Executive Brief

Addressing the Top 4 Data Protection Pain Points Using Intelligent Backup and Deduplication

In This Paper

• Midmarket companies find their data volumes are rapidly growing
• Traditional approaches to backup and data protection are falling short
• IT needs backup appliances that tightly integrate deduplication and data protection

An HP Executive Brief. © 2013
Addressing the Top 4 Data Protection Pain Points Using Intelligent Backup and Deduplication

Introduction

In today's world of explosive data volumes, midmarket companies face a growing number of challenges when it comes to safeguarding data.

Specifically, traditional approaches to backup, recovery, data protection and disaster recovery/business continuity are falling short. This is forcing companies to re-evaluate their strategies in these areas.

Simply put, the status quo is no longer adequate. For years, companies have used disk-based systems to improve backup and recovery efforts. Such systems helped in several ways. They could be used as essentially a cache to buffer the flow of data to tape drives. In this way, backup windows were more flexible. They could also be used to keep data online that would traditionally be saved to tape libraries. This allowed for faster recovery of files and data compared to retrieval from tape, but this method is expensive.

More recently, there has been an increased use of deduplication technology to help reduce the volume of data that needs to be backed up.

However, as data volumes continue to grow, and more of that data must be retained, preserved and easily retrievable, a more robust and sophisticated backup and recovery solution is needed.

To fill that need, the market for purpose-built backup appliances that offer simplified administration, as well as tightly integrated deduplication and data protection, is on the rise. The drivers for adoption of these devices include the need to improve backup window time, to provide faster restore and recovery times, to simplify administration by enabling seamless integration with existing backup applications and to provide a hands free data protection solution for remote and branch offices.

To get a better understanding of the benefits of such a solution, this Executive Brief discusses the top four pain points midmarket companies face today with their data backup, restoration and retention efforts, and the capabilities these newer backup solutions with intelligent deduplication deliver to address these challenges.

Pain point 1: Growing data volumes increase backup windows

The first backup pain point relates to the amount of data that must be backed up.

A recent industry survey found that in 2011 one in ten companies was managing an unprecedented petabyte (PB) or more of data. Many more companies are likely to be in that range now. Data volumes have been growing, on average, by about 20 to 40 percent per annum over the last few years. And in the past year, the growth rate itself increased. At these growth rates, a company that had 500 terabytes (TB) of data two years ago will need to manage over a petabyte in 2013.

The main problem here is that while there is more data that
must be backed up, the backup window remains the same. Typically, companies back up their data at night. If it took the full overnight shift to back up a database, file servers or storage systems last year, and data volumes grew, say 30 percent, something must change or all the data will not be backed up overnight.

A common approach that companies have tried in the past would be to add more tape drives. This increases the aggregate data backup rate, allowing more data to be backed up in a given time period. However, the additional drives add operating costs as each drive must be powered, maintained and the large volume of tape cartridges managed.

A better approach is to use a solution that combines disk-based backup and deduplication.

To benefit from this approach, IT managers must first determine what role disk-based backup will play. Disk systems can serve either as the primary backup (with tape being used for disaster recovery, compliance and archival purposes) or as a holding device that buffers data intended for tape cartridges.

In both scenarios, tape and disk-based backup hardware systems must be physically integrated with other storage systems and disk-based systems must be incorporated into existing backup and recovery procedures and software systems.

The key here is selecting the right system. For instance, some disk-based backup systems use existing backup software and act as virtual tape libraries or file shares. Others simply offer relatively low-cost disk storage capacity (as compared to a primary storage system) that holds data intended for backup on tape drives. However a disk-based system is used, the system must have very high performance and throughput so very large amounts of data can be backed up within tight time frames.

Deduplication addresses the ‘too much data for a fixed backup time window’ issue in a different manner. Much of the data growth in midmarket companies today is from unstructured data. And this type of data is ripe for deduplication.

In fact, deduplication is very effective in applications where many copies of similar or even identical data are stored on disk. This is a quite common experience in most companies. The bottom line is that data deduplication reduces the amount of storage needed for a given set of files; thus it reduces the amount of data that needs to be backed up. As a result, data deduplication can help reduce the time needed to backup files and systems.

However, to realize the full potential of deduplication technology, a solution must leverage intelligence and awareness at the source, backup server, and storage device. It is not enough to just deduplicate data sitting on a storage device, which is what many companies already do. A system that has an awareness of what data is already in the deduplicated storage, and can intelligently decide at the production server level whether to send new data, will improve backup speed since only changed data is transmitted from the production server to the storage solution.

When looking for a solution to alleviate the pain of too much data for a given backup window, the best choice is a disk-based backup system with tightly integrated and intelligent
deduplication. One additional point to consider is whether the deduplication works on data stored on one drive or across multiple locations.

**Pain point 2: Disaster recovery and backup are complicated and takes a lot of IT resources**

Most companies are operating on tight budgets with lean IT staffs. They are resource constrained, spending a great deal of time and money on day-to-day operations. This reduces IT’s capacity, and thus the company’s ability, to react to new market opportunities, develop systems that will improve internal operations and grow the business.

The direct link between resource issues and the challenges of managing backup is that the common approach to dealing with the growing volume of data has been, and remains, to throw raw storage capacity at the problem. This is a bad approach for several reasons.

A quick look at some market statistics puts the problem into perspective.

Budgets remain tight. One industry survey of 200 companies found that, on average, total IT budgets increased by only 1.8 percent in 2013 versus 2012. While most companies are turning to consolidation and virtualization to reduce the number of servers they need to purchase and manage, the same is not happening with storage. Total disk storage systems capacity grew 24.4 percent year over year in the third quarter of 2012.

Adding storage capacity to match data growth incurs a capital expenditure. But worse, each new drive adds operational costs. The devices take up data center floor space, require electricity to power and cool and carry annual warranties and service contracts. With tight budgets, money allocated for storage device CAPEX and OPEX costs simply drains dollars from other ventures.

Furthermore, with IT being asked to do more with fewer resources, any additional time needed to manage growing storage arrays cuts into the work an IT staffer can do on another project.

Compounding matters, adding raw storage capacity does nothing to help reduce backup challenges. The additional data stored on the drives still needs to be backed up, thus requiring more staff the time to perform the longer backups.

Here again, intelligent deduplication software that is tightly integrated with storage and backup solutions can help. Such a solution significantly reduces the amount of data that needs to be backed up and stored resulting in reduced OPEX costs, lower power and cooling expenses, and fewer service and warranty contracts. IT staff have fewer devices to manage, lowering administration costs and freeing up their time to work on other projects. Additionally, backup requirements are reduced helping companies meet tight backup windows.
Pain point 3: Safeguarding data in geographically dispersed offices

To compete with the geographical reach of their enterprise counterparts, most midmarket companies need branch offices to provide local services for their customers. Whether it is a regional service center, a sales office, or walk-in storefront, the employees in that location need the same level of IT services as those in the main office. And just like their colleagues in headquarters, workers in remote offices are generating large volumes of data and working with the same constantly growing corporate databases. As a result, the data stored locally must be afforded the same level of protection as the data housed at the corporate data center. In particular, the data must be routinely backed up, and there must be a disaster recovery/business continuity plan in place.

In this regard, the problem most midmarket companies face is that many of their remote offices have either very small or no IT staffs. Those offices with staff often rely on only one or two people to be IT jacks of all trades, managing all aspects of the IT operations. This leaves little time for managing backup and disaster recovery efforts.

What can help is the use of intelligent storage and backup solutions that automates processes and can be managed from a central location. An ideal solution would allow IT staff in headquarters to get a backup and a disaster recovery copy of data at any location without the need for onsite expertise. As is the case in the main office, a solution that combines data replication and intelligent deduplication would help reduce the volume of data that needs to be backed up and subsequently moved across the WAN to the main data center.

Pain point 4: Data protection and disaster recovery needs on the rise

Backup is becoming ever-more critical as the number of declared natural disasters that have wiped out businesses has trended upward over the past few years. In addition, data corruption due to the work of hackers is on the rise, and companies must deal with increasingly stringent data retention laws and regulations for many types of data.

To put the issue into perspective, consider that over the past few years, natural disasters caused a record amount of property damage around the world. In the United States-between hurricanes, tornados, wildfires, and floods--few parts of the country have been immune. So companies must have data from different offices protected to survive a site or regional disaster.

With hackers, the main intent is typically to access data for identity theft and fraud or to steal the intellectual property of a company. In the process, data is sometimes either intentionally or accidently corrupted, rendering it useless.

And on the data retention front, rules and regulations like Sarbanes-Oxley, HIPAA, and eDiscovery laws mean more data than ever before must be kept for longer periods of time in case of audits or civil suits.

Investing in backup is futile if a company cannot restore its data fast enough to meet its business needs. Companies thus need robust backup AND recovery solutions. The reason:

“To realize the full potential of deduplication technology, a solution must leverage intelligence and awareness at the source, backup server, and storage device.”
Many companies have very little tolerance for downtime, meaning systems and their associated data need to be restored quickly after any outage or disruption. Adding more pressure for rapid recovery capabilities is the growing need to retrieve archived data to satisfy the increased number of eDiscovery requests. In such events, companies need to produce subpoenaed information such as email exchanges between two parties over several years or they risk losing a civil suit.

What is needed is a solution that combines deduplication and data replication capabilities. The deduplication is used to reduce the volume of data that must be preserved and saved at different locations. The replication technology is used to move data from one site to another. In this way, a company might choose to have copies of critical data at several locations to support its disaster recovery efforts.

The solution must also offer a way to quickly restore data and systems in case a building or site is destroyed by floods, fire or winds. Similarly, the solution must provide a way to quickly retrieve data from old backups to audits and eDiscovery.

**HP as your technology partner**

Midmarket companies need intelligent storage and backup solutions to address these pain points. This is an area where HP can help.

HP StoreOnce Backup systems with StoreOnce Catalyst software are high-speed disk-based backup solutions that with tightly integrated intelligent deduplication. The integration of the backup and deduplication capabilities makes HP StoreOnce solutions easy to manage, provides automated backup and supports disaster recovery operations.

HP StoreOnce Backup systems reduce the amount of backup data companies need to store by up to 95 percent and in many cases more. HP StoreOnce systems use an HP-developed single deduplication algorithm, which works across all of HP’s data deduplication hardware and backup software platforms. In contrast, most other solutions rely on multiple products. This bolting together of dissimilar solutions is in stark contrast to the single, unified architecture that HP offers.

Leveraging the hardware and deduplication software, HP StoreOnce Catalyst delivers industry leading backup speeds to meet shrinking backup windows. The software offers Federated Deduplication across the company and remote office data protection, managing the movement of data between remote offices and main office data centers. Best of all, HP StoreOnce Catalyst allows management through a single console, thus offering a single pane of glass to ease backup administrative tasks.
With these capabilities, HP StoreOnce systems let midmarket companies backup more data in the same time, restore data faster, support remote office data protection and enable disaster recovery. Most important, HP StoreOnce fits into an existing environment. The HP StoreOnce products are designed to work with all the popular backup applications, so companies do not need to rip and replace.

For companies that want to replace their current backup application, HP offers HP Data Protector. HP Data Protector uses the same HP StoreOnce deduplication algorithm enabling it to share deduplicated data with HP StoreOnce appliances in a deduplicated state. Additionally, Data Protector supports context-based searching of backed-up data. This means companies can recover files by searching for what is in them rather than by file name. This is possible because Data Protector is now powered by Autonomy’s Intelligent Data Operating Layer (IDOL) platform. One important application of this capability is eDiscovery.

Taken together, the HP technologies in the HP StoreOnce product line ease the pain of managing data protection, backup and restoration. In particular, the systems help companies deal with growing data volumes and tight backup windows, ease management burdens by simplifying operations and reducing the number of devices required for storage, provide data protection for branch offices, and aid in regulatory data retention and retrieval operations.

For more information about HP StoreOnce solutions with integrated deduplication, visit www.hp.com/go/storeonce

1 http://h20195.www2.hp.com/v2/GetPDF.aspx/4AA4-1785ENW.pdf


3 http://www.analytics-magazine.org/special-articles/546-idc-forecasts-big-growth-for-big-data-

4 http://www.executiveboard.com/it-blog/2013-it-budget/

5 http://www.nasdaq.com/article/idc-q3-worldwide-external-disk-storage-systems-factory-revenues-rise-33-20121207-00004#.UO3adm_AfJY

6 http://www.fema.gov/disasters/grid/year