WHITE PAPER

InMage
Business Recovery... Simplified.

InMage: A Compelling Alternative to Deduplicating Virtual Tape Libraries

Introduction

For most organizations, data is growing at 50% - 60% a year or more, and this rapid pace is expected to continue for at least the next five years. High data growth is uncovering a number of storage management challenges, while at the same time administrators are faced with evolving business and regulatory mandates to remove the business impacts of data protection operations, minimize data loss on recovery, shorten recovery times, improve recovery reliability, and provide higher availability for mission-critical applications. Implementing disaster recovery is also a top priority for many organizations, although administrators are understandably concerned about possibly increasing the cost and complexity of their storage infrastructures.

Two new technologies have become available in the open systems market within the last several years that can potentially help administrators deal with the problems of high data growth and increasingly stringent recovery requirements. Virtual tape libraries (VTLs) use disk as the storage media, but use a software layer to make that disk appear to backup software applications like tape. This "tape emulation" allows customers to leverage the random access and reliability characteristics of disk to shorten initial backups, improve recovery times, and increase recovery reliability while continuing to use familiar, backup software-based processes with minimal change. Disk can be quite a bit more expensive than tape on a $/GB basis, however, and this is where the second technology – "data deduplication" – comes in. Data deduplication decreases the amount of storage capacity required to store a given amount of data by using a very specific process to remove redundancies in stored data sets. Data deduplication breaks data down into small "chunks", fingerprints those chunks using hash algorithms, and stores those chunks in an index. As new data is stored and duplicate chunks are found, they are replaced by pointers that reference a master index (the deduplication repository). Certain data sets, like those comprised of weeks or months of daily backups, have very high levels of redundancy, allowing deduplication technologies to achieve data reduction ratios of 95% or more over time. When deduplication is combined with VTL technology, it can make this use of disk as a backup medium much more affordable since deduplication can allow a given terabyte of data to store as much as 20TB or more of backups.

If you are considering a deduplicating VTL, InMage Systems would encourage you to consider the following: our technology offers a compelling alternative to deduplicating VTLs that provides the same business benefit (leverage the concept of data redundancy to decrease the cost of storage capacity), is not limited to specific backup software or configured target storage subsystems, and provides a solution not only for disk-based recovery, less expensive backups, and remote disaster recovery, but also offers other advantages of interest to storage administrators like improved recovery granularity, complete elimination of backup windows, higher application availability, improved security for data in-flight, and I/O trend analysis capabilities that span multiple servers.

Defining The Business Benefits You Want

If you're thinking about a deduplicating VTL, then chances are that you're experiencing problems completing backups and/or performing recoveries. Here are some of the problems you may be facing:

- You aren't able to meet backup windows, either because your data sets have grown too large or because the available windows within which you can perform backups are shrinking
- Your backup frequency is resulting in too much data loss when recoveries are required
- Your tape based infrastructure requires too much time to perform recoveries, or results in recovery reliability problems that you're ready to spend money to solve now

You may also be considering how to implement a disaster recovery plan that supports better recovery point (RPO) and recovery time objectives (RTO) than your existing approach that relies on regularly shipping backup tapes to an offsite storage location. Replication technologies offer some significant benefits here, but your data must be sitting on disk to replicate it. Many organizations are suffering from one or more of the above problems, and the fact that deduplicating VTLs may also offer replication is an added incentive that increases the value proposition they can provide.

Because it leverages disk, a VTL will provide better performance than tape for initial backups, for object-level recoveries (which is what most restore requests are for), and possibly for disaster recovery (if their replication capabilities are used). A VTL will also likely provide better recovery reliability. These benefits accrue because VTLs are based on disk, and have nothing to do with their tape emulation capabilities. Despite these advantages, note that VTLs still treat backup as a discrete operation; they can make backups faster, but as data grows they will ultimately succumb to the same problem that forced you to consider moving away from tape: you won't be able to complete your backups within the available windows.

Note also that deduplication technology is exclusive in how it is deployed; if you get it as part of a backup software package, it puts a heavy load on the backup client as backups are deduplicated before they're sent out across the wire, and it is a solution specific to that particular backup software. If you have other backup software you're using, you'll need to buy a second deduplication technology. If you purchase it as a storage target (NAS or VTL), then it is typically deployed in front of a particular storage array whose storage it is deduplicating – if you want to leverage deduplication against another array, you'll need to purchase another appliance. This can make target-based deduplication a good approach if you have all the data you want to deduplicate in a single, large array, but it is not cost-effective if your data is spread out across a variety of different storage subsystems and architectures (NAS, DAS, SAN).

Storage Realities for Mid Market Enterprises

If you're like most mid market enterprises, you have a highly heterogeneous environment that is composed not only of different server and storage equipment, different storage architectures, and various applications, but also several different backup software packages. Your backup is probably tape-based today, and you may think that implementing replication technologies to support DR is too complex or expensive. Management is probably pushing you to meet more stringent recovery requirements, both locally and remotely, as well as to manage critical applications to achieve higher overall availability. You may have considered conventional shared disk clustering technologies, but find them too complex and expensive. Your IT organization is likely oversubscribed, leading to a strong preference for solutions that are easy to deploy and manage, and underbudgeted, which makes you very cost-conscious as you consider how to address your most pressing problems.

A VTL could help you address some of these issues. If you are currently using a tape-based infrastructure, implementing a VTL could decrease backup windows, improve recovery granularity by using disk-based snapshots, shorten recovery times, help address recovery reliability concerns, and possibly enable a DR configuration. Note, however, that you will still have a backup window, you will still have data loss on recovery, you will have to buy at least one VTL in each location to support replication for DR purposes, you'll have implemented a solution that works for only one storage array target (depending on where and how you deploy the unit), and you will have done nothing to help manage applications for higher availability.

InMage: A Compelling Alternative

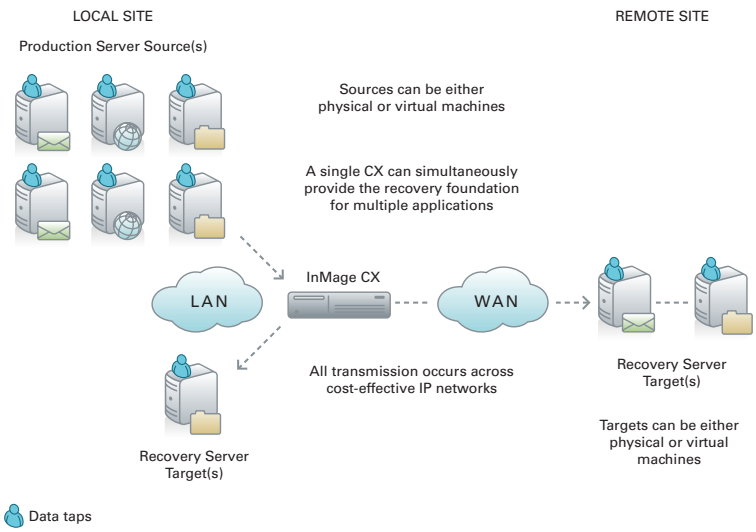

Figure 1. *Unlike a VTL, InMage can simultaneously solve application and data recovery problems, both remotely (for DR) and locally (for backup), for multiple storage repositories, backup software products, and application environments.*

InMage offers a software-based recovery solution deployed in the network which combines remote DR, local backup, and application failover (both remote and local) into a single platform. Leveraging a foundation of disk-based recovery technology, InMage definitively resolves backup window, RPO/RTO, recovery reliability, and automated application recovery issues with a solution that can stand alone or be seamlessly integrated as a "front end" to one or more existing backup software products and tape-based infrastructures. Application-aware functionality is available for a number of key enterprise applications, plugging into the InMage core platform (called the CX) and leveraging a set of centralized management practices that are common across all protected application environments. InMage's architectural design flexibly and simultaneously accommodates Windows, Linux, and Unix servers, physical and/or virtual servers, heterogeneous storage and storage architectures, and any application environment on a platform with a low entry price point that scales cost-effectively to configurations protecting hundreds of servers across multiple locations.

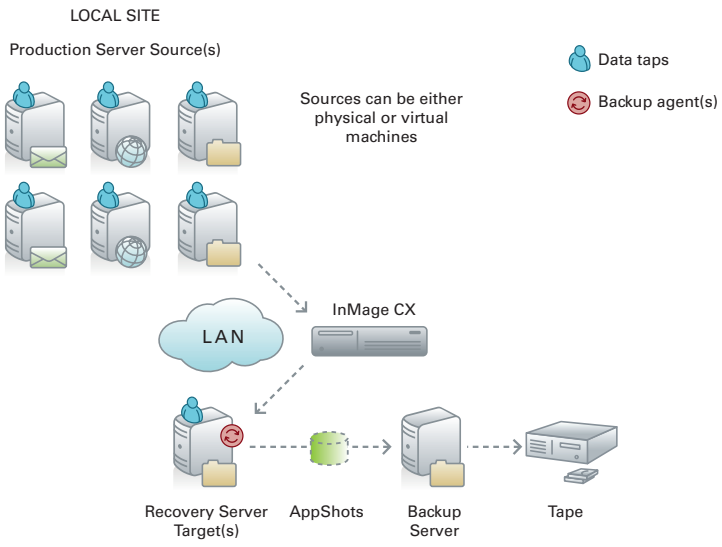

Figure 2. *InMage can effectively "front end" existing backup infrastructures. Here, AppShots for different applications can be created and mounted on the Recovery Server target, then backed up using existing backup software to dump the data to tape. This can be done either at the local site (as shown) or at the remote site.*

InMage does not implement index-based data deduplication, but achieves the same benefits by applying multiple technologies. When first implementing either InMage or a VTL, a baseline must be performed that is akin to a full backup. In InMage's case, after this baseline is performed, we only ever send change data across the network. If you attempted to deduplicate this data stream, you would realize very little data reduction: we have effectively already removed all the redundancies. On top of that, we leverage WAN acceleration technologies such as TCP optimization and compression, implemented in the CX, to cut down on the amount of data we have to send across the WAN. We also employ bandwidth-shaping technologies that allow administrators to implement policies that can guarantee quality of service metrics are met for other applications sharing the network with InMage. Finally, in cases where you have failed over to a remote site, run production there for several days, and then want to fail operations back over to the original site, we leverage delta differencing to minimize the amount of data that has to be sent back across the WAN to re-synchronize the sites.

Because of the way that index-based deduplication works, it can take from several weeks to several months for these technologies to reach their advertised data reduction ratios. This is because the ratios get higher as the deduplicating VTL sees more backups that only change 3-5% per day. Contrast this with InMage which takes no time to "ramp up" to its data reduction ratios: within minutes of deploying it in production, it is achieving its maximum data reduction ratios.

Creating Application-Consistent Recovery Points

Crash-tolerant applications use journaling or logging techniques to maintain high performance while still being able to reliably recover from failures. Reliable recovery is defined as maintaining data integrity for all recovered data. Crash-tolerant applications may lose some data (generally this is limited to only the most recent data for transactions that were in process at the time of the failure), but any data that they are able to recover will not be corrupted.

Crash-tolerant applications will recover more rapidly from application-consistent recovery points. Many enterprise applications provide APIs that allow external applications to call for applications to put themselves into a consistent state. In Windows, this API is called Windows Volume Shadowcopy Services (VSS), in Oracle it is called RMAN (or "hot backup"), while other applications have their own APIs. At a high level, these calls will force the application to stop accepting new transactions, complete all writes in process and flush all buffers to disk. Conventional backups are done from application-consistent recovery points. A common practice with disk snapshot technology has been to put applications into a consistent state, take a disk-based snapshot of that point (which may take only on the order of minutes), and then perform a backup to tape from that snapshot at a later point in time. VTLs can use snapshots as a backup source.

InMage can also interact with various applications through their APIs to put applications into a consistent state. But instead of waiting for a disk snapshot to be taken, InMage just inserts a marker into the data stream at that point, a task that takes less than a second. If an administrator wants to reference that particular point for recovery (or other) purposes, they just refer back to it using the InMage GUI, retroactively creat-

ing a disk-based image of that data state (called an AppShot) if and when it is needed. If it never is needed, then that image is never created.

InMage leverages the same APIs that VTLs do to ensure that application-consistent recovery points are available for use, but InMage offers a better model. Inserting a marker takes less time than creating a real snapshot, which is what occurs when VTLs are used, thereby impacting production applications less. InMage doesn't take up space for snapshots that never get used, because they never get created unless they're needed. You also don't have to manage a snapshot bank, pruning older snapshots to make room for newer ones. If you create a snapshot, modify it, and then want to refer to it in its original state again, a VTL will not let you do this unless you consciously make at least two copies. With InMage, just select the appropriate point in the data stream again and re-create it – no extra up front planning is required. And with InMage, you additionally get immediate access to every other possible recovery point, which is very valuable in performing root cause analysis of data corruption or other application failures. A VTL does not provide that.

Point By Point Comparison: InMage vs Deduplicating VTL

Backup windows. InMage uses continuous data protection (CDP) technology to capture writes to protected applications as they occur in real time, effectively spreading the "backup" out across the entire day. For most applications, this traffic, which flows from the production server source(s) to the CX across a LAN, is not even noticeable. We don't shorten the backup window, we remove it entirely. VTLs can complete backups faster than tape-based approaches can, but backup is still a discrete operation where you try to bundle up all the changes since the last backup and push them through your LAN all at once.

Data loss on recovery (RPO). Since InMage captures and retains a record of every write, it can retroactively create any desired recovery point, providing for infinite recovery granularity. This level of recovery granularity translates to the least amount of data loss on recovery possible, since InMage can always present the most recent reliable recovery point. VTLs perform backups at discrete intervals, and those intervals determine which recovery points are available. For a VTL, RPO is determined by backup frequency. For example, if a VTL is used to provide 2 backups per day, the RPO is 12 hours; 4 backups per day would result in an RPO of 6 hours. How many business transactions does 6 hours of lost data translate to? InMage supports an RPO that will be no more than a few seconds to possibly a few minutes old at the most.

Note, however, that for DR RPOs, there may be a considerable difference. Some deduplicating VTLs do not replicate data sets until an entire backup (composed of multiple scheduled backup jobs) is complete. If it takes 6 hours to complete the backup, VTL replication cannot start until that is complete. And even for some deduplicating VTLs that claim to start replication before backup jobs complete, they can't support a specific RPO until all the metadata for that particular set of backup jobs has been replicated to the remote site. If you are considering a deduplicating VTL to support DR requirements, make sure you understand how the options you are considering operate. InMage replicates data captured by the local CX continuously using asynchronous replication over IP. Once a particular write in the data stream hits our recovery target (regardless of whether it is remote or local), then the recovery window has been extended to include it. This allows us to support RPOs of seconds to minutes from the remote site.

Because InMage can retroactively create one or more recovery points reflecting previous points in time, we offer additional advantages that VTLs cannot. We will always offer the optimum recovery point for any scenario; with a VTL, you will be limited to choosing the closest recovery point from available backups. If that particular backup was somehow corrupt, then you will have to go back to the previous backup, which could easily result in the loss of another 6 to 12 hours of data, depending on your backup frequency. With InMage, we can go back to the previous write in the data stream (before the corruption occurred), and perform the recovery from there. The result: significantly lower data loss on recovery.

The flexibility to retroactively create multiple recovery points offers some nice advantages for root cause analysis that VTLs just cannot support. When attempting to determine the cause of a problem, it can be very helpful to be able to examine multiple recovery points around a particular event suspected of causing the problem. InMage can provide those points, a VTL cannot.

Shortened recovery times (RTO) and recovery reliability. Once a recovery point has been identified, a VTL and InMage will support comparable RTOs. Recovery reliability should be comparable as well. This is because both approaches leverage disk as the underlying recovery media. A deduplicating VTL may introduce a small amount of additional time to convert deduplicated data back into its original format so that it can be read by an application, but this additional time probably won't noticeably impact RTOs.

Support for existing environments. Some VTLs (but not all) can be used with any backup software. Similarly, VTL appliance gateway products can generally use any back end storage. InMage can be used with any backup software to create the "front end" shown in Figure 2, and can leverage not only any heterogeneous storage, but multiple different storage architectures (DAS, NAS, SAN, iSCSI, SCSI, FC) simultaneously.

In addition, InMage supports encryption of data in-flight for improved security. InMage also supports a unique I/O profiling tool, called the InMage Analyzer, that can accurately map RPO capabilities given existing network bandwidths, allowing you to deploy already knowing the costs that will be associated with meeting a given set of recovery requirements. And because the I/O streams of potentially multiple servers all flow through the central CX, we can provide I/O profiling that can help determine I/O peaks and valleys across networks during the day and track I/O growth rates over time. This data is valuable in planning virtual server deployments as well as forecasting increased network bandwidth requirements over time. Because of its architectural design, InMage offers additional benefits that deduplicating VTLs cannot. Here is a short summary of these features with their associated benefits:

- A deduplicating VTL provides for data recovery, but InMage provides for more comprehensive recovery capabilities that cover both applications *and* data
- Replication is part of the base InMage solution, but it is an expensive add-on for many deduplicating VTLs that support it; depending on configurations and number of sites, a choice to deploy InMage can result in significant savings
- InMage works transparently with any backup software, while some deduplicating VTLs that are "content aware" have to be qualified to work with particular backup software products which may or may not cover the ones you're using
- InMage's design supports a "host off-load" feature that allows you to remove backup agents from production servers yet still use backup software to perform backups against application-consistent recovery points (AppShots) if you so desire, although with the data sitting in the InMage disk-based repository and available for recoveries, most customers are at least minimizing tape-based operations if they don't entirely eliminate them; this feature minimizes overhead on production servers by off-loading all data protection operations to non-production servers, and can lower backup agent costs (since you can just run the backup agents on one or a few servers that you are mounting AppShots on to perform off-host backups for compliance purposes) as well as the business impact of backup agent maintenance operations

Conclusion

If you are considering a deduplicating VTL to reduce backup impacts on business operations, meet more stringent RPO/RTO requirements, improve recovery reliability, or help in implementing DR capabilities, then you should consider InMage. InMage provides the flexibility to continue to use your existing backup software and storage subsystems, preserving existing investments, but also offers the freedom to buy and/or use different products going forward. Leveraging an underlying foundation of disk-based recovery, it will eliminate backup windows and meet more stringent RPO requirements than deduplicating VTLs can while providing the same RTO and recovery reliability improvements. InMage offers data reduction advantages comparable to that achievable by deduplicating VTLs, but on top of that provides for automatic application failover/failback that increases application availability, works with multiple heterogeneous environments simultaneously, and offers other advantages that deduplicating VTLs cannot match such as host off-load, better root cause analysis, encryption, and I/O profiling to aid in capacity planning. You'll get a more comprehensive solution that deploys more cost-effectively and offers greater flexibility in accommodating your existing processes for recovery than deduplicating VTLs can provide.

InMage 3255-1 Scott Boulevard, Suite 104, Santa Clara, CA 95054 | 1.800.646.3617 | P: 408.200.3840 | F: 408.588.1590 | info@inmage.com
www.inmage.com

2010V1