

Massively Scalable NAS – Pre-Emptying Tomorrow's Data Overload with Today's Technology

An ENTERPRISE MANAGEMENT ASSOCIATES® (EMA™) White Paper
Prepared for Hewlett-Packard

May 2008



Table of Contents

- Executive Summary 1
- Introduction 1
- What Drives Data Growth? 1
- What is Massively Scalable NAS?..... 2
- EMA’s Perspective 5
 - Why Scalable NAS? 6
 - Buying a Preconfigured System -- Does It Make Sense? 6
 - Who Should Consider Massively Scalable NAS?..... 7
- About Hewlett-Packard..... 7

Executive Summary

NAS has always been simple...unless IT managers wanted to grow their NAS storage significantly. Then the simplicity evaporated. Now, with massive storage build-outs a necessity, a simply managed, highly scalable NAS environment is a must. For the first time, storage administrators are thinking in terms of managing *petabytes* of data. Fortunately, new technologies make such systems available even to IT shops with limited budgets. Now smaller firms as well as large enterprises can scale their storage to meet demands, and can align their storage investment to accommodate a need for more storage, for faster storage, or both.

Introduction

No data center manager needs to be reminded about how data is growing at enterprise sites. Regardless of what business their company is in, it’s a rare IT manager indeed who doesn’t budget for *quarterly* data growth rates approaching and often exceeding 25%. The cost of this may be painful, but IT managers rarely perceive it as being as painful as the consequences of not being able to adequately store and protect a company’s intellectual property. Running out of storage capacity or falling down on the job when it comes to protecting data are perceived by many IT staffers as being the ultimate career-limiting events.

What Drives Data Growth?

Several easily identifiable trends are causing the increase in both personal and corporate data. Personal data (much of which is maintained at corporate sites) sees larger file sizes from applications, cameras, and cell phones. More users watch and share streaming video. Broadcast and social network sites demonstrate dramatic growth.

Storage is being consumed faster than ever by private users, and is being consumed just as dramatically by business users, particularly in the following areas:

Increased use of rich media by mainstream applications. Rich media has proven to be extremely useful in mainstream corporate applications. Examples of this are streaming video embedded in computer-based training programs, webinars that are stored on corporate websites for later use, and the like. Customer-facing Web applications in particular often require substantial amounts of data and high-performance storage to house them.

Web 2.0. As the Internet increasingly becomes a platform for mass collaboration and computing, the rich applications that support these operations may well place the greatest demands on storage. It remains to be seen where “Web 2.0” will eventually go, but Web syndications such as RSS and numerous other applications that transfer data require more than just bandwidth; there is also a requirement for a place to store whatever the bandwidth has transferred between one collaboration node and another.

Storage is being consumed faster than ever by private users, and is being consumed just as dramatically by business users

Digital archiving. For a number of reasons, most companies find that it makes good sense to convert their analog archives (usually consisting of myriad boxes, cases and cartons of documents of all sorts) into digital archives. Such data is typically scanned and then stored on digital tape. Storage costs are less of course, but the real reasons behind this shift have much more to do with satisfying a need to access and identify off-line data in a timely fashion. Properly filed and accessible data, even old data, is part of a company’s intellectual property and should be viewed as an asset to be leveraged whenever possible.

More granular data. Many applications have increased the granularity of the data they use. The most obvious example of this is the digital camera, which five years ago provided pictures in a range of one-to-two megapixels, and which today creates consumer photographs of eight megapixels, 10 megapixels, and beyond. Similar increases in granularity are seen in the medical and scientific fields (with significantly more accurate CT scans and digital x-rays), and, perhaps most dramatically, in commercial television, where the shift to digital (and beyond that, to high definition) formats now means 13.5 GB of storage space is necessary for each hour of encoded video. Even more traditional applications such as databases are making use of new technologies such as visualization and data cubes which can add substantially to their storage requirements.

Extremely scalable NAS should offer advantages over more traditional storage in the areas of scalability, manageability, and affordability.

Whatever the cause for such growth, the storage must be managed efficiently so that the data is adequately protected and always available whenever and wherever it is needed.

What is Massively Scalable NAS?

Massively or extremely scalable NAS is storage capable of accommodating file-based content that is always growing and that must be pervasively available. Furthermore, such devices must do this in a way that is affordable in terms of a device’s initial cost, the incremental costs associated with each scale out operation, and of course, the mix of daily expenses that compose the OPEX. All storage must be fully visible all the time both to the clients accessing the data and to the management console that monitors the system. The system should be rapidly scalable, and should be capable of moving any data to an application in both small increments and very large increments (hundred-plus terabytes at a time).

Extremely scalable NAS should offer advantages over more traditional storage in the areas of scalability, manageability, and affordability.

Scalability. Such systems are extremely dense in terms of both computing power and storage capacity; that is, they are able to put a lot of computing power and storage capacity into each cubic foot of data center space they occupy. Not all dense systems are scalable however; many require a data center to “start big”, and then to increase computing power and storage capacity in large increments. Scalable systems operate differently, providing minimal configurations and then allowing scale out operations in small, affordable increments that match performance and capacity demands. Because of this capability,

they allow a data center to start small and then to scale out to develop a massive storage infrastructure capable of supporting even the most data-hungry applications. Ideally, performance and storage capacity should scale independently of one another, allowing the system managers to add performance without spending money on added storage capacity, or to add storage capacity without the necessity of budgeting for unneeded extra performance. Systems scale as needed, in an affordable and predictable fashion.

Manageability. The maximum value of massive scalability – particularly in on-demand environments -- can only be realized if the newly added content and the devices that hold it can be managed easily and with efficiency. This means several things. Wizards and/or policy-based decision making are fundamental. Scaling operations, no matter how large they may be, must never impose substantial planned downtime. Initial deployment and day-to-day management of both the storage and server components should be done through a single, unified interface that streamlines all aspects of provisioning and configuration, and which allows servers and storage to be scaled both together and separately, as requirements demand. Adding additional storage modules or server blades must be a simple, automated affair that can be handled with relative ease by any level of staffer. Advanced systems using more sophisticated management tools -- high availability (clustering) software for example – should have those tools integrated within the management console. Systems with capabilities such as these will enable even relatively inexperienced administrators to scale their management capabilities as well, eventually enabling them to manage petabytes rather than terabytes of storage.

Affordability. Clearly, massive scalability and highly available systems will be of little use to IT managers if they are not also affordable. Systems should have no hidden costs, should scale out in a granular and affordable fashion, allowing the system manager to invest only where investment is required, and should be managed by a single interface that addresses both performance and capacity management. A single easy-to-use management interface such as this has the potential to drastically lower administrative expenses.

HP StorageWorks 9100 Extreme Data Storage System (ExDS9100)

HP’s answer is the ExDS9100, a system designed for use in a wide variety of application environments that require high data throughput and extreme scalability. Importantly, scaling storage capacity is done easily and economically, and because it requires no “forklift upgrades” allows for quick scaling of operations in response to customer demand loads. The system has management features built-in to enable rapid turnkey deployment of both the initial installation and any upgrades that become necessary to accommodate new storage requirements. It is designed to provide high throughput for content hungry applications, and is highly scalable. Perhaps most importantly, the system emphasizes affordability both in terms of initial acquisition costs and the day-to-day operational expenses required to maintain it.

These systems arrive from the vendor preconfigured and pretested (an easy value-add for HP as the vendor builds all components of the system itself). The result is that downtime -- both planned and unplanned -- is kept to a minimum, and content remains highly available for the applications that need it. This adds further value when customer support is required; because there are no third-party products, a single phone call to HP support is all that is required to access support for any software or hardware within the system. No time will ever be wasted determining who the right support person is.

The system is designed from the ground up for easy scalability, allowing extreme build-outs that enable small initial installations to grow easily, manageably, and affordably into the massive systems needed to support Web 2.0, digital media, application hosting, and other content hungry applications. Significantly, this ease of scale out applies to more than just adding capacity; it applies to systems management as well. IT managers will find administrators using the ExDS9100 are now able to manage petabytes of information rather than terabytes. Lower per-terabyte management costs will significantly lower operational expenses, and rapid deployment of added capacity in response to customer demand will be less stressful.

The system. The ExDS9100 is based on Hewlett-Packard’s c-Class BladeSystem, which combines the power and space efficiencies of blade-based processing with high-density storage. The base system consists of four processor blades, each with multi-core processors, and 246 TB of online storage. Both storage and processing power may be scaled independently of one another, depending upon the demands of the application. Processor blades are expandable from four to 16 blades, while storage capacity is expandable to 820 terabytes, resulting in a storage density of 12 TB per “U”. The processor blades and the storage devices may be scaled either together or independently of one another, depending upon application requirements.

A single GUI manages all aspects of the ExDS9100. Configuration may be done using either wizards or command line interfaces.

Use. The ExDS9100 accesses data via NFS, HTTP, and PolyServe’s Direct IO. Because this is an open system built on industry-standard parts, applications may run directly on the blades, each of which has fully parallel, symmetric access to the file-based data residing in the virtualized storage pool.

Rapid deployment. Storage capacity and processing power are easily upgraded on the system. To add storage capacity the process is:

1. Add a storage block to the system by connecting two cables.
2. Type one command and within minutes, file system capacity is automatically grown.

To add processing power, the process is:

1. Slide in a performance block (blade server).
2. Type one command and within a few minutes, you have additional performance in the form of another node in the cluster.

High-availability and data protection that is built-in. The ExDS9100 comes with its components preconfigured for high availability. In the event of either hardware or software failure, HP’s software provides a high availability option that allows seamless failover to take place, ensuring continuous client access. It consolidates islands of data from file servers into a single, clustered storage pool.

Additionally, because the system supports rolling upgrades there is no need for any planned downtime for maintenance.

The system also comes with RAID 6 built-in.

For several years a generally accepted standard for data protection at most enterprise data centers has been RAID 5, a technology that provides a high level of data protection by extending a parity stripe across each element in the RAID set. A difficulty with RAID 5 however is that when any disk within the RAID set fails, even if that failure was but a single unrecoverable read error, the system loses its ability to protect data -- should a second device fail, all data is irretrievably lost. Unfortunately, the new larger disks that are now available will eventually fail action in, and while these failed disks are being replaced within systems, all data is vulnerable.¹ RAID 6 was developed as a direct response to this situation.

RAID 6 adds a second parity stripe to the RAID set, which provides a second level of protection. A RAID 6 configuration tolerates two device failures within each RAID set, a significantly lower probability than a single device failure. In the event of disk failure, the array can be rebuilt without sacrificing protection for data that is on it.

¹ Industry data indicates a high degree of data loss occurs during RAID 5 rebuilds, primarily due to operator error (typically, the wrong disk being removed, or a misconfiguration of the RAID set), but also due to power problems (outages, brownouts, and spikes).

EMA’s Perspective

The highly competitive nature of modern business demands that all of a company’s assets be immediately available when they are required. Nowhere is this more important than when it comes to corporate data. The challenge of course is that the total amount of data being handled by almost every company has increased at a frightening rate, and the need for speedy access to that data – on-demand in the course of daily business, during an e-discovery exercise, or in response to regulatory demands -- continues to increase. The challenge becomes all the more problematic because budgets rarely increase sufficiently to accommodate these demands. Increased requirements for high-performance data access coupled with the anticipated sluggish budgetary response to those requirements mean IT managers will have to be particularly careful when it comes to spending for storage systems. Every dollar spent must be directed towards systems that will provide appropriate levels of performance, that will scale out easily when more storage is needed, that can be easily managed, and that can be run with economy.

...the total amount of data being handled by almost every company has increased at a frightening rate

Why Scalable NAS?

IT managers have always liked NAS solutions because they are easy to manage when compared against the alternatives provided by DAS (direct attached storage) and SAN. NAS has fallen short however in its ability to scale capacity simply and economically, and as a result the NAS market is essentially split into two separate segments – low-priced, low-performance, non-scalable devices that are best suited for home offices and the like, and much more expensive enterprise class devices that offer high-performance and scalability but which also typically require special management skills. Obviously the ideal is to take the best from both of these worlds, and provide a reasonably priced, highly scalable, high-performance machine.

If in addition the system combines capacity management and performance management into a single tool, users can also expect a lessening of the operational expenses associated with managing and monitoring the system.

In many cases scalable NAS fits the bill quite nicely. Sites anticipating a high degree of data growth can expect to get good value from these systems if they are aggressively priced in terms of acquisition costs and hardware scale out costs. If in addition the system combines capacity management and performance management into a single tool, users can also expect a lessening of the operational expenses associated with managing and monitoring the system.

Finally, well-designed scalable NAS systems will be extremely dense, making use of blade and other technologies to concentrate significant amounts of compute power and capacity in a small area. In many cases this will result in a significant decrease in power consumption requirements.

Buying a Preconfigured System -- Does It Make Sense?

EMA believes that well-chosen preconfigured systems deliver particular value to any IT environment. While many IT organizations must make significant commitment to acceptance testing prior to the operational cutover of any new system, appliances and other preconfigured products should arrive on site pre-tested and pre-configured. This is no trivial thing, as interoperability, performance, and regression testing are extremely time- and labor-intensive.

Consider, for example, the case of interoperability testing, done to ensure that systems and services operate effectively with one another. In a system certified as being interoperable, every reasonable, and in some cases unreasonable, combination and permutation of the system's components is tested prior to certification. Not only must the server and storage be tested together, but every change to a component (firmware upgrades, for example) must also be tested prior to release. With preconfigured systems, this is a relatively straightforward task, but the reality for multi-vendor environments is quite different. Where there are many devices from many vendors, it is rarely practical to test every possible device in the mix; in a preconfigured environment, much (although certainly not all) of that testing is no longer necessary, so things happen faster and more reliably.

What other values do pre-configured systems deliver? Pre-tested systems with known capabilities ensure predictable operations, easier upgrades and efficient maintenance. Their consolidated architecture provides an easy route to unified management. The consolidated monitoring, analysis and management capabilities that come from this can form

The ability to simply uncrate a system, roll it out onto the data center floor, and then fire it up is a scenario most IT managers will find highly compelling

the basis of automated, on-demand IT services that deliver increased availability, improved performance and a much greater likelihood of meeting agreed-upon service levels. And of course, a system that arrives on the loading dock preconfigured and ready to go offers markedly faster time to value. The ability to simply uncrate a system, roll it out onto the data center floor, and then fire it up is a scenario most IT managers will find highly compelling

Who Should Consider Massively Scalable NAS?

Likely candidates for scalable NAS (network attached storage) systems such as the one described in this paper obviously include providers and distributors of high definition content. This of course includes firms involved in any form of rich media such as video pre- and post-production, pre-press, 3D modeling, satellite imagery, and many aspects of high-performance computing. An aggressively priced, highly scalable NAS system would have applicability far beyond those segments however, and would certainly deliver value both to service providers and to any company involved in:

- Delivering streaming video to users (perhaps as part of an educational or sales outreach program).
- Providing Web 2.0 file servers.
- Digital archiving, including companies whose strategy includes migrating legacy information into a digital environment (such analog-to-digital conversions including scanning paper documents and digitizing data that currently resides on analog videotapes).
- Storing massive amounts of data, either as large individual files (“binary large objects”) or as a great number of smaller files.

Companies providing storage services to the broadband, cellular, broadcast, life sciences, and oil and gas industries all are likely users but, with across-the-board increases in storage demand it is clear that most companies will benefit from an ability to add storage in a rapid, well-managed, and affordable way. It's hard to imagine a corporate IT environment that would not benefit from files that were more accessible and easier to migrate, and it is hard to imagine an IT manager who would appreciate a simpler growth path to increasing storage capacity.

About Hewlett-Packard

HP focuses on simplifying technology experiences for all of its customers – from individual consumers to the largest businesses. With a portfolio that spans printing, personal computing, software, services and IT infrastructure, HP is among the world's largest IT companies, with revenue totaling \$107.7 billion for the four fiscal quarters ended Jan. 31, 2008. More information about HP (NYSE: HPQ) is available at www.hp.com.

About Enterprise Management Associates, Inc.

Founded in 1996, Enterprise Management Associates (EMA) is a leading industry analyst and consulting firm dedicated to the IT management market. The firm provides IT vendors and enterprise IT professionals with objective insight into the real-world business value of long-established and emerging technologies, ranging from security, storage and IT Service Management (ITSM) to the Configuration Management Database (CMDB), virtualization and service-oriented architecture (SOA). Even with its rapid growth, EMA has never lost sight of the client, and continues to offer personalized support and convenient access to its analysts. For more information on the firm's extensive library of IT management research, free online IT Management Solutions Center and IT consulting offerings, visit www.enterprisemanagement.com.

This report in whole or in part may not be duplicated, reproduced, stored in a retrieval system or retransmitted without prior written permission of Enterprise Management Associates, Inc. All opinions and estimates herein constitute our judgement as of this date and are subject to change without notice. Product names mentioned herein may be trademarks and/or registered trademarks of their respective companies.

©2008 Enterprise Management Associates, Inc. All Rights Reserved. EMATM, ENTERPRISE MANAGEMENT ASSOCIATES®, and the mobius symbol are registered trademarks or common-law trademarks of Enterprise Management Associates, Inc.

Corporate Headquarters:
5777 Central Avenue, Suite 105
Boulder, CO 80301
Phone: +1 303.543.9500
Fax: +1 303.543.7687
www.enterprisemanagement.com



1623.050508