

# cloudera<sup>®</sup> CONNECT



**PENTAHO**

**INDUSTRY**

Data Integration  
Business Analytics

**WEBSITE**

pentaho.com

**SOLUTION HIGHLIGHTS**

- > Visual interfaces for data ingest, data manipulation, data integration, task orchestration, visualization, exploration and prediction that integrate directly with Cloudera's EDH.
- > Visual development interfaces enable up to 10x boost in productivity.
- > Highly interactive visualizations, advanced and predictive analytics and reporting, including production, ad-hoc and operational

**ENTERPRISE DATA HUB (EDH)**

- > Cloudera's EDH, built on Apache Hadoop, provides one place to store all data as long as required in its original fidelity, allowing enterprises to be truly information driven with data at the center of business.
- > Run a variety of workloads—including batch processing, interactive SQL, enterprise search and advanced analytics—with enterprise-grade security, governance, data protection, and management.

## Cloudera and Pentaho

### Accelerating Time to Big Data Value

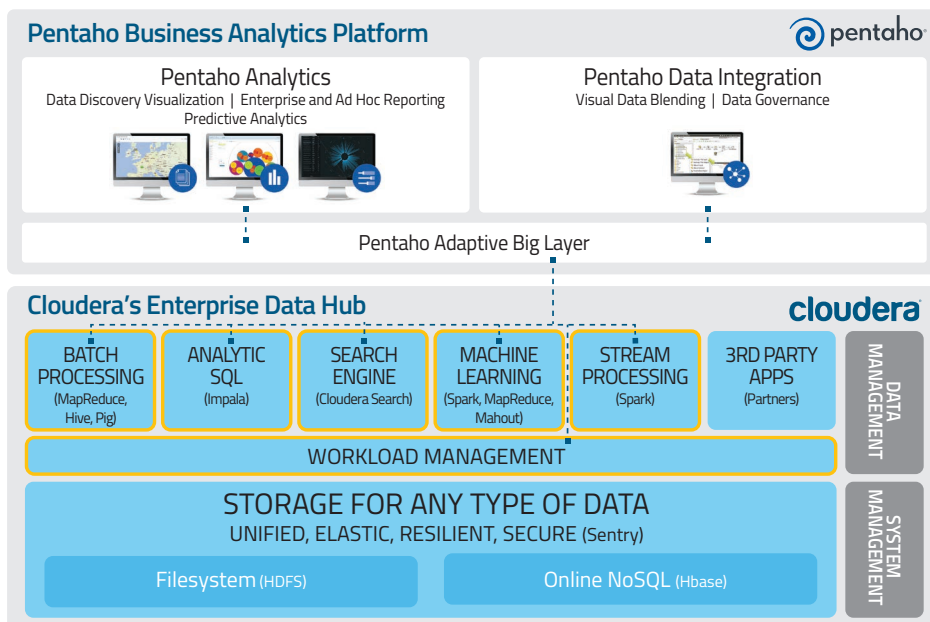
Enterprises today collect and generate more data than ever before. Business users ask more sophisticated questions and explore data at a level of detail that can tax existing data management systems. To win in a competitive marketplace—obtaining insights to improve products, understanding customer sentiment, optimizing marketing promotions, and detecting and preventing fraud—organizations must combine a variety of data from many sources to extract real, actionable insight.

### Empowering the Data-Driven Enterprise

With Pentaho and Cloudera, organizations get a comprehensive view of data across their business. Cloudera's enterprise data hub (EDH), powered by Apache Hadoop, delivers a complete data management platform—one place to store, process, and analyze data in full fidelity that complements existing storage strategies and analytic infrastructure investments. The EDH offers the flexibility to run a variety of enterprise Big Data workloads, including batch processing, interactive SQL querying, enterprise search, and advanced analytics, together with the robust security, governance, data protection and management that enterprises require. Machine generated data, web logs, and social media data can be leveraged in concert with operational data, making disparate data silos a thing of the past.

With an EDH, leading organizations are changing the way they think, transforming data from an expense to an asset. Pentaho's comprehensive analytics platform, data integration, and a spectrum of data visualization and analysis capabilities can bring this data to life. Deeply integrated with technologies such as Impala, Search and YARN, Pentaho's Analytics Platform is optimized for Cloudera's enterprise data hub.

### Joint Solution Integration



Pentaho Advanced Integration

## Benefits of Cludera

- > **Powerful** – Store, process, and analyze all your data in one location to drive competitive advantage.
- > **Efficient** – Unify compute and data to improve operational efficiency. Scale linearly on industry-standard hardware.
- > **Open** – Cludera is the world's most popular open source Hadoop distribution. An open platform ensures a choice of tools, including over 200 ISV products that work with Cludera today.
- > **Industry-leading Management and Support** – Easy to deploy and operate with centralized end-to-end administration supporting configuration, monitoring, and issue resolution.

## Benefits of Pentaho

- > **Comprehensive** – Complete analytic offering including enterprise ready data integration and blending, reports, ad hoc visualizations, advanced and predictive analytics.
- > **Modern** – 100% java-based architecture with REST services enables simple systems integration.
- > **Open** – Technology founded on open source principles, supported by Pentaho and thousands of customers.
- > **Future-Proofed** – Modern architecture built to support current business needs as well as future requirements and data sources.



## Solution Benefits

**IT Managers** – Decrease costs and reduce time in data preparation and management

- > Pentaho's Visual MapReduce enables ETL via its intuitive drag-and-drop interface, simplifying and reducing time and complexity of data integration.
- > Infrastructure costs are reduced as ETL workloads previously run on expensive systems can migrate to a Hadoop active archive in a Cludera enterprise data hub to run at very low cost, in parallel, with increased performance and accessibility.

**Developers** – Accelerate big data adoption with familiar tools

- > Pentaho's visual drag-and-drop ETL handles Hadoop data sources in the same way as traditional relational and file-based data sources. Instead of writing Java programs, MapReduce jobs or Pig scripts, Pentaho empowers developers to design and develop big data jobs using visual tools—resulting in greater team productivity and efficiency.
- > With Instaview for Cludera Impala, the world's first open source MPP SQL query engine expressly built for Hadoop, large volumes of disparate data can be analyzed faster than was previously possible with Hive alone.

**Business Users** – Uncover previously unseen data patterns

- > With Pentaho's Analytics Platform users discover deep insight into data stored in the entire enterprise data hub to quickly spot patterns and identify insights.
- > Pentaho provides highly interactive visualizations, advanced and predictive analytics. Self-service and production reporting including production and dashboards complete the analytics capabilities.

## About Cludera

Cludera is revolutionizing enterprise data management by offering the first unified Platform for Big Data: The Enterprise Data Hub. Cludera offers enterprises one place to store, process and analyze all their data, empowering them to extend the value of existing investments while enabling fundamental new ways to derive value from their data. Founded in 2008, Cludera was the first and is still today the leading provider and supporter of Hadoop for the enterprise. Cludera also offers software for business critical data challenges including storage, access, management, analysis, security and search. With over 40,000 individuals trained, Cludera is a leading educator of data professionals, offering the industry's broadest array of Hadoop training and certification programs. Cludera works with over 900 hardware, software and services partners to meet customers' big data goals. Leading organizations in every industry run Cludera in production, including finance, telecommunications, retail, internet, utilities, oil and gas, healthcare, biopharmaceuticals, networking and media, plus top public sector organizations globally. [www.cludera.com](http://www.cludera.com)

## About Pentaho

Pentaho is delivering the future of business analytics. Pentaho's open source heritage drives our continued innovation in a modern, integrated, embeddable platform built for the future of analytics, including diverse and big data requirements. Powerful business analytics are made easy with Pentaho's cost-effective suite for data access, visualization, integration, analysis and mining. Pentaho provides analytics to over 1,200 customers across industry verticals such as healthcare, retail, gaming, finance, government and software (SaaS). For a free evaluation, download Pentaho Business Analytics at [pentaho.com/testdrive](http://pentaho.com/testdrive).