HITACHI
Inspire the Next

Achieve Digital Transformation with Data Analytics

Pentaho Data Platform

Hitachi Vantara Federal
Date: August 2023

Table of
Contents

Federal agencies now have more data than ever. Today, the challenge for many organizations is how to leverage cutting-edge technologies to transform all their data into strategic assets providing real-time intelligence.
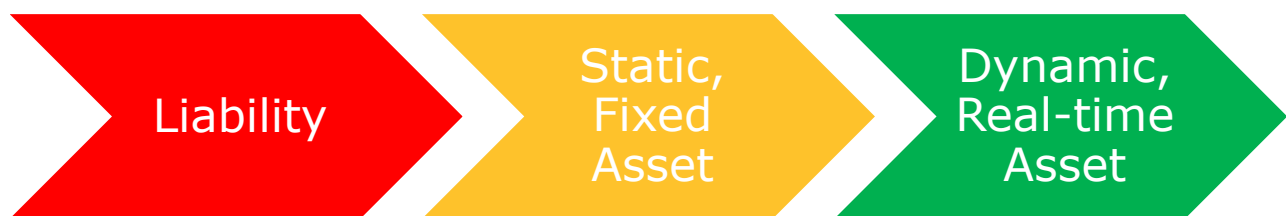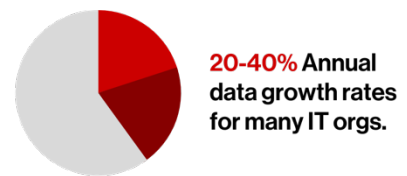
Digital Transformation

As recent as 2005, many organizations viewed data as a liability – content to be archived and kept safe for compliance without any return on investment. With advancements in computation, memory, and algorithms, the ability to rapidly convert data to information to aid decision-making processes led to a change in attitude. For the first time, data began to be viewed as an asset rather than a liability.

The ability to turn this data into a practical asset is even more important because businesses are capturing, storing, and analyzing more data than ever before. This presents new problems to overcome, as the data growth rates are exploding, and much of the new data produced is complex and unstructured.

20-40% Annual data growth rates for many IT orgs.

Even if organizations can utilize business intelligence and incorporate data analytics into their digital strategy, they can still be handicapped by their inability to leverage AI and machine learning (ML), extract real time analytics, and drive operational efficiency. In other words, they are only able to use data as a fixed asset rather than a current asset.

70-80% of that data will be unstructured.

Realizing the power of data as a real-time asset to continuously improve processes is called _Digital Transformation_.

Liability	Static, Fixed Asset	Dynamic, Real-time Asset

Value of Data as Organizations Adopt Next-Gen Data Strategies that Enable Digital Transformation

How can organizations transform their data from a static, unused asset into a powerful, real-time analysis enabler?

This is the challenge of **Digital Transformation** that Hitachi Vantara Federal addresses.

HITACHI
Inspire the Next

The Four Pillars of Digital Transformation

A successful DataOps-driven data analytics strategy relies on the people, processes, and technologies of an organization. All three must work together in harmony to develop, enable, and maintain a successful and continuously evolving Digital Transformation strategy.

At Hitachi Vantara Federal, we see successful Digital Transformation strategies rely on four core pillars: **Data Governance, Data Catalog, Metadata Management, and Data Storage**.

Pillar	Features	Benefits	Hitachi Solution
Data Governance	• Shared Language • Universality • Repeatability • Security • Processes	• Encourages collaborative environment • Builds a community of users • Enables data federation • Standardizes language and process • Enables visibility to available resources	*Professional Services*
Data Catalog	• Lineage • Curation • Federation • Ease of Access and Use	• Provides data provenance to expeditiously track data sources and changes to data • Federated information access • 360-degree view for Chief Data Officers / Data Stewards / CXOs to address changing needs of organization • AI/ML-based solutions reduce manual effort extensively and increase accuracy with usage over time	*Pentaho Platform*
Metadata Management	• Ingestion • Support for Structured and Unstructured Data • Automation (AI/ML)	• Enriches complex data with contextual and summary information • Enables usage of complex non-tabular data • Frees time for data scientists • Makes disparate types of data usable as part of a data science solution	*Pentaho Platform & Hitachi Content Intelligence (HCI)*
Data Storage	• Object Storage • Flexibility • Backup • Retention and Versioning • Immutable • Security	• Provides the backbone of a strong data science solution • Accelerates policy implementations around Archival, Replication, Retention, Disaster Recovery • Strengthens regulatory compliance and associated business and data protection requirements • Roll back data if necessary • Protects against viruses and ransomware	*Hitachi Content Platform*

Data Analytics Solutions from Hitachi

Hitachi Vantara Federal offers an entire suite of data analytics products and services to bring innovation to federal agencies' Digital Transformation journeys:

Hitachi Vantara Federal is a data-driven company. Storing, processing, and analyzing data is inherent in everything we do. Hitachi is also at the cutting edge of data analytics, using artificial intelligence (AI) and machine learning (ML) and integrating these technologies throughout our broad corporate product portfolio.
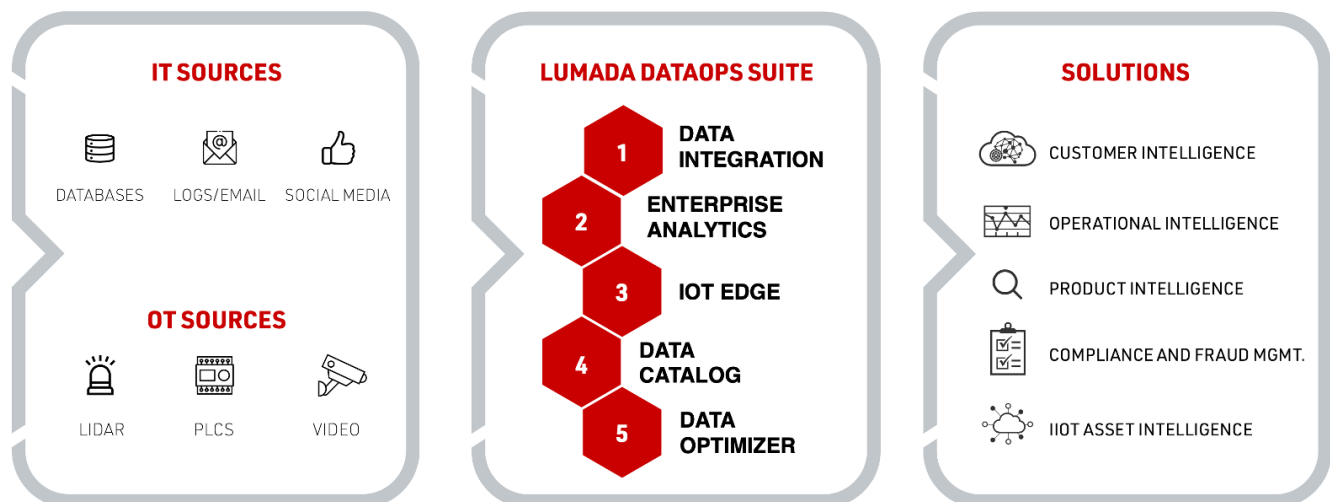
- **Pentaho Data Catalog** is a self-service platform. It utilizes machine learning to perform automatic discovery and tagging of data, including semantic inference, and can refine tags based off analyst feedback then use the same feedback to refine future tagging suggestions.

- **Hitachi Content Intelligence (HCI)** can tie into existing structured and unstructured data sources and perform transformations to enhance the searchability and compliance of the data.

- **Pentaho Data Analyzer** enables organizations to access, prepare, and analyze all data from any source, in any environment, between data managers and consumers.

- **Hitachi Content Platform (HCP)** is an object storage solution designed to store, share, sync, protect, preserve, analyze, and retrieve file data from a single system. HCP offers mobile access, archiving, backup, custom metadata, and easy scaling.

- **Hitachi Professional Services (PS)** provide organizations with dedicated expertise around Digital Transformation and Hitachi analytic products. Our experts are dedicated to working with our clients to further their vision and Mission by understanding their data environment. We develop next-generation strategies, plans, policies, and implementation frameworks.

IT SOURCES

DATABASES   LOGS/EMAIL   SOCIAL MEDIA

OT SOURCES

LIDAR   PLCS   VIDEO

LUMADA DATAOPS SUITE

1 DATA INTEGRATION
2 ENTERPRISE ANALYTICS
3 IOT EDGE
4 DATA CATALOG
5 DATA OPTIMIZER

SOLUTIONS

CUSTOMER INTELLIGENCE

OPERATIONAL INTELLIGENCE

PRODUCT INTELLIGENCE

COMPLIANCE AND FRAUD MGMT.

IIOT ASSET INTELLIGENCE

Figure 1. Raw Data is ingested into the system and enhanced with AI/ML for optimized analytics.

As depicted in ***Figure 1***, this solution allows data scientists, analysts, and users to reduce the time it takes to process captured data by accelerating metadata discovery and categorizing data. It permits data curators and stewards manage sensitive data. It also enables IT administrators to eliminate unintended redundant data.

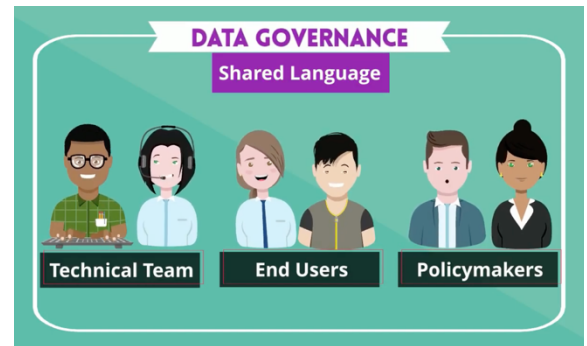A Digital Transformation solution from Hitachi Vantara Federal:

- **Establishes a unified comprehensive data store** – Hitachi Digital Transformation solutions simplify preparing and blending all data types and provide a broad spectrum of self-service analytics including dashboards, reports, data visualization, discovery, and predictive analytics.
- **Provides Governed data delivery** – The ability to blend trusted and timely data to power analytics at scale for all users in all environments.
- **Is easy to embed** – The platform uses a multi-tenant architecture, making it easy to embed analytics into any website, workflow application or deployment model, including public cloud, on premise, mobile and hybrid models.
- **Has the power to access, prepare and blend all data** – The data integration engine delivers accurate, "analytics ready" data to end users from any data source. Visual drag and drop tools eliminate coding and complexity, easily putting big data and traditional data sources at the fingertips of analysts.
- **Leverages Kubernetes as a Service** – An open hybrid and multi-cloud management service for enterprise-grade Kubernetes implementations.
- **Enables native integration to all big data sources** – Regardless of data source, analytics requirements, or deployment environment, the Hitachi's Digital Transformation solutions create "future-proofed" access to disparate data sources.
- **Provides object storage** – Object storage can significantly enhance data management and governance while drastically lowering storage costs. HCP is not only a modern object store but also a data management platform to enrich data with the context necessary to secure, protect, control, and analyze it.

As federal agencies perform their core missions and provide services to the nation, the data infrastructure and analytics solutions supporting these missions must remain readily available and scalable to meet growing and future demands. To modernize their data-centric solutions, agencies will need an industry partner that brings cutting edge technology, world class services, and experience in effectively managing data environments. Hitachi Vantara Federal is that partner.

Hitachi Vantara Federal helps data-driven leaders find and use the value in their data, to innovate intelligently and reach outcomes that matter. We combine technology, intellectual property, and industry knowledge to deliver data-management solutions to help the U.S. Government innovate and leverage data to create new efficiencies.

Notes

- Data Governance (Policy, holistic, all products combined)
  - Shared language
  - Useful (and useable) for everyone in org
    - Builds community of users
  - People, process, and technology
  - Creates repeatable ops
  - Supports strong security (permissions)
  - Data preparation, ingest, and transformation
- Data Catalog (Pentaho)
  - Lineage
    - Provenance. Easily track source and changes to data
  - Curated
    - Data stewards/owners/CDOs can address changing needs of the organization and answer questions from users. Data does not just sit there "as a mystery" – adds order to the information overload.
    - Talk to an actual human (human augmentation) – human insight into data
  - Federation – single point of information availability
    - Consolidated view of data
    - Encyclopedia of all data assets
  - Ease of access/user friendly/visual UI
    - Useable by people with any level of technical expertise
    - Collaborative
    - Searchable
- Metadata Management (Pentaho, HCI)
  - Ingestion with HCI
  - Structured and unstructured data
    - Supports virtually all data types
    - Can organize disorganized and messy data
  - Automation (AI/ML)
    - Frees time for the data scientists
    - "Learns" over time
    - Can be just as accurate as human analysis
  - Trustworthiness of data/data sources
  - Enriching and adding value to existing data
- Data storage (Object Storage, HCP)
  - Backbone of data solution
    - Feeds into all aspects of solution
  - Backups (replication, DR, search SNMP)
    - Addresses compliance and other business requirements
  - Retention and versioning
    - Long term retention, low cost
    - Versioning so you can roll back data if necessary
  - Immutability – protects against viruses and ransomware

HITACHI
Inspire the Next

- o Security/Permissions/RBAC
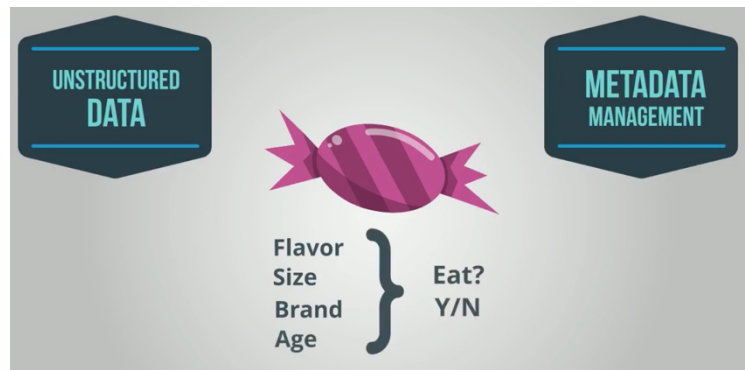 - ▪ Supports governance and compliance

The first step is the creation of a shared language between the technical team, end users, and policymakers. Mapping the digital lifecycle from these multiple perspectives requires a common framework, which we call data governance.
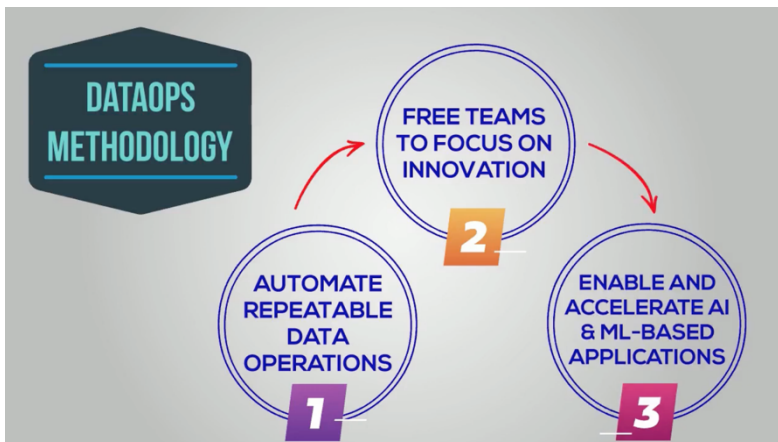

Implementation of certain tools can also help on a journey towards digital transformation. Data Catalogues provide a medium for curation of data and provides the basis for a shared language. A Data Lineage provides much needed provenance and transparency around your data, improving trust in data and empowering confidence when utilizing data for decision-making.

The challenge of leveraging unstructured data also needs to be addressed. Metadata management serves a key role in dealing with non-tabular data. Think of metadata like a candy wrapper. When you look at a candy wrapper, you should be able to tell the flavor, size, brand, and age of the candy to determine whether it is worth eating or not. Metadata provides this insight into unstructured data, allowing you to sort, manipulate, and leverage unstructured datasets which are relevant to your analysis.

Finally, self-service analytics can play a key role in empowering end users to leverage data without involving IT. Allowing decision-makers access to curated datasets or realtime analytics dashboards can create a data-driven culture across all department in your organization, and contribute to engaging with data as a current asset rather than a fixed asset.

Another mechanism for the transformation from data as a fixed to a current asset is called DataOps. DataOps recognizes data as assets for business initiatives and empowers individuals across the organization to leverage this data. It encourages participation at all levels in a feedback loop that strives for continuous improvement in the processes that make up the overall data architecture.

DataOps empowers data consumers as citizen data scientists, involving them in the process of data science as much any technical person in your organization who may have a better understanding of your data. It brings them into the fold under a shared language to leverage data as a tool the entire team can get benefit from.

A key component to a DataOps methodology is the automation of repeatable data operations. This frees up the teams responsible for your data operations to focus on more specific questions and problems requiring their attention rather than manipulating data, which can be automated and streamlined. This in turn enables and accelerated AI/ML applications.
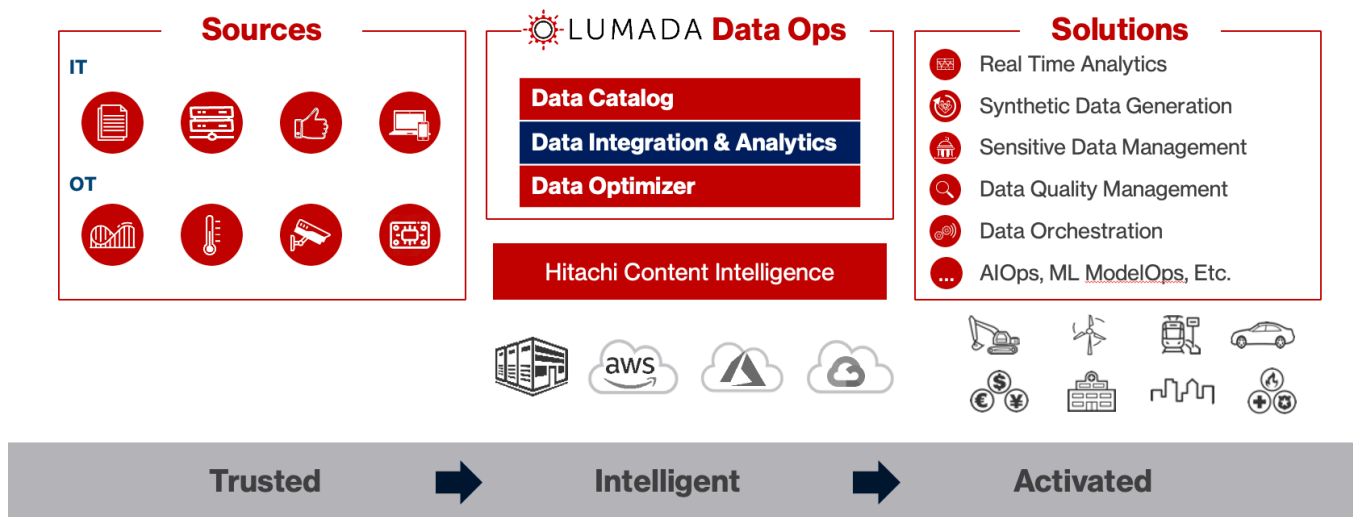

Figure 2. Pentaho DataOps features end-to-end data governance.

About Hitachi Vantara Federal

Hitachi Vantara Federal is the most reliable data solutions provider with the most complete set of hybrid cloud and data fabric solutions in the Federal IT market . We're a collaborative, full-service company empowering data-driven insight with a deep bench of integrated partners – advancing Federal customer missions regardless of their data maturity levels.

We're more than just a leader in digital infrastructure. We're constantly evolving our data management, analytics, and Internet of Things (IoT) solutions to service Federal leaders' mission needs today, and in the future. All of our Federal customers get the platinum treatment as a standard offering. From civilian agencies to the Department of Defense (DoD) and Intelligence Community (IC), we're known for our agility, dependability, and excellence.

As the trusted leader in mission-centric data solutions for the Federal government, Hitachi Vantara Federal is dedicated to providing the solutions and support our customers need today, and in the future. We power digital transformation from the core to the edge, offering a full-service experience for integrated digital solutions spanning the entire data fabric. Balancing a century of tradition with constant innovation, we stand for agility, dependability, and excellence –guaranteed.