



White Paper

Cloud Computing Needs Availability@Performance™

99.99% availability for cloud
architectures is not enough.

Emulex OneCommand™ Vision is.

<http://www.emulex.com/vision/>

Table of contents

Executive Summary 3

Traditional Cloud Availability 3

Availability@Performance 4

Where is A@P Beneficial? 4

How is A@P Achieved? 5

 A@P Process..... 5

 A@P SLA Guidelines 6

I/O Management Tools..... 6

 Emulex OneCommand Vision 6

 A@P SLA Examples..... 7

 IaaS 7

 PaaS..... 7

 Storage Infrastructure 8

 Application I/O Perspective..... 8

Migrating Applications to the Cloud..... 9

Conclusion..... 10



Executive Summary

Cloud providers focused on delivering “4 Nines” quickly realize that availability alone is not enough to meet service level agreements (SLAs) of enterprise customers. High-end applications require Availability@Performance (A@P) - that is, they require measures of resource availability as well as specific I/O performance metrics. This white paper explores the method for setting, achieving and maintaining A@P SLAs for cloud architectures using the award-winning Emulex OneCommand Vision, an I/O management application.

Traditional Cloud Availability

Cloud providers offer SLAs for “uptime” and “availability”. While these availability SLAs meet the needs of many cloud-compatible applications, as enterprises move high-end applications into the cloud, including private clouds, availability requirements alone are not enough.

Traditional public cloud availability delivers “Triple Nines,” or 99.9% availability, which equates to nearly nine hours of downtime per year. Businesses and operations, such as 24x7 online retailers, e-mail applications and providers and CRM systems, cannot withstand that level of downtime. They require 99.99% uptime (“4-Nines”) -- 53 minutes of downtime per year -- or better.

Most cloud providers offer sub 4-Nines guarantees, but none offer performance guarantees based on measurable performance metrics, such as:

- **I/O Load:** The amount of load required for a particular application and its I/O infrastructure, defining a class-of-service for I/O availability and performance
- **I/O Performance:** The application's view of the I/O response time, at varying levels of service requirements
- **I/O Availability:** An end-to-end view of the redundancy built in between the application and its I/O resources

For high-end applications, availability is not enough. Enterprise customers need to define a set of application-specific performance metrics that govern performance and enforce them.

Consider an online retailer with 4-Nines availability built into their SLAs. Even when system availability requirements are met, if the performance is unacceptable just 3% of the time, this could translate to the equivalent of 11 *days* per year. Depending on which applications are affected, this level of resource downtime can have negative revenue impacts and also raise expenditures. Customers impacted by performance issues are left feeling frustrated, alienated and angry.

Cloud providers need to deliver availability and performance - Availability@Performance.



Availability@Performance

Cloud providers setting and meeting SLAs need to focus on the percent of time a system is up *and* performing at an acceptable level. A@P goes beyond traditional availability SLAs to consider actual application performance requirements. Defining A@P SLAs is a matter of defining availability metrics, measures of network, system and application uptime, as well as metrics designed to measure performance, such as IOPS, latency and throughput.

Cloud providers and enterprise data centers need improved tools aimed at proactively monitoring A@P SLAs. In all cloud offerings, I/O performance is best monitored close to the application. Emulex OneCommand Vision, an I/O management application, leverages Emulex's 15-year history of server-side I/O management. This includes the high-performance host bus adapter (HBA) and converged network adapter (CNA) technologies, as well as the broad range of operating system (OS) and server platforms supported. With its visibility into each I/O path through the I/O network, OneCommand Vision has an application view of the performance metrics essential for managing A@P SLAs.

This paper takes a closer look at A@P for cloud providers and enterprise data centers, with the following key sections:

1. Where is A@P beneficial?
2. How is A@P achieved?
3. I/O management tools
4. Migrating applications to the cloud

Where is A@P Beneficial?

Today, many cloud providers deliver traditional SLAs focused on availability only. As more enterprise customers migrate into that market, the need for A@P increases. Here are some examples of how the following service providers can benefit and profit from well-defined A@P SLAs:

Infrastructure as a Service (IaaS) leases infrastructure to customers at a negotiated usage rate for compute, networking and storage. Customers use this Infrastructure as a platform to host their own custom OS and application operation. Clear availability and performance requirements enable providers to innovate around how they can best raise utilization, lower capital expenditures and meet customer expectations.

Platform as a Service (PaaS) leases application environments to customers, typically based on virtual server architectures. These PaaS deployments raise the stakes on A@P as providers push to increase the utilization of shared resources, particularly storage resources. Increased



sharing can result in decreases in performance, especially when the performance is not actively monitored.

Software as a Service (SaaS) leases entire applications to customers, actually sharing one instance of an application among several paying customers. In these environments, it is critical for the service provider to ensure that the application is providing very good performance, lest the customers consider competitive solutions.

Storage as a Service is an emerging arrangement whereby data centers are outsourcing their on-premise storage infrastructure to a third party. In these cases, it is very important that specific A@P SLAs are defined to support the application's performance requirements. Also important is the need to intelligently monitor the performance of the outsourced storage cloud from the application's (host) perspective.

Defining and monitoring appropriate A@P SLAs provides important benefits for most cloud service providers. The challenge is defining the correct mix of I/O availability and performance required for each deployment or application. For vendors who take the time to intelligently define A@P SLAs with their customers, the rewards include additional service revenues, lower costs and increased levels of customer satisfaction.

How is A@P Achieved?

Existing and emerging cloud architectures leverage new virtualization technologies to increase capital and operational efficiencies. As more virtualization layers are inserted between the application and its resources (e.g., network, storage), it becomes increasingly important to understand the application's view of both performance and resource availability. A@P addresses this requirement.

A@P Process

Well-defined A@P SLAs are typically produced when cloud providers and service customers understand and observe the following best practices:

1. Manage I/O availability and performance at the origin of the I/O path (the server)
2. Establish application performance baselines and use them as input into SLA definitions
3. Enforce SLAs by deploying active, intelligent I/O monitoring solutions
4. Analyze performance trends to reveal small issues before they become big problems
5. Apply forensics to assist problem resolution
6. Enlist expert I/O management support as needed



A key piece of achieving A@P involves the last two steps. I/O monitoring is not enough. Analytics and forensics are required to detect negative availability and performance trends that help to proactively and efficiently address issues.

A@P SLA Guidelines

Cloud providers and customers can use the following guidelines for setting SLAs:

Infrastructure as a Service (IaaS)	Customers should baseline the A@P requirements for existing in-house applications before migration to cloud infrastructure services. Baseline metrics should be passed on to providers to be incorporated into service agreements.
Platform as a Service (PaaS)	Baseline application performance metrics, monitoring A@P SLAs at the provider's platform layer.
Software as a Service (SaaS)	Redefine A@P SLAs using available software performance metrics, such as transactions per second or response time.
Storage as a Service (Internal or Cloud)	Baseline and measure storage A@P SLAs at the origin of the I/O traffic, typically the server (or virtual machine). Monitor health of I/O path as well as historical performance of storage services.

I/O Management Tools

I/O management tools are critical to cloud providers and their customers. Clouds have to do two things very well. They have to economically scale-up resource performance at times of peak demand and dynamically scale-down resources when they are no longer needed.

Scaling up the cloud infrastructure includes deploying intelligent I/O management tools to ensure that additional resource loads are being managed within the A@P requirements. To accomplish this economically, cloud providers need to understand the minimum infrastructure requirements to meet the A@P requirements. Scaling down the cloud is similar in requiring the need to power down or de-commission infrastructure, while still meeting performance and availability requirements with the reduced resource set.

Emulex OneCommand Vision

OneCommand Vision is an enterprise I/O management application that assists cloud administrators in defining and monitoring I/O availability and performance metrics. Combined with I/O industry best practices, administrators can achieve the following:



1. **Development:** Deploying OneCommand Vision, cloud providers can baseline application performance and use that baseline to create A@P SLAs that guarantee applications, and thereby business operations, have the performance required to be successful. It is also beneficial to develop class-of-service criteria such that SLAs can be applied to the appropriate class-of-service.
2. **Monitoring and Enforcement:** OneCommand Vision monitors performance metrics essential to ensuring that A@P SLAs are met. OneCommand Vision sends proactive alerts when either an SLA or the agreed-upon class-of-service is violated.
3. **Application I/O Chargeback:** I/O is expensive. OneCommand Vision extracts the cost of the I/O and provides chargeback mechanisms to the appropriate line of business.
4. **Dynamic I/O Balancing:** OneCommand Vision detects I/O balancing issues (overloads and underloads) and feeds that information into management frameworks, allowing for real-time control of infrastructure balancing.
5. **Scale-Up Performance Validation:** As cloud infrastructures scale up, increased resource sharing begins to threaten performance and availability. OneCommand Vision helps plan and validate scale-up operations.
6. **Scale-Down Utilization Validation:** Scaling down is easy, if one doesn't care about costs. The challenge with scale-down is maintaining high utilization (i.e., use less infrastructure), while meeting new and existing I/O performance and availability requirements.
7. **Problem Detection and Forensics:** Despite all the best practices, sometimes things just go wrong. OneCommand Vision assists with both proactive and reactive problem resolution.

A@P SLA Examples

Setting A@P SLA metrics requires application performance baselining, which can be accomplished using OneCommand Vision, as described above. With a performance baseline, the next step is set the A@P SLAs. Below are a few examples of A@P SLA metrics for specific cloud service providers:

IaaS

- Resource Availability: 5-Nines (99.999%)
- Resource Performance: Up to eight cores of dedicated compute, with network throughput of no less than 10 Gbps and average I/O latency of no higher than 30 milliseconds (ms) for workloads up to 3K IOPS

PaaS

- Resource Availability: 4-Nines (99.99%)
- Resource Performance: Up to four CPU cores of compute at up to 80 percent sustained, 95 percent bursts, with average I/O latency of no higher than 60 ms for workloads up to 4K



IOPS for shared storage and average I/O latency of less than 20 ms for workloads up to 2K IOPS for dedicated storage

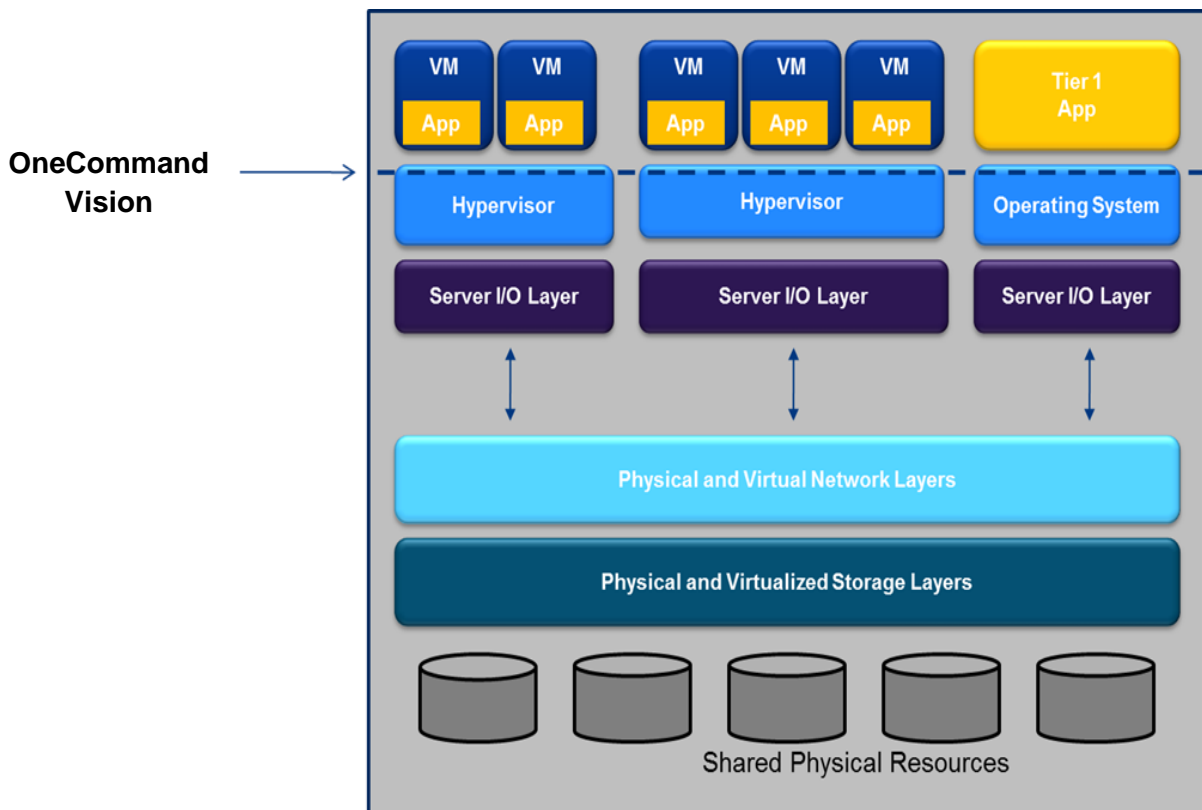
Storage Infrastructure

- Resource Availability: 4-Nines (99.99%)
- Resource Performance: Average I/O latency of no higher than 20 ms for workloads up to 1K IOPS

Application I/O Perspective

OneCommand Vision is the first I/O management tool that proactively monitors I/O from the application's perspective. Figure 1 outlines how OneCommand Vision measures I/O performance from the origination of the I/O request, just below the application. This approach also helps when monitoring I/O availability and redundancy, as nothing is better than monitoring the performance from the perspective of the application.

Figure 1. OneCommand Vision monitors I/O availability and performance from the application's perspective



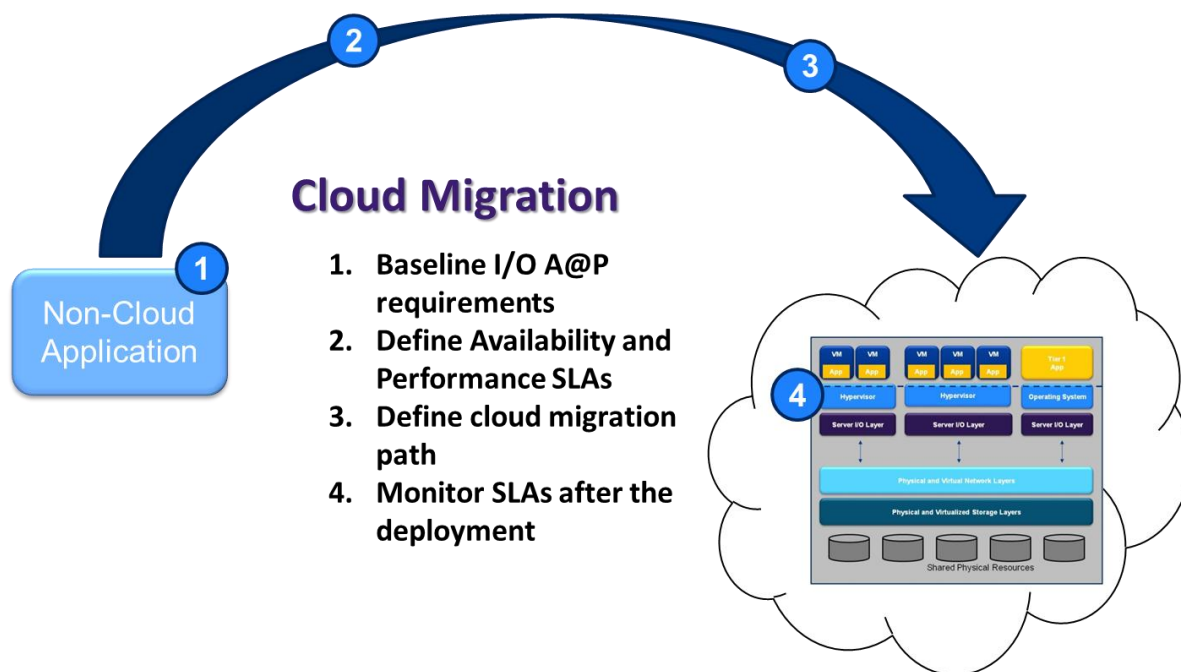


Migrating Applications to the Cloud

Many applications currently operating in a non-cloud environment can benefit from being migrated to a cloud infrastructure. However, to maximize the benefits and avoid application downtime and brownouts, administrators need to consider carefully what level of service and performance they require for each application. Capturing an accurate baseline of the key performance experienced today in the non-cloud environment is a main starting point in the planning process.

As shown in Figure 2, migration of a non-cloud application requires a robust performance and availability baseline that considers several critical performance factors. This baselining allows for a real-world measure of current performance for each server to establish A@P SLAs required for success in the cloud. The SLA requirements will then define the cloud migration potential for the application. Once an application is moved into the cloud, the SLAs need to be monitored to ensure that I/O availability and performance requirements are being met.

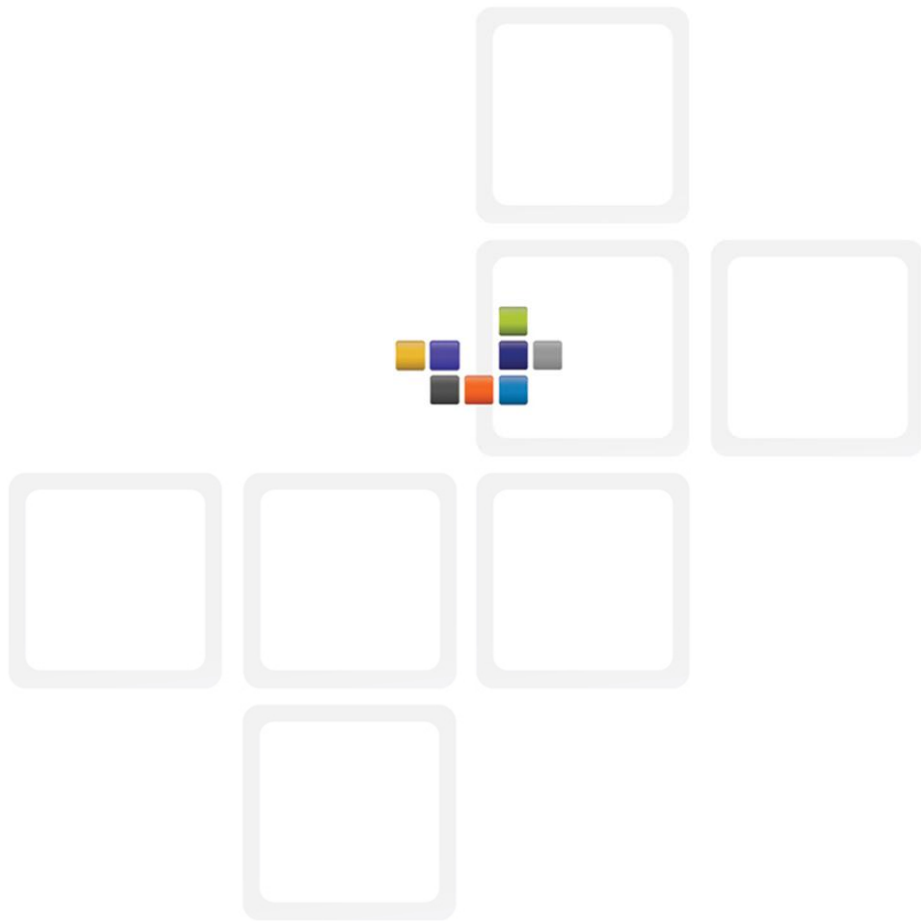
Figure 2: Application migration to the cloud





Conclusion

Availability@Performance SLAs focus not only on traditional availability metrics, but also on performance, creating new A@P SLA requirements for cloud service providers and customers. To set and achieve them, improved tools aimed at proactively monitoring A@P SLAs are needed. OneCommand Vision is a unique I/O management tool that collects, analyzes and delivers valuable I/O information to fully understand what A@P SLAs make sense for each cloud offering and then proactively monitor them. By migrating applications to the cloud, not only are cloud infrastructure benefits achieved, but more accurate A@P SLAs are set that address the availability *and* performance levels required for each application. Existing and emerging cloud architectures can all benefit from A@P with OneCommand Vision.



www.emulex.com

World Headquarters 3333 Susan Street, Costa Mesa, California 92626 +1 714 662 5600
Bangalore, India +91 80 40156789 | **Beijing, China** +86 10 68499547
Dublin, Ireland+35 3 (0)1 652 1700 | **Munich, Germany** +49 (0) 89 97007 177
Paris, France +33 (0) 158 580 022 | **Tokyo, Japan** +81 3 5322 1348
Wokingham, United Kingdom +44 (0) 118 977 2929

© Copyright 2011 Emulex Corporation. The information contained herein is subject to change without notice. The only warranties for Emulex products and services are set forth in the express warranty statements accompanying such products and services. Emulex shall not be liable for technical or editorial errors or omissions contained herein.

12-0128 August, 2011