WIKIPEDIA

Continuous Data Protection

Continuous data protection (**CDP**), also called **continuous backup** or **real-time backup**, refers to backup of computer data by automatically saving a copy of every change made to that data, essentially capturing every version of the data that the user saves. In its true form it allows the user or administrator to restore data to any point in time.[1] The technique was patented by British entrepreneur Pete Malcolm in 1989 as "a backup system in which a *copy* [editor's emphasis] of every change made to a storage medium *is recorded as the change occurs* [editor's emphasis]."[2]

In an *ideal* case of *continuous data protection,* the recovery point objective—"the maximum targeted period in which data (transactions) might be lost from an IT service due to a major incident"—is zero, even though the recovery time objective—"the targeted duration of time and a service level within which a business process must be restored after a disaster (or disruption) in order to avoid unacceptable consequences associated with a break in business continuity"—is not zero.[3] An example of a period in which data transactions *might* be lost is a major discount chain having card readers at its checkout counters shut down at multiple locations for close to two hours in the month of June 2019.

CDP runs as a service that captures changes to data to a separate storage location. There are multiple methods for capturing continuous live data changes involving different technologies that serve different needs. *True* CDP-based solutions can provide fine granularities of restorable objects ranging from crash-consistent images to logical objects such as files, mail boxes, messages, and database files and logs.[4] This isn't necessarily true of *near*-CDP solutions.

Contents

Differences from traditional backup

True continuous data protection is different from traditional backup in that it is not necessary to specify the point in time to recover from until ready to restore.[5] Traditional backups only restore data from the time the backup was made. *True* continuous data protection, in contrast to "snapshots", has *no* backup schedules.[5] When data is written to disk, it is also asynchronously written to a second location, either another computer over the network[6] or an appliance.[7] This introduces some overhead to disk-write operations but eliminates the need for scheduled backups.

Allowing restoring data to any point in time, "CDP is the gold standard—the most comprehensive and advanced data protection. But 'near CDP' technologies can deliver enough protection for many companies with less complexity and cost. For example, snapshots ["near-CDP" clarification in the section below] can provide a reasonable near-CDP-level of protection for file shares, letting users directly access data on the file share at regular intervals—say, every half hour or 15 minutes. That's certainly a higher level of protection than tape-based or disk-based nightly backups and may be all you need."[1] Because "near-CDP does this [copying] at pre-set time intervals",[8] it is essentially incremental backup initiated—separately for each source machine—by timer instead of script.

Continuous vs near continuous

Since *true* CDP "backup write operations are executed at the level of the basic input/output system (BIOS) of the microcomputer in such a manner that normal use of the computer is unaffected",[2] *true* CDP backup must in practice be run in conjunction with a virtual machine[6][9] or equivalent[10]—ruling it out for ordinary *personal* backup applications. It is therefore discussed in the "Enterprise client-server backup" article, rather than in the "Backup" article.

Some solutions *marketed* as continuous data protection may only allow restores at fixed intervals such as 15 minutes or one hour or 24 hours, because they automatically take incremental backups at those intervals. Such "near-CDP"—short for **near-continuous data protection**—schemes are not universally recognized as true continuous data protection, as they do not provide the ability to restore to any point in time. When the interval is shorter than one hour,[11] "near-CDP" solutions—for example Arq Backup[12]—are typically based on periodic "snapshots"; "to avoid downtime, high-availability systems may instead perform the backup on ... a read-only copy of the data set frozen at a point in time—and allow applications to continue writing to their data".

There is debate in the industry as to whether the granularity of backup must be "every write" to be CDP, or whether a "near-CDP" solution that captures the data every few minutes 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* must be continuously *automatically scheduled*, which is often sufficient to achieve the benefits cited above, or whether the ability to *restore* from the backup also must be continuous. The Storage Networking Industry Association (SNIA) uses the "every write" definition.[5]

There is a briefer sub-sub-section in the "Backup" article about this, now renamed to "Near-CDP" to avoid confusion.

Differences from RAID, replication or mirroring

Continuous data protection differs from RAID, replication, or mirroring in that these technologies only protect one copy of the data (the most recent). If data becomes corrupted in a way that is not immediately detected, these technologies simply protect the corrupted data with no way to restore an uncorrupted version.[13]

Continuous data protection protects against some effects of data corruption by allowing restoration of a previous, uncorrupted version of the data. Transactions that took place between the corrupting event and the restoration are lost, however. They could be recovered through other means, such as journaling.

Backup disk size

In some situations, continuous data protection requires less space on backup media (usually disk) than traditional backup. Most continuous data protection solutions save *byte or block-level* differences rather than *file-level differences*. This means that if one byte of a 100 GB file is modified, only the changed byte or block is backed up. Traditional incremental and differential backups make copies of entire files; however starting around 2013 enterprise client-server backup applications have implemented a capability for block-level *incremental* backup, designed for large files such as databases.

Risks and disadvantages

When real-time edits—especially in multimedia and CAD design environments—are backed up offsite over the upstream channel of the installation's broadband network,[14] network bandwidth throttling[15] may be needed to reduce the impact of *true* CDP.[14] An alternative approach is to back up to a separate Fibre-Channel-connected SAN appliance.[7]

See also

- Quest AppAssure
- Cofio Software
- Disaster recovery
- EMC RecoverPoint
- FalconStor
- InMage DR-Scout
- List of backup software
- List of online backup services
- Single instance storage
- CloudEndure

References

1. Behzad Behtash (2010-05-06). "Why Continuous Data Protection's Getting More Practical" (https://www.informationweek.com/why-continuous-data-protections-getting-more-practical/d/d-id/1088883). *Disaster recovery/business continuity*. InformationWeek. Retrieved 2011-11-12. "A true CDP approach should capture all data writes, thus continuously backing up data and eliminating backup windows.... CDP is the gold standard—the most comprehensive and advanced data protection. But "near CDP" technologies can deliver enough protection for many companies with less complexity and cost. For example, snapshots can provide a reasonable near-CDP-level of protection for file shares, letting users directly access data on the file share at regular intervals--say, every half hour or 15 minutes. That's certainly a higher level of protection than tape-based or disk-based nightly backups and may be all you need."
2. Peter B. Malcolm (13 November 1989). "US Patent 5086502: Method of operating a data processing system" (http://www.google.com/patents/US5086502). Google Patents. Retrieved 29 November 2016. "Filing date Nov 13, 1989 ... a backup system in which a copy of every change made to a storage medium is recorded as the change occurs ... backup write operations are executed at the level of the basic input/output system (BIOS) ..."
3. Richard May (November 2012). "Finding RPO and RTO" (https://web.archive.org/web/20160303224604/http://www.virtualdcs.co.uk/blog/business-continuity-planning-rpo-and-rto.html). Archived from the original (http://www.virtualdcs.co.uk/blog/business-continuity-planning-rpo-and-rto.html) on 2016-03-03.

4. Pat Hanavan (2007). "An Overview of Continuous Data Protection" (https://web.archive.org/web/20190617161626/http://www.infosectoday.com/Articles/Continuous_Data_Protection.htm). Infosectoday.com. What is Continuous Data Protection?, Can CDP be leveraged for backing up and recovering email?. Archived from the original (http://www.infosectoday.com/Articles/Continuous_Data_Protection.htm) on 2019-06-17. Retrieved 2011-11-12. "... may be block, file-, or application-based and can provide fine granularities of restorable objects to infinitely variable points in time.... New granular recovery technologies have emerged that enable mail messages, mailboxes, and folders to be restored individually without having to restore an entire email database, and without separate and redundant mailbox backups."
5. "Data Protection Best Practices" (https://www.snia.org/sites/default/files/DPCO/Data%20Protection%20BP%20White%20Paper%20Final%20v1_0.pdf) (PDF). *SNIA*. Storage Networking Industry Association. 23 October 2017. 2.1.4 Continuous Data Protection (CDP). Retrieved 27 June 2019. "...pros to the use of snapshots:[new paragraph]Allows for the recovery of files from a specific point in time (based on snapshot schedule)....CDP can provide the ability to restore to any previous point in time, since the backups are taking place near-instantaneously; therefore, the potential for data loss is very small."
6. Wu, Victor (4 March 2017). "EMC RecoverPoint for Virtual Machine Overview" (https://wuchikin.wordpress.com/2017/03/04/emc-recoverpoint-for-virtual-machine-overview/). *Victor Virtual*. WuChiKin. Retrieved 22 June 2019. "The splitter splits out the Write IOs to the VMDK/RDM of a VM and sends a copy to the production VMDK and also to the RecoverPoint for VMs cluster."
7. Wendt, Jerome M. (21 September 2009). "Symantec Brings RealTime CDP into NetBackup Data Management Fold" (https://www.dcig.com/2009/09/symantec-realtime-cdp-netbackup.html). *DCIG*. DCIG LLC. Retrieved 5 August 2019. "NetBackup RealTime is an appliance-based CDP solution intended for the protection of multiple hosts. Residing in corporate FC-SANs as a side-band appliance, it sits outside of the data path between application servers and their assigned storage to eliminate any possibilities of application disruption."
8. "Continuous data protection (CDP) explained: True CDP vs near-CDP" (https://www.computerweekly.com/Continuous-data-protection-CDP-explained-True-CDP-vs-near-CDP). *ComputerWeekly.com*. TechTarget. July 2010. Retrieved 22 June 2019. "... copies data from a source to a target. True CDP does this every time a change is made, while so-called near-CDP does this at pre-set time intervals. Near-CDP is effectively the same as snapshotting....True CDP systems record every write and copy them to the target where all changes are stored in a log. [new paragraph] By contrast, near-CDP/snapshot systems copy files in a straightforward manner but require applications to be quiesced and made ready for backup, either via the application's backup mode or using, for example, Microsoft's Volume Shadow Copy Services (VSS)."
9. "Zerto or Veeam?" (https://resqdr.com/zerto-or-veeam/). *RES-Q Services*. March 2017. Retrieved 7 July 2019. "Zerto doesn’t use snapshot technology like Veeam. Instead, Zerto deploys small virtual machines on its physical hosts. These Zerto VMs capture the data as it is written to the host and then send a copy of that data to the replication site.....However, Veeam has the advantage of being able to more efficiently capture and store data for long-term retention needs. There is also a significant pricing difference, with Veeam being cheaper than Zerto."
10. "Agent Related" (https://docs.cloudendure.com/Content/FAQ/FAQ/Agent_Related.htm). *CloudEndure.com*. 2019. What does the CloudEndure Agent do?. Retrieved 3 July 2019. "The CloudEndure Agent performs an initial block-level read of the content of any volume attached to the server and replicates it to the Replication Server. The Agent then acts as an OS-level read filter to capture writes and synchronizes any block level modifications to the CloudEndure Replication Server, ensuring near-zero RPO."

11. Pond, James (25 May 2013). "FAQ 13. How are [Time Machine] backups scheduled (and can I change that)?" (https://www.baligu.com/pondini/TM/13.html). *Apple OSX and Time Machine Tips*. Baligu.com (as mirrored after James Pond died in 2013). Retrieved 4 July 2019. "Time Machine was designed and optimized to do backups hourly.... You cannot change the schedule within Time Machine. You must use a 3rd-party app, or manually alter some system files."
12. Reitshamer, Stefan (5 July 2017). "Troubleshooting backing up open/locked files on Windows" (https://www.arqbackup.com/blog/troubleshooting-backing-up-openlocked-files-on-windows/). *Arq Blog*. Haystack Software LLC. Retrieved 25 June 2019. "Arq uses Windows Volume Shadow Copy Service (VSS) to back up files that are open/locked. [Reitshamer is the principal developer of Arq Backup]"
13. Mayer, Alex (6 November 2017). "Backup Types Explained: Full, Incremental, Differential, Synthetic, and Forever-Incremental" (https://www.nakivo.com/blog/backup-types-explained-full-incremental-differential-synthetic-and-forever-incremental/). *Nakivo Blog*. Nakivo. Full Backup, Incremental Backup, Differential Backup, Mirror Backup, Reverse Incremental Backup, Continuous Data Protection (CDP), Synthetic Full Backup, Forever-Incremental Backup. Retrieved 17 May 2019.
14. Carter, Nick (5 August 2010). "Off-Site Backup – The Bandwidth Hog" (https://web.archive.org/web/20110707080023/http://www.accel-networks.com/blog/index.php?q=%2F2010%2F08%2Foff-site-backup-bandwidth-hog.html). Accel Networks. Archived from the original (http://www.accel-networks.com/blog/index.php?q=/2010/08/off-site-backup-bandwidth-hog.html) on 2011-07-07. "In a true CDP environment, whenever large files are saved – images, audio, video, CAD or 3D models – the data is transmitted over the same broadband connection that feeds users' email and internet, not to mention back-end business-critical processes. Moreover, these transmissions rely on the scarcer of the two channels, the upstream channel. The result for many companies is an erratic broadband performance, and even server slow-down."
15. David Pogue (4 January 2007). "Fewer Excuses for Not Doing a PC Backup" (https://www.nytimes.com/2007/01/04/technology/04pogue.html). *The New York Times*. "options like "Enable Bandwidth Throttle" and "Don't back up if the CPU is over this % busy.""

Retrieved from "https://en.wikipedia.org/w/index.php?title=Continuous_Data_Protection&oldid=1038815187"

This page was last edited on 14 August 2021, at 22:42 (UTC).