) virtualized platform resources for the development of cloud based software applications in a capacity known as platform as a service (PaaS); and 3) virtualized software which executes on cloud based computing resources in a capacity known as software as a service (SaaS). It is claimed that the benefits that cloud computing brings to the enterprise are broadly threefold [4]. First, it simplifies the management of IT resources spanning hardware, middleware and software. Second, it provides enterprises with the ability to rapidly scale available resource capacity dynamically and on demand. Last, cloud computing reduces and simplifies the enterprise's IT cost base. A final important point to note regarding the delivery of cloud services is that there are currently two dominant deployment models, or templates of technical architecture, for cloud computing. One deployment model is termed public cloud and is defined by NIST as cloud infrastructure which is "made available to the general public or a large industry group and is owned by an organization selling cloud services" [12, P3]. Public cloud is characterized by its hosting and processing of customer data off site on physical resources that are shared with many other commercial entities that may be unknown to each other in a multi-tenancy architecture. The other major deployment model is termed private cloud and is define by NIST as cloud infrastructure which is "operated solely for an organization. It may be managed by the organization or a third party and may exist on premise or off premise" [12, P3]. In this way it is characterized by the customer choosing whether data is hosted and processed on or off site, and if done off site, then selecting whether this is done on physical resources that are shared with other entities or whether this is done on its dedicated resources in a single-tenancy architecture. In practice, there are two other cloud deployment models known as community and hybrid cloud. They can, however, be viewed as architectures that share characteristics of both public and private cloud, and are currently less significant in the enterprise market.

The expectation is that cloud computing will become a significant part of the computing industry, as organizations migrate their IT systems to the cloud and consumers increasingly use virtualized cloud based computing resources. But at the present time, it is claimed that the industry is still highly immature, and some commentators and analysts [11] suggest that less than 5 per cent of the world's total IT budget is now devoted to public or private cloud. Cloud computing demonstrates further characteristics of an early stage emerging industry as discussed above. For example on the supply side, numerous architectural varieties exist within the Public & Private deployment models and no dominant design [8] has yet to emerge. On the demand side enterprise consumers are uncertain about their needs and the ability of the solutions on offer to meet those needs.

Following on from this background, it is interesting that commentators [13] note that the industry has become polarized between the two deployment models. On the one hand, the supply of public cloud is dominated by large platform owners such as Amazon, Google and Microsoft and is appealing more to SMEs rather than larger enterprises. On the other hand, the supply of private cloud is dominated by an alliance of vendors, systems integrators and hosting service providers and is more attractive to larger enterprises. This observation forms the basis of a set of interesting research problems. What is of interest is not just enterprises' reaction to the differences in the broad templates of systems architectures that the different cloud deployment models represent. It is also interesting to consider that enterprises might be attracted to one cloud deployment model over another, in terms of the structure of industry actors within each model and the way that ownership of elements of cloud functionality is distributed amongst them. Following on from this, it is also of interest to identify and understand factors that explain how these different arrangements of industry architecture, or sector wide templates describing the division of labor, have come about.

Since cloud is a new and emerging phenomenon, there has been little opportunity to collect empirical data concerning the development, implementation and management of cloud computing solutions in practice. Instead, research to date has preferred to focus on forward looking issues such as design in cloud computing [14], as well as on enterprise cloud computing service needs and requirements [4]. It is only recently that research has emerged based on actual experiences and events that have happened in practice [15]. To that end, and given the research interest in emerging industry architectures in cloud computing, this paper documents research that attempts to answer the following research questions:

1. How can industry architectures emerging in cloud computing be characterized?
2. How do these different industry architectures attract enterprise cloud consumers?

Our research is intended to contribute primarily to an understanding of the digital economy. We believe an explanation of how certain factors influence the emergence of particular configurations of industry architectures in digital infrastructures, such as cloud computing, facilitates an understanding of value creation and capture in these emerging industries.

The remainder of this paper is structured as follows. Section two conceptualizes industry architecture [5] which is later used to provide insight into our empirical data. Section three explains our methodological approach to gathering and analyzing empirical data based around a case study comparing two instances of different cloud industry architecture. Section four analyses our empirical data using our theoretical concepts in order to explain how the two industry architectures are shaped and how they attempt to attract an installed base. Finally we discuss our research in the context of a rapidly evolving industry.

2. Industry Architectures

An industry architecture (IA) is a sector wide template that describes the division of labor across an industry wide network of relationships between actors, rather than in simple dyadic relationships. IA describes complex structures of co-specialized actors and their associated assets and capabilities as architecture. In this way Jacobides et al. [5] claim that their theory builds on the foundations laid by Teece [16, 17] concerning how organizations create and capture value from innovation.

IA can be used to describe how activities and value are divided amongst industry participants in terms of who does what and who gets what [18]. In this way its focus is on the roles of actors, the interdependencies between actors and the rule governing their arrangement, interconnections and interdependencies.

IA explains how some firms can capture more value than others through architectural advantage based on the complementarity and mobility of assets. The complementarity of assets describes the returns from a primary asset and distinct secondary assets that when combined create synergistic value. The mobility of assets describes the number of secondary assets of the same sort that can potentially combine with a primary asset. In instances where the mobility of assets is limited, their owners capture increased value on account of their increased bargaining power. It is on this basis that organizations engage in vertical integration in order to benefit from co-specialization [17]. However, Jacobides, Knudsen al. [5] move beyond this logic by claiming that organizations do not necessarily need to engage in vertical integration to derive value. Instead, they can enhance the complementarity of their assets with secondary assets, owned by another actor residing in another area of the industry architecture where they are not directly active. Jacobides, Knudsen et al. [5] go on to describe how owners of assets can capture value in their relationships with other owners of assets by generating architectural advantage by creating bottlenecks. A bottleneck describes an asymmetric dependency between two sets of assets. The asymmetry is caused by limiting the mobility of a primary asset, by excluding the entry of potential competitors into the bottleneck segment. These barriers to entry can be created for example through technical attributes, superior capabilities, driving the need for large capital expenditure or generating high positive network externalities. Value can then be driven to the bottleneck asset by facilitating high mobility in neighboring segments, which can be achieved through manipulating architectural and governance regimes defining who can participate. In this way complementarity and competition is encouraged in neighboring segments rather than in the bottleneck segment and value is pulled in.

As digital infrastructures, cloud deployment models are layered modular architectures and can be expressed as a set of design rules [19]. Similarly theoretical conceptions of industry architecture have a focus on governance and architectural arrangements [5]. In this way, this approach may lend itself to an understanding of how a variety of industry actors that make up the cloud are brought together and how their different capabilities are orchestrated into an effective and coherent whole, which addresses our first concern.

However, this approach does not help in explaining how a particular industry architecture is able to evolve in such a way that it is able to attract an installed base of consumers, which is the focus of our second research question. But, as "unbounded, evolving, shared, heterogeneous, and open installed bases of capabilities" [3, 20] digital infrastructures are able to draw upon information infrastructure theory [21]. We use this theoretical perspective to explain how architectural and governance arrangements can facilitate the bootstrapping of an installed base as part of a process strategy. Our interest here in architecture concerns the arrangement of functional elements in cloud deployment models, the mapping of these functional elements to particular actors who enable them within an industry architecture, and finally the flexibility and control an enterprise consumer has over the architecture and the functionality that it delivers. Similarly our interest in governance regimes concerns the degree to which those involved with a cloud deployment can make decisions about the way that it is organized and managed. In this way we focus on how standards are used to organize of the structure of an IA, and we focus on the nature of contractual agreements [22] between members of an IA and enterprise customer as a means to manage their obligations to each other. Finally our interest in process strategies is concerned with the way that the architectural and governance features embodied within an IA are used as attractors in order to bootstrap [20, 23] an installed base and ignite network externalities [10] to further propel the growth of a particular instantiation of an industry architecture.

3. Methodology

The approach taken to collecting and analyzing qualitative data was to conduct a comparative case study [24] of an example of the public cloud deployment model and an example of the private cloud deployment model. The intention was to collect data concerning these two examples and then to analyze and compare the industry architectures supporting each example.

3.1. Data Collection

In order to build a corpus of relevant qualitative data [25] upon which to conduct our analysis we first identified the types of data that we sought, and then identified sources of data that we yield the information we needed. Given the scope of data that we needed to collect, and the timescales within which we had to collect the data, we turned to online sources for our empirical evidence in an approach consistent with other studies of digital infrastructures [2, 3].

As a result of our data collection we collected: 471 blog entries written by respected industry experts commenting on both the supply and demand side of the cloud computing industry; web pages from cloud providers concerning their cloud services; and documents from the US Government's National Institute for Standards and Technology (NIST) a neutral source for definitions and reference architecture models concerning cloud computing.

3.2. Data Analysis

Our approach to analyzing our qualitative data was broadly interpretive and hermeneutic [26]. In this way we cycled iteratively between identifying and coding concepts in our data, and analyzing and developing concepts through deductive reasoning informed by our chosen theoretical concepts. We continued until a coherent picture emerged, that we could agree upon, of the broad IAs that supported each of the two opposing examples of cloud deployment models that we investigated.

4. Results and Analysis

We chose to study Amazon Web Services (AWS) to inform our case study of an example of the Public cloud deployment model. We also chose cloudstack, which is an open source software solution, managed by the Apache Software Foundation, to form the basis of our case study of an example of the private cloud deployment model. First, we identify the architectural and governance arrangements for the examples of both deployment models. Second, we examine how these arrangements are used to build and shape different types of industry architecture. Last, we explain how these architectural and governance arrangements are employed in order to attract different types of cloud Consumers and to establish an installed base.

4.1. Cloud Architecture and Governance

To provide focus to our study, we limit our architectural analysis to the primary functions concerning the orchestration of cloud services, an area where we can make clear distinctions between our two cases.

As a major functional component of their Cloud Computing Reference Architecture, the US Government's National Institute for Standards and Technology (NIST) refer to Service Orchestration as "the composition of system components to support the cloud providers activities in arrangement, coordination and management of computing resources to provide cloud services to cloud consumers" [27, P12]. Within this overall function, NIST describe broad sets of activities within three distinct layers; the service layer; the resource layer; and the physical layer. The service layer is where interfaces are provided to cloud consumers to access cloud computing software services in the form of IaaS/PaaS/SaaS. The resource layer provides two main functions. The first concerns resource abstraction, which is where virtualized computing resources such as storage and processing are both provided and managed. It is on these virtualized resources that the elements of the service layer depend for their functioning. The second function of the resource layer is that of resource control which is responsible for the allocation of resources to service instances, consumer access to those instances and the distribution of those instances across physical resources. In this way, the function of resource control acts as "the glue" of Service Orchestration. Finally the physical layer consists of the hardware upon which cloud functionality resides as well as the facilities in which it is located, managed and hosted.

An outline understanding of the cloud reference architecture, focusing on Service Orchestration, is essential for two reasons. First, it enables a comparison of the cases respective arrangements of architecture and governance of interfaces within and around the architecture. Second, in addition to representing functionality, each of these activities can represent independent commercial entities, as the cloud provider, and the service orchestration functionality does not have to be black boxed into one monolithic organization. Based on this understanding, figure 1 illustrates the composition of the Service Orchestration of our two cases using the NIST reference architecture as a template.

Starting with the architectural arrangements of Amazon's model of public cloud, we can see that they offer up APIs to cloud consumers so that they can access Amazon IaaS (e.g. Elastic Cloud Compute (EC2) and Simple Storage Service (S3)) as well as some PaaS and SaaS. It must be noted that the cloud services on offer to cloud consumers are owned, managed and controlled by Amazon alone. Moving on to the resource and physical layers Amazon's Service Orchestration functionality becomes a closed and tightly integrated system stack. These layers are effectively black boxed and closed to third parties. As a large proprietary system it is strictly owned, managed and controlled by Amazon. That is not to say that individual modules within their architecture are developed from first principles in house. For example, the hypervisor module in the resource layer, which runs virtual machines, is based on XEN, an open source collaborative project.

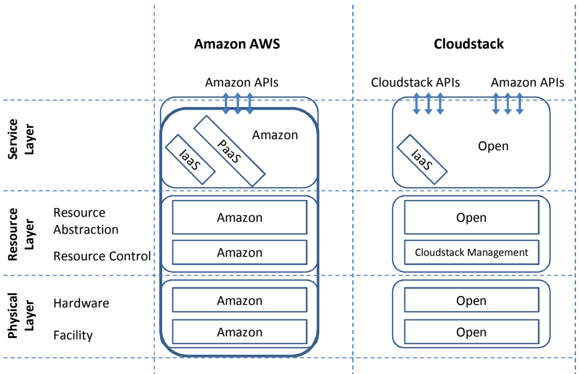

Figure 1. Composition of Amazon’s and Cloudstack’s cloud orchestration architecture.

In this way the functional and industry architecture of Amazon's Service Orchestration stack seems to be representative of other larger public cloud deployments, which appear to be run as platforms by organizations like Google and Microsoft. In this way a single public cloud provider owns and operates the whole stack as a proprietary architecture. Whilst the deployment has open interfaces at the top of the service layer thus enabling cloud consumer access, the remainder of the architecture is closed. The reason for these organizations operating these types of architectures may come down to legacy. Much of the service orchestration stack was first built by cloud providers such as Amazon and Google to support their core business. In Amazon's case this was e-commerce, in Google's case this was internet search and online advertising. Given the scale of their operations, these organizations built their businesses around orchestrating web based services over massive data centers, long before commercializing these capabilities as cloud services. This focus provided them a means to optimize their architectures both technically and operationally through techniques such as a multi-tenancy, enabling them to achieve superior computing performance on their infrastructures whilst minimizing cost. These organizations then extended into offering up public cloud services, in order to monetize spare capacity in their infrastructures, which was not being used by their core businesses.

Turning our attention to the architecture of Cloudstack's model for private cloud deployment we can see a very different arrangement. First of all, and at the service layer, the architecture can be accessed by cloud consumers using both Cloudstack APIs and Amazon AWS APIs. Furthermore, whilst Cloudstack's model for deployment is currently limited to offering IaaS alone, it is agnostic as to whose IaaS it runs as long as it is compatible with the rest of the architecture. At the resource layer, Cloudstack's architecture is completely open to different vendors of resource abstraction functionality, as long as they are compatible with Cloudstack's Management Server software, which is responsible for the function of Resource Control. This means that Cloudstacks's architecture is not just layered, but it is a layered modular architecture [28], as modular elements such as hypervisors running virtual machines are interchangeable, allowing cloud consumers a choice between implementations such as XEN, KVM and vSphere. This flexibility is replicated at the physical layer where cloud consumers are free to use a range of hardware vendors as well as deciding whether they want to implement and host a deployment themselves or whether they want to outsource elements of their private cloud solution. Last of all, it must be noted that the one element that Cloudstack does control, is the resource control function that orchestrates and manages the functionality of the stack as a whole. The rest of the stack in this private cloud architecture is open, to the extent that Cloudstack, as an open source software project, can be influenced by outsiders to support the necessary interfaces.

In this way the functional architecture of Cloudstack's service orchestration stack is representative of the many private cloud deployments, although though some private cloud stacks, such as VMWare's, may be a little more closed and proprietary at the level of resource abstraction. In general though, private cloud deployments appear more architecturally open than most public cloud deployments. Private cloud deployments are generally layered modular architectures, enabling cloud consumers to choose between different combinations of software providers, hardware vendors, managed hosting providers and system integrators. Cloud consumers are free to configure private clouds in such a way that they are closed to other cloud users through employing single-tenancy architectures, and hosting virtualized resources on their own premises or at least on the premises of tightly controlled hosting providers. However the loose coupling of the layered modular architecture of this deployment model tends to mean that the overall cost is higher and performance lower than in the public cloud. On the supplier side the private cloud deployment model is popular with vendors such as HP, software providers such as Oracle, managed hosting providers such as Rackspace and systems integrators such as IBM for two possible reasons. First, they have so far been excluded from the monolithic public cloud deployments of Amazon and Google. Second, the private cloud model is an extension of their traditional IT outsourcing businesses.

Turning our attention now to comparing governance regimes between the two cloud deployment models, we first explore the example of Cloudstack. In examples of the private cloud deployments model, such as Cloudstack, we see judicial use of both standards and contracts. At the service layer the standards governing interfaces with cloud consumers through APIs are based on both open and what have become de facto standards. Cloudstack's own APIs are based on open standards which mean that they are both transparent and open to influence. Cloudstack also allows Amazon AWS APIs to interface with its stack. These de facto Amazon standards have built up a significant installed base, upon which Cloudstack is attempting to capitalize through bridging. The interfaces that the vital orchestrating functionality of the Cloudstack Management Server (Resource Control function) has with the rest of the functionality at the resource and physical layers are also open. This means that if a third party provider of a complementary cloud asset is unable to interface at the appropriate layer, then the appropriate code allowing interoperability can be submit to the Cloudstack project. In general, the standards governing interfaces in private cloud architectures are quite open. Some private cloud deployments, such as VMWare, may seek to control virtualized assets, meaning that the standards controlling the interfaces between the resource control and resource abstraction functions will be closed, but the interfaces into the overall resource abstraction and control layer are open. The contractual relationship that the cloud consumer has with members of the private cloud deployment stack depends on the type of deployment that the Consumer chooses. If the cloud consumer elects to outsource the hosting and management of their cloud solution then the arrangement between consumer and system integrator or managed hosting provider can be based on a robust legacy outsourcing contract.

Moving onto the governance regime in public cloud deployments, the arrangement is quite different to that of private clouds both at the level of the use of standards and of contracts. With respect to the use of standards, Amazon has a set of APIs available to cloud consumers at the service layer to call upon its IaaS offerings, which are based on proprietary standards, as they are not open to influence. Furthermore, given the success of Amazon AWS in markets beyond enterprise IT, these interfaces are establishing themselves as de facto standards. However since the rest of Amazon's cloud stack is black boxed, it does not share the standards it uses for its internal interfaces in order to exclude others and keeps its architecture closed. In general the same strategy with respect to the use of standards is followed by the other owners of monolithic public cloud deployments, such as Google. With respect to contractual agreements managing the obligations of cloud provider to cloud consumer, Amazon, like other public cloud providers, takes a completely different stance to organizations that enable hosted private cloud deployments. Rather than put into place robust outsourcing style contracts which place strict obligations on the organizations enabling a private cloud, public cloud providers typically provide their customers with simple "software style" customer agreements, where their obligations to the cloud consumer are limited.

4.2. Shaping Cloud Industry Architecture

We now proceed to the second part of our analysis where we examine how the distinct architectural and governance arrangements of our examples are used to build and shape different types of industry architecture.

Whilst only being present in one part of the their private cloud deployment stack, Cloudstack have deployed Jacobedian [5] complementarity to great effect in order to orchestrate an industry architecture that creates and captures value from enterprise cloud consumers. Cloudstack provides the Cloudstack Management Server which provides the crucial Resource Control functionality that orchestrates the functionality of the whole stack. As the primary asset in the stack it encourages complementarity with secondary assets by enabling a broad set of open interfaces to a wide range of complementary assets such as hypervisors, hardware and hosting facilities owned by other actors. Private cloud deployments like

Cloudstack are composed of an open layered modular architecture. This enables assets as modules within the functional layers of the service stack to be easily mixed and matched, which facilitates high mobility amongst these complementary assets. The utility that Cloudstack provides its cloud consumers is the aggregate of its assets combined with the assets brought by the other actors taking part in the stack. Cloudstack has created a bottleneck by generating asymmetric dependencies between it and secondary assets. Not only does it provide the vital orchestration functionality for the stack, but it limits the mobility of this function by putting in place barriers to entry for competing alternatives. The first barrier is created by the fact that it is free to deploy and use. The second barrier is created by the positive network effects that it is creating and the value of its increasing network of complementors and users. As a bottleneck it is able to capture value. However, as an open source project, rather than capture value for itself, the value is created for the whole industry architecture and distributed amongst it, thus perpetuating its own value to its complements, and the cohesiveness of the assemblage as a whole.

In contrast Amazon and other monolithic public cloud providers have pursued a very different strategy to engineering an industry architecture to create and capture value. They exemplify Teecian vertical integration [17] whereby complementary assets are owned and controlled by one actor in a "black boxed" architecture that delivers the whole cloud service orchestration function. By owning complementary assets that can be tightly integrated, Amazon et al are able to derive synergies and optimize performance. They are able to derive further benefits from the scale of their operation, which drives further optimization through capacity utilization and load balancing. As a result of owning and controlling the whole service orchestration stack, Amazon is in effect its own industry architecture and is able to capture all the value that it creates.

4.3. Adopting Cloud Industry Architecture

In the final part of our analysis, we take concepts from II theory [21] in order to explain how the primary actors in each industry architecture use architectural and governance arrangements in order to bootstrap [20, 23] and attract an installed base of cloud consumers.

Before moving on with this analysis, it is worth briefly considering the broad requirements of cloud C consumers. The primary consumer focus of this study is on enterprise needs, particularly those concerns of the enterprise IT department. Enterprises are largely concerned with retaining control [29] over their cloud deployments, especially if elements of their solution are outsourced. In this way they demand transparency [30] in these solutions so that they can minimize various risks that they see in the cloud, not least in terms of a perceived lack of security [31] in some cloud solutions. Furthermore enterprise cloud consumers demand flexibility in their implementations so that they can overcome issues of lock-in [32]. Addressing these concerns is more important to enterprise cloud consumers than the performance and cost advantages offered by deployments of public cloud which seem to be relatively more appealing to SMEs and start-ups [33].

In this way private cloud deployments seem better suited to fulfilling enterprise consumers requirements than public cloud deployments. The layered modular architecture of deployment models like Cloudstack provides the flexibility and transparency that they desire. The open Cloudstack architecture means that it is easy to interoperate with a range of different providers in each of the functional layers. In this way enterprises are presented with a choice of virtualization solutions, hardware vendors, managed hosting providers and systems integrators, which in turn provides them with the flexibility to reduce lock-in and capacity to reduce other risks. Furthermore enterprises have more control in these deployments over architectural decisions such as implementing a single-tenancy computing environment on hosted servers, which may help to build trust in this model and reduce perceived levels of risk in the cloud. The confidence of enterprise cloud consumers in private cloud deployments is further enhanced by the governance arrangements in these regimes. The readiness of models of deployment like Cloudstack's to enable standards of interfaces other than their own encourages further adoption. Like a number of other models of private cloud deployment, Cloudstack allows cloud consumers to access its resources using Amazon AWS APIs thus enabling bridging between the two architectures [34]. However, perhaps the most reassuring governance element of private cloud deployments to enterprises is the potential to establish robust outsourcing style contracts with those organizations responsible for the managed hosting elements of the solution. Enterprise IT departments are historically used to detailed contracts for traditional managed hosting solutions and expect equivalence in contracts for cloud solutions. It is therefore in their interest to gain a contractual agreement with a cloud provider with service levels at least equivalent to what the customer already has [35].

In contrast, public cloud deployments, such as Amazon's, may not be well suited, in their current arrangement, to meet enterprise cloud consumers needs. Whilst the tightly integrated proprietary architecture of Amazon's public cloud deployment may deliver superior performance and significant cost reductions it does not yet sufficiently address enterprises concerns about transparency, control and flexibility. As a closed system, other actors attractive to enterprises, such as systems integrators and managed hosting providers are shut out. Bound by legacy decisions concerning architecture, Amazon is committed to multi-tenancy solutions, which is still perceived as a security risk by many enterprises. Decisions that Amazon has made with respect to its governance regime have not attracted enterprises. By refusing to offer enterprises robust contracts, which take on obligations regarding quality of service, enterprises are not reassured. These concerns have been perpetuated by well publicized outages on Amazon's cloud infrastructure [36]. Consequently enterprise cloud consumers may be less attracted by public cloud deployments, like Amazon's, especially for hosting services that are core to their needs. Instead public cloud deployments seem relatively more attractive to smaller organizations that are drawn by superior computing performance at low cost.

5. Discussion and Conclusion

This research investigates the apparent polarization of the cloud around two emergent industry architectures. It addresses two specific research questions concerning how industry architectures in cloud computing can be characterized, and how these different industry architectures attract enterprise cloud consumers. The research goes on to analyze Amazon AWS, an example of the public cloud deployment model, and Cloudstack, an example of the private cloud deployment model, in order to compare their respective supporting industry achitecture. We classified Amazon's model as a black boxed architecture. A set of APIs provide access to its cloud services, but none of the underlying middleware and hardware is externally visible. Amazon owns and controls all the assets in its vertically integrated stack. As sole actor in its industry architecture, Amazon is able to capture all the value from the synergies that this level of integration creates. We classified the model enabled by Cloudstack as an open layered modular architecture. Cloudstack, whilst only present in part of its architectural stack, was able, as sole owner of the resource control function, to act as a bottleneck and thereby attract additional actors with secondary assets. In this way it established and orchestrated its own industry architecture. In general, enterprises seem attracted to the architectural and governance arrangements of the model of private cloud deployments enabled by an industry architecture, embodied by Cloudstack. The open layered modular architecture thus enabled provides enterprises the freedom and flexibility to choose the configuration of functional modules and suppliers that they desire. Enterprises were further attracted by the governance arrangements that this model allows for in terms of robust outsourcing based contracts, which provided them with more reassurance and control. In this way it may seem that the cloud industry is polarized as much between groups of cloud providers and groups of cloud consumers who want to retain control over their assets in the cloud, as it is polarized between deployment models.

Never the less, the cloud industry is immature and it is evolving [37]. A view is emerging [38] that enterprises are gradually experimenting with using the public cloud, as they recognize its cost/performance advantages. Consequently they are experimenting with outsourcing non-core functions to the public cloud, such as SaaS based CRM solutions like salesforce.com. On the supply side, cloud architectures are emerging which may encourage a change in consumer tastes. For example hybrid cloud deployments enable an enterprise to "cloud burst" [39] in order to periodically switch to using public cloud capacity, on occasions when their private cloud capacity is temporarily exhausted. It might therefore seem that an overall dominant industry design [7] has yet to emerge. So we naturally ask, what might happen next?

Perhaps it is a sign of the immaturity of the industry, and the fact that different organizations are competing to achieve dominant design, that numerous attempts at formal "de jure" standardization in order to achieve interoperability in the industry have failed [40]. Whilst some commentators note that formal standards may commoditize the industry and blunt innovation [41], it is also possible that cloud providers perceive formal standards as a threat as they may facilitate the loss of control over an installed base that they have acquired. Rather than seeing architectural standards as a means to unite and integrate, some cloud providers may see standards as a means of control in order to attract, retain and exploit an installed base on closed proprietary platforms, such as Amazon's and Google's. These tactics, along with capitalizing upon both brand name and first mover advantage are moves of aggressors in wars of architecture [9] and wars of standards [10]. It is also interesting to note that private cloud deployments, such as Cloudstack, have employed bridges and adaptors [42] to enable Amazon APIs to function on their deployments as means to retain and even grow their installed base. This action is associated with a defensive rear-guard action in standards wars [10]. Unfortunately, whilst this may attract an installed base to Cloudstack in the short term, it may facilitate their exit should Amazon's cloud become comparatively more attractive. Given the actions of the participants, there is much to suggest that the conditions are emerging for a war of architectures in the cloud industry, particularly with respect to the enterprise market.

The views of commentators [43] is that the public cloud will ultimately win in a war of architectures. Given the value of enterprise computing, and given enterprise cloud consumer's apparent preference for private cloud deployments, it is interesting to consider how public cloud providers, such as Amazon, might respond. There may be many possible courses of action, but we present three that we think are likely. First, we think that public cloud providers may simply try to envelop the benefits of the private cloud [5, 44]. In this scenario public cloud providers attempt to persuade enterprises of the integrity of their architectures, or even make improvements to demonstrate their security. To a degree this is seen in Amazon's efforts to market its virtual private cloud. Furthermore public cloud providers neutralize other benefits of the private cloud by, for example, offering enterprises high quality "outsourcing" grade contracts. Enveloping is a strategy documented by Morris and Ferguson [9] whereby general purpose architectures absorb special purpose solutions. In this way public cloud providers would retain control of their cloud assets, and enterprise cloud consumers would give up some control of theirs. A second option is for public cloud providers to open up part of their architecture. This might involve allowing third parties to host Amazon's resource layer middleware, which is used to abstract and control virtualized resources as well as to manage hardware. The advantage of this would be that third parties would take on the burden of offering enterprise cloud consumers robust contracts, and the cloud provider would be able to keep a key asset (the resource layer) closed and intact. This option would entail the cloud provider giving up control over some cloud assets, and the enterprise cloud consumer retaining control. A third strategy might be to simply do nothing, in the hope and expectation that enterprise cloud consumers' attitudes to retaining control of cloud assets will gradually change.

Turning our minds now to how this research can be further developed, we believe that there are numerous possibilities for further work. First, there is scope to improve our research design. The study would be strengthened by taking a longitudinal perspective in order to examine how industry architectures emerge, rather than providing a snapshot as we have, although the industry may need to mature more before we can do this. The study could also be improved by gathering data concerning directly from enterprise cloud consumers rather than from the reports of industry experts. Second, the study has paid little heed as to how the unique digital characteristics [45] of cloud computing as an IT artifact, such as virtualization, have influenced the evolution of the industry. Last, the study has paid little attention as to how the service nature [46] of cloud computing has influenced its emergence.

Beyond these shortcomings, our research provides an explanation of how certain factors influence the emergence of particular configurations of industry architectures in digital infrastructures such as cloud computing. By analyzing the architectural and governance arrangements of emerging industry architectures in cloud computing, we conclude that they currently appear to be polarized between those arrangements that offer cloud providers control over assets in the cloud, and those that offer enterprise cloud consumers control over assets in the cloud. In this way we believe that our research generates insight into the creation and capture of value in emerging industries enabling digital infrastructures, particularly with respect to those serving enterprises.

6. References

[1] Tilson, D., Lyytinen, K., and Sørensen, C., "Digital Infrastructures: The Missing Is Research Agenda", Information Systems Research, 21(5 - 20th Anniversary Special Issue of Emerging Challenges), 2010,

[2] Ghazawneh, A., and Henfridsson, O., "Balancing Platform Control and External Contribution in Third-Party Development: The Boundary Resources Model", Information Systems Journal, 22(2), 2012, pp. Forthcoming.

[3] Tilson, D., Sørensen, C., and Lyytinen, K., "Change and Control Paradoxes in Mobile Infrastructure Innovation: The Android and Ios Mobile Operating Systems Cases", 45th Hawaii International Conference on System Science (HICSS 45), 2012, pp. 10.

[4] Venters, W., and Whitley, E.A., "A Critical Review of Cloud Computing: Researching Desires and Realities", Journal of Information Technology, 27(3), 2012, pp. 179-197.

[5] Jacobides, M.G., Knudsen, T., and Augier, M., "Benefiting from Innovation: Value Creation, Value Appropriation and the Role of Industry Architectures", Research Policy, 35(8), 2006, pp. 1200-1221.

[6] Abernathy, W.J., and Utterback, J.M., "Patterns of Industrial Innovation", Technology Review, 80(7), 1978, pp. 40-47.

[7] Anderson, P., and Tushman, M.L., "Technological Discontinuities and Dominant Design: A Cyclical Model of Technological Change", Administrative Science Quarterly, 35(1990, pp. 604-633.

[8] Clark, K.B., "The Interaction of Design Hierarchies and Market Concepts in Technological Evolution", Research Policy, 14(5), 1985, pp. 235-251.

[9] Morris, C.R., and Ferguson, C.H., "How Architecture Wins Technology Wars", Harvard Business Review, 71(1993), pp. 86-86.
[10] Varian, H.R., and Shapiro, C., "Information Rules: A Strategic Guide to the Network Economy", Harvard Business School Press, Cambridge, 1999,
[11] Darrow, B., "'How Amazon's Cloud Competitors Are Trying to Find Cracks in Aws's Armor'", in GigaOm blog, 2013,
[12] Mell, P., and Grance, T., "The Nist Definition of Cloud Computing", NIST special publication, 800(2011), pp. 145.
[13] Leong, L., "'Ecosystems in Conflict – Amazon Vs. Vmware, and Openstack'", in Gartner blog, 2012,
[14] Armbrust, M., Fox, A., Griffith, R., Joseph, A.D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., and Stoica, I., "A View of Cloud Computing", Communications of the ACM, 53(4), 2010, pp. 50-58.
[15] Tee, R., and Woodard, C.J., "Value Migration and Architectural Design Choices in Innovation Ecosystems: The Case of Cloud Computing", DRUID 2013, 2013
[16] Teece, D.J., "Profiting from Technological Innovation: Implications for Integration, Collaboration, Licensing and Public Policy", Research Policy, 15(6), 1986, pp. 285-305.
[17] Teece, D.J., Pisano, G., and Shuen, A., "Dynamic Capabilities and Strategic Management", Strategic management journal, 18(7), 1997, pp. 509-533.
[18] Tee, R., and Gawer, A., "Industry Architecture as a Determinant of Successful Platform Strategies: A Case Study of the I-Mode Mobile Internet Service", European Management Review, 6(4), 2009, pp. 217-232.
[19] Baldwin, C.Y., and Clark, K.B., Design Rules: The Power of Modularity, Volume 1, MIT Press, 2000.
[20] Hanseth, O., and Lyytinen, K., "Design Theory for Dynamic Complexity in Information Infrastructures: The Case of Building Internet", Journal of Information Technology, 25(1), 2010, pp. 1-19.
[21] Hanseth, O., and Henningson, S., "Towards an Information Infrastructure Theory.", Submitted for review, 2013,
[22] Reid, S., "'The Cloud-Compliance Challenge'", in Forrester blog, 2010,
[23] Hanseth, O., and Aanestad, M., "Design as Bootstrapping. On the Evolution of Ict Networks in Health Care", Methods of information in medicine, 42(4), 2003, pp. 384-391.
[24] Yin, R.K., Case Study Research: Design and Methods, Sage, Thousand Oaks, CA, 1994.
[25] Bauer, M., and Gaskell, G., Qualitative Researching with Text, Image and Sound: A Practical Handbook for Social Research, Sage Publications Ltd, 2000.
[26] Myers, M., and Avison, D., Qualitative Research in Information Systems: A Reader, Sage Publications Ltd, 2002.
[27] Liu, F., Tong, J., Mao, J., Bohn, R., Messina, J., Badger, L., and Leaf, D., "Nist Cloud Computing Reference Architecture", NIST special publication, 500(2011, pp. 292.
[28] Yoo, Y., Henfridsson, O., and Lyytinen, K., "The New Organizing Logic of Digital Innovation: An Agenda for Information Systems Research", Information Systems Research, 21(5 - 20th Anniversary Special Issue of Emerging Challenges), 2010,
[29] Verstraete, C., "'The Business Aspects of Cloud: Part 5'", in Cloud Source blog, 2012,
[30] Verstraete, C., "'The Business Aspects of Cloud: Part 6 '", in Cloud Source blog, 2012,
[31] Plummer, D., "'Legal Eagles and the Cloud: What Would Denny Crane Do?'", in Gartner blog, 2010,
[32] Perry, G., "'Integration-as-a-Service: Ibm+Cast Iron'", in Thinking Out Cloud blog, 2010,
[33] Miller, P., "'Hewlett Packard: A Tale of Many Clouds'", in Cloud of Data blog, 2012,
[34] Perry, G., "'Evaluating Cloud Computing Services: Criteria to Consider'", in Thinking Out Cloud blog, 2010,
[35] Leong, L., "'The Cloud and Customized Contracts'", in CloudPundit blog, 2011,
[36] Kepes, B., "'On Cloud Outages (Yeah, They Happen)'", in CloudAve blog, 2012,
[37] Leong, L., "'Getting There from Here'", in CloudPundit blog, 2011,
[38] Verstraete, C., "'The Business Aspects of Cloud: Part 7'", in Cloud Source blog, 2012,
[39] Verstraete, C., "'The Business Aspects of Cloud: Part 9'", in Cloud Source blog, 2012,
[40] Miller, P., "'Tosca May Prove a Prescient Name for New Cloud Standards Effort'", in Cloud of Data blog, 2012,
[41] Leong, L., "'Are Multiple Cloud Apis Bad?'", in CloudPundit blog, 2009,
[42] Hanseth, O., "Gateways—Just as Important as Standards: How the Internet Won the "Religious War" over Standards in Scandinavia", Knowledge, Technology & Policy, 14(3), 2001, pp. 71-89.
[43] Perry, G., "'10 Predictions About Cloud Computing'", in Thinking Out Cloud blog, 2011,
[44] Eisenmann, T., Parker, G., and Van Alstyne, M.W., "Strategies for Two-Sided Markets", Harvard Business Review, 84(10), 2006, pp. 92.
[45] Kallinikos, J., Aaltonen, A., and Marton, A., "The Distinct Ontology of Digital Objects: A New Frontier for Is Research", in The Distinct Ontology of Digital Objects: A New Frontier for Is Research, forthcoming, 2012
[46] Vargo, S.L., and Lusch, R.F., "Evolving to a New Dominant Logic for Marketing", Journal of Marketing, 68(1), 2004, pp. 1-17.