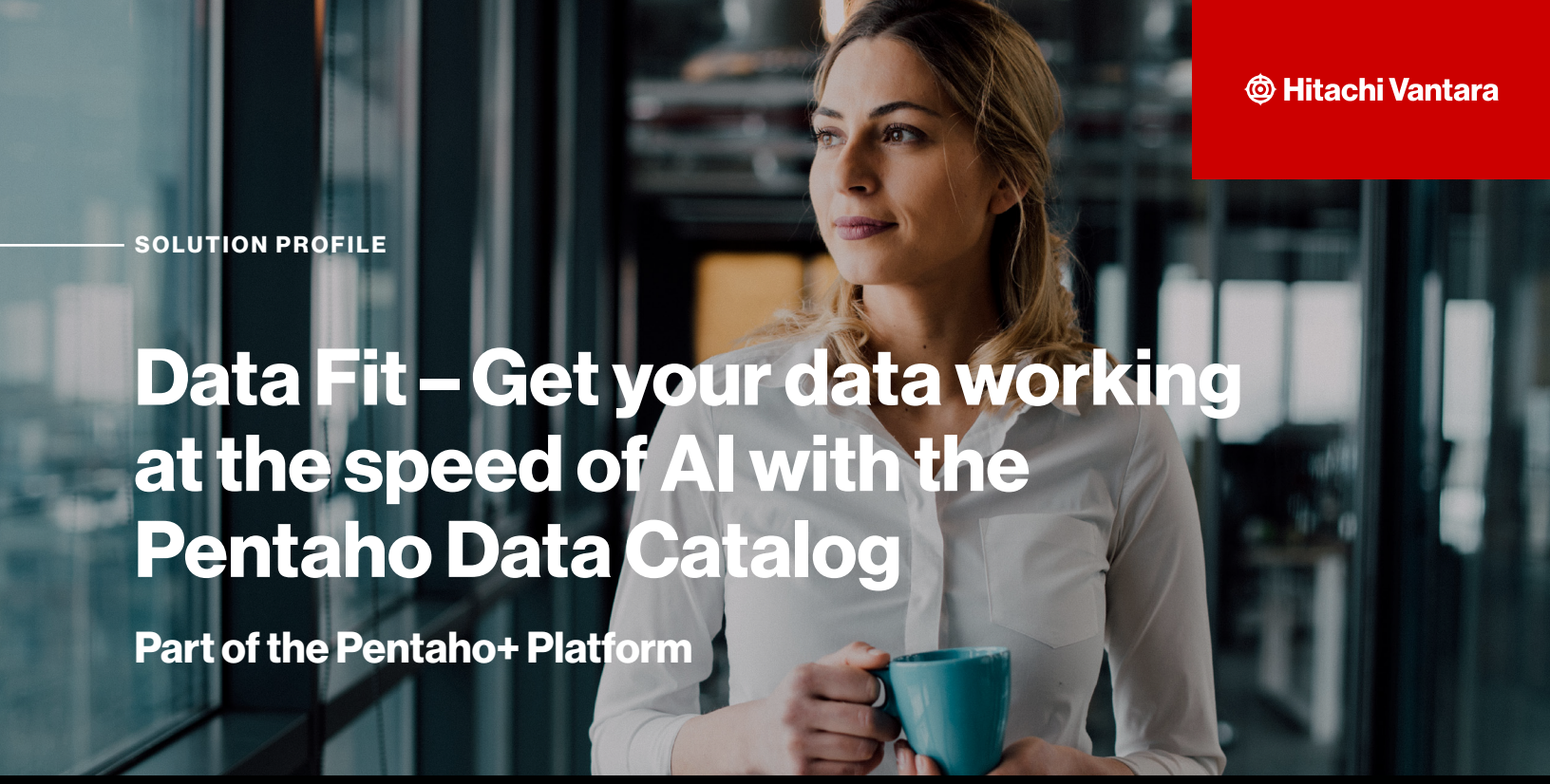


Hitachi Vantara

SOLUTION PROFILE

Data Fit – Get your data working at the speed of AI with the Pentaho Data Catalog

Part of the Pentaho+ Platform

Know Your Data. Trust Your Data. Build Your Data Products

Understand Data

Automatically find, analyze and tag structured and unstructured data across. Contextualize with business glossary and governance policies.

Activate Metadata

Observe data to define measures for data over time. Monitor metadata and act upon changes, trends and anomalies in data. Leverage event-driven architecture to apply remediation before any impact is noticed downstream.

Data Fit for Business

Trusted and high-quality data is made available to decision makers with a shopping experience. Graph view helps see relationships across data elements to further improve contextualization. User feedback helps democratize the process of data curation.

Optimization and Compliance

Measure data utilization, value and aging to make optimized storage decisions. Automated classification and characterization enable the application of life cycle, governance and access policies.

Pentaho Data Catalog Helps Business Users Search and Understand Structured and Unstructured Data Everywhere

A modern organization must be data fit. As data grows, so does the need and cost of maintaining the data in business-ready shape. To leverage data for business decisions and enable AI, data must be trusted, high quality and seamlessly available to the data users. Now more than ever, there is a need to discover content across structured and unstructured, on-prem and cloud. Organizations must monitor their data to spot trends and anomalies and maintain data hygiene at the speed of data growth. Policies for governing, life cycle and quality need to be enforced to ensure appropriate high-quality data is available to consumers. Data users and models can easily find and use data via the data catalog, a necessity for modern data-driven organizations.

Pentaho Data Catalog (PDC) rapidly ingests, profiles and curates structured and unstructured data with both automation and machine learning. Fingerprinting of data and metadata rules are used to contextualize data with the language of the business documented in the business glossary. Its policy manager enables the implementation of governance and security policies.

A powerful rules engine helps determine quality, sensitivity and usage patterns. Activate your metadata by leveraging monitoring and notification capabilities in the product. Construct a relationship graph across business entities and terms to add semantic understanding to data.

Data fingerprints are analyzed to determine potential duplicates, copies and similarities across data stores to assess data movement, optimization and mastering needs. Data lineage support for Open Lineage provides the ability to track data as it flows through your organization, building trust and enabling a left shift of data quality and remediation activities.

A powerful observability stack captures popular assets, popular searches and trends, helping stewardship organizations focus their energy on the right data. Collaborate in the context of your data and business vocabulary to capture tribal knowledge, recommendations and user's perspective on data assets.

Hitachi Vantara Pentaho Data Catalog

Manage Data from Anywhere

Manage structured and unstructured data connected to multiple and disparate data stores, such as RDBMS Systems, File Systems (NFS, HDFS, SMB) and Object Stores.

Make Better Business Decisions with Better Data

Governed data with more depth and accuracy leveraging machine learning from customizable NLP classifiers and semantic technology that puts IT and OT data into context.

Flexibility through Modular Extensibility

Choose the applications you need and build from there – with apps for privacy, security and governance.

Enterprise-Class Security

Get enterprise-class security from roles-based access controls (RBAC), Password Vault Support, Minimum privileges, multifactor authentication, secure cloud deployments and eliminated data deduplication.

Cloud Ready

Deploy when you want in the cloud, on-premise, or across a hybrid environment – agentless and automated.

Enterprise Scale

Modern architecture is designed to scale with your data at petabyte scale – without affecting business or systems.

Open Ecosystem

API-powered integrations with critical components and applications across your data stack, such as NetApp, SAP Hana, S3 and SQL views for interoperability and many more.

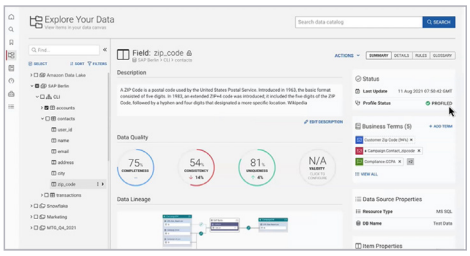


Fig.1 At-a-Glance Data Discovery & Metadata Catalog

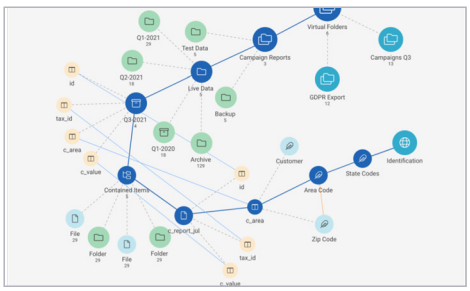
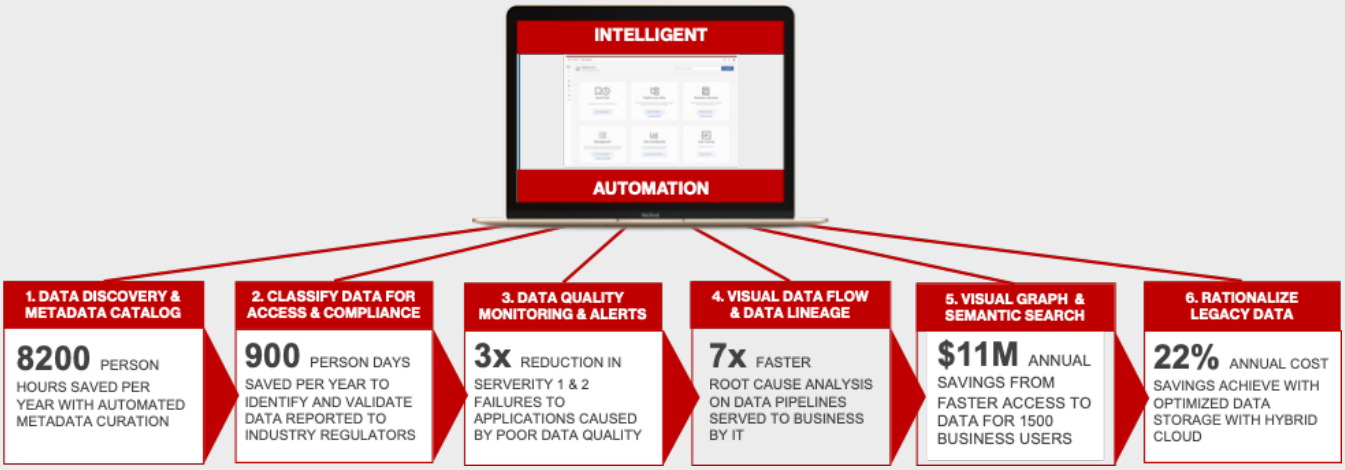


Fig.2 At-a-Glance Visual Galaxy Graph & Semantic Search

> "Pentaho Data Catalog allows our staff to find the data they need so they can spend less time on data and more on water. Thank you for making your tool so easy to use!"

Lisa Williams, CDMP
Manager and Founder, Office of Data Management at Arizona Department of Water

Hitachi Vantara Pentaho Data Catalog



Source: Hitachi Vantara (2022)
Based on Hitachi Vantara's use case across a typical US financial institution fully implemented running on hybrid cloud (AWS, Snowflake, Oracle, SalesForce).

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.

Discover Pentaho Data Catalog →

Let us help with your data discovery and classification today.

Hitachi Vantara
A Hitachi Group Company

Corporate Headquarters
2535 Augustine Drive
Santa Clara, CA 95054 USA
hitachivantara.com | community.hitachivantara.com

Contact Information
USA: 1-800-446-0744
Global: 1-858-547-4526
hitachivantara.com/contact

HV-BTD-SP-Pentaho-Data-Catalog-20Oct23-A