

Towards an Open Cloud Marketplace

Vision and First Steps

Azer Bestavros and Orran Krieger
Boston University

October 2013

Abstract

As one of the most promising, emerging concepts in Information Technology (IT), *cloud computing* is transforming how IT is consumed and managed; yielding improved cost efficiencies, and delivering flexible, on-demand scalability by reducing computing infrastructures, platforms, and services to commodities acquired and paid-for on-demand through a set of cloud providers. Today, the transition of cloud computing from a subject of research and innovation to a critical infrastructure is proceeding at an incredibly fast pace.

A potentially dangerous consequence of this speedy transition to practice is the premature adoption, and ossification, of the models, technologies, and standards underlying this critical infrastructure. This state of affairs is exacerbated by the fact that innovative research on production-scale platforms is becoming the purview of a small number of public cloud providers. Specifically, the academic research communities are effectively excluded from the opportunity to contribute meaningfully to the evolution – not to mention innovation and healthy mutation – of cloud computing technologies.

As the dependence on our society and economy on cloud computing increases, so does the realization that the academic research community cannot be shut out from contributing to the design and evolution of this critical infrastructure. In this article we provide an alternative vision – that of an *Open Cloud eXchange (OCX)* – a *public cloud marketplace*, where many stakeholders, rather than just a single cloud provider, participate in implementing and operating the cloud, thus creating an *ecosystem* that will bring the innovation of a broader community to bear on a much healthier and more efficient cloud marketplace.

A shorter version of this article will appear under the “View from the Cloud” Section of *The IEEE Internet Computing Magazine, January 2014 issue.*

The Cloud Marketplace

The notion of *computing as a commodity* that can be bought and sold is not new. As early as the 1960s, luminaries, such as John McCarthy and Douglas Parkhill, foresaw the potential for computation to "*someday be organized as a public utility*" that parallels the electricity industry and its complex set of marketplace players. In recent years, the advent of virtualization and the emergence of many public cloud offerings at various levels of the "cloud stack" – from Infrastructure as a Service (IaaS) to Platform as a Service (PaaS), and to Software as a Service (SaaS) – have fueled significant interest and new research in mechanisms for supporting the cloud marketplace.¹ Yet, much of this research remains only of "academic value" given constraints imposed by providers that make it difficult, if not impossible, for customers to leverage such innovation and to benefit from the efficiencies of a truly *open* marketplace.

The Challenges of Today's Public Clouds

Today's public cloud is dominated by a small number of IaaS providers willing and able to make the massive capital and automation investments needed to create an at-scale offering. This natural oligopoly stifles innovation and limits research; only a small number of public cloud providers have access to operational data about their clouds, have rich insights into the applications running on their clouds, can perform at-scale experiments on large-scale computing infrastructure, and have a sufficiently large share of the cloud market to implement sustainable economic models of operation. Three core elements of this *status quo* fundamentally limit competition and research. First, the clouds are largely homogeneous and are operated by a single provider. Second, all operational data is limited to the provider. Third, rather than competing solely at the infrastructure level, IaaS providers, operate higher level services as well. We examine these limitations below.

The first two properties mean that research and innovation is limited largely to cloud providers who tend to design their own hardware and develop their own software. Not only is innovation limited, but compatibility with private clouds (clouds internal to large companies) is sacrificed. In the long run, if only a handful of major providers continue to dominate the public cloud marketplace, then any innovation will need to be exposed through one of those providers. Thus, even if (say) a new hardware platform has huge value for a community of users, it seems likely that unless it is of broad utility, none of the big providers will deploy it.

The single-provider model of today's public clouds results in a strong degree of homogeneity in the cloud providers offerings. This homogeneity not only limits research, but also results in a monoculture which is susceptible to security threats. Security concerns also arise because public clouds are designed under the assumption that the cloud operator is fundamentally trusted. There are no documented technologies or even policies that keep a cloud provider, or even a disgruntled employee of the provider, from

¹ Cloud economics are compelling, both because consumers only pay for used capacity, and because producers benefit from extensive automation and economy of scale. As a result, many predict that all computing will eventually move to the public cloud [<http://beta.fool.com/marascobn1/2013/02/08/cloud-computing-future-and-your-money/23615/>].

introspecting on a customer's network traffic, computers or even private data sets. In many ways, it could be argued that an open-cloud operation would deliver some of the same security benefits that have been long-recognized for open-source software development.²

The lock that cloud providers have on operational data not only limits the ability of third-parties to innovate, but also it prevents or at best significantly encumbers important services such as those offered in other marketplaces by value-added resellers, aggregators, and auditors. The lack of such services is recognized by many as one of the main hurdles for an even broader embrace of cloud computing in many sectors of the economy.

With exclusive access to design and operational data, cloud providers gain another significant advantage: they can offer highly-competitive higher-level services such as big data analytics and other PaaS platforms, and applications. A competitive advantage for one party (resulting in an effective monopoly) tends to drive researchers and innovators away from working in related problems. In addition, since each of the big cloud providers develops different higher-level services, the big public clouds tend to become increasingly incompatible. A customer dependent on a service from one of the clouds will find it difficult to move to another cloud.

Given that each cloud provider operates its own isolated vertical silo, with no internal competition, the economic models developed by providers are designed to pull customers into a single cloud. For example, Amazon's networking charges incentivize customers to put more and more of their computing resources in the cloud to avoid external networking charges. Richer pricing models in more open marketplaces (such as those used to trade in electricity) have yielded enormous efficiencies that are not being explored in today's public clouds.

While the public cloud model is projected by many to eventually dominate all computing, the above limitations have resulted in big customers (*e.g.*, government, large enterprises, universities) standing up their own private clouds. In fact, by most estimates, the private cloud market today is far more important than the public one. Therefore, if the vision of public clouds is to be realized, we must somehow find a way to engage a broad research community and allow the resulting innovation to be relevant.

The Open Cloud eXchange

The cloud could have as big an impact on our society as the Internet has had. Some of the key architects of the Internet have argued³ that a key factor in its success has been its accommodation of "tussle" between conflicting interests, playing out in an arena which is neither unregulated, thus favoring only the largest actors, nor over-constrained with a pre-determined winner. That tussle is limited in today's public cloud: very few companies have the resources to deploy a competitive public cloud, and vendor lock-in imposes high costs on customers wishing to switch providers. Can a dynamic similar to what fueled internet

² For a summary of the arguments regarding the security implications of open-source software, see Brian Witten, Carl Landwehr, and Michael Caloyannides. 2001. Does Open Source Improve System Security?. IEEE Softw. 18, 5 (September 2001), 57-61. DOI=10.1109/52.951496 <http://dx.doi.org/10.1109/52.951496>

³ Clark, D. D., Wroclawski, J., Sollins, K. R., and Braden, R. 2005. "Tussle in cyberspace: Defining tomorrow's Internet". IEEE/ACM Trans. Netw. vol. 13, num. 3, Jun. 2005.

innovation be reproduced in support of cloud computing? Can a broad community participate and innovate as was critical to the success of the Internet?

We believe that the answer to these questions is *yes*. We posit that the public cloud can evolve into a marketplace where many actors can compete and cooperate with each other, and where innovation can flourish. We refer to the marketplace model we propose for the public cloud as the *Open Cloud eXchange* (OCX). In an OCX many stakeholders, rather than just a single provider, participate in implementing and operating the cloud. This creates a multi-sided marketplace, in which participants freely cooperate and compete with each other, and in which customers can freely choose which set of services to use from among any number of competing services and solutions that are differentiated not only by the features they offer, but also by their applicability to particular applications, by the reliability, price or security of their offerings, among other characteristics.

Our vision and ongoing work on setting up an Open Cloud eXchange at scale will expose researchers to problems that are currently invisible (because they require access to operational data available only to providers in today's closed cloud model) as well as to opportunities for innovations that are currently only of academic value (because they cannot be implemented and experimented with at scale).⁴

The OCX Ecosystem

The OCX is a *marketplace* open to vendors of software, hardware and services who collectively participate in operating open-cloud solutions; we call these participants eXchange Service Providers (XSPs). For example, the OCX might have a number of hardware companies as XSPs, who would install their equipment in the OCX and make them available to other partners or customers, on a pay-as-you basis. The OCX might also have a number of software companies as XSPs, who provide and manage IaaS management stacks or PaaS solutions that make use of equipment provided by other (not necessarily fixed) set of XSPs.⁵ The OCX might also have a number of solution providers as XSPs, who offer customers tailored SaaS solutions.

From an end-customer's perspective, navigating the OCX marketplace with all its potential XSPs can be seen as daunting.⁶ One way that this challenge has been addressed in other marketplaces is through the use of third-party intermediaries, which we call Value-Added Providers (VAPs), who offer *new* services, capabilities, or different pricing models by integrating resources and services from XSPs and other VAPs, by acting as agents for multiple customers, or by acting as brokers connecting customers having specific needs to

⁴ The development of the Internet stack and of the Linux operating system are but two examples of the innovation unlocked by providing the research community with the opportunity to contribute to open systems operating at scale.

⁵ For example, an XSP may expose a simple IaaS abstraction (just like today's single-provider IaaS clouds) that exploits whatever capacity is cheapest in other underlying IaaS offerings, or even within a single IaaS offering (*e.g.*, AWS EC2 versus Spot instances). In the OCX, multiple such XSPs can compete, thus providing customers (or other XSPs) with a range of options as opposed to the limited possibilities baked into a single infrastructure offering.

⁶ Indeed, it could be argued that one of the values of today's public clouds is that they offer a simple economic model and a small number of offerings that a cloud consumer can easily understand.

XSPs with resources that meet (or which could be made to meet) those needs. Thus, VAPs could be seen as providing a simple consumption model for a particular class of application or customers, while spanning and integrating many lower-level services, thus hiding the complexities of the OCX marketplace from end-customers.

Our concept of VAPs as third-party intermediaries can be seen as exploiting and taking to a different level an emerging trend. In today's cloud marketplace, offerings such as Heroku, Cloud Foundry, Rightscale, Engine Yard, OpenShift, packaged Hadoop, and StarCluster can be seen as XSPs that provide customers with new or specialized abstractions. We believe that the role of intermediaries and the opportunity for innovation in their creation is even more important in an Open Cloud, where there is a rich diversity of underlying services, and thus an opportunity for intermediaries to go well beyond the typical IaaS, PaaS, and SaaS landscape. It is in that sense that VAPs are distinguished from XSPs; they offer services that *cannot* be offered in today's highly stratified IaaS, PaaS, and SaaS public cloud marketplace. Among other examples of capabilities that could be offered by VAPs, we single out third-party auditing, brokerage, aggregation, and auctioneering.

A salient feature of our envisioned OCX ecosystem is that it provides an *equal-opportunity* platform not only for production, but also for experimental solutions. In particular, while well-established industry players may provide production-level services, based on private cloud software that they have developed, researchers and aspiring software entrepreneurs are able to offer alternative services on an "experimental" or "trial" basis to customers, who are given the opportunity to compare and contrast, and ultimately choose the solutions that fit them best. Within the OCX marketplace, a researcher or innovator can be seen as yet-another XSP or ASP, thus offering a path for evaluation and adoption of new solutions.

The operation of the OCX will be subject to rules and protocols that ensure a level-playing field. As an open platform, the OCX will make all operational data available to stakeholders (unless otherwise spelled out in specific provisions of a Service Level Agreement between a provider and a consumer of a service). This includes the operational data of what is running on the cloud as well as performance and problem reports. This data is critical to allow research in how clouds are used and in the limitations of existing cloud hardware and software infrastructures. It is also critical to allow customers to assess the value proposition of competing platform or solution providers. Finally, the availability of such data would make it possible for third-party auditors to engage in this cloud ecosystem.

The OCX Testbed at Boston University

A problem that has crippled many areas of systems research in the past is that researchers could only focus on applications they had direct access to. For the cloud, the community of users must include not only scientific researchers, but also companies and even individuals. They must be able to trust their computation and their data sets to the platform. Hence, while research can go on at all levels, and while experimental services should be exposed, at-scale production-level services must also be available. To tackle this challenge, we have embarked on the design and implementation of an OCX testbed that would involve researchers with both vendors and customers in a production setting.

While the concept of an OCX marketplace may have merit, bootstrapping it and setting it up as an operational testbed is a major undertaking that requires a public/private partnership to succeed. In particular, an OCX testbed must provide value not only to researchers, but also to a broad community of technology vendors and users.

The initial vision for this partnership is being driven by a multi-university collaboration anchored in the Massachusetts Green High-Performance Computing Center (MGHPCC) and incubated at Boston University. The MGHPCC⁷ is a nationally-unique model of collaboration among that brings together industry, state government, and public and private universities around a single project – to build and operate a research computing data center supporting the growing research computing need of five of the most research-intensive universities in Massachusetts: Boston University, Harvard University, Massachusetts Institute of Technology, Northeastern University, and the University of Massachusetts.

To set up the OCX testbed, we are inviting multiple technology vendors⁸ to set up their (primarily open stack) IaaS offerings on top of their own racks of equipment, and to make such services available (either granted or on a pay-as-you-go basis) to the participating universities and regional users. The OCX testbed is also leveraging some shared MGHPCC research computing capacity, which would be managed as a cloud.

As a regional initiative, the OCX testbed is giving us insights into both the model and underlying technologies of our envisioned OCX marketplace, and is providing us with a unique opportunity to house the OCX testbed in a modern at-scale data center.

⁷ See <http://www.mghpcc.org> for more details on the MGHPCC.

⁸ Technology vendors are willing, and in fact eager, to participate in setting up the OCX testbed. The interest of much of the computer industry is, fortunately, aligned with the interests of the research community in the OCX concept. This can be understood by noting that not only the research community, but also industry is excluded from today's public cloud marketplace. Efforts by major vendors to stand up their own public clouds have been largely unsuccessful, partly because few customers are interested in locking themselves into another vendor-specific offering. At the same time, major strides were made in developing IaaS platforms for the private cloud that offer services similar to those needed to stand a major public cloud. The OCX concept offers a path forward for deploying such services.

Box 1: The OCX as a Prerequisite for Rational IaaS Pricing

Today's vertically integrated public clouds offer simple fixed prices for a (virtual) computer. While researchers have explored other alternatives that can offer much greater efficiency, or encourage performance or power optimization in the applications, none of this research has been relevant to large public clouds. In addition, the kind of economic models possible in other marketplaces – e.g., electric grid – is not possible in today's public clouds.

As an example, consider the current state of affairs in pricing cloud resources. While the operational costs incurred by providers for operating such resources depend on variables such as energy cost, cooling strategies, and aggregate demand, the pricing models extended to customers do not reflect these variables – at least not directly. This disconnect between “supply” and “demand” results in an economically inefficient marketplace. In particular, it does not provide incentives for customers to express workload scheduling flexibilities that may benefit them as well as providers. In recent work of ours, we showed how a dynamic pricing model that leverages Shapley valuation⁹ can address this inefficiency by giving customers the opportunity and incentive to take advantage of any flexibility they may have regarding the provisioning of their workloads. In the absence of an OCX, the implementation of such an approach will remain mostly an academic exercise.

In an OCX setting, every research and industry partner can specify how their service is priced. For example, some partners may charge a fixed rate for the resources used, while others may charge based on demand, on power used, or based on specific workload constraints (e.g., colocation), or flexibilities as alluded to above.

⁹ Vatche Ishakian, Raymond Sweha, Azer Bestavros, and Jonathan Appavoo. [CloudPack: Exploiting Workload Flexibility Through Rational Pricing](#). In *Proceedings of the ACM/IFIP/USENIX Middleware Conference*, Montreal, Canada, December 2012. (Best Paper Award)

Box 2: The OCX as a Platform for Value-Added Data Services

Data collection is becoming "standard operating procedure" for businesses and organizations, alike. And, while a given data set (or stream) is typically collected for a particular purpose, such data could be of significant value for other purposes as well – especially if combined in creative ways with data from other sources. This is precisely the value proposition from Big Data analytics. But, for such creative uses of (big) data to flourish – and for the associated economic development to materialize – it is necessary to develop open platforms that provide the incentives and means (including secure/private access/mining platforms that are currently only of academic value) to "monetize" data assets. In an OCX setting, data as well as computational resources become assets that could be combined in creative ways by value-added providers (VAPs). More importantly, lowering the barrier to entry for new value-added data services, and consequently increasing competitiveness among providers of such services, will ensure that the true market value of data assets is realized.

Box 3: OCX-Enabled Research

Addressing the challenges of cloud computing will engage many areas of computer science, engineering and many other disciplines. It is not our role to define the solutions, but rather pose some of the important questions. A production level service comparable to today's public clouds is the platform needed to address these challenges; we do not need to wait to solve these problems before developing the OCX testbed.

Data driven analytics: One of the core value propositions of the cloud is that it will allow data sets to be shared and for public data sets to be made broadly available. Can big data platforms be made performant on top of existing virtualization technology? What networking technologies and assumptions (e.g., non-oversubscribed networks) are necessary? What storage services are necessary? What kind of provenance information should be built into storage services? At what levels should availability and fault tolerance be provided?

Security: Can we protect the user not only from each other but from individual malicious providers? Can we distribute data or compute across multiple providers without trusting any of them? How do we expose operational and performance data in a privacy-preserving fashion? Can we expose performance data while protecting vendor proprietary information? How can we federate across untrusting data centers managed by different institutions?

HPC: Can clouds provide the kinds of high performance computing capacity that traditional HPC clusters have been used for? What networking requirements are needed? Can we exploit the massive computational capacity of a cloud to move towards a more interactive rather than batch HPC model?

Software Defined Networking: A key enabling technology of the cloud is software defined networking (SDN). How do we integrate across multiple heterogeneous SDN solutions? Can an SDN approach be defined so that third parties develop services? Can we use SDN techniques across heterogeneous federated clouds? Does the customer have to trust the SDN solution?

System Software: How do we deal with accelerators, like GP-GPUs, Intel Phi processors... in virtualized platforms? How do we efficiently support high-performance RDMA in a virtualized platform? Can we support highly elastic applications? Can we develop a lingua franca for the services available at different federated sites? Can we provide end-to-end lossless SLA for distributed applications over circuit switched WANs? What are the key federation services we need to design, e.g., replicated storage across availability zones, appliance marketplace, content distribution, content ingestion...? What control does each data center need to cede to the larger cloud?

Economics: How do we expose the characteristics of different services to aggregators? Can we develop a lingua franca for this purpose? How do we represent other constraint, e.g., capacity provided for specific customer segments? What regulation can/should we impose at the marketplace level to ensure free competition? How do we keep the users from gaming the marketplace?

Domain Specific Platforms: Most of the value in the cloud is in the innovation that will occur on top of it. We have already seen platforms like CloudMan¹⁰ have a dramatic effect on particular domains. What turnkey services can be developed for other environments, e.g., fisheries, brain mapping, chemistry? What community services can be developed? What lower level platforms and interfaces will enable these platforms?

¹⁰ <http://wiki.galaxyproject.org/CloudMan>

Box 4: The OCX OpenStack Prototype

While much of the value of an OCX depends on its research and vendor partners, the OCX must provide a core infrastructure shared by all the participants in the marketplace. This common set of services is necessary for multiple partners to participate in the cloud, and for customers to move between one partner and another, if needed. Below, we summarize this set of core functionalities, which we have embarked to implementing within the OpenStack framework.

Automation: This is a cross-cutting concern for the OCX, as for any public or private cloud, as the economics depends on minimizing the need for human intervention in all but the most extreme situations. We single out three aspects of automation below.

Orchestration: This includes configuring, monitoring and upgrading the systems. OpenStack contains mature support for these functions; however it will need integration and development for the use cases and additional services required for the OCX. This is an ongoing task, since the components of OpenStack are constantly evolving, and entry of new industrial participants in the OCX testbed will bring new services requiring integration.

Automated Request generation¹¹, handling and tracking: The OCX will need to handle a large class of requests in an automated fashion, dispatching most to the appropriate entity controlling the infrastructure that is the source of the problem and notifying affected users of problem detection and resolution.

Automated problem detection, diagnosis and finger pointing: Whenever possible, the cloud needs to identify problems, especially performance problems, long before they become visible to end users. Just as with human generated requests, the system will need to forward the request automatically to the entity controlling the service that is the source of the problem.

Network Operations Center (NOC) Software: The OCX needs its own NOC for its operators to control core infrastructure and services; this operations center will in turn provide information to partners controlling their own services in the shared cloud.

Scheduling: Infrastructure introduced into the OCX by different partners may be subject to specific constraints that need to be observed and satisfied by scheduling and provisioning services. For example, an academic institution introducing federally-funded hardware into the OCX may impose constraints on its use, *e.g.*, restricted for use by federally-funded research projects, or restricted for use in support of specific subjects. Similar constraints may be imposed by vendors to limit liability or access to their technologies.

Metering and Billing Infrastructure: The OCX will leverage the rich metering functionality recently added to OpenStack as well as the rich billing solutions available from third parties. However, these solutions must be integrated with core OCX services, likely requiring modifications, *e.g.*, to handle scheduling constraints described above.

¹¹ *E.g.*, extending <http://bestpractical.com/rt/>

Shared Provisioning: With multiple administrative entities deploying OCX services, it is critical to have a layer of software that allows an entity to provision a part of the physical infrastructure, installing whatever OS and higher-level services they want on that hardware without interference with or risk to other entities. This includes both production and experimental services. We are developing a prototype of such a service to allow trusted administrators to partition the infrastructure and to allow credentials to be distributed to partners and aggregators who would then each be able to provision appropriate subsets of the physical infrastructure.

Compatibility Testing: A key attribute of the OCX as a platform for innovation is that new implementation of services will need to be constantly incorporated into its offerings. Thus, at-scale testing of these new services will be needed before they are made broadly available to customers. This is an area of significant current effort in both industry and academia, but additional work is required to support the OCX multi-provider environment.

OpenStack Services Directory: The services directory in OpenStack is used to locate instances of compute, storage, and other services to fulfill a customer request. The current OpenStack services directory assumes a homogeneous cloud, with a single instance of each kind of service. This directory will need to be extended to support multiple instances, along with a means of describing the different instances available. On top of this work, there is a need to explore how characteristics such as pricing, SLA, performance and security should be exposed, creating a service definition language for this purpose.

User Interfaces: The changes above to the OpenStack services directory will make the heterogeneity of the platform visible to programs. It will also be necessary to make this heterogeneity visible to users, along with higher-level information such as reputation, ranking, price information, *etc.* A significant body of open-source software packages is available for this purpose; candidates will be evaluated and suitable solutions deployed.