SPOTLIGHT

Sponsored by: Interlock Technology

Data silos, data volumes, and other causes lead to "trapped" data making it difficult for organizations to extract data value. Making that data easy to utilize begins with simple methods for migrating data using common protocols and standardization.

Improve Data Mobility by Solving Data Migration Challenges

June 2023

Written by: Phil Goodwin, Research Vice President, Infrastructure Systems, Platforms and Technologies Group

Introduction

The ability to access and use data to generate value for an organization has never been more challenging. Two key barriers to data value creation are the massive amount of data that organizations must manage and much of that data is trapped in "data silos."

According to IDC research, the amount of data captured and stored by IT organizations will grow from 6.3ZB in 2022 to 14.9ZB in 2026 (source: *Worldwide Global DataSphere and Global StorageSphere Structured and Unstructured Data Forecast, 2022-2026,* IDC #US49084022, May 2022). Less than a decade ago, petabyte-scale data environments were rare and the province of only the largest organizations. Today, petabyte-scale data estates are commonplace.

"Data silos" further stymie effective use of the data. A data silo is isolated data that is difficult to move or utilize due to logistical challenges. Due to these silos, data becomes "trapped," which diminishes its usefulness to the organization.

AT A GLANCE

WHAT'S IMPORTANT

- IT organizations commonly manage 14–20 data silos.
- Data silos make data value creation difficult, lead to unnecessary human labor, and increase data management cost.

KEY TAKEAWAY

Efficient migration is an often underappreciated technology for improving data use and simplifying data management and governance while eliminating complications from proprietary data formats and obsolete applications.

Some of the factors causing data silos include:

- **Data location:** Data may be distributed across on-premises private clouds, application-specific clouds, and multiple public clouds, with different protocols and access methods that make data movement and cross-access difficult.
- **Data type:** Data may be structured in various file system formats without common toolsets to move and manage data uniformly or easily to a desired repository or the metadata may be recorded differently depending on the application.
- **Data owner:** Data may be managed by a specific group or business unit and simply not shared with other groups or may have different movement and access policies that are incompatible with other groups.

- **Operating environment:** Data may be captured on physical Windows, virtual infrastructure, Linux, legacy UNIX, and other legacy environments with proprietary protocols that stymie data movement and use.

The results of trapped data can impact business costs, operation efficiency, and the ability to extract value from the data. For example, data can become trapped due to becoming "orphaned" when applications are decommissioned or users leave the company. Data may also be trapped in "legacy" applications using obsolete file types, operating systems, and protocols. This data may be on an expensive primary disk and, unless it is migrated to lower-cost media, will needlessly cost the company money. In addition, when data is trapped, it can only be used for the single purpose of its application and cannot be leveraged for other purposes such as analytics and value extraction.

IDC predicts that by 2027, the need for faster, higher-quality data-driven decisions will cause 80% of G2000 CIOs to mandate companywide data logistics strategies for data management, protection, and integration.

Many of the problems associated with siloed and trapped data can be addressed by modernizing data logistics. IDC defines data logistics as "the movement of data through an organization to satisfy its business needs." In a nutshell, data logistics describes how data is captured, stored, moved, protected, governed, and deleted at end of life. IDC predicts that by 2027, the need for faster, higher-quality data-driven decisions will cause 80% of G2000 CIOs to mandate companywide data logistics strategies for data management, protection, and integration.

The analogy compares package logistics and how goods flow through the supply chain; data logistics can be thought of as the data supply chain. Just as package logistics cannot function without efficient, reliable, and cost-effective transportation, data logistics cannot function without efficient, reliable, and cost-effective data migration. Package logistics companies with the most efficient transportation have a competitive advantage. Similarly, organizations that can move data where and when it is needed gain advantage through superior use of data. Data migration, when applied on an enterprise basis, can improve data utilization, reduce storage costs, and improve data management practices.

Benefits

Properly applied data migration can yield benefits in several use cases as follows:

- **Value extraction from data:** By consolidating data and eliminating silos, more effective analytics can be applied to extract value from data.
- **Data consolidation:** Data may need to be consolidated for many reasons: better utilization of disk capacity; consolidation of NAS servers for better scale; mergers, acquisitions, or divestitures; and reduction of silos to improve data management operations.
- **Better data governance:** Emerging regulations are now requiring specific data to be stored for longer periods of time. Often, however, that data is rarely accessed. Migrating it from primary storage can offer both cost savings and improved performance on primary systems, provided it does not create another silo.
- **Faster, simpler system migration:** Migrations from one system to another (i.e., arrays or operating platforms), whether this is due to a technology upgrade or competitive trade out, can be both time-consuming and challenging

due to data volumes and data format/protocol differences. An efficient data migration tool can make this transition much less difficult with faster data transfers and a new system cutover, which in some cases can take days to execute.

- **Cost optimization:** Fallow or trapped data may remain on primary storage, occupying expensive capacity, simply because it is so much trouble to move it off. Efficient and automated data migration can move data by policy to the lowest cost storage that will still meet data access SLAs, whether on premises or in the cloud
- **Selective data repatriation:** Organizations routinely repatriate applications from the cloud to on premises for various reasons. Efficient data migrations that minimize egress fees are critical to making these repatriation efforts effective — combined with data discovery engines that can help determine "move this and not that" means data can be stored in it optimally.
- **Data location optimization:** Data may need to be in a particular place for performance, data sovereignty, and other reasons. Automated data migration can ensure that data is reliably located at the right place at the right time with application compatibility.

To attain optimum benefits, data migration should be considered at an enterprise level to break down silos and gain efficiencies. Moreover, doing so early in the storage system selection and deployment process can avoid problems and ensure application and user success.

Considering Interlock Technology

Interlock Technology is a global organization specializing in data migration and data extraction software and services, supporting a wide range of unstructured data protocols and data types. The company is focused on enabling more ubiquitous data mobility at scale across regulated industries, organizations with large volumes of data, enterprises seeking to leverage the cloud, and application owners needing to adopt new data formats or protocols.

Interlock Technology has developed a four-step approach to data migration and data extraction that addresses the variability of each environment. The four steps are:

1. **Data Discovery:** Interlock takes inventory of all data residing on the source system. This assessment is intended to identify types of data being stored, volume of data, data ownership, retention requirements, application dependencies, and level of activity.
2. **Environment Assessment:** Evaluation of both source and target systems informs the amount of time it will take to perform the migration, how much bandwidth will be dedicated and available, what source system resources will be available for the migration, if the target system supports real-time hash verification, if the migration will be local or remote or to the cloud, and the cutover windows for each data set being migrated. Movement of a large number of small files will take longer than a large number of large files.
3. **Extract, Migrate, Verify:** Data is extracted from an application or proprietary protocol and moved to a target system. During the migration process, data may be transformed to accommodate application, security, and user requirements. Data is migrated from storage system to storage system, bypassing the application for faster completion. Wherever possible, Interlock leverages system APIs to expedite migration and verification of data. If data can be verified during the actual migration, the company claims total time to migrate may be cut by as much

as 50%. At the end of the migration, Interlock can provide data verification and an audit report that can withstand regulatory scrutiny.

4. **Manage Cutover:** Cutovers, especially in active environments, cause disruptions in operations and must be orchestrated to minimize impact. Interlock tracks file access trends to determine which files are static and which are accessed/modified frequently. This heatmap of activity enables Interlock to manage cutover with great precision without data loss.

Interlock addresses the need for data mobility across several use cases:

- Extraction and migration of "trapped" data (hardware and services refresh requiring compliance with applications, users, and regulations)
- Data transformation across storage protocols (NFS, SMB, CIFS, S3, LTO, REST)
- Movement of data to the cloud and repatriation from the cloud
- Data consolidation or divestment locally or over WAN

Challenges

Data migration is not a new technology, nor is the shortage of systems that include data migration competitors. Most IT staff members are familiar with data migration technology and therefore may assume they already know all about it. Interlock must break through any preconceived notions to get IT professionals to take a fresh look and be open to new methods of greater operation efficiency. There will be an inherent trade-off between what products IT administrators already know, even if they are platform specific and limited in application, versus bringing in a new vendor for specialized purposes. Interlock must show clear advantages to the latter.

Array vendors usually have data migration tools in place – and may even be free — but these may be optimized for their own systems. In cases where vendors offer a data migration tool to their platform, the tool is usually unidirectional from the competitor to themselves; they are not designed for general-purpose use cases. Interlock must educate the market regarding the value of using a third-party tool to avoid vendor lock-in and deliver enterprise data movement.

Conclusion

The massive amounts of data stored by organizations are becoming so unwieldy that current data migration models are being strained to deliver the needed performance and agility. Organizational leaders should think of their data management infrastructure like a supply chain, where superior data logistics leads to better data use and therefore relative competitive advantage in the market.

As transportation is key to efficient package logistics, efficient data movement is key to data logistics. Without simple, reliable, and cost-effective data movement, organizations cannot hope to get data where and when it is needed. Many data migration tools are available in the marketplace, and some are even free from storage vendors. However, these products may be limited in scope or intended for a specific purpose or contribute to vendor lock-in. By implementing a general-purpose data migration tool, organizations can apply an enterprise approach to data management. This approach can help deliver greater economies of scale, standardize data management operations, and reduce or eliminate data silos. This means getting data to the right place at the right time at the lowest possible cost.

About the Analyst

Phil Goodwin, *Research Vice President, Infrastructure Systems, Platforms and Technologies Group*

Phil Goodwin is a Research Vice President within IDC's Infrastructure Systems, Platforms and Technologies Group, with responsibility for IDC's infrastructure software research area. Mr. Goodwin provides detailed insight and analysis on evolving infrastructure software trends, vendor performance, and the impact of new technology adoption. His focus is on multicloud data management, data logistics, on-premises and cloud-based data protection as a service, cyber protection and recovery, recovery orchestration, and more.

IDC Custom Solutions

The content in this paper was adapted from existing IDC research published on www.idc.com.

IDC Research, Inc.
140 Kendrick Street
Building B
Needham, MA 02494, USA
T 508.872.8200
F 508.935.4015
Twitter @IDC
idc-insights-community.com
www.idc.com