COHO DATA

Enterprise Storage Guide
www.EnterpriseStorageGuide.com

Architecting Large Virtualization Environments Using Web-Scale Storage

How It Can Benefit the non-Web-Scale Enterprise

By Scott D. Lowe and David M. Davis
Co-Founders, ActualTech Media

September 2014

Introduction

When virtualization first came on the scene earlier this millennium, projects were generally limited to non-strategic server consolidation efforts and ancillary services. Mission critical services remained firmly in the physical world. Over the next few years, though, as the hypervisor realm became more trusted, more agile, and more dependable, IT decision makers started to view the hypervisor environment with a "first look" attitude. Rather than just defaulting to running new applications on physical systems, the virtual environment was first considered for the task.

Today, that "first look" is the de facto standard in IT, with more workloads being deployed on a hypervisor than on physical servers. However, with all of the benefits that virtualization has brought, increased use of virtualization has also created a number of challenges. Most notably, organizations have discovered that they need to rethink how they build their data center environments. Further, as the business landscape becomes increasingly competitive, companies have looked for new services that can streamline operational budgets and improve efficiency and agility.

At the same time, cloud providers have created infrastructures that are giving enterprise IT professionals new ideas for architecting data center environments. Companies such as Amazon, Google, and Facebook have storage environments that operate with scale in mind – to grow the environment – as a normal part of operations. Scaling in these kinds of environments is a routine task. Rather than scaling being something unusual, these companies can scale on demand for performance and capacity with low management overhead. These companies achieve this feat through the use of web-scale storage architectures that easily grows. These architectures are managed by software that automates configuration, load balancing, and the additional of new resources. An emerging segment of storage and infrastructure market are companies that are productizing the lessons learned in these large scale environments and delivering it in a products that enable enterprises to utilize it for environments that require scaling dynamically for business needs. In this paper, we will cover the keys to architecting large virtualized environments using web-scale storage.

Review: Key Tenets of Web-scale Storage

Before continuing with the discussion, make sure there is understanding around the key tenets that comprise web-scale storage:

1. Horizontal scaling enables capacity and performance on demand
2. Shared nothing storage architecture eliminates risk of complete storage outages
3. Multiple tiers of storage with automated management of tiers minimizes administrative overhead and lowers total cost of ownership
4. Commodity hardware helps bring down total cost of solution, leading to improved return on investment
5. Minimal management of storage further reduces total cost of ownership and helps IT respond to business needs with more agility

To learn more about these characteristics, read the paper Five Key Characteristics of Web-Scale Storage.

Architecting Storage for Virtual Environments

Before the onslaught of virtualization, storage was a far different world. Many companies ran SAN-based storage and many more leveraged storage directly attached to the server. This server-based storage eventually became relatively inefficient as organizations are looking for ways to optimize the environment and reduce costs.

SAN-based storage helped companies achieve their efficiency goals, but SANs had their own challenges, many of which are still suffered today. Many SAN environments lack high availability, disaster recovery mechanisms and scaling is a major hassle. Especially in terms of capacity and performance. In the sections below, we will cover what your storage environment needs to be capable of meeting the business needs of today and tomorrow.

Availability

As virtualization environments grow, the need for high availability of storage becomes critical since hardware can and does fail. This key point drives enterprises to demand storage that survives failures and continues to operate without any downtime to the applications as downtime could reflect loss in revenue.

Most traditional storage systems have the ability to survive a drive loss and controller loss. But when both storage controllers go down data becomes unavailable. The result: an inability to read, write or save data and a direct impact on the business. It's important to make sure the storage architecture is designed to handle both storage controller failures and drive failures while continuing to operate.

The ability to handle multiple storage controller failures without losing business continuity is what web-scale storage architecture was designed to do. With web-scale storage, the architecture was designed with high availability as a key tenant by leveraging the distributed nature of the storage architecture. Thus web-scale storage automatically handles the hardware failure and routes the data accordingly without requiring any manual management of the storage.

Performance and Capacity

Estimates place enterprise storage growth at a whopping 25% year over year. End users contribute significantly to this and don't want to constantly focus on keeping their storage use in check. Given such massive growth, adopting a traditional storage system will certainly require significant capacity-based reconfiguration as organizational data grows.

The need for performance has become critical in the storage world, equal to the requirement of maintaining sufficient storage capacity. Modern workloads, such as VDI and big data analytics, are driving IT organizations to rethink how they provide storage that can scale with performance to meet these and other virtualization use cases. With legacy storage systems, administrators would add different storage tiers to address individual application needs but this lead to constant management issues, such as:

- The need for manual intervention to rebalance and reallocate workloads each time storage is added. The result: Increased cost, as additional administrative overhead is required.
- A requirement to individually scale each storage performance tier as capacity needs dictate. The result: Islands of trapped storage and the potential to over or under allocate storage capacity.

Today's agile businesses require a storage environment that can keep pace with the demands of performance or more capacity and do it in a way that doesn't carry additional overhead. As such, a virtual storage environment needs to operationalize capacity and performance scale; just like Facebook, Google, etc. do in their environment. Furthermore, by operationalizing scale, organizations can achieve savings through eliminating management activities relating to creating volumes, LUNS and storage tiers.

Moreover, storage performance and capacity increases shouldn't have to be tied to the expansion of other resources in the environment, such as compute. Customers should have the opportunity to deliver on-demand storage performance to applications regardless of load placed on compute nodes. In other words, the ability to scale storage independently of compute is a key need in many enterprises.

Scaling storage shouldn't require someone to deliver a forklift to remove existing storage, either. Scaling storage to meet applications demands without having to forklift upgrades to current storage is critical to operationalizing scale, as this reduces management overhead.

Web-scale storage systems solve the capacity and performance scaling issues introduced by broad virtualization initiatives while eliminating the need for forklift upgrades. With the ability to simply add a node to boost both capacity and overall performance, web-scale storage systems bring simplicity to data center operations and allow a focus on the application rather than on the underlying infrastructure.

Cost

The biggest cost in a virtualization project isn't the hypervisor. In fact, it's not uncommon to see storage as eating up the biggest chunk of the budget – sometimes at much as 50% to 70% of that budget. As a result it's imperative to get the storage part right the first time. When implementing storage for a virtualization initiative, consider how your environment will grow over time and whether the storage architecture you choose will offer hardware flexibility. And continue to deliver the right amount of capacity and performance scale over the next 18-24 months from the initial deployment. This includes not just the storage, but flexibility of the interconnect. For example, does the storage leverage commodity Ethernet for converged interconnect or does it require expensive Fiber Channel or Infiniband connectivity? The latter per-port costs can add up really quickly.

As a part of the cost equation, consider also the effect that the wrong storage has on productivity in terms of both administration and the end user experience. Of course, CIOs want to see reduced levels of hands-on administration for storage systems, but when bad storage negatively impacts user productivity, operational costs are increased drastically and productivity drops as well. Thus it's key to deploy a storage architecture that can deliver ROI, while eliminating the need for constant management.

Web-scale storage is designed to reduce the overall cost of storage, from both CAPEX and OPEX views. On the CAPEX front, web-scale storage enables companies to take a "just in time" approach to buying storage. There is no more need to do a large storage buy once every few years and sink money into storage that may not be leveraged for 2 or 3 years. Buy just what you need right now and easily scale storage over time. Further, web-scale storage enables the use of Ethernet, further reducing costs. Finally, enjoy OPEX savings by reducing the sheer manpower necessary to manage the storage environment.

How Web-scale Storage Supports Virtualization Initiatives

With an understanding for how traditional storage impacts overall virtualization and how web-scale storage can help address these challenges can assist with specific kinds of virtualization initiatives.

Enterprise Server Virtualization

Even though newer applications, such as VDI and big data analytics, get a lot of the press these days, the fact remains that enterprise server virtualization remains the top use case when it comes to the need for web-scale storage. Even within the silo of the virtualization space, applications have suffered over the years as they've been starved of performance resources needed to deliver performance to the end user. Because web-scale storage systems add significant performance – in the forms of both IOPS and throughput – each time a node is added, enterprise applications – SQL Server, Oracle, Exchange, SharePoint, etc. – enjoy a storage resource that provides predictability that has been absent in the past.. With web-scale storage, the business is assured that these mission-critical applications get the resources they need.

On the capacity front, the ease of scaling capacity with a scale-out architecture with web-scale storage means as you grow from 20% to 50% to 80% virtualized, you can manage a single logical storage cluster. That can support a range of performance requirements with built-in tiering across petabytes of capacity. There is no more worry about hitting a scalability ceiling or worrying about what happens when shared resources become overwhelmed, as the architecture has no shared resources in the web-scale storage world.

Finally, ease of management enables administrators to use storage without having to worry about constant manual intervention. Web-scale storage architectures enables you to deploy storage to your virtual machines without requiring LUNs and without the need to create individual volumes. Further, web-scale storage allows you to jettison mundane tasks that are often associated with storage expansion, such as enabling load balancing and improving resiliency. All of these availability and data protection tasks are handled behind the scenes in an automated way.

VDI/Desktop as a Service

Virtual desktop infrastructure (VDI) projects were, for many, the last straw when it came to realizing that storage performance was simply not cutting it for the enterprise any more. Poorly planned storage environment brought failure to a number of VDI initiatives, but also forced CIOs and others to take a hard look at the storage environment. Unfortunately, for many, the choice was to implement VDI on separate specialized storage, thus negating potential economies of scale in storage and vastly complicating an already complex storage situation. For others, the choice was to implement a separate tier of storage in an existing environment and manually place volumes that required high levels of performance on new storage tiers and place data volumes on other tiers.

With web-scale storage, organizations get performance and capacity results that create success in VDI initiatives. Web-scale storage with automated storage tiers enables virtual desktops to have per virtual machine granular storage SLAs without requiring the manual migration of storage from different tiers. This means that disks that need performance – boot disks, to avoid "boot storms" – and disks that need capacity

– user data disks – can reside on the same storage with the system making automated decisions about where best to place data. These decisions are based on algorithms that take historical needs into consideration as a part of the placement process.

Further, web-scale storage for VDI/Desktop allows you leverage a shared nothing distributed architecture that avoids boot & login storms. Since desktop data is dispersed across multiple storage nodes to deliver performance and data redundancy.

Infrastructure as a Service/Service Providers

The crowded service provider space is one in which margins are critical to ensuring ongoing success as a company. As such, there is a need to keep operating costs low, but customer satisfaction high. This drives the need to automate the environment as much as possible so it's beneficial to both the customers and business.

A service provider requirement to the infrastructure vendors is to require as little manual configuration as possible even while application requirements change in real time. In the past, adjusting to changing application needs would require manual reconfiguration of the storage – LUNS, volumes, load balancing, tiering – but by leveraging web-scale storage you can deliver performance and capacity on demand without the overhead of spending countless hours on these value-minus activities. This is where web-scale storage really shines as it makes it easier to provide an infrastructure as a service offering without the additional overhead of man hours needed to constantly monitoring and reconfigure storage.

Infrastructure as a service offerings, by their very nature, must be ultra-reliable and not carry with them single points of failure that could jeopardize multiple customers in the event of an outage. A key characteristic of web-scale storage is a shared nothing architecture that ensure that there are no critical is no single points of failure. Web-scale storage provides customers with a globally distributed system with multiple data copies to ensure data availability for resiliency in the event of a failure. The software layer monitors the hardware to make sure there is always built in resiliency of the data to prevent data loss.

Let's wrap up with a discussion about why legacy storage deserves the boot due to sheer level of administrative overhead. With traditional storage, any time an application's needs change – perhaps there is a spike in use, for example – an administrator must be notified and then, with time, reactively move applications away from any developing hot spots that might arise. This leads to a lot of IT hours spent manual moving data around rather than focusing on solving important business issues. By leveraging web-scale storage, the system itself rebalances data across the shared nothing architecture to provide the best performance for applications without requiring manual intervention. This allows the environment to be dynamic and deliver performance to applications on demand without creating bottlenecks in the storage.

Closing Thoughts

You already know that virtualization brought to compute significant flexibility in terms of management. It also brought to compute major efficiency gains as administrators centralized their activities and focused less on the underlying hardware and more on the virtual machine layer. With web-scale storage, the same kinds of gains are brought to the storage realm, making the architecture a perfect complement to modern compute.

Remember, scale isn't necessarily all about sheer capacity. Instead, think of "scale" as the process of growing an environment easily using the power of distributed systems software on commodity hardware with automated storage management that is well integrated into applications like VMware.

Web-scale storage provides organizations with a storage solution that provides significant ease of scale and just the right levels of capacity and performance while reducing storage management and driving down overall storage costs. Coho Data, a leading web-scale storage vendor, provides organizations with solutions to meet even the most complex storage needs and at prices that simply make sense.

About the Authors

Scott D. Lowe is Senior Editor of EnterpriseStorageGuide.com. Scott has been in the IT field for close to twenty years and spent ten of those years in filling the CIO role for various organizations. Scott is also a micro-analyst for Wikibon and an InformationWeek Analytics contributor. In addition, Scott has also written thousands of articles and blog postings and has regularly contributed to such sites as TechRepublic, Wikibon, and virtualizationadmin.com. Because of his unique blend of skills (CIO/strategic & Engineer/tactical) Scott is also a sought-after resource for speaking engagements, and marketing collateral, including white papers and e-books, for a variety of technology firms.

David M. Davis is a well-known virtualization and cloud computing expert, author, speaker, and evangelist. David's library of popular virtualization video training courses can be found at Pluralsight.com. He holds several certifications including VCP5, VCAP, CCIE, and has been awarded the VMware vExpert award 6 years running. His website, covering the latest in virtualization is VirtualizationSoftware.com.

About Coho Data

Led by a team of XenSource/Citrix virtualization and storage industry veterans, Coho Data is enabling businesses of all sizes to build their own high performance web-scale storage for their environments. Coho Data has developed a scale-out storage architecture designed for virtual environments that delivers high performance and simplified management with ease of deployment.

COHO DATA