zimory
going beyond...

END-TO-END CLOUD

White Paper, Mar 2013

TABLE OF CONTENTS

ABSTRACT

The use by enterprises of infrastructure cloud computing is constrained by the gaps between complex and demanding requirements and the inadequacies of existing solutions. In this paper we present a solution for an end-to-end enterprise-grade cloud infrastructure, comprising a distributed and dynamic compute and data cloud interconnected by a high-performance, dynamic infrastructure network.

This combination enables businesses to create a distributed cloud with defined and assured performance characteristics, increasing the applicability and utility of cloud computing principles and services in the heart of the enterprise-grade IT infrastructure.

INTRODUCTION AND **PROBLEM DESCRIPTION**

Infrastructure cloud computing substantially has impacted the data center services market in recent years. Many large-scale web applications are enabled by the use of infrastructure cloud services. Early infrastructure cloud market participants like Amazon, with its Amazon Web Services (AWS) portfolio, targeted such consumer web applications. These cloud services have allowed application developers to reach new levels of agility and cost efficiency, as the applications could scale automatically with demand on the user side.

Naturally, enterprise IT departments are seeking to improve their agility and cost structures and thus seek to exploit cloud computing technologies and principles for their internal systems. But introducing cloud principles in the enterprise IT environment runs into a number of obstacles, including complex legacy application architectures, particular legal requirements governing various IT operational considerations, and strategic supplier dependency realities. To align with these requirements more complex cloud delivery models have been developed, such as distributed hybrid clouds and community clouds. These models introduce further levels of technical complexity and constraints that restrict the span of service utility from an application or 'use case' perspective, and reduce the net benefit of cloud to the enterprise.

In this paper we present a solution that shows how the integration of dynamic networking and dynamic computing can extend the utility and feasibility of cloud to include and encompass enterprise-grade distributed data center environments.

This paper is organized as follows. In the next section we discuss in more detail the challenges enterprises face when introducing cloud services into their IT environments. In section III we present the alternative setup scenarios for the solution, and in section IV we explain the solution in detail, describing an example application. We conclude the paper with a consideration of the market outlook for the solution.

ENTERPRISE CHALLENGES

Enterprises are faced with three main challenges when cloud computing services and technologies are being evaluated.

First, enterprises have many existing and interconnected applications that come with specific technical requirements. Moving all these into an infrastructure cloud does not necessarily make sense, as any single cloud cannot be set up to support all of the different detailed technologies. Whenever not all applications can be run in a single cloud out of the box, the enterprise has two options: either ‘decouple’ and localize the applications more, or invest development effort to adjust existing applications for operation in the cloud. Both approaches require a potentially risky development project, negatively impacting the cloud computing business case for the enterprise. A third option would be setting up or using a hybrid cloud, but this requires interconnecting the cloud segments with sufficient performance and quality guarantees. Nonetheless, in many respects this is the most promising option, and it can be realized with the solution stack we present here.

Second: the enterprise has legal requirements it must satisfy concerning information management. Often enterprises have global business structures and thus need to adjust to different local legal requirements. Many countries for example require that specific enterprise data needs to be stored and processed within the country. Requirements in terms of backup, data security and accessibility also differ among countries and applications. This factor, too, leads to a distributed application landscape that needs to be powered by a distributed cloud.

Third: enterprises are forced to use multi-supplier strategies. Especially larger organizations need to decouple their core systems from a single supplier, to avoid single sources of supply or performance failure and to reduce strategic dependencies. Enterprise applications must operate in a multi-vendor cloud environment.

All these requirements create the demand for a distributed and heterogeneous IT infrastructure environment. Combining these requirements with the cloud principles of a homogenous and multi-tenant operating environment, it becomes clear that enterprises will need to consider their situations carefully. In some scenarios a private cloud or even a dedicated environment will be the required solution, whereas in others a fully multi-tenant cloud can be the optimal solution. These different models and the available combinations will be reviewed in detail in the next section.

SETUP ALTERNATIVES

There are three known, defined variations of cloud setup: Private Cloud, Provider Cloud (also often called Public Cloud) and Hybrid Cloud. These various setups deliver different advantages and disadvantages. In this section the three models will be briefly presented and the pros and cons of each will be considered.

Private Cloud

A private cloud is a non-shared cloud that is implemented for a single customer. Often this cloud is implemented using capacity from pre-existing virtualized data centers. It normally consists of a compute pool, a storage pool, local network components and a self-service facility.

Table 1 Assessment of Private Cloud Model:

Pro	Contra
- No security risk - private structure - Update cycles can be defined by the customer - Customization is easily possible	- No pay-as-you-go pricing model possible - Cloud must be built to cover peak demand capacities

Provider Cloud

A provider cloud (or public cloud) is a fully shared infrastructure that is shared among higher numbers of customers. These infrastructures include technically the same components as the private cloud, but add network access infrastructures (firewalls or VPN Gateways) and all components must support managed multi-tenancy. This model mostly reverses the pros and cons of the private cloud model.

Table 2 Assessment of Public Cloud Model:

Pro	Contra
- Pay-as-you-go cost models work best - Platform development is supported by all users thus potentially a faster overall platform advancements	- Capacity shortage can be induced by other users - Potential security threat due to shared infrastructure - Trust involved - Cloud provider works towards lock-in

Hybrid Cloud

A hybrid cloud combines a private and a public cloud in a common resource pool. This model tries to combine the advantages of the provider and the private cloud. This carries great advantages – effectively summing the 'pros' and negating the 'cons' – but the hybrid model introduces significant further complexity and requires additional management control.

Table 3 Assessment of Hybrid Cloud Model:

Pro	Contra
- Workloads can be placed according to varying requirements in different parts of the hybrid cloud (purely private vs. purely public vs. 'mixed') - Peak demand dimensioning of the private cloud infrastructure is not required - No single supplier dependency	- Additional complexity due to multiple resource pools - Potentially uneven workload distribution between the pools

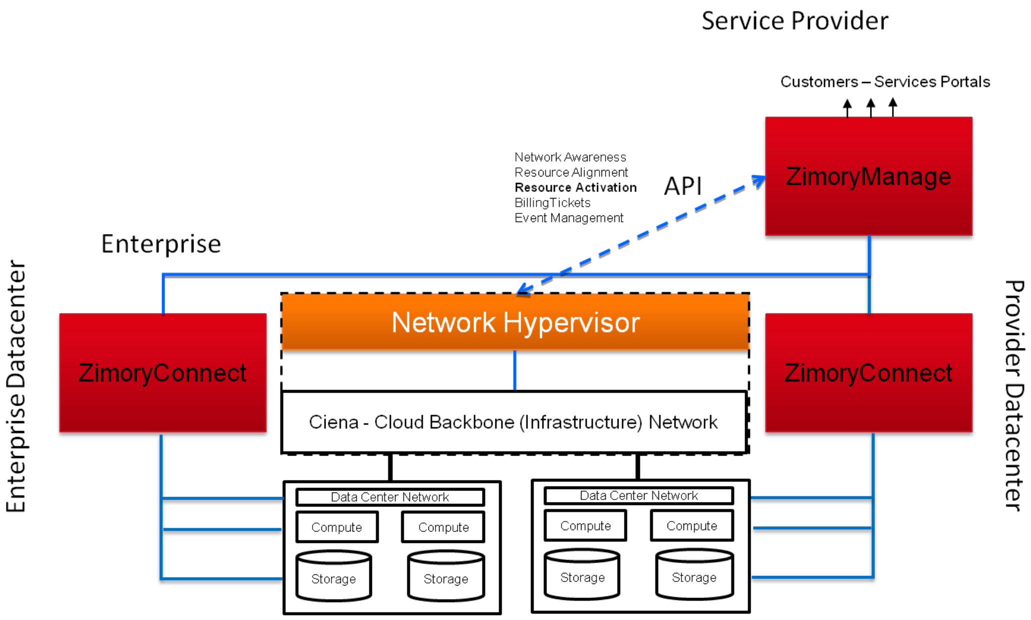
This overview shows that the hybrid cloud model generically suits enterprise needs best – provided the specific challenges it poses can be overcome. By adopting it, enterprise cloud customers effectively must come to manage three different types of resource pools: non-cloud, private cloud and public cloud. Of each of these types there can be multiple instances with different characteristics.

All this creates additional complexity - and a major success factor for such a distributed cloud environment is to be able to interconnect the pools and distribute workloads across the collective environment as efficiently as possible, with the fewest constraints and restrictions, with minimal performance impacts and encountering no real operational complexities.

SOLUTION DETAILS

The solution combines a dynamic computing and storage cloud with a dynamic, high-performance network component. These solution components are integrated and their operations orchestrated through APIs that interconnect them.

The general architecture is depicted below. In this simplified view, two independent cloud environments are connected to construct a hybrid cloud infrastructure. Whereas the resources in the left-hand side data center are private, the resources in the right-hand side data center are shared resources in a provider cloud.



The two main system components are depicted in red and in orange – we have reflected a specific vendor implementation from the perspective of named architectural components, but the principles are generic. The cloud computing stack allows the enterprise user to orchestrate workloads across the distributed resource pool and also to trigger scaling actions within the compute and storage clouds. The cloud computing stack also establishes billing metrics for the different cloud segments. For the 'umbrella' cloud administrator the distributed resource pool is driven through a single service console – specifically, here, the zimory®manage component.

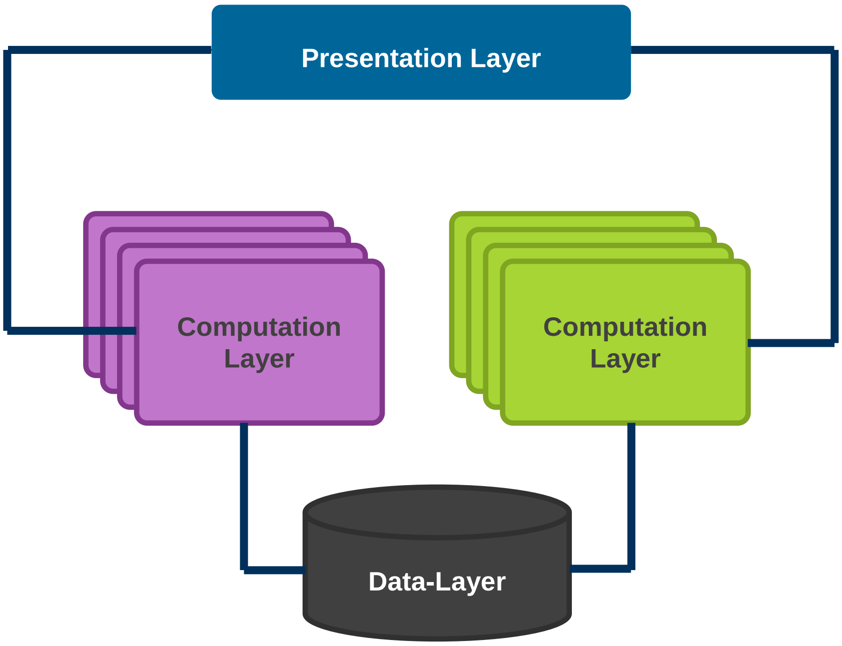
The different resource pools are interconnected through (here) a Ciena packet optical and Carrier Ethernet network infrastructure. This network provides high connection performance from latency and virtual circuit quality and availability perspectives, and delivers significant bandwidths to support mobility and related cloud operations. A core packet optical mesh provides a liquid 'pool of bandwidth' that is driven by the cloud computing stack. The interface between the network and the cloud stack is an abstracted, service-oriented API northbound from what amounts to a network hypervisor.

This hypervisor leverages control plane-driven network intelligence and also facilitates cost-based 'negotiation before provisioning' and pay-per-use bandwidth billing operations with the cloud computing stack.

Users can easily choose (or the cloud stack can choose on their behalf), in combination with a compute or storage workload, what bandwidth and other virtual connection characteristics should be available between the workload and some other application component at a point in time. Overall then, the user can choose end-to-end performance targets and the integrated cloud solution will provision as needed across the entire infrastructure.

Example Use Case

As an example use case let us consider a computational application with periodically very high compute power demands. The application consists of a data backend, a frontend presentation layer and different number of compute nodes – see the figure below. The network bandwidth relative requirements are suggested by the relative weights of the



connection lines. The data resides in the enterprise data center as does the presentation layer. In this example there are two different types of computation nodes illustrated, distinguished by various characteristics of the application kernels being run on each. When computing power requirements spike high, the distributed cloud would select – e.g., based on security requirements – the red nodes to remain within the enterprise's data center while the green compute nodes can be moved to or further instances instantiated within the service provider cloud. In parallel with the movement or creation of those 'remote' compute nodes, the cloud stack 'orders' new network virtual connections, or increases the size (bandwidth) of existing connections, between the enterprise and service provider data centers to facilitate effective 'reach back' to the datastore as well as to provide high-performance connections, if and as needed, among the compute virtual nodes in the two locations. This simple example shows nicely how the infrastructure can support running the application in the hybrid cloud. The workload distribution mechanisms of the cloud computing stack work hand-in-hand with the dynamic network capabilities of the 'cloud backbone' network. Without dynamic network provisioning, the distributed application would be not consistently and assuredly function at target performance levels, as the network requirements between the computational layer and the data layer could not be met generally and at reasonable total cost points. Without the distributed workload scheduling mechanisms of the cloud computing management stack, the application could not be placed in the different cloud segments simultaneously to satisfy technical/performance, legal and other administrative and other operational constraints and requirements.

CONCLUSION AND OUTLOOK

The hybrid cloud model must be made to work if enterprises are to tap significantly further into the promise and benefits of the infrastructure cloud. As we have seen, the hybrid cloud poses a number of technical challenges. We have shown in this paper how these challenges can be resolved. A cloud computing 'stack' – effectively a cloud 'OS' – must be able to create and manage the various components of a distributed data center architecture that crosses 'ownership' lines (enterprise vs. service provider) and also crosses geographical lines, with implications on information management practices.
A high-performance network inter-connecting the data centers within this distributed cloud infrastructure architecture is equally important: this network lies within the distributed computer itself, and every operational characteristic of that computer: performance, reliability, control of performance, etc., will be tied to and limited by the equivalent characteristics of the network. Tight latency control is essential if performance is to be maintained over significant inter-data center distances. And bandwidth on demand, matched to the on-demand characteristics of the cloud services themselves, is essential in terms of constraining costs while allowing for scalable high performance across a large base of users.

We have described here relatively simple scenarios from a network connectivity perspective: time-variable 'size' of inter-connections limited to data center pairs. However, over time, we expect many forces to drive the cloud infrastructure naturally to significantly more 'meshed' multi-data center architectures. For example, load balancing or follow sun/moon operations would shift active workload among provider data centers – driving new requirements for transient high bandwidth connections among various data centers. 'Proximity' constraints usefully would be relaxed with high-performance, low latency cloud backbone network implementations, increasing the permutations and combinations of provider data centers involved in specific hybrid cloud service instantiations and requiring scalable network inter-connections. Overall, the future and full promise of the cloud will depend on leveraging the ‘toolkit’ that we have described here.

CONTACT **INFORMATION**

Zimory GmbH
Alexanderstrasse 3,
10178 Berlin
Germany

Email: info@zimory.com
Tel: +49 (0)30 609 85 07-0

For the latest information, please visit www.zimory.com