SearchWinComputing.com E-Guide

Rein In File Server Proliferation: The Guide to File Virtualization

Times are changing and file virtualization is poised to go mainstream. Get up to speed on File Virtualization with this handy E-Guide from TechTarget and Brocade.

Sponsored By: BROCADE

TechTarget
The Most Targeted IT Media

SearchWinComputing.com E-Guide

Rein In File Server Proliferation: The Guide to File Virtualization

Table of Contents:

Windows drives storage virtualization trend

By Joan Goodchild, News Writer

The growth of data coming from the vast number of Windows files has contributed significantly to a surge in popularity of file virtualization in enterprise storage, according to a recent study.

A study published this month by TheInfoPro (TIP) Inc., a New York-based consultancy, looked at purchasing and implementation plans among 155 storage professionals from the largest U.S. companies. Survey data for the study, titled Wave 7 of the Storage Management Report, was collected between January and March.

The survey asked storage professionals to rank in order of priority the technologies they plan to install in the near term and found that file virtualization ranked higher than block virtualization. This is a first, said Robert Stevenson, managing director of TIP's research.

File virtualization, which lets administrators create and manage application-specific storage pools, is less invasive and a more attractive option for storage management than block files, which are single pools of virtual storage. Block files are usually large and tied to specific applications, he said.

"File virtualization tends to be more software-based and is easier to implement," Stevenson said. "Block virtualization forces you to change your infrastructure." Regulatory compliance has spurred an upswing in network-attached storage (NAS) for businesses, he said, resulting in three times the number of NAS footprints over the last 18 months. Continued growth of data from Windows operating system files, given that it is the most prevalent OS, is a significant driver in the trend toward NAS and file virtualization, he added.

"With file virtualization existing outside of the actual data path, there's no dependency on a hardware vendor," said Chris Wolfe, a virtualization and storage expert. "You can switch around your storage infrastructure without any impact on your file-level virtualized resources."

Stevenson also pointed to numbers in his research that he said further reveal why file virtualization is outpacing block virtualization. When storage professionals were asked what they considered to be the most important piece of Information Lifecycle Management—which is the idea of managing data throughout its useful life—40% said data classification was the most important function in ILM.

"If you look at file virtualization products on the market, many are starting to incorporate data classification capabilities," Stevenson said. "That is one reason why we see file virtualization products exceed block virtualization in demand."

Rein in NAS with file virtualization

By Jerome M. Wendt

File virtualization appliances change the dynamic of how organizations manage their network file servers. For example, the appliance aggregates stranded storage capacity on network file servers, simplifies file migrations among the network file servers and virtualizes network file shares. However, these benefits depend heavily on the underlying architecture of the file virtualization appliance. Key attributes such as high availability, the ability to scale capacity and performance, and integration with third-party indexing and search engines are handled differently by the various file virtualization appliances.

Typically, organizations initially bring file virtualization appliances into their storage environments to manage file servers more efficiently. Approximately three years ago, Ubicom Inc., a developer of communications and media processor and software platforms in Sunnyvale, CA, was running out of space on its Network Appliance (NetApp) Inc. filer. However, Jim Poehlman, Ubicom's chief information technologist, didn't have the $100,000 required to upgrade the filer. While he had other network filers available with excess capacity, he needed to migrate files in real-time without disrupting production applications because downtime would cost the company tens of thousands of dollars per hour in lost productivity and missed service-level agreements.

Poehlman deployed NeoPath Networks' File Director, which "didn't require us to reconfigure our network or bring our applications down," he says. "NeoPath Networks also gave us a price break by selling it to us for $15,000, though at the time we bought it they were still in beta."

Kevin Hayes, technology security manager for the City of Jacksonville, FL, also needed to recapture stranded storage capacity. Hayes had approximately 15 Microsoft Corp. file servers with varying amounts of excess capacity. As part of the reconfiguration, Hayes needed to improve application file performance, an issue he couldn't resolve with a simple file migration. To address these issues, he chose Maestro File Manager from Attune Systems Inc., which allowed him to create one logical file share that aggregated and virtualized filer volumes across different network filers. "Migrating files to one virtual file share allowed me to stripe files across my entire network file server farm...and increase application file performance," says Hayes.

Ibis Consulting Inc., which provides electronic discovery and compliance solutions for corporations involved in litigation, had a different set of requirements. As part of its discovery process, the Providence, RI-based firm receives large amounts of data from multiple clients in different file formats that need to be uploaded to its NAS storage. Ibis then culls and classifies files from its different clients and moves the files to different storage tiers for attorney review.

Shane Lennon, vice president of strategy and market development at Ibis Consulting (now a Pitney Bowes company), selected an Acopia Networks Inc. ARX6000 because it helped to "track and place files across multiple NAS systems and a large number of CIFS shares without causing a bottleneck to the Ibis process."

Review of file virtualization appliances

File virtualization appliances

		APPLIANCE						VIRTUALIZATION				MIGRATION					
Vendor/Product	List price*	Appliance OS	CLUSTERING		Custom hardware	Ethernet ports Min./Max.	In-band	PROTOCOLS		Global name-space	Policy-based storage tiering	Max. number of files supported	REPLICATION		Performance thresholds	Preserves file security	Scheduling
			Supported	Max. nodes				CIFS	NFS				Asynch	Synch			
Acopia Networks Inc. ARX500	$24,995	Proprietary	✓	2	✓	2/2	✓	✓	✓	✓	✓	128 million	✓		✓	✓	✓
Acopia Networks ARX1000	$55,000	Proprietary	✓	2	✓	6/6	✓	✓	✓	✓	✓	384 million	✓		✓	✓	✓
Acopia Networks ARX6000	$175,000	Proprietary	✓	2	✓	6/24	✓	✓	✓	✓	✓	2 billion	✓		✓	✓	✓
Attune Systems Inc. Maestro File Manager	$44,995	Windows Embedded Server 2003	✓	2		10/10	✓**	✓	✓	✓	✓	200 million			✓	✓	✓
Brocade Communications Systems Inc. StorageX	$10,000***	Windows Server 2003	✓	2		1/1		✓	✓	✓	✓	No limit	✓	✓		✓	✓
EMC Corp. Rainfinity Global File Virtualization	$80,000	Proprietary Linux	✓	2	✓	4/12	✓**	✓	✓	✓	✓	No limit	✓	✓	✓	✓	✓
NeoPath Networks File Director 220	$45,000	Proprietary Linux	✓	2		6/6	✓	✓	✓	✓	✓	Acts as a switch				✓	✓
NeoPath Networks File Director 7200	$90,000	Proprietary Linux	✓	2		4/4	✓	✓	✓	✓	✓	Acts as a switch				✓	✓

*For one appliance with hardware and software
**Can operate in-band or out-of-band
***$10,000 is the price for the software; users will need to provide the platform (Windows server, OS and, if desired, clustering software), which raises the price of the total solution by another $15,000 to $20,000

BROCADE

TAKE CONTROL

Take a more automated approach to file data management. Brocade File Area Network (FAN) solutions enable you to structure your unstructured file data. Now you can consolidate, migrate, and manage your file data like never before.

To get the complete story on how to centralize management of your unstructured file data, download *INTRODUCING FILE AREA NETWORKS*, a free 250-page eBook from Brocade, at www.brocade.com/bookshelf.

File area networking (FAN) to unify disparate network-attached storage (NAS) resources

By Chris Evans, Contributor

File area networking (FAN) is a relatively new concept that represents a collection of integrated resources and services used to provide a single file-based network. Most companies will have network-attached storage (NAS) resources dispersed around their enterprises. A FAN can bring these disparate resources together and provide your customers with added value—but how?

Move NAS resources to a unified address space

Most organizations will have implemented many NAS servers over the years. Typically, new servers are deployed to replace legacy equipment. However, the legacy equipment is kept longer than expected. As a consequence, many big companies will have NAS shares located in multiple datacenters with inconsistent naming standards. In addition, business units may have implemented their own NAS infrastructures.

The solution is to implement a unified address space (UAS). FAN technology can be used to plug existing NAS resources into a single unified file system address space. This not only has the benefit of enabling NAS resources to be easily located, it also positions the environment for utilizing other FAN functionality, which may include planning for future upgrades and expansions.

You should work with your customers to move their NAS environments into a single unified address space. Many products on the market offer this, either by providing a virtual access point that manages all storage requests, or by integrating with Active Directory and Microsoft's Distributed File System (DFS). Look at your customer's NAS infrastructure requirements. Do they require integration with Active Directory or are they a major network file system (NFS) user?

Optimize FAN storage

Once data has been brought together under a unified address space, you have options to optimize and manage FAN storage more effectively. Once data is managed under UAS, a number of benefits may be exploited.

Migration: FAN technology enables data to be mapped to a UAS and migrated transparently between NAS devices. This can be used for hardware consolidation, to replace aging legacy equipment, and to manage workloads and performance across the enterprise.

Replication: FAN enables data replication between NAS resources that may be located in the same or remote datacenters. Through a FAN infrastructure, replication can be used to improve data availability and user response times.

Storage tiering: File-based data is ideal for the application of information lifecycle management, as it already has many attributes associated with it and most data is usually written once and reread infrequently. Under a FAN

infrastructure, data can be moved transparently between tiers, ensuring data is located on the most cost-effective hardware platform.

Data availability: In conjunction with replication and migration, a FAN can be used to improve the availability of user access to file-based data. This is achieved through redundant FAN hardware that can be deployed in geographically dispersed datacenters.

Developing FAN services and solutions

Clearly FAN offers significant functionality over normal file-based NAS offerings. Once a unified file system is in place, you can offer your customers additional features to help them improve their existing NAS infrastructure. Look to evaluate and target key NAS issues that can be addressed with FAN, which may include:

- Retiring legacy equipment
- Implementing a tiered NAS infrastructure
- Implementing a highly available NAS infrastructure
- Implementing a unified NAS security policy
- Reducing storage costs
- FAN is an emerging technology. Positioning for a FAN infrastructure now can reap rewards for your customers in the future.

How to get the most from Windows storage consolidation

By Rick Cook

With the release of Windows Server 2003, Microsoft has made it significantly easier to consolidate storage. But getting the most out of a storage consolidation project requires some thought and planning.

Here are some points to keep in mind as you embark on combining your data into a more manageable and efficient arrangement.

Tip 1: Plan an "invisible" process
As usual, many of the critical issues revolve around organizational and managerial acceptance rather than the technology. The benefits of storage consolidation, such as reduced administrative workload and better use of storage, flow almost entirely to the IT department. Any negative effects, such as performance reductions because of network congestion or running low on storage or service disruptions, are going to be felt by everyone. Because of the potential negative effects on the rest of the enterprise, it is important to make sure that the storage consolidation process proceeds invisibly to anyone outside IT.

Tip 2: Expect circuitous improvements
While simply moving storage onto fewer arrays behind fewer servers provides basic benefits, there are other advantages to a properly implemented storage consolidation project. One big one is process improvements.

For example, consolidating storage eases storage management because there are fewer platforms to manage. However, Windows Server 2003 also makes it much easier to implement advanced storage management features from Shadow Copy to facilitating remote storage. While you're consolidating storage, it makes sense to make intelligent use of these new features as well.

At the policy level, storage consolidation offers an ideal opportunity for changes like establishing a consistent platform build on all servers and for establishing a consistent methodology for adding storage capacity. Such changes will help to hold down storage administration costs in the future.

Tip 3: Identify and reinforce potential weaknesses
Moving storage to fewer, centrally located servers offers many advantages, but it also puts some new demand on your IT infrastructure.

For example, storage consolidation increases the importance of network reliability and throughput, especially if you are moving from servers with direct attached storage to a SAN-NAS environment. It's important for you to examine your storage-related networking to make sure it is robust and flexible enough to handle the new demands.

Tip 4: Make haste…slowly

Storage consolidation doesn't have to be done all at once. In the beginning, it makes sense to "cherry pick" your projects by looking for areas where storage consolidation will provide big payoffs for relatively modest efforts.

As you progress you can refine your infrastructure and policies to get the most out of storage consolidation for your organization.

GET YOUR LIFE BACK

AUTOMATE DFS-RELATED TASKS WITH BROCADE STORAGEX

If you use or plan to use Microsoft Distributed File System (DFS), you'll love Brocade StorageX.

Brocade® StorageX® takes a good thing and makes it even better—building on DFS to give you the industry's most comprehensive file management solution. It can help you automate key tasks such as:

- Easily migrating and consolidating data
- Optimizing backup/replication at remote offices
- Providing high availability for critical data

Brocade StorageX provides a wide range of additional capabilities, such as:

- Building and managing enterprise-class DFS roots
- Integrating with both CIFS and NFS
- Integrating with and managing NetApp filers
- Auditing actions, operations, and administrative policies
- Integrating logical-physical file views
- Supporting namespace backup/restore
- Automating creation of the DFS consolidation root
- Replicating to shared cluster storage
- Providing local group processing for files and directories
- Automating storage tiering

"Brocade StorageX is a simple, elegant solution to the complex problem of file system management. It merely rides on top of what is already there. The price is such that a quick, back-of-the envelope ROI calculation should yield a solid multiplier."

— Michael Fisch, Director of Storage and Networking, The Clipper Group

TRY BROCADE STORAGEX YOURSELF!

E-mail bcrgroup@brocade.com today to get your free evaluation download!

BROCADE

Replication upgrades make DFS ideal for low-bandwidth branch offices

By Laura Hunter, Contributor

While the Distributed File System (DFS) has been around since the days of Windows NT 4.0, improvements made to DFS in Windows Server 2003 R2 make it much more attractive for Windows systems administrators. In particular, there have been advances in the areas of replication and file availability that can make DFS an ideal solution for low-bandwidth branch office environments.

In Windows Server 2003 R2, the DFS service has been split into two components: DFS-Namespace (DFS-N) and DFS-Replication (DFS-R). The DFS Namespaces feature in R2 offers three updated capabilities: target priority, client failback and delegated authority. While these three capabilities are quite useful, the most exciting advances for DFS in R2 are in the realm of DFS-Replication (DFS-R).

It is here that the new DFS really begins to shine, with a new replication algorithm that provides incredible performance gains for bandwidth-challenged environments such a branch office separated from its corporate headquarters by a low-speed or heavily-utilized WAN link.

Prior to R2, DFS used the File Replication Service (FRS) to replicate files between multiple link targets. FRS is the service used to replicate the information that's stored in the Active Directory SYSVOL share: logon/logoff scripts and Group Policy Objects. FRS uses RPC over TCP/IP to replicate files within a single site as well as between sites; FRS creates its own replication topology with its own schedule and connection objects that are controlled separately from Active Directory replication.

FRS will trigger replication whenever a file is closed, with changes held in a 3-second aging cache to allow for files that are being changed frequently. Once this 3-second "waiting period" is up, the FRS service on the server hosting the changed file notifies its FRS replication partners, and the file gets replicated across the FRS replication topology. For small files stored on lightly utilized servers, this process works quite well. But because FRS traffic is not compressed even when traversing site boundaries, replication of large files can be a tricky process. Often, replication errors or inconsistencies are created.

DFS in R2 changes all of this by introducing a new replication algorithm called Remote Differential Compression (RDC). RDC functions by breaking files up into small "chunks"; it will then only replicate the individual chunks of a file that have changed from one replication cycle to the next. For example, consider a Microsoft Word file that contains the line of text: The quick brown fox jumps over the lazy white dog.

If someone changed the first few words of the sentence to read "The slow black fox…" RDC would replicate only that particular chunk, rather than sending the entire .DOC file across the wire. RDC does this by computing MD4 hashes of these small chunks of files, then comparing those hashes between servers that are attempting to replicate.

If a particular chunk of a file has changed, then the MD4 hash of that chunk will change while the hashes for the remaining chunks in the file remain the same. This allows an RDC-enabled server's replication partner to request only those chunks whose hashes have changed since the last time replication took place. For larger files that only need to replicate small changes, this provides a drastic reduction in replication time and improved performance for your users.

Imagine a 4MB Visio document where you need to change the title of one or two sections. In the FRS world, this would prompt the entire 4MB file to replicate, whereas RDC will only need a few seconds to replicate the changed sections of the file. For environments with branch offices to support, particularly where bandwidth is at a premium, DFS-R in R2 can more than justify making the move to the new operating system.

> NOTE: Because DFS Replication is triggered on file close, it's not good for replicating files that are always locked and in use, like in the case of a database or another file that's used by an "always-on" service. Nor can you use DFS-R to replicate the Active Directory SYSVOL share: Logon scripts and Group Policy Objects still need to be replicated via FRS. However, FRS and DFS-R can co-exist quite comfortably on the same server.

More replication improvements for DFS in Windows Server 2003 R2

By Laura E. Hunter, Contributor

While the Distributed File System (DFS) has been around since the days of Windows NT 4.0, improvements made to DFS in Windows Server 2003 R2 make it much more attractive for Windows system administrators.

In a previous article I discussed how advances in the areas of replication and file availability in particular make DFS an ideal solution for low-bandwidth branch office environments. In that article we focused on target priority, client failback and delegated authority. In this article we'll look at some of the other improvements in DFS-Replication (DFS-R) in Windows Server 2003 R2:

Bandwidth throttling and replication scheduling. To further enhance your control over the use of your bandwidth, you can specify replication schedules similar to those you'd set up between sites in Active Directory. You can specify these schedules for an entire replication group, or create a custom schedule for an individual replication connection. You can also put a cap on the amount of bandwidth that DFS-R replication can take up.

Support for replication groups. You can configure one or more sets of data and servers as a replication group with a common configuration for replicated folders, replication schedule, and bandwidth throttling. Each DFS server can support a maximum of 256 replication groups, and each of these groups can contain up to 256 replicated folders.

Collecting data for backup purposes. You can also use replication groups to collect data from branch sites to perform centralized backups. Rather than relying on remote sites to maintain their own backup hardware and perform their own backups, you can create a separate replication group to replicate their data to a central location. By disabling replication from the hub site back to the branch server, you'll create a "one-way" replication agreement that will prevent any inadvertent changes made at the backup site from replicating back to the remote server.

> NOTE: DFS-R can replicate data across multiple forests within the same forest; you're not restricted to replicating within a single domain.

Cross-file RDC. This takes the performance improvement of RDC to the next logical level. Say you have a file stored in a DFS namespace called "2006 Board of Directors.doc," detailing the names and biographical information of your company's board for that year. You need to create a similar file for the 2007 board, so you do a File/Save. As on the 2006 file, saving it as "2007 Board of Directors.doc" and making a few changes to reflect two new board members.

Now there's a new file that needs to be replicated within the DFS namespace…but is it really brand-new? By using cross-file RDC, DFS can use the contents of the 2006 Board of Directors file to seed replication for the new file, using the "chunking and hashing" method we've already described to only send over the wire the information that's

different between the two files. (This feature is possible because comparing the MD4 hashes created by two files is far more efficient than comparing the actual contents of the files.)

File and subfolder filters. You can specify individual sub-folders or filenames that should not be included in DFS replication, either by explicitly listing the name of the file or folder or by using the * wildcard symbol. By default, DFS-R will not replicate any folder that begins with the tilde (~) character, as well as any files with a .TMP file extension.

The following files and file types will always be excluded from DFS Replication:

- Any EFS-encrypted files.
- Any file that has had the temporary attribute set.
- Any reparse points used by Single Instance Storage or Hierarchical Storage Management. (The reparse points used by DFS itself are not affected by this.)
- Any NTFS-mounted drive paths where you've added a new drive to a system and assigned its space as a folder within an existing drive letter, rather than assigning it a drive letter of its own.

Resources from Brocade

Brocade File Virtualization & FAN Resources: Free Software Downloads, Product Demos, White Papers, Case Studies, Analyst Papers, and Webcasts

About Brocade

Brocade® (Nasdaq: BRCD) provides the industry's leading platforms, solutions, and services for intelligently connecting, managing, and optimizing IT resources in shared storage environments. By developing innovative Storage Area Network (SAN) and File Area Network (FAN) solutions, Brocade helps organizations increase efficiency, reduce costs, and maximize their data assets.