

Pentaho Data Catalog

Automated Discovery, Classification, and Optimization.
Know Your Data. Trust Your Data. Build Your Data Products.

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

Pentaho Data Catalog changes how your business discovers and manages data, ensuring seamless scalability across all data types and volumes. Simplify data observability with a unified business glossary and advanced metadata management to enhance lineage, trust, and quality. Embrace a smarter way to handle your data, making it easier to search, validate, and derive insights, all tailored to your unique business needs.

Get Faster and More Meaningful Data to Users

A modern organization must be data-fit to support core operations and AI use cases. As data grows, so does complexity and cost of maintaining data in business-ready shape. To confidently leverage data for business decisions and enable AI efforts, 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 lineage, 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 a modern data catalog, a foundation for modern data-driven organizations.

Powerful Business Glossary Using Machine Learning

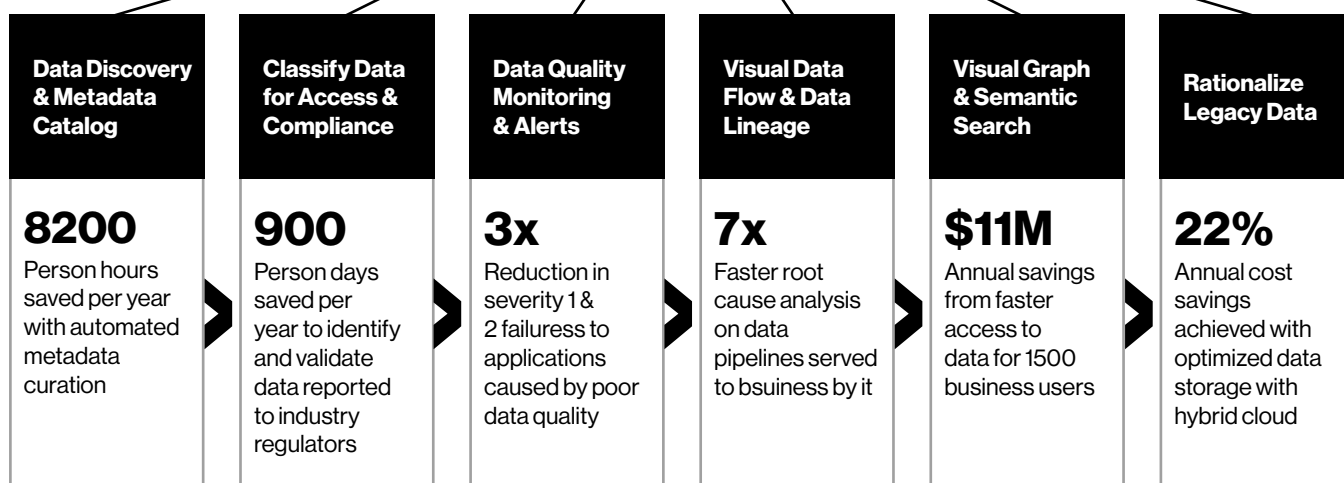
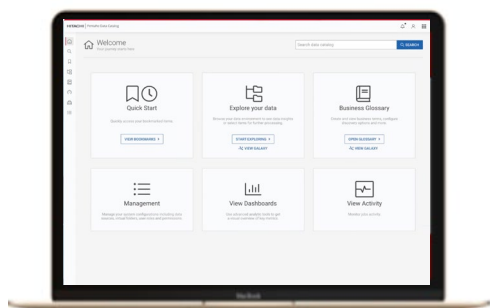
Pentaho Data Catalog (PDC) rapidly ingests, profiles and curates structured and unstructured data with both automation and machine learning. Fingerprinting of data and metadata rules are used to contextualize data with the language of the business documented in the business glossary. Its policy manager enables the automation and scalable 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 with lineage and terms to add semantic understanding to data.

“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

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.



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 AI

Trusted and high-quality data is made available to decision makers with a shopping experience. Catalog users deliver data to the desired destination with a No Code Data Pipe build experience.

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, regulatory compliance and access policies.

Manage All Your Data

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

Governance for the Enterprise

Bring business and governance vocabulary, policies and standards and their application to data and applications and reports. Determine lineage and usage.

Feature Rich

Bring in reference data, usage characteristics, view semantic relationships, customize properties.

Enterprise Scale

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

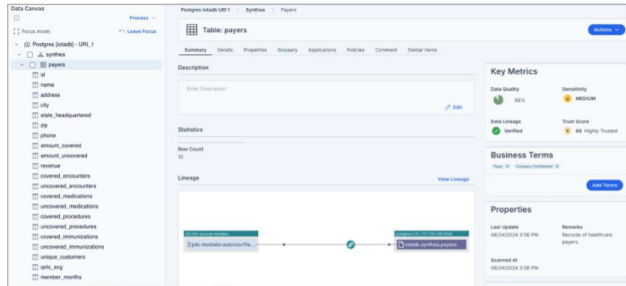


Fig.1 At-a-Glance Data Discovery & Metadata Catalog

Flexibility through Modular Extensibility

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

Make Better Business Decisions with Better Data

View data with its full context through Pentaho's Galaxy View. Stay informed of data's characteristics, measuring and monitoring for accuracy, sensitivity, freshness to drive better business decisions.

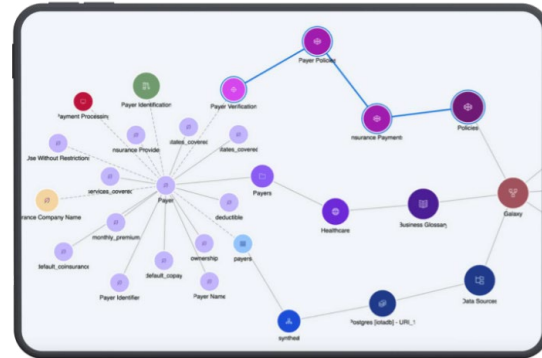


Fig.2 At-a-Glance Visual Galaxy Graph & Semantic Search

Discover Pentaho Data Catalog →

Pentaho Data Catalog Technical Specifications and Compatible Components [Learn More](#)

About Pentaho

Trusted by over 70% of the Fortune 100, Pentaho's end-to-end data integration and intelligence platform enables the foundational and AI-grade data fitness that drives faster and more confident decision-making.

Pentaho

pentaho.com

© Pentaho® is a registered trademark of Pentaho Canada Inc. All other trademarks, service marks and company names are properties of their respective owners.

P-BTD-DS-Pentaho-Data-Catalog-25.June26