



Continuous Data Protection — Now Ready for Prime Time

Analyst: Dianne McAdam

Management Summary

When it was first introduced, many vendors proclaimed that *Continuous Data Protection*, commonly referred to as *CDP*, is the next evolutionary step in data protection. In fact, CDP technology *does* take data protection to that next level. Traditionally, data protection schemes back up data at specific intervals. **CDP differs from these other technologies by continuously backing up data, giving IT administrators many granular recovery points. This allows administrators to think differently about protecting and restoring data. With CDP, data can be restored from many different points in time.**

Years ago, IT organizations were limited in data protection choices. Online applications were brought down at night and the volumes were backed up to tape. If data needed to be restored, IT administrators had one choice – they had to restore from the previous night’s backup tape. A commonly used metric called *Recovery Point Objective (RPO)* measures the age of the data used for the restore operation. In the case of nightly backups, RPO can be many hours; in fact, it can be as much as 24 hours. More recent innovations, such as disk copies, or *snapshots*, allow IT administrators to back up critical data several times a day. If disk copies are created every four hours, then RPO is reduced to four hours or less.

Traditional data protection techniques, such as tape and disk copies, force IT administrators to determine ahead of time when backups must be scheduled. Administrators are restricted to restoring from the last point in time copy. This can lengthen the recovery process. A second metric, *Recovery Time Objective (RTO)* describes how long the recovery process takes. If a database is restored via a copy taken four hours early, then the last four hours of transactions have to be reapplied to the database, extending recovery time. Since CDP continuously protects data, it eliminates the need to schedule point-in-time copies, and reduces RPO and RTO. **Today, many IT organizations support business applications that must be available all of the time. The demands of these business applications can be best protected by CDP solutions, but CDP solutions should not be viewed as a replacement for more traditional data protection techniques.**

IT organizations need to understand CDP technology to make the right decisions about where it fits (*and does not fit*) into their environment. In fact, CDP allows IT administrators to think differently about how they can support business processes. Read on to learn how.

IN THIS ISSUE

➤ How CDP Works	2
➤ Applications That Require CDP	2
➤ Application-Intelligent CDP.....	3
➤ Thinking Outside the Box	3
➤ Conclusion	4

How CDP Works

Today, several vendors offer CDP solutions. CDP technology comes in two basic flavors – *block-based CDP* and *file-based CDP*. Block-based CDP operates at the logical volume level and records every write, while file-based CDP operates at the file system's level and records any changes to the file system. The point of CDP is to present a recoverable image of the data at the selected point, and to ensure that this can be done both approaches maintain write order fidelity (the write ordering established by the application from which the data is being collected). While each vendor's solution may differ from the others, most operate in a similar way.

Take the example of block-based CDP. Here, vendors may supply agents that are installed on the servers, or have appliances that are installed in the data path. The agents or appliances examine every write or update that is destined for a CDP-protected volume and send a copy of that write to another disk storage system, usually containing less-expensive SATA disks. A CDP recovery engine manages this second disk system and time-stamps every write. Later the CDP engine can create any image of the volume for any point in time that it was "recording", which then can be presented to any server connected to this pool of CDP disk storage.

Since CDP solutions continuously capture any updates (changes), the application never needs to be interrupted to accommodate a backup application. When data corruption occurs and the data needs to be restored, IT administrators can request an image of the data that was written seconds prior to the corruption. RTO is reduced to the time required to mount the image and restart the application. RPO is reduced to any updates recorded between the time corruption entered the system and the time of the restored image.

Applications That Require CDP

Should all applications be protected by CDP solutions? Absolutely not! In the days of tape-only backups, there was only one data protection method to protect every application. Now we have several different choices – tape-based, disk-based, and CDP, each delivering different RPO and RTO objectives.

Today, enterprises run many different types of applications. Some mission-critical applications, such as internet-based commerce applications, email or customer support applications are updated frequently and must be available all of the time. These types of applications must be able to

recover quickly when problems arise. Extended outages of mission-critical applications can cost companies thousands of dollars (or more) in lost revenue and customers. These applications benefit from CDP solutions. Other applications, such as accounts payable, payroll, and human resources, may only have to be available during prime shift or may not have to be updated frequently. Here, more traditional data protection solutions, such as tape-based or disk-based copies, can meet RTO and RPO objectives. **Simply put, each different class of application should be matched to the right level of data protection.**

CDP should not be considered a replacement for existing data protection processes, but as a complement to them. CDP provides local data protection, but does not protect against a data center outage. CDP works with existing backup processes. A CDP image may be used as the source for an existing backup process. This backup copy can then be transmitted offsite for disaster recovery purposes.

The Need for Application Awareness

Pure block-based CDP solutions protect applications without any understanding of application objects (e.g. files, records) or application processes (e.g. file delete in a file system, transactions in a database, etc.) – they just track block-level changes. Because write-order fidelity is maintained, a recoverable image can be presented at any point in time, but it is important to note the distinction between *crash consistent* and *transaction consistent* images. Let's take an example from the database world. If a database server crashes unexpectedly, transactions are in mid flight and there will be data in both the data volumes and the log volumes that must be interpreted by the database on restart to produce a reliable data state. The data state presented to the database on restart by the CDP-generated image is said to be *crash consistent*. On the other hand, if a database is shut down in an orderly manner, data from the log volumes is purged as data is written to disk and all transactions in flight are allowed to complete and are also written to disk. This is a *transaction consistent* state, and you'll note that there is no data in the log volumes that must be interpreted to complete a reliable recovery. Attempting to restart a crash tolerant application (e.g. journaling file system, database) from either type of image will result in a reliable recovery, but recovery from a crash consistent image is likely to take longer. The following example shows the shortcomings of some block-based CDP solutions that only support *crash consistent* images.

Do I Feel Lucky?

Let's assume that a database system, which is protected by block-based CDP product, is accidentally corrupted at approximately 11 in the morning. The IT administrator wants to restore the database to a point prior to the corruption. Since the database is protected continuously, there are many different time-based images from which to choose. *So, which one is the right one?* Is it the image from 10:59 AM, or 10:58 AM or 10:30 AM? The administrator chooses the image from 10:58 AM and mounts that image to the application. The recovery process is not yet complete. Since transactions span multiple blocks, it is very likely that the administrator will choose an image with partial, or incomplete, transaction data. After mounting the chosen point in time image to the application, the application must now check the database and back out any partial transactions; this takes valuable recovery time. In fact, it can take a great deal of time mounting versions from different time periods, backing out incomplete transactions only to determine that the corruption still exists in those versions. This haphazard approach to data recovery reminds me of the scene in the *Dirty Harry* movie when Clint Eastwood asks:

*You've got to ask yourself one question:
Do I feel lucky?*

Administrators should not have to ask themselves "how lucky do I feel" when recovering data. Some CDP vendors have a remedy for this problem – application aware CDP.

Application-Aware CDP

Application-aware CDP allows customers to annotate the recovery timeline (which includes all possible *crash consistent* points) with markers indicating *transaction consistent* points. For example, a DBA may perform a database checkpoint, putting the database into a *transaction consistent* state while it is still online. Administrators can now mark this point as a *transaction consistent* (usable) application image. Now, when this image is used to recover a database, there are no incomplete transactions to "back out" and recovery is much quicker. Using event markers is a common way among block-based CDP vendors to provide awareness of critical process points in an application's recovery timeline. Some CDP vendors allow administrators to implement policies to periodically generate and mark a database checkpoint.

This approach gives administrators two choices - they can recover from *very granular*,

time-based images or *less granular, but transaction consistent, images*. Creating many transaction consistent images during the day has a slight disadvantage. Checkpoints require the application to stop processing any incoming work for a few minutes in general, which can create temporarily elongated response times, particularly during periods of high database activity.

Application-Intelligent CDP

The next generation of CDP products eliminates the performance problems of issuing checkpoints while maintaining application consistency. This next generation is not only *application aware (transaction consistent)* but understands the application structure. We call this generation of CDP solutions *application-intelligent CDP*. Application-Intelligent CDP requires close integration with the application. It must understand the structure of the database and understand where transactions begin and end. It can then equate its block structure to a transaction and create numerous consistent recovery points without quiescing (stopping) the database. **Application-intelligent CDP solutions are the optimal solution for applications that demand fast recovery times.**

Thinking Outside the Box

CDP changes the way IT administrators think about backups. They are no longer forced to recover from scheduled point-in-time copies but can choose any point in time. CDP also changes the way IT administrators support business processes. **In fact, CDP lets administrators think outside the box to solve common business problems.** CDP images can be used in other ways than protecting data. For example, the image can be mounted on other servers to test new versions of applications, providing development personnel with a complete copy of the database for thorough testing. Or, the image can be used to refresh data warehouses more frequently, providing more current and timely business information. Images can also be mounted on other servers to perform compute intensive data analysis and queries without affecting the performance of online applications. Images of financial databases can be mounted to produce reports of the financial state of the company at a particular month, week, or quarter. Images of third-party software applications can be used to test new patches or maintenance updates without affecting the version running in production. All of this can be accomplished without scheduling time to create point-in-time copies and doesn't require that we determine - in advance - which point-in-time copies we may

need in the future. Different uses for CDP images are only limited by our imagination. In fact, the name Continuous Data Protection restricts us to thinking about CDP as a data protection tool. CDP is more than that. It is a technology that allows us to think differently about supporting many business processes without requiring we periodically create and maintain lots of different copies of data.

Maybe a better name would have been *Continuous Access to Data (CAD)*, or *Continuous Images of Data (CID)*, or even *All Different Views of Data (ADVD)*. Bottom line, **CDP is a great technology for protecting mission critical data. It is an even better technology for supplying different views of data to any business process that needs accurate and timely data.**

Conclusion

There are numerous CDP products available today from both small and large vendors. Some are integrated with specific applications, such as Microsoft *Exchange*, that allow IT organizations to recover quickly when a virus infects the email system. Others are more general solutions that can protect many different applications. **IT organizations that are struggling to meet stringent recovery time objectives should consider CDP products.**

CDP solutions are ideal for large enterprises with mission-critical applications. Smaller enterprises can also benefit from specific solutions, such as those that protect Microsoft *Exchange*, which allows them to recover faster when email systems become corrupted.

CDP provides a new tier of data protection but should not be viewed as a replacement for existing backup or disaster recovery processes. All applications must still be backed up, and copies sent off site for disaster recovery. CDP changes the way IT administrators support backup, recovery and business processes. Don't let the name CDP lull you into thinking this is just a new way to protect data. It is a great new technology to protect mission-critical data, but it is much more than that. It gives us flexibility to support business applications with images of data created from any point in time, when we need it, as often as we need it.

Is CDP ready for prime time? Absolutely! It is also ready for prime-time thinking.



About The Clipper Group, Inc.

The Clipper Group, Inc., is an independent consulting firm specializing in acquisition decisions and strategic advice regarding complex, enterprise-class information technologies. Our team of industry professionals averages more than 25 years of real-world experience. A team of staff consultants augments our capabilities, with significant experience across a broad spectrum of applications and environments.

- ***The Clipper Group can be reached at 781-235-0085 and found on the web at www.clipper.com.***

About the Author

Dianne McAdam is Director of Enterprise Information Assurance for the Clipper Group. She brings over three decades of experience as a data center director, educator, technical programmer, systems engineer, and manager for industry-leading vendors. Dianne has held the position of senior analyst at Data Mobility Group and at Illuminata. Before that, she was a technical presentation specialist at EMC's Executive Briefing Center. At Hitachi Data Systems, she served as performance and capacity planning systems engineer and as a systems engineering manager. She also worked at StorageTek as a virtual tape and disk specialist; at Sun Microsystems, as an enterprise storage specialist; and at several large corporations as technical services directors. Dianne earned a Bachelor's and Master's degree in mathematics from Hofstra University in New York.

- ***Reach Dianne McAdam via e-mail at dianne.mcadam@clipper.com or at 781-235-0085 Ext. 212. (Please dial "212" when you hear the automated attendant.)***

Regarding Trademarks and Service Marks

The Clipper Group Navigator, The Clipper Group Explorer, The Clipper Group Observer, The Clipper Group Captain's Log, The Clipper Group Voyager, and "*clipper.com*" are trademarks of The Clipper Group, Inc., and the clipper ship drawings, "*Navigating Information Technology Horizons*", and "*teraproductivity*" are service marks of The Clipper Group, Inc. The Clipper Group, Inc., reserves all rights regarding its trademarks and service marks. All other trademarks, etc., belong to their respective owners.

Disclosure

Officers and/or employees of The Clipper Group may own as individuals, directly or indirectly, shares in one or more companies discussed in this bulletin. Company policy prohibits any officer or employee from holding more than one percent of the outstanding shares of any company covered by The Clipper Group. The Clipper Group, Inc., has no such equity holdings.

Regarding the Information in this Issue

The Clipper Group believes the information included in this report to be accurate. Data has been received from a variety of sources, which we believe to be reliable, including manufacturers, distributors, or users of the products discussed herein. The Clipper Group, Inc., cannot be held responsible for any consequential damages resulting from the application of information or opinions contained in this report.