An *IT Briefing* produced by SearchStorage.com

How to Build a Superior File Serving Environment Without Costly NAS Appliances

Sponsored By: PolyServe™ Server and storage clustering software

How to Build a Superior File Serving Environment Without Costly NAS Appliances

By Michael Callahan

© 2006 TechTarget

BIO

Michael Callahan—Currently the PolyServe Chief Technical Officer, Michael was formerly the head of Advanced Development at Ask Jeeves. He has more than 15 years of research experience in computer vision, scientific visualization, differential geometry and topology, networking, and distributed file systems. The Linux networking software he wrote is in use at millions of sites worldwide. He was a Rhodes Scholar and a Junior Research Fellow in Mathematics at Oxford University and holds a BA from Harvard University.

This *IT Briefing* is based on a PolyServe/TechTarget Webcast entitled "How to Build a Superior File Serving Environment Without Costly NAS Appliances." To view this Webcast online, please click the link:

Contents

Copyright © 2006 PolyServe. All Rights Reserved. Reproduction, adaptation, or translation without prior written permission is prohibited, except as allowed under the copyright laws.

About TechTarget *IT Briefings*

TechTarget IT Briefings provide the pertinent information that senior level IT executives and managers need to make educated purchasing decisions. Originating from our industry-leading Vendor Connection and Expert Webcasts, TechTarget-produced IT Briefings turn Webcasts into easy-to-follow technical briefs, similar to a white paper.

Design Copyright © 2004 - 2005 TechTarget. All Rights Reserved.

For inquiries and additional information, contact:
Dennis Shiao
Director of Product Management, Webcasts
dshiao@techtarget.com

How to Build a Superior File Serving Environment Without Costly NAS Appliances

Introduction

A number of problems have emerged in some I.T. environments as file serving has grown in its applicability and scope. With file serving's growth has come a dramatic increase in individual machines or appliances that are providing file-serving capabilities.

The following document will explain an alternative to the traditional approach of managing file serving and its resources. This new approach, deemed the "UnAppliance," uses software – rather than yet another appliance to tie the many strengths of what the industry-standard hardware community is bringing to the market to build a more flexible file-serving environment.

This document will describe the UnAppliance and its applications, offer some examples of customers using this approach, and explain how the approach uses a unique technology called "shared-data clustering."

A Different Approach: The UnAppliance

Very often, the traditional file-serving environment – both appliances and regular standalone file services – has difficulties with overall levels of performance. As more clients are connected to a given file-serving device, that device can become a bottleneck, as shown in Figure 1. The appliance or server represents a choke point, a maximum level of performance that can be achieved in supplying data.

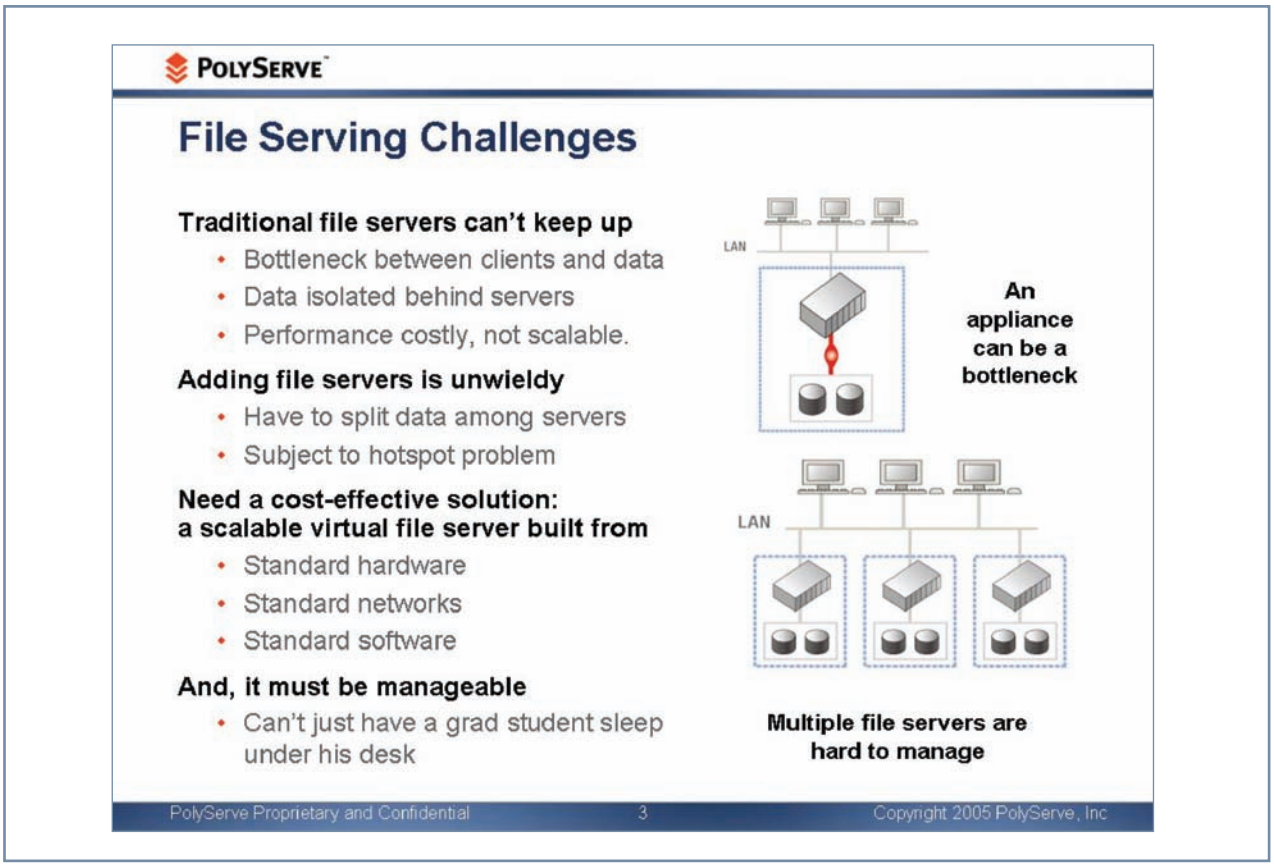

Figure 1. File Serving Challenges

One solution in some environments is to implement multiple file servers and multiple appliances that split the data among the individual machines to spread the load. This presents a challenge by significantly complicating the administrative environment while increasing the cost. Also, splitting the data with incorrect partitioning may significantly overload some of the additional file-serving devices, while other devices are not being heavily used. This has been called the *hotspot problem,* where some parts of the data are heavily used and those servers that are handling those hotspots become overloaded, while other parts have a large amount of spare capacity.

The ideal solution would use entirely standard hardware, standard Linux, and standard software environments that could be managed in an easier manner than the sort of proliferate environment that is so common.

Goals of the UnAppliance

The goals of the UnAppliance include:

- Harnessing multiple hardware components into a single unit so they scale performance by aggregating, not partitioning
- Working with industry-standard hardware and interfaces, including equipment preferred by clients
- Taking advantage of spare CPU capacity on existing equipment to provide high availability for certain data, without resorting to a second device in crossover-failover configuration or duplicate hardware
- Leveraging the expertise, processes, and support software clients already have

The Benefits of the UnAppliance (the PolyServe File-Serving Utility)

Figure 2 shows an environment (described as Windows, but it can also be Linux) with a set of 16 servers connected in a network, and all connected to a common pool of shared storage on a storage area

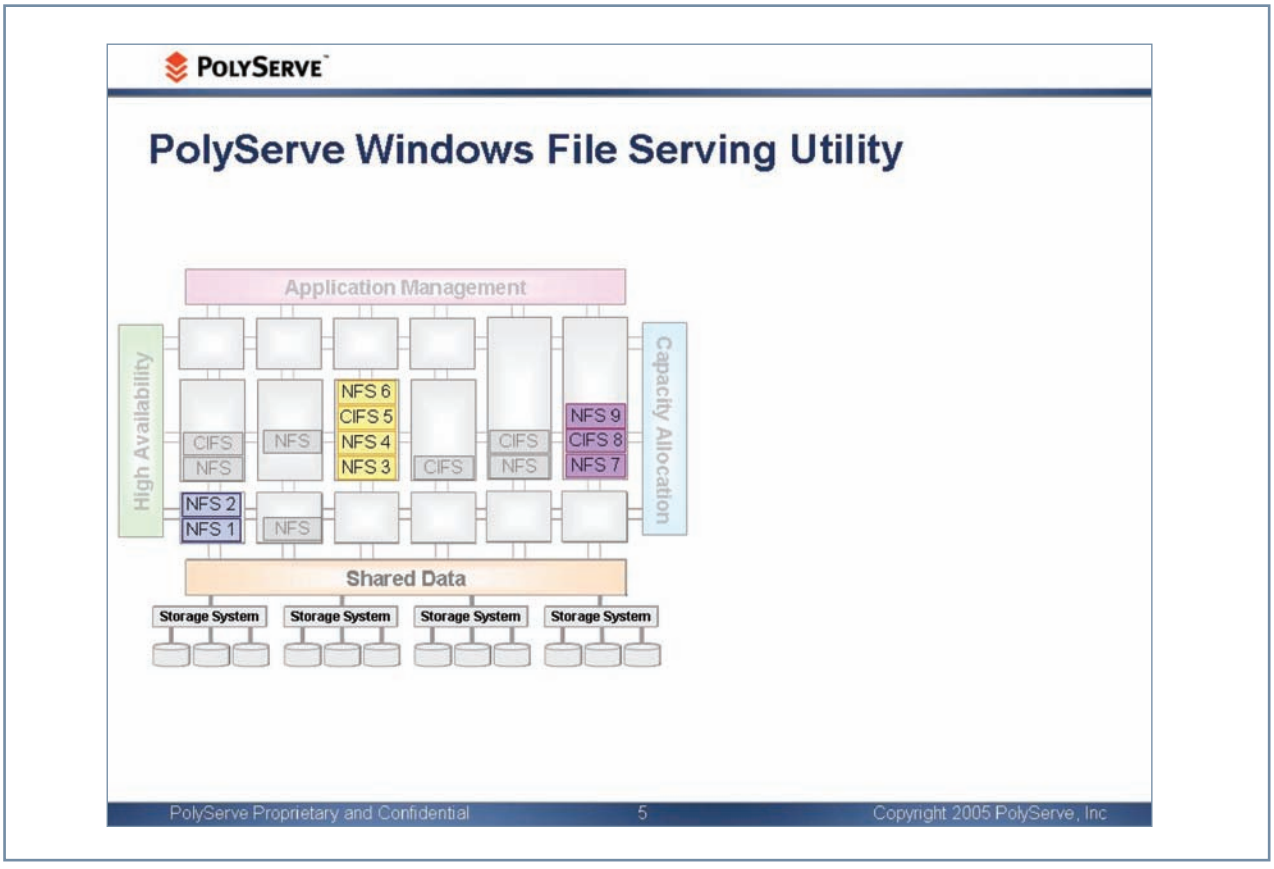

Figure 2: PolyServe Windows File Serving Utility

Sponsored by PolyServe Server and storage clustering software

network (SAN). The different sizes of the server representations show that these servers are configured differently: some of them could be dual-, tri-, or four-process servers, and some of them might be traditional single course servers. The exact layout is not important; the important point is that the sum collection of servers chosen are running a set of different file-serving tasks.

The PolyServe File Serving Utility provides fault tolerance and high availability. If one of the servers failed, the UnAppliance will transition the client communicating with that server over to another server, reconnecting immediately and continuing to get access to the data needed. Therefore, if, as shown in Figure 3, the NFS server in the lower left corner were to fail, those NFS services will automatically become available on another machine.

In an environment where individual servers provide individual file exports, some servers get overloaded. The PolyServe File Serving Utility enables multiple servers in the environment to export the same shares or exports at the same time. Data no longer needs to be partitioned among individual file servers, since multiple file servers export the same files at the same time, which scales the throughput of access to a given set of data. As shown in Figure 4, servers or storage arrays can be added while the system is running if a larger data footprint is needed.

The PolyServe File Serving Utility also provides simplified administration. If, for example, as in Figure 5, the server shown with yellow exports needs a hardware upgrade, the administrative tools can fail-over the client sessions to other servers. This frees the server needing the upgrade to be taken offline with no interruption of service, providing a comprehensive means of using standard servers that run Windows or Linux and use standard Intel or AMD chips to provide an aggregate file-serving capability.

Performance

The UnAppliance liberates the connection between an individual file server and the piece of data, allowing multiple file servers to export the same data simultaneously. The performance for those exports scales simply through the addition of servers to the

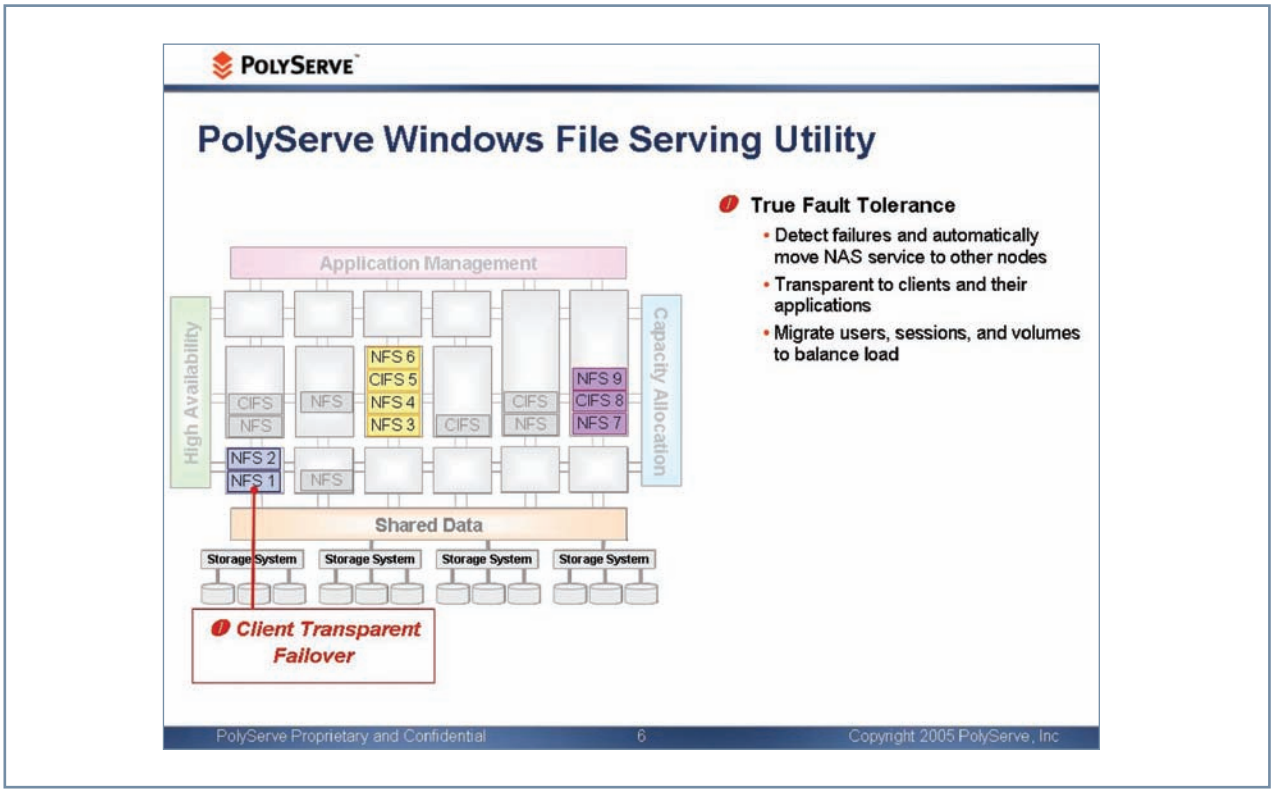

Figure 3: PolyServe Windows File Serving Utility: True Fault Tolerance

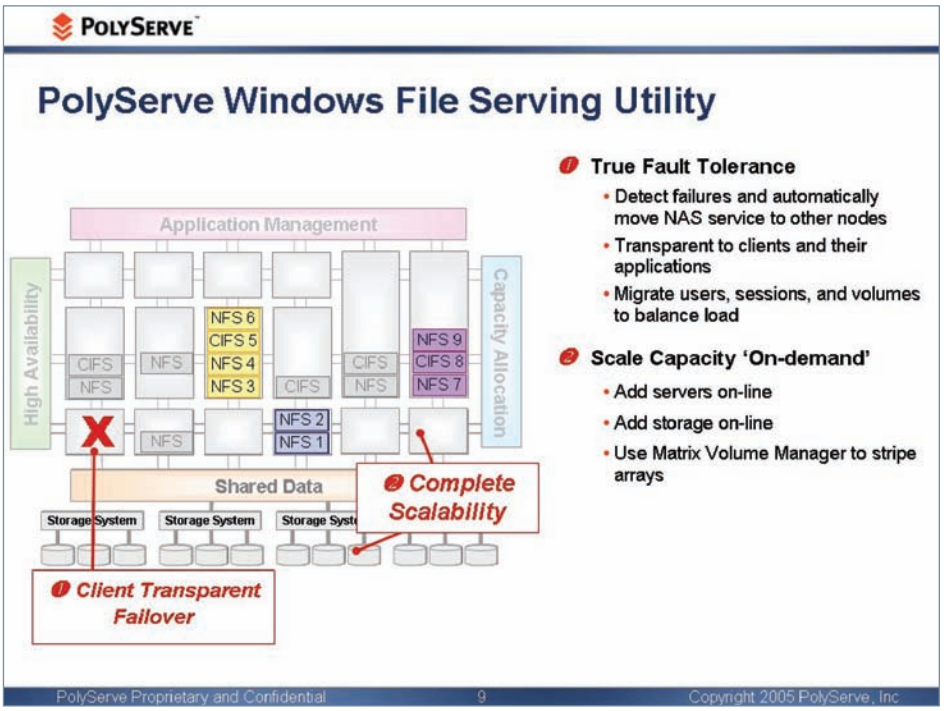

Figure 4: PolyServe Windows File Serving Utility: Scale Capacity On-demand

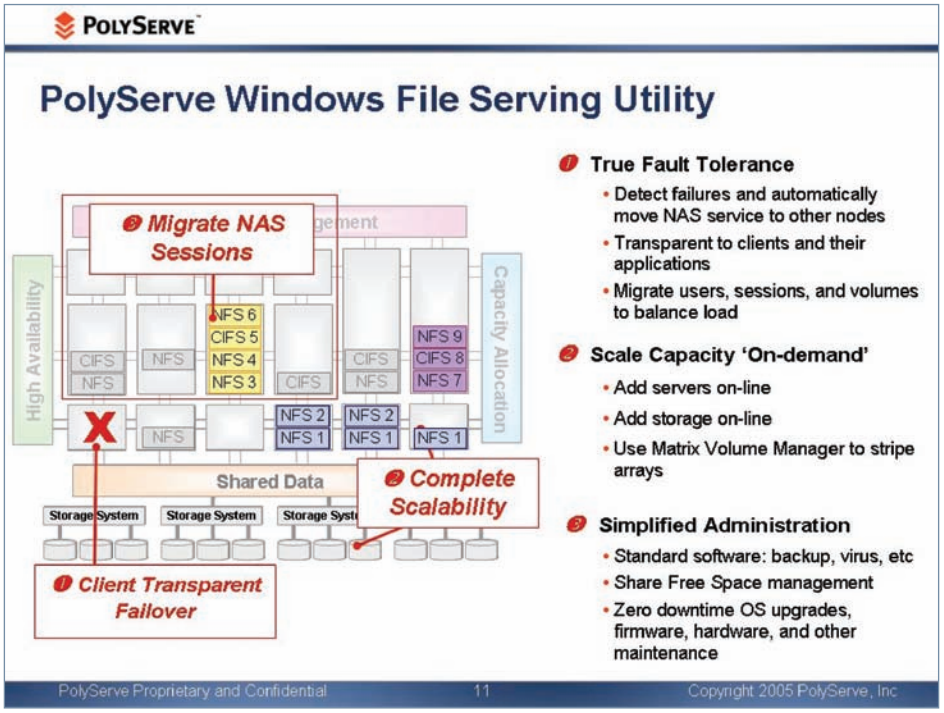

Figure 5: PolyServe Windows File Serving Utility: Simplified Administration

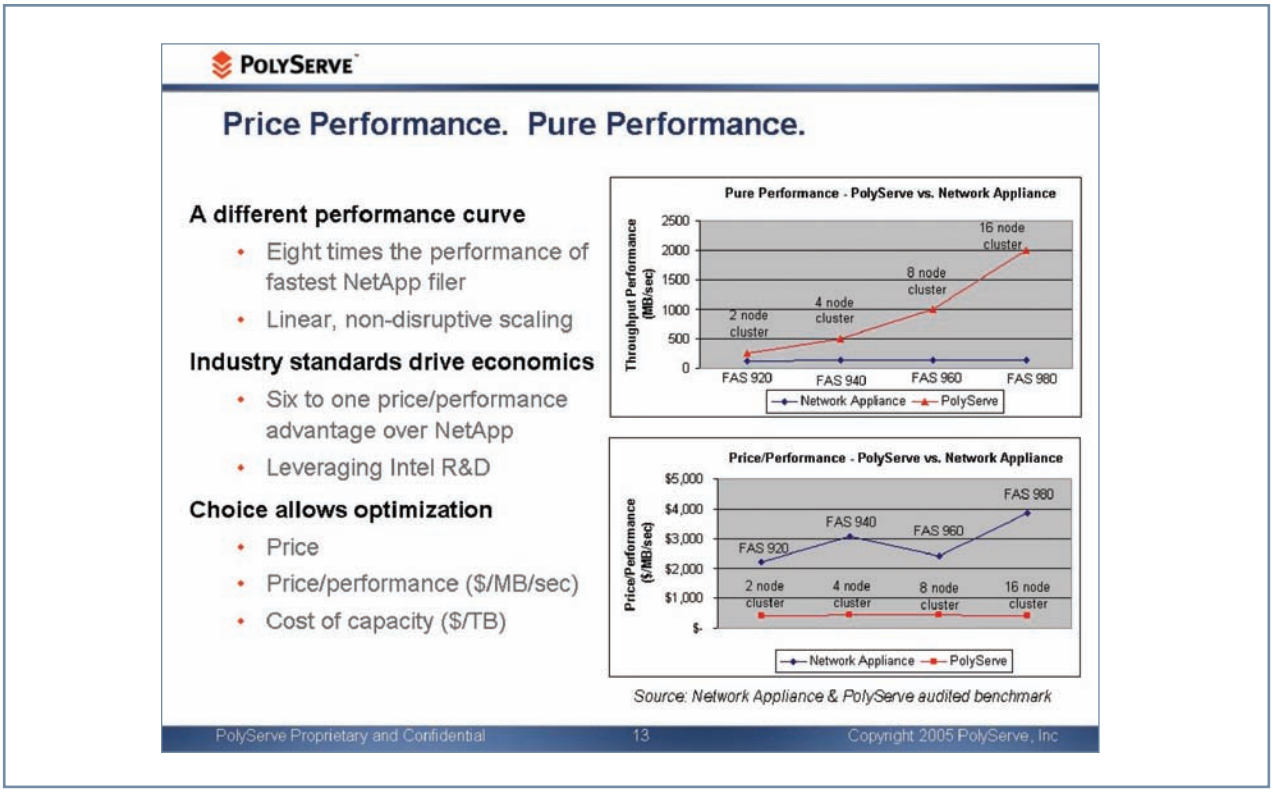

Figure 6: Price Performance. Pure Performance.

environment. In demonstrating environments using standard servers, aggregating multiple servers together has delivered eight times the performance of the largest reported filer average, as shown in Figure 6.

Entirely standard hardware provides high levels of absolute performance in a very cost-effective way. The PolyServe File Server Utility can address larger problems that require more capacity at a lower price point.

Customer Case Studies

The following are two customers with very different goals and environments who have used PolyServe software to achieve those goals.

Fidelity Investments

One excellent example of a company implementing this approach is Fidelity Investments, the largest manager of mutual funds in the world. They have a very large internal IT environment to support their financial services business, and they have recently focused on reducing the complexity of that environment. They were eager to consolidate by almost half, their number of file servers.

Concurrently, they wanted to elevate the quality of service to the file serving clients in their organizations and simplify the operational requirements to enable staff to spend less time managing the facility.

As shown in Figure 7, PolyServe shared data clustering software was part of this solution.

Amerada Hess

Amerada Hess is a large oil company with a significant discovery division. One major challenge is handling and processing very large seismology datasets with large groups of servers that perform the numerical analysis. To feed this process, there must be very high bandwidth access to the large pools of data. Hess wanted to dramatically improve the level of performance, and achieved this by implementing a scalable UnAppliance approach, as shown in Figure 8. Over time, they developed an environment with 11 servers linked together as a single virtual UnAppliance. Hess has nearly 150 terabytes of data attached to 11 machines. In turn, there are 1200 servers that connect, using NFS, SYS, and similar protocols, to retrieve the data for processing from this HPC focused cluster.

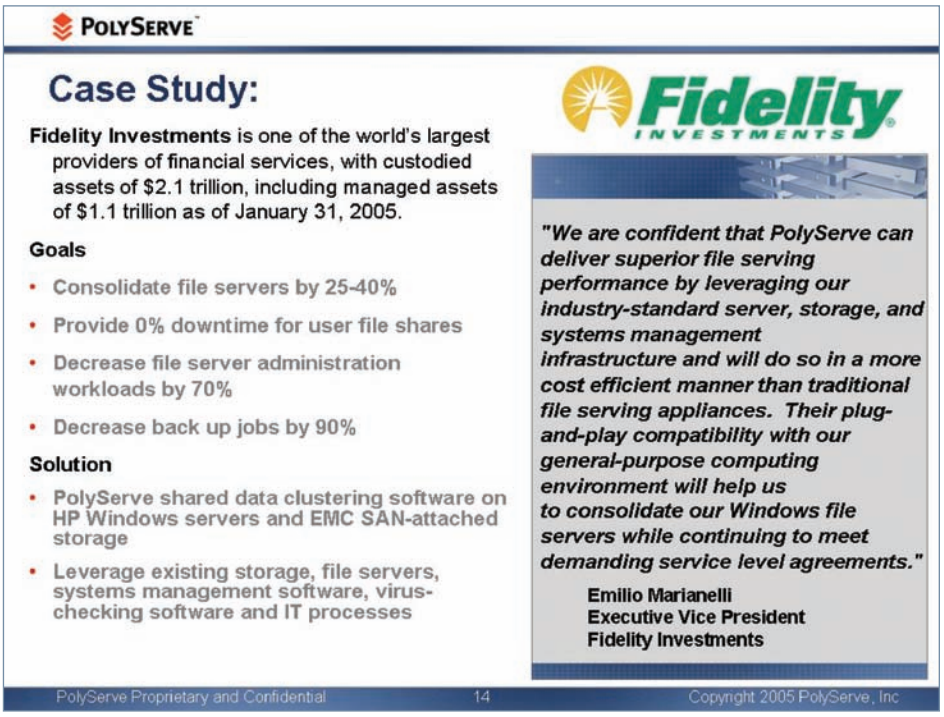

Figure 7: Case Study of Fidelity Investments

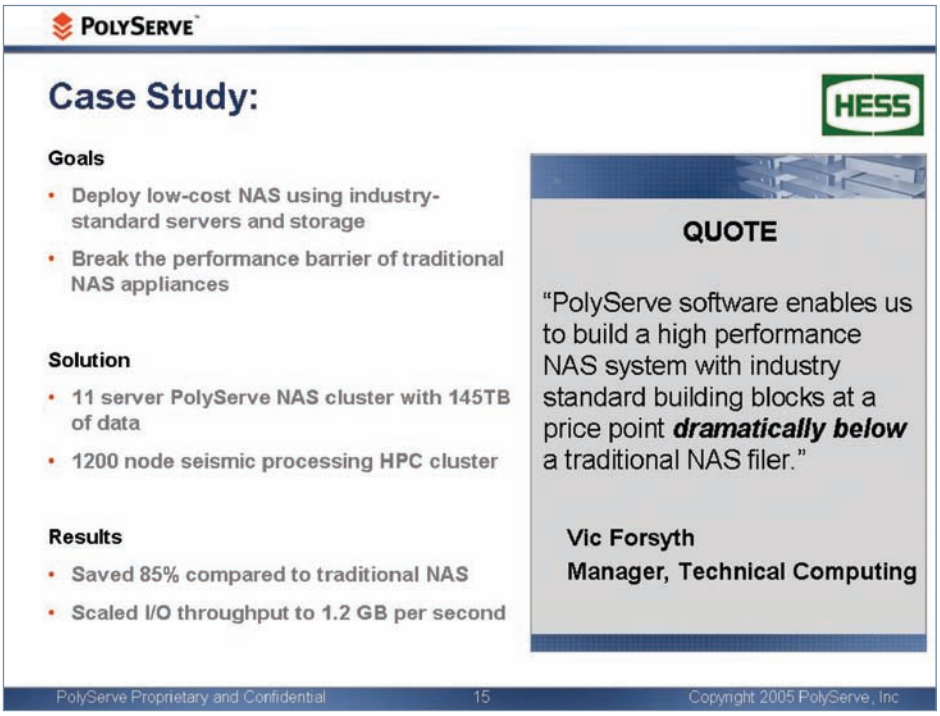

Figure 8: Case Study of Amerada Hess

Using the shared data approach, Hess was able to demonstrate that, as they scaled the performance of this environment, they could go from relatively lower levels of throughput. These levels of throughput were slightly comparable to what the company has seen in the past from appliances (up to 1.2 gigabytes per second), which was a multiple that they had never been able to achieve with their previous approach of adding servers into the environment. When Hess added those servers into an 11-node cluster they actually saw 85 percent better performance scalability.

Case Studies: Conclusion

These two clients demonstrate different goals. Fidelity is an example of consolidating and simplifying by going to the standard UnAppliance approach. Hess is an example of the need and fulfillment of carrying out bigger tasks than old approaches can handle.

How the UnAppliance Works

Shared Data Clustering

Many people have experience with clustering in the past, where, in a set of multiple servers, one server can take over for another server if it fails. In this traditional approach, whatever storage had been connected to the failed server must be reconnected, remounted, and rechecked.

Shared data is a very different approach, where all of the servers have access at all times to storage. With shared data, all servers can read and write files at the same time, in the common storage array, and in a common volume or alone. Software that enables these benefits must fully maintain data integrity and allow good scalability, while keeping the servers from interfering with one another.

The UnAppliance approach is based on a shared-data clustering technology that allows multiple servers to access the same files simultaneously, as shown in Figure 9.

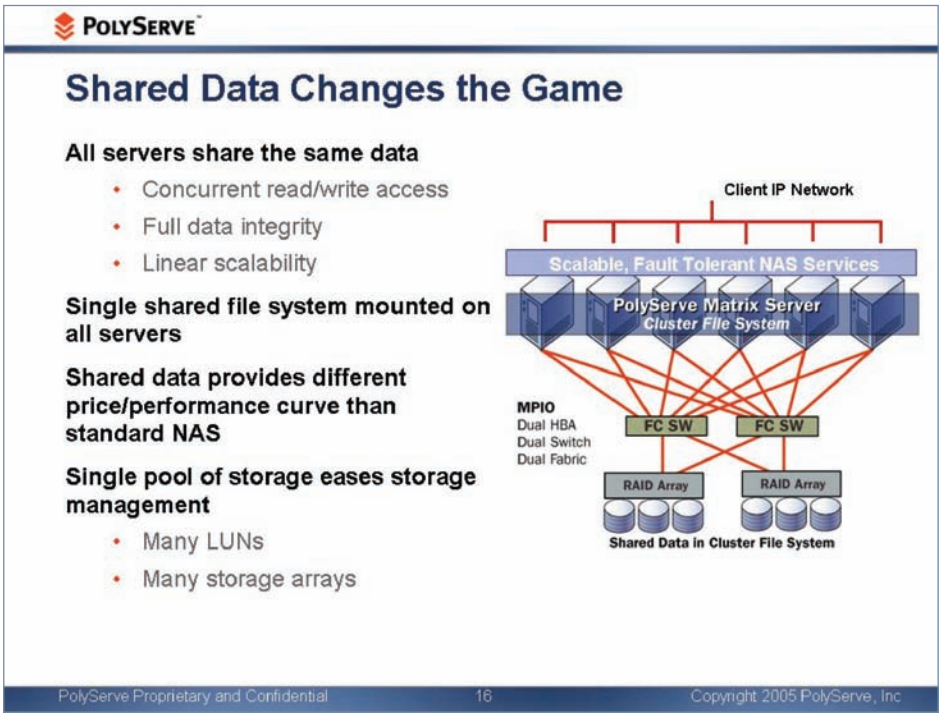

Figure 9: Shared Data Changes the Game

Overcoming the limitation of the traditional approach that allows for only one server at a time to access a given pool of storage opens up some very interesting possibilities:

- Providing high levels of performance by having multiple servers share data simultaneously, and doing it at a very low price point with completely standard servers, as in the case of Hess

- Dramatically simplifying an environment, as in the case of Fidelity, by handling multiple servers as a single entity, thus creating a single pool of storage managed as a single object

Components

The PolyServe Matrix Server contains:

- A cluster file system that allows multiple servers to access data simultaneously

- A high availability capability so that, if a server has failed – whether that failure was caused by hardware, networking, or software – the system locates another server in the environment to handle whatever clients were connected to the failed server and orchestrates that transition transparently

- The cluster volume manager, which allows multiple storage arrays to be put into a single pool that can be managed as a single unit

- Tight integration with the native implementations of NFS and CIFS

PolyServe does not provide its own implementation of networking software or file-serving software. PolyServe uses the standard implementations that integrate with all standard tools, and provides the ability to spread those capabilities across multiple servers in a transparent way.

The other important software component is the client's standard operating system, which can be either Linux or Windows. PolyServe supports the usual Linux implementations, RedHat and SuSE, as well as the usual Windows implementations, 2000, 2003, and, soon, R2. PolyServe software can use any 32 or 64-bit X86 Xeon or Opteron server in any kind of storage array networks for the shared storage environment underneath.

Client Allocation

As shown in Figure 10, new clients are distributed in a round-robin fashion across the multiple servers that are exporting the same data, resulting in each server only handling a subset of clients. The clients actually receive much higher levels of performance than they would have if they were all trying to run on one highly congested individual machine.

Cluster Management

Although there are multiple individual physical servers that make up an UnAppliance environment, they do not differ much from one another. They all have access to the same data, yet they can be configured to handle, for instance, different sets of clients. However, it is very simple to treat all servers in the UnAppliance environment as essentially the same thing: a single unit.

One PolyServe customer has more than 1000 servers set up with the UnAppliance approach. This customer has a good deal of experience running a larger number of servers, and normally allocates one person to manage 50 to 100 servers in their old configuration. Using PolyServe, only two people manage the environment. This customer stopped thinking in terms of servers, and started thinking about clusters, because every server in a given cluster is basically identical to the other servers in the cluster. As each cluster is made up of 10 to 16 servers, each administrator is managing about 50 clusters. That means more servers managed per administrator, but no more burden than with the old configuration of 50 servers per administrator.

The environment can be managed as if all of the servers were a single, virtual, logical entity, which can dramatically simplify the administrator's task. One specific example pertains to backups. In the case of Fidelity, the company wanted to reduce the number of individual backup jobs and, therefore, the number of backup jobs that need to be verified, tracked, and checked by a factor of 10. How could they do this? In this environment, each group of servers in a cluster making up an UnAppliance can be backed up with a single job, as shown in Figure 11.

With the UnAppliance approach, 10 servers do not need to be backed up with 10 separate jobs because

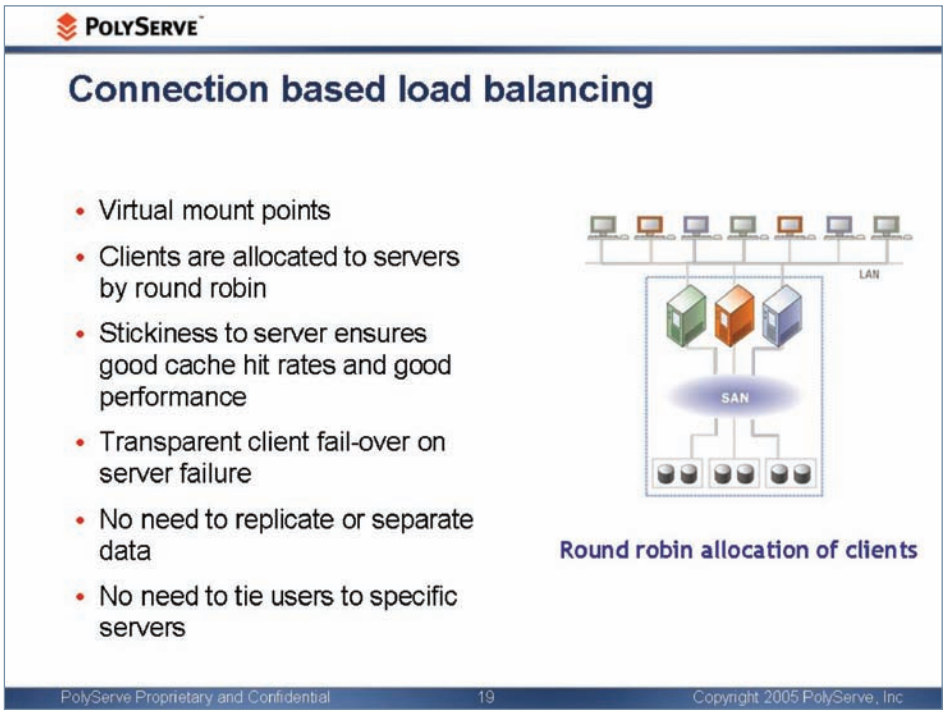

Figure 10: Connection-Based Load Balancing

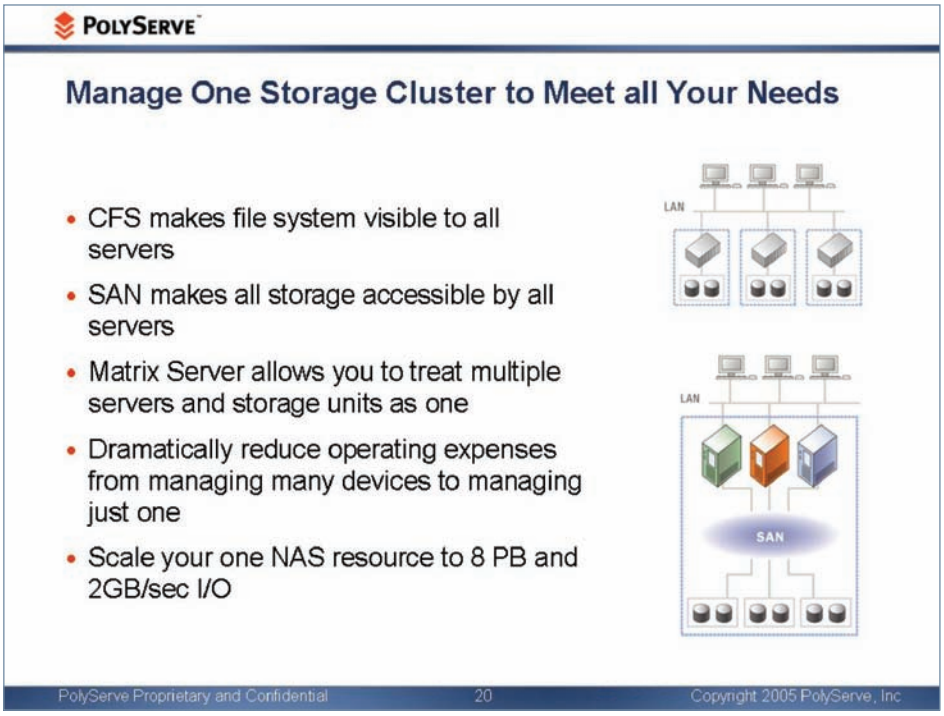

Figure 11: Manage One Storage Cluster

10 separate portals of data do not exist. One large pool of data can be backed up with a single job. In fact, there is no need for an installation of backup software on all the servers; the backup software can be installed on as few as one server in the cluster. That server can perform the backup of the entire shared data pool. Completely standard backup software for Linux or Windows servers is sufficient for the task because this file system looks like any other software system.

Also, a server in this environment can be dedicated to backup, without providing file services to clients. With this configuration, the backup job does not contend for capacity with the file serving functionality. This isolates that load from the file serving workload and provides independent capacity for those two tasks, which can be a dramatic improvement in the operational environment for administrators.

Volume Manager

In addition to treating multiple servers as a single virtual file server, it may be valuable to treat individual, physical storage and storage arrays as a single pool of storage. Volume Manager allows this storage virtualization, as shown in Figure 12.

Volume Manager can aggregate free space or multiple storage arrays together, treating them as a single pool of storage for handling a given set of files. This can provide higher levels of performance by striping data automatically across multiple arrays. This can take pieces of storage on multiple physical storage arrays and automatically stripe a given file system or a collection of file systems across those arrays and thereby have much higher overall levels of read and write performance.

Management Console

The PolyServe management console allows administrators to view all the hardware of the environment from a single place, and allows them to manage the hardware in a simple, unified way. As shown in Figure 13, the basic console shows each server – where each column represents an individual server – that is part of an UnAppliance self-serving cluster. In this set of servers, there is also a set of different exports that are being exported,

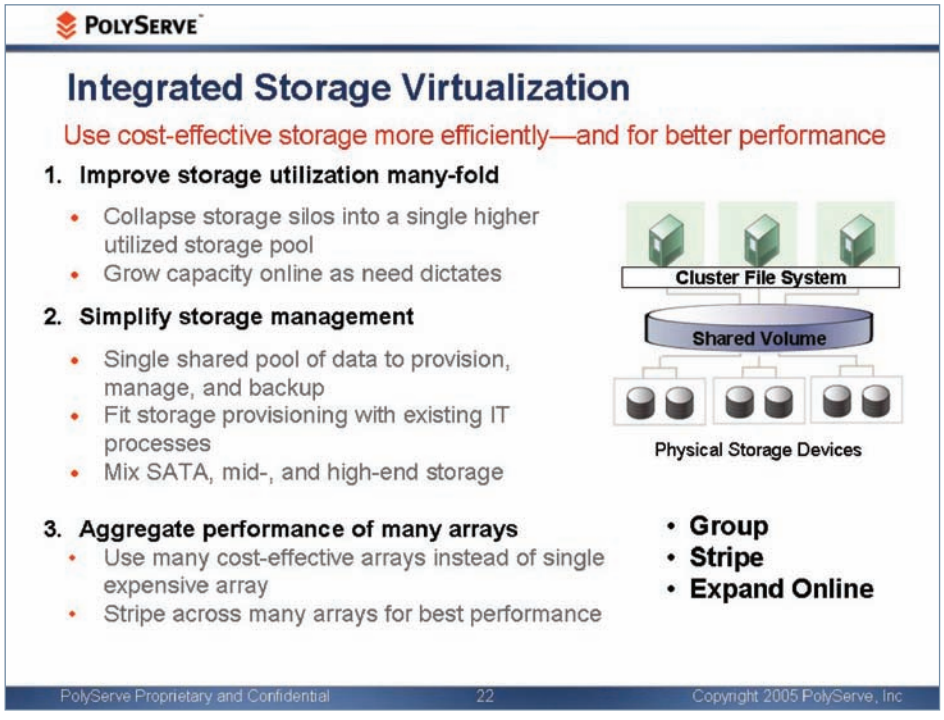

Figure 12: Integrated Storage Virtualization

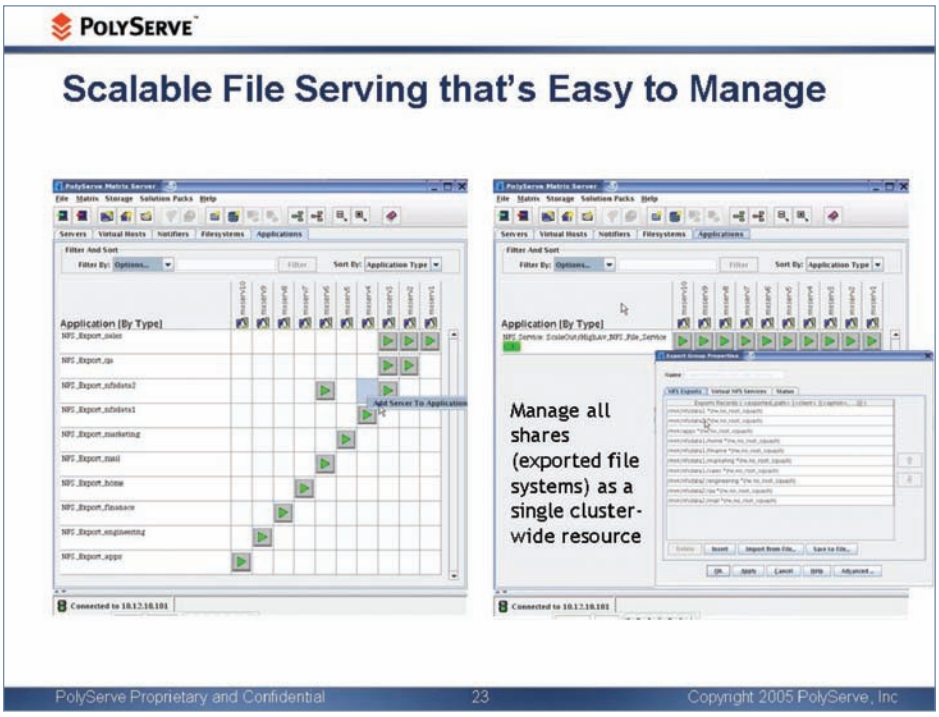

Figure 13: Scalable File Serving That's Easy to Manage

and the matrix cells show which exports are being run from which server.

Although the administrator in Figure 13 has chosen for the most part to export only a given piece of data from a given server, the top row shows three green arrows that indicate that a particular export is being made available from three servers at once. In the right window, a particular export is being made available across all servers in the environment using a dialog box that allows the management of exported data across the whole environment at the same time.

Conclusion

Figure 14 shows some of what the industry is saying about PolyServe and the UnAppliance approach.

The UnAppliance liberates data from being tied to one individual server. It scales performance as needed, and that scaling can be done with the most cost-effective hardware available. As new servers come out, going from single core to dual core to a future quad core, the UnAppliance allows clients to take advantage of the new hardware and build clusters that may include some old hardware and some new hardware without constraint.

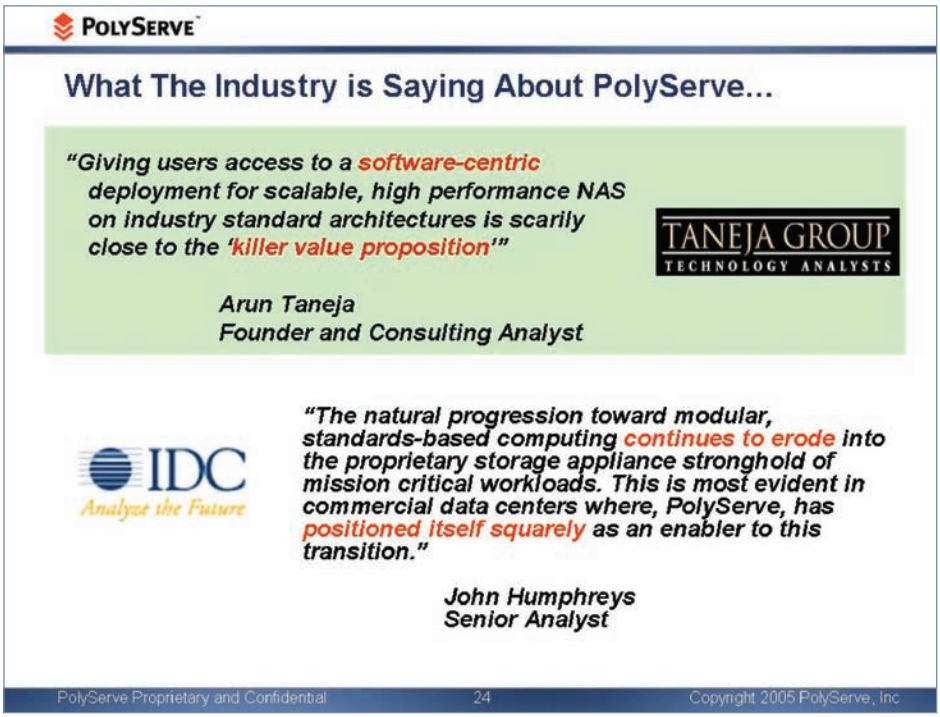

Figure 14: What the Industry Is Saying About PolyServe

Common Questions

Question: What vendors support PolyServe software?

Answer: Hewlett-Packard, IBM, and Dell all sell PolyServe software. HP also sells a product based on this software that is entirely pre-configured, sold as a unit with the software already set up: the HP Enterprise File Services Clustered Gateway (EFSCG). Some of the storage vendors, such as Data Domain, sell PolyServe software as well.

Question: What kind of servers and storage can be used with the PolyServe system?

Answer: Any server that runs either 32-bit or 64-bit X86; Xeons, Opterons, or similar servers from almost any vendor. Any kind of SAN storage works. Currently, PolyServe only supports fiber channel, but support for iSCSI will be added in the coming months.

Question: What about MSA1000?

Answer: MSA1000 works with PolyServe software; several customers use it. Direct connections should work as well. There are some configuration details, but it can work very well.

Question: Can workload be moved from one platform to a different kind of platform, as long as the hardware is supported by PolyServe?

Answer: Yes. For example, workload can be moved from an old Xeon server to a new dual-core Opteron server. Each server may be from a different vendor as well. There can be a mix of different kinds of hardware, from different vendors, in any given cluster.

Question: What kind of backup software can be used with PolyServe software?

Answer: Any kind of backup software that can run on Linux or Windows will work.

Question: Does running backup for the entire cluster from one server represent the risk of over-running backup windows and raising the level of congestion in the backup process?

Answer: It is true that having one server do backup means that the environment is only getting the performance of one server backing up. However, a server can be dedicated to backup only, so the backup is not competing for server capacity with other kinds of work like file serving, which can improve performance. If one server is not sufficient, the environment is in no way limited to having one server for backup. A backup job can be added to another server, as well as the backup server, providing two backup jobs for the environment. The important point is that backup is analogous to the file serving approach: if a given server is overloaded, another server can be added to receive load-improving performance.

Question: Would it be possible to use direct attached storages instead of SAN storage?

Answer: The basic approach of PolyServe is one where data are no longer associated with an individual server. The shared data approach relies on having servers able to access a common shared pool of storage at the same time. In beginning to support iSCSI, one of the expectations is that customers will use servers with direct attached storage (DAS), even storage inside the server box, and turn those into iSCSI storage arrays using these software approaches. Those arrays can be used as building blocks for building up a larger UnAppliance-type file serving cluster. The ability to integrate direct attach storage in this approach is coming in the near future.

About TechTarget

We deliver the information IT pros need to be successful.

TechTarget publishes target media that address your need for information and resources. Our network of industry-specific Web sites give enterprise IT professionals access to experts and peers, original content and link to relevant information from across the Internet. Our conferences give you access to vendor-neutral, expert commentary and advice on the issues and challenges you face daily. Practical technical advice and expert insights are distributed via more than 100 specialized e-mail newsletters and our Webcasts allow IT pros to ask questions of technical experts in real time.

What makes us unique?

TechTarget is squarely focused on the enterprise IT space. Our team of editors and network of industry experts provide the richest, most relevant content to IT professionals. We leverage the immediacy of the Web, the networking and face-to-face opportunities of conferences, the expert interaction of Webcasts and Web radio, the laser-targeting of e-mail newsletters and the richness and depth of our print media to create compelling and actionable information for enterprise IT. For more information, visit www.techtarget.com.

PolyServe_02_2006_0007

Sponsored by PolyServe™ Server and storage clustering software