COHO DATA

Enterprise Storage Guide
www.EnterpriseStorageGuide.com

Five Key Characteristics of Web-Scale Storage

How It Can Benefit the non-Web-Scale Enterprise

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

July 29, 2014

Introduction

While every aspect of the data center resource spectrum — networking, compute, storage — is critical when it comes to completing the full infrastructure picture, storage has emerged in recent years as the current challenge for many organizations — and many new trends — that have hit the IT landscape. One of the most prevalent trends is what Gartner has coined web-scale IT and, with that, the rise of what's become known as web-scale storage.

In Gartner's estimation, by 2017, web-scale IT will be found in 50% of enterprises. Cameron Haight, research vice president for Gartner, has this to say about web-scale IT:

"Web-scale IT looks to change the IT value chain in a systemic fashion. Data centers are designed with an industrial engineering perspective that looks for every opportunity to reduce cost and waste. This goes beyond redesigning facilities to be more energy efficient to also include in-house design of key hardware components such as servers, storage and networks. Web-oriented architectures allow developers to build very flexible and resilient systems that recover from failure more quickly."

In this paper, you will learn about the **five characteristics that embody web-scale storage** and gain an understanding for how the architectural option provides companies with heretofore unseen levels of flexibility, agility, and cost control and enables IT to focus their efforts more on service delivery than technology.

STATE OF WEB-SCALE STORAGE REPORT

Coho Data partnered with **ActualTech Media** to conduct a market survey to gain information regarding the overall views of the web-scale trend. This report provides results and analysis around the feedback provided by respondents with a focus on web-scale storage.

Download your own copy of the free report here.

Agenda

In this paper, we will focus on the key characteristics that define 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

What Is Web-scale Storage?

First off, understand that there are two things that do *not* define web-scale storage. First, the word "web" in the trend name creates confusion. It has nothing to do with HTML or with web sites. Further, web-scale storage is not just about capacity as measured in gigabytes. Also understand that web-scale storage itself is not really a new concept. Companies like Amazon, Facebook, and Google have leveraged web-scale resources for years as the fundamental architecture for their data centers. It's only more recently that web-scale storage has moved to the enterprise as a truly viable architectural option, providing organizations with what amounts to storage-as-a-service without a lot of the baggage that hinders traditional storage.

Dynamic, On-Demand Growth

Web-scale storage is about dynamic growth. Storage scales independently from application servers to respond to storage growth needs, without being tied to other resources. While other architectures are a good fit for homogeneous workloads, many enterprises see compute and storage needs growing at different times based on various business triggers. Web-scale storage enables enterprises to buy storage when they need storage and compute when they need compute on demand.

New Economic and Operational Model

We**b-scale storage is about changing the model for sizing, purchasing and provisioning storage**. In traditional environments, decision makers often operate under an "estimate & buy" mentality for a resource that has to operate for 3 years or more. Under a web-scale storage model, decision makers operate under a more efficient and predictable "just in time" model that aligns with business needs, whether that business needs 100TB or 100PB of capacity they can scale as demand grows.

Scalability in Both Capacity and Performance

Web-scale storage is about operational scalability. Adding storage shouldn't be a burdensome process that requires manual reconfiguration or rebalancing of current storage. Every minute spent on these kinds of tasks is a minute spent wasting valuable time and resources. Inefficiencies in scale result in an overall high cost of managing the storage, even despite falling hardware prices. If the system is not automated to scale seamlessly, then it works against an organization's goals.

Five Web-scale Storage Characteristics

When compared to legacy storage models, web-scale storage features a number of unique characteristics that make it a compelling choice for the enterprise.

1. Horizontal Scaling

The days of vertical scaling — sometimes referred to as scale up — of storage systems are quickly fading as organizations seek ways to ensure that all storage metrics — capacity, networking/communications throughput, data processing speed, and performance/IOPS — grow as the organization expands its use of a storage device. In legacy scale up storage systems, customers had the ability to easily scale the capacity and IOPS portions of their storage systems, but while doing so, more and more load was placed on the single processing units and network connections that these devices leveraged. This approach eventually leads to network- and CPU-based performance challenges while, at the same time, the environment is exposed to a single point of failure in what is basically a shared architectural component: the processing unit, which connects to the storage communications fabric.

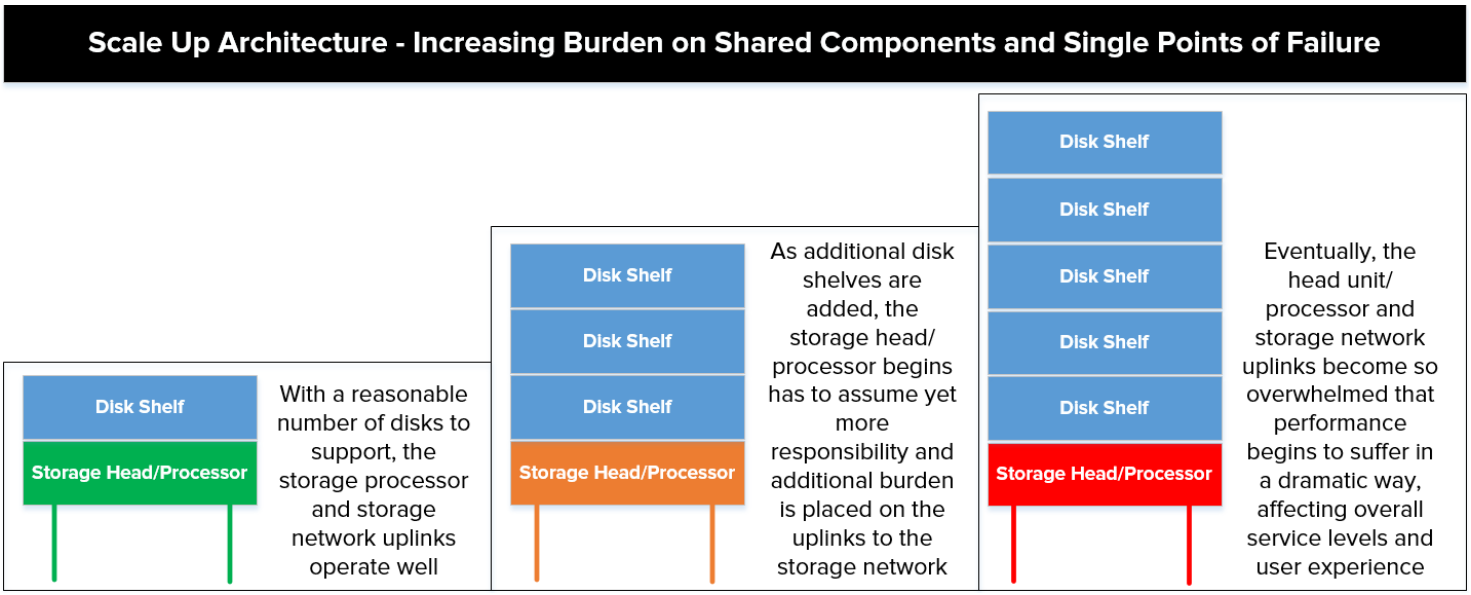

Figure 1: Scale up architectures eventually result in serious resource overloads

Web-scale storage relies on a different architectural paradigm known as *horizontal scaling*, which is often referred to as a scale out architecture. In horizontally scaled environments, new resources — generally in the form of new clustered nodes — are added in a linear fashion and, as added, each new node adds all necessary resources — storage, IOPS, networking, and CPU — necessary to stand alone. These individual nodes, although they include everything necessary to function on their own, participate as members of the clustered environment and contribute their individual components to the whole.

The cluster management software itself controls where data is placed on these individual nodes using globally defined granular policies that enable high levels of data protection and availability. If a node experiences a fault or other availability issue, another node in the cluster can transparently assume the responsibilities for the failed node. The fact that this failover process is transparent is a key characteristic of web-scale storage scenarios.

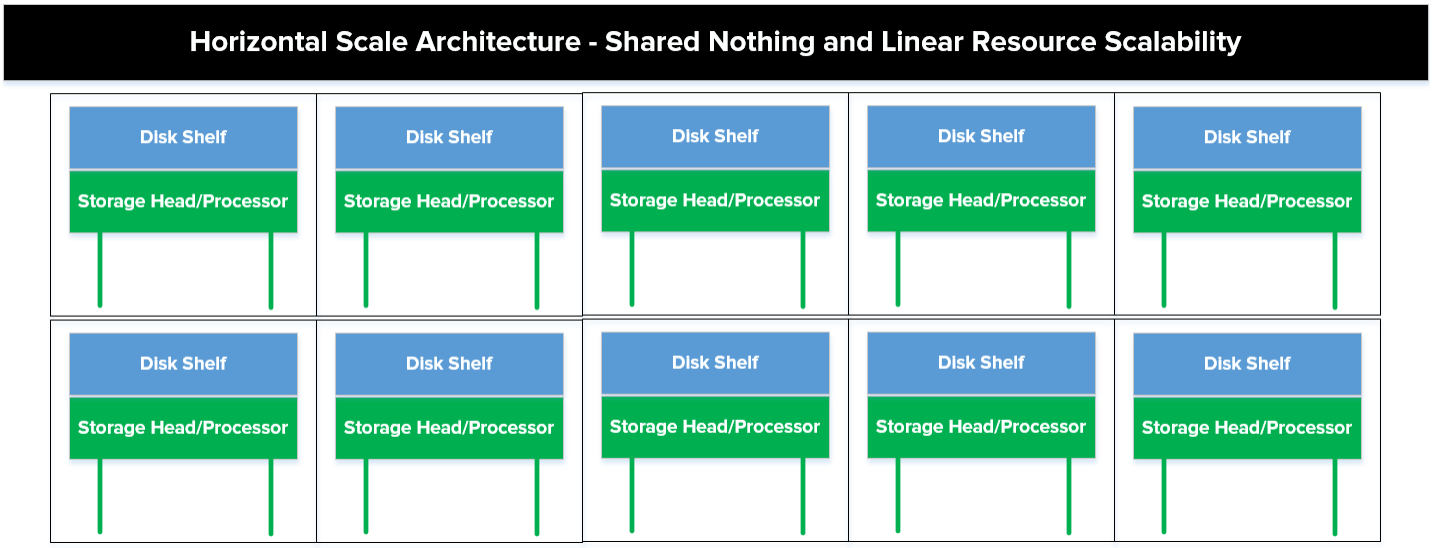

Figure 2: Horizontal scale architectures enable linear scalability

As new nodes are added to these kinds of clusters, the cluster management layer takes control of the new resources. This process includes adding the individual node resources to the global cluster and, if necessary, rebalancing other cluster nodes to leverage any performance or availability opportunities that may be afforded by the new resources. Again, these processes are generally transparent to both administrators and end users in Web-scale storage.

2. Shared Nothing Storage Architecture

In the data center, great effort has been taken over the years to reduce reliance on shared components. The reason is simple: When a shared component fails, any second order services that depend on that component will also fail, resulting a wider impact. Vertically scaled environments are particularly susceptible to failures that result from shared component failures. Moreover, such environments are susceptible to performance issues for the reasons described in the previous section — at some point, shared resources will become overwhelmed and failure can result in major outages.

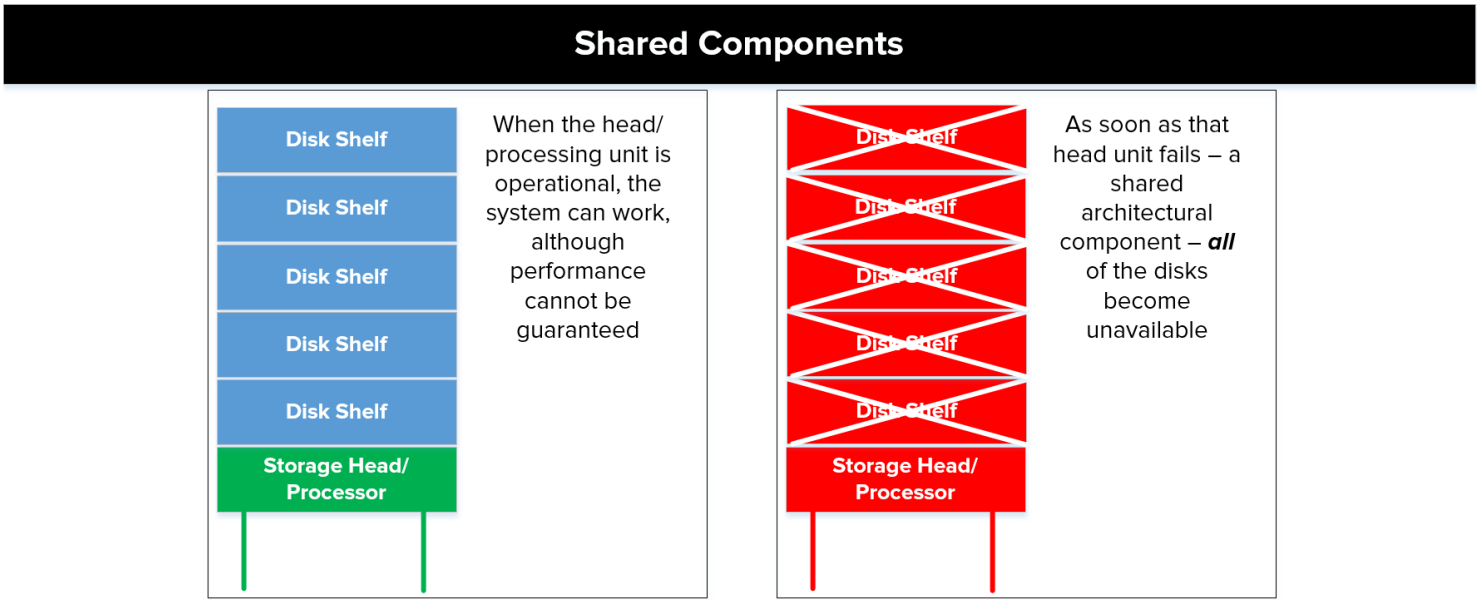

Figure 3: Shared components can result in widespread outages

A shared nothing architecture implies that there are no single points of failure in the data center. A shared nothing architecture that relies on commodity hardware controlled by a comprehensive software layer carries with it the understanding that hardware components can fail and they will fail. This should not be construed as a weakness in the system. In fact, it should be considered a strength as the software layer carries with it necessary intelligence in order to transparently handle hardware failures in a way that maintains the integrity of the stored data as well as expected performance levels. As is apparent in the figure above, none of the storage nodes share memory, disk, CPU, or network resources. If a component fails, the impact is very localized and very minimal. Each node in a shared nothing storage architecture operates independently without a single point of contention between nodes even during times heavy loads on the system. This is key to businesses, as it increases uptime and translates into more revenue generated.

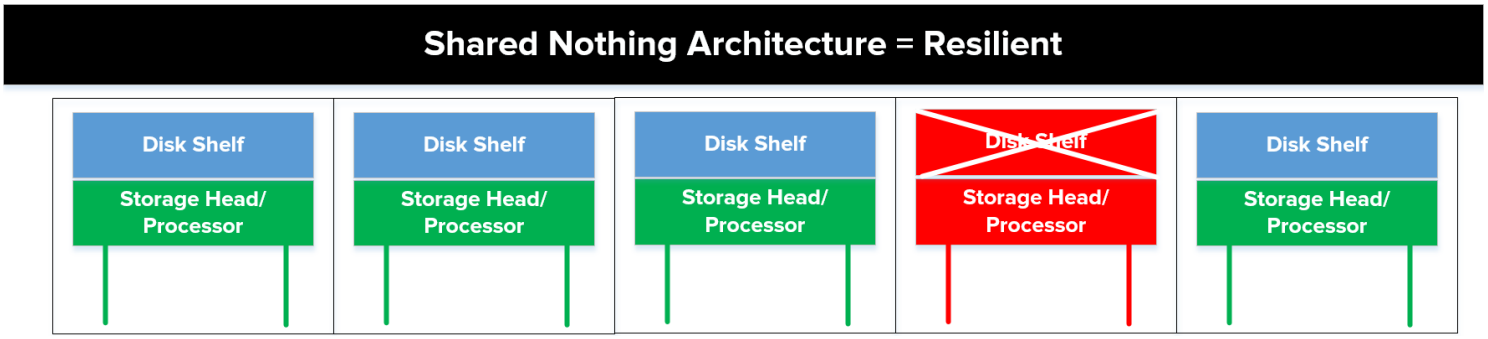

Figure 4: A shared nothing architecture compartmentalizes potential failures

3. Hands Off Management with Many Storage Tiers

Fact: Administrators should not need to manually control storage tiering. For those unfamiliar with the term storage tiering, this is a process by which storage is carved up into smaller chunks. Each of these chunks carries with it individual parameters and characteristics. For example, an administrator may create one storage tier for general purpose file storage. This tier would be configured with high capacity and relatively low performance specifications. Another tier might be created for database applications and may have only modest capacity needs but require significant IOPS to keep up with with business needs. Further, storage tiers may be created based on how data is accessed. Manual moving data around a storage system's tiers is a tedious process that is operationally expensive and that limits IT's short-term flexibility. With simpler and automated storage tiering, if data has been accessed recently, it may reside on a hot tier, while stale data may be moved to a cold tier.

When it comes to opportunities to scale the storage environment, breaking apart storage in these kinds of separate constructs begins to add complexity to the environment. Some storage systems do a good job at handling this complexity behind the scenes so that the administrator doesn't need to constantly be involved in tiering decisions. However, these auto-tiering solutions aren't available in every storage environment, particular in environments in which the storage architecture has been kludged together over the years. Moreover, even automated tiering systems can be more reactionary in nature than proactive. With a web-scale storage environment, the storage system tracks historical data usage patterns in order to predict when particular data will be used This enables such systems to take a much more proactive approach to storage management.

The bottom line is that web-scale storage provides multiple tiers of storage with automated handling of these tiers based on end user and business requirements of data and without requiring manual configuration of what data should reside in the individual (hot/medium/cold) tiers. The tiering of data is

performed based on business requirements rather than application demands as these are not one and the same. Legacy systems react to application performance demands whereas web-scale storage systems react to business-driven performance demands. This is a key distinction between legacy tiered storage systems and web-scale storage systems.

4. Commodity Hardware

The phrase *economies of scale* plays a major role in a web-scale storage environment. These environments are highly standardized, which has positive impact on a customer's total cost of ownership. Standardization also improves the overall administrative paradigm and improves outcomes provided by vendor support personnel. The more standard — or normalized — an environment, the easier it is to administer. Consider for a moment the plight of the legacy storage administrator. In many cases, as an environment has had to grow, organizations may have opted to move to vertical scale solutions simply to address immediate capacity or IOPS needs. Worse yet, they may have found it necessary to adopt storage systems that fell into one of the categories below:

- Were from different vendors.
- Were from the same vendor, but fell into different product lines due to different storage needs or vendor product availability.
- Were not intended to scale horizontally and were managed with per unit or per node administrative tools.

However, with web-scale storage, the software and processes layer is not the only element that carries with it a high degree of standardization. The hardware itself is also highly standardized and is often based on commodity off-the-shelf hardware. Besides being less expensive than systems that require high levels of custom hardware engineering, this kind of hardware is readily available from original equipment manufacturers, making creation of web-scale systems easier. Customers do not have to wait inordinate amounts of time for new nodes to be delivered, meaning that customer scaling needs can be accomplished relatively quickly. Being able to leverage hardware that requires no custom ASICs, board layouts or FPGAs is critical to creating web-scale storage that isn't hardware dependent. Over time, hardware dependent solutions can quickly become outdated and not integrate well with future generations of hardware.

People often believe that the word commodity relates directly to the word cheap, as in low quality. While it may be true that commodity hardware is in use in many modern data center architectures, including those that take a web-scale approach to storage, what is not true is that these components are necessarily low quality. A commodity hardware-based approach to web-scale storage enables storage vendors to focus more on building the comprehensive software layer that is going to become more important as businesses continue to push IT to greater limits. This approach untethers storage vendors from specific hardware and enables what would have been hardware engineers to focus on building and improving the software layer. As time goes on, IT will live in an increasingly software-defined world, a trend in and of itself. That has the potential to reap significant benefits for both IT and the business.

Commodity hardware makes it possible to mix and match hardware. This opportunity to mix and match hardware in a web-scale storage solution without requiring forklift upgrades is key to businesses as it eliminates down time which results in driving revenue and also enables economics of scale that are not possible with proprietary hardware solutions.

5. Minimal Storage Management Required

Legacy storage environments can be beasts to manage. Every vendor puts their own spin on the various constructs that need to be managed. Sometimes, it seems as if an administrators needs a Ph. D. in storage just to manage a new array. Remember this key fact: every time an administrator has to spend time on some mundane task is additional time that it takes to meet the demands of the business. This increases costs and extends the time-to-value for storage purchases.

Perhaps even more important that being able to be more hands off about things is the reason. Think about that legacy storage environment for a minute and consider some of the tasks that are performed on a pretty regular basis:

- Creating LUNs
- Managing those LUNs
- Creating and managing RAID groups
- Tuning storage on an ongoing basis
- Moving stored files to different tiers to change performance characteristics
- Expanding storage as new needs arise

Although these administrative tasks may become second nature to some, for many, they are simply time-consuming hindrances to the ultimate goal: enabling the business. Web-scale storage systems eliminate the need for many of these legacy administrative tasks. A key characteristic to web-scale storage is the ability to eliminate the mundane management tasks of storage such as creating LUNs, creating volumes, rebalancing storage, and tiering workloads. By letting the storage assume automatic responsibility for these tasks, it decreases operational expenses while delivering the same – or better – services to business. All of these mundane tasks are handled automatically behind the scenes in a well-designed web-scale storage system. Administrators focus on service delivery rather than on complex technology constructs. A focus on service delivery over technology is critically important for modern technology departments.

Action Plan

Can you relate to any of the following situations?

- I am finding it difficult to maintain necessary capacity and performance levels in my storage environments.
- I have found it necessary to deploy multiple islands of storage in my data center to meet individual application needs.
- My current storage environment has single points of failure that could result in a widespread outage.
- It takes too long to provision new business services due to the delays imposed in creating storage environments for those services.
- I am avoiding upgrading my current storage environment due to the expense and complexity of doing so.
- I have to buy bulk storage at the front end of my replacement cycle (i.e. buy more than I may need) because upgrading mid-stream is prohibitively expensive.

If these issues sound familiar, web-scale storage may be just the solution that solves – once and for all – the storage challenges being faced. Even if these situations aren't familiar, it's time to consider web-scale storage's ability to focus on end users applications with ability to have quality of service across the infrastructure stack.

Web-scale storage provides unprecedented ease of scale and performance and eliminates storage management while driving down costs as a business benefit. 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