



WHITE PAPER



Windows Recovery Solutions For Today's Environments

Introduction

Microsoft Windows is in wide usage in enterprises of all sizes, and moving forward this penetration is only expected to increase. Applications like Exchange, SQL, and SharePoint back mission-critical services that many enterprises cannot afford to be without for even short periods of time. Administrators are required to protect these environments, and must be able to respond quickly when recovery is required. As Windows environments become more mission-critical, recovery must include not just local plans but disaster recovery plans as well.

While many administrators equate "recovery" with "backup", there is a strong argument to expand the definition. While data is important, it is useless without the applications which drive the business-critical services on which end users rely. Recovery is really a continuum that spans the range from file-level recovery to system-level recovery to site-level recovery. By definition this includes both data and applications. If a conscious decision is made to leave application recovery as a manual process, it introduces additional risk, making the recovery process time-consuming and very dependent on the skill set of the administrator(s) performing the recovery.

The requirements along the recovery continuum are consistent: administrators must be able to rapidly and reliably perform recoveries on demand for data and applications, both locally and remotely. Yet in most Windows environments, there are a number of point products, along with manual processes, used to meet recovery needs. This leads to additional complexity, higher cost, and increased risk.

High data growth rates are already making it very difficult for conventional backups to provide adequate protection for the more mission-critical Windows applications. Administrators have a number of different products and tools from Microsoft and third party vendors from which to choose, but putting the right recovery strategy in place for your enterprise requires a strategic re-consideration of the problem. This white paper is intended to help you think about that problem with knowledge of pertinent market trends and next generation recovery technologies.

Re-Considering the Problem

Traditional recovery strategy has involved periodically making a copy of production data and storing that copy separately from the production servers. For "local" recovery, daily copies were often made to tape, while for "remote" or disaster recovery, weekly copies were often sent to an off-site location far enough away from the primary data center that a disaster, such as a hurricane or earthquake, would not impact them.

When data sets were small, the tape-based approach was feasible and cost-effective. But as the Internet grew in popularity, data sets – particularly unstructured data like files – began to grow at rates never before seen. And applications like Exchange, SQL, and SharePoint are growing rapidly as well, all adding to the data deluge. According to IDC, data has been growing at 50% - 60% a year or more since 2005, and they are forecasting those growth rates to continue. What has become

increasingly clear is that tape-based data protection strategies, if they are not already obsolete, are quickly becoming so for many production environments.

Integrating Disk As A Backup Target

Replacing tape with disk as a backup target makes it easier to deal with the data protection requirements of larger data sets. Random access disk is much better suited to the requirements of backup and recovery of data than serial access tape, and it is much more reliable. The use of disk allows initial backups to complete much faster and more reliably, and enables the use of snapshot technology to provide recovery copies. Disk-based snapshots of incremental backups, enabled through change based tracking in key Microsoft applications, allows "backups" to be completed much more quickly than in the past. This means that instead of just one backup per day, you may be

able to use snapshots to make several recovery copies per day, lowering the average expected data loss on recovery. Disk also supports faster and more reliable recoveries than tape for the types of recoveries (brick, object or file-level) that Windows administrators spend most of their time doing.

If you are not already thinking of integrating disk as a backup target, you should. Most backup software can accommodate disk as a backup target, and there are a number of snapshot products on the market that leverage disk as the recovery media. But if you are integrating disk and making no other changes, you are selling yourself short by just using disk as a short term, tactical response to your recovery problems.

The use of disk opens up the possibility of moving from traditional "point in time" based data protection regimens to continuous ones, a critical consideration given the industry's high data growth rates.

From "Point in Time" to Continuous

Think about the long term impact of high data growth rates. The performance characteristics of tape have not been able to keep up, and moving to disk improves backup and recovery performance. It does not, however, change the nature of "point in time" based backup. As data continues to grow, the backup window, data loss on recovery, and recovery time issues will re-appear. This is because the real problem is "point in time" backups. Even though technologies like incremental backups and the use of disk targets for backup do help to ease the pain of "point in time" backups, they are ultimately just a stopgap. Deduplication technology, by the way, does nothing to improve backup. But it does make the use of disk, which does improve backup, more affordable.

Changing from a "point in time" to a "continuous" approach has a real strategic impact. Continuous approaches include technologies like replication and continuous data protection (CDP), and are game changers for two reasons. First, because they collect changes continuously from applications, in real time, they completely eliminate the idea of performing a discrete backup. Dealing only with changes to an application between backups (instead of a full backup) has already become a best practice because it significantly reduces the amount of data that has to be processed. But it still imposes a lot of overhead on production servers, networks, and storage subsystems while the backup is being done. For example, if a 400GB database has only a 5% daily change rate, then at some point each day you will have to back up 20GB of data. Dealing with that 20GB of data all at once impacts the performance of production applications and can cause quality of service problems on your networks. Because of the impacts that backups impose, performing more than one backup per day is likely not feasible. But if you *could* do more than one backup per day, it would decrease the amount of data that

has to be processed during each "backup" cycle, lowering the overhead. Continuous protection technologies take this to the extreme by updating the backup copy at the same time that the primary storage is updated when an application write occurs. Transferring the 20GB of data from our example is now spread out across a 24 hour period each day, effectively making the data protection regimen transparent from the point of view of the load on servers, networks, and storage subsystems.

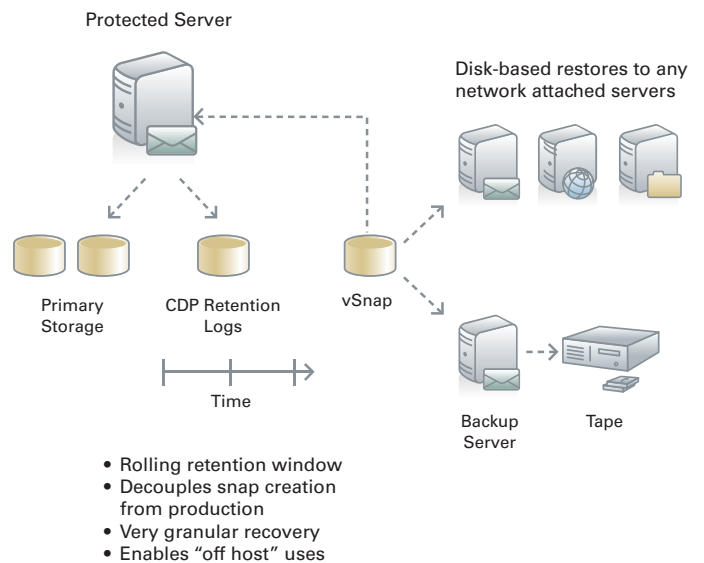


Figure 1. Think of CDP as "tivo" for your data. The steady stream of writes is captured in the CDP retention log, allowing the creation of virtual, disk-based snapshots representing data states at any previous points in time with zero impact on production servers. Snapshots can be discarded after use and re-created again if needed.

Second, continuous protection technologies make data recoverable as soon as it is written, not just once it's backed up. This minimizes data loss on recovery in many cases to zero and even in the worst cases to near zero. With a single backup each day, the average amount of data you can expect to lose on a recovery is 12 hours. For many applications, that may not be a problem but for mission-critical ones, that may be a big problem. For this reason, many Windows shops use disk-based snapshots, taking several a day to minimize data loss on recovery. Because of the Windows Volume Shadow-copy Services (VSS) API, disk-based snapshots can be taken much more quickly and with less impact than conventional backups – even incremental backups. If two recovery points a day are better than one, the reasoning goes, then four must be even better. Continuous protection technologies again take this to the extreme, offering a recovery point at literally any previous point in time (within a defined window) where a write occurred.

While this may sound good from a recovery granularity point of view, what about storage capacity costs and the overhead associated with taking “snapshots” that often? Interestingly, continuous protection technologies like CDP don't take snapshots – they record every write in a log and retain the appropriate metadata to re-create any selected previous point in time, retroactively and on demand, as a snapshot only when it is needed. CDP dispenses with all the hassle of managing snapshots (pruning, etc.) and puts less overhead on production applications than disk-based snapshot approaches do. And because CDP effectively deals only with incremental changes, studies have shown that it requires roughly as much storage as a backup target that is deduplicating at a ratio of 20:1.

The take-aways from this section are:

- Disk should be used strategically, not tactically, in data protection for best results
- High data growth rates are making point-in-time backup approaches obsolete for many mission-critical applications

Continuous approaches enabled by disk, like replication and CDP, offer a more strategic resolution of data protection issues with less overhead and better granularity for less data loss on recovery (improved RPO), with excellent advantages in terms of speed of recovery (faster RTO) and recovery reliability as well.

Application-Consistent Recovery

In the past, when applications were shut down to perform a backup, the backup copy was application-consistent. As global, 24x7 operations moved the industry to on-line backups, software vendors had to modify their applications to enable application-consistent, on-line backups. Application snapshot APIs were the answer to this, with the most popular ones in Windows environments including Microsoft's own VSS, Oracle Recovery Manager (RMAN), and SAP's *backint*, among others. Application-consistent recovery is important because applications can be recovered faster and more reliably from these types of recovery points than they can from crash-consistent recovery points (the only other alternative for reliable recovery). When you need to quickly restore an application service that is down, you will want to recover it from an application-consistent point in time.

The use of disk as a recovery medium is a pre-requisite to using application snapshot APIs. Using these APIs is a data protection best practice in Windows environments today. VSS is a public, documented API that allows third party products that support snapshot technology to request that major Windows applications such as Exchange, SQL, and SharePoint put themselves into an application-consistent state so that a snapshot can be taken. Data protection products (i.e. backup software) have traditionally interacted with production servers through backup agents which work with VSS.

As virtual server technologies penetrate the enterprise, they are being used more and more in production environments. Virtual servers tend to be provisioned at much higher utilization rates than physical servers, making the use of backup agents a non-starter in most situations due to overhead issues. To offer a lower overhead option, hypervisor-level APIs have been introduced by some server virtualization platform vendors that allow disk-based snapshots to be created without having to directly interact with virtual servers. It can be tricky to use hypervisor-level APIs to create application-consistent snapshots of applications running in the guests, depending on which server virtualization platform is being used. If you are limited to the use of crash-consistent snapshots in virtual server environments, you can expect recovery to be more complex and in some cases less reliable. It is very important to understand how you can create application-consistent recovery points for Windows applications. To ensure the best recovery options, you will want to generate at least several application-consistent recovery points per day per application environment.

Asynchronous Replication

Replication creates and maintains a copy of production data between a “source” and “target” volume across a network. Synchronous replication will keep the two volumes in exact lockstep at the expense of production application performance, while asynchronous replication can support very long distance configurations without impacting production applications. The “target” volume in asynchronous replication configurations may lag slightly, depending on the network latencies introduced by the distance, which may affect the achievable RPO at the target site. Heterogeneous, asynchronous replication over IP has become very popular as a way to create cost-effective disaster recovery configurations.

Replication by itself is not a viable data protection scheme since it only maintains the latest point in time. With replication, if the source volume becomes corrupted, then the target volume becomes corrupted as well. Replication has been combined with snapshots to address the data corruption issue. If the latest point is corrupt, then an administrator can go to the next most recent snapshot and attempt recovery from there.

Replication can be performed at the block, file, or transaction level and can be continuous or scheduled. Each of the approaches has implications in terms of efficiency (network bandwidth utilization) and RPO. When creating DR configurations, the cost of the bandwidth between sites is often the largest cost component in a multi-year total cost of ownership (TCO) calculation, so efficiency is a prime consideration. Continuous replication will deliver better RPO performance than scheduled replication.

Shared Nothing Clustering

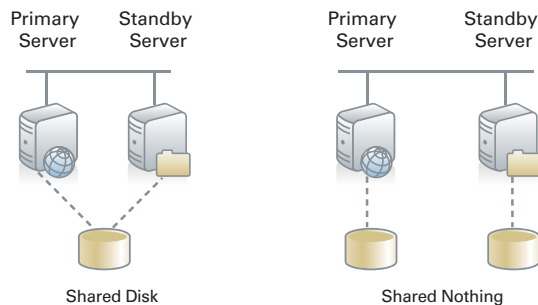


Figure 2. In a shared disk cluster, both servers are physically attached to the same disks whereas in a shared nothing cluster, each server has its own dedicated disks which are kept in sync through replication.

Technologies like application-consistent snapshots, CDP, and replication all deal with data recovery issues, but Windows administrators must also plan for application recovery. When application high availability (HA) is a requirement, some form of application failover/failback has often been employed. These approaches automate application recovery, providing a less riskier, more predictable recovery process than manual approaches. Early HA configurations used shared-disk cluster architectures, but these have been largely superseded by shared nothing configurations which are simpler to deploy, handle wide area failover much better, and are more reliable. Shared nothing clustering depends on some form of underlying replication to keep source and target volumes in sync, and can recover an application on top of a target volume if necessary.

Microsoft has really embraced the shared nothing clustering architecture in applications like Exchange and SQL. Using native tools like Cluster Continuous Replication (CCR) and Standby Continuous Replication (SCR) on Exchange and log shipping on SQL, customers can build their own HA and DR configurations. SharePoint relies on underlying SQL databases, so log shipping can be used in SharePoint environments.

By providing application-specific approaches to shared nothing clustering, Microsoft is proliferating point product solutions. Shared nothing clustering has real merit, but enterprises may consider the added complexity and cost of using application-specific tools to build their own recovery solutions.

Evaluation Points for Windows Recovery Solutions

In strategically crafting recovery solutions for Windows environments, administrators will want to keep the following planning points in mind:

- How quickly is your data growing and what are your daily change rates? What are your objectives in terms of backup windows (data protection overhead), RPO, RTO, and recovery reliability?

- If you are implementing disk as a recovery target, are you looking for tactical or strategic solutions for recovery?
- Are you trying to solve backup, disaster recovery, or both sets of problems? If disaster recovery is part of the problem you are trying to solve, do you have sufficient wide area network (WAN) bandwidth available to meet your target RPOs?
- Do you have physical *and* virtual servers that need protection in your environment? Are all the servers you want to protect Windows or do you have any other platforms (Linux, Unix, etc.)? Which server virtualization platform are you using? If you are using more than one and have applications on all that need protection, this is a critical point to consider.
- Do you plan to leverage application snapshot APIs and application-consistent recovery as an option in your environment?
- How much data protection overhead can you incur on physical and especially virtual servers without impacting the performance of production applications?
- Are there any production servers on which you want to leverage automated application recovery to meet RTO and/or recovery reliability targets? Are both automated failover and failback important to you, or do you care mostly about just automating failover?
- How many separate products do you need to deploy to meet your recovery requirements?

InMage Windows Recovery Solutions

InMage offers an integrated recovery solution that covers both data and application recovery, locally and/or remotely, that leverages next generation disk-based recovery technologies. InMage provides a single platform that can replace backup agents and tape-based infrastructure, application-specific replication products like CCR, SCR, and log shipping, and availability clustering solutions for a cleaner, simpler, less expensive approach to recovery. Customers solve local backup, disaster recovery, and application HA requirements with solutions from InMage.

As a Microsoft Certified Partner, InMage supports rich recovery functionality for Windows environments:

- Application-consistent recovery options using Windows VSS for application snapshot creation
- Automated recovery (failover/failback) for key Windows applications like Exchange, SQL, SharePoint, and Oracle as well as NTFS file services on Windows Server 2000, 2003, and 2007

- Extremely low overhead data protection which works well for physical server environments and is particularly suited for virtual server environments like Microsoft Hyper-V and others
- Easy to use, browser-based management GUI leveraging wizard-driven installation for the same key Windows applications mentioned above

InMage combines CDP, heterogeneous asynchronous replication over IP, application failover/failback, WAN optimization, and disk-based recovery into a comprehensive recovery solution for Windows environments that provides the same simple deployment model for all application environments, regardless of whether they are running on physical or virtual servers.

Using block-based CDP for maximum efficiency, InMage captures data from production servers using a data tap (filter driver) that generates only 1-2% overhead (much lower than the backup agents which it replaces). Application-consistent recovery points can be marked based on policies established by the administrator using VSS with much lower overhead than conventional disk-based snapshot products. Instead of copying change data like a disk-based snapshot does, InMage just inserts a "bookmark" into the CDP data stream, a task which takes only a few seconds and moves no data. This very low overhead approach can meet very stringent RPO and RTO requirements, and make copies of production data sets available for many purposes besides just recovery without impacting production servers in any way.

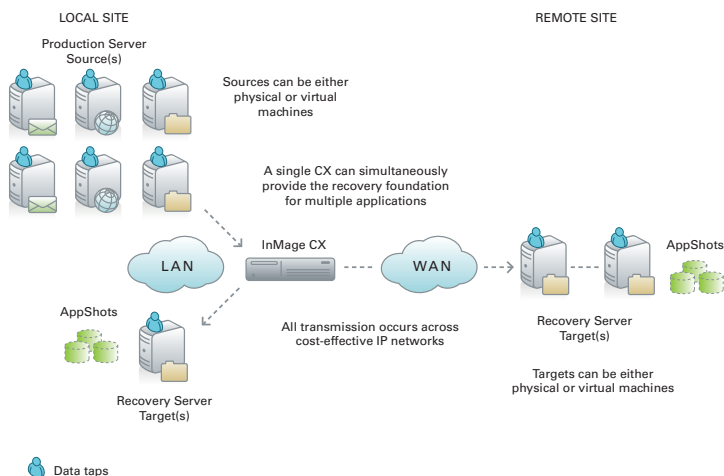


Figure 3. In the InMage architecture, data is captured from production servers and transmitted to the CX, a LAN-based appliance, using CDP. From there it is distributed to one or more targets using asynchronous replication. AppShots are the application-consistent recovery points most often used to support rapid, reliable recovery. Automated failover can occur at the push of a button at any location where a target resides.

InMage's asynchronous replication is flexible, supporting N into 1 and 1 to N topologies, and scales without adding any additional overhead on production servers. It leverages IP networks for all data transmission, and includes integrated WAN optimization to help keep network bandwidth requirements low. Our WAN optimization technologies include TCP optimization, various compression modes, and I/O throttling to help preserve quality of service for other applications that may be sharing the same networks.

Application failover and failback is driven by a set of production, fully supported recovery templates that can handle failure identification, recovery point selection, application restart, and network client re-direct for a variety of different applications. These templates can be customized to support other applications as well. Recovery automation provides faster recovery than manual processes, offers predictable recovery performance, and removes risk since success does not depend on the skill set of the administrator.

InMage supports the heterogeneity that is a fact of life in most IT shops. Although it offers strong support for Windows environments, InMage provides the same set of industry-leading recovery functionality for both Linux and Unix environments. All the major server virtualization platforms are supported, including Microsoft Hyper-V, VMware, Citrix, and KVM, as are all types of storage, including DAS, SAN, and NAS.

If you are implementing a disaster recovery configuration, understanding the server, network, and storage resource requirements up front allows you to accurately gauge recovery performance metrics, storage capacity, network bandwidth requirements, and therefore costs. InMage uniquely provides a tool, called the Analyzer, which will profile the needed information so that you know this information up front. The Analyzer is a small software component that is installed prior to deployment and left to run in the background for several weeks. At the end of that period, a number of reports can be generated that map everything from application growth and change rates, achievable RPOs based on existing bandwidth (and any "what if" analyses you may want to run), and how network load varies over time, to storage capacity requirements for the CDP Retention Log. This ensures that there are no surprises when the solution is deployed to production.

Conclusion

With evolving business and regulatory mandates, planning to meet recovery requirements is a moving target. High data growth rates, concerns about production server overhead, the insertion of virtual server platform technology, and the increasing number of recoveries driven by corruption problems (viruses, etc.) are driving a sea change in how recovery is handled for

mission-critical applications. There are some incremental improvements that can be made within your existing processes to tactically address protection issues such as backup window, RPO, RTO, and recovery reliability, but these are stopgap measures that delay a strategic re-thinking of how your recovery solutions need to be designed.

Looking forward, continuous protection technologies offer much better options than "point in time" based approaches across all major recovery metrics, for both data and applications. Other complementary technologies, such as recovery automation and WAN optimization, also need to be considered to create the most reliable and cost-effective solutions to keep businesses up and running. Microsoft provides some very valuable native tools, like the VSS API, that can be combined with next generation technologies (disk-based recovery, CDP, asynchronous replication) and architectures (shared nothing clustering) to create reliable, enterprise-class recovery capabilities for even the largest Windows environments.

Complexity drives cost, and it is on the increase with the varieties of storage architectures, different operating system platforms, the introduction of virtual server technology, and the number of different recovery products required to protect them. InMage offers unique value in that it provides a single, integrated recovery solution that can provide industry-leading recovery capabilities for Windows environments, regardless of whether they are based on physical or virtual servers, cover key Windows applications like Exchange, SQL, SharePoint, Oracle, Blackberry Enterprise Server, SAP, and NTFS, among others, and extend those capabilities to Linux and Unix environments as well. InMage simplifies the deployment of business recovery solutions, replacing multiple existing recovery products while driving down licensing, maintenance, and training costs.

