

Asigra Televaulting

CDP makes backup: better, faster, cheaper

by
Mark Ferelli



The Experts in
Agentless Multi-Site Backup/Recovery Software

www.asigra.com

White paper

Asigra, Asigra Televaulting, and the Asigra logo are trademarks of Asigra Inc. All other brand and product names are, or may be trademarks of their respective owners. Asigra Inc. reserves the right to change or modify any of the product specifications or features described herein without notice. This document is for information only. Asigra Inc. makes no express or implied representations or warranties in this document.

©2007 Asigra Inc. All Rights Reserved.



CONTENTS

<i>TABLE OF CONTENTS</i>	2
Introduction.....	3
Continuous Data Protection (CDP) Gains Momentum	3
What about Recovery?.....	4
Implementing CDP	5
Asigra Televaulting with CDP Feature	6
Selecting a CDP Product at a Glance.....	7
The Best Reason of All.....	8
Conclusion.....	8
About the Author	9



CDP makes backup: better, faster, cheaper **asigra**

Introduction

The data backup world has changed dramatically in recent years. No change has been more dramatic or rapid than the shift from traditional tape-based backup technology to disk-to-disk (D2D) backup. Disk-based backup has enabled shorter backup windows and more rapid data recovery which has opened the way for more sophisticated backup and recovery software technologies that were not possible with tape backup systems. Software vendors have responded to the technology potential of disk-based backup with new enhanced functionality, such as point-in-time snapshots and local and remote replication in an effort to reduce the vulnerability of data loss in between scheduled backup sessions.

Beyond the pure speed advantages, disk backup is also the right technology at the right time to address the convergence of two business trends: the necessity for 24/7 data access in a global wired economy and the increasing use and importance of remote offices, meaning that more remote data is at risk than ever. According to the Enterprise Strategy Group an estimated 60% to 70% of mission-critical data is stored and used at offsite locations. Enterprise IT managers face the challenge of how to protect and manage all remote data in an era of tight budget constraints and the reality that the geographically distributed locations typically lack the IT staff to manage and monitor backup systems and to verify that the backup operations were completed successfully.

Against this backdrop of vulnerable data stored at remote offices, IT managers are also facing a fundamental change in the concept of backup data. Disk-based backup has largely removed the issue of the backup window and the focus has moved to what has always been the most critical aspect of the backup process: data recovery. Driven by business objectives in terms of application and data availability, and the disaster recovery and business continuity planning, as well as compliance with government regulations and legal discovery, data backup processes are now focused on two new metrics: Recovery Time Objective (RTO) and Recovery Point Objective (RPO). The RTO establishes a maximum duration for how long the restore process will take while the RPO sets a goal for the maximum age of the data that will be used for a restore operation. The challenge is to drive the RTO to as close to zero as possible and to have the RPO be minimal, while being able to afford it.

Continuous Data Protection (CDP) Gains Momentum

The disk-based backup and recovery strategy gaining traction in data centers of various sizes to meet the RPO and RTO challenge is CDP. The traction is especially visible among users of Exchange, where the management and compliance challenges are driving elements of the CDP marketplace.

As with any new technology there is not yet universal agreement in how CDP is designed, deployed or even an exact definition of the technology. International Data Corporation defined it this way:

“Continuous data protection (CDP), also referred to as continuous backup, pertains to products that track and save data to disk so that information can be recovered from any point in time, even minutes ago. CDP uses technology to continuously capture updates to data in real time or near real time, offering data recovery in a matter of seconds. The objectives of CDP are to minimize exposure to data loss and shorten time to recover.”



CDP makes backup: better, faster, cheaper **asigra**

While many echo this definition, not all makers and users of CDP are onboard. The underlying problem, it seems, is the issue of "granularity" - how many points are needed (in a given time) for protection to be considered "continuous." A competing definition from the Storage Networking Industry Association's (SNIA) reads:

"A methodology that continuously captures or tracks data modifications and stores changes independent of the primary data, enabling recovery points from any point in the past."

A CDP product is one that will continuously monitor an object for changes and will preserve copies of all prior versions of the object. The user will have the ability to view and access these prior versions as required. The time to perform recovery changes from hours or days to seconds or minutes. The backup window is no longer a problem because there is no longer the concept of a backup window.

Specifically, though, CDP is a cross between disk-based backup and replication. CDP continually captures all changes made to a file, and engages in tagging (versioning) objects so that they can be specifically rolled back to a particular point in time. The business value of CDP lies in the ability to restore data objects to a point before a data corruption or interruption event takes place. If you experience a data loss at 12:34:14 am, a storage admin can restore back to that particular time, or somewhere very close. CDP protects/captures data as it is written to disk. One of the great myths of CDP is the unspoken assertion that CDP is for every kind of data, all the time. This is of course untrue, since the value of data changes as a matter of time, urgency, and business dynamics. A vital document today could well be worthless three days from now, and important again in another six months.

There are two approaches to CDP object protection: file-system centric or block-based. Historically, the CDP file system approach started on the Windows environment, where most applications utilized files to hold their data. Block based CDP started in the UNIX (and now Linux) community, where database applications traditionally bypassed the file system and operated directly at the disk/block level.

Block-based CDP products started in intelligent arrays such as those from EMC and HDS and soon moved to host-based or appliance-based platforms. These CDP products operate as a layered feature of the underlying storage infrastructure, and usually operate independent of the host's file system and volume manager. Recovery is typically storage or database administrator driven, is provided through capabilities outside of the platform being protected, and is managed by the CDP implementation.

One important scenario to keep in mind when considering the implementation of CDP is that of centralized backup for the remote or branch office. Too often, basic IT tasks like monitoring the backup server and changing tapes can be missed when assigned to remote office clerical staff not skilled in IT. Using a CDP strategy over the WAN to protect branch office file servers removes the requirement for tape drive and media handling at the remote site.

What about Recovery?

There are two general principles that govern all recovery policy-making: the recovery point objective (RPO) and the recovery time objective (RTO). The RPO defines how much data you are willing to lose when you recover data. For example, if you back up twice daily your RPO would be 12 hours, which is the maximum amount of data loss that could occur between



CDP makes backup: better, faster, cheaper **asigra**

backup images. The RTO defines how long it will take to recover your business processes from a data failure. This includes not only the data recovery, but restarting the servers or applications that depend on that data. These recovery considerations must also be applied to local and remote recovery strategies.

A true CDP product protects every data change as it takes place, and the RPO approaches zero. On the other hand, with the vast amount of data being recoverable, how you choose the recovery point effects your RTO.

Some recovery points are based on time, a particular hour or minute. More usefully, however, they are event-based. Since every data change is protected, a loss event can be absorbed and yield a recovery event.

Implementing CDP

Along with the debate over an exact definition of CDP there is no agreement on the best method to deploy CDP within an organization's storage and data protection infrastructure, how to approach what is actually protected and the level of granularity required. CDP solutions are designed to be block-based, file-based or application-based. Block- and file-based CDP solutions have the advantage of functioning with a range of different applications, while application-based CDP is optimized and tightly integrated with a specific application, such as Microsoft Exchange. Potential CDP buyers should also be aware of the level of recovery granularity a particular CDP solution provides, as all CDP products are not created equal on this issue. Some products only support recovery of servers, volumes or folders and lack the granularity to recover a single file or email message.

CDP is deployed most frequently either as an appliance or as a software solution running on a server or switch with agents. A dedicated CDP appliance can deliver good performance without impacting application servers, but the hardware cost can be extremely high, especially when an enterprise needs to scale its CDP capabilities and add more appliances. Appliances can carry price tags around \$50, 000. The software solutions are billed using various licensing strategies, frequently per server. License fees of \$5000 to \$25,000 per server can be readily found. The software solutions also involve agents that must reside on each server to be protected. The more servers a user has, the more agents that have to be purchased and managed...a stumbling block to the SMB, a potential struggle for the enterprise, that might have to manage agents on hundreds (or thousands) of servers. There is a significantly better way.

Host-based CDP software eliminates the hardware expense of CDP appliance but comes with its own set of cost and complexity issues. The software solutions require that agents be installed on each server to be protected, creating management overhead and additional costs. The pricing model for this type of CDP is typical of most enterprise backup software that charges a license fee for each server or database that is protected, regardless of how often the CDP functionality is actually used by each server.

The third CDP architectural alternative is to simply incorporate the CDP functionality as a feature in a full-featured backup and recovery software suite, the simplest, most cost-effective and most practical approach to CDP. To date, only one product is available in this category: Asigra TeleVaulting.



CDP makes backup: better, faster, cheaper **asigra**

Asigra Televaulting with CDP Feature

Asigra's implementation of CDP is the first to focus on remote office CDP and is designed to work over a WAN. The CDP functionality is simply integrated as a feature of Asigra Televaulting software with no additional cost to customers and no separate CDP application or appliance to purchase. The functionality is available for file data (Windows and UNIX file systems) with agentless CDP and for Microsoft Exchange emails.

Apart from offering the CDP functionality in the software core, Asigra Televaulting software also offers a robust feature set, including retention policies management and the ability to perform data restores without interrupting CDP backups. The company has structured their software pricing on a 'pay as-you-grow' model where users pay only for the actual amount of de-duplicated, compressed data stored in the repository.

When CDP is enabled, the Asigra DS-Client monitors the specified data source; when changes are detected, data is automatically backed up offsite. The client also provides automatic retention policy enforcement, an essential feature for a business whose retention practices are dictated by regulatory requirements. Retention enforcement is enabled for both regular and CDP backup. Separate retention configurations can be established for both local and offsite data storage and eliminates the accumulation of vast amounts of data requiring storage. CDP without retention rules could necessitate an additional investment in disk hardware.

The Asigra solution uses native remote API to access data, providing continuous 24/7 protection. CDP should not be confused with mirroring or snapshot technology. Mirroring will safeguard data only from catastrophic or accidental hardware failures. If data is corrupted or removed on the primary system, it will be on the mirrored disks, too. CDP covers both hardware and data episodes.

Snapshot technology captures changes at a given point or points in time, with every snapshot checked for consistency before the next one is taken. Since the points of time must be designated in snapshots, undesirable time gaps can develop between each snapshot. CDP products capture changes continuously -- without any gaps or missing data.

Asigra's CDP implementation is a two-stage continuous backup that without agents backs up any changes on Windows or Unix or Linux servers to a local server or servers as they occur. Backup starts with this change event, and granularity is available to the pace at which data is written to disk in a consistent state. The local servers aggregate the changes, deduplicates, compresses, encrypts with up to AES 256 prior to offsite transit to the centralized storage repository at a data center and the data remains encrypted at-rest. This DS-System piece of software at the repository automatically checks and verifies data for consistency and recoverability. If Autonomic Self Healing determines a file can't be recovered, it automatically asks for it again from the local server. This provides an established consistency point for all recoveries, again providing a quick recovery time.

The agentless nature is significant. The disadvantage to an agent-based architecture is that you have to manage agents installed on perhaps hundreds or even thousands of client machines. Agentless solutions do not require another application running in the background greedily consuming IT resources, such as memory and CPU cycle time, or prone to being used in a hack.

The Asigra DS Client allows setting up and performing CDP backup as a part of its core functionality, essentially making CDP a free bonus for customers. No additional licensing costs are associated with the CDP functionality and the software licensing is purely based on the de-



CDP makes backup: better, faster, cheaper **asigra**

uplicated compressed amount of data stored offsite. Regardless of the number of servers or client workstations in use.

Selecting a CDP Product at a Glance:

One way to determine if CDP is right for your remote office or branch office locations is to ask yourself a set of qualifying questions. For example, are you worried about meeting the business SLAs established by the CIO which include remote sites? Perhaps you are being asked to measure business impact of downtime at remote sites or are looking to modify or improve your remote branch office backup site and strategy? Do you have rapidly changing data at remote sites that is critical to business operations, and are you worried about shrinking backup windows to protect that data? If you answered yes to one or more of these questions, you should be seriously investigating CDP technology for your remote sites.

There are several features a robust CDP product brings to the market. These include:

- Support of heterogeneous storage and server environments. Today's customers are refusing to be locked into a single vendor for their storage and server solution. Users should select a CDP product that doesn't restrict them to only a subset of their possible storage and server environments.
- Awareness of applications and their environments. Application recovery is becoming more complex and time consuming, users should chose a product that integrates application specifics into the CDP recovery process.
- Non-invasive to the application or server that is being protected. A CDP product should attempt to minimize any impact to the application's I/O throughput or CPU load. This is best done by keeping the CDP footprint on the application server to a minimum, and moving any 'heavy-lifting' to an external server or appliance.
- Built on a scalable, reliable platform. If the CDP product is hosted on an appliance platform, the user should have the ability to add additional appliances that can scale their CDP capacity as the data protection needs grow.
- Supports a federated application environment. Many of today's complex applications (such as SAP R3) utilize servers and storage that span multiple hosts. Customers should choose a CDP product that supports these systems, as it provides the user with a consistent, federated image for recovery.
- Supports business policies and SLAs. Companies assign different values to their different applications. A CDP product that is flexible in its support of differing protection and recovery policies can provide a better overall solution.
- Can be extended. Look for a CDP product that has functionality that can be easily extended by the customer to meet business needs.
- Tightly integrated with business continuity technologies. A CDP product that supports application clusters and remote replication provides a stronger solution then a CDP product that only provides a stand-alone solution.



CDP makes backup: better, faster, cheaper **asigra**

The Best Reason of All

CDP as a feature in a backup/recovery software package carries the obvious financial advantage and the ease of use features that Asigra incorporates into its product. But perhaps the most important reason of all for CDP to be part of a data protection package is that CDP is not a universal replacement for other forms of backup, replication or disaster recovery. CDP isn't RAID, replication, or mirroring. Each of these technologies has its place in the data center data life cycle, but copy-based strategies only allow you to recover the most recent copy of data. CDP allows you to restore previous versions back to a specific time or event.

CDP has an important place in the data protection hierarchy and ensures the recoverability of all crucial business information. But it is not for data that is relatively static or is less than mission-critical. In some cases, CDP does not provide the user any advantage over conventional backup and recovery. If the data is not changing regularly, it doesn't matter if it is continuously backed up.

If data is not changing enough to require continuous backup, making a business case for a CDP appliance is difficult to impossible. But if CDP is a feature in existing backup software, it can be enabled or disabled as the IT manager thinks best depending on the importance of the data. If an enterprise or SMB is experiencing dynamic data growth (the average is over 100% every 6 to 18 months) CDP can be a lifeline, keeping data not only protected but up to date. CDP is a powerful, in many cases an essential, data protection feature, but it should be just that: a feature integrated with powerful, flexible, backup and recovery software. Asigra is the sole vendor offering this data protection integration.

Conclusion

Continuous Data Protection is the latest piece in an enterprise's data protection arsenal. Asigra has redefined the CDP model and driven the cost of CDP deployment to zero by integrating CDP as a feature of its award-winning TeleVaulting remote office/branch office (ROBO) distributed backup software platform. CDP is not a replacement for other data protection technology. Instead, it complements existing backup, replication and snapshot technology to bring advanced backup and recovery capabilities to improve the protection of customer data. Asigra is the first company to recognize the complementary role of CDP and to integrate it into existing backup software. CDP as a feature in a backup-recovery software package carries the obvious financial advantage and the ease of use features that Asigra incorporates into its product. CDP will quickly become an integral part of a total backup solution for business and on that count, Asigra is ahead of industry with the only CDP solution integrated as a piece of total backup software suite. And with a customer cost of entry of zero, Asigra CDP offers customers a combination of unique capabilities, ease of use, integration and value.

About the Author

Mark Ferelli is an independent technology journalist and commentator specializing in computer storage. He has been an industry analyst, and was editor of Computer Technology Review for 17 years. His articles on storage technology number in the hundreds, and he has taught or moderated seminars at trade events nationwide.