near-Continuous V.S. True CDP

near-Continuous V.S. True CDP

There is a big debate between the advantages of "True" CDP vs. near-Continuous. R1Soft's Continuous Data Protection products are technically near-continuous. A good definition is available at http://en.wikipedia.org/wiki/Near_Continuous_Backup

True CDP is a variation of replication technology and to be a True Continuous Backup, an application must provide at least one-second granularity between recovery points. This is why this True CDP Technology is only available to Storage Area Networks and also has limited application particularly to streaming type applications.

The reason True CDP is not useful for most servers is that most servers tend to be transaction or file oriented. For example a user edits a file or sends an email. Or a person purchases something online which involves one or more database transactions. All of these end up as File System I/O.

The file system maintains a cache of changes made to files in memory. This is commonly called the System Cache in Windows and buffers/cache in Linux. More accurately this is a Page Cache. This in-memory representation of files that is found on all modern operating systems like Windows and Linux is only flushed to disk periodically and this technique is called lazy write. A great explanation can be found here: http://en.wikipedia.org/wiki/Page_cache and a highly technical explanation is available here.

The important thing to understand is that a critical part of server performance is that writes to disk don't actually get written to the disk platters except periodically and this period is determined by the operating system. True CDP applications do not see changes written to the Storage Area Network until the operating system flushes its Disk Cache. This means the view of the raw SAN disk is in the average case corrupt and out of date compared to the actual state of the system.

The only way to release these changes form memory to disk in order to get a consistent state of data on disk is to Flush dirty or changed memory pages to disk. This can be done manually in Linux using the "sync" utility. There is a utility written by Mark Russinovich for Windows with the same name available here as part of the MS SysInternals tool kit.

If you have dealt with performance tuning on a server you probably have experienced that frequently flushing the Disk cache can seriously degrade performance. So a flush done every second on the order of so called "True CDP" is completely impractical.

Near-Continuous backup applications like R1Soft in contrast to True CDP have user scheduled

synchronizations. Realistically these can only be performed as frequently as every 15 minutes as that is about as often as you can safely flush the disk cache without losing any performance.

	True CDP	near-Continuous (R1Soft)
Recovery Point Granularity	1 Second or less playback	User scheduled e.g. every 15 minutes
Requires SAN to Work	Yes must have SAN	No any block device, real or virtual
File System Consistency	No	Yes
Application Consistency	Some Cases	Yes
Can Backup Files on a NAS Appliance through CIFS or NFS	No	No
Best Suited	Streaming Disk I/O	All File I/O