The Massachusetts Open Cloud (MOC)

October 11, 2012

Abstract

The Massachusetts open cloud is a new non-profit *open* public cloud that will be hosted (primarily) at the MGHPCC data center. Its mission is to accelerate innovation in and broaden the applicability of cloud computing. In creating an open alternative to today's public clouds, the MOC will create an environment that will foster innovation, job creation, new companies and industries, and more research and development in both the production and the scientific, big data, and business uses of cloud computing.

1 Introduction and executive summary

Cloud computing is having a huge impact on the computer industry. Rather than buying a computer, a cloud consumer gets on-demand access to computational capacity and pays only for what is consumed. A producer operates a shared service that benefits from extensive automation and economy of scale.

We introduce in this document the concept of an *open* public cloud where many *partners*, rather than just a single provider, participate in implementing and operating the cloud. An open cloud model will unlock the innovation of a much broader community to address the challenges of cloud computing. An open model will also allow us to extend cloud computing to address a broader set of applications, including high performance and big-data applications.

Our realization of an open cloud is called the Massachusetts Open Cloud (MOC); a new non profit we are creating. Partners will incur most of the development and operational responsibilities. The MOC will be responsible only for those services that span partners, such as metering, billing and defining common standards.

While the costs and risks of the MOC are limited, it will deliver much greater value than the sum of its partner provided parts. The MOC will provide a testbed and meeting ground for vendors, researchers and users of cloud computing. Vendors will be able to demonstrate and evolve their products. Customers will be able to evaluate different technology for their problems even if they eventually deploy in a public or internal cloud. Entrepreneurs and researchers will have the data and be able to do the fundamental research and development needed to realize the full potential of cloud computing.

As the first truly open public cloud, the MOC will provide the ecosystem and testbed where the various components of open and closed cloud stacks will be refactored and where we will learn how to make a complex heterogeneous cloud with contributions from many players work. It will define the reference implementation that will be repeated in other clouds around the world.

The MOC both benefits from and will act as a further engine of growth to the region. The deep regional expertise in, for example, big data, will help us enormously in defining and evolving the MOC. Further, because of the concentrated local community, we can effectively harness a relatively small development team to help local users harness the cloud and to take back learnings to our partners; modest investment is all that is required to make this concentrated community largely self supporting.

We first describe in more detail our concept of an open cloud. We then discuss the value that an open cloud could provide to both consumers and partners. The remainder of this document discusses more specifics of the MOC, our realization of an open cloud.

2 An *Open Cloud*?

An open cloud differs from existing public clouds in three major respects. First, it is open to software, hardware and services vendors and innovators to participating in its implementation; we call these participants *partners*. For example, an open cloud might have both IBM and HP providing hardware while both Red Hat and VMware provide IaaS management stacks. An open cloud is a multi-sided marketplace, like a shopping mall, where the different participants in the marketplace can interact and compete in a level playing field. In contrast, with existing public clouds the provider is responsible for all aspects of the implementation.

Second, all stakeholders have open access to rich information about the cloud operations. In particular, all partners in implementing the cloud have access to the operational data of what is running on the cloud as well as performance and problem reports. This data is critical to allow these partners to enhance existing and evolve new products. Cloud consumers will have access to performance and operational data about the resources their applications use; data that is critical for performance sensitive applications. In contrast, with existing public clouds, much of the operational and performance data is only visible to the cloud provider.[1]

Third, the cloud has an open economic model, where each partner can decide themselves how to charge for their services. For example, some partners may charge a fixed rate for the resources used, while others charge based on demand. An open economic model is necessary to allow each component of the cloud to be charged for in a fashion that reflects its costs and value and to enable customers to optimize their use of the underlying system.

While an open cloud sounds like it could be complicated, that complexity is important to allow performance and cost sensitive applications and middleware to optimize appropriately. However, we expect that most customers will see a model that is simpler than that provided by existing public clouds. For each class of consumer, services, specific to its target users will provide a model specialized to those users. In many cases we expect these higher level services will be *aggregators*, that provide a simple model for a particular class of application or customer, while spanning and integrating many lower level services.

The open nature of the MOC is critical to enabling a level playing field, which is in turn critical for innovation. Today's large public clouds, much like previous operating system platforms, offer the single provider enormous advantages which we believe in the long term stifle innovation. Two examples are the accretion of features into the cloud offering and the confidential treatment of operational data. When a consumer of a cloud develops technology that is of value to many other customers, the cloud vendor often adds the functionality to the core platform; much like operating systems became extended with browser functionality. This bloating of the platform limits innovation both by reducing the incentive of customers to provide innovative features that the platform could incorporate, and because the platform tends to get more complex as it incorporates more and more features. Also, the operational information of todays large public clouds is considered a competitive advantage, where the large cloud operators use the information to design new services and hardware. An open cloud would expose that information to a large audience to allow many partners to attack the challenges faced by cloud applications.

In today's public clouds, the cloud provider operates the cloud, assumes all the risks, develops much of the core technology, and of course retains much of the economic benefit of success. In an open cloud, the cloud provider provides the marketplace, but it is the participants in the

[1] Note, while operational data can and should be exposed, this must be done without sacrificing the privacy of the customer's data or operational information about individual customers.

marketplace that incur mosts of the costs and get most of the rewards. In particular, the partners will each control and assume most of the costs of providing and operating their own service. Each partner also determines what customers should pay for using their service and receives most of the revenue generated by a successful service.

3 The value to cloud consumers

We believe that an open cloud will be appealing to many customers. Probably the most important reason is that we expect, at least in the long term, that innovation will occur at a much more rapid rate when there are many partners innovating rather than just a single cloud provider. Also, the heterogeneity of an open cloud will be attractive to some customers. No one solution works perfectly for all needs, but a heterogenous cloud can have aspects that satisfy many more users. Heterogeneity is also the norm in large enterprise data centers, a community that has been slow to adopt existing public clouds.

An open public cloud, with a diverse set of partners, will naturally address a more diverse set of users. In fact, an open cloud will allow for more innovation than traditional data centers. For example, consider an exotic machine that might have two orders of magnitude advantage for a particular class of application. Today's public clouds would be unlikely to deploy a machine that targets a niche audience due to the complexity to integrate it into their automation. Traditional data centers would find it difficult to justify the cost of acquiring and managing an exotic machine. On the other hand, such a machine is relatively easy to deploy in an open cloud since the partner developing the machine is the one responsible for its operation. The machine is available to all users of the cloud, so it is cost effective to develop and operate such a machine if even a small percentage of applications are greatly advantaged by the machine's unique characteristics.

Another reason that an open cloud will be appealing for consumers of the cloud is because it avoids vendor lock in. The specifics of the services and capacity a customer uses are visible to them; a consumer can use multiple types of capacity concurrently and/or can switch from one type of capacity to anther depending on price, or other characteristics. Because the different platforms co-exist in the same data center, it is much easier for a customer to move the data, or access it remotely, from an alternative technology. Also, the consumer can use the same services and capacity if they shift to another internal or external cloud.

The information exposed by an open cloud will force it to deal with security up front, which is far better than relying on security through obscurity. Also, the information will allow customers and third party security firms to audit the cloud. We expect that some of the cloud partners will provide highly security capacity and services, while others will focus more on other characteristics like performance or elasticity. The information they expose should allow the informed customer to make the appropriate tradeoff.

Another key value of open clouds to customers (and vendors) is that it allows for novel economic models (e.g., using colocation-as-a-service, Shapley pricing, etc.) These models are impossible to deploy on existing public clouds because they are inconsistent with the controls, architecture or terms of use of the cloud. The use of such economic models have been shown to result in more efficient marketplaces, with significant payoffs to customers.

4 The value to partners

We expect partnerships with a large community, ranging from existing hardware and middleware vendors that target the enterprise, to the scientific computing community, to new PaaS and Big Data vendors, to cloud researchers all the way to the existing large public clouds. For each of these communities there is a benefit in participating in an open cloud.

The open cloud is attractive to hardware, middleware, and appliance vendors that today focus on the enterprise customer; customers that largely use internal rather than public clouds. The open cloud provides these vendors with a way to address startups that will grow to become future enterprises. The realities of the startup market today is that these companies need to rely on a public cloud, and hence they have little exposure to enterprise vendors technologies. If the startup instead grows up on an open cloud, they may use vendor exposed technology, and are more likely to adopt that technology when they move in the future to an internal cloud.

Partnering in an open cloud is also critical in exposing the enterprise vendors to the fast changes taking place in the cloud world. Today they are often challenged when an enterprise customer demands some feature or application made popular in a public cloud. Having access to operational data about the open cloud will expose the vendor to the trends far earlier. Understanding emerging workloads early, and gaining data about the requirements and characteristics of the workload will allow the vendors with deep expertise to develop new products for these workloads. This will give them a competitive advantage not only within the open cloud, or to enterprise customers, but also to offer technology to the massive public clouds that they are locked out of today.

Partnering in an open cloud will also give these vendors a target to point their customers to in order to deal with short term or cyclic changes in workload, or to deal with disaster recovery. In this case, the open cloud would be used to supplement an enterprise's internal cloud. Today's public clouds are being touted extensively for such "burst" workloads, but the utility is limited only to services that are built into or above the public cloud.

An open cloud is clearly valuable an enterprise vendors that provides only part of the technology stack for a cloud to showcase their technology. We also believe that it is critical for the very large enterprise vendors (e.g., IBM, HP...) that have the capability of deploying their own clouds. Even though they have the capability, these vendors have been slow to develop their own public clouds for a number of reasons. First, many of their customers rely on heterogenous data centers with many vendors technology; it is difficult both technically and politically for any single vendor to deploy such a data center. Second, many customers would be concerned about vendor lock in. Third, operating their own production cloud is a radical departure from the business model of selling technology to customers. Supporting their technology in an open cloud is much less of a change of model.

Vendors are already making huge investments in technology to enable internal clouds, such as the vCloud Director and Open Stack IaaS software. This is critical to us in developing an open cloud, since most of the pieces are already in place. The open cloud will then provide the ecosystem where vendors and the larger community can experiment with how all the parts work together, and where the reference implementation and definition of how complex heterogenous clouds that satisfy many users will evolve.

We expect that the enterprise vendors will all line up to support open clouds much like they lined up in the late 90s to support open source software like Linux. The value of engaging are large, the risks of engaging are relatively small, and the risk of being on the sidelines are great.

While the open cloud is of broad value to partners, partners for which it is essential are: 1) systems researchers, and 2) researchers and innovators in performance demanding areas like scientific computing and big data. Systems researchers is enormously difficult to make relevant when there is no visibility into the applications running on the cloud, and when the system researcher has no ability to experiment with systems of relevant scale. For performance sensitive applications, today's clouds are appropriate only for the simplest applications. However, the value proposition of an elastic, secure, and scalable shared infrastructure is at least as important for these applications as for the existing cloud workloads. Enormous innovation will be unlocked if any scientist can gain access to thousands of cores, on demand, while only paying the cost for the short period of time that they are using the resources. While we will support a broad spectrum of partners, we expect extensive collaboration with partners focused on research and developing HPC and big data

services.

For large public cloud operators, the primary benefit of becoming partners is to have access to the innovation that is going on in the open cloud. The focus of the public clouds is to provide a service at massive scale, and many kinds of experiments in their own clouds can be disruptive to their data centers processes and procedures, automation, and/or business models. The open cloud gives them access to a testbed to explore new ideas in cloud computing at lower risk.

The large public cloud operators do not view open clouds as a competitive threat, since open clouds will not develop the scale and geographical distribution soon, if ever. The public cloud operators are also supportive of the open clouds because it will help train a generation of students and researchers to address issues like scale and security that are critical to them. A public cloud provider can hence gain a great deal of value by participating in an open cloud with very low risk.

5 The MOC

There are reasons why today's clouds are homogenous and relatively simple. The scale of today's public clouds require a huge degree of automation. This automation has only been practical for a single cloud vendor to achieve if they are highly prescriptive in the hardware chosen, in the computational model exposed to users, and in the economic model. The cloud provider can't develop the automation to work with many different hardware platform, to work with different base capabilities, or to provide a wide variety of business models.

The MOC will be a heterogenous cloud from its inception, with multiple hardware vendors for compute, storage and networking, with multiple partners of IaaS stacks, multiple higher level aggregators... Automation continues to be important as the cloud increases in scale, however, each partner is responsible for automation at their level. While the aggregate cost of automating a heterogenous open cloud is much larger than a homogenous closed one, that cost is distributed across vendors that are, in many cases, incurring that cost already for other markets.

The responsibility of each partner in the cloud will be: 1) to develop the technology they are deploying, 2) to maintain the operations of their service, 3) to deal with service failures, 4) to provide the resources to integrate their services with other services that will use it, 5) define the economic model for charging for their service, 6) help with the tooling for the monitoring and metering of their service, and 7) to pay other costs incurred by the MOC to deploy their service. For example, a system hardware vendor that expects their hardware to be used by OpenStack would be responsible for deploying their systems into the data center, would implement the integration into OpenStack and the MOC's monitoring and metering infrastructure, and would define the model of how they want to charge for the use of their hardware.

Since most of the work is deferred to partners, the team involved with the MOC will be comparatively modest. However, some shared services and functions are essential, and these will be managed, contracted by or where necessary developed by a small team in the MOC. These include: 1) defining the rules for accepting a partners contribution (e.g., all operational data exposed), 2) a base service to report or identify problems and forward them to the appropriate partner, 3) a common metering, monitoring, and operational data retention infrastructure, 4) shared security analysis and auditing to ensure that no partner's service can impact the larger service, 5) aggregation services, 6) outreach and educational services to expose the MOC to critical user communities, help port applications to the services, and bring back the shared learnings to the partner community, 7) project management of services that cross multiple partners, and 8) manage networking services that span services and connect those services to the external world.

The MOC will lease space from the MGHPCC data center, and to start with, most of the physical capacity will be located there. It will, however, federate to other clouds as possible. This will be important both to study the issues of geographical distribution, and to enable applications to span geographies for locality or to tolerate failures. We are currently in conversation with several

academic/research clouds about federation.

The MOC differs fundamentally from today's clouds in that it has neither a profit nor a research agenda itself. Its goal is to create an environment to enable profit, innovation, and research by others. We have talked in the previous pages about how it will create this innovation, and how it will generate direct and indirect profit for its partners. The MOC is also directly impacted by its location in Massachusetts and will result in long term advantage to the region.

The Commonwealth is uniquely suited for the MOC for a number of reasons. Clearly, the investments by the state in building the MGHPCC are critical to enabling the MOC, and the MOC will help exploit those investments. There is very strong local industry and research universities that will participate as partners and users of the MOC. This highly concentrated community means that a much more modest investment is required to get the local community both highly engaged and largely self sufficient in exploiting the MOC. Finally, Massachusetts is developing a strong ecosystem around big data. The MOC could become the central shared data repository for New England academic, R&D, commercial, government and community organizers.

One way the MOC will impact the regional economy is that partners will tend to locate or hire staff involved in operating their portion of the MOC locally. At least two of the partners we have discussed the MOC with are already discussing how they can build up a local team.

Another way that the MOC will affect the local economy is by enhancing businesses and local scientists that will use the MOC. While we will not limit consumption to regional users, we will build a local team of programmers that can directly interact with local potential users. While the fundamental motivation of this local engagement is to achieve a critical mass of knowledge that can be in part self supporting, the side effect is that local users will be most advantaged by the MOC. Also, our programming resources will bring back to the MOC's partners information about the computational requirements of local companies, such as the strong regional focus on big data. This means that the MOC will help advance regional priorities.

The MOC will also advantage regional investment because we expect to issue the state and regional university sponsors credits for computational capacity to compensate them for their investments. These credits will be, in turn, be provided to local industry and researchers.

Educationally, we will involve regional students through courses and internships in the MOC. While students in most universities are only exposed to traditional computer systems, students in the involved institutions will be among a very small number of students in the country exposed to modern large-scale datacenter. We expect that these students will both be highly sought after world wide and will be naturally creators of new startups.

6 Getting there

It will be a substantial challenge to get the MOC to the level where it will be self sustaining, and where it will generate the degree of innovation required for the value propositions described above. We first describe our plans of building out the service, and then discuss how at each phase the service will be funded and generate value.

The development of the MOC will go through three basic phases: 1) exploratory and initial development, 2) expansion and stabilization, and 3) self-sustaining. The first exploratory phase will last for one year. We will in that year build a modest scale offering with two partners in HW, two partners in IaaS management stacks, and we will focus the MOC on a small number of target applications domains. During that exploratory phase we will build the rules, select or build the key shared functionality described above, and build the required development team. By the end of the first year, the service will be of sufficient scale and stability to be of interest to a community of users.

In the expansion and stabilization, we will focus on four things: 1) expanding the partners to address a larger set of applications, 2) building or supporting aggregation services targeting

important customer segments, 3) expanding the capacity, and 4) federation to other clouds. Service in this phase will be relatively stable. We will continue to identify key applications areas, and deliver ongoing feedback to the partners about application and cloud requirements. Partners will derive ever increasing value from participating in the cloud, and we we will work with our partners to refactor their offerings and define the reference implementation for other open clouds. This phase will last two years and by the end of this phase both the partners and the MOC should be able to fully pay for operations through the fees charged to users.

By the fourth year, the service should have achieved significant enough scale and stability to be of interest. The operating data will be widely disseminated and the basis for a very large number of academic publications. The public clouds will start getting involved.

While it may seem aggressive, we believe that we can create something of this scale and get to a profit in a short time because most of the partners have already done many of the initial investments. IaaS offerings are reaching maturity for the service provider and internal cloud markets. Two examples are vCloud Director and OpenStack. The vendors have started building into their offerings differentiation for key cloud applications.

We don't foresee a problem with getting vendors to stick with us through the first three years. While their service will not be operating at a direct profit in this phase, indirectly they will already start reaping the value through understand applications, marketing, etc...

We will focus in the first two phases on customers that will get value, especially Big data, HPC and enterprise customers. Even if the scale or stability of the MOC will not be sufficient to host the production use, the MOC allows customers to experiment and compare the various architectures under the same conditions. The MOC will hence acts as a proving ground for applications and for the right technology on which those applications should be deployed.