HOW VIRTUALIZATION CUTS CDP COSTS BY 80%

WRITTEN BY:

DAVID DAVIS
VEXPERT, VCP4, CCIE

CONTENTS

CONCEPT

Continuous data protection (or CDP) has been used by large enterprises for some time for disaster recovery and to achieve the highest level of business continuity. CDP synchronizes tier 1 data between sites to ensure availability of the most critical enterprise applications. Until recently, CDP has been used only by the largest companies and only for their most critical data. Even though these enterprises would like to use CDP for lower priority tier 2 data—and small and medium businesses (SMBs) would love any type of CDP—it was simply cost-prohibitive.

Thanks to advances in virtualization, advanced disaster recovery methods that result in fast recovery and low system downtime are affordable by businesses of all sizes. Near-CDP (defined below) is now available to all organizations, from the largest enterprise to the smallest business, for a very reasonable cost. This option should be appealing to IT directors, CIOs, and especially CFOs.

Before I explain how, let me first offer a definition of CDP and explain why your company (large or small) needs CDP.

5 THINGS YOU NEED TO KNOW ABOUT CDP

Whether you are already paying big bucks for a disaster recovery solution that uses SAN (storage area network)-based CDP or you aren't sure what CDP is, here are 5 fundamental concepts that we need to lay out before moving forward.

1. **CDP synchronizes data between two or more sites**, either across a wide-area network or WAN (offsite) or within the same data center (onsite). CDP is used onsite to make mission-critical applications and storage systems highly available, and it's used offsite for disaster recovery purposes in the event an entire data center is lost.

2. **CDP can provide offsite backup for disaster recovery AND replication for high availability**. So unless your IT policy dictates otherwise, if you're replicating offsite, you don't need to also back up the data and you don't have to manage backup tape media, which can result in large cost and time savings.

3. **CDP takes business continuity and high availability to the next level**. Unlike most traditional high availability (HA) solutions, CDP performs data synchronization that results in HA for the underlying data storage system. For example, HA solutions like VMware High Availability and VMware Fault Tolerance (FT) only provide HA in order to keep the virtual machine (VM) up and running, not HA for the underlying storage. With CDP, if your primary SAN storage becomes unavailable, protected VMs can be pointed to the secondary SAN storage with the replicated data and be quickly brought back up.

4. **CDP is synchronous data replication** where data is written to both the primary and the secondary disk storage at the same time, before the database or application is notified that the data has been written. Synchronous data replication requires expensive storage hardware and software, as well as low-latency, high-bandwidth WAN links.

5. **Near-CDP is asynchronous data replication** where data is written to only the primary disk storage before the database is notified that the data has been written, and the data is later sent to the secondary disk storage. With virtualization, low-cost storage and lower bandwidth/higher latency network links (such as an Internet VPN tunnel) can be used for secondary storage.

3 OPTIONS FOR CDP

As an IT manager, you know that there is always more than one way to get a job done. CDP is no different. Here are 3 options for implementing CDP.

Option 1: "Legacy CDP" or Classic SAN-based CDP

- **Implementation**: High-dollar SAN hardware and associated CDP software are used at each site to replicate entire SAN disks.

- **Pros**
  – Has an RPO (recovery point objective) of zero (that is, no data loss)
  – Can work with VMware Site Recovery Manager (SRM)

- **Cons**
  – Most expensive option
  – Must have identical SAN hardware at each site
  – Must replicate entire SAN disk
  – Cannot fail over an individual VM
  – Requires manual registration of all VMs in the secondary site, which can jeopardize RTOs (recovery time objectives)
  – Does not provide true CDP for applications that require Windows VSS (Volume Shadowcopy Services) for transaction consistency
  – Requires low-latency/high-bandwidth connection that is difficult to achieve with offsite replication and can impact performance of the production SAN

- **Estimated Cost**: CDP software options of $20,000-$60,000 for each site (depending on the size of the SAN), plus identical SAN hardware for the secondary site (which can easily cost $100K or more).
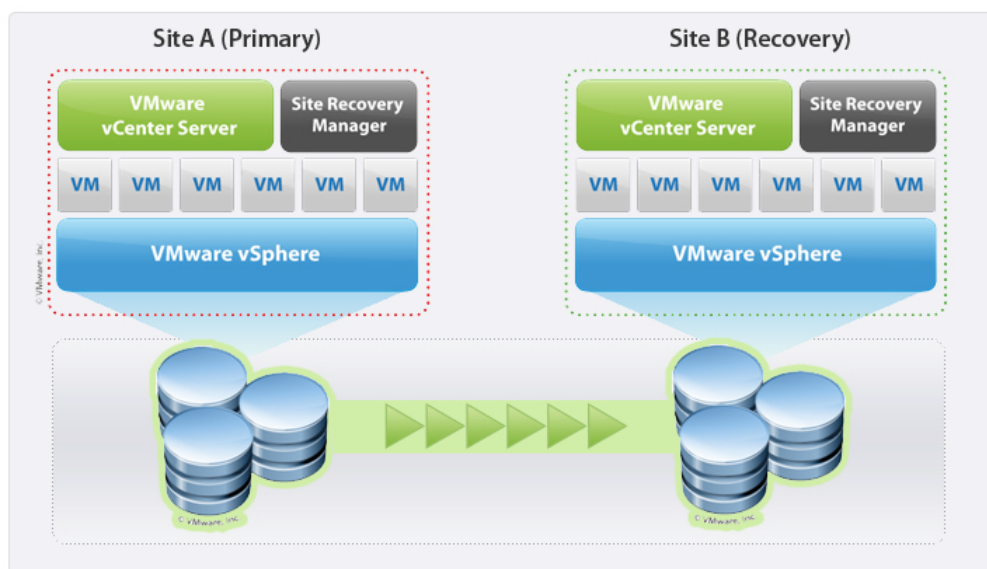
- **Vendor Example**: EMC RecoveryPoint, NetApp SnapMirror

Figure 1. SAN-based CDP.
Entire SAN disks are replicated, and identical hardware
is required for the recovery site.

Option 2: Software-based Server Replication

- **Implementation**: Software is loaded on each physical (or virtual) server, and all data on that server is replicated using asynchronous replication to an identical server running the same software at the secondary site.

- **Pros**: Lower cost option for a handful of physical servers that need data replication.

- **Cons**:
  - Oriented to physical servers (but can be installed in a VM)
  - Does not replicate VM configuration (for example, VMX files) or other VM "metadata"
  - Doesn't scale well to a large number of servers
  - Has an RPO of approximately 5 minutes (that is, up to 5 minutes of transactions can be lost)
  - Doesn't take advantage of VMware vSphere Changed Block Tracking (CBT)
  - Can put significant load on a host if multiple VMs on the host are replicated
  - Still costly: requires identical hardware for the secondary server, requires replication software licenses on both the primary and secondary servers, and requires VM replicas to be powered on, which increases license costs

- **Estimated Cost**: $2,000-$10,000 per physical server.

- **Vendor Example**: Double-Take Software Availability
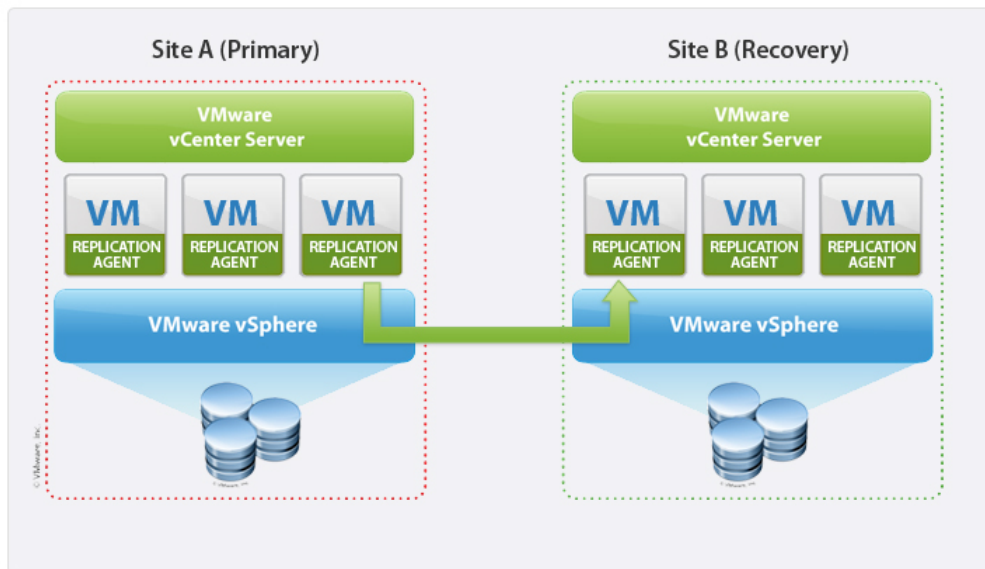
Figure 2. Software-based server replication.
Asynchronous replication is performed via a software agent inside the VM.

Option 3: Virtualization-enabled "Near-CDP"

- **Implementation**: Servers running VMware vSphere track changed data in VMs, and that changed data is sent using asynchronous replication across a LAN or WAN.

- **Pros**
  - Lowest cost option
  - Affordable for all tiers of data
  - Can replicate between different types of hardware
  - Doesn't require VM replicas to be powered on
  - Is application aware and can use Windows VSS to quiesce transactional applications to ensure data consistency
  - Uses the latest VMware vStorage technology

- **Cons**
  - Has an RPO of approximately 5 minutes
  - Potential for slight delay in transaction processing when application consistency options such as Windows VSS are used

- **Estimated Cost**: Several hundred dollars to a few thousand dollars per host, depending on the size (number of CPUs) of the host.

- **Vendor Example**: Veeam Backup & Replication
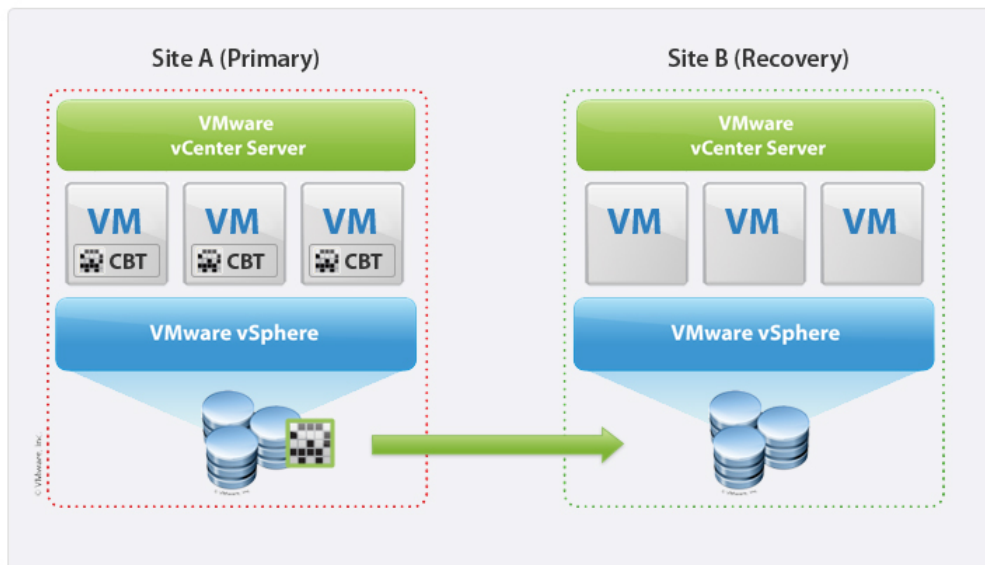
Figure 3: Virtualization-enabled "Near-CDP".
Only changed blocks are replicated via the storage layer. Allows 5-minute RPO and supports replication between different types of hardware.

Summary Table

3 Options for CDP			
Feature/Benefit	SAN-based	Software-based	Virtualization-enabled
Supports onsite and offsite replication	Available	Available	Available
Can replicate between dissimilar storage hardware	Not available	Available	Available
Replicates entire VM	Available	Not available	Available
Designed for virtualization	Partially available	Not available	Available
Application-aware	Partially available	Available	Available
Uses Changed Block Tracking	Not available	Not available	Available
Recovery point objective (RPO)	0[1]	5 min.	5 min.
Cost	$$$$	$$	$

Available   Partially available   Not available

HOW VIRTUALIZATION MAKES CDP AFFORDABLE

With companies moving more and more toward 100% virtualization of servers, it makes sense to take advantage of the many benefits gained from using virtualization. One of these benefits is that your virtual servers are hardware independent. This provides portability from the primary site to a secondary site where the physical server and storage hardware can be totally different.

When VMware released vSphere 4 in 2009, they made some big changes to the vStorage APIs for Data Protection that the backup and replication vendors use. The most notable change was in the area of CBT. With CBT, the vSphere ESX(i) 4.x server keeps track of blocks changed on a VM's virtual disk.

---

[1] An RPO of zero is possible only if synchronous replication can be maintained. This is difficult if replicating offsite and is impossible if using VSS.

With this change, virtualization-enabled CDP—or Near-CDP—becomes the most affordable and most logical disaster recovery choice for businesses of all sizes. Here are 3 reasons why:

1. **More Dependable**: With Near-CDP, replicated VMs are consistent and dependable. Database transactions and file systems can be put into a consistent state when backups or replicas are performed, and you have full confidence that you can quickly get these critical applications up and running without lengthy database consistency or file system checks. Legacy CDP solutions may have optional VSS agents for transaction consistency, but if you use these agents, replication is asynchronous and you lose the zero RPO that you paid so dearly for.

2. **Less Cost**: The cost for Near-CDP is 80% less than legacy CDP.[2] Near-CDP doesn't require you to have the same expensive SAN equipment at the primary and secondary sites. For example, you could have a low-cost iSCSI SAN appliance at the secondary site where all VMs are replicated to. Additionally, you aren't tied to the expensive CDP solution from your SAN vendor and instead are able to choose from a variety of lower cost solutions.

3. **Most Scalable**: With Near-CDP, you replicate exactly what needs to be replicated—namely, the VM disks where your critical data is stored. With legacy CDP, you are forced to replicate entire SAN disks. With Near-CDP, as the number of VMs grows, you can easily opt to replicate all VMs, no matter what type of storage they are contained on.

CONCLUSION

Thanks to advances in virtualization, the highest level of business continuity and disaster recovery is now affordable. Companies of all sizes should consider implementing Near-CDP for their virtual infrastructure. With Near-CDP offering higher reliability, lower cost, and more scalability, using Near-CDP for all VMs is the smart thing to do. Large enterprises should re-evaluate their use of legacy SAN-based CDP for tier 2 data and consider Near-CDP with virtualization. At the same time, SMBs should evaluate virtualization-enabled data replication to protect critical company applications and implement Near-CDP.

---

[2] 80% is a conservative estimate that comes from a consolidation ratio from the primary (production) site to the secondary site of 5x. In other words, a server in the secondary site can hold 5 times more VMs than a server in production. If you take into account other costs like expensive SAN hardware and more complex management, the cost savings are even greater.

About David Davis

David Davis is Director of Infrastructure at TrainSignal.com, the global leader in video training for IT pros. He has a number of certifications including vExpert, VCP, CISSP, and CCIE #9369. Additionally, David has authored hundreds of articles and six different video training courses at Train Signal with his most popular course being the VMware vSphere 4 video training course. His personal website is VMwareVideos.com. You can follow David on Twitter or connect with David on LinkedIn.

About the Sponsor

Veeam Software, a premier-level VMware Technology Alliance Partner and member of the VMware Ready Management program, provides innovative software for managing VMware vSphere 4 and Virtual Infrastructure 3. Veeam offers an award-winning suite of tools to assist the VMware administrator, including:

- **Veeam Backup & Replication:** #1 for Virtualization Data Protection
- **Veeam Reporter Enterprise:** for VMware performance, storage, and capacity reporting and chargeback
- **Veeam Monitor:** for vSphere performance monitoring and alerting across multiple vCenter Servers
- **Veeam Business View:** for business-aligned management of virtual machines, independent of the virtual infrastructure.

With its acquisition of nworks in June 2008, Veeam's products also include the nworks Smart Plug-in and the nworks Management Pack, which incorporate VMware data into enterprise management consoles from HP and Microsoft. Learn more about Veeam Software by visiting http://www.veeam.com/.