Managing the information that drives the enterprise

STORAGE

new tools for better backups

Next-generation backup tools for virtual machines and data deduplication increase data protection.

what's inside:

TechTarget
The IT Media ROI Experts

Backup for the 21st century

By Rich Castagna

WE HAVE THE SURVEY DATA to prove it, but just ask any group of storage managers what their main concern is and you're likely to hear a chorus of "backup." For those lucky enough that backup isn't front of mind, it's probably running a close second to whatever tops their list.

What's most troubling is that the backup dilemma has a history—and a pretty long one at that. For most shops, it was also job No. 1 last year and the year before and so on.

But this cloud has a silver lining. Twenty-first century backup tools have filtered into the storage ecosystem and, in the slowly evolving storage market, are starting to take hold. There are now mature, practical alternatives to backup as we have known it, and products that can augment traditional backup systems. The introduction of disk into the backup process has served as the catalyst to this reformation. It's the archetypal game changer.

Technologies like virtual tape libraries, continuous data protection and data deduplication are poised to forever change how we do backup. And, even better, they can make restores an almost sure thing, putting to bed forever the old joke: "Our backups run fine—it's the restores that don't work."

Getting these new backup technologies off the drawing board and into the data center has been an arduous process.

But getting these new backup technologies off the drawing board and into the data center has been an arduous process. Initially, many of these tools were point products offering the unpleasant prospect of managing yet another app in an already harried backup environment. And a fair number of early entries were limited by scalability or performance issues.

Most of those stumbling blocks have been overcome, but the fear of upsetting the current backup process, even if it's only marginally workable, remains. The truth is, if you haven't hit the backup wall yet, you will.

That's why we put together this special collection of articles focused on new backup technologies and techniques. With capacities growing at remarkable rates, it's time to get off the backup treadmill. ⊙

Rich Castagna is Editorial Director of the Storage Media Group.

ALTERNATIVE THINKING ABOUT VIRTUAL STORAGE:

VIRTUALIZE STORAGE NOW.

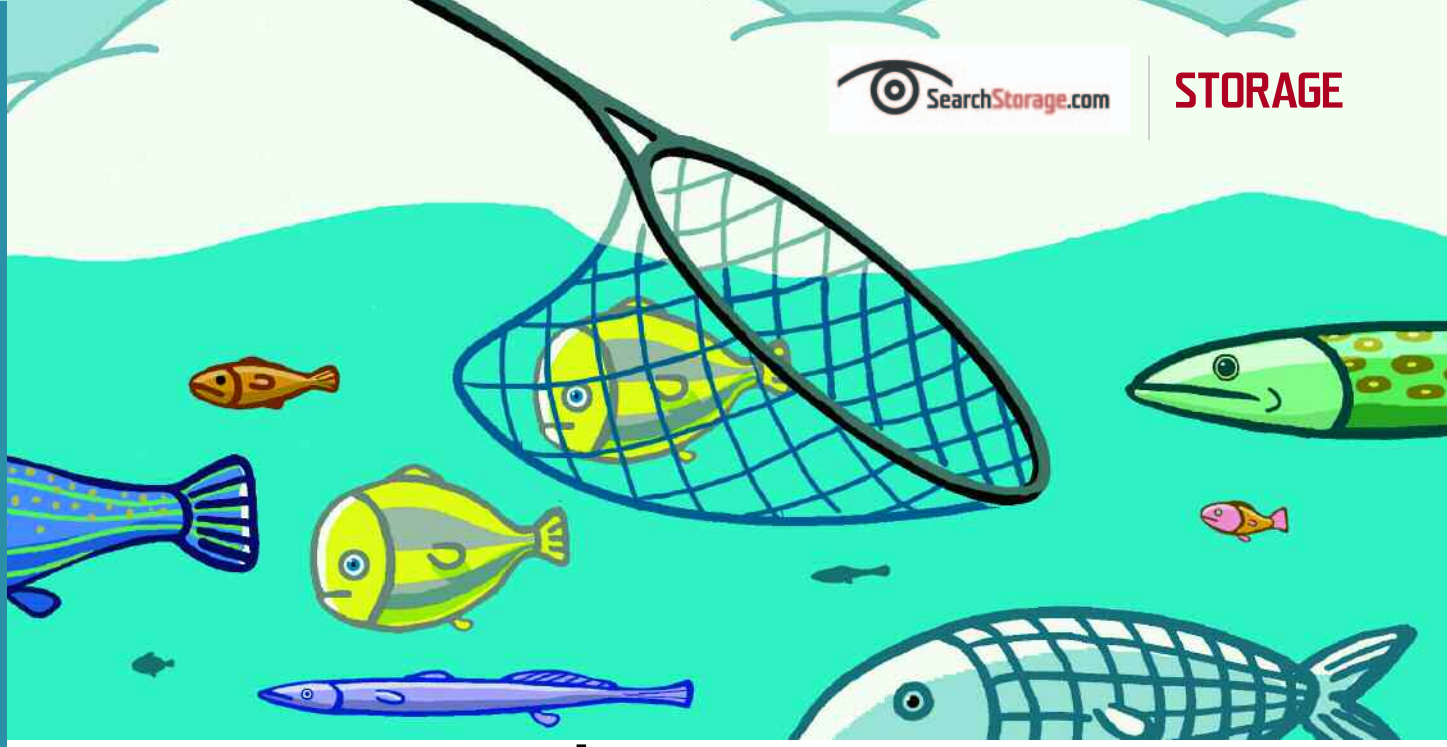
A powerful business innovation in data storage is now within your reach. The new HP StorageWorks 4400 Enterprise Virtual Array is here. It virtualizes up to 96TB of storage—across numerous storage servers and platforms—simplifying storage management and speeding access. Less limitations. More freedom. Technology for better business outcomes.

HP STORAGEWORKS EVA4400

Up to 96TB virtual storage capacity.

- Enterprise-class performance
- Over 30% better capacity utilization*
- Up to 75% less time needed to configure and manage*
- Easy application integration

Now's the time for virtual storage.
Visit hp.com/go/virtualstorage



catch up with dedupe

Deduplication backup products differ in how they recognize and reduce duplicate data. Here's how to pick the product that will best fit into your environment.

By Jerome M. Wendt

BACKUP HAS SEEN THE FUTURE, AND IT IS DISK. As the backup target gradually, but persistently, changes from tape to disk, data deduplication is becoming a key component of the backup process. Because vendors implement deduplication differently, the fear, uncertainty and doubt surrounding deduplication products has increased as have the questions about when to deploy what product.

Deduplication resides in the backup process in two primary places: backup software and disk libraries. Asigra Inc.'s Televaulting, EMC Corp.'s Avamar and Symantec Corp.'s Veritas NetBackup PureDisk are backup software products that deduplicate data at the host level, minimizing the amount of data that needs to be sent over corporate networks to backup targets or replicated to disaster recovery sites. Disk libraries from Data Domain Inc., Diligent Technologies Corp., Quantum Corp. and Sepaton Inc. deduplicate data at the target, which allows companies to deploy disk libraries without disrupting current backup processes.

Illustration by Alison Seiffer

With the underlying deduplication algorithms essentially the same across both sets of products, the real issues are how each product implementation impacts performance and data management in the short and long term. Neither approach is yet ideal for all backup requirements, so a crossover period is emerging in which some storage managers will likely use backup software and disk library methods for specific needs. Hidden issues like undeduplicating data to store it on tape, integration with enterprise backup software products, and the ability to selectively turn off deduplication to accommodate specific compliance requirements and pre-existing encryption conditions should be evaluated closely to determine whether those issues outweigh the benefits of deduplication.

DATA REDUCTION AND COMPRESSION ALGORITHMS

Backup software and disk library products deduplicate data in similar ways, with most using a combination of data-reduction and compression algorithms. Both types of deduplication approaches initially identi-

HOW TO ESTIMATE YOUR DEDUPLICATION RATIO

REDUNDANT DATA. The more redundant data you have on your servers, the higher the deduplication ratios you can expect to achieve. If you have primarily Windows servers with similar files and/or databases, expect to achieve higher ratios of deduplication. If your servers run multiple operating systems and different files and databases, expect lower deduplication ratios.

RATE OF DATA CHANGE. Deduplication ratios are related to the number of changes occurring to the data. Each percentage increase in data change drops the ratio; the commonly cited 20:1 ratio is based on average data change rates of approximately 5%.

PRECOMPRESSED DATA. Data compression is a key component in every vendor's data-reduction algorithm. Vendors base their advertised data-reduction ratios on the premise that compression will reduce already deduplicated data by a factor of 2:1. In a case where data dedupe achieves a 15:1 ratio, compression could take that ratio up as high as 30:1. However, users with large amounts of data stored in compressed formats such as jpeg, mpeg or zip aren't likely to realize the extra bump that compression provides.

DATA-RETENTION PERIOD. The length of time data is retained affects the data-reduction rate. For example, to achieve a data-reduction ratio of 10 times to 30 times, you may need to retain and deduplicate a single data set over a period of 20 weeks. If you don't have the capacity to store data for that long, the data-reduction rate will be lower.

FREQUENCY OF FULL BACKUPS. Full backups give dedupe software a more comprehensive and granular view into the backup. The more frequently full backups occur, the higher the level of deduplication you'll achieve.

fy whether chunks of data or files are the same by first performing a file-level compare or using a hashing algorithm such as MD5 or SHA-1. Unique files or data chunks are preserved, while duplicate files or data chunks may be optionally rechecked. This recheck is done using a bit-level comparison or secondary hash to ensure the data is truly a duplicate and not a rare hash collision. This first stage in the deduplication process typically reduces data stores by factors of approximately 10 or more over time.

To achieve data-reduction factors of 20 times or greater requires the product to compress the unique deduplicated files or data chunks. To accomplish this, vendors use a lossless data compression algorithm, such as Huffman coding or Lempel-Ziv coding, which executes against the unique file or deduplicated data chunk. Compression squeezes out items like leading zeros or spaces to reduce the data to its smallest possible footprint before it's stored.

However, using deduplication at the source or target introduces performance and management issues. Backup software-based deduplication products introduce a heavy initial processing toll on the host. In addition, users should carefully examine how swapping current backup software with a deduplication backup product, or running two backup software products concurrently, will affect server and application performance, as well as their stability.

Conversely, deduplicating data on a disk library may require users to deploy multiple disk libraries to handle the performance overhead created during peak backup periods. This creates more management overhead as each disk library creates its unique deduplicated data store; admins must also manage and direct backup jobs to multiple physical disk libraries as opposed to a single logical one. Determining which backup software, disk library or combination of them to select, and under what circumstances, is how they handle these potential bottlenecks.

BREAKING THE BOTTLENECKS

Asigra Televaulting attempts to break the management bottleneck by taking an agentless approach that expedites deployments while minimizing user involvement. Users initially install the Asigra Televaulting gateway software on a Windows or Linux server. The Televaulting backup software accesses client files over the internal network using CIFS, NFS or SSH (SSH allows for security but is slower), and reads the files. As it reads each file, the Asigra Televaulting server performs a hash on the file. If the file is determined to be unique, the file is chunked up with its unique blocks stored, while redundant blocks are indexed and thrown away.

All hash processing takes place on the Asigra Televaulting server, which maintains a database of all of the unique file blocks on the different servers it's assigned to protect. Once the initial backup and index is done, subsequent server backups execute faster because they can use this common repository of unique blocks created from the first server's backup.

This approach still doesn't completely eliminate the performance toll of deduplication. By running the deduplication on a central server, the Televaulting software transfers the performance overhead from the client servers to the Televaulting server. Multiple servers with unusually large daily data change rates (more than 10%) or large numbers of servers (100 or more) needing to run backups at the same time could impact backup times and force the deployment of more Asigra Televaulting servers to manage the overhead.

EMC Avamar and Symantec Veritas NetBackup PureDisk take a slightly different approach to address the performance issue. They use agents that utilize computing resources on each client server to do the initial file hash. As part of this process, the agents communicate with the main backup server, which maintains a central database of the unique file hashes. As the Avamar or PureDisk agents on the servers hash the files, they check with the central server to see if the generated hash already exists. If the hash exists, the agent ignores the file; if it doesn't exist, it breaks the file into smaller segments and looks for new unique file segments to store. From that point, EMC Avamar and PureDisk deviate in their product implementation.

> Compression squeezes out items like leading zeros or spaces to reduce the data to its smallest possible footprint.

EMC Avamar allows server storage capacity to grow to approximately 1.5TB in size. Although Symantec Veritas NetBackup PureDisk servers can grow to manage nearly 4TB of PureDisk storage capacity, EMC Avamar uses segment sizes that are about one-fourth the size of PureDisk's. This allows it to better identify redundant data in files, asserts Jed Yueh, EMC Avamar's VP of product management. If users should need to grow in capacity and scale, EMC Avamar uses a redundant array of independent nodes (RAIN) clustering architecture. This allows organizations to add more server nodes into the RAIN cluster to increase server capacity and performance by striping the data across multiple nodes.

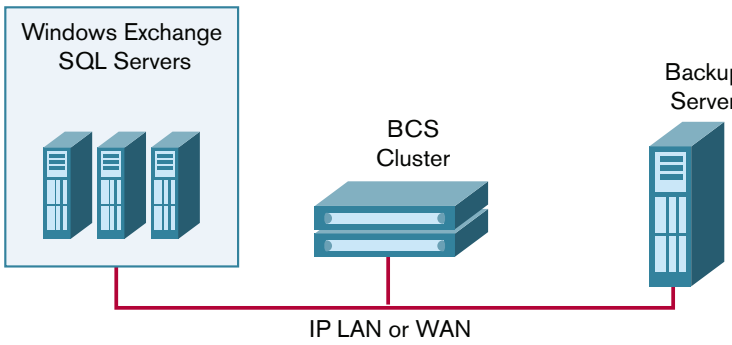
No Backup Windows.
No Data Loss.
No Corruption.
No Multiple Recoveries.
No Kidding.

ASEMPRA

In our 24x7 world, downtime and data loss are simply not options. You need a backup plan that provides complete data protection and availability of critical business applications.

Asempra Business Continuity Server™ (BCS) is your no-hassle solution. Record every Microsoft® Exchange, SQL, or Windows application data event in real-time, filling in the gaps left by traditional backup windows. Recover data in seconds or minutes, so your users are back up and running immediately. Plus, you get an all-in-one solution that addresses your backup, disaster recovery, and continuous availability needs.

Asempra BCS continuously captures and indexes events, ensuring 100% usable data on first recovery. Use the BCS to generate backups, eliminating server backup windows.



Asempra Business Continuity Server:
Continuous Availability and Data Protection for Microsoft Exchange, SQL and Windows

Find out for yourself how Asempra can make your life a lot easier.

Visit us on the web at www.asempra.com/backup/

ASEMPRA

Phone: 408.215.5800 • Email: info@asempra.com • www.asempra.com

In a PureDisk environment, a single server can manage 4TB of PureDisk storage and up to 100 million files, which equates, according to Symantec, to a little more than 80TB of source data. Additional servers can be added to expand PureDisk's storage capacity or to handle larger numbers of files.

PureDisk manages file meta data outside of the file system using MetaBase Server and MetaBase Engines. As an environment grows, a storage manager uses PureDisk to add new instances of MetaBase Engines; because the MetaBase Server controls communication to all MetaBase Engines, expanding the deduplication environment is a relatively simple process. This separation of the file meta data from the file system allows PureDisk to improve search- and maintenance-related activities on the underlying storage system, grow to hundreds of terabytes and billions of files, and retain a single logical instance of deduplicated data across the enterprise.

EARLY ADOPTERS

Early adopters of EMC Avamar and Symantec Veritas NetBackup PureDisk report minimal issues with installing backup software agents or server performance hits, but there are some specific circumstances that they monitor more carefully: the initial round of backups and the age of the server on which agents are deployed.

Jim Rose, manager of systems administration with the State of Indiana's Office of Technology, installed PureDisk at branch state offices as part of a mandate by Indiana Governor Mitch Daniels to centralize certain IT functions. At each of the 80 offices he manages, Rose installed a Microsoft Windows server with PureDisk software, as well as PureDisk agents on each of the servers targeted for backup. Rose found backup of the initial server took between 24 hours and 36 hours, while the second backup took about half that time; by the second or third day, backup windows across all of these servers were almost back to normal, he says.

"Symantec PureDisk backed up new servers without any discernible performance hit," says Rose. "[But] servers older than three years took longer to complete the initial scan."

Michael Fair, network administrator, information technology division at St. Peter's Health Care Services in Albany, NY, finds that the performance and management overhead associated with EMC's Avamar is almost nothing vs. what he encountered when backing up his servers with CA BrightStor and Symantec Backup Exec. "I eliminated domain controllers in eight sites and can now run backups during the day if the

need arises with no discernible impact to server applications," says Fair.

Introducing PureDisk allowed the State of Indiana's Rose to back up 300 servers across 80 sites in six hours, and he now has a demonstrable, working recovery plan for those sites. However, as the individual responsible for both remote offices and enterprise data centers, he recognizes the limitations of backup software deduplication. Taking 24 hours to 36 hours to complete an initial backup, coupled with high change rates on central databases, precludes Rose from deploying PureDisk in his core data center environment. For these more mission-critical servers, he looks to disk libraries to keep processing off the hosts.

INLINE DISK LIBRARIES

Disk libraries perform deduplication in two general ways: inline and post-processing. With inline processing, the disk library processes backup streams and deduplicates the data as it enters the disk library. Inline disk libraries use three general deduplication methods to minimize the performance impact: hash based, inline compare and grid architecture.

Integrating deduplication into the backup process is rapidly evolving from a nice-to-have capability to a must-have capability.

Data Domain's DDX disk library uses a hash-based technique. DDX takes an 8KB slice of the incoming backup data and computes a hash or fingerprint value. If the fingerprint value is unique, it deduplicates and stores the data. The main issues with this approach are the performance requirements to compute the hash and keeping the hash index in memory; as the hash index grows, it spills over from memory onto disk. To mitigate the performance overhead associated with retrieving the index from the disk, Data Domain developed a technique called stream-informed segment layout (SISL) that minimizes seeks to disk so the performance is CPU-centric; the faster the disk library CPU, the better the performance.

Diligent Technologies' inline ProtecTier Data Protection Platform attempts to avoid the performance penalty required by hash lookups by doing a computational compare. Using its proprietary HyperFactor technology, it avoids opening the backup data stream to examine content and instead scans and indexes the data stream, looking for data that's similar to data already stored.

When the ProtecTier Data Protection Platform finds data it considers similar to data already stored in its index, it does a byte-level compare

of the two sets of data; if it matches, it discards the match and references it. Diligent claims this compare-and-compute technique allows its ProtecTier Data Protection Platform to scale to manage hundreds of terabytes. However, this technique still requires some processing power on the part of the disk library to do the computational compare and to compress the data after it has been deduplicated.

NEC Corp. of America's Hydrastor also uses an inline approach, but it employs two different techniques to offset the performance overhead. In the first phase, Hydrastor deduplicates larger, variable-sized chunks of data to eliminate large pieces of redundant data. In the second phase, Hydrastor analyzes smaller, variable-sized chunks of data. In both cases, unique data is compressed.

To compensate for the performance overhead this multiphased approach creates, Hydrastor uses a grid architecture. This allows users to add additional nodes to the cluster at any time, which are designed to deliver additional performance or capacity. Unlike some other disk libraries, Hydrastor doesn't offer an option to present itself as a virtual tape library. Rather, it presents itself to hosts as a NAS filer using standard NFS and CIFS interfaces and creates one large storage pool on the back end. The Hydrastor architecture may present a problem for those enterprises that need to allocate and reserve certain amounts of storage for specific departments or business units.

POSTPROCESSING DISK LIBRARIES

With postprocessing, the disk library stores the data in its native format before deduplicating it, which allows the disk library to dedupe the data during nonpeak backup times. Vendors implement postprocessing in a variety of ways.

For example, Quantum's DXi-Series deduplicates data after it's stored, but initiates the deduplication process without waiting for the entire backup job to finish. By starting deduplication and then compressing the data while the backup is still running, it overcomes one of the principal downsides of postprocessing—the requirement for sufficient capacity to house the native backups. However, deduplication requires use of the DXi-Series' cache and processor, which can potentially slow the backup process because the backup job may need to write the data directly to slower responding disk instead of storing it in the DXi-Series' cache.

To avoid that scenario, ExaGrid Systems Inc.'s ExaGrid and Sepaton's S2100-ES2 execute only on backup sets that have completed, so dedu-

plication doesn't impact backup and restore performance. On the first analysis of backed up data, ExaGrid and S2100-ES2 only compress the data and don't deduplicate. When a second backup completes, ExaGrid does byte-level delta differencing while Sepaton uses its ContentAware software to compare objects in the first backup at byte level against similar objects in the second one. Like objects in the first backup are then deleted and replaced with pointers to objects in the second backup, with objects in the second backup then compressed but not deduplicated. This deduplication and compression process repeats with subsequent backups.

The difference between the two determines what size environments they best fit. You can't add more controllers to ExaGrid to allow it to deduplicate the large amounts of data that enterprise backups generate. Sepaton uses a grid architecture in S2100-ES2 so additional controllers for more processing and capacity can be added as deduplication requirements grow.

HIDDEN ISSUES

Regardless of the deduplication approach, there are some hidden issues. For postprocessing disk libraries, as the amount of data increases, it may take much longer to deduplicate the data once the backups are complete. If the deduplication takes longer than the time between the end of one backup window and the start of the next, all of the data from the first backup won't be deduplicated. Users will need to ensure they can add more processing power to handle this load.

Another potential problem may arise with inline or postprocessing disk libraries that aren't replicating the data to a remote disk library: the need to create tapes. The disk library needs sufficient time to first deduplicate the data and then undeduplicate a copy of the data to be spun off to tape. Both ExaGrid Systems' ExaGrid and Sepaton's S2100-ES2 avoid this undeduplication overhead because the last backup is only compressed, not deduplicated, so users can copy the job directly to tape.

Other postprocessing disk libraries like Spectra Logic Corp.'s nTier appliance allow users to run a local master or media server within their nTier appliance, which alleviates some of the pain of this process. The nTier appliance eliminates the need to move data from host to media server to deduplication box to media server to tape, and allows the data to move from host to nTier appliance to tape. This design also eliminates the need to undeduplicate the data before storing it to tape.

Deduplicating backup software products that must operate in conjunction with enterprise backup products like Symantec Veritas NetBackup or EMC NetWorker face a different problem—allowing the enterprise backup software product to recognize and catalog the data it has backed up. While neither Asigra Televaulting nor EMC Avamar have any formal integration in place with any enterprise backup software product yet, Symantec Veritas NetBackup PureDisk includes a NetBackup export engine that allows an admin to copy a backed-up data selection from a PureDisk content router to NetBackup. NetBackup then catalogs the data and copies it to tape or disk and, from the NetBackup administration console, the storage admin can treat those files as if they were native NetBackup files. Both EMC and Symantec anticipate tighter integration between their enterprise and deduplicating backup software products in the near future.

A final area of concern is the ability to selectively turn off deduplication for specific files or servers. This is important when compliance is an issue because the authenticity of data may come into question if it's deduplicated in any way. Also, if data is encrypted before the disk library receives it, deduplication provides no additional space-saving benefits; users should identify ahead of time what data is encrypted before it's stored or sent to a disk library.

As data stores continue to soar, integrating deduplication into the backup process is rapidly evolving from a nice-to-have capability to a must-have capability for most corporate environments. The good news is that for small- to medium-sized businesses managing 10TB of data or less, using either type of deduplication product—backup software or disk library—will significantly shorten backup windows. The decision then becomes what product best fits your environment.

At the enterprise level this isn't the case yet. Though promising work is occurring with inline approaches such as Diligent Technologies' ProtecTier Data Protection Platform and NEC's Hydrastor, most enterprises will find a postprocessing disk library such as Sepaton's S2100-ES2 a safer choice for now until all of the costs, risks and processing overhead associated with inline deduplication are better understood and documented. ⊙

Jerome M. Wendt (Jerome.wendt@att.net) is lead analyst and president of DCIG Inc.

Get "Friendly Advice" on your most pressing data concerns...

from a not so friendly chap.

Are you looking to gain control of your company's digital information?

Or maybe your sanity?

Well, you're in luck because now you can get "Friendly" advice from a professional who's truly passionate about helping you manage your data — Dr. Harold Twain Weck.

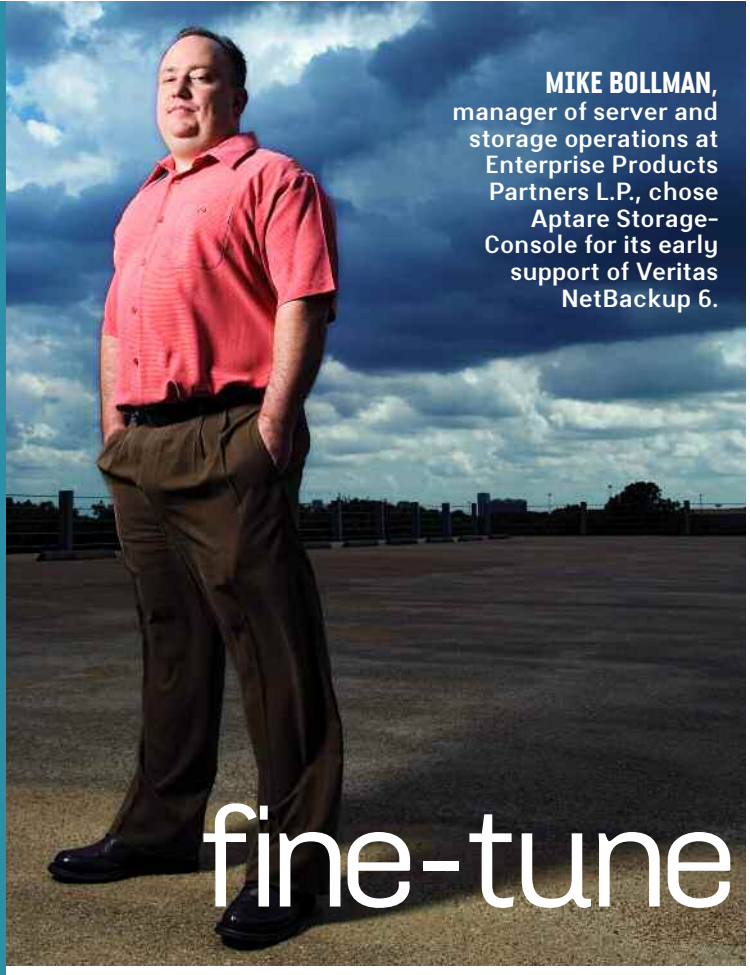
That's right, John Cleese is at it again as Dr. Twain Weck to give you friendly advice on your most critical digital information protection and storage challenges.

Just what do we mean by "Friendly?" Well, you'll just have to see for yourself by visiting John Cleese's Friendly Advice Machine. It's the new Web destination for IT people, legal professionals and business executives seeking guidance on today's most pressing and complex information protection and storage concerns.

Visit the Web site. And let Dr. Twain Weck set you on a path that leads you away from digital data madness, and on a path to just plain madness, but in a friendly sort of way.

Visit John Cleese's Friendly Advice Machine today!
www.friendlyadvicemachine.com

DIGITAL™ IRON MOUNTAIN



MIKE BOLLMAN, manager of server and storage operations at Enterprise Products Partners L.P., chose Aptare StorageConsole for its early support of Veritas NetBackup 6.

fine-tune backups

Reports from specialized tools that work alongside major backup apps help you predict usage patterns and troubleshoot issues.

By Jacob Gsoedl

BACKUP AND RECOVERY (B/R) applications do a good job of managing tasks such as job scheduling, tape management, library support, tracking backup data in catalogs and supporting a variety of backup media, including disks. But B/R applications fall short when it comes to advanced capacity reporting, predicting usage patterns, performance tuning, troubleshooting and cost management.

Because backup applications deliver some of these advanced features, especially reporting, storage managers often find it difficult to justify bringing in another product to overcome the limitations of their backup apps. Product choice is highly subjective and dependent on the storage and backup environment in place. Storage managers deploy data protection and recovery management (DPRM) products that best address the needs of their unique requirements and environments, such as increased visibility about backups, compliance requirements, reporting consolidation and a need to cover the entire backup cycle, including snapshots and replicas.

Photograph by Matthew Mahon

Douglas Bovie is head of infrastructure management at Orange Business Services, a global provider of business and communications services in Eschborn, Germany. Bovie chose WysDM for Backups from WysDM Software Inc. for its cross-domain reporting and predictive analysis capabilities. It gave him reporting data for all elements involved in the backup process, including EMC Corp.'s NetWorker, his ADIC libraries, filers from Network Appliance (NetApp) Inc. and switch ports from Brocade Communications Systems Inc.

A lack of visibility into backups for information technology and business process owners caused Mike Bollman, manager of server and storage operations at Enterprise Products Partners L.P., a Houston-based energy services provider, to use Aptare Inc.'s Aptare StorageConsole.

"We selected Aptare StorageConsole for its early support of [Symantec Corp.'s Veritas] NetBackup 6 and Aptare's apparent close relationship with Symantec/Veritas," says Bollman.

KEY FEATURES OF DPRM TOOLS

When considering a DPRM tool for your environment, keep in mind the following product capabilities:

Support for heterogeneous products. DPRM tools vary in the number of backup apps and disk arrays they support. Bocada Inc., Tek-Tools Inc. and WysDM have the largest backup application support.

Cross-domain reporting. Being able to report beyond the backup app and include all of the elements the B/R process depends on, including servers, network, switches and libraries, is crucial for performance analysis and troubleshooting. "When backups run slow or fail, it's essential to be able to look beyond the tape drive and backup software," says Greg Schulz, founder and senior analyst at StorageIO Group, a technology analyst and consulting firm in Stillwater, MN.

Vendors have taken two approaches to reach beyond the backup app. Products from Aptare, Bocada, Servergraph Inc. and WysDM report on the full data path within a single application. Hewlett-Packard (HP) Co. (with its HP Storage Essentials Backup Manager) and Tek-Tools (BackupProfiler) defer to other modules or products to get to related information. As a component of a larger storage resource management (SRM) application, HP Storage Essentials Backup Manager may require the licensing of additional SRM modules to get all of the relevant information. Tek-Tools defers to other tools within the Tek-Tools Profiler family, such as AppProfiler, ServerProfiler and StorageProfiler, to report on infrastructure components not directly related to the backup application.

DPRM vs. SRM

Storage resource management (SRM) applications haven't lived up to their promise from the late 1990s, namely to be a single tool that manages all storage and backup assets. There are several reasons for this: SRM applications are expensive and complex, and a lack of standards limits their capabilities. Element managers and specialized programs, such as data protection and recovery management (DPRM) tools that are good in one area, have done well because they're simpler, less expensive and typically outperform comparable features of SRM applications.

SRM apps are monoliths, with a single vendor implementing all features. But monolithic app designs are no longer state of the art and are challenged by service-oriented architectures (SOAs) where loosely coupled apps form a larger app framework. "The new trend is to have best-of-breed tools that work together based on an SOA-type architecture," says Nancy Hurley, VP of marketing and business development at Bocada Inc.

Predictive analysis. DPRM tools are changing from backward-looking instruments that help users understand past backups into forward-looking tools that foresee backup needs and problems, such as storage consumption, media storage availability and anomaly detection. All of the DPRM tools we looked at have some predictive analysis capabilities in the form of trending reports. “StorageConsole uses its predictive analysis engine to determine the ‘backup heartbeat’ for any given environment, and reports and alerts on abnormalities or early hot spots,” says Rick Clark, Aptare’s president and CEO. Similarly, WysDM for Backups’ Predictive Analysis Engine runs 24/7 to check the backup environment and identify anomalies.

Policy enforcement and real-time alerting. Similar to predictive analysis, alerts are triggered if a policy violation is detected. For instance, you may have a policy to be notified if a server doesn’t back up within its backup window for two nights in a row. WysDM leverages its Predictive Analysis Engine to define and enforce policies. “You can configure rules to enforce policies required for compliance,” says Alan Atkinson, WysDM’s CEO. In a similar manner, Servergraph Data Protection Expert provides a rules engine through its prediction module, while Aptare StorageConsole supports policy-based, real-time monitoring based on exception thresholds across backup domains.

Which backup reporting tool is the best fit?

The ability to verify the success of backup and recovery (B/R) applications is the top priority of any IT organization. Backup reporting comes in three flavors:

- Reports integrated in B/R applications
- Data protection and recovery management (DPRM) tools with multivendor B/R application support
- Storage resource management (SRM) applications with integrated backup reporting

If your company has standardized on a B/R product, the integrated reporting option may suffice or you may opt for an advanced reporting add-on from your B/R vendor. If you have multiple B/R products deployed, a DPRM tool with multivendor B/R support may be your best best. For a complex storage and backup environment, an SRM product might be best. Finally, if your backup environment has grown organically over time, B/R consolidation should be part of your strategy.

"Over time, we will see more focus on real-time behavior and event correlation," predicts StorageIO Group's Schulz.

Support for replicas, snapshots and clones. Triggered by snapshot and replication features in storage arrays, disk-based data protection adoption has proliferated and DPRM tools are playing catch up. In many environments, snapshots and replicas are eventually backed up to tape, but what if the replica itself has problems?

Exposure analysis and recoverability. Proving recoverability and having provisions in place to detect unprotected data assets has been the focus of compliance audits, especially Sarbanes-Oxley. Backups need to be complete and consistent to be recoverable; a simple "job succeeded" doesn't suffice unless it implies completeness and consistency. To take this a step further, estimating recovery time helps storage managers live up to their advertised recovery time objectives (RTOs).

Service-level compliance and risk and gap analysis are the focus

of Symantec's Veritas Backup Reporter. "The backup of an app like SAP may consist of different backup policies for the application, database and files; for the backup to be successful, all related jobs must succeed to claim success," says Erica Antony, product manager at Symantec. "Veritas Backup Reporter correlates related backup policies and takes a holistic approach when assessing the success of backups."

Aptare and Tek-Tools also concentrate on exposure analysis and recoverability. "We're comparing hosts in the backup policies with the actual backup and discovering clients that aren't backed up, as well as unprotected volumes and mount points of partially protected clients," says Aptare's Clark.

Business-centric reporting. Aligning storage with business objectives is part of all DPRM tools. Topping the list are compliance reports, triggered mainly by audit requirements of public companies. "Data protection in audits is all about proving that you follow defined policies and that you take [the] best effort to be recoverable," says Nancy Hurley, Bocada's VP of marketing and business development. "Bocada Enterprise maintains all relevant information and delivers the evidence required during audits," she says. Besides compliance, reports on service-level management, cost management, chargeback and billing are finding their way into DPRM tools.

PRODUCT ASSESSMENT

DPRM tools are constantly evolving, with some products overlapping others in features and functionality. However, each product listed below has carved out its unique niche.

Aptare StorageConsole has a compelling list of next-generation DPRM features beyond backup reporting. Functions include a sophisticated predictive analysis engine, policy enforcement, cross-domain reporting, file-level reporting, and a three-tier architecture tailored for distributed and hosted environments. Aptare is working on increasing the backup apps it supports and improving its lack of disk-based data protection reporting.

Bocada Enterprise 5 has strong backup app support, a highly scalable architecture, an extensive list of historical and trending reports, and an impressive list of Fortune 500 accounts.

HP Storage Essentials Backup Manager is one of the many modules in the HP Storage Essentials SRM suite. Although it can be licensed separately for backup reporting, it depends on other Storage Essentials modules for cross-domain reporting. A limited number of backup prod-

ucts are supported and there's a lack of next-generation DPRM features. "Typically, you look at Storage Essentials Backup Manager if you're looking for an SRM tool that also does backup reporting," says StorageIO Group's Schulz. "But if you're only looking for a DPRM tool, you probably would look somewhere else."

Servergraph Data Protection Expert is a comprehensive tool focusing on real-time reporting and flexible report presentation through easily configurable reporting dashboards and custom message boards. Servergraph, a division of Rocket Software, is able to leverage other Rocket Software products, making it a viable DPRM player with a product that has evolved beyond historical reporting.

Symantec Veritas Backup Reporter focuses on trending and analyzing historical backup data to ensure backups occur in line with defined policies. For real-time activities like monitoring the backup environment, Symantec defers to Operations Manager in Veritas NetBackup 6 or equivalent modules in other supported backup apps. Although Veritas Backup Reporter was designed for NetBackup, it supports other backup applications.

Tek-Tools BackupProfiler, which is part of the Profiler family, addresses storage reporting and management needs beyond backup. Extensive backup app support, the ability to report on NetApp filer snapshots, and a comprehensive list of historical and real-time reports make BackupProfiler a competitive reporting offering.

Disk-based data protection adoption has proliferated and DPRM tools are playing catch up.

"Tek-Tools is very broad when it comes to backup and infrastructure reporting and management, but the tools don't go as deep on data protection reporting as some of the other tools," says Schulz.

WysDM for Backups has cross-domain reporting capabilities, and a predictive analysis engine capable of carrying out complex logical and mathematical operations on data to spot trends and proactively alert on future error conditions. "WysDM goes very deep when it comes to backups, but they need to expand to replication," says Schulz. WysDM for Backups and Aptare StorageConsole are the two products in this roundup that have expanded the furthest beyond backup reporting.

BACKUP APPLICATION REPORTING FEATURES

DPRM tools supply some of the reports and features that major backup vendors can add fairly easily to their apps. For example:

BakBone Software Inc. offers the NetVault:Report Manager family of reporting tools, but they report only on products within the BakBone NetVault family.

EMC opted against developing its own DPRM tool and instead licenses and resells WysDM for Backups as EMC Backup Advisor (EBA), a heterogeneous, highly customizable enterprise reporting tool. With EMC NetWorker Dashboard, EMC offers an optional scaled-down version of EBA for EMC NetWorker customers. Unlike EBA, EMC NetWorker Dashboard isn't customizable, comes with approximately 150 reports and works only with EMC NetWorker.

IBM Corp. depends on the reporting capabilities in Tivoli Storage Manager (TSM), which is limited to mostly operational reporting. For more advanced reporting, IBM defers to third-party tools like Bocada and WysDM. "We're actively looking at features of other DPRM tools to get more trending reports," says Tricia Jiang, technical evangelist for TSM products at IBM.

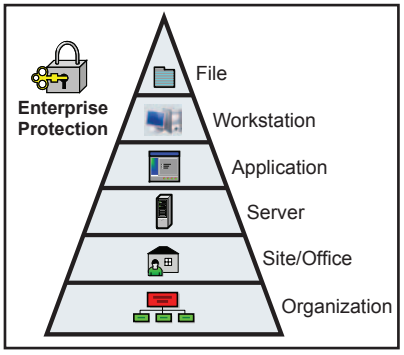
Symantec has developed its own DPRM tool with Veritas Backup Reporter, which has its roots in the Veritas CommandCentral Service.

DPRM is evolving from passive reporting tools to proactive instruments that help storage managers better manage their backup environments with features like predictive analysis, cross-domain reporting, real-time monitoring and alerting, and reporting on the complete backup cycle. Although all of the tools are works in process to some degree, Aptare StorageConsole and WysDM for Backups appear to have the edge when it comes to next-generation DPRM features beyond reporting. ⊙

Jacob Gsoedl is a freelance writer.

How to Instantly Recover Applications, Data, Servers, and Sites

The risks and pitfalls facing today's information management landscape demand more than a scatter-shot approach to data protection and business continuity. A fully integrated paradigm is needed to unify all the aspects of enterprise protection. By deploying a scalable enterprise-wide approach to disaster recovery, you ensure that your organization will survive any disastrous loss — of data, applications, servers or an entire site — without impacting operations.



Tiered Risks, Single Enterprise Solution

The unique new XRServer™ technology available in the latest version of Syncsort Backup Express offers the following benefits in a single, easily managed package:

Business continuity. Instant Availability of critical data and applications — including Oracle, Exchange and SQL Server — occurs without a time-consuming data transfer because your backed up snapshot immediately becomes a live application after a failure.

Meet RPOs and RTOs. Extremely fast recovery to any past point in time provides for little or no data loss, and minimal downtime.

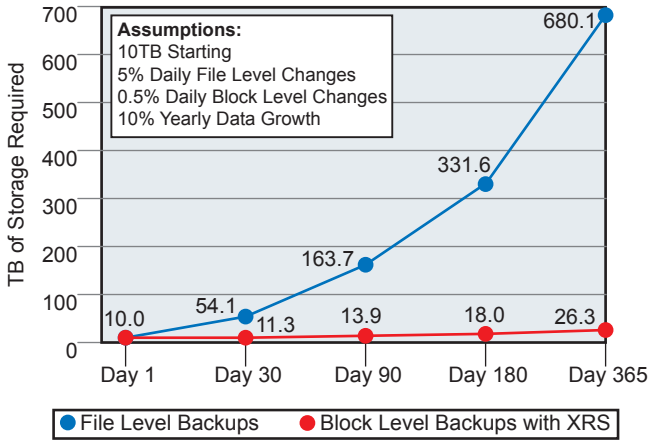
Fast backup. Non-intrusive, block-level incremental backups are extremely fast, require no backup window, and use minimal bandwidth and CPU. Fast backups have extremely high success rates and can be scheduled to run very frequently — yielding nearly continuous protection.

Fast and simple disaster recovery. After a server failure, disk-to-disk full server recovery, using the regular backups, enables entire servers to be fully restored in minutes, instead of in hours.

Enterprise-wide coverage. The XRServer™ solution supports not only critical applications, but all levels: file, server, and site. XRS file transfers serve as routine file backups, disaster recovery backups, and everything in between.

Reduced complexity. One catalog and management interface control all levels of backup and recovery — saving time and making life easier.

Data deduplication saves storage. Block-level technology is not only fast, but uses much less space and CPU than standard backup techniques. Over time, the storage savings through source-side deduplication is enormous. Using one product for all protection needs also reduces the costs associated with multiple products.



Storage Savings add up with XRS — a 96% reduction in first year!

Backup Express XRS — Express Recovery Server — is a single solution to multiple challenges. Disk-to-disk, block level incremental technology provides fast backup, equally fast restore and Instant Availability of applications.

You need to find out more:
www.syncsort.com/tge38ab

Syncsort Incorporated
50 Tice Boulevard • Woodcliff Lake, NJ 07677 • 201.930.8200

syncsort Backup Express

demystifying dedupe

Data deduplication offers an attractive solution to the problem of data growth. But will deduplication work in your environment?

By Stephen J. Bigelow

DATA DEDUPLICATION can be implemented using software or a dedicated appliance, and each approach has a number of benefits and drawbacks. This article evaluates each approach and outlines other important points to consider before deploying deduplication.

EVALUATING SOFTWARE-BASED DEDUPE PRODUCTS

Data deduplication technology can be implemented with software products installed on a dedicated server, or integrated into backup or archiving software. Software-based data deduplication is typically less expensive to deploy than a dedicated hardware appliance, and it should require no significant changes to the physical network. However, software-based deduplication can be more disruptive to install and harder to maintain. The disruption can be even greater if you're replacing the backup engine with an entirely new product because the backup administrator must recreate backup job configurations, schedules and alerts from scratch. Deduplication at the host backup server is also processing intensive, so the server must be configured for the task.

EMC Corp.'s Avamar software product performs in-band deduplication at the host server (the source) using the SHA-1 algorithm. Avamar employs a central management scheme to inspect data in the entire environment, but the actual deduplication is performed at each server before being sent to the backup storage platform. This saves storage space at the backup target and reduces network congestion. EMC reports plans to incorporate Avamar technology into its own backup software and virtual tape library (VTL) system in the near future.

Symantec Corp. provides software-based deduplication in its Veritas NetBackup product through a feature called PureDisk, which uses a proprietary hash algorithm to perform deduplication inline at each host server. NetBackup PureDisk 6.2 supports tape targets and the Backup Reporter monitoring tool. NetBackup 6.5 offers even better integration and support for deduplication, VTL and third-party appliances.

Sepaton Inc. implements deduplication using DeltaStor software, an option on its S2100-ES2 VTL hardware product. Like PureDisk, DeltaStor uses a proprietary hash algorithm, but the S2100 deduplicates data at the VTL (the storage target), so there's no decrease to network traffic. Sepaton also works differently than other deduplication schemes. Typically, the first iteration of data is written and later iterations receive pointers; DeltaStor writes the latest version and replaces the previous iterations with a pointer.

EVALUATING HARDWARE-BASED DEDUPE PRODUCTS

Hardware-based data deduplication can be performed in-band or out-of-band. In-band deduplication reduces data while it's being written to storage. This approach can be efficient because it's performed only once, though the additional processing power needed to handle the process may actually extend the backup window.

Hardware-based data deduplication can be performed in-band or out-of-band.

Out-of-band deduplication is performed after data has been stored. This approach doesn't affect the backup window and alleviates concerns about CPU processing creating a bottleneck between the backup server and the storage. However, out-of-band deduplication uses slightly more disk space during the data deduplication process. Out-of-band deduplication may also take longer than the actual backup window. Disk contention is another problem, reducing disk performance as users attempt to access storage during the deduplication process.

Hardware-based deduplication doesn't offer the bandwidth savings it might receive by deduplicating at the source, but compression levels are often better and hardware-based data deduplication products require less maintenance. Hardware-based data deduplication appliances are noted for their high performance, scalability and relatively nondisruptive deployment. Backup software will normally see dedicat-

COMPRESSION, ENCRYPTION AND DATA DEDUPLICATION

One of the stickiest issues with data deduplication is the relationship between compression, encryption and deduplication. Traditional compression works by eliminating redundancy in files. Deduplication can eliminate redundant files, blocks or bits, and encryption turns that data into a data stream that's random by its nature. So if you encrypt data first, it may be impossible to compress or deduplicate it. Ideally, data should be compressed and deduplicated first, and then encrypted as needed. This isn't difficult when compression and deduplication are performed at the host server using backup software, and the resulting data stream is encrypted on the way to the backup target using a dedicated appliance or at the tape library or tape drive. However, this may present difficulties when deduplicating at the target storage system. For example, if the backup data is encrypted by an inline appliance and then sent to a deduplication-capable storage system like the Sepaton S2100, it may be impossible to further compress or deduplicate the encrypted data.

ed appliances as a generic “disk system” and remain totally unaware of the deduplication processes taking place under the covers.

Hardware-based deduplication may also be incorporated into other storage (target) platforms. For example, data deduplication is often a feature of VTL systems. VTLs speed backup tasks by utilizing disk rather than tape for storage, and adding deduplication allows the VTL to maximize disk usage. In many cases, VTL deduplication is implemented as an out-of-band process. This is an advantage because all of the VTL’s contents can be deduplicated to achieve very good compression ratios. The downside is that data deduplication isn’t immediate. However, some VTLs do incorporate the processing power to deduplicate backup data in-band as data is received from the backup server.

Data Domain Inc. touts one of the most diverse product lines intended for VTL and NAS systems. Appliances range from the branch-office DD410 to the enterprise-class DDX series. All deduplication is performed in-band using the SHA-1 algorithm along with a second proprietary algorithm to prevent hash collisions. The index itself is maintained on nonvolatile RAM within the appliance. Data Domain appliances are relatively slow, offering a throughput of only 110MB/sec, but the company claims it’s working to improve those data rates through clustering.

The enterprise-class ProtecTier VTL from Diligent Technologies Corp. also performs in-band deduplication using a single proprietary algorithm. The index is then stored on a Fibre Channel disk that can potentially improve indexing performance. The results are shown in Diligent's performance numbers that achieve up to 400MB/sec. Similarly, the DXi3500, DXi5500 and DXi7500 appliances from Quantum Corp. perform in-band indexing and data deduplication using a patented algorithm that has also been added to Quantum's StorNext file system. By comparison, the Single Instance Repository (SIR) on FalconStor Software Corp.'s VTL uses out-of-band indexing with SHA-1 and MD5 algorithms.

For backup appliances, ExaGrid Systems Inc. includes an out-of-band deduplication feature with its NAS backup appliance. ExaGrid works with bytes rather than bits, so the indexing is simpler, leading to faster search performance. ExaGrid also examines the common data patterns in backup software products, aiding search and indexing performance. The Hydrastor grid backup appliance from NEC Corp. of America uses a proprietary process to deduplicate data at the subfile level. NEC claims a 75% reduction in storage utilization without impacting storage performance.

When implementing a data deduplication system, it's important to consider scalability.

Network Appliance (NetApp) Inc. performs block-level data deduplication in its NearStore R200 and FAS storage systems. Deduplication is based on NetApp's Advanced Single Instance Storage (ASIS) feature that uses 16-bit checksums already stored with each data block to look for redundancy candidates. Those blocks are then compared at the bit level, and identical blocks are discarded. NetApp's storage systems will deduplicate primary storage.

ENSURE SCALING AND RELIABILITY IN DEDUPE SYSTEMS

When implementing a data deduplication system, it's important to consider scalability. Performance should remain acceptable as storage capacity and deduplication granularity increase. Data deduplication should also be unaffected by data loss due to errors in the hashing algorithm.

When the system processes new data elements, their resulting hash numbers are compared against the hash numbers already in the hash index. If a new data element produces a hash number identical to an

entry in the index, the new data is considered a duplicate, and it's not saved to disk—only a small reference "stub" that relates back to the identical data has been stored. If the new hash number isn't in the index, the data element is considered new and stored to disk normally.

A data element can produce an identical hash result even though the data isn't completely identical to the saved version. Such a false positive, also called a hash collision, can lead to data loss. There are two ways to mitigate false positives. The data deduplication vendor may opt to use more than one hashing algorithm on each data element, or a single hashing algorithm may be used if you perform a bit-level comparison of data elements that register as identical.

The problem with both approaches is that they require more processing power from the host system, reducing index performance and slowing the deduplication process. As the deduplication process becomes more granular and examines smaller chunks of data, the index becomes much larger, and the probability of collisions increases and can exacerbate any performance hit.

Another issue is the relationship between deduplication and more traditional compression and encryption in a company's storage infrastructure. Ordinary compression removes redundancy from files, and encryption "scrambles" data so that it's completely random and unreadable. Both compression and encryption play an important role in data storage, but eliminating redundancy in the data can impair the deduplication process. If encryption or traditional compression are required along with deduplication, the indexing and deduplication should be performed first. ⊙

Stephen J. Bigelow is the features writer for SearchStorage.com.

Deduplicate.

Minimize your data to maximize your storage.

Enterprise deduplication means more than saving disk space. It means keeping petabytes of data at your fingertips for instant recovery. It means meeting SLAs and regulatory requirements – all while cutting space, energy and disk costs. It means regaining control of your growing data infrastructure.

Learn how DeltaStor™ from SEPATON can meet your enterprise deduplication needs. Visit www.SEPATON.com/deduplicate.
Or call **866.SEPATON**.

SEPATON®

VM backup tips

Virtual server backups can be accomplished using conventional backup software. However, this approach has some drawbacks. Learn your options for backing up virtual servers.

By Stephen J. Bigelow

AS SERVER VIRTUALIZATION assumes a greater role in the enterprise, administrators face a proliferation of virtual machines residing on the same physical server. Each virtual machine uses a portion of the physical machine's processing, memory and I/O resources. Ideally, server virtualization provides a means of increasing hardware utilization.

But as more "logical" servers are consolidated into fewer "physical" computer systems, it's important to protect each virtual machine's data against failure or loss. Virtual server backups are the key to providing this protection. This article examines how virtual server backup can be achieved using a mix of traditional backup techniques and specialized virtualization tools. It also highlights important deployment issues.

A virtual machine is a complete logical environment that exists as a separate entity on a physical server. Each virtual machine is treated and perceived as if it's physical. A user can't tell the difference between a real machine and virtual one. A data center may host thousands of virtual machines running on only a fraction of that much hardware, and this presents a serious problem for storage or backup administrators. Data loss on a virtual server can be just as catastrophic as data loss on a physical server, so every virtual server must be backed up as part of a company's backup regimen.

Virtual server backups can be accomplished using a traditional approach with conventional backup software. The backup software is installed and configured on each virtual machine, and backups run

RESTORING FILES
If you have an agent-based backup system, you can restore a single file back into the virtual machine (VM) just as in the past. But if you're doing system-level backups where you're backing up the entire VM image, you have to restore the entire VM image. Once you recover it, you have to mount it somewhere in a virtualized environment, typically not the production environment. You can then recover files from that point and transfer them back to where they need to be.

normally to any conventional backup target, including tape drives, virtual tape libraries (VTLs) or disk storage.

"That's probably the most popular way that people do it today because it's familiar," says Lauren Whitehouse, an analyst at Enterprise Strategy Group (ESG), Milford, MA. "It ensures a consistent backup, will give you the granular recovery that you're looking for and it's application specific."

However, applying traditional backup tactics to virtual server backups does have its drawbacks. The most significant problem is resource contention. Backups demand significant processing power, and the added resources needed to execute a backup may compromise the performance of that virtual machine and all virtual machines running on the system. "Don't go for 100% utilization," says Greg Schulz, founder and senior analyst at StorageIO Group, Stillwater, MN. Leave some server resources unused to accommodate backup tasks and stagger backup processes so that only one virtual machine is being backed up on any physical system at one time.

BACKUP PROCESS MORE COSTLY IN VIRTUALIZED ENVIRONMENTS

Installing backup software on every virtual machine can make your backup process far more costly. Also, traditional backups will copy programs and application data but don't necessarily capture the entire virtual machine state. This may be fine if your only goal is to preserve an application, such as a database, but a failed virtual machine may need to be recreated and reconfigured from scratch before the backup can be restored.

Virtualization-specific tools, such as VMware Consolidated Backup (VCB) or Microsoft's Virtual Machine Manager (VMM), interface directly with their respective virtualization platforms—VMware's Virtual Machine

Disk Format (VMDK) or Microsoft's Virtual Hard Drive (VHD). Virtual server backup tools like VCB or VMM can capture the entire virtual machine state quickly, and the virtual machine typically doesn't need to be acquiesced or taken offline. Not only does this allow for fast, complete system restorations, but complete snapshots can be uploaded to new virtual machines, allowing system administrators to "clone" virtual servers on demand.

The downside to virtual server files is a potential loss in granularity. With traditional backups, it's easy to restore a single application or data file. When there's one single VMDK or VHD file, you typically have to restore the entire snapshot to recover, even if only one file is lost or corrupted. "Some snapshot vendors have figured out how to take that image-level backup and break it down into the granular single files that people need to recover," says ESG's Whitehouse. "Not everyone has done that, though."

A virtual machine is a complete logical environment that exists as a separate entity on a physical server.

IMPLEMENTING VIRTUAL SERVER BACKUP

Storage space poses a particular challenge for virtual machine files. The virtual snapshot is always seen as a new file, so it's backed up in its entirety, regardless of how much data has changed since the last snapshot. Snapshots will continue to use the full backup window and consume the same amount of disk/tape space. Data deduplication, also called single-instance storage, can reduce these storage demands. Deduplicating at the storage system doesn't shrink the backup window because data must still be transferred across the network prior to deduplication. Experts suggest deduplicating through an appliance or at the source to save backup media while minimizing the backup window.

Virtual server backups have no specific affinity for backup targets. Traditional backups can go to tape, VTL or other disk systems as they do now, though most performance-minded users will back up to some form of disk storage first and then offload the backup to tape later. VCB or VMM backups are almost universally sent to disk, then later

Problems don't have to be **Disasters**

Introducing the Operations IT
Advanced Recovery & Continuity Appliance (ARCA)

Starting at
$299/month

Keeping Your Business ON

The ARCA provides the speed and reliability of enterprise-class business continuity and recovery solutions…without the enterprise-class price.

With RTO and RPO of five minutes, the ARCA transforms the way your IT department can address anything from a server crash to a lost email.

And with the support of Operations IT regional partners and our global NOC, we're there from restart to finish.

Learn more at www.OperationsIT.com

- Quick Standby Server
- Bare Metal Restore
- File & Database Recovery
- Mailbox & Message Restore

OperationsIT
Powering Continuity

www.OperationsIT.com | info@OperationsIT.com

DEDUPE AND VMs
When backing up virtual machines (VMs), you have to save multiple copies of files, whether it's an OS file, a patch or an application. So, the benefit of data deduplication can be enormous. However, experts warn that this can put excessive workload on the VM, so a standalone deduplication appliance is recommended.

replicated to offsite disk storage or sent to tape. Backup media is then retained or stored in exactly the same way as conventional backups. However, retention periods should be evaluated carefully; it may not be necessary to save every snapshot for a prolonged period. Consult your local retention experts or legal counsel for their recommendations.

Virtual server backups should also be verified and tested periodically to ensure that the required suite of data has been captured adequately, but this typically involves restoring the backup to another virtual server and verifying normal operation. For some shops that perform frequent restorations, the "testing" process is ongoing; backups are tested each time a file or application needs to be restored. Other virtualized shops have auxiliary machines available for testing purposes that allow administrators to periodically test backups without taking the original production machines offline.

COMPANIES PERFORMING VIRTUAL SERVER BACKUPS

For Young America Corp., the customer fulfillment business generates a great deal of customer data. Close to 20TB of production data and another 10TB of development and test data is spread across several EMC Corp. platforms running under VMware Inc.'s Infrastructure 3 virtualization software. Virtualization has proven its benefit to the organization. "The No. 1 [benefit] is efficient use of resources," says Dan Thompson, network engineer at Young America. "Secondary reasons include ease of backups and disaster recovery."

Thompson backs up virtual machines using VCB operated in concert with EMC's Legato backup software. Virtual server backups are performed nightly along with the entire backup process, and are also performed on demand. The entire backup process takes about six to seven hours each night; but with approximately 160 servers to contend with (half of them virtual servers), it's difficult to say exactly how long a single virtual machine backup takes.

In addition to protecting existing virtual servers, Thompson uses virtual snapshots to clone new servers. "You can use VCB to actually save

a copy of a virtual machine 'hot,' then you can restore it to another virtual machine and bring it up as a clone of the first one," he says.

An EMC Clariion Disk Library (CDL) provides virtual tape support. "The backup application backs up to that and also to actual [IBM] tape, so we go to both," says Thompson at Young America, noting that the current LTO-3 tape drives will soon be upgraded to LTO-4. Although Thompson has never needed to restore a virtual machine failure, the restoration process has been thoroughly proven and is tested monthly or even more frequently.

Thompson notes that virtualization has proven reliable since the resolution of some early difficulties. "We had virtual machines lock up when VCB is executed that we attributed to outdated VMware drivers and tools," he says. "With that updated, those virtual machines haven't had a problem since." This underscores the importance of software maintenance and version control in the virtual environment.

Storage volumes will continue to grow, which will inevitably lead to a demand for more network storage for virtual machine backups.

Next to efficiency, flexibility in integrating infrastructures is probably the most important benefit gained from server virtualization. For Kroll Factual Data, the flexibility afforded by Microsoft Virtual Server 2005 R2 proved critical when integrating data centers. "We were moving an acquired company and their technology infrastructure into our data center, and the virtual environment was really the only way that we could be flexible enough to tackle the integration in a timely manner," says Christopher M. Steffen, principal technical architect at the information services business.

Once the benefits of storage virtualization became clear, the entire infrastructure was migrated to a virtual server environment, supporting more than 600 virtual machines in production (80% to 85% of the production environment). In addition, there are approximately 400 virtual machines in disaster recovery and another 400 virtual machines in development. "It's a hardware-agnostic point of view," says Steffen. "Any platform that runs a Windows server can support full virtualization and really utilize your hardware to its fullest potential." Today, Kroll Factual Data operates about 60TB of

storage on an IBM FAStT storage server.

Steffen uses the VMM utility to manage and back up Microsoft virtual machines. Not only does VMM help to configure and optimize the virtual environment, it also creates backup snapshots of the VHD file. Steffen also uses VMM to create standard server "images" that speed the deployment of new virtual servers, while helping to prove the compliance of software/driver versions across the environment. "Instead of configuring a new server from scratch, which can take two to four hours, just take and copy the hardened image that you've already created and patched correctly up to the host machine—that takes 10 to 15 minutes," he says.

Almost all virtual machine backups are performed through VMM, though there are still some manual backup processes to accommodate mission-critical processes that haven't been virtualized yet. The actual time needed to back up a virtual server depends on the size of the VHD file and the bandwidth available to pass the backup data to the target. Backups are always sent to disk first, then offloaded to tape as a separate process.

The ability to configure disaster recovery sites virtually anywhere, where power and Internet access are available, was an important benefit, according to Steffen. "Virtualization makes the whole disaster recovery 'mess' actually something that is manageable," he says. "And VMM helps with configuration management, update migration and so on." VMM provides load-balancing recommendations that can optimize the number of virtual machines on each particular server.

THE FUTURE OF VIRTUAL SERVER BACKUPS

Storage volumes will continue to grow, which will inevitably lead to a demand for more network storage for virtual machine backups. This will also usher in greater application awareness and data deduplication with virtual server backups. The real challenge will be to implement deduplication without compromising virtual machine performance.

"If you run dedupe on a VM, you'll put more workload on the VM [CPU]," says StorageIO Group's Schulz. In the near term, an external data deduplication appliance may be necessary to achieve performance goals. There are other performance issues with server virtualization that will increasingly be addressed using optimized hardware

chipsets, such as Intel Corp.'s vPro processor technology and Q35 Express Chipset.

While conventional backups will rely upon backup software for proper restoration, affording a small amount of native security, virtual machines are complete, self-standing system snapshots that are far simpler to restore than a backup volume. Encryption is another component in the virtual backup environment, but few virtualization users have made security a major priority yet.

Ultimately, the future of such tools remains murky. Experts note that virtualization vendors may shift the backup burden to third-party developers. "I think the first step for them [virtualization vendors] would be to create APIs for backup vendors," says ESG's Whitehouse, noting that backup vendors could then build new applications or add features to their existing backup products that would utilize those APIs to provide better and more refined backup products. ⊙

Stephen J. Bigelow is the features writer for SearchStorage.com.

Top 5 Reasons you need a solid Business Continuity Plan

5. Downtime is more expensive than Paris Hilton's bar tab.
4. Sleepover parties are no fun when you're rebuilding a server.
3. "My dog ate the backup tapes" won't fly as an excuse anymore.
2. Your CEO won't learn your name for the first time during an outage.
1. You might just make it home for dinner tonight. And every night.

The success of your company rides on your ability to keep vital business operations running in the event of an infrastructure failure. Extended downtime can be fatal, and the longer you wait, the more damage it can do to your business processes and more importantly, your bottom line.

Overland Storage solutions keep key operations going strong. Our rock-solid tiered data protection works at every level of defense—all within your current infrastructure.

Let Overland Storage help you create a winning Business Continuity plan that makes sense for your business and gives you peace of mind.

Get your FREE white paper today!
www.overlandstorage.com/bc

Overland STORAGE®

Check out the following resources from our sponsors:

ASEMPRA

Register to Attend an Upcoming Asempra Product Demo Webinar
Asempra Business Continuity Server Overview
Asempra Business Continuity Server Datasheet

hp invent

HP StorageWorks EVA 4400
TCO White Paper: EMC, NetApp, and HP Midrange Storage Arrays
StorageVirtualization and the HP StorageWorks Enterprise Virtual Array

IRON MOUNTAIN DIGITAL™

Storage-as-a-Service: Data In, Value Out
Regulatory Compliance: How Digital Data Protection Helps
Lessons Learned

OperationsIT

Enterprise-class Business Continuity Without the Enterprise-class Price
View a Demo of Our Advanced Recovery & Continuity Appliance
Operations IT Webinars on Business Continuity

Overland STORAGE

Business Continuity Planning Resource Page
REO Series is the World's Best-selling Disk-based Backup and Recovery Appliance
Calculate Your Cost of Downtime Now!

SEPATON

The Forrester Wave: Enterprise Open Systems Virtual Tape Libraries, Q1 2008
TCO Report: Comparing Costs of Tape, VTL and Deduplication Solutions
Weekly Webcast: Sepaton Enterprise VTL and Deduplication Solutions

syncsort

Fast and Simple Disaster Recovery with Syncsort ExpressDR
Innovative Business Continuity Solutions with Syncsort XRS
Near-instant Oracle Cloning with Syncsort AdvancedClient Technologies