WHITE PAPER

optimizing service levels in public cloud deployments

keys to effective service management

agility
made possible™

technologies

table of contents

executive summary 3

Introduction: Cloud deployment models 3

Private clouds

Public clouds

Hybrid clouds and other cloud models

Implications of the cloud for service level management 5

Support for highly dynamic environments

Focus on end user experience monitoring—the best measure of service

Seamless integration between multiple cloud and data center environments

Simplicity of deployment and use—zero-touch configuration and deployment of monitoring

Maximum coverage of technologies and application service delivery capabilities

Key requirements for monitoring instances in public clouds 6

Common requirements

SaaS monitoring requirements

Detailed monitoring of PaaS

Monitoring IaaS infrastructures

CA Nimsoft Monitor: Service level management for public cloud environments 8

Complete public cloud coverage

Unified monitoring coverage of the IT landscape

Automated monitoring of dynamic cloud environments

Conclusion 10

executive summary

When organizations migrate vital business services to the cloud, they confront an entirely new paradigm when it comes to monitoring and managing service levels. This paper looks at the unique service level management challenges cloud environments present, and it then offers detailed insights into the capabilities and approaches that will be required to ensure optimal service levels in public cloud deployments.

Introduction: Cloud deployment models

Migration to the cloud offers the potential for tremendous business benefits, but it also presents a host of potential risks. For example, once organizations migrate to the cloud, how can IT track and manage service levels to ensure critical business services continue to run optimally? This paper offers a detailed look at how organizations can address this challenge.

In this section, we start with some definitions of cloud models. With all that's being said and written about the cloud, it's important to start with some context as to the many deployment scenarios that fit under the "cloud" umbrella, as each can have different implications from a service level management perspective.

Private clouds

A private cloud is considered anything an organization deploys on its own premises, on its own infrastructure, and for its own use. Typically, private clouds are a highly virtualized infrastructure that an internal IT team deploys and "rents" out to various business groups. Another common deployment model within this category is the virtual private cloud. In this model, a service provider dedicates a set of resources—such as servers, storage, virtualization, or converged infrastructure—within their cloud environment to a single customer. Users within these customer environments connect to these dedicated resources using a secure connection, just as they would access resources that are part of their company's internal IT infrastructure.

Public clouds

These are services that customers access via the Internet. Most fundamentally, there are three types of public cloud offerings:

- **Software as a service (SaaS)**. These are complete service offerings delivered over the Internet. Representing the majority of cloud deployments to date, SaaS represents a culmination of the cloud's opportunity for businesses, where not only the infrastructure, but the development, deployment, and maintenance of the entire business application is outsourced. Salesforce, RightNow, and Microsoft, with its Live Office offering, are a few of the most recognized SaaS vendors.

- **Infrastructure as a service (IaaS)**. When signing up for IaaS services, organizations get computing power, storage, and other infrastructure services on demand. Sample service providers in this category include Amazon Web Services and Rackspace.
- **Platform as a service (PaaS)**. Like IaaS, PaaS offerings enable organizations to outsource the hosting and administration of the entire infrastructure—all the underlying network devices, hardware, operating systems, and other elements that comprise the computing infrastructure. Unlike IaaS, PaaS also delivers an additional layer of intelligence that facilitates the development and deployment of applications. Two of the leading providers in this group include Google App Engine and Microsoft Azure. In addition, Force.com, an application development platform from Salesforce, is another prominent example.

"Representing the majority of cloud deployments to date, SaaS represents a culmination of the cloud's opportunity for businesses, where not only the infrastructure, but the development, deployment, and maintenance of the entire business application is outsourced."

Hybrid clouds and other cloud models

As the adoption of the cloud continues, the concepts outlined above have been merged in many different ways. Today, there are a host of other approaches, some more established, others recently emerging. Following is an overview of some of these options:

- **Hybrid deployments**. Cloud customers and providers can implement hybrid cloud approaches, which can span public and private clouds, or the traditional data center and the public cloud. An enterprise business infrastructure may be comprised of both public and private clouds. For example an organization may have its Web servers and application servers hosted at two different public cloud vendors, while maintaining and hosting its database servers in the internal data center.
- **Cloud bursting**. An approach known as cloud bursting is one example of a hybrid approach. Here, organizations rely on their internal data center to accommodate a base level demand, and when spikes in demand occur, they migrate processing to an outsourced cloud. Examples of cloud bursting may include a retailer that uses an outsourced IaaS vendor to accommodate the spike in online transactions that occur during the holiday season. An online media company can employ cloud bursting on a daily basis to accommodate increased streaming requirements during peak viewing hours.
- **Hybrid cloud offerings**. In other cases, cloud service providers are merging approaches. For example, IaaS providers may offer an enterprise customer a complete, dedicated computing "stack", an entire computing infrastructure that effectively functions as a private cloud, while still being completely outsourced. On the other hand, some PaaS vendor offerings are now available as appliances that an enterprise customer can deploy in their data center.
- **Community clouds**. A community cloud can be viewed as one blending public and private cloud approaches, where several organizations with similar requirements can sign on for a given cloud service. While in effect these are multi-tenant environments, they are tailored, and perhaps restricted, to a specific type of organization. Google Apps for Government, a cloud offering that adheres to government security mandates, is one example of this type of offering.

Implications of the cloud for service level management

When organizations migrate to the cloud, service levels remain just as vital as they've ever been. However, there are several implications for organizations looking to gain the visibility they need to track and manage service levels.

Support for highly dynamic environments

The advent of cloud computing will create a significant challenge to traditional service level management solutions, which have relied on knowledge of the infrastructure components and their relationships to perform their functions correctly. This applies particularly to centralized, server-based solutions that rely on a schedule of polling activities to monitor the underlying infrastructure. Given that there is no way to predict which cloud computing infrastructure components should be accessible at any point in time, the only reliable mechanism for monitoring cloud components for performance data will be through on-board agents that gracefully register with distributed collection points and, more importantly, gracefully de-register when the cloud computing components close themselves down. This mechanism will be particularly vital in IaaS environments and will require secure authenticated communication between agents and collection points, all configured in a highly resilient topology.

Focus on end user experience monitoring—the best measure of service

Cloud deployments create an urgent need to monitor service performance and availability in a more direct manner that more accurately mirrors how real users access their applications. The cloud computing model also means that there is no single point in the corporate network from which to monitor these applications. Solutions for the cloud will need to be simple to deploy across multiple monitoring points, including beyond the bounds of the corporate offices. For example, an organization using SaaS may want to conduct synthetic transactions from various endpoints outside of corporate offices that measure transaction times. Organizations may opt to measure round trip response times of actual user transactions from PaaS- or IaaS-based services. The ubiquitous consumer to ubiquitous provider computing model poses serious challenges to monolithic, high-touch service level management solutions. Highly distributable, light weight monitoring components will be required to get a high level and comprehensive view of the end user experience.

Seamless integration between multiple cloud and data center environments

The use of cloud services will not be a singular activity. Organizations are likely to use multiple deployment models, including the traditional data center, public clouds, private clouds, and hybrid approaches. Consequently, IT organizations must be able to monitor elements of their services, even if they span multiple environments. For example, if an organization has a multi-tier Web site that feeds data to Salesforce CRM, administrators may need to monitor the following areas:

- The IaaS cloud hosting the Web site.
- The operation of all the Web site elements housed in the IaaS cloud, including Web servers, application servers, and database servers.

- The operation of the Web site as a complete application.
- End user response times for transactions conducted on the Web site.
- Latencies and availability metrics for the Salesforce integration.
- The performance of the connection between the Web site and the back-end mainframe in the data center.

The ability to monitor all of these disparate areas and bring all the data into a normalized monitoring database will be a challenge for the majority of legacy service level management solutions, as they were built based on the assumption of a single data center with a secure and reliable network connecting all the components back to a monolithic systems management server. However, given any service may be based on multiple cloud and data center environments, the cloud will require that administrators can model a service and its underlying applications and infrastructure in a highly dynamic way. Key to this will be the normalization of performance data in a logically consistent form and the ability to build highly dynamic service models on top of this data.

Simplicity of deployment and use—zero-touch configuration and deployment of monitoring

Traditionally, there was an accepted process cost associated with adding monitoring to the IT infrastructure. In the cloud computing paradigm, that accepted cost will no longer be sustainable. Systems management solutions will need to match the dynamic, on-demand model of the cloud. These solutions will need to be "zero-touch"—that is, when an infrastructure component is instantiated, no human intervention should be required for that component to also become part of the monitored system. Toward that end, the monitoring solution will have to be embedded inside the cloud components and activate and register itself at the time of component instantiation.

Maximum coverage of technologies and application service delivery capabilities

To be viable, a service management solution must adapt to the "unknown unknowns" of the cloud's evolution. Many solutions have struggled to keep up with the pace of technological change during the virtualization phase of the last few years. The pace of change is not going to slow and the surviving systems management solutions of the future will be those that have been built with adaptive capabilities in mind. Strong application programming interfaces that allow rapid development of new technology monitoring will be key to success in the cloud computing world.

Key requirements for monitoring instances in public clouds

As organizations move to expand the use of cloud offerings, increasingly critical business services rely on the performance and availability of cloud-based infrastructures. Consequently, service level monitoring in the cloud will be an increasingly vital mandate. Note, given the evolution and variants outlined in the introduction, it's clear distinctions can be blurry, however, for the purposes of this section, the focus will be on requirements for monitoring services based in public cloud environments.

Common requirements

Following are some of the requirements of monitoring any public cloud service:

- **Understand global service data**. Administrators and business managers need to have visibility into the performance and availability of the cloud service as a whole. This includes basic up/down status, performance by location (if the service is hosted in multiple locations), as well as performance of specific offerings within the cloud environment. For example, this could include specific modules a company uses within a SaaS framework.
- **Services usage**. Businesses need to have visibility into their actual usage of a cloud service. This can include such metrics as the number of users during a given time period, usage metrics, and more.
- **Monitoring access**. In order to ensure adequate service levels are maintained, organizations need to have visibility into the specific public cloud instances they rely on. This includes the location of the instances, and detailed performance information about each instance. While various cloud providers offer some dashboards, these tend to provide only basic monitoring, for example, indicating whether a service is up or down. To get more detailed and useful insights, businesses typically need to use a cloud provider's APIs to get more detailed monitoring information.

SaaS monitoring requirements

When it comes to monitoring SaaS-based applications, administrators need insights on the following metric types:

- **Average transaction speeds and latencies**. Administrators should be able to track specific speeds and latency figures for a host of common interaction types.
- **Status of SaaS-based resources**. This includes storage, data, files, concurrent users, login speeds, query times, and response times.
- **End user experience and Web services response**. This should include capabilities for tracking end user response times from various geographic locations.
- **Compliance with SLAs and SLOs**. Organizations need to be able to gain definitive measures of their SaaS provider's compliance with service level agreements (SLAs) and service level objectives (SLOs). This includes capabilities for historical reporting as well as real-time alerts that enable administrators to be notified of a breach or if metrics are trending towards a breach.

Detailed monitoring of PaaS

When a business employs the services of a PaaS vendor, it requires specific monitoring capabilities in the following areas:

- **Status of dedicated instances**. This includes the up/down status, storage utilization, and task queues for individual instances.
- **Subscription status**. Businesses need to be able to monitor the actual status of their subscriptions. This includes comparing actual users versus the number contracted, as well as usage patterns over various time periods.
- **End user experience**. Organizations need to be able to track the responsiveness that end users experience. This includes monitoring transactions from various locations, and measuring the times of

multiple common transaction types. This also involves specific measures of the page load times of specific URLs and response times of specific Web services.

- **Compliance with SLAs and SLOs**. Organizations need to be able to gain definitive measures of their PaaS provider's compliance with service level agreements (SLAs) and service level objectives (SLOs). This includes capabilities for historical reporting as well as real-time alerts that enable administrators to be notified of a breach or if metrics are trending towards a breach.

Monitoring IaaS infrastructures

When migrating to IaaS, businesses need to ensure the vital business services that rely on the IaaS structure ultimately deliver the uptime and availability required. Toward that end, businesses must have concrete, reliable measures that support the setting and tracking of service level agreements. Following are some of the specific metrics that must be available through the IaaS provider's cloud APIs:

- Virtual server instances—including detailed metrics on network performance, CPU utilization, storage utilization details, read/write response times, and more.
- Global IaaS offering performance—including subscription data, server start up times, and the availability of servers and instances, instance groups, and instance types by location.
- Further, administrative teams will need the same levels of monitoring coverage they rely on for their internally hosted data center, which includes coverage of the following areas:
  - Applications, including Exchange, Apache Tomcat, Notes, SharePoint, Active Directory, IIS, databases, and more.
  - Web environments, including those based on WebSphere, WebLogic, and more.
  - Multi-tier Web application views.
  - End user experience and transaction times.

In addition, just as in the data center, to realize maximum value from monitoring investments, the monitoring data drawn from IaaS environments needs to be integrated with existing workflows, service desk solutions, usage metering, configuration management databases, and more.

CA Nimsoft Monitor: Service level management for public cloud environments

With CA Nimsoft Monitor, IT teams can monitor and manage business applications, from the data center to the cloud, including SaaS, hosted, and virtualized environments—all with a single product, architecture, and console. With its combination of comprehensive coverage, ease of use, and scalability, CA Nimsoft Monitor enables organizations to leverage existing and emerging technologies and services, with improved agility and ROI. CA Nimsoft Monitor offers open APIs, a flexible architecture, and out-of-the-box third-party integrations and gateways—making it easy for organizations to adapt the solution to their other management tools, monitoring processes, and operational preferences.

Following is an overview of the comprehensive capabilities that set CA Nimsoft Monitor apart.

Complete public cloud coverage

CA Nimsoft Monitor collects detailed metrics and quality of service information for cloud health and for specific instances within a range of public cloud environments:

- SaaS, including accounts on Salesforce.com and Google Apps.
- IaaS, including servers and storage on Amazon Elastic Compute Cloud (EC2), Amazon Simple Storage Services (S3), and Rackspace.
- PaaS, including applications built on Microsoft Azure and Google App Engine.

In addition, CA Nimsoft Monitor offers complete coverage of the critical applications installed in IaaS environments—including databases, email, Web environments, services (including FTP, HTTP, DNS, and DHCP), end user experience, servers, and custom applications.

Unified monitoring coverage of the IT landscape

CA Nimsoft Monitor represents a single platform that enables centralized, cohesive monitoring of all IT systems and services—across private clouds, data centers, and public cloud environments. With one unified solution and architecture, organizations can gain insights into the performance and availability of IT environments, including physical and virtual infrastructure, applications, power usage and data center environments, transaction times, end user response, servers, VoIP, networks, and more. In addition, CA Nimsoft Monitor can track private clouds, including converged infrastructure packages like Vblock and FlexPod. As a result, with CA Nimsoft Monitor, administrators can effectively monitor any environment that supports a business service.

Automated monitoring of dynamic cloud environments

CA Nimsoft Monitor offers the virtual capabilities required for effectively and efficiently monitoring dynamic cloud-based environments. It enables administrators to configure templates that automate configuration and deployment of monitoring agents when new resources come on line, and the graceful de-registration of agents when resources are taken off line. Further, the solution automates alerts and the display and updating of reports and dashboards, ensuring administrators get the information they need, when they need it.

Conclusion

Service level management in cloud environments is vitally important, and poses a host of significant challenges. CA is a vendor that has the products and expertise that can help deliver effective service level management that today's cloud computing environments demand. CA Nimsoft products offer the automation and comprehensive coverage, both of cloud environments and the entire IT infrastructure, that are critical requirements for organizations looking to sustain optimal service levels in their cloud-based services.

For more information, visit ca.com/nimsoft.