READER/EXPERT I/O

What *is* CDP?

Q: How is CDP (continuous data protection) different from traditional host-based replication software?

■ BY JACOB FARMER

CDP is the latest three-letter acronym to bring mass confusion to the mass storage marketplace. "CDP" stands for continuous data protection, which presumably refers to data-protection technology that safeguards data on an ongoing basis, as opposed to running batch-based backup jobs once a day.

No *one* organization owns the definition, so it's up to the marketing departments at the various vendors to haggle over what constitutes *true* CDP. So, depending on whom you ask, you might hear that replication and CDP are one in the same. You might also hear that replication is merely a component of CDP, or that replication and CDP are distinct technologies. In other words, everyone's got his own definition, and in the meaningless world of three-letter storage technology acronyms, everyone is right.

However, since you asked, I'll take this opportunity to put forward my own definition and then spend the next few weeks answering hate mail from every vendor pushing CDP products.

By my definition, CDP is a hybrid of replication, snapshot, and backup/restore technologies. Some products replicate data and then take snapshots of the replica to create point-in-time representations of the data.

CONTINUED ON PAGE 10

Pillar adds management software

■ BY KEVIN KOMIEGA

Pillar Data Systems is adding a new set of management features to its software suite to provide better management for its "all-in-one" SAN and NAS tiered storage systems, including support for virtual tape, a new continuous data protection (CDP) option, and secure write-once, read-many (WORM) technology. The company, which is backed by Oracle CEO Larry Ellison's investment firm Tako Ventures, is also preparing to launch a lower-cost, midrange version of its Axiom hardware product.

Pillar calls the new software AxiomONE and bills the bundle of tools as a full-service storage management and data services suite. Pillar's senior director of marketing and strategy, Russ Kennedy, says AxiomONE brings together SAN and NAS management for data protection and archiving, guided maintenance and support, and a full-featured operating system under a single management interface.

"The software suite reduces the number of [systems] you have to manage and enables users to deliver tiers of storage in a single, unified platform," says Kennedy.

The software optimizes application performance via its Dynamic Performance Manager, a policy-based information management engine that enables multiple performance priority levels over both Serial ATA (SATA) and Fibre Channel disk drives.

The core of AxiomONE comprises an operating system, file system, storage services manager, and dynamic performance manager at a starting list price of approximately $40,000, while the newer modules such as AxiomONE File Replicator, Volume Replicator, Continuous Data Protection, and Virtual Tape Library are available at an additional fee.

"All of the Pillar software comes in the software suite for a one-time license fee regardless of the capacity of the system, while the volume replication, CDP, and VTL modules are licensed on a capacity basis," says Kennedy.

Kennedy says the AxiomONE software will also be shipped with Pillar's upcoming hardware platform, called the Axiom 300, for small and medium-sized businesses (SMBs). The Axiom 300 will share the same architecture as Pillar's larger disk systems. All versions of the Axiom systems accommodate multiple tiers of block- and/or file-level storage in one box.

Data storage on the Axiom platform is laid out by its performance priority. High-priority data is written to a narrow band on the outermost tracks of the hard drives for increased performance and optimized seek times, while data with lower-performance requirements is laid out on the inner tracks.

Pillar's modular system also allows for the configuration of a NAS-only system, SAN-only system, or a combination of SAN and NAS on one platform. Kennedy estimates that 50% of Pillar's end users are running SAN-only configurations, 20% are NAS-only, and 30% are running both SAN and NAS on their systems.

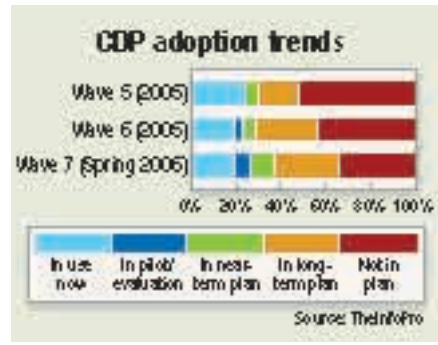
The Axiom storage system is priced from about $50,000. ❑

IBM brings CDP to SOHOs

■ BY KEVIN KOMIEGA

Continuous data protection (CDP) technology was arguably the most hyped storage software innovation of 2005, but once reality set in, CDP quickly went from being a must-have product to something enterprise users expect to be offered as a standard feature of any backup/recovery product.

In response, IBM is taking a new approach to selling CDP software. The company recently announced that it has signed a multi-year deal with e-commerce outsourcing firm Digital River to help Big Blue sell its IBM Tivoli Continuous Data Protection for Files software to PC users and small businesses via the Internet.



According to the deal, Digital River is making the CDP for Files software available for online purchase and digital download through the Digital River oneNetwork marketplace, which includes online retailers OfficeMax, Staples, and Circuit City, among others.

IBM's CDP for Files software will also be for sale via download from IBM's Website and through other sales channels at a cost of $35 per laptop or desktop PC.

So why bring a seemingly advanced backup product to the average remote worker and consumer users? For one, IBM CTO Chris Stakutis has doubts about the potential market for selling CDP—at least block-level CDP—in large enterprises.

"I have my reservations about CDP at the block level. There is a large amount of CDP-like technology already built into databases and other applications," says Stakutis. He claims that the data-pro-

CONTINUED ON PAGE 19

HP ships hybrid array

■ BY KEVIN KOMIEGA

Hewlett-Packard was expected to redouble its efforts in the small to medium-sized business (SMB) storage market this month with shipments of a new line of hybrid arrays that promise block-based (SAN) and file-based (NAS) storage with integrated data protection in one box.

The new family of StorageWorks hardware, which had not been named at press time, will be based on Microsoft's Windows Storage Server 2003 R2 operating system and HP's ProLiant hardware with a new set of software tools for "those who don't understand storage," according to Debbie Young, worldwide marketing manager for HP's StorageWorks division.

Young claims the new products will provide integrated data protection features and redundancy with disk-based snapshot technology. A complete hardware and software solution will cost approximately $5,000 for a 1TB configuration.

HP estimates that up to 70% of SMBs still have not moved to networked storage. "Today's SMB products are great for SMBs that already understand storage," says Young, "[but] many small companies are staying with direct-attached storage because they understand it. You almost need an MBA in storage to make the simple move to networked storage today."

The announcement of the new product family raises questions about possible cannibalization of HP's current SMB storage products, most notably the iSCSI-based StorageWorks 1510i Modular Smart Array (MSA1510i). However, HP officials say that the new systems will complement the existing MSA arrays.

Aside from possibly competing with its own portfolio, HP is competing with other major vendors for SMB storage dollars. For example, EMC and Intel announced a partnership earlier this year under which Intel is selling a line of entry-level storage arrays based on EMC's Clariion AX150 disk arrays (see "Intel to sell entry-level EMC arrays," *InfoStor*, May 2006, p. 1).

And more recently, Network Appliance made its move into SMB storage with the launch of the StoreVault S500, which supports NAS, iSCSI SAN, and Fibre Channel SAN connectivity (see "NetApp targets SMBs with hybrid array," *InfoStor*, August 2006, p. 1). The goal is to combine iSCSI, SAN, and NAS into a single storage pool, and that can be managed from a Windows-based user interface.

NetApp's S500 scales to 6TB and can be configured with either 250GB or 500GB Serial ATA (SATA) hard drives. The array supports Microsoft's iSCSI Software Initiator for IP SAN connectivity. The system can also use Silverback Systems' iSNAP2110 iSCSI host bus adapter (HBA) initiator cards to boost TCP/IP speeds in iSCSI SAN environments.

NetApp's S500 is priced at $5,000 for a 1TB configuration, similar to HP's hybrid array. ❑

VENDORS MENTIONED
EMC, Hewlett-Packard, Intel, Microsoft, Network Appliance, Silverback Systems

READER/EXPERT I/O from page 8

Others take snapshots first and then replicate. Take your pick. There are subtle advantages to each approach. The key is that the product offers a restore interface, much like that of traditional backup software.

The best way to understand my definition of CDP is to look at the shortcomings of traditional backup software, traditional replication, and traditional snapshots. Backup software is cumbersome: It results in unchanged data being backed up over and over again. A file that has not changed in five years might have been backed up 260 times (52 weeks times five years). Meanwhile, the backup system administrator complains about missing backup windows.

Replication, by contrast, moves data to a separate storage system as data changes. There are dozens of replication schemes on the market. Some are synchronous; others are asynchronous. Some work continually as files or blocks are updated; others send deltas in small batches, maybe every hour or on some other practical interval.

There are two main reasons why replication has not replaced traditional backup. First, a corruption or deletion of the primary storage gets propagated to the replica. Second, replication software does not offer granular restore capabilities such as that offered by backup software.

There are dozens of ways to do snapshots, but they all more or less give you logical representations of your data at different points in time. Snapshots have not replaced backup software for three main reasons: First, snapshots do not protect you from a catastrophic hardware failure. If the disk array crashes, the snapshots as well as the primary representation of the data are gone. Second, snapshots lack a user-friendly restore mechanism. Many snapshot implementations require you to mount an entire volume on a dedicated server to pick off a select file or object. Some snapshot systems put files in hidden directories off the original directory, but this is still a bit cumbersome and requires a sophisticated user to retrieve the files. Third, few snapshot systems offer any meaningful retention policies. Usually, you are limited to a fixed number of snapshots and you cannot save them forever—at least not without other negative consequences.

Now imagine software that is smart enough to back up only block-level or transaction-level changes, much the same way that replication software does. In other words, you do a full backup once and thereafter you only back up the changes. You then combine the power of snapshot technology to save representations of your primary data at different points in time without consuming insane amounts of disk. Finally, you add a familiar point, click, drag, drop restore interface so that you can restore whatever you need from whatever point in time. As an added bonus, you might get advanced features such as mailbox- and message-level restore, bare-metal restore, remote site replication, etc. This is my definition of CDP.

I am in the minority in that I do not believe that data changes need to be recorded continuously, down to the individual transaction. Similarly, I do not believe that most people want or need to restore from any point in time down to the nanosecond. Rather, I believe that *continuous* simply means "more often than once per night."

Moreover, I believe that the real value of CDP technology is not that it is continuous; the advantage is that it is super-efficient. CDP eliminates backup windows. It removes the system overhead of batch-based backups. It economizes on secondary disk storage by only storing deltas rather than a series of full backups and incremental backups. As an added bonus it allows more-granular recovery points.

Jacob Farmer
CAMBRIDGE COMPUTER

So, to answer your question: Conventional replication software is a component in CDP, but I think the marketing folks at the replication vendors who have jumped on the CDP bandwagon are taking too much poetic license in labeling their products as CDP.

In fact, it's too bad the industry chose the acronym "CDP" to describe the next generation of backup/restore technology. In so doing, it failed to call attention to the shortcomings of traditional backup software and the true path to salvation. Meanwhile, the acronym invited all kinds of vendors to redefine their existing products as CDP, causing further confusion and sending the message to consumers that "CDP" is just a new term for old technology. Follow my definition. You will find that real solutions exist on the market, and they do work wonders! ❑

Jacob Farmer is chief technology officer at Cambridge Computer (www.cambridgecomputer.com) in Waltham, MA. He can be contacted at jacobf@cambridgecomputer.com.