TANEJA GROUP
TECHNOLOGY ANALYSTS

P R O D U C T P R O F I L E

Extending the Benefits of Real-Time Replication to Small and Medium Business: SteelEye DataKeeper

February 2009

Tape-based backup and recovery infrastructure is a mainstay in most small and medium business (SMB) environments. SMBs using tape for backup are very likely dealing with issues around shrinking backup windows, data loss on recovery, recovery times, and recovery reliability. Technologies are available which can integrate with existing tape-based infrastructure to address these issues, but SMBs may be concerned about cost and complexity.

Replication technologies can resolve these data protection issues in tape-based infrastructures, but also offer other benefits in terms of improving the overall availability of computing resources (servers and storage). Newer replication solutions are surprisingly cost-effective and easy to implement, particularly if they come from vendors who are focusing on providing replication solutions for the SMB. In this Product Profile, we'll take a look at some of the issues with which SMBs are dealing that can be addressed with replication, discussing value propositions associated with their use. Because Windows is a mainstay platform amongst smaller companies, we'll focus on Windows ecosystem considerations. Then we'll provide a spotlight on SteelEye Technology, a leading vendor of replication solutions that are well-suited for deployment in SMB environments.

Replication As A Data Protection Tool

Any enterprise that collects and retains data, regardless of size, must protect that data. One of the central tenets to data protection that has transcended all eras is that of regularly creating and maintaining secondary copies of critical data that can meet the need for data recovery, if and when it should arise. Manually creating secondary copies of data on physical tape has been a mainstay of data protection and disaster recovery (DR) plans for decades. As data stores that must be protected have grown larger, tape-based approaches have been challenged to meet evolving backup and recovery requirements.

With data growing at 50% to 60% a year or more, enterprises of all sizes have struggled to meet backup window, recovery point objective (RPO), recovery time objective (RTO), and recovery reliability requirements. SMB customers, with their limited IT budgets and relative lack of sophisticated IT resources, have been even more hard pressed than larger enterprises to design and execute adequate data protection and DR plans.

Being forced to do more with less, many enterprises have explored automation, hoping that it would allow them to improve the reliability of operations, increase the span of administrative control, and simplify the daily life of existing IT managers. When

87 Elm Street, Suite 900 Hopkinton, MA 01748 Tel: 508-435-5040 Fax: 508-435-1530 www.tanejagroup.com

it comes to data protection and DR, replication is one technology that can automate the creation and ongoing maintenance of secondary copies of critical production data. Simply stated, replication is a process that keeps a "source" disk and a "target" disk in sync across a network connection. Source and target devices may reside within the same computer room, or they may reside in different computer rooms, as long as those two facilities are connected across a network.

The Value Add of Replication

Many SMBs use an exclusively tape-based infrastructure today to handle data protection requirements. Critical servers in central locations generally have good local backup and recovery capabilities, but dealing with distributed data (in remote office/branch office (ROBO) environments) and DR requirements pose difficult challenges. Backup windows are getting increasingly smaller (if they exist at all), and even those servers that are backed up regularly may not support the desired RPO/RTO capabilities as evolving business and regulatory mandates demand faster recovery with less data loss. By judiciously adding replication technology to such an environment, SMBs can achieve improvements on all fronts. Replication should be thought of as a complement to rather than a replacement for backup operations.

First, replication can be used to shift backup operations to non-production servers. Replication is a process that occurs transparently and in real-time so that data can be collected from production servers as it is created and/or modified without impacting production applications. Replication can be paused whenever desired so that the target disk can be used as a source for backups. Once the backup is complete, the source and target devices are re-synchronized across the network. The production server stays on-line during the entire operation. When used in this way, replication can effectively make backup windows on production servers disappear. And because replication can occur across WANs as well as LANs, this approach can be used to perform backups of distributed data in centralized locations. This frees up personnel at ROBO locations from having to deal with backup operations, and allows centralized administrative resources that have more data protection expertise to manage backups and restores.

Replication by itself is not a sufficient backup solution because it makes only the latest copy of the data available for recovery. If this copy is for some reason corrupt, then recovery cannot be completed. For this reason, replication is often combined with snapshot technologies to provide multiple recovery points. What replication adds to snapshot technologies, however, is the ability to recover from the most recent point in time rather than the most recent snapshot, which may be hours old. If RPO is an issue, replication, and other technologies which rely on its tenets such as continuous data protection (CDP), should be considered.

Second, if replication is used to collect data for backup purposes, then by definition that data is retained on disk for at least some period of time. The use of disk as a primary

87 Elm Street, Suite 900 Hopkinton, MA 01748 Tel: 508-435-5040 Fax: 508-435-1530 www.tanejagroup.com

backup target offers a number of advantages in terms of RPO, RTO, and recovery reliability. Since data is being continuously collected, the latest recovery point will be either current (if synchronous replication is used) or likely no more than several minutes old (if asynchronous replication is used). Compare this to the once daily backups that are in common use in SMBs today. With replication, you'll be able to recover from much more recent data, minimizing data loss on recovery. Since most recoveries draw from the most recent copy of the data, most recoveries will come directly from disk. Disk supports much faster restores because the data does not have to be converted from a tape format prior to a restore and because it allows random access to data, features that can save minutes to hours for the types of object-level restores that comprise 99% of restore operations. Disks are also much more reliable than tape for initial backups and object-level restores because of their random access method, and this translates into improved recovery reliability. Industry data suggests that tape-based recovery attempts fail as much as 20% or more of the time, due primarily to tape handling and media issues. Disk does not suffer from these problems and with newer technologies such as SATA is getting very inexpensive.

Third, replication offers significant advantages in establishing both local and remote recovery solutions. Disks are often mirrored to increase data availability, but if disk mirrors exist within a single cabinet (server or storage), availability risks still exist. Replication can automatically create and maintain a separate copy of that data in a separate server or array, hosted either locally or remotely, to ensure that data is available even if the source is not.

If a DR plan exists at all today, most SMBs are physically transporting tapes by truck to one or more remote locations. Tape handling at ROBO locations is often a cause of problems, and lack of local administrative resources with backup expertise can lead to inconsistencies in completing and shipping backups to remote locations. Replication offers the ability to transmit that data electronically across secure networks with no local operator involvement. Electronic distribution minimizes backup data loss, offers much better RPOs for disaster recovery (since data arrives at the remote locations within minutes or hours rather than days), and provides the option of using any network-accessible location as a DR site (other facilities you may own, managed service providers, cloud-based storage, etc.). For DR purposes, customers can combine replication with local high availability solutions to provide for off-site recovery of not only data but also application services. In these configurations, issues such as disk, server, or software failures can be quickly recovered from locally, while site disasters can be recovered from using the remote site. In this latter case, replication ensures an up to date copy of the data at the remote site from which to recover.

To add value, replication does not need to be used against all data sets; it can be targeted at selected data, such as your most mission-critical data, or data that is housed in ROBOs that have no on-site IT expertise. Replication can be used to establish a disk-based backup tier for initial backups and very recent

87 Elm Street, Suite 900 Hopkinton, MA 01748 Tel: 508-435-5040 Fax: 508-435-1530 www.tanejagroup.com

backup data that will complement a secondary tape-based tier (which leverages your existing tape-based infrastructure) for much more cost-effective long-term storage of your backup data.

SMB Replication Requirements

When evaluating solutions for integration into their environments, SMBs typically look for ease of use, cost, and functionality – in that order. When considering replication tools, Taneja Group recommends that SMBs look for the following:

- Synchronous and asynchronous replication mode options
- Cross-platform solutions that support heterogeneous servers and storage and any and all application environments
- Automatic re-synchronization between source and target devices when replication is paused or network partitions occur
- GUI and CLI management options, with wizard-driven installation, configuration, and management in the GUI
- Features to minimize bandwidth requirements and make the most of available bandwidth
- Support for virtual machine (VM) environments

Cost is a key consideration as well, but it is impacted by several factors beyond just the cost of the licenses. Does the solution require dedicated hardware or can it be used with pre-existing hardware? Will the solution require network infrastructure expansion and if so, how much? Realize that, if you are using replication to place secondary copies of data on separate servers, storage, or sites, this implies certain hardware redundancies. Replication products which provide options to re-deploy existing infrastructure to meet redundant hardware requirements will require less capital outlay.

Windows Considerations

Microsoft does not provide a general-purpose replication product that can be used across all Microsoft environments, and encourages the use of third party solutions when customer needs require comprehensive replication capabilities. Microsoft certification programs ensure that third party products are tested and work reliably in Windows environments, providing peace of mind to customers using certified solutions. Microsoft Windows Server Failover Clusters (WSFC), the new successor to Microsoft Cluster Server, is in common use, and is a prime platform for extension with replication technologies to remove single points of failure such as shared storage disk arrays. Replication can extend WSFC to support multi-site DR configurations as well. Server virtualization can offer significant management and cost benefits in SMB environments, and Microsoft's offering in this space is Hyper-V. When evaluating third party vendor Windows compatibility, Taneja Group recommends customers look for the following:

- Microsoft Gold Certified Partner status
- Cluster Enabler packages which simplify the integration of replication technologies into WSFC configurations

87 Elm Street, Suite 900 Hopkinton, MA 01748 Tel: 508-435-5040 Fax: 508-435-1530 www.tanejagroup.com

- Support for real-time replication of Hyper-V virtual machines (VMs) as well as compatibility with key Microsoft utilities (e.g. VSS, etc.)
- Third party tool management through the Microsoft Management Console (MMC), i.e. tool GUI is an MMC snap-in

Introducing SteelEye

SteelEye is a Menlo Park, California-based company that provides data and application availability management solutions for business continuity and DR for Windows, Linux, and VM environments. Focused primarily on the SMB market, SteelEye boasts a suite of solutions based on almost twenty years of technological innovation that are cross-platform and architected with the ease of use advantages important to smaller enterprises. SteelEye sells direct but also has a thriving channel and OEM business. SteelEye Strategic Technology Partners include Citrix, EMC, HP, IBM, Microsoft, Novell, Oracle, Platespin, Red Hat, SAP, and VMware while their Technology Enablement Partners include Dell, Fujitsu Siemens Computer, SGI, Sun (MySQL), and Sybase.

SteelEye's product suite includes LifeKeeper, a high availability solution that supports either a shared-disk or a shared-nothing architecture, DataKeeper, a host-based replication solution that replicates at the block level, and several product suites targeted for DR purposes. SteelEye also markets a number of Application Recovery Kits which layer on top of LifeKeeper and provide application-specific high availability solutions for the most popular SMB applications including Microsoft Exchange and SQL Server, Oracle, IBM DB2 and WebSphere MQ, MySQL, and SAP.

DataKeeper for Windows

DataKeeper for Windows is SteelEye's pure-play replication solution, targeted for use in Windows ecosystems. In 2H08, SteelEye announced enhanced DataKeeper solutions that provide support for the latest Windows platforms, including Windows Server 2008 and WSFC. SteelEye has created customized DataKeeper packages that are specifically targeted to integrate in the field with key Microsoft technologies such as Hyper-V and Microsoft WSFC. In this Product Profile, we'll take a close look at DataKeeper, evaluate it against SMB data protection and DR requirements, and identify the value propositions associated with its deployment.

Mapping DataKeeper to SMB Replication Requirements

Replication mode options. DataKeeper supports both synchronous and asynchronous options. Having both available is important since they meet different requirements. In synchronous mode, DataKeeper updates both the source and the target in lockstep as writes occur, ensuring that both disks reflect the same data state at all times. Recovery from the target device will result in no data loss. Synchronous replication is best used to keep two devices that are relatively close together in sync, such as when replicating to a separate server or array within the same computer room, building, or campus environment. As the source and target devices get farther apart,

87 Elm Street, Suite 900 Hopkinton, MA 01748 Tel: 508-435-5040 Fax: 508-435-1530 www.tanejagroup.com

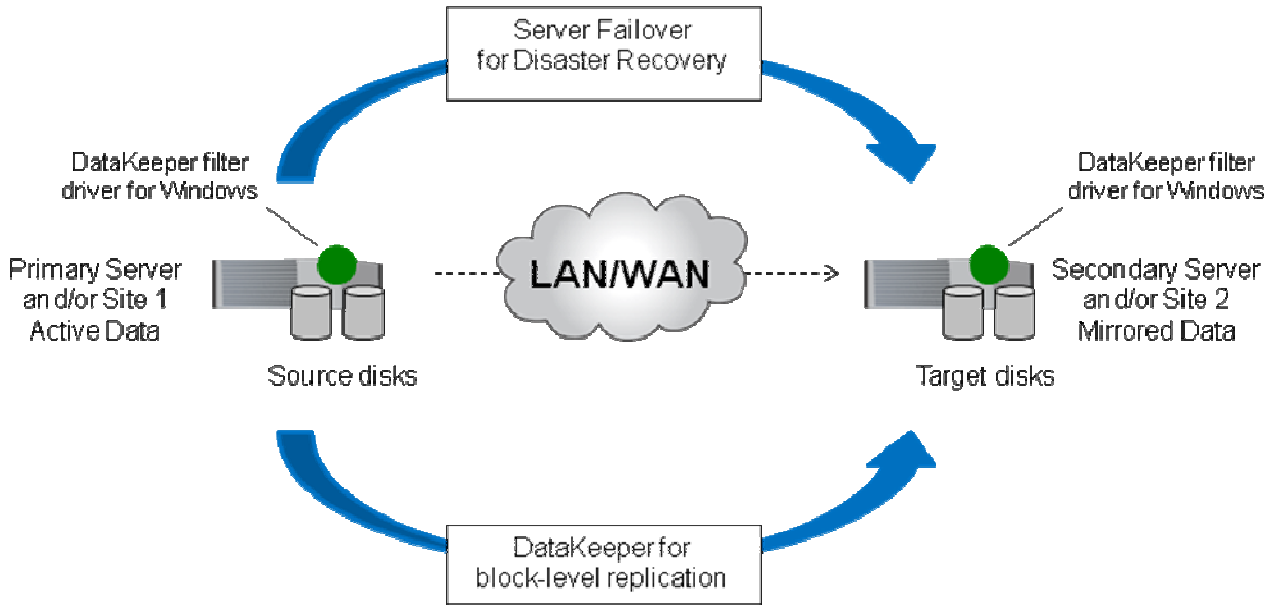

Figure 1. DataKeeper is a host-based replication solution that keeps source and target disks in sync across any IP-based network, and can be used in data protection and disaster recovery solutions.

however, network latencies may impact primary application performance. Synchronous replication is rarely used across distances greater than 30 kilometers (km).

When replication is required over longer distances, asynchronous replication may be a better fit. In asynchronous replication, writes are still sent to both the source and the target as they come in, but a write acknowledgement is sent back to the application after it has written it to the source without waiting for acknowledgement back from the target. A local intent log comes into play only during initial sync or re-sync operations. Asynchronous replication supports unlimited distances, making it a natural fit for DR requirements, but the data state of the target device may lag the data state of the source. With asynchronous replication, recovery from the target device may result in a small amount of data loss.

For additional flexibility, DataKeeper supports 1 to 1 replication as well as n into 1. Replication is set up between pairs of devices (source and target) but a single server can house multiple target devices. N into 1 configurations can be used in DR solutions just like 1 to 1, but may also support backup consolidation projects that stream data from multiple ROBOs into a single centralized site.

Cross-platform capabilities. As a host-based replication solution, DataKeeper's source and target devices for a given replication pair must be hosted on servers running the same operating system (i.e. Windows to Windows) but not necessarily the same hardware. As long as they are running Windows, DataKeeper can replicate between Dell, HP, IBM, and/or white box x86 servers.

87 Elm Street, Suite 900 Hopkinton, MA 01748 Tel: 508-435-5040 Fax: 508-435-1530 www.tanejagroup.com

DataKeeper can replicate between different versions of Windows though (e.g. Windows 2003 to Windows 2008). Replication between heterogeneous storage is also supported; all DataKeeper requires is defined source and target volumes of the same size.

Because DataKeeper replicates at the block level, it can be used to support any application. Contrast this with replication tools that leverage log shipping and are specific to a certain application (such as SQL Server or Exchange). Block-based replication is a single approach that can be used with any and all applications, resulting in less management overhead and cost in environments where data from multiple applications is being replicated.

Automatic re-synchronization. As part of their normal operation, networks can occasionally experience "transient network partitions". These are brief periods where network connections are lost, and then re-established automatically by the network. Replication depends on a network connection between the source and target devices, and must include an automatic re-synch-ronization capability to handle these types of failures without operator intervention. DataKeeper includes such a capability, and this capability is used not only to handle network partitions and/or failures, but any situation where a source and target device lose their connection. Once the connection is re-established, DataKeeper will run a quick comparison of the data states of the two devices, and then update the target device so that it reflects the state of the source device.

Management. DataKeeper supports both GUI and CLI options on Windows. If the GUI is used, management is streamlined through the use of wizards that make most installation, configuration, and management tasks point-and-click operations.

Network bandwidth optimization. Data-Keeper optimizes network bandwidth usage through four distinct features. First, block-level replication sends only changed blocks. Contrast this with file-based replication products which need to send more data to the target to ensure that the source and target files are kept in sync. Second, DataKeeper uses compression technologies to minimize the amount of bandwidth needed to transmit a given amount of data. Third, DataKeeper's replication engine has undergone a number of enhancements that allow it to parallelize replication operations, fully using available bandwidth, while still maintaining write order fidelity. Write order fidelity is critical, since if data is not written to the target in the same order in which it was written to the source, data integrity problems may arise. DataKeeper supports bandwidth throttling so that administrators can limit the amount of bandwidth that is consumed by replication, but once that limit is established, DataKeeper will take full advantage of the available bandwidth, making replication as performant as possible. Finally, block-level replication solutions like DataKeeper can use a bitmap to minimize the amount of data that needs to be sent across the network to re-synchronize source and target devices. Contrast this again with file-level replicators that use a journal. With a bitmap, only the latest state of the bit is transferred across the network during re-

87 Elm Street, Suite 900 Hopkinton, MA 01748 Tel: 508-435-5040 Fax: 508-435-1530 www.tanejagroup.com

synchronization so there is no chance that it will run out of space no matter how long the outage is. With a journal, all operations which have modified a particular bit are logged, and depending on the length of the outage and the amount of write activity, a journal can run out of space. If a journal runs out of space, then a complete re-synchronization is required, adding possibly days to a recovery effort, depending on the amount of data which must be transferred.

Support for VM environments. DataKeeper has been qualified for use and supports real-time replication of both Citrix XenServer and Microsoft Hyper-V VMs. VMware environments are supported as well.

In terms of cost considerations, DataKeeper pricing varies by server virtualization platform but in all cases is priced for SMB budgets. Support for heterogeneous servers and storage helps to minimize any capital outlay during deployment, while its network bandwidth optimization features allow DataKeeper to be deployed with a minimum of additional network infrastructure. DataKeeper's integrated network bandwidth optimization features support deployments that do not require WAN acceleration devices like some file-based replication products do – just one more way it provides a more cost-effective solution.

Windows Considerations

SteelEye is a Microsoft Gold Certified Partner and has certified its products for use in Windows Server 2003 environments (with Windows 2008 certification pending). DataKeeper is available in a Standard Edition as well as a Cluster Edition that is designed for use with WSFC, and supports integrated management, Quick Migration, and multi-site clustering for DR purposes. DataKeeper continuously replicates live running Hyper-V VMs from primary to secondary servers either locally, remotely, or both (using its 1 to n replication feature). This allows administrators to eliminate shared storage as a single point of failure, manage both planned and unplanned downtime, set up highly resilient DR configurations, and easily perform DR testing (using target VMs) without impacting production VMs. Finally, DataKeeper's Windows GUI is an MMC snap-in, providing a single pane of glass to manage integrated solutions that leverage DataKeeper with Windows-managed storage, WSFC, and Hyper-V.

Taneja Group Opinion

SMB data protection and DR requirements are becoming more sophisticated over time, and they are dealing with many of the same backup window, RPO/RTO, and recovery reliability issues that larger enterprises do. But solutions designed for use in their environments must meet additional requirements for ease of use and low cost. As replication technologies have moved down-stream and become much more cost effective, they have become very viable options for SMBs to use to improve data protection and add better DR capabilities. We have been strongly urging customers of all sizes to consider complementing their existing tape-based backup and recovery infrastructures with disk-based backup, and affordable replication solutions are one way to do that.

87 Elm Street, Suite 900 Hopkinton, MA 01748 Tel: 508-435-5040 Fax: 508-435-1530 www.tanejagroup.com

For SMBs that are concerned about both RPOs and adding resiliency to their data protection infrastructure that does not require much ongoing management, there is no better technology for that than replication. SteelEye offers a solid replication capability that they have enhanced with key features such as network bandwidth optimization, heterogeneous support, and compatibility with server virtualization technologies, in packages that are optimized for use in Windows environments. SteelEye's expertise in meeting the requirements of SMBs results in replication solutions that are cost-effective and easy to deploy and manage. If you are an SMB struggling with backup window, RPO/RTO, recovery reliability, or DR requirements, we would urge you to evaluate what host-based replication solutions have to offer. And if you're looking for a vendor that has had a lot of experience meeting the needs of the SMB segment, SteelEye is a vendor you should put on your short list.

87 Elm Street, Suite 900 Hopkinton, MA 01748 Tel: 508-435-5040 Fax: 508-435-1530 www.tanejagroup.com