Copy Data Management:  
In-Place vs. Rip and Replace  

by George Crump, Lead Analyst

Organizations are increasingly identifying the need to address today’s challenges with managing Copy Data. Copy Data is the term used to describe the copies of production data made for various business functions, including data protection, development and testing, archives, eDiscovery and analytics. It is growing at an alarming rate as critical IT and business functions beyond data protection demand access to versions of production data.

As a sign of the growing recognition for the need to better control and leverage secondary data, a fast-growing category of solutions had arisen focused on Copy Data Management (CDM). CDM technologies allow organizations to more efficiently leverage secondary copy data resulting in reduced data across the organization through better control. Multiple CDM approaches have appeared on the market falling into one of the following categories reflective of their approach; in-place CDM and rip-and-replace CDM (off-host).

Copy Data Core Tenets

There are two core tenets of CDM that are interdependent. First, curtail the growth of secondary data, which in some cases is now consuming as much as 60% of data center storage capacity. Second, improve IT’s ability to deliver core services that rely on secondary data to application owners and users by way of simplified management and delivery of useable data copies.

The Snapshot Foundation

Most CDM software leverages snapshot techniques as their foundational components – But snapshots, without some core added capabilities, have significant limitations. While snapshot technology is available on almost every storage system, its potential remains greatly unrealized.

Today in most environments, snapshots are little more than the first line of defense in data protection. Should a file get deleted, or a database be corrupted, the IT team’s first step is to roll back to the most recent snapshot and re-instantiate the application at that point in time, accepting some potential data loss for any new or modified data since the most recent snapshot.

Meanwhile, when IT is asked to run a DR test, or create a new Dev/Test environment, snapshots are rarely used; instead other technologies, primarily backup and replication, are used to create brute force data copy and movement. The root of the problem is that today snapshots lack a management framework, making them difficult to search, and difficult to re-purpose to support other data access and data availability needs.

CDM technologies should be designed to fill these gaps. Essentially CDM should enhance snapshots and make them more readily accessible, allowing them to be valuable across a
broader range of use cases. How the CDM solution creates and takes advantage of this
snapshot technology is a key differentiator between solutions. Some CDM solutions leverage the
snapshot technology that is already in place in the customer’s existing storage infrastructure,
while others build out a new set of snapshot services, replacing what is already there.

**Leveraging Existing Infrastructure With In-Place CDM**

In-Place CDM leverages the existing production storage infrastructure, and the storage system’s
native snapshot and replication capabilities. As a starting point, such a CDM approach can
organize the existing, historical snapshot data providing valuable visibility and insight into the
status of the secondary data across the enterprise before any new snapshots are taken. Once
implemented the CDM solution can execute, organize and automate all future storage system
snapshots.

The snapshot infrastructure is more than just copies of data though. Most organizations have an
extensive list of policies and procedures around the snapshot processes like scripts that interface
with an application before and after a snapshot is taken. A key advantage for an in-place CDM
solution is that it becomes the management framework for the existing infrastructure. The
organization’s existing policies can be migrated and owned by the in-place CDM platform, while
new policies are created by the solution. The result is that processes today that rely on a
hodgepodge of tools and scripts become centralized and automated. The use and re-use of data
in support of the various functions becomes simplified.

Because in-place CDM relies on the underlying capabilities of existing infrastructure, the
challenge becomes delivering support to the various storage systems on the market today. Each
storage vendor has implemented snapshots differently, so the in-place CDM vendor has to adapt
their solution to each specific storage platform as well as maintain compatibility with updated
hardware.

**Starting Over with Off-Host CDM**

Off-host CDM implements a whole separate tier of storage hardware and software, essentially
duplicating the customer’s existing investment in storage and the associated storage services—
snapshots, compression, replication, etc. These solutions interface at either an operating system
or hypervisor level to capture snapshots or net new writes at given intervals. The copy data is
then written to either a hardware stack that the CDM vendor provides or the customer has to put
together. In either case, it creates yet another storage hardware silo that has to be independently
managed.

Once the copy is created on the new platform, the CDM solutions then provide software that
manages the presentation of copy data in similar fashion to the in-place solutions mentioned
above. If the customer can work through the potential turmoil that the off-host solution can create
the final comparison comes down to the capabilities of the software.

An advantage of the off-host approach for the vendor is that it does not need to worry about
developing code that will support the existing storage infrastructure. This approach is a negative
for the customer since they need to replace their entire data protection process with a new one
that is incompatible with their current software, hardware and existing processes.

**Comparing In-Place to Off-Host Copy Data Management**

Time to Implement - An in-place solution can leverage the storage services that are already in
place, by using APIs to bring those services within the overall copy data management
framework. Aside from the virtual appliance, no other software needs to be installed or enabled
on the physical or virtual servers in the environment. With an off-host solution, software needs to be installed and new hardware configured or at least old hardware needs to be repurposed. In a shared storage environment, this can mean additional configuration of the storage area network to accommodate a new storage system.

Time to Learn - With in-place solutions, initially nothing needs to change. Storage administrators can leverage the existing processes they have and then slowly, as they have time, begin to leverage the functionality of the CDM. With in-place solutions there are new capabilities to take advantage of but learning how to do that is non-disruptive. Off-host CDM requires a hard cut over to new processes and procedures with an immediate sharp learning curve.

Hard Costs - In-place solutions have a clear advantage from a cost perspective since they leverage the existing, already paid for, hardware and storage software. Off-host requires a completely new way of performing data protection and other copy creation functions, and a new storage hardware silo to be implemented at additional cost.

Even if the off-host CDM vendor allows the customer to select their own hardware, this can’t be the cheapest commodity storage hardware on the market. Given the processes that the CDM solution will now host, and that the CDM solution as one of its core tenets will limit the number of secondary data copies available, the performance and reliability of this secondary storage system is more critical than ever.

Time To Value - The single biggest advantage that an in-place CDM solution has is its time to value. As mentioned above, the in-place can leverage existing snapshot histories. The off-host CDM solution needs to build up a history of snapshot data in order to make available any historical use cases for this data.

Conclusion

The decision between in-place and off-host copy data management solutions seems obvious, in-place has all the advantages, the impact of a rip and replace implantation is too severe for most data centers to tolerate. The only exception is if the in-place solution does not support the current storage hardware infrastructure or if the off-host CDM solution has dramatically better software. The in-place CDM vendors are quickly closing the gap in storage platform support. While a separate examination of software functionality is the focus of an upcoming Storage Switzerland article, it’s fair for now to say that in-place vendors are at least equal to off-host vendors in features and capabilities.

Sponsored By Catalogic

About Catalogic

In business since 1996, Catalogic is a Copy Data Management vendor that provides in-place and off-host CDM solutions, with over 1,000 customers world-wide. Their in-place solution provides a non-disruptive path to copy data management as well as complete automation and orchestration of copy data instances. Use cases for its solution include enhanced recovery, automated DR, enhanced Test/Dev, DevOps automation and near real-time data access for operational analytics.