

WHITE PAPER

# How To Optimize Your Hybrid Cloud With Dynamic Data Migration

# Table of Contents

- 03** **Executive Summary**
- 03** **The Challenges of Data Management Are Deep and Widespread**
- 04** **A Bright Future With Dynamic Data Migration**
  - 04** **Analytic Innovation**
  - 05** **Security**
  - 05** **Resource and Expertise Efficiency**
  - 05** **Cloud Spend Management**
  - 05** **Governance and Compliance**
  - 05** **Data Mobility**
  - 05** **Technical Debt Reduction**
- 06** **Your Ideal Dynamic Data Migration Solution Is Ready: Pentaho**
  - 06** **Pentaho Data Catalog**
  - 06** **Pentaho Data Integration and Analytics**
  - 06** **Pentaho Data Storage Optimizer**
- 07** **Dynamic Data Migration Optimizes Your Hybrid Cloud**



## Executive Summary

The promise of cloud computing was enticing, but incomplete. Theoretically, who wouldn't want near-infinite scalability with pay-as-you-use pricing? But with concerns over security, unforeseen costs, and vendor lock-in, most organizations have already moved to hybrid cloud. Ironically, the result is still more silos in more locations: on premises, in private clouds, and in public clouds.

The thing is, data must move freely and continuously among your platforms so that sensitive data moves inside the firewall, cold data shifts to inexpensive stores, and relevant data goes near analytic applications – at scale, at speed – and back again when needed. This complexity begs for automation, but most tools are lacking. In fact, many legacy data integration tools actually hamper data agility in hybrid clouds.

IT leaders are demanding change, and change is already emerging. In this white paper, we explore the challenges addressed and practical application of dynamic data migration for three of your major IT objectives: 1) Data governance, sovereignty, and compliance; 2) Cloud cost control; 3) Analytics performance.



## The Challenges of Data Management Are Deep and Widespread

Recent studies reveal that IT is dealing with significant issues in security, performance, and cost. IT leaders worry about major data breaches within their cloud-centric ecosystems. Many report that analysis is hampered by data structures and formats that trap data and limit its movement. Industry analysts find that most infrastructure and operations (I&O) leaders have experienced public cloud cost overruns that impacted their on-premises budgets.

Many of the reasons behind these data management issues are well-entrenched. Lack of resources and expertise is one. The current crop of cloud-based tools seem attractive but each one has a steep learning curve. The more clouds and tools you employ, the harder it gets for your teams.

### The Data Struggle

IT leaders must manage critical data issues.



#### Security

- Breaches in cloud-centric ecosystems.
- Data location bound by privacy and regulatory requirements.



#### Cost

- PData dispersed across cloud providers.
- Siloed data and data sprawl.



#### Performance

- Data must be physically located closer to analytics workloads.
- Data structure and format preventing analysis.

It's also difficult to manage cloud spend. Pay-as-you-use cloud offerings and apparently low storage costs are attractive. But examine the details such as costs that are affected by usage characteristics. For example, cloud data egress is often far more expensive than writing data.

Governance and compliance continue to challenge IT because, today, anyone with funds can add new cloud services to your organization's portfolio. Also, your IT and business operations must coordinate joint goals for both governance

and business outcomes. Data hygiene is vital and migrated analytic data must maintain concurrency with its source, which means your migration projects may never end.

Don't forget the risk of lock-in because data migration tools for each cloud vendor are often exclusive. Migration can be the first step in a journey toward limited options. It may also be the start of growing technical debt because once you start migration, the original systems often cannot be switched off, committing you to ongoing, costly maintenance.

# A Bright Future With Dynamic Data Migration

It's no secret that today's successful businesses ride on digital transformation and data. Lots of data. Frankly, the not-so-hidden danger is that data dependence has been rushing forward faster than data management can keep up. But now data management is finally accelerating to the benefit of your business and your customers.

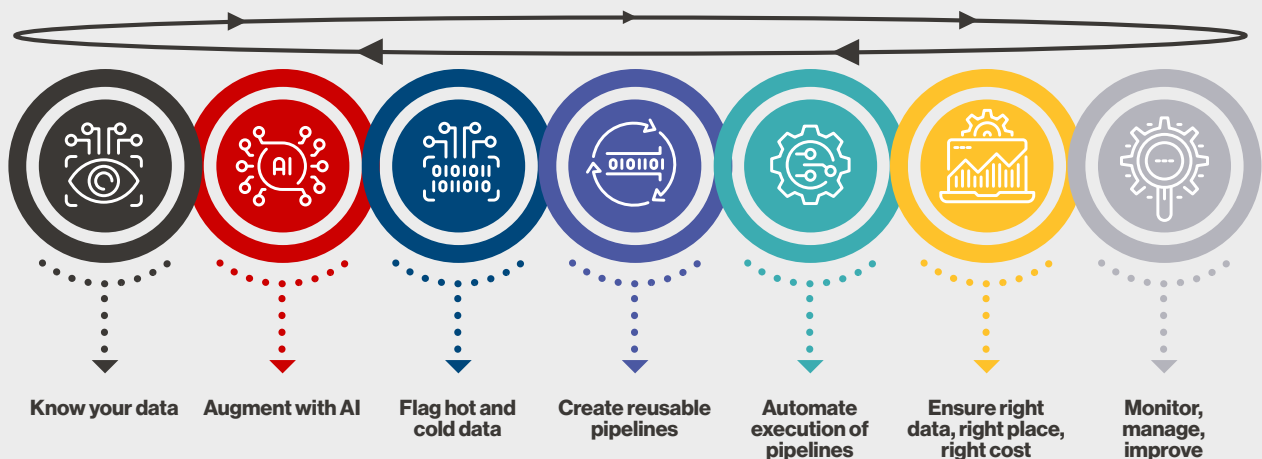
The more data you must manage, the more you need to move it around for purpose and for profit. To help you evaluate your next move, we propose seven criteria for the next generation of data migration, which we call "dynamic data migration":

## 1. Analytic Innovation

Modern competitive advantage for your business comes from deriving more value from your data. It means putting systems in place to accelerate discovery, curation, classification, contextualization, and access to data. In other words, putting the right data in the right place at the right cost, consistently over time.

The foundation is an analytic data platform that takes advantage of new cloud-based analytic capabilities to extract value from data in combination with traditional on-premises solutions. It's a system that allows different personas to work with relevant data and appropriate analytics, defining optimal placement of the data and analytic workloads in support of FinOps goals.

The user experience is enhanced by a single user-interface with an intuitive, shopping-like experience for discovering data



## Dynamic Data Migration Cycle

Institute these steps for always-on high-performance data migration.

across the edge, core, and cloud, with the ability to easily verify its quality, sensitivity and fit for purpose. Users also benefit from accelerated enablement through self-service access to datasets. At the same time, IT administrators can monitor and measure key performance indicators (KPI) for data migration activities for business impact, project timeline efficiency, and enhanced user experience and adoption.

## 2. Security

To mitigate risks and ensure that all data across the hybrid cloud is protected and safe, all data must be secured with access controls in place. They must conform to the organization's policies and work with desired security tools. This includes data at rest wherever datasets are hosted as well as any migration or other tools that need to connect, read, change, or move data.

## 3. Resource and Expertise Efficiency

Accelerate time to value by minimizing the need for specialist resources and expertise. Keep to a minimum the number of tools that your staff must learn. Use tools that work across platforms, including on-premises and cloud, and all data types. These tools must have an intuitive, GUI based interface to accelerate their adoption and make their operation less prone to error. In addition, it is essential that efficient data migration must provide a self-service, shopping-like experience for finding and validating data wherever it resides.

## 4. Cloud Spend Management

Minimizing total cost of ownership (TCO) cannot be at the expense of delivering other key outcomes. Modern data migration must see that not all data is equal. IT must be able to automatically tier data to storage that is appropriate for its usage profile. For example, it must recognize and act on data that is hot or cold, high- or low-sensitivity, and so on.

Migration must also recognize that not all data is needed everywhere. Any dataset should contain only the data needed to support its purpose. In other words, move, host, and manage only what is needed, but retain the flexibility to adapt to changes.

Avoiding unnecessary cloud egress costs is also essential. Finding and validating data for analytics purposes is mostly accomplished using metadata, which can be stored in places where egress penalties do not apply, so that data is only egressed and charged for when needed.

## 5. Governance and Compliance

Regulatory and other requirements are more easily achieved by reducing risks and making data inherently more usable.

These policies must be supported proactively by technology and processes to enhance customer satisfaction and avoid reputational damage.

IT must have access to tools and storage systems to ensure governance and compliance policies are followed, including mechanisms that expose relevant data. This also means commercially sensitive data, personal identifiable information (PII), confidential data, medical information, and more.

Infrastructure must allow finding and validating data based on an agreed business glossary and other metrics such as sensitivity and quality. Another vital metric is metadata, which must be generated automatically with the ability for data stewards to vet, qualify, and modify it.

## 6. Data Mobility

Efficient data mobility, built on repeatability and automation, optimizes cost and value. Flexible, repeatable processes enable easy migration of datasets between cloud providers and on-premises to support business needs and to minimize lock-in to any particular cloud vendor. But throughout, the ongoing concurrency of datasets must be maintained automatically. What's more, policy, and usage-based tiering of data to optimize storage must be automated and nondisruptive to any consumers of that data. Tiering must also support governance requirements of privacy, sovereignty, confidentiality, and more.

## 7. Technical Debt Reduction

No-value data and "dark data" that is redundant or superfluous to business needs should be identified by ongoing processes. This supports remediation decisions of leave, move, archive, and delete as part of an automated information lifecycle management process. Legacy storage systems that are freed up by the information lifecycle must be de-commissioned to reduce ongoing costs. And, of course, the infrastructure should accommodate nondisruptive refreshes of technology. All of this reduces technical debt, a key component of total cost of ownership (TCO).

# Your Ideal Dynamic Data Migration Solution Is Ready: Pentaho

The Pentaho Intelligent DataOps Platform optimizes data modernization and cloud migration while providing flexibility and choice through intelligent automation. It consists of three primary modules: Data Catalog, Data Integration and Analytics, and Data Storage Optimizer.

## Pentaho Data Catalog

Enable analytic innovation with trusted data through a self-service metadata catalog for profiling, discovery, privacy, and search. It profiles data and automates the generation of metadata to perform at enterprise scale, spanning hybrid and multi-cloud.

It establishes flexible, customizable rules for data quality and defines rules to help identify sensitive or personal information, to flag "cold" data that has not been accessed for some time, and to discover and flag duplicate copies of data.

## Pentaho Data Integration and Analytics

Ensure your pipelines keep pace with business growth by embedding powerful integration and analytics into everyday workflows. You can streamline data management from onboarding to analytics with a proven no-code, drag-and-drop interface for developing data pipelines for connecting, copying, transforming, enriching, and blending data.

It creates re-usable data pipelines and accelerates connections and interactions with old and new data sources through templated, drag and drop "steps" to develop no-code pipelines quickly. To mesh well with your existing environment, Data Integration and Analytics includes steps for augmented interactions with Snowflake, MongoDB, Salesforce, SAP, Hadoop, and many more. It automates processes using capabilities such as conditional logic, Metadata Injection and Streamlined Data Refinery.

The module also integrates with Hitachi infrastructure solutions such as Hitachi Content Platform and operationalizes data science for Machine Learning Orchestration.

## Pentaho Data Storage Optimizer

Reduce infrastructure costs with an intelligent data tiering solution that uses object storage for seamless application access. It identifies "cold" data and background processes to automatically "stub" cold data and tier it to lower cost object storage.

## Case Study: Banking

A large regional bank struggled with hundreds of data sources, often of low quality. They used Pentaho to unify data across a hybrid cloud of on-premises systems and Snowflake Data Cloud.



### Data discovery

**8,200** person hours saved.



### Data classification

**900** person days per year saved.



### Data search

**\$11M** annual savings.



### Storage optimization

**22%** cost savings.

## Dynamic Data Migration Optimizes Your Hybrid Cloud

Data movement is an ongoing process that must be automated and managed. It's not a one-time "lift and shift" data migration activity. Data needs to be maintained and remain fluid so that it is available for analytic processing wherever and whenever it is needed. Hybrid cloud migration solutions must – and can – enable ongoing intelligent data mobility, putting the right data in the right place at the right cost.

### ABOUT HITACHI VANTARA

Hitachi Vantara, a wholly-owned subsidiary of Hitachi Ltd., delivers the intelligent data platforms, infrastructure systems, and digital expertise that supports more than 80% of the Fortune 100. To learn how Hitachi Vantara turns businesses from data-rich to data-driven through agile digital processes, products, and experiences, visit [hitachivantara.com](https://hitachivantara.com).

[Learn More](#) →

Contact Hitachi Vantara for more information about dynamic data migration for hybrid cloud.



**Corporate Headquarters**  
2535 Augustine Drive  
Santa Clara, CA 95054 USA  
[hitachivantara.com](https://hitachivantara.com) | [community.hitachivantara.com](https://community.hitachivantara.com)

**Contact Information**  
USA: 1-800-446-0744  
Global: 1-858-547-4526  
[hitachivantara.com/contact](https://hitachivantara.com/contact)

© Hitachi Vantara LLC 2023. All Rights Reserved. HITACHI and Pentaho are trademarks or registered trademarks of Hitachi, Ltd. All other trademarks, service marks and company names are properties of their respective owners.

HV-BTD-WP-How-To-Optimize-Your-Hybrid-Cloud-3Jul23-A