



Closing the Backup Gap

Continuous data protection gives organizations peace of mind through granular backup and reduced recovery time.



What Is CDP?

Continuous data protection is a storage system in which all organizational data is backed up whenever any change is made. In effect, CDP creates an electronic log of complete storage snapshots — one storage snapshot for each incident of data modification.

Disasters are indiscriminate. Sometimes they can be anticipated, as was Hurricane Katrina's slow churn over the Gulf Coast. But most frequently they are unexpected and much less dramatic, such as when a disk drive fails in a computer or file server. In both cases, the normal response to the disaster is to rebuild the lost systems using a backup made prior to the disaster. But the real question is "How much information was lost between the time of the last backup and the disaster?" If your organization has implemented continuous data protection (CDP), the answer could be "None."

CDP is growing in popularity because it closes the gap between backup cycles by continuously copying changes made to data files to a second storage device. If this sounds familiar, it's because the concept of reproducing data to guard against system failure has been around for years. PC servers touted mirrored disk drives in which everything written to the primary drive also was written to a secondary drive. And current replication technologies copy data from one system to another, possibly in a different location.

Both of these strategies have their place in protecting data from loss, but most implementations of this technology suffer from the same defect: Any corruption introduced to the primary storage location is efficiently duplicated to the secondary location. The result is that the data can be restored if the primary storage fails, but it will restore the same corruption that existed on the primary storage.

For this reason, replication systems have their place in the overall scheme of disaster recovery, but organizations that rely on immediate and continuous access to their data are finding ways to leverage combinations of traditional backups, replication and CDP.

RPO and RTO

Enterprise service level agreements (SLA) typically include both recovery point objectives (RPO) and recovery time objectives (RTO). The RPO specifies the maximum difference measured in time between any failure and the previous backup. In essence, it is a measure of the amount of information that is likely to be lost between the last backup and any failure. The RTO specifies the maximum length of time that will be required to complete a restoration, bringing the system to its usable state, which is the most recent backup.

For a growing number of organizations, any RPO greater than zero and any RTO longer than immediate are unacceptable. CDP holds the promise of being able to meet these tighter requirements because changes are recorded at fine levels of granularity. Recovery times can be reduced because smaller quantities of data need to be restored in order to return the data back to its desired state.

Dave Russell, vice president of storage technologies and strategies for research firm Gartner, in Tucson, Ariz., says, "The market for CDP is growing, but we are finding that organizations are taking advantage of the concept primarily with regard to their Microsoft Office-type files and Microsoft Exchange databases." The reason for the initial acceptance of CDP for these types of data stems from the fact that they have not typically been the center of attention in the same way that the enterprise databases have been. Russell explains, "These Windows data-types have been either unprotected or underprotected. E-mail, and Exchange in particular, is no longer a tier-2 application. It has risen to the critical level in nearly all organizations." ▶

Most Exchange servers have been managed outside the realm of the database administration where enterprise application databases receive intense scrutiny and constant attention. This has left an opening for server administrators to find solutions to protect their Exchange databases. The same holds true for Office-type documents. As Gartner's Russell says, "E-mail administrators are crying for help. Exchange is entirely separate from the database administrators."

State of Change

The issue of recovering lost or damaged data is always a matter of "when" and "how much." Data loss always will happen because of software malfunction, hardware failure or user error. While traditional backups are still largely stored on tape, which remains the predominant backup medium, studies show that failure rates for recovery from tape range from 17 percent to as much as 40 percent. This level of uncertainty leads administrators tasked with delivering immediate recovery of data with few options other than to increase the frequency of their backups.

The size of data stores, such as those for Microsoft Exchange, continue to expand, meaning that the time required to perform a standard backup increases. Shrinking backup windows become more troublesome when the backups have to be used to restore their contents.

A study performed by Enterprise Strategy Group showed that the No. 1 concern among administrators was the ability to manage backup and restore functions for Microsoft Exchange Server and Microsoft SQL Server. Michael Parker, global product marketing manager at Symantec, says that managing Exchange backups is one of the main focuses of Symantec's latest release, Backup Exec 11d.

"This release provides granular backup and recovery, making it possible to create a restore point that exactly matches what needs to be restored," Parker says. The alternative without CDP is to restore the previous backup, or if daily mailbox backups are being made, to restore just the affected mailbox.

Performing standard restoration of even a single mailbox still presents problems because both more and less information is restored. More is restored because most restorations are done to retrieve only a single accidentally deleted e-mail

message. At the same time, less is restored because messages received after the last backup are not included in the retrieved mailbox contents.

CDP can significantly alter the landscape when it comes to meeting an organization's RTO and RPO, according to Parker. The restoration process is nearly instantaneous for two reasons. "Since the system is making incremental copies of each change, the segments of changed data are very small," he says. "Secondly, backups are made to disk drives, so the medium is immediately accessible and data transfer speed is fast. This combination eliminates the backup window and makes RTO and RPO negligible at the same time."

EMC's RecoverPoint CDP is an appliance-based application that can be used in the data center and can provide CDP for a heterogeneous storage environment. Rick Walsworth, director of product marketing for EMC, points to the trend in providing advanced data protection.

"The traditional backup with daily, full and incremental backups is no longer sufficient for transactional systems in the government and educational environments. For most applications that are transaction based, a four-hour level of granularity is unacceptable," he says.

Walsworth points to CDP's finer granularity, giving it the ability to go back through the data to a point in time before the particular corruption was present. "This minimizes the amount of data lost and is the main driver for the technology from the end users' perspective," he says.

Expanding Markets

The need for faster response and less intrusive recoveries is not specific to any sector. However, the market for dedicated CDP technologies is concentrated around a specific set of applications. In particular, file-based applications such as word processing and spreadsheets, Microsoft SQL Server, and Microsoft Exchange are widely used and generally maintained outside the data center.

It isn't that applications that use databases such as Oracle and SAP are immune to data loss, but as Gartner's Russell explains, "Database providers like Oracle have already done a lot with their transaction logs, and recovery is interwoven in their products."



“We are seeing increased interest in a variety of market sectors,” explains EMC’s Walsworth. “The education market, particularly in research centers, is moving to protect their valuable data housed in SQL Server databases.”

At the same time, government agencies needing to meet regulatory requirements regarding availability and retention of e-mail are looking at CDP solutions. As Walsworth puts it, “For many agencies, the acronym ROI has taken on an entirely new meaning. Where it used to mean ‘return on investment,’ it now can be read as ‘risk of incarceration.’” The inability to provide required access to critical documents and data has risen to a top organizational concern.

For its part, Microsoft’s System Center Data Protection Manager (DPM) Version 2 is available as a beta release. The product adds protection for SQL Server, Exchange Server and Microsoft SharePoint, in addition to its existing CDP function for files. It was released in late 2006 and addresses the same vulnerabilities that both Symantec and EMC protect.

Combined Functions

CDP is important as an independent function because it provides reliable mechanisms for assuring access to data even after disastrous occurrences. But the integration of CDP with traditional backup facilities is already part of the plan, if not reality, for many implementations.

Because CDP copies changes to data at the moment it occurs, it requires high-speed systems to transfer and store the changes. This level of access is made possible by copying from the data’s primary disk system to secondary disk drives — called “disk-to-disk copying.” Depending on the size of the source data stored and the frequency of change, it is conceivable that the secondary storage will continue to grow without limit. In practice, various methods are provided by different manufacturers to limit the size of the CDP storage space and simplify the recovery of data.

The trend is to combine the disk-to-disk copies created by CDP with tape backup facilities, creating a disk-to-disk-to-tape migration sequence. Symantec’s Backup Exec 11d includes CDP as a configurable addition to its traditional backup functions, providing a single interface to manage the full set of services.

Part of the process of progressively moving backups to tape is

the rolling up of CDP incremental changes into larger segments of data. Symantec’s Parker explains, “If you’re capturing every change in an Exchange Server database, you can end up using a tremendous amount of storage space. Our product is configurable so that administrators can define their own roll-up periods for Exchange databases.”

Consolidating granular changes into larger batches has the effect of reducing the size of transaction logs, which can mean a reduction in the processing overhead on the server. Microsoft’s DPM product also provides integrated CDP with tape backup.

Storage Systems Will Grow

In order to accommodate CDP, organizations will need to add disk capacity to their already expanding storage. Parker recommends that organizations plan on adding capacity in the ratio of 1:1.5 or 1:2 primary storage to CDP storage.

“Fortunately, disks are cheap, and we expect to see the majority of CDP storage in NAS appliances,” he says. NAS storage is particularly well suited to housing CDP because it is relatively less expensive than SAN storage, and the relatively faster access provided by SANs is not required for CDP.

Slow to Adopt

According to Gartner’s Russell, organizations are still slow to adopt CDP. “As more large manufacturers get more established and build more trust in CDP, we will see more adoptions in the database arena on Unix/Linux environments,” he says. “But we still expect it to be two to three years before there is significant movement. There is very little tolerance for risk in those shops with existing backup and recovery schemes.”

Russell points out that while there continues to be talk about disk-to-disk backup strategies, tape continues to be the predominant backup for many medium or large organizations. These organizations understand their existing risks and contingencies and have stable environments even if there may be more promising options by way of CDP implementations.

“The value proposition is definitely there and is dramatic,” Russell says. “The issue is integrating CDP with existing backup technology and hitting the right purchasing cycle. What we have is a smoldering fire that isn’t yet ignited.” ■



Did you know that CDW•G offers configuration, product support and customized professional services? Call your account manager for details.