Pentaho Big Data Integration and Analytics

Within one platform, Pentaho provides big data tools to extract, prepare, and blend your data, plus the visual analytics that will change the way you run your business. From Hadoop and Spark to NoSQL and analytic databases, Pentaho allows you to turn big data into big insights.

Broad and Adaptive Big Data Integration

Deep native connections and an adaptive big data layer accelerate access to the latest versions and capabilities of popular big data stores.

- Graphical extract-transform-load (ETL) designer to simplify the creation and orchestration of big data pipelines
- Ability to access data once and then process, combine, and consume it anywhere, providing insulation from changes in the big data ecosystem
- Support for the latest Hadoop distributions, NoSQL stores, and analytic databases, as well as Spark and Kafka
- Automation to rapidly accelerate the ingestion and onboarding of hundreds or thousands of diverse and changing data sources into Hadoop
- Visual design tools that empower faster development in Spark or MapReduce versus hand coding or other design tools
- Seamless switching or combining of data processing engines with in-cluster execution, load balancing and container support to maximize existing processing capacity

PENTAO INTEGRATES WITH A FULL SPECTRUM OF BIG DATA SOURCES

Visual design tools that run in-cluster
Power a Complete Analytic Data Pipeline

A tightly coupled data integration and business analytics platform accelerates the realization of value from blended big data.

• Full array of analytics: from data access and integration to data visualization and predictive analytics
• Empowers users to architect big data blends at the source and stream them directly for more complete and accurate analytics in the form of interactive visualizations, dashboards, and reports.
• Integration of advanced analytic models from R, Python, Spark MLlib and Weka to operationalize predictive analytics while reducing data prep time
• Open, standards based architecture makes it easy to integrate with or extend existing infrastructure

An Expert Partner with Proven Implementations

Technology only gets you so far. People, experience, and best practices are the most important drivers for project success with big data technologies.

• Big Data Blueprint design patterns that have allowed customers to optimize infrastructure, refine big data for self-service analytics, create a complete view of customers, and monetize analytics as a service
• Successful big data projects across verticals, including fraud detection in financial services with FINRA, infrastructure optimization in hardware with NetApp, and cyber security data integration with BT
• A dedicated team of big data services experts with offerings for every phase of the implementation lifecycle, including training workshops, technical account management, solution delivery, and engineering services

SAMPLE END-TO-END BIG DATA ARCHITECTURE WITH PENTAHO

Other Data Sources
Web
Social Media
Transactions – Batch and Real-time
Sensor Data
Internal Data

Self Service Data Request, Architected Data on Demand
Pentaho Analyzer
Pentaho Reports
Pentaho Data Integration
Hadoop / Spark Cluster
Pentaho Data Integration
Analytical Database

© 2017 Pentaho Corporation. All rights reserved. 17-085v6