In-place advantages

In-place copy data management offers many advantages over out-of-band approaches

Highlights

- Build more efficient copy data solutions using in-place copy data management (CDM)

- Maximize existing storage investments with IBM® Spectrum™ Copy Data Management

- Help lower storage costs by avoiding the use of additional storage resources for copy data

- Increase the efficiency of databases such as Microsoft SQL Server and InterSystems Caché

CDM is increasingly recognized as a must-have technology for the modern data center. But there remains a fundamental difference in approach among vendors: "In-place" CDM leaves data on primary storage and leverages existing storage services and functionality; "out-of-band" (OoB) CDM moves data off primary storage to a CDM appliance. IBM Spectrum Copy Data Management uses in-place CDM and offers some distinct advantages over OoB approaches.

CDM means efficiency

CDM involves the creation, use, access, distribution, retention and clean-up of data copies. Different feature sets may or may not be added on top of CDM solutions. These features can include automation, reporting, user self-service options, and representational state transfer (REST) application programming interfaces (APIs) for application development and operations (DevOps).

The goal of CDM is to eliminate the many manual steps and reduce delays currently associated with creating and delivering data copies, as well as to centralize operations into one tool rather than several. Results typically include faster delivery of copy access, reduced storage capacity consumption, dramatically lower operational overhead, improved reliability and lower storage costs. Currently, CDM solutions leverage one of two approaches: OoB or in-place architectures.
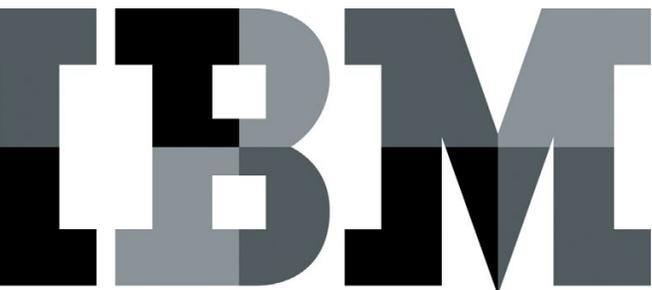
OoB CDM

In the OoB model of CDM, data is moved from the source (production) storage to a target storage device—a storage stack separate from primary storage. The data is typically moved by a client agent residing on the host server. The copy method is usually some form of changed-block model. While this reduces the impact on the host compared to traditional backup,[1] the impact is not zero, because OoB methods do affect host CPU utilization and bandwidth.
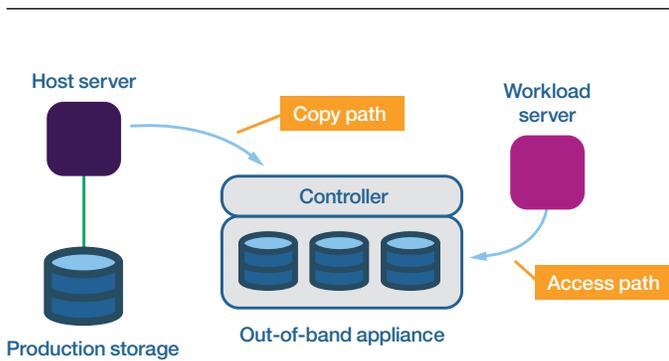

Figure 1. Out-of-band CDM architecture

The target device is usually an appliance, though some vendors offer "bring-your-own-storage" software versions or virtualized appliance options. It is important to realize that the appliance is a separate storage device with its own controller and firmware. This can add significant cost and complexity.

When data copies are used for reporting, DevOps or other tasks, a user workload server connects to the OoB appliance. If data changes need to go back to the production side, they are copied back over the network. Multiple copy workload servers can connect to a single copy, which is parceled out via snapshot images. This is how efficiency benefits are realized. But the performance characteristics of OoB appliances may be different from those of the hardware used for production storage. Figure 1 illustrates the OoB approach.

In-place CDM

IBM Spectrum Copy Data Management uses an in-place CDM model. This approach makes use of primary storage array copy processes—including snapshots, replication and clones—as well as data reduction inherent in the array, such as deduplication or compression. IBM Spectrum Copy Data Management functions as a *control plane* and doesn't replace any storage array features; it makes use of them, providing features including:

- Automated copy delivery
- Automated copy refresh
- Copy lifecycle management
- Self-service and API copy creation and management
- Application integration (for issues including data consistency and log management)
- Hypervisor integration
- Database integration
- DevOps tool integration

With in-place CDM, there is no separate storage device involved. IBM Spectrum Copy Data Management communicates to both the host server and the production storage, driving copy processes.
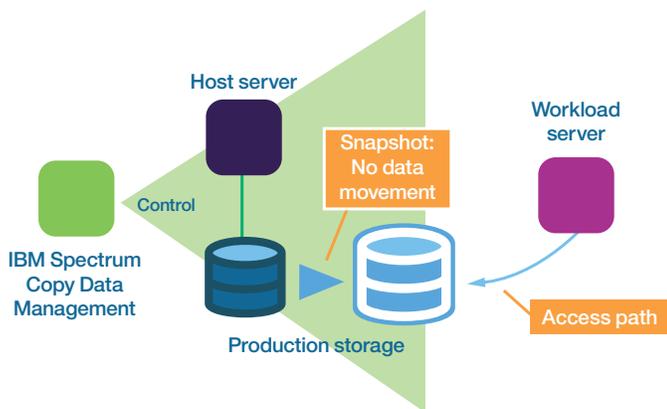
Figure 2. In-place CDM architecture

In this model, there is no data movement, because storage snapshots are created in place. Access to the copy workload server is provided using existing storage cloning features and by connecting the snapshot clone to the copy workload server. With the in-place approach, recoveries or data copying are far quicker than the OoB approach because little or no data movement is involved, and nothing traverses the network. Figure 2 illustrates in-place CDM.

In-place CDM is superior to OoB approaches because it:

- Uses existing storage resources and features
- Maximizes the current hardware investment
- Reduces total number of copies under management
- Avoids the use of additional storage and eliminates management of separate storage devices
- Minimizes data capture impacts on the host
- Maintains data reduction (there is no need to "rehydrate" data for copying to an alternate tier)

Resolving the use of production storage

There is one aspect of the in-place model that should be discussed. Not everyone is comfortable with running non-production workloads on production storage. Even though all-flash arrays have plenty of performance to spare and typically can run non-production workloads without performance impacts (unlike spinning disk arrays), many organizations simply won't allow non-production work to take place on production storage.
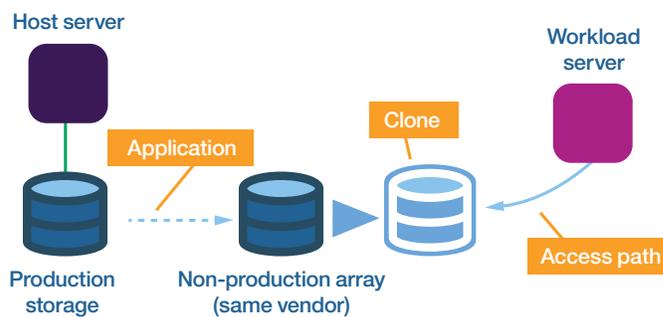

Figure 3. Replicating production data to non-production storage

IBM Spectrum Copy Data Management provides an easy resolution for this issue. Because the IBM solution can manage replication as well as snapshots, it's a simple enough task to replicate data from production to alternate storage and run non-production workloads on the alternate array. Figure 3 illustrates this solution concept.

This simple approach provides complete array separation. A major benefit is that by using array replication, you maintain all the data reduction from the primary array, thereby greatly decreasing the amount of data replicated. And no data goes through the host, which limits network and host impact and reduces complexity.

Test/Dev benefits from in-place CDM

CDM end users have highlighted another benefit of in-place CDM. Consider application development and testing (test/dev) use cases. In the OoB model, test/dev work is being conducted on storage that may be *fundamentally different from production storage*. It may have different software, a different file system or different disks, any one of which could mean different performance characteristics.

This means that if you are doing application development work on a different system, when you move the application back into production *you can't be certain it will work properly*. An element of risk remains. And when something *does* go wrong, it inevitably leads to finger-pointing as development blames IT for an infrastructure problem and IT blames development for problems with the code. Troubleshooting this kind of problem can be nightmarish, because the source of the problem could be anywhere on either side of the divide.

In the IBM Spectrum Copy Data Management approach, the same array type is used for production and non-production use cases, so test/dev is conducted on a storage stack *that is similar to the production stack*. Even the copy processes use the native array software. IBM Spectrum Copy Data Management *does not introduce any software changes*; it only manages the copy process by talking to the array APIs. This is a low-to-no-risk infrastructure model for application development.

Superior database copy management

Relational database management systems (RDMS) are a foundational component of most IT environments, managing and delivering the data used by the business applications that power enterprises around the world. Just like the applications they support, databases also produce a number of data copies used for local and remote recovery, development and test, reporting, compliance, and many other functions. Most RDMS environments involve multiple database instances, multiplying the number of database copies that consume valuable data storage resources.

For these reasons, CDM is a valuable tool for increasing the efficiency and reducing the overall system costs of RDMS environments, as well as of the business application portfolio. As noted, in-place CDM solutions such as IBM Spectrum Copy Data Management leverage the capabilities of existing storage systems. This means that performing effective copy creation and management requires knowledge of the RDMS and specific points of integration to ensure that the copy process has minimal impact, completes successfully and results in application-consistent copies every time.

IBM Spectrum Copy Data Management leverages database tools and functionality in the same manner as it does other relevant storage system features. Two of the most common RDMS solutions—Microsoft SQL Server, for general application environments, and InterSystems Caché, used to manage Epic electronic health record (EHR) data—provide excellent examples. IBM Spectrum Copy Data Management integrates storage array copy processes such as snapshots, replication and clones with SQL Server and Caché, allowing enterprises to leverage their existing storage infrastructure to eliminate copy sprawl. At the same time, IBM Spectrum Copy Data Management helps dramatically improve the solutions' capabilities to deliver fresh copies of key databases to those functions and business units that demand them.

IBM Spectrum Copy Data Management supports SQL Server and Caché instances running on either physical servers or VMware virtual machines. As in any other application environments, administrators create copy policies that define key copy metrics such as how often to make copies, where the copies will

reside (local, remote or both) or how long copies are retained. Once a policy is created, it can easily be applied to any SQL Server or Caché workload. No more complex scripting or working with multiple tools—just an easy click-and-apply function. IBM Spectrum Copy Data Management provides all the same advantages to databases as it does to any other application environment—automation and self-service, simplified data protection and disaster recovery, complete application development and test copy automation, DevOps tool support via APIs, and simplified licensing.

The better choice

In some ways, the choice between OoB and in-place CDM is reminiscent of the early days of data deduplication. IT decision-makers actively debated the merits of each data deduplication approach: inline dedupe during backup versus post-processing dedupe after backup, for example. The debate went on for some years, until hardware improvements made it impossible to argue anymore. In-line dedupe won—because it was clearly more efficient. A similar shakeup is happening with copy data. OoB was the first approach that came to market, but in-place CDM now offers more advantages.

The differences are straightforward. The OoB model requires deployment of additional storage hardware and software that may have different operational and performance characteristics. Plus, OoB CDM appliances can be quite expensive. But the costs are more than just the initial purchase price. With OoB, you have to manage and maintain a separate storage environment just for the sake of your copies.

Scalability can also become an issue. Deployments of IBM Spectrum Copy Data Management have encountered users with thousands of database instances, each of which spins off eight or 10 copies for non-production workloads. How many copy appliances will you need in order to address this level of workload demand?

The in-place approach provided by IBM Spectrum Copy Data Management lets you maximize the storage investments you've already made in both money and operational know-how. If required by budget limits and allowed by policy, you can run non-production workloads on production storage, which means that with an easy, nondisruptive software tool you can bring automation, self-service and other operational enhancements to *your existing storage footprint*. If you need to deploy alternate storage for CDM purposes, IBM Spectrum Copy Data Management fully manages operations across nodes: production array snapshots for recovery and the replication needed to move data onto separate storage.

After careful comparison, it becomes clear that the in-place CDM approach used by IBM Spectrum Copy Data Management is the better choice.
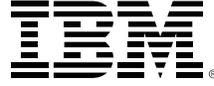
For more information

To learn more about IBM Spectrum Copy Data Management, please contact your IBM representative or IBM Business Partner, or visit:

ibm.com/us-en/marketplace/spectrum-copy-data-management

Additionally, IBM Global Financing provides numerous payment options to help you acquire the technology you need to grow your business. We provide full lifecycle management of IT products and services, from acquisition to disposition. For more information, visit: **ibm.com**/financing

[1] "IBM storage and Actifio copy data virtualization," *IBM Corp.*, October 2015. https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=TSW03322USEN

Please Recycle

TSW03543-USEN-00