

STORAGE SWITZERLAND

Cloud Environments have Familiar Storage Challenges - just More of Them



Eric Slack, Senior Analyst

“Cloud” can have many definitions and how it is defined depends in large part on the use case. In the enterprise the goal of a cloud is to provide dynamic IT services that improve IT responsiveness by leveraging server, storage and potentially I/O virtualization. It’s a ‘shared everything’ environment that includes shared storage, using either a SAN or NAS. An enterprise private cloud can also have a multiple-tenant nature to it, similar to a public cloud, but instead of separating different businesses it separates different departments. Cloud environments have many storage challenges that are familiar to data center managers; but due to their density and abstraction, may just have more of them.

More Activity on Less Infrastructure

Whether it’s called a cloud or just a large, virtualized IT infrastructure, this environment brings with it a number of challenges, and risks. Compared with a ‘traditional’ data center, cloud environments frequently have more users, more applications, more data and typically, a tighter set of operating standards to meet. But thanks to virtualized servers, storage and networking, they’re usually running this increased level of activity on less hardware. This abstraction makes it more difficult to address performance

problems and troubleshoot root causes to issues that can bring applications down. And the density increases the impact of a failure, as more users and applications are affected by these performance problems and downtime.

Consolidation rarely made anything simpler, it can add complexity and single points of failure, compared with a legacy, distributed infrastructure. But like was the case with the adoption of fibre channel SANs to replace direct attached storage, it’s a risk worth taking if the right tools are used and right steps are employed. Consolidated storage infrastructures brought flexibility, simplified management and economies of scale to large corporate data centers, benefits that would have been all but impossible with a direct attached storage architecture.

Virtualization brings similar benefits to the cloud infrastructure with the ability to balance workloads and reduce management points to fewer physical servers, storage and networking devices. But this consolidation also brings more risk, again, more activity on less infrastructure. What’s needed is comprehensive, real-time information about the network, storage and compute systems, and how they interact so that IT can troubleshoot problems efficiently and resolve issues fast.

Resource optimization is also important in a cloud environment. Storage and switch utilization in the 10% range, not uncommon in the pre-virtualized data center and in many SAN environments, becomes even more expensive in a private cloud infrastructure. Unfortunately, the complexity and abstraction that accompanies virtualized resource pools can make it more difficult to identify the causes of storage inefficiency, without better tools.

Clear the Fog

Managing the cloud requires a true understanding of the environment instead of vague rules of thumb to allocate resources and accommodate growth. What's needed is to clear out the 'fog' that clouds can create in an abstracted, complex SAN and storage infrastructure and improve the accuracy of the data that's used to manage and maintain that infrastructure. Otherwise, asset utilization and efficiency won't reach the levels required to make the economics of the cloud work. Or worse, a problem could take a critical application down and the troubleshooting exercise that ensues could take hours, days, or longer to fix it.

Trial and Error too Slow

Typically, when an issue is reported, the user doesn't know where the problem is, it could be in the network, in the storage infrastructure or in the server hosting the virtual machine that's running their application. In the less virtualized physical environment, typically connected to FC SANs, the system administrator may know the cause of the problem or at least where to look. Thanks to the abstraction of the cloud environment, they, like the users, are unsure where to begin. There are tools available from most of the device manufacturers, like SRM software from the array vendor or fabric monitoring programs for the switches and HBAs. But these 'element' tools require that

you know the general location of the problem, which is often not the case. These tools also don't provide a comprehensive, cross domain view of the environment, and troubleshooting with them can be ineffective, like hearing only half of a phone conversation. The result can be extended periods of trial and error, hunting for the root cause, or creating routines to track specific links and waiting for the problem to reoccur. In a cloud environment, where more data is at risk and more users can be affected, this may not be good enough.

Need Real-time Information

Solutions like [VirtualWisdom from Virtual Instruments](#) use physical layer taps and fibre channel frame inspection to gather real-time data from every device and the network itself, not relying on periodic sampling of status codes. They can capture and record transaction information showing the relationships between all the devices in the environment and present a moving picture of the entire infrastructure that admins can 'rewind' to the specific point in time when a problem was reported.

Without real-time information, you're relying on averages. Imagine an average latency being reported as 20ms over a typical 5 minute polling period, an acceptable number for most applications. But that 20ms could represent 60% of your I/O at 10 seconds, and 40% at 100 seconds. So 40% of your users could be very unhappy, but you can't see it.

As an example, a check of the network at that time could reveal a 'loss of signal' or 'loss of sync' condition caused by a failing SFP or a cable disruption that's intermittent. These kinds of errors are fleeting and almost impossible to detect with standard element tools. But they can lead to aborts or long timeouts, or force devices to reinitiate or resend data which were corrupted, all of which degrades performance. With this information these components can be checked or replaced first, resolving the problem quickly.

Throwing Storage at the Problem

Besides performance and troubleshooting, asset utilization is critical in a cloud environment. Economies of scale, which drive these large infrastructures, require that storage arrays be run at high capacity levels. But many large SAN environments are seeing storage port utilization percentages in the single-digit range. This is often due to a propensity to ‘throw storage at the problem’ when an issue occurs. For example, server-based performance monitoring tools can report on latency experienced by an application, but they don’t tell you which of the LUNs supporting that server is producing the slow transactions. They also don’t show you which LUNs are experiencing low utilization and could be candidates for load balancing the slower LUNs.

Instead, what’s needed is information that identifies the source of a performance problem and helps administrators balance workloads. Solutions like VirtualWisdom can capture network transaction data to show the highest latencies between host HBAs and storage LUNs across the environment, not just for one server. It can show exchange completion times for server HBAs and the storage LUNs that are supporting them to determine if there’s a correlation between application latency and demand for storage. If there is, the cause may be an unbalanced load on the storage LUN or a network bottleneck, issues that can be quickly resolved by moving workloads to less busy LUNs, redesigning LUNs so they don’t share the same hard drives or rerouting network traffic.

Know when Storage is not the Problem

If no correlation exists, the cause could be settings like HBA queue depth or buffer credits or an intermittent network physical layer problem. These conditions can be identified easily and addressed. The point is that storage may *not* be the problem, but administrators don’t have the information to confirm that. In addition to resolving functional issues and safely maximizing usage of existing resources, managers need the ability prove when storage isn’t the cause to a performance issue and keep from adding unnecessary capacity.

Large enterprise infrastructures, often called “private clouds” are created to provide IT services for large numbers of users, usually in multiple departments and often in dispersed locations. While these environments share many characteristics with a typical corporate data center, they usually have more data and more applications running on virtualized servers and storage. This fundamental abstraction causes more complexity and the increased number of users and applications causes more risk. Keeping these cloud infrastructures up and running efficiently requires better information than is available from the usual data center tools. What’s needed is real-time information taken continuously from all devices and network components in the environment. Such data, that can be recorded and replayed from any point in time, can help to quickly resolve problems and keep system utilization and availability high.

About Storage Switzerland

Storage Switzerland is an analyst firm focused on the virtualization and storage marketplaces. For more information please visit our web site: <http://www.storage-switzerland.com>

Copyright © 2011 Storage Switzerland, Inc. - All rights reserved