IDC TECHNOLOGY SPOTLIGHT

Optimizing Approaches to Enterprise Backup and Recovery

February 2014

Adapted from *Worldwide Disk-Based Data Protection and Recovery 2013–2017 Forecast* by Robert Amatruda and Laura DuBois, IDC #242452

Sponsored by ExaGrid

IT organizations are faced with ensuring that backups occur in the shortest amount of time and are not operationally disruptive as well as maintaining a timely backup window as data continues to grow. However, because companies are demanding faster backup, restore, and recovery, many are foregoing investments in physical tape infrastructure in favor of general-purpose disk storage systems and/or purpose-built backup appliances (PBBAs). Traditionally, there has been reliance on approaches involving the physical removal of media such as tape. However, increasingly, disk-based technology is being used for backup and restoration, and the advent of data deduplication has greatly accelerated the transition from tape to disk. In addition, the accelerated adoption of virtual servers and desktops is causing IT organizations to review and modernize their data protection architectures and processes with disk-based solutions. Deduplication reduces the amount of disk required and allows organizations to eliminate tape, but how deduplication is implemented greatly impacts the backup window. This Technology Spotlight examines these trends and the role that ExaGrid's disk-based backup with deduplication solution plays in these rapidly transitioning markets.

Introduction

The use of disk in the data protection process has accelerated in recent years because it relieves many of the backup bottlenecks associated with using traditional storage methodologies such as tape. Companies are demanding faster backup, restore, and recovery to meet aggressive backup windows and recovery times.

As a result, many are foregoing further investments in physical tape infrastructure in favor of general-purpose disk storage systems or purpose-built disk storage systems, which IDC calls purpose-built backup appliances. The advent of data deduplication has greatly accelerated the transition from tape to disk. However, as discussed in this Technology Spotlight, data deduplication may come with a price.

Backup Requirements

IT organizations are faced with a number of challenges. The first is ensuring that backups occur in the shortest amount of time and are not operationally disruptive. Another is maintaining a timely backup window as data continues to grow, thus increasing the backup window time. IDC expects that data growth will continue to grow exponentially and will increase sixfold in the next five years. It's particularly challenging to manage and safeguard critical data being created outside the confines of the central datacenter.

In addition, data has to be readily available for rapid restoration. Traditionally, there has been reliance on approaches involving the physical removal of media such as tape. However, increasingly, disk-based technology is being used for backup and rapid restoration of data in the event of a disruption, whether caused by natural disaster or human error. If this happens, companies need to recover their

data in 24–72 hours to ensure uninterrupted operations. In the past, companies relied solely on tape for the operational recovery of their data. However, recovering data from tape can be a very cumbersome process. In many cases, companies need quite a bit of infrastructure to support a restoration process that can take days to complete.

Drivers for increased investment in disk-based data protection solutions include the need to improve backup window time, provide faster recovery times, enable seamless integration with backup applications, improve performance of backup resources, and lower operational and capital costs. The accelerated adoption of virtual servers and desktops is causing IT organizations to review and modernize their data protection architectures and processes with disk-based solutions.

The basic backup requirements for media storage are as follows:

- Shortest possible backup window
- Backup window that stays fixed in length with data growth
- Fast local restores
- Fast instant recovery for virtualized environments
- Fast offsite tape copy (if tape is still used for offsite disaster recovery)
- Fast disaster recovery if the offsite location has a replicated disk-based backup appliance
- Low cost up front and over time; buy what you need as you need it
- Low cost to operate
- No future hidden costs
- No forklift upgrades
- No product obsolescence

Given this context, the sections that follow analyze each of these requirements against the different storage technologies: tape, general-purpose disk storage systems, and purpose-built backup appliances.

Achieving the Shortest Backup Windows

When backing up to tape, companies are limited to the number of concurrent backups (backups run in parallel) by the physical number of drives in the tape library. In addition, tape is a serial read-and-write media, and with backup, it suffers from the "shoeshine effect," meaning tapes in the tape library go forward and reverse as data is written, which slows data writing. Backup to tape is not the most optimal for backup performance, resulting in longer backup windows.

Backing up to disk staging is faster for several reasons. First, disk is a random access medium and the data can be written quickly. Second, you can set up virtual concurrent jobs using multiple volumes or NAS shares. The limit is not physical. Disk staging allows for faster and more reliable backups and restores. However, the cost of disk quickly becomes expensive after two weeks of retention, which is why disk staging fronts tape and tapes are still used for onsite retention.

Deduplication reduces the amount of disk required, which allows you to eliminate tape, but how deduplication is implemented greatly impacts the backup window. Deduplication in the backup software is performed at either the client or the media server. In either case, there are not enough compute resources to power high-ratio deduplication.

Backup Windows That Stay Fixed in Length

The starting backup window length may or may not change as data grows based on the storage media used. Using tape, you can add drives as data grows, but once all drive slots are filled, the backup window will expand with data growth. If the backup window encroaches on user production hours, then a larger tape library needs to be acquired.

However, using disk staging, you can set up as many virtual targets as you desire to have concurrent backups. As data grows, more storage and more targets can be set up, allowing the backup window to stay fixed. Economically, you can afford one to two weeks of backups on disk and then, because of economics, the rest is stored on tapes. For backup software that has deduplication, the more data you add, the slower backup becomes unless you continue to add more media servers.

Inline deduplication appliances are the most challenged. As data grows, there is more data to be deduplicated. Inline approaches have a front-end controller architecture, and as the data grows, so does the amount of deduplication that needs to occur. Since you cannot add additional compute power, the backup window grows until it encroaches on production. The only way to bring the backup window back in line is to replace the front-end controller with a bigger and more powerful front-end controller (forklift upgrade).

On the other hand, post-backup (i.e., post-process) deduplication appliances with a scale-out architecture have full servers, each with processor, memory, and bandwidth as well as disk. As the data and resulting deduplication load grows, full servers are added, bringing more processor, memory, bandwidth, and disk to the task at hand. As a result, the backup window does not grow.

Data Restores, Instant VM Recoveries, Tape Copies, and Disaster Recoveries

There are four different situations where data restores and recoveries are needed: local restores, instant onsite recovery when a VM goes awry, weekend offsite tape copy and, in the event of a site disaster, offsite disaster recovery. All of these are different needs requiring access to backup data.

Tape is the most difficult for all restores and recoveries. The tape needs to be located and loaded. The media is serial read, so you have to stream to the position on the tape where the desired data resides to perform a restore or recovery. Tapes are often damaged by wear, dirt, humidity, or temperature, and restores or recoveries aren't always possible.

Disk staging solves all the problems of tape restores. The disk is fast random access, and the data is reliably there when needed. The only challenge is due to economics; tapes are still required for longer-term restores and recoveries as the cost of disk for more than one to two weeks of backups is cost prohibitive.

Inline deduplication appliances reduce the data on the way to the disk and therefore only store deduplicated or dehydrated data. Whenever a restore or recovery is required, the data has to be rehydrated in order to restore or recover it. One hundred percent of tape copies and instant recoveries come from the most recent backup, and 80% of onsite restores come from the most recent backup. To accommodate the need for fast restores, fast offsite tape copies and instant recoveries, the most recent backups need to be kept in their full un-deduplicated form. Appliances that have a landing zone and post-process deduplication that retain the most recent backups in their full form and then store longer retention in a deduplicated form offer the best of both worlds: fast restores, tape copies, and instant recoveries as well as low-cost disk with deduplication to replace tape. Since post-process backups using deduplication have the most recent backup in its full form, all copies, restores, and recoveries are as fast as reading from disk. The benefit of this approach is that it provides the performance of disk staging. In addition, one can store weeks to years of data on disk, alleviating tape altogether.

Cost Considerations: Up Front and Over Time

There are many cost considerations that need to be factored in for each of the different storage technologies:

- Cost per terabyte
- How much capacity must be bought up front
- Cost to operate
- Future hidden costs
- Forklift upgrades
- Product obsolescence

Tape backup can be the least expensive up front. The first challenge with tape backup is what size tape library to buy up front. If you buy a tape library that has been sized for your current data volume and your data grows, you will run out of capacity and need to buy a bigger, faster tape library. Handling tapes is much more manual than any other process, so the IT staff costs for tape are higher than for any other process.

Tape has other costs as well, such as the requirement for companies to continue buying more tapes over time. In addition, tapes wear out as they are reused, so they need to be replaced with new tapes. There is also a cost to transport the tapes to an offsite location for disaster recovery. And last, tape libraries are constantly being upgraded, resulting in obsolescence of a tape library that is only a few years old. Tape libraries have a large number of moving parts and, as a result, wear out faster than disk, resulting in increased maintenance costs typically in year four or five. For disk staging to tape, the costs of the tape library are the same; however, straight disk is an added cost. Disk staging in front of a tape library onsite is the most expensive option as you have to add the cost of the disk to the cost of the tape library, plus the cost of the tape and the cost to transport the tape. If you have more than a couple weeks of disk staging, the cost of disk rises quickly.

Deduplication in backup software to straight disk appears to be less expensive up front as you simply add straight disk, which is low cost per terabyte, behind the backup media server. However, since deduplication is compute intensive, and since the backup software has other tasks to run, it cannot afford to provide much processor or memory to the deduplication task. Therefore, the deduplication found in backup software is less aggressive than a dedicated disk-based backup appliance.

Deduplication rates in backup software are typically 2:1 up to 7:1. As a result of lower deduplication rates, there are two hidden costs. The first is that as retention goes up, the disk required goes up at a rapid rate. At a 7:1 deduplication rate, the amount of disk required goes up at a rate of 3x disk-based backup appliances with deduplication.

There is a crossover point where straight disk behind a backup application with deduplication costs more than a dedicated disk-based backup appliance. The crossover point where straight disk is more expensive is typically between three and six weeks of retention. Over this retention, dedicated disk-based backup with deduplication appliances is less expensive.

Considering ExaGrid's Approach to Data Backup and Recovery

ExaGrid offers a scalable disk-based backup with deduplication solution. Its scale-out GRID architecture is designed to constantly adjust to growing data backup demands to permanently shorten backup windows and eliminate expensive forklift upgrades. In addition, ExaGrid's landing zone retains the most recent backups in their full non-deduplicated form to provide fast restores, tape

copies, and recoveries. With offices and distribution worldwide, the company serves customers across a variety of major industry sectors. Key features include:

- Fast backups up front with permanently short backup windows as data grows
- Fast restores and offsite tape copies
- Instant recovery of full systems, VMs, and files to minimize downtime
- Low total cost over time designed to eliminate "forklift" upgrades and product obsolescence

The company's zone-level deduplication is designed to reduce the disk space needed by a range of 10:1 to 50:1 by storing only the unique bytes across backups instead of redundant data. Post-process deduplication delivers the fast backups. As data grows, the solution avoids expanding backup windows by adding full servers in a GRID. The landing zone keeps a full copy of the most recent backup on disk, delivering the fast restores, instant VM recovery, "instant DR," and fast tape copy.

The company claims that using the solution to replace tape in the nightly backup process can reduce backup windows by up to 90% and a typical 12-hour backup window can be decreased to as little as 2 to 3 hours. It includes support for advanced virtualized server recovery techniques such as instant VM recovery.

For offsite long-term retention or disaster recovery, the ExaGrid solution offers the ability to transfer backup data to an installed system at a remote location to supplement or eliminate offsite tapes. It also supports multisite topologies where multiple locations can transfer backup data to a centralized site for DR protection. Costs can be contained because deduplication only moves changes, requiring minimal WAN bandwidth. The costs and reliability issues associated with tape handling, shipment, and storage are significantly reduced or eliminated.

The system includes standard appliances along with software to deliver a complete turnkey solution for disk backup with data deduplication. The appliance is rack-mountable and uses standard components, including Intel processors, enterprise SATA/SAS drives, and Gigabit Ethernet connection(s). The system works seamlessly with popular backup applications to preserve investments in backup applications and processes.

Scalable next-generation GRID architecture with full servers provides plug-and-play expansion. Multiple appliances allow full backups of 1TB, 2TB, 3TB, 4TB, 5TB, 7TB, 10TB, 13TB, or 21TB with corresponding raw capacity of 5TB, 7TB, 9TB, 11TB, 13TB, 16TB, 26TB, 32TB, and 48TB, respectively. Any size appliance can be mixed and matched in multiple different configurations, with up to 10 servers combined into a single GRID of up to 480TB raw capacity, allowing full backups of up to 210TB.

Challenges

ExaGrid's key strengths are the company's GRID-based highly scalable and flexible architecture focused on midmarket and SME customers. ExaGrid maintains a strong new customer acquisition strategy against larger, more capitalized competitors. ExaGrid has invested integration with fast-growing third-party application vendors such as Veeam in the virtual protection market. ExaGrid has a strategic co-marketing relationship with Veeam and in addition recently partnered with CA Technologies to offer a combined value proposition for CA ARCserve D2D and ExaGrid for customers looking to alleviate the constraints of tape-based backup and recovery.

ExaGrid is an adroit competitor in the PBBA market, and we expect it will continue to execute crisply against larger companies and competitors. The biggest challenge the company will face is growing its installed base of new customers and displacing larger more formidable competitors in the marketplace. We believe ExaGrid is well positioned to gain share in the PBBA market, with a growing list of partners and new customers alike.

Conclusion

Today, customers are still grappling with subpar backup performance as backups outstrip the allotted backup window time. Customer strategies for data protection and recovery continue to be dictated by aggressive SLAs, rapid recovery, and ease of integration in existing environments. As a result, firms are embracing more disk-based data protection technologies, including purpose-built backup appliances to protect and recover data and applications. These appliances include features such as data deduplication, compression, encryption, and replication. Meanwhile, unabated data growth continues to pressure IT staff and protection and recovery processes, leading customers to consider alternative backup methods and targets.

The evolution of backup processes has provided storage administrators some relief, moving from tape for backup and recovery to using general-purpose disk with software and now PBBA systems utilizing storage optimization such as data deduplication. IDC has discovered that PBBA systems provide measurable benefits to customers looking to improve their existing data protection and recovery processes. Furthermore, PBBAs, as the moniker suggests, are purpose-built highly tuned turnkey solutions that have preconfigured hardware, interconnects, interfaces, and software. ExaGrid offers a disk-based backup with deduplication solution that has a key role to play in these rapidly transitioning markets. To the extent that the company can address the challenges described in this Technology Spotlight, IDC believes that the company is well positioned for success.

A B O U T T H I S P U B L I C A T I O N

This publication was produced by IDC Custom Solutions. The opinion, analysis, and research results presented herein are drawn from more detailed research and analysis independently conducted and published by IDC, unless specific vendor sponsorship is noted. IDC Custom Solutions makes IDC content available in a wide range of formats for distribution by various companies. A license to distribute IDC content does not imply endorsement of or opinion about the licensee.

Global Headquarters: 5 Speen Street Framingham, MA 01701 USA P.508.872.8200 F.508.935.4015 www.idc.com