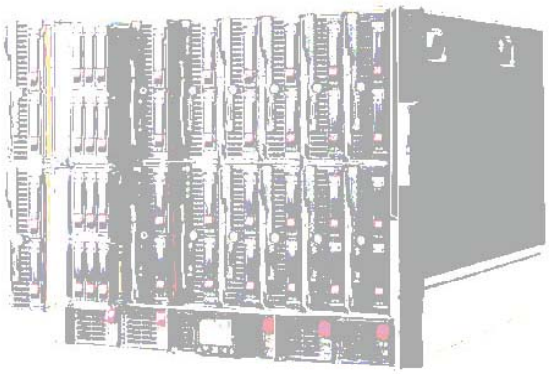
Acronis®
COMPUTE WITH CONFIDENCE
WWW.ACRONIS.COM

# CONTINUOUS DATA PROTECTION

## AN ACRONIS WHITE PAPER



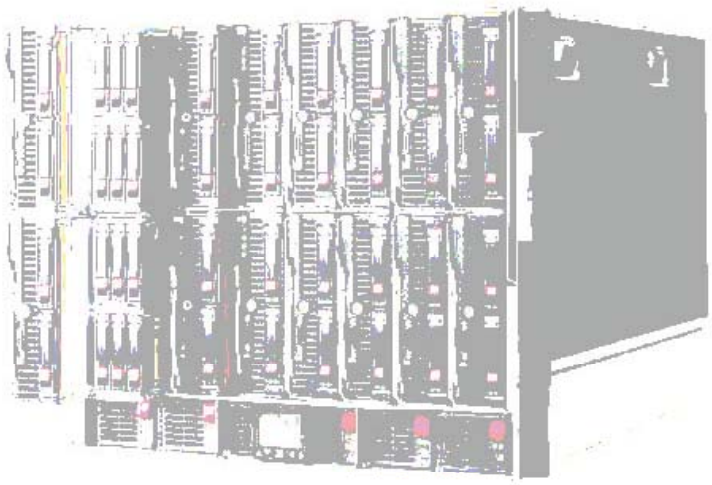
# Table of Contents



**Who should read this paper?** Any IT manager who wonders whether continuous data protection is necessary to meet an organization's recovery point objective.

# Continuous Data Protection

*Do your Recovery Point Objectives Support the Investment?*



## Introduction

With all of the recent industry interest in *continuous data protection* (CDP), you may very well be investigating how CDP may fit into your own organization's data recovery objectives.

Buzz it has, and value for some, but is CDP worth its high cost in your organization? As this white paper shows, CDP is just one of three data protection plans you can choose from. The one or ones you choose depends on how much data you can afford to lose.

Some companies who claim to offer CDP are, technically, offering near-continuous data protection products. To clear the air, we'll define CDP, reveal when it's an appropriate solution to consider, and sketch out system costs.

We'll follow with an examination of the two data protection solutions that will always complement a CDP investment – near-continuous data protection (near CDP) and traditional file-and-folder – and show where each is best applied.

## Executive Summary

Do you need *continuous data protection* (CDP)? One thing is for sure: the buzz surrounding CDP demands close attention. Unfortunately the freewheeling marketing arms of most companies offering backup and recovery products don't agree on what CDP really is.

Sitting at the top of the data protection pyramid, true CDP is easy to separate from the rest of the market's data protection offerings. Usually purpose-built, expensive when compared to other data protection systems, and boasting ultra-high performance, CDP protects data by writing *every* data change to storage as *soon as it happens.* Because it is a continuous process, a CDP system administrator can choose any point in time from which to restore the data, rather than being locked into recovering from the time of the last backup. The advantage of CDP? No data is lost, not even a single transaction.

The primary application for true CDP is very narrow: large financial firms where a single service outage in their transaction systems will put the company at serious risk. Large volumes of data are involved, and the biggest systems can support many dozens of terabytes and cost well into six figures. Any company that uses CDP will not do so exclusively. They will also use image-based approaches to protect frequently changing data stores. Meanwhile, traditional file-and-folder backups continue to be a major data protection approach in large and small organizations alike for protecting infrequently changed data stores, and they're sometimes used exclusively in businesses and organizations that are small enough to need nothing more.

### True CDP

**Frequency of use:** rare, used in mission-critical high-volume financial transaction environments.

**Cost:** many multiples of either image-based near-CDP or traditional file-and-folder backup-and-recovery approaches.

You'll need to establish a Recovery Point Objective (RPO) for each type of data you have and determine which data protection approach or approaches you should take. For instance, if you are running high-speed financial applications where thousands of transactions are completed every minute, and your RPO is measured in zero transactions lost, you may very well need CDP. Where data changes frequently, but a single block of lost data doesn't pose a problem, a near-continuous data protection solution will be ideal. In cases where data changes very infrequently, a standard 'file-and-folder' backup and recovery schemes may be the answer to your data protection concerns.

**The bottom line:** CDPs may be an important technology to consider if part of your overall IT program includes very heavy transaction processing. The relatively few companies that adopt CDP will invest in it when the cost of not doing it will jeopardize their business existence. But no business can afford CDP throughout its organization; it will also use a combination of near-continuous data protection and traditional backup and recovery products for the bulk of its data.

## The promise of CDP

Continuous Data Protection (CDP) has become part of today's management jargon, but it's not always easy to define in relation to near CDP offerings which are often confused with it.

### CDP Defined

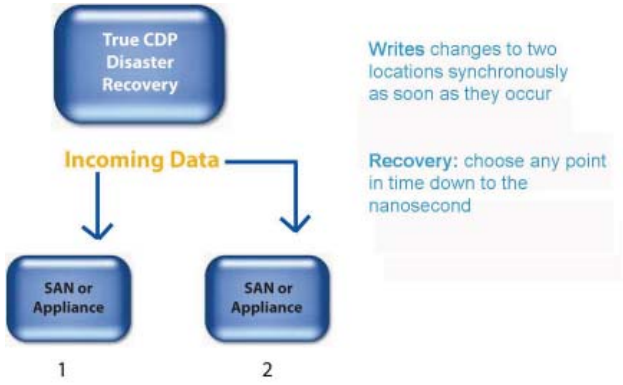
- Data changes are continuously captured or tracked. Every single modification will be written to storage.
- All data changes are stored in a separate location from primary storage.
- Recovery point objectives are arbitrary and need not be defined in advance of the actual recovery.

The Storage Networking Industry Association (SNIA) uses the "every write" definition to distinguish true CDP from other forms of backup and recovery. This approach introduces some overhead to disk-write operations, but eliminates the need for scheduled backups. CDP systems may be block-, file- or application-based and can provide fine granularities of restorable objects to infinitely variable recovery points equivalent to about a nanosecond. For instance, if it is 11:40 am and a file is corrupted, you can pick any time prior to that – say 11:35 am – before the corruption occurs to use as a restore point.

Many products marketed as CDP are actually near CDP and are often based on imaging technologies. They provide point-in-time recoveries rather than the unscheduled, arbitrary recovery points that CDP is capable of. Still, the points in time can be very closely spaced so that little data is lost. Some vendors stretch the definition of CDP to include image-based approaches, but they are not true CDP solutions.

### What is True CDP?

- Cutting-edge technology used sparingly in organizations where capturing every transaction in a high-speed environment is a matter of organizational survival
- Writes every change to storage as it happens
- Allows recovery to any point in time
- Doubles your hardware investment in servers, requires premium performance disk storage, and demands a very large investment in high-speed communications to remote locations
- Requires a substantial increase in experienced staff to acquire and manage
- Presents your organization with substantial technical issues,
- Can't support certain security measures such as encryption
- Cannot support data compression
- Is rapidly evolving and largely proprietary, so the solution you buy today may not be transferrable to the next stage of CDP evolution
- Doesn't exist in a vacuum. Any company using it will always complement it with some form of near-CDP and/or traditional file-and folder data protection



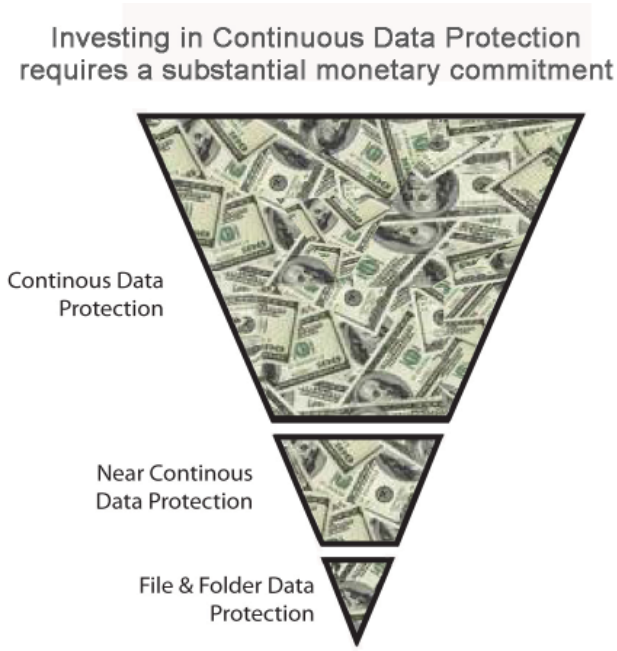

Where is CDP used? It is almost exclusively specified for use on very high-volume-transaction financial servers. The perfect example is a major Boston-based financial services company using CDP to protect their financial database, which changes at a rate of thousands of transactions a minute. However, they protect *all the rest* of their data, located on non-financial servers, with Acronis True Image, an image-based backup and recovery product, operating in near-CDP mode for frequently changed data, and with traditional file-and-folder recovery for data that doesn't change frequently.

## What does CDP cost?

One of the distinguishing features of a true CDP product is its eye-popping price. If you hire a company like EMC, they will build a purpose-built solution and plug it in. When the primary machine fails, every bit of data written into storage can be retrieved by an administrator. When you buy an EMC RecoverPoint or a Mendicino Software Infiniview product, you will start from a $50,000 base price and climb rapidly from there as you add heterogeneous SAN-attached storage, communications links, and a highly educated support staff to support it. EMC-based systems can cost $200,000 or more when fully configured, but they also have an enormous capacity of dozens of terabytes.

High-end CDP machines often combine clustering with CDP, where a SAN is replicated to a different offsite location for as much as $150,000 per SAN. You'll need two. Each location's configuration can consist, for instance, of an EMC dual-core, quad-processor system at headquarters with 16 GB of RAM and five terabytes of hard drive.

Second on the price scale are appliance-based CDP systems from companies like Exigrid and Quantum. These have base prices starting at $25,000 and range upwards of $100,000. Overall capacities of appliance-based systems are somewhat lower, but they can still range into multiple terabytes.




But there is more to pay for. Since the remote SAN has to be written to synchronously, and the distance will almost always be greater than the 1,000 feet that can be accommodated by CAT-6 copper cable, the customer has the expensive choice of operating over optical fiber (measured in dollars per foot)

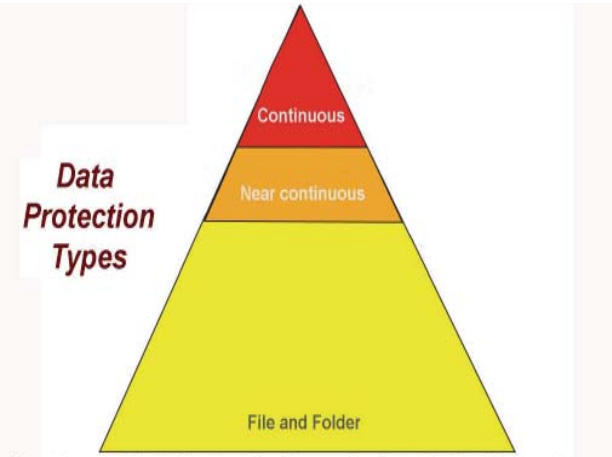
if guaranteed high-speed communications are required, or using maximized (and somewhat less expensive) Web-based communications, if they are not.

## What CDP can't (or shouldn't) do

Would you use CDP for your HR server? Certainly not, because the data changes too infrequently to warrant such a high level of protection. Neither would you use it for your Web server for the same reason. For your Exchange server you will choose near CDP because it can get nearly every message back with high efficiency. Besides, Exchange cannot be configured to support true CDP. But if you're hosting a high-volume shopping cart on your Website, you might very well specify CDP for that component of the business. You'll consider it seriously for any financial server running a high volume of transactions, where not even a single transaction can be lost.

CDP, near-CDP, and standard periodical backups can coexist very nicely in a single organization, each allocated according to the data protection requirements of each function they're protecting. Perhaps your company has a total of 200 servers. Even if high-speed financial transactions are an important part of your IT structure, only six or seven of those servers may need to be protected by CDP.

While CDP can ensure you don't lose fast-moving data, it won't necessarily protect it from prying eyes. Encryption is not an option because it is not possible to append, unencrypt, and append to the remote failover machine fast enough to deliver continuous data protection.




*Continuous Data Protection is used primarily for high-volume financial database transactions. All other data protection falls under either near-continuous or file and folder categories.*

Neither does CDP support data compression. If you have to be able to retrieve data down to the last nanosecond, there is also no time to compress and uncompress it and still protect every single one of thousands of transactions that can be written each second. As a result, CDP requires massive disk storage requirements that can range into several dozen terabytes for some applications. Moreover it must store full-size files on expensive, high-speed network disks capable of keeping up with CDP's speed of operation.

Near-CDP imaging technology, on the other hand, opens the door to a disk-saving file compression rate averaging between 50% and 60% using Acronis compression technology. Moreover, the data can be put on lower-priced disks. If your company deals with five terabytes of data, and it costs $25 a gigabyte to maintain a hard drive a year, a near-CDP product will enable you to cut the size of the backup store to roughly 2.5 terabytes, saving 2.5 terabytes of disk storage over a CDP solution, saving perhaps $54,000.

Continuous data protection products cannot automatically protect themselves against data corruption because anything that is written locally is also written to a remote location. While CDP administrators return back to any point in time to restore a previous, uncorrupted version of the data, they will lose any transactions that took place between the corrupting event and the restoration point. Administrators can search through changes in the journal file system containing the changes that were going to be made to the data, but it is labor intensive and success isn't guaranteed.

To summarize, most operations will need to demonstrate a significant need for true Continuous Data Protection before they will make this major IT purchase. And if an organization can justify CDP, it will always share data protection duties with near CDP and/or file-and-folder backup products like Acronis True Image Echo Enterprise Server or Acronis True Image Server for Windows.

## The promise of Near CDP

Many of the companies that you think must use CDP do not, including giant online auction services and the biggest online booksellers. Instead, they are more likely to employ a combination of cluster-based servers and near CDP to back up and recover transactions.

The terms near CDP and CDP are hard to pin down because they have been used interchangeably by most vendors, leaving customers confused as to what either means, as this Wikipedia entry shows:

> *There is some debate in the industry as to whether the granularity of backup needs to be "every write" in order to be considered CDP or whether a solution which captures the data every few seconds is good enough. The latter is sometimes called Near Continuous Backup. The debate hinges on the use of the term continuous: whether only the backup process needs to be continuous, which is sufficient to achieve the benefits cited above, or whether the ability torestore from the backup also has to be continuous.*

Here's how Acronis defines it.

Near CPD can be accomplished with any image-based backup and recovery product by asynchronously writing data on a secondary SAN or NAS drive, deleting it and immediately repeating the process. For example, Acronis True Image can be configured for near CDP by capturing incremental changes to disks or partitions as frequently as an organization desires. It's not a synchronous (simultaneous) write because there's a step in between – it could be a file copy or a snapshot – but data loss is minimal and recoveries can be completed rapidly.

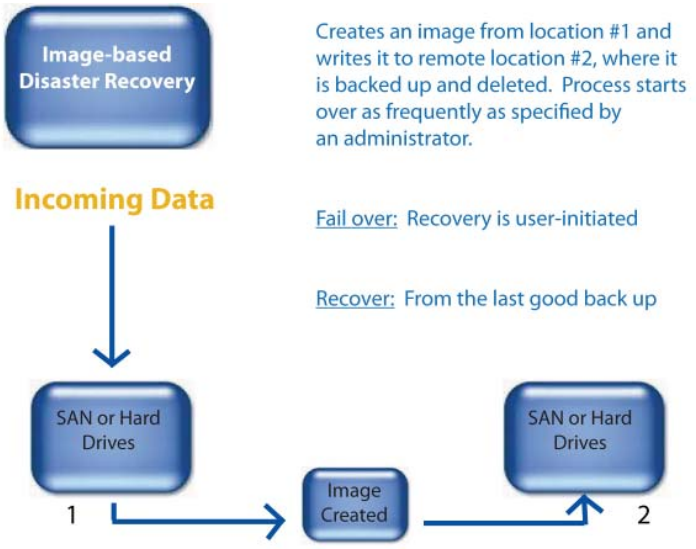
With near CDP, changes are stored in log files. They have limited size, perhaps 5MB, and they are only kept for a limited amount of time until they are backed up. Theoretically, near CPD can be configured on any image-based backup and recovery product by asynchronously writing data on a secondary SAN or NAS drive, deleting it and immediately repeating the process.

For example, if you're using Near CDP for your Exchange email system, a vendor like Acronis Recovery for Exchange allows you to script incremental backups as frequently as one minute apart for very intensively used mail systems, so very little can be lost if there is a failure. It uses asynchronous writes rather than synchronous (simultaneous) writes because there's a backup step in between, but data loss is minimal, recoveries can be completed quickly, and its relatively low cost (compared to CDP) makes it a very cost-effective purchase decision.

While an initial full backup will take quite a while to complete in a near-CDP system, all incremental backups on the remote SAN or NAS drive from that point on will be carried out in seconds. A remotely located offline machine, running some form of boot media, will continuously carry out recoveries. The system has to be configured to delete the image and the incrementals as soon as the recovery is completed to be set up as a near-CDP system. The result is a very low or no-data-loss backup and recovery program that:

- Costs a fraction of the cost of true CDP
- Saves perhaps 50% of the disk space required for CDP
- Can be encrypted for security
- Protects against data corruption

Since Acronis True Image stores the time-stamped images for as long as they are needed, it is easy to roll back the disk or partition to any of the incremental backups. Acronis True Image also provides an easy way to browse any of the files and folders in any stored image to make sure the data is accurate before restoring it.




## What is Near-CDP?

- Image-based
- Recovery to the last completed backup
- User-initiated failover to a second location
- Cost-effective approach to disaster recovery when your server is too important to depend on file-and- folder backups
- An increasingly common approach used across large and small enterprises for email and most commercial applications where data changes very frequently
- Based on Windows commodity servers and disk hardware
- Can be set up and administered with relative ease
- Supports data compression, saving 50% to 60% of disk space
- First line of defense against data corruption, allowing return to the last good state
- Can be used with ordinary SAN and NAS platforms
- No more difficult to set up than image-based backup and recovery systems you may have in place already
- Very stable and supported by the largest companies

Is near-CDP the right solution for your organization? If both data and the equipment it resides on are important for business continuity, near CDP is an excellent choice as these three customer scenarios show.

*Customer Scenario #1:* One of the largest construction companies in the United States, headquartered in Arizona, runs SAP on an aging Compaq 580 server, but no one on the current IT staff knows how to work with it. To protect the application in case of a server failure, they imaged the drive using Acronis True Image. When they later migrated to a newer platform, they simply wrote the image on a new server and were able to continue running the application without any loss of productivity.

*Customer Scenario #2:* Data doesn't change very frequently on a Connecticut-based electronics company's Dell PowerEdge server which runs a VoIP-based customer help-desk application, but a server failure can have catastrophic effects. That's why the company imaged the server using Acronis True Image. If something goes wrong, they can image its contents onto another machine and keep down time to a minimum.

*Customer Scenario #3:* A major Canadian wire and metal products company based in Terrebonne, Quebec wanted to ensure that the accounting, email, and engineering functions accessed by its 6,000-dealer network could be backed up without having to take them offline. Acronis True Image was specified to place all essential servers on a schedule of snapshot-based image backups to while allowing full or partial restores as needed.

Perhaps your company has a huge global Exchange-based email system and an active directory tree. In this case you won't need CDP, but you *will* want to consider near CDP, which can carry out incrementals frequently enough to ensure that few – in many cases *none* – of your emails are lost in the event of a failure.

## A place for file-and-folder backups

The traditional file-and-folder backup and recovery approach to data protection remains a major component in many companies' backup strategies whether or not the organization is large, small, or anywhere in between. With file-and-folder backups, recovery involves locating the last backup, loading it onto a server and searching for the appropriate folder and its file contents. File-and-folder backups are done regularly, but rarely are they performed more frequently than once a week. This backup scheme works well for data that doesn't change frequently or where the server on which it resides on can go offline for a period of time without affecting company operations.

Generally speaking, file-and-folder backups:

- Are perfect for protecting data that doesn't change frequently, like Websites and HR databases
- Allow automated failover from one location to another
- Have the lowest cost
- Can be used with ordinary SAN and NAS platforms
- Can be run effectively by less experienced administrators
- Require minimal administration
- Are very stable and are supported by the largest companies

## Conclusion

- **CDP may have a place in your company.**
- **But no company can afford to use CDP for everything.**
- **Both near-CDP and file-and-folder data protection programs continue to provide the overwhelming basis for any organization's data protection needs.**

Users need to consider these issues and decide where each of three levels of data protection can be applied with the highest level of efficiency. Most companies who use CDP will find that near CDP and file-and-folder backup and recovery are complementary solutions within a single organization.

# # #

### Learn more

1) **CDP and near CDP definitions and discussion.** Go to: http://en.wikipedia.org/wiki/CDP

2) **Storage Networking Industry Association** (SNIA). They cover a broad range of issues related to online data storage and protection.
http://www.snia.org/home

2) **Near CDP and Exchange server protection.** Download this Acronis white paper: http://www.acronis.com/enterprise/download/docs/whitepaper/?f=ARMSExchange_whitepaper.en.pdf

### About Acronis

Acronis is a global provider of storage management software that enables corporations and individuals to move, manage and maintain digital assets. Acronis sells innovative solutions for disaster recovery, server consolidation and virtualization migration, which allow users to maintain business continuity and reduce downtime in computing environments. Acronis software products are sold in more than180 countries and are available in 13 languages. For additional information, please visit www.acronis.com.



Author: Jay Woodruff Oct. 2008. Thanks to Alain Gentilhomme, Andrey Moruga, Chris Scozzari and Jerome Boutard and for their assistance.