Global Enterprise Cloud Transformation: centralize, distribute or federate?

David Breitgand   Alex Glikson
Cloud Operating System Technologies
IBM Haifa Research Lab, Haifa, Israel
Email: {davidbr, glikson}@il.ibm.com

***Abstract*—Cloud computing rapidly gains popularity as a cost-efficient way of provisioning and consuming IT services. There are two fundamental cloud computing deployment paradigms: public and private. While public cloud has gone a long way since the concept inception, it still fails short on a number of enterprise requirements. We believe that in the foreseeable future the private cloud model, in which the cloud is fully owned by the company and serves users from this company only, will remain the dominant one in the enterprise. While agility is achieved in a private cloud by using virtualization, automation, and self-service, the economic model of a private cloud is not as clear as that of the public one.**

In this position paper, our focus is on qualitatively analyzing organizational and business factors affecting the cost-effectiveness of private clouds for global enterprises. We analyze the global enterprise structure, which often comprises an overwhelmingly complex and intertwined networks of global organizations, and suggest a simple classification model, reflecting the private cloud maturity levels. We discuss technological and organizational issues arising at each level of maturity and outline a research agenda to address pain points in private cloud computing.

I. Introduction

Public cloud providers offer virtualized IT resources as a metered service over the Internet. Over the last few years the popularity of the public clouds increased tremendously with more and more vendors offering cloud services and more customers choosing IT consumption through Cloud over other alternatives.

According to recent Gartner's latest quarterly IT spending report, the worldwide market for public cloud services (IaaS, PaaS SaaS combined) was 91 billion USD in 2011 [1]. Gartner expects it to increase steadily in the next few years and to top 200 billion USD by 2016. Other analysts provide prognoses ranging from half of Gartner's predictions to twice as high ones. The general agreement seems to be that "Cloud Wave" will continue to grow for a few years to come.

Some reality check is required with respect to enterprises, however. Namely, how much of this growth can be attributed to the enterprises? According to Deloitte, between 2010 and 2012 the Cloud-related IT spending of the enterprises are between 0.75% and 1.75%, respectively, out of the global enterprise IT budget [2]. The same study expects this percentage to grow up to 15% by 2020. Other studies give slightly different figures, but they seem to be in agreement that enterprises are slow adopters of the public cloud paradigm in spite of its obvious advantages allowing to reduce CAPEX, software, and licensing costs.

Does this slow adoption rate simply reflects inertia inherent to large enterprises or is this rooted in some real problems with the public cloud model when applied to enterprises? Among the top requirements on the cloud cited by the CIO Cloud Computing Adoption Survey (Jan 2011) [3], were high performance (including reliability and availability), enterprise level security, enterprise level quality of service, vendor-independence, and regulatory and standards compliance, which is imperative for successful integration with the business internal systems. In a recent State of the Cloud survey by InformationWeek [4] virtually identical set of concerns with respect to the cloud computing was cited. Moreover, they only marginally change since 2010. Unfortunately, public clouds are still deficient with respect to these requirements even though they steadily improve.

Private cloud is currently being explored at various stages by as many as 80% of the enterprises [5]. It turns out, however, that transforming an enterprise IT to a cost efficient private cloud is far from being straightforward. The complexity stems from the fragmentation of business processes, complex organizational structure of the global enterprises, lack of the practical economic models for global capacity sharing, and the need to simultaneously support clouds at different levels of maturity. The global enterprise scale and organizational structure mandate a decentralized architecture for resource sharing coupled with globally unified business processes.

Thus, in this position paper we raise the following questions. What future global private cloud architecture will look like? What will be its impact on the enterprise business processes? What architectural building blocks are missing? What organizational changes may boost the adoption and efficiency of private clouds? While public cloud model received a lot of attention from research community recently, there is virtually no literature systematically exploring requirements and architectural alternatives for the global enterprise private model in spite of its importance. We regard this position paper as one of the first attempts in this direction. Our study is an architectural, qualitative one. We defer the quantitative evaluation of our propositions to a later stage.

II. Related Work

The bulk of research efforts on decentralized clouds focus on federated and distributed public cloud models. The private cloud model received considerably less attention. An extensive literature exists on facilitation of resource discovery. monitoring, orchestration, brokerage in the context of the federated [6],

978-3-901882-50-0

[7] and distributed [8] in the public cloud. Initial modeling of business process impact is given in [9].

In the context of the public cloud model, the researchers focus on profit maximization for providers. This is not the goal in the private cloud, where the same enterprise executes the roles of cloud provider and cloud consumer under different organizational "hats". In the private cloud model, the focus is on systematically reducing the costs associated with CAPEX and OPEX of IT, improving the overall enterprise profitability.

In [10], cooperative virtual machine management for multi-organization cloud is considered. The authors creatively apply cooperative gaming theory mechanisms, such as core solution and Shapley value [11] to apportion costs in a coalition of symmetric organizations to obtain stable and cost efficient collaborations. This study is an important first step in researching the fundamentals of the private cloud transformation. Further research is required to achieve models of practical importance.

III. Global Enterprise Structure

The organizational structure of enterprises can be extremely complex. In the simplest case, a global enterprise comprises a hierarchy of *global organizations*. Each of these organizations can be either a profit or a cost center directly influencing the profitability of its parent organization. For example, an organization may correspond to a product or a service of the company, generating external revenues, or be a cost center, having a business enabling role, such as organization's IT services. Each organization comprises what we term *organizational cells*, which are 'leafs' in the hierarchy. The plurality of IT resources, IT personnel, and users belonging to the organization at a local premises, operating within the budgetary constraints imposed by the parent organization, is the cell of this organization at this premises. Cells are not cost or profit centers on their own, but contributors to their parent organizations. Since premises are shared between organizational cells, there are matrix relationships between organizations in managing and governing premises.

Figure 1 shows the geographical structure of a fictitious Acme company, which is incorporated in the USA (shown as a larger circle), having multiple foreign subsidiaries around the globe (shown as smaller circles) with each subsidiary owning one or more premises. Acme has a hierarchical organizational structure that we do not show explicitly due to the lack of space.

IV. Enterprise Cloud Transformation

We differentiate between the three basic maturity levels that offer increasing value: intra-cell, intra-organizational, and inter-organizational enterprise private clouds. As the level of maturity increases, providing higher value, the technological and business challenges also increase.

A proper transformation to the cloud model requires (a) business process transformation and (b) technological transformation, which are equally important. Figure 2 summarizes the Operation Support System (OSS) and Business Support System (BSS) components required to implement an enterprise cloud, focusing on those which are most relevant for the enterprise cloud transformation.

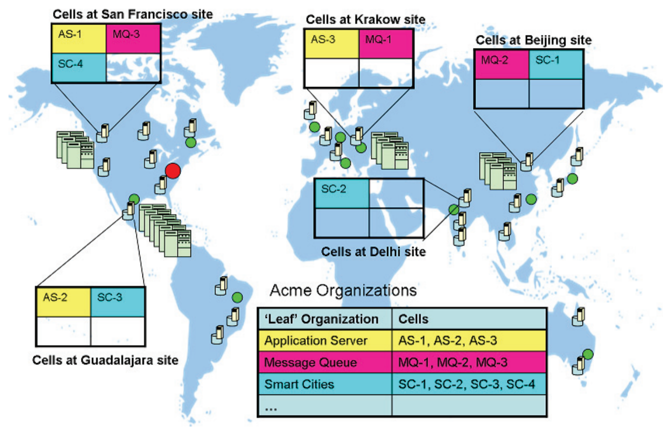

Fig. 1. Acme corporate structure: subsidiaries in multiple geographies with multiple premises in each geography.

From the business process transformation perspective, the challenge lies in the fact that the private cloud model intrinsically blurs distinctions between the consumer and provider roles. Cells acting as providers bear the costs of provisioning and effectively reduce profitability of their parent organizations if these costs are not recovered. The consumers, on the contrary, might potentially increase their profitability on expense of the intra-enterprise cloud providers, if usage is free of charge or pricing does not reflect the actual costs of the providing cells. This situation is not remedied trivially by introducing inter-cell chargeback mechanisms, since careless introducing of chargeback might create unplanned and undesirable profit and cost centers without bringing about a real benefit to the enterprise as a whole.

From the supporting technology perspective, agile resource sharing through efficient resource discovery and brokerage, coordinated capacity and workforce planning, scalable usage metering and associated reporting, accounting to enable proper cost sharing at the business level come to the forefront as the more important technical challenges. Some of these building blocks already exist and should be integrated into an enterprise private cloud architecture, yet other, such as coordinated capacity and workforce planning, still need to be developed.

We now specify an informal short-list of global enterprise private cloud requirements relevant to this discussion.

- RR1: "Fairness": no profit center utilizing cloud model services should be able to unilaterally improve its own profitability in a way that reduces profitability of any other profit center in the company.
- RR2: "Efficiency": overall IT costs, computed at the parent organization level, should be strictly lower in the enterprise cloud model than those attainable without it, for the same resource demand.
- RR3: "Rationality": all cells/organizations that contribute resources to the enterprise private cloud are better off doing that than not participating in the private cloud.
- RR4: "Manageability": there should be simple business processes allowing to plan and operate global enterprise private cloud and easily understand, track, and visualize

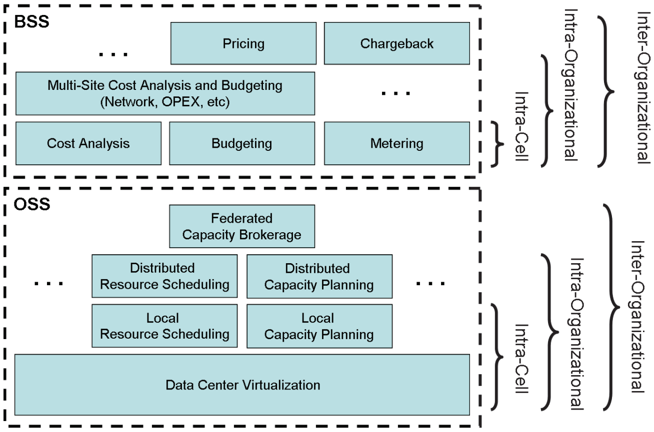

Fig. 2. OSS and BSS components required at different levels of maturity of the enterprise private cloud.

benefits accrued to the company and its constituents, as well as identify inefficiencies.
- RR5: "Economic Incentives for Consumers": proper economic incentives should be set to influence users behavior in such a way that they use only the needed amount of resources.

V. Enterprise Private Cloud Maturity Levels

Consumers: / Providers:	Single cell Single org	Multiple cells Single org	Multiple cells Multiple orgs
Single cell Single org (centralized)	Provider and consumers come from the same cell within the same organization (*)	Provider is a single cell, while consumers within the same organization are spread across different cells (**)	Provider is a single cell within an organization, consumers are multiple cells from multiple organizations (***)
Multiple cells Single org (distributed†)	Providers are multiple cells within the same organization, while consumers come from a single cell within the same organization (**)	Providers and consumers are spread across multiple cells within the same organization (**)	Providers are multiple cells within a single organization, while consumers are spread across multiple cells from multiple organizations (***)
Multiple cells Multiple orgs (federated)	Providers are multiple cells within multiple organizations, while consumers are a single cell in a single organization (***)	Providers are multiple cells within multiple organizations, while consumers are spread across multiple cells within a single organization (***)	Providers and consumers are multiple cells within multiple organizations (***)

Maturity levels: intra-cell* (Section V-A), intra-org** (Section V-B), inter-org*** (Section V-C)
† in some cases a federated approach between cells within the same organization might be also applicable

Fig. 3. Enterprise private cloud taxonomy

Different approaches to enterprise private cloud architecture can be compared along multiple dimensions. We believe that there are two architectural aspects that allow classifying solutions for qualitative (and subsequently quantitative) study: (a) inclusiveness of consumer/provider pools and (b) centralization level. Providers pool specifies cells that provide private cloud services to the users from the consumers pool. Consumers pool specifies the cells, whose users are eligible to obtain services from the providers pool.

With respect to centralization level, the following models can be applied to the providers pool: "centralized" (resources and personnel are logically and physically centralized at a single cell on a specific site), "distributed" (resources and personnel are logically centralized, while physically distributed over a number of cells, which are not independent in their resource and workforce management and budgetary decisions), and "federated" (resources are logically centralized, while physically distributed, with each cell exercising its own autonomous resource and workforce management decisions and enjoying a measure of fiscal autonomy delegated to it by its parent organization).

Figure 3 shows the private cloud architectural classification resulting from different combinations of the aspects above.

A. Intra-Cell Cloud

A single organizational cell is a cost center of a parent organization. Transforming the cell into a private cloud by introducing virtualization does not alter this. Since both consumers and providers belong to the same cost center, chargeback in this model corresponds to an associative cost reapportionment (equivalent to budgeting) among the users of the same cost center and will not have any effect on profitability of the parent organization (profit center).

The fact that the chargeback is not required simplifies cloud transformation. On the down side, it reduces the observed elasticity of workloads, since the users do not have proper incentives of using only the resources they need, when they need them. This problem can be mitigated by introducing quotas and policies on resource usage. Intrinsically, these means are imprecise and imply enforcement delays.

Another strong disadvantage is that capacity planning is done locally in each cell on the relatively small statistical population, typically targeting the peak load plus some headroom. This makes cost reduction offered by the intra-cell model smaller than what would be possible to achieve if a more global capacity planning procedure was utilized.

Summarizing our analysis of the intra-cell private cloud, we note that this model does not address Requirements RR2 and RR5. Thus, on the business side, the main challenge is to properly budget non-cooperative cells in terms of capacity and workforce to bring about global cost improvement. On the technological side, the main challenge is to plan capacity such that this budget would be minimal, yet sufficient to meet the requirements of the users.

B. Intra-Organizational Cloud

While simple intra-cell private cloud is a realistic first step in the enterprise private cloud transformation process since it allows different cells having different requirements to gain experience with the cloud technology quickly and make plans for the next steps, in the long run multiple isolated intra-cell clouds are inefficient. Henceforth, the next maturity level of the private cloud model is the intra-organization private cloud where providers and consumers can come from multiple cells within the *same* organization. Unfortunately, due to the lack of space, we will limit discussion of this maturity level to a short summary as follows.

The intra-organization model addresses requirement RR1 similarly to intra-cell cloud and addresses requirement RR2 better than the intra-cell model (thanks to leveraging of statistical multiplexing) across cells. With respect to cross-cell resource sharing model, two architectural approaches are conceivable: federated and distributed. Addressing RR3 in the federated model in the intra-organizational cloud is an open question. Addressing RR4 might be more difficult than in the distributed case due to management overhead implied by $n * (n-1)$ inter-cell relationships that need to be maintained.bOn the positive side of this model is that there is no central agency that has to make decisions and the system is fully symmetric, which makes it robust and easy to seamlessly absorb new cells.

In a distributed model, a distinguished provider cell is vested with policy making decisions for all provider cells. We refer to this as *managing cell*. The managing cell receives inputs (e.g., demand figures, cost of production) from the provider cells and exercises a cross-cell capacity and workforce management process, instructing cells how much capacity should be where, how the intra-organization network serving the cloud traffic should be managed and planned and how the capacity of the different provider cells should be provisioned to user from different cells within the organization. The distributed model has even better potential to reduce costs (RR2), it has much better manageability (RR4) as an organization grows. Requirement RR3 might be more easily addressed in distributed model by applying core solution or Shapley value techniques, but this requires additional investigation. It seems to be equivalent to the federated model with respect to other requirements. Similarly to the previous maturity level, requirement RR5 is not addressed by the intra-organizational cloud.

C. Inter-Organizational Cloud

Technological means developed at the previous maturity levels, namely data center virtualization and capacity planning at the intra-cell level, accounting, distributed or federated resource scheduling and capacity brokerage at the intra-organizational level, are reused at the inter-organizational level to obtain a company wide private cloud. In addition to those, a new technological means, chargeback, is required to satisfy RR1. Introducing chargeback in a federation of symmetric organizations, while preventing unintended consequences, such as creation of unplanned internal profit centers, is a challenging open question. RR2 is better addressable in this model because company-wide statistical multiplexing can be leveraged. Efficiently addressing RR3 appears to be doable [10], but additional research is required to develop practically applicable coalition formation schemes. By introducing an appropriate chargeback mechanism, RR5 can be satisfied as well. Naturally, the only resource sharing model applicable in this case is federation, because provider cells may belong to different organizations. The drawbacks of federation are exacerbated at this level, which poses challenges and suggests an architecture based on distributed model, as we outline in the next section.

VI. Conclusions and Future Work

We argue that the best approach to designing a global enterprise private cloud is to establish a global provider organization, which would build and maintain a distributed multi-cell provider architecture, serving users from all the organizations and cells, company-wide. The locations of the provider cells will be determined from the plurality of all the company sites (as well as new locations, such as those available via off-shoring). These cells' CAPEX and OPEX are no longer managed by other organizations that become consumers paying only the OPEX costs resulting from the metered cloud usage according to the classical "pay-as-you-go" chargeback model even though the infrastructure still resides on the local premises of various organizations. The calculation of business indicators of each organization would be simplified. The global cloud organization will be managed as a cost center exercising organization-wide workforce and IT resources management and capacity planning with cells of this organizations having local presence at different premises. This will leverage the existing structure of the enterprise and create a next generation enterprise wide virtual IT department that will also take care of the infrastructure, such as networking that is not cooperatively managed by cells and organizations today. The global private cloud organization will benefit from the economy of scale and will be able to develop cloud-related unique capabilities, which can be translated into competitive advantage for the company as a whole.

Our next steps include formalization of the presented model and obtaining quantitative results in addition to the qualitative ones using data from the enterprise. One of the more important and challenging topics we will consider next are economic models for the private cloud at different maturity levels. We plan implementing the technical enablers of the proposed model on top of OpenStack [12], which is gaining popularity as an open source cloud stack for building massively scalable private and public clouds.

References

[1] David Mitchell Smith, "Gartner Hype Cycle for Cloud Computing, G00230930," Aug 2012.

[2] L. Columbus, "Roundup of Cloud Computing Forecasts and Market Estimates, 2012," http://softwarestrategiesblog.com/2012/01/17/roundup-of-cloud-computing-forecasts-and-market-estimates-2012/, Jan 2012.

[3] CIO, "CIO Global Cloud Computing Adoption Survey Results," http://download3.vmware.com/interactive_brochure/docs/CIO-Global-Cloud-Computing-Adoption-Survey-Results-PP-EN.pdf, Jan 2011.

[4] M. Healey, "InformationWeek 2012 State of Cloud Computing Survey ," reports.informationweek.com, Dec 2011.

[5] J. Feldman, "Cloud ROI Calculation," reports.informationweek.com, Mar 2012.

[6] Rochwerger, B., Breitgand, D., Epstein, A., Hadas, D., Loy, I., Nagin, K., Tordsson, J., Ragusa, C., Villari, M., Clayman, S., Levy, E., Maraschini, A., Massonet, P., Munoz, H., and Tofetti, G., "Reservoir – when one cloud is not enough," *IEEE Computer*, vol. 44, no. 3, pp. 44–51, Mar 2011.

[7] Sundareswaran, S., Squicciarini, A., and Lin, D., "A brokerage-based approach for cloud service selection," in *IEEE 5th Conference on Cloud Computing (CLOUD)*, Jun 2012, pp. 558–565.

[8] P. T. Endo, A. V. de Almeida Palhares, N. N. Pereira, G. E. Goncalves, D. Sadok, J. Kelner, and J. Melander, B.; Mangs, "Resource allocation for distributed cloud: concepts and research challenges," *IEEE Network*, vol. 25, no. 4, pp. 42–46, 2011.

[9] K. A. Beaty, V. K. Naik, and C.-S. Perng, "Economics of cloud computing for enterprise IT," *IBM Research and Development Journal*, vol. 55, no. 6, Dec 2011.

[10] D. Niyato, Z. Kun, and P. Wang, "Cooperative Virtual Machine Management for Multi-Organizational Cloud Computing Environment," in *VALUETOOLS'11*, Brussles, Belgium, 2011.

[11] B. Peleg and P. Sudholter, *Introduction to the Theory of Cooperative Games*. Springer, 2010.

[12] "Openstack: Open source software for building private and public clouds." www.openstack.org.