Designing a Responsible Cloud Infrastructure

An ENTERPRISE MANAGEMENT ASSOCIATES® (EMA™) White Paper
Prepared for Avocent®

July 2010

IT & DATA MANAGEMENT RESEARCH,
INDUSTRY ANALYSIS & CONSULTING

Table of Contents

Executive Summary

Cloud computing implementations have the potential to drive significant improvements in operational efficiency and reductions in operational expenditures. In order to ensure reliability in a Cloud infrastructure, however, organizations need to adopt management practices and infrastructure designs consistent with what ENTERPRISE MANAGEMENT ASSOCIATES® (EMA™) analysts have referred to as a "Responsible Cloud" deployment that is secure, reliable, economical, and rapidly extensible. Enterprises can achieve a Responsible Cloud with the introduction of key practices and enabling management tools that aid in appropriately sizing the infrastructure, ensuring high-availability, and minimizing operating expenses.

Introduction to a Responsible Cloud Architecture

Fueled by increased desires for greater efficiencies, supportability, extensibility, and cost effectiveness in Information Technology (IT) infrastructures, Cloud computing has developed into an almost mythic status as a powerful consolidated architecture for delivery of enterprise IT services. Although commonly built on long-existing technologies – such as virtualization, grid computing, and clustering – the concepts behind Cloud implementations are still fairly young. Despite this newness, they are seeing relatively rapid adoption. Cloud computing services constitute somewhere between a $40 and $95 billion dollar industry, according to EMA research, with an expectation for it to quadruple to $160 billion by 2015, and yet it is clear we are still only on the cusp of Cloud computing proliferation. EMA survey results identify only 16% of organizations have as yet considered Cloud solutions for their IT services, indicating a wide potential for future growth as the value proposition becomes more widely accepted.

Cloud computing services constitute somewhere between a $40 and $95 billion dollar industry, according to EMA research, with an expectation for it to quadruple to $160 billion by 2015, and yet it is clear we are still only on the cusp of Cloud computing proliferation.

Part of the education process that must occur for broader adoption is a standardized definition of what constitutes a Cloud architecture. Cloud computing, as defined by the US National Institute of Standards and Technology (NIST), includes several key defining characteristics that should be used as a guide for ensuring relevance in an infrastructure deployment:

- Self-service – end users should be able to initiate and alter Cloud services without interacting with administrative or management personnel.
- Network accessible – any network will suffice to meet this criteria, so the Cloud service can be dedicated to local users and does not necessarily need to be accessible over the Internet or WAN.
- Resource pools – critical to the concept of a Cloud environment is the idea that infrastructure services are consolidated to provide a seemingly endless supply of computing resources. Naturally, it is impossible to maintain infinite disk space and processing power, but the end users should have the illusion that they can provision any resources they require. This illusion is established by connecting and load-balancing IT services in shared pools.
- Rapid elasticity – provisioning of Cloud services should be nearly instantaneous. In fact, Cloud implementations are able to provision services such as virtual desktops, storage, and applications up to 240 times faster than more convention methods.

- Measured service – NIST does not define what actually needs to be measured with this service, leaving it open to the providers to define the type of metering system most appropriate to their service delivery models.

Cloud implementations that meet these criteria can be offered publically over the internet, privately to a dedicated business, or as a community service to a group of businesses. Naturally, public Clouds are always offered off-premise from the end users, but private and community Clouds can be accessible either on-premise or off-premise from the users. Although public Cloud services have garnered the most publicity, private Clouds have seen by far the largest adoption – with 75% of all Cloud adoptions being a dedicated service of which 84% include on-premise deployments, according to EMA research.

When developing on-premise Cloud deployments, however, careful planning should be employed to achieve a Responsible Cloud – that is, a Cloud implementation that is well-managed to provide secure, compliant, and high-quality business services. Responsible Cloud environments deliver more secure, reliable, and flexible IT services to meet organizational requirements and reduce both capital and operational expenses. Lacking a responsible approach to Cloud development can result in a failure to meet expectations. In an EMA survey of businesses that have deployed a Cloud infrastructure, 28% actually saw an increase in operational expenses and 30% reported decreased flexibility, principally due to inadequate management practices. Ensuring value in a Cloud investment begins at deployment and continues through the entire lifecycle of the infrastructure.

Constructing a Responsible Cloud Infrastructure

Whether building a Responsible Cloud infrastructure from existing computing resources or an entirely new environment from the ground up, three areas of focus should dominate design considerations: effectively size the infrastructure, ensure high availability, and minimize operating expenses.

Whether building a Responsible Cloud infrastructure from existing computing resources or an entirely new environment from the ground up, three areas of focus should dominate design considerations: effectively size the infrastructure, ensure high availability, and minimize operating expenses.

The first of these focuses on supporting the maximum amount of computing services with the minimum number of computing resources. New environments will reduce capital expenditures by investing in a smaller number of more powerful resources, and established environments can more efficiently utilize existing systems through consolidation efforts or by maximizing system utilization with grid computing. The fewer physical components there are in an infrastructure, the fewer systems there are that must be deployed, configured, and maintained.

Additionally, the number of potential security and failure points is proportionally decreased, improving overall reliability. Consolidated servers are also more easily pooled and can better accommodate the use of shared services, such as with commonly accessible storage and clustering. This allows new resources to be more rapidly provisioned to meet service demands. For example, as requirements for Cloud services grow, new disk drives can easily be added to an existing storage pool and new servers can be easily introduced to a cluster without affecting service availability.

Capacity planning is critical to effective infrastructure sizing. Systems that consolidate a large number of resources, such as blade servers and mainframes, are physically larger and more powerful than standard servers. Organizations must ensure data centers do not exceed environmental conditions – such as weight restrictions, power availability, and network bandwidth – and that sufficient environmental resource are always available to support rapid expansion. Automated tools should be employed to track system and environmental resources availability so that quick and informed decisions can be made on the most effective ways to size and expand Cloud services to meet an expected growth in service requirements.

Since Cloud services are expected to be continuously accessible, high availability is another critical infrastructure requirement that needs to be addressed. Clustered servers have already been identified for their ability to be rapidly expanded, but their primary value comes from their ability to load balance system resources across cluster members and to provide uninterrupted fail-over services in the event an individual server experiences a catastrophic failure or requires downtime for maintenance.

Clustered environments are typically contained within a single physical location so that they can share storage systems and do not have any performance latency due to WAN traffic. Large Cloud implementations typically have multiple clustered environments at multiple facilities. This allows failover of a Cloud service in the event of a site disaster, such as a flood or fire. Individual Cloud instances can be expanded to operate across multiple clustered environments, both local and remote, to create a "hub and spoke" architecture that ensures highly available and highly reliable compute services. Even by eliminating the possibility of a single point of failure, however, catastrophic system or site failure can impact performance of Cloud services, so automated tools should be employed to monitor the health of these systems as well as the availability of support services, such as power and network connectivity.

A core expected value from Cloud adoption is the reduction of business IT expenses. In fact, EMA research has confirmed that roughly 76% of Cloud implementations have resulted in significant and measurable cost reductions. Consolidation efforts can certainly help to reduce capital expenditures, but operational cost reductions can be achieved with additional planning and consideration. Administrative staff, for instance, can be reduced and yet still sufficiently sized to meet support requirements by simplifying the environment to be managed.

EMA research has confirmed that roughly 76% of Cloud implementations have resulted in significant and measurable cost reductions.

Unlike traditional data center models, Cloud infrastructures provide the opportunity to standardize on only one or a few system architectures. The fewer server, configuration, and application types there are to support, the less specialized knowledge is needed to support staff pool. Similarly, standardizing management practices and automating procedures reduce the amount of administrative tasks and effort that must be performed. Monitoring and asset discovery tools can also reduce the administrative burden by making it easier and faster to identify root causes of problems and to proactively prevent potential problems.

Cost efficiencies can also be achieved by reducing data center energy consumption. EMA research indicates that, on average, more than 20% of data center electricity costs can be eliminated with basic power management practices, and more than 80% reduction in consumption has been achieved by the most energy-efficient data center reconfigurations. The most effective server level power management practices involve monitoring the environment to look for opportunities to reduce consumption during low use periods. Significant cost can also be achieved by reducing data center cooling requirements. By identifying "hot spots," data centers can be reconfigured to ensure proper airflow and heat dissipation so that chillers do not need to be run more often than necessary.

Key Resources in Cloud Infrastructure Management

As noted, automated tools are essential for the design and maintenance of a Responsible Cloud infrastructure. They enable quick identification of the environmental health and status to facilitate effective environment planning, problem resolution, and capacity expansion to rapidly adjust to changing support requirements. In-depth knowledge of data center assets and performance also enables accurate cost assessments that can ensure cost-effectiveness is achieved in the Cloud deployment.

Modeling solutions in particular can provide easily digestible infrastructure intelligence which is critical for enabling rapid cloud expansion to meet changing customer requirements. Data center asset and configuration information is collected into a centralized repository that provides a holistic view of the Cloud implementation. Not only does this provide a single point of access for viewing details about individual IT components, it also provides insight into how those components interact. Intelligent modeling solutions go further to allow organizations to generate hypothetical scenarios so that quick and informed decisions can be made on infrastructure growth and improvement.

For instance, before actually adding a new server to an existing cluster, a modeling solution can identify if there is sufficient rack space, power, networking, and structural support for the new server. If any discrepancies should be discovered, the theoretical environment in the modeling software can be rearranged to establish a more optimal configuration. In this way, environments are streamlined before any equipment is purchased or cables run. Since Responsible Cloud environments must rapidly adapt to support requirement changes without diminishing infrastructure reliability, a modeling solution can be an indispensable tool.

> IT management solution provider, Avocent, offers several tools that can achieve this level of infrastructure insight.

IT management solution provider, Avocent, offers several tools that can achieve this level of infrastructure insight. The Avocent MergePoint® Infrastructure Explorer software is a visual modeling tool that provides a graphical representation of an IT infrastructure, such as a data center. The display can zero in on views of specific IT components, such as servers, with details including available ports and cabling; specific racks to identify available space or power; or of the entire end-to-end architecture for a holistic layout of infrastructure-wide concerns such as heat dissipation and weight distribution. The MergePoint Infrastructure Explorer allows hypothetical modeling, so that new components can be visually represented in the environment and optimally configured before investing in hardware purchases and administrative staff time.

Avocent also offers DSView® 3 management software, which integrates directly into MergePoint Infrastructure Explorer to automate the asset discovery process for modeling and enable reliable and secure access to the managed end points. DSView 3 software is an IP-based KVM solution that provides a centralized console for accessing and managing all components in a Cloud infrastructure, whether local or remote. The solution connects administrators to both physical and virtual devices at the console level, so connectivity is maintained even when the servers are powered off or rebooting. This level of access and the ability to initiate scripts from a single interface to managed systems across the support stack simplifies administrative tasks. High-availability is also facilitated with a failover access authentication feature that ensures users are always connected to a fully operating environment. Additionally, DSView 3 software is integrated directly with power distribution units to monitor and meter actual power consumption at each individual rack, so planning can be made for energy reduction programs.

Responsible Cloud deployments are carefully planned for optimal effectiveness with minimal resources, and yet decisions on infrastructure changes often must be made in extremely short intervals of time. Success under these conditions demands prompt, detailed and accurate IT infrastructure intelligence, and holistic infrastructure management resources, such as the Avocent MergePoint Infrastructure Explorer and DSView 3 software, to enable a simplified platform from which prompt and informed decisions can be made.

EMA Perspective

Cloud computing is a relatively young paradigm in IT service delivery, and yet it is built on concepts and technologies that have been evolving for decades toward the goal of providing greater data center efficiencies. In fact, the very concept of a consolidated IT infrastructure that maximizes resource utilization is at the very heart of Cloud concepts, so it should be no surprise that EMA primary research clearly identifies service improvements and cost reductions as the most important drivers for Cloud adoption. Undoubtedly, the motivation to achieve this value in Cloud implementations will drive the development of even more comprehensive management tools than are currently available, but the foundation for end-to-end Cloud infrastructure intelligence is being established today.

> Visual modeling tools, such has Avocent's MergePoint Infrastructure Explorer, have the potential to revolutionize the concepts of IT management.

Visual modeling tools, such has the Avocent MergePoint Infrastructure Explorer, have the potential to revolutionize the concepts of IT management. By transitioning from a passive monitoring solution to an active management interface, the solution can be transformed into a centralized console for unified service management. Certainly, the Avocent DSView 3 software brings some of these capabilities to the MergePoint Infrastructure Explorer platform, but integration with more extensible automated systems management processes – such as for provisioning, configuration, security, patching, and remediation services – can provide a very intuitive virtual data center representation where individual services can be modeled at the process level as well as at the physical hardware level. We're not there yet, but for current Cloud infrastructure developers, adoption of tools for infrastructure intelligence today not only provides immediate benefits with improved efficiencies, it also gives their organizations a leg up on the technologies that will shape enterprise IT infrastructure management in the future.

About Avocent

Avocent (a wholly owned subsidiary of Emerson Electric Co.) delivers IT infrastructure management solutions that reduce operating costs for IT environments via integrated, centralized in-band and out-of-band hardware and software. Additional information is available at www.avocent.com.

About Enterprise Management Associates, Inc.

Founded in 1996, Enterprise Management Associates (EMA) is a leading industry analyst firm that specializes in going "beyond the surface" to provide deep insight across the full spectrum of IT management technologies. EMA analysts leverage a unique combination of practical experience, insight into industry best practices, and in-depth knowledge of current and planned vendor solutions to help its clients achieve their goals. Learn more about EMA research, analysis, and consulting services for enterprise IT professionals and IT vendors at www.enterprisemanagement.com or follow EMA on Twitter.

Corporate Headquarters:
5777 Central Avenue, Suite 105
Boulder, CO 80301
Phone: +1 303.543.9500
Fax: +1 303.543.7687
www.enterprisemanagement.com

2129.081810