



I D C T E C H N O L O G Y S P O T L I G H T

Solving the Copy Data Problem with In-Place Copy Data Management

October 2015

Adapted from *The Copy Data Problem: Analysis Update* by Phil Goodwin and Ashish Nadkarni, IDC #254354

Sponsored by Catalogic

Increasingly, IT organizations are taking a hard look at the management of their copy data, and for good reason. It is easy to recognize that today, managing copy data is characterized by significant cost and complexity. Yet despite devoting significant resources to the processes involved in creating and managing data copies, businesses are struggling to achieve the data access and data availability needed to meet their service-level agreements (SLAs). This recognition is driving IT organizations to seek a holistic approach to copy data management (CDM) in order to reduce cost and complexity and improve the ability of the IT team to deliver the required data access for the various functions that require it. This paper examines how better management of copy data can offer immediate benefits in terms of capex reduction and opex savings. It also looks at the critical role of copy data across a myriad of business and IT functions and how better CDM is one of IT's most powerful levers for dramatically improving the delivery of IT services to the business.

Consider the role of copy data today: Companies make copies of data for disaster recovery, backup, archiving and compliance, development and testing, analytics, reporting, and more. Interestingly, despite the similarities in the processes for creating and using these data copies, many different roles in the organization are making copies: database administrators (DBAs), quality assurance (QA) teams, virtual machine (VM) administrators, business analysts and, of course, storage and backup administrators. Moreover, these different groups typically utilize different hardware, software, and processes to create and manage copies. This growing complexity is driving the need for a new approach to more holistically manage, protect, and share corporate data.

Putting the CDM challenge in context highlights the urgency of addressing it. IDC estimates that overall, data will grow at a 34.8% CAGR through 2019 while storage budgets will increase only 3.6% in 2015. During this period, staffing levels are expected to remain flat. At the same time, SLAs are becoming more stringent, as users expect "always on" systems.

Industrywide, 60% of storage capacity is consumed by copy data. In other words, more than half of an enterprise's storage hardware budget goes toward storing and managing redundant data. Excessive copy data impacts IT organizations in two ways. First is the added cost of disk capacity. IDC estimates that this cost will top \$50.63 billion in 2018. This represents hard IT dollars that are being spent annually. The number is staggering and exposes the tip of the iceberg of how serious this problem is.

The key question is, How can IT organizations manage more data and deliver better service levels without adding budget or IT staff? Obviously, IT must find better ways to manage data, and specifically copy data.

By achieving better data management, IT organizations that maintain a 50–55% CAGR of storage growth can achieve year-over-year hardware budget neutrality. A CAGR higher than 55% will require an increased budget, while a CAGR below 50% will shrink the budget. Therefore, eliminating excess copy data is the "low-hanging fruit" for reducing the data CAGR.

Identifying smarter ways to manage and share copy data represents a potential goldmine for many IT organizations. Industrywide, a 20% reduction in data can return \$10 billion to IT, which could be used for more important business priorities.

The second way excessive copy data impacts IT organizations is that poorly managed copy data is making it increasingly difficult for IT to deliver on its SLAs for data availability, uptime, and protection. Business customers are driving increased demand for rapid access, but data sprawl and overall infrastructure complexity are forcing IT into longer delivery times and worse overall performance.

Copy Data Today

As with many scenarios in IT, today's situation is the result of many years of IT growth and the evolution of various domains within the datacenter. The result is a mixed bag of tools and technologies that create, store, and manage multiple different silos of copies of production data.

Data replicas were originally intended for data protection and business resiliency, whereby the IT team creates copies as a source for recovery in the case of either a logical event or a physical event. Focusing narrowly on this function demonstrates a wide range of technologies in play today: storage snapshots for point-in-time recovery, mirroring for system failures, synchronous or asynchronous storage replication for disaster recovery and archive, VM-level copies or clones and, of course, traditional agent-based backup. A much wider range of business and IT functions that rely on access to recent copies of production data can be added to the mix: development and testing, business analytics, and archiving and compliance.

One of the chief drivers of copy data generation is that various business units have autonomy over their own data sets. Certain business units are willing to dictate the creation of multiple copies under the premise of compliance or service quality or simply because they have the budget to do so. Depending on procurement practices, these business units end up paying for such excesses, and many times, they're ignorant about the amount of money that they're wasting in storage costs and data management costs (time and risk).

The consequences of today's crisis in copy data management are significant. First, IT is at risk of failing to deliver on its commitments to the business. Because of the clutter of data, disparate products and systems, and even disparate organizations, the ability of the IT team to deliver on one of its primary missions — keeping data secure and protected — is at risk. Backup windows are growing, not shrinking, and IT is struggling to keep up. When the ever-growing demand for copy data from other IT and business functions is added to the mix, the risk of failure increases rapidly.

A second, related consequence is that data recovery or data access times are growing longer. Whereas the business expects immediate access and an "always on" environment, the ability of IT organizations to meet the demand is actually decreasing. Third, a by-product of all the storage capacity is the overhead that it brings in terms of space, power, and cooling. Of course, all of these consequences drive increases in capex and opex.

Copy data management is a relatively new category of storage management that is evolving to address the problem by enabling the utilization of application-consistent data for a multitude of purposes.

Benefits

Most organizations are experiencing a pressing need to curb the overall IT budget that is allocated to storage. One method would be to create fewer data copies, which is not likely to happen overnight. But by deploying a CDM solution to holistically manage the life cycle of copy data, IT teams can find immediate efficiencies that allow them to quickly reduce the burden of copy data from 60% of total capacity to a more reasonable 40% or less.

As the IT organization's use of a central CDM solution grows, the benefits increase. A firm won't eliminate redundant hardware and software spending all at once, but over time, investments in dedicated silos of infrastructure for backup, archive, business continuity, and test and development can be eliminated. CDM solutions can provide storage efficiency and application-aware data management services that eliminate the need to procure separate infrastructure for each of a business' use cases. The economics of this cannot be overlooked.

By integrating with applications, CDM can gain a level of awareness that storage virtualization solutions lack, reducing the actual physical copies of data while allowing virtual copies of data to be leveraged for multiple business purposes.

Armed with effective CDM tools, IT organizations should be in a better position to rein in runaway copies of data while improving their ability to support key IT operations and business SLAs that depend on better data access and greater data availability. The benefits are significant: reduced capex in the storage budget, slower data growth while enhancing business operations, improved service-level delivery, and less downtime.

Effective CDM addresses many use case scenarios, but most companies start by implementing a solution to address the following areas:

- More efficient data protection (including local and remote backup)
- Improved disaster recovery and business resiliency in the form of shorter RTOs and RPOs
- Active archive in response to eDiscovery, compliance, and long-term retention
- More agile and efficient application development, testing, and DevOps
- Faster data insights/big data analytics

IDC sees CDM as a critical segment of the storage infrastructure software, and it has become one of the key functional markets that IDC tracks. The primary value of an enterprise CDM solution is to allow IT managers to easily align application SLAs directly with business requirements while enabling business agility and cost savings. Therefore, CDM solutions must be considered not just another storage management layer but also a core component of IT operations.

CDM Requirements

A holistic approach to copy data management has the potential to dramatically reduce both opex and capex by helping organizations get a greater hold on all of the data copies across their different functions and divisions. Through IDC's work with organizations that are seeking to capitalize on the promise of copy data management, a clear set of requirements has emerged.

Today's IT reality drives a lot of the requirements. Given that budgets and IT staffing levels are flat, organizations need solutions that deliver clear economic results in the following areas:

- **Nondisruptive:** IT organizations don't want to create another storage silo because beginning a project with a large capital investment eliminates any potential for near-term capital savings. In-place solutions work with the existing storage infrastructure to provide a comprehensive CDM solution.
- **Optimization:** The CDM solution should allow the IT organization to create only those copies that are required and nothing more.
- **Existing environment insights:** The CDM solution should enable a clear view of the current state of the copy data environment.

- **Correlation to business requirements:** The CDM solution should enable IT to clearly associate any copy with the business function that may need access to it.
- **Repurpose of copies:** This is at the heart of the value of a CDM solution. Efficiencies are gained when any given copy can be accessed and leveraged for multiple use cases.
- **Monitoring and reporting:** A strong CDM platform must have rich reporting and monitoring capabilities, enabling IT to keep its finger on the pulse of the overall copy data environment, including SLA compliance and exception reporting.
- **Quantifiable results:** In this environment, spending must be justified. IT needs a solution that can provide quantifiable results and demonstrate economic savings.

Most organizations have heterogeneous storage infrastructures, and when the complexity of cloud storage is added to that, the back-end repository details may be unknown. Thus, IT managers will want to consider software products that can manage copy data independently of the storage media, whether on-premise or in the cloud, or a hybrid of the two.

Considering Catalogic

Catalogic Software's flagship offering is ECX, a software-defined CDM solution that allows IT teams to take advantage of their copy data by enabling them to manage, orchestrate, and analyze copy data across an entire organization. One of the primary unique attributes of Catalogic's solution is that it is designed to integrate with a client's existing storage infrastructure, taking advantage of the underlying copy services of the storage and the hypervisor. The software uses the public storage and virtual infrastructure APIs to deliver its enterprise CDM platform. This "in-place" CDM approach allows clients to manage the full life cycle of copy data without the need for a new storage silo and the associated capital expenses that it would necessitate.

As a solution that is entirely software based, Catalogic ECX is simple to deploy. Organizations need only download a virtual machine using hypervisor infrastructures, such as VMware CBT, and then register the relevant storage arrays and virtual systems within the ECX GUI. Because ECX leverages the APIs of the infrastructure components, no agents are required.

As a "full life-cycle" solution, Catalogic's ECX contains policy engines, storage workflows, and wizards that allow IT to manage the creation of copies and to orchestrate the use of the copies in conjunction with the applications with which the copies are associated. ECX also ensures that IT manages the deletion of expired copies and the capacity return after use — a key part of solving the copy data sprawl.

In a copy creation workflow, Catalogic's ECX allows IT to define the detailed policy for how, when, and where copies will be created and stored, in addition to how long each copy will be retained. For example, in a NetApp environment, ECX enables IT to create Snapshot, SnapMirror, and SnapVault policies for various use cases.

To automate and simplify data access, the software orchestrates the use of copies in conjunction with the relevant VMs to bring up live environments in support of the various business functions — whether for operational recovery, testing, analytics, or another function. ECX also shrinks capex and opex by reducing or eliminating the need for additional data copies, thereby culling data sprawl and the need to manage it. By reducing the amount of data to manage, and by simplifying the management of existing data, ECX allows IT to focus on other, higher-value activities.

ECX also offers a wide set of data analytics and reporting features that allows IT teams to better understand the state of their environment and all of the functions that rely on copy data. This dramatically simplifies IT's ability to manage copy data compared with today's status quo, which relies on a complex mix of tools and scripts.

Catalogic's latest version, ECX 2.2, released in September 2015, enables both in-place and "off-host" CDM. Unlike in-place CDM, which collocates data copies in the production environment, the off-host functionality of ECX uses VMware's VADP to nondisruptively copy VMs from any storage environment into an ECX-supported storage system, which means critical business operations are not impacted. The combination of ECX's in-place and off-host deployment enables better service levels and more agility at a lower cost.

Challenges

Catalogic does face challenges, however. For example, while the product can yield significant benefits, its deployment is very hardware/software specific. ECX 2.2 provides "in-place" CDM for NetApp and IBM storage, as well as for VMware. Unsupported configurations cannot benefit from the in-place CDM capabilities of ECX. Catalogic recognizes this and has put significant emphasis and resources into expanding the product's platform coverage as quickly as possible, with support for new storage platforms delivered frequently throughout the year.

The ever-increasing adoption of cloud computing can also present a challenge for Catalogic as clients demand a single solution that can manage data across all storage environments. Similar to the company's need to qualify additional storage platforms, Catalogic must also support the primary options that enterprises choose for public and hybrid clouds.

Conclusion

Copy data management is a growing problem that is sapping IT budgets and adding complexity at the same time that SLAs are becoming more stringent because today's users expect systems to be up 24 x 7. A centralized CDM platform is needed to eliminate duplicate data, thereby lowering both opex and capex by reducing management complexities across the different functions and divisions within the organization.

To meet the requirements and avoid creating additional data silos, a CDM platform should run in your current production storage environment and wherever duplicate data ends up. CDM should also enable IT to clearly associate any data copy with the business function that may need access to it and in doing so enable it to be leveraged for any number of business use cases.

Catalogic Software is a CDM vendor with a unique software-only approach that focuses on leveraging a client's existing infrastructure, with the promise of significant capex and opex savings. Clients evaluating CDM solutions should consider Catalogic Software's flagship CDM offering, Catalogic ECX.

ABOUT THIS PUBLICATION

This publication was produced by IDC Custom Solutions. The opinion, analysis, and research results presented herein are drawn from more detailed research and analysis independently conducted and published by IDC, unless specific vendor sponsorship is noted. IDC Custom Solutions makes IDC content available in a wide range of formats for distribution by various companies. A license to distribute IDC content does not imply endorsement of or opinion about the licensee.

COPYRIGHT AND RESTRICTIONS

Any IDC information or reference to IDC that is to be used in advertising, press releases, or promotional materials requires prior written approval from IDC. For permission requests, contact the IDC Custom Solutions information line at 508-988-7610 or gms@idc.com. Translation and/or localization of this document require an additional license from IDC.

For more information on IDC, visit www.idc.com. For more information on IDC Custom Solutions, visit http://www.idc.com/prodserv/custom_solutions/index.jsp.

Global Headquarters: 5 Speen Street Framingham, MA 01701 USA P.508.872.8200 F.508.935.4015 www.idc.com