Pentaho Business Analytics Platform

Ingest, Prepare, Blend and Analyze

DATASHEET

By tightly coupling data integration with business analytics, the Pentaho platform from Hitachi Vantara brings together IT and business users to ingest, prepare, blend and analyze all data that impacts business results. Pentaho's open source heritage drives continued innovation in a modern, unified, flexible analytics platform that helps organizations accelerate their analytics data pipelines.

Pentaho's open, embeddable platform (see Figure 1) supports flexible analytics that both leverage existing data infrastructure and future-proof deployments against tomorrow's inevitable changes. Intuitive data integration and preparation capabilities drastically reduce the hand coding required to bring data together for insight. At the same time, Pentaho Business Analytics provides a spectrum of analytics for all user roles, from visual data analysis for business analysts to tailored dashboards for executives. Pentaho is fast to deploy, easy to use, and purpose-built for the future of analytics.
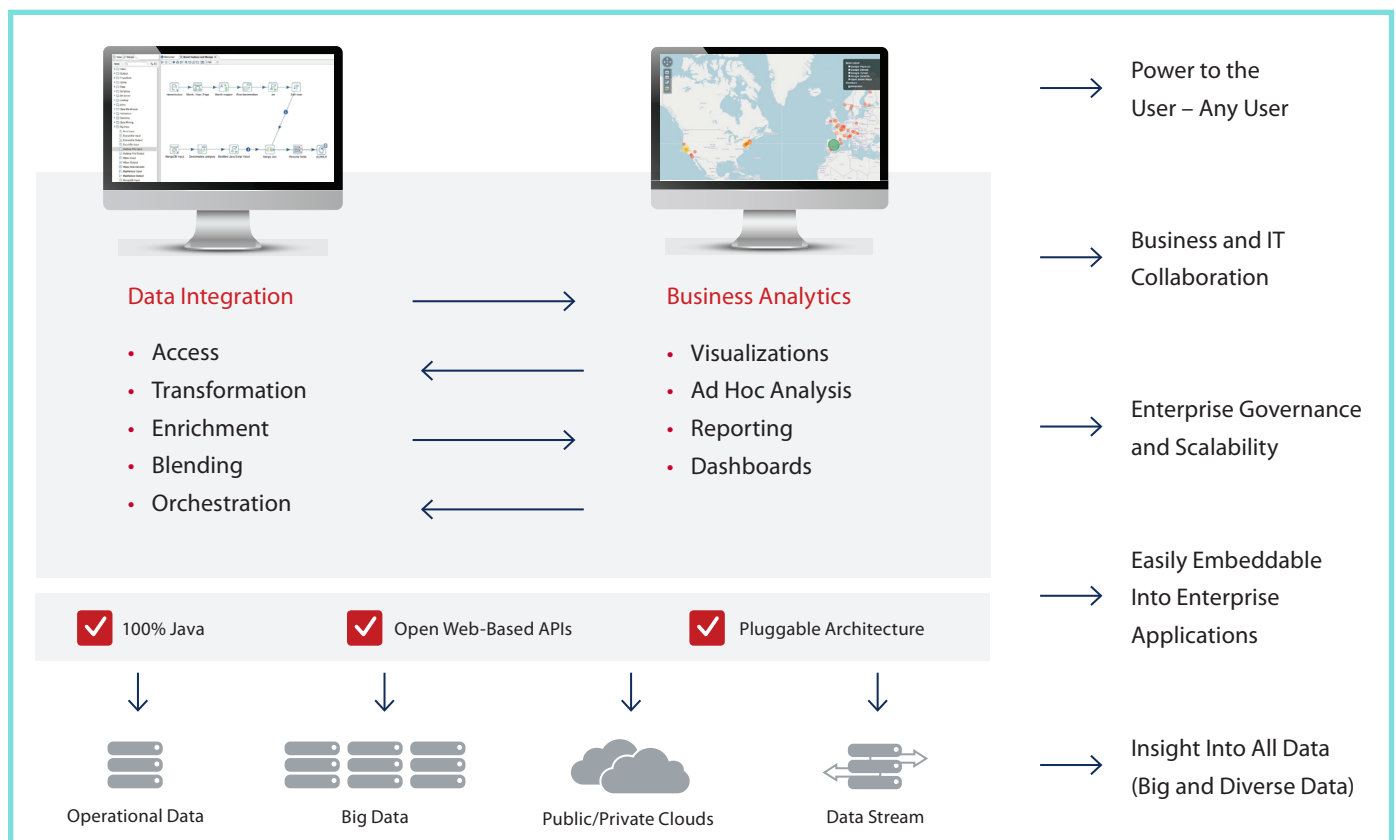

Data Integration

- Access
- Transformation
- Enrichment
- Blending
- Orchestration

Business Analytics

- Visualizations
- Ad Hoc Analysis
- Reporting
- Dashboards

✓ 100% Java   ✓ Open Web-Based APIs   ✓ Pluggable Architecture

Operational Data   Big Data   Public/Private Clouds   Data Stream

Power to the User – Any User

Business and IT Collaboration

Enterprise Governance and Scalability

Easily Embeddable Into Enterprise Applications

Insight Into All Data (Big and Diverse Data)

Figure 1. Pentaho Business Analytics Platform

Data Integration

Organizations face an increasing challenge to manage and extract value from a growing variety and volume of data across their edge-to-cloud infrastructure. With Pentaho Data Integration (PDI), organizations can access data from complex and heterogeneous sources and blend it with existing relational data to produce high-quality, ready-to-analyze information — all without writing a line of code. A rich graphical user interface paired with a powerful multithreaded transformation engine offers high-performance ETL (extract, transform and load) capabilities that cover all data integration needs, including big data ingestion and processing.

Pentaho Data Integration Features

- Intuitive drag-and-drop interface to simplify the creation of analytic data pipelines. (see Figure 2).
- Broad connectivity to virtually any data source either on-premises or in the cloud, including flat files, relational database management systems (RDBMS), big data, object stores, APIs and more.
- Integration with transactional databases, including Oracle, IBM® DB2®, PostgreSQL, MySQL and others.
- Access to data in enterprise applications, including SAP, Salesforce.com, Google Analytics and more.
- Rich library of prebuilt components to access, prepare, blend and cleanse data.
- Direct access to complete analytics, including charts, visualizations and reporting from any step of PDI.
- Robust orchestration capabilities to coordinate complex workflows, including scheduling and alerts.
- Integration of advanced analytic models from R, Python, Scala and Weka that incorporate libraries, such as scikit-learn, Spark MLlib, Tensorflow and Keras, into the data flow.
- Enterprise-grade administration, scalability, load balancing, containerization and security capabilities.

Big Data

The Pentaho platform enables companies to realize business value from large volumes of diverse data by dramatically reducing the time and complexity required to design, develop and deploy big data analytics. Pentaho covers the entire big data life cycle, from data extraction and preparation of diverse data, to scalable processing on Spark and Hadoop, leading to end-to-end analytics solutions.

Pentaho Is the Leading Solution for Big Data Integration and Analytics

- Visual design environment for blending multiple big data sources (see Figure 3) and processing data at scale.
- Integration with leading Hadoop distributions, object stores, NoSQL stores and analytic databases, as well as log file data and JSON/XML formats.
- Code-free data transformation design that empowers 15 times faster productivity versus hand coding and executes Spark or Hadoop jobs in clusters for high performance.
- Operationalize with Spark stream and batch job execution, SQL on Spark connectivity, Kafka access and more.

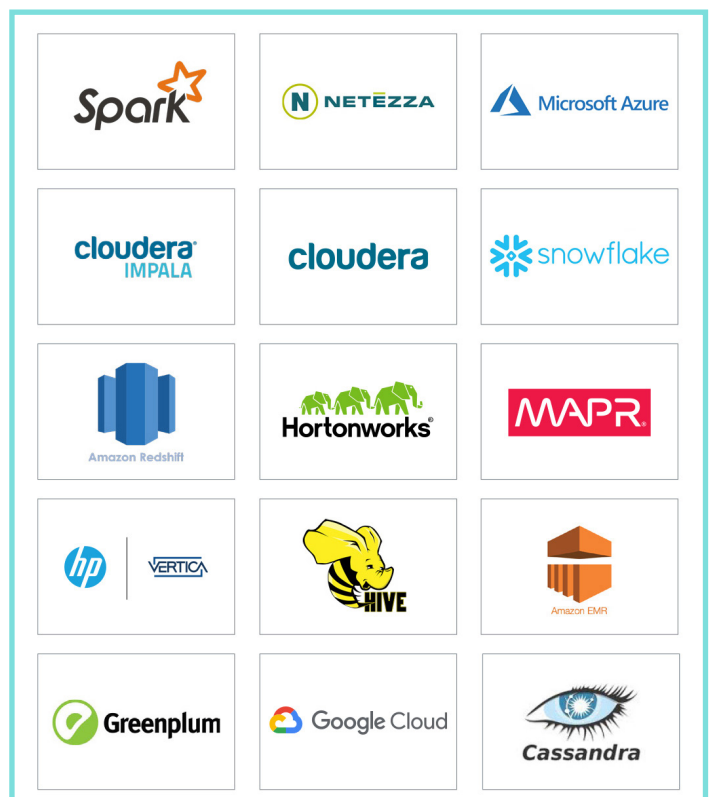Figure 2. Drag-and-Drop Data Transformation in Pentaho Data Integration

Figure 3. Variety of Big Data Sources Supported by Pentaho

"Using Pentaho, we are now helping clients blend a 360-degree view of all equipment data sources to enable early prediction of potential machinery failure."

– Caterpillar Marine Asset Intelligence

- Seamlessly switch between execution engines such as Spark and Pentaho's native engine to fit data volume, velocity and transformation complexity.
- Template-based approach to rapidly onboard data sources into Hadoop via metadata injection feature set.
- Adaptive big data layer that enables smooth portability of transformations across different Hadoop and Spark distributions.
- Pragmatic solutions to deliver on-demand data marts in a big data environment.

Multicloud Support

Extend the benefits of the open, extensible Pentaho platform to address a wide range of needs in multi, hybrid and private cloud deployments. Pentaho's modern data architecture simplifies management of your increasingly distributed data architecture with a single data management tool.

End-to-End Data Platform

- Connectivity to cloud storage and computing in AWS, Google Cloud and Microsoft Azure.
- Support for popular cloud data warehouses, including Amazon Redshift, Snowflake and Google BigQuery.
- Big data processing in Amazon EMR and HDInsight environments.
- Filtering and contextual analysis of streaming data in AWS Kinesis and Kafka.
- Bulk loading for popular cloud data warehouses.

Business Analytics

Pentaho Business Analytics provides a modern, highly interactive, and intuitive web-based interface for business users to discover and explore virtually any data. With a full spectrum of analytics tools, users can create reports and dashboards as well as visualize and analyze data across multiple dimensions without dependence on IT or developers. Meanwhile, IT benefits from secure, scalable and governed analytics for the whole enterprise. Pentaho can be deployed on premises or in the cloud and can be seamlessly embedded into other software applications. Pentaho Business Analytics provides:

Ad Hoc Analysis and Visualization

- A rich library of interactive visualizations such as geographic maps, heat grids, bubble charts and more (see Figure 4).
- Extreme scale of in-memory data caching for speed-of-thought analysis on large data volumes using a drag-and-drop paradigm.
- Visual lasso filtering and zooming to understand or exclude outliers.
- Attribute highlighting for better visual contrast among data displays.
- The ability to drill down into supporting reports for detailed data.

Flexible Dashboards

- Web-based drag-and-drop dashboard designer for business users.
- Portal and mashup integration to seamlessly integrate business analytics with other web applications.

- Rich visualizations with navigation, drill-down capabilities and a library of filter controls.
- Advanced dashboard framework for 100% tailored user experiences.

User-Driven Reporting

- Full support for operational reports, parameterized reports (see Figure 5), and interactive self-service reporting against transactional databases.
- Intuitive web-based interactive reporting for business users.
- Rich graphical pixel-perfect report designer for power users.

Mobile Business Analytics

- Ability to access data discovery, interactive analysis and visualization on mobile devices.
- Optimized mobile experience with native gestures, such as touch filtering, drill through and touch-enabled drag and drop.
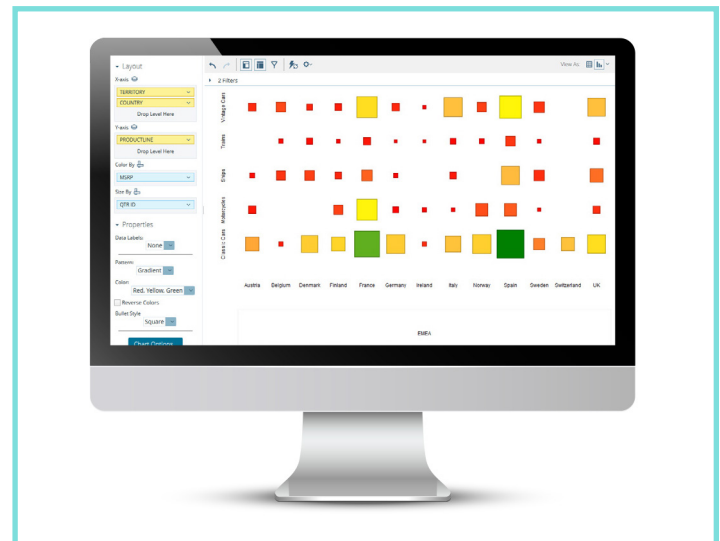

Figure 4. Heat Grid Analysis in Pentaho Business Analytics

Embedded Analytics

Pentaho's flexible cloud-ready platform is purpose-built for embedding into and integrating with your applications, portals and processes. Our powerful analytics and extensible architecture ensure that you can get to market quickly and delight your customers. Our embedded analytics solution offers:

- The ability to seamlessly embed real-time visualizations reports and dashboards into existing applications. (See Figures 5 and 6.)
- Highly customizable web-based user interface and robust web APIs that offer maximum control over the look, feel and user experience.
- Flexible capabilities for multitenant deployment as well as rich single sign-on and security integration.
- Tailored training and access to architect-level staff who have made hundreds of organizations successful.

Any Analytics, Any Data, Simplified

Pentaho addresses the barriers that block organizations from getting value from all of their data. Our platform simplifies the process of preparing and blending any data and includes a spectrum of tools to easily analyze, visualize, explore, report and predict. Open, embeddable and extensible, Pentaho is architected to ensure that each member of your team — from developers to business users — can easily translate data into value.

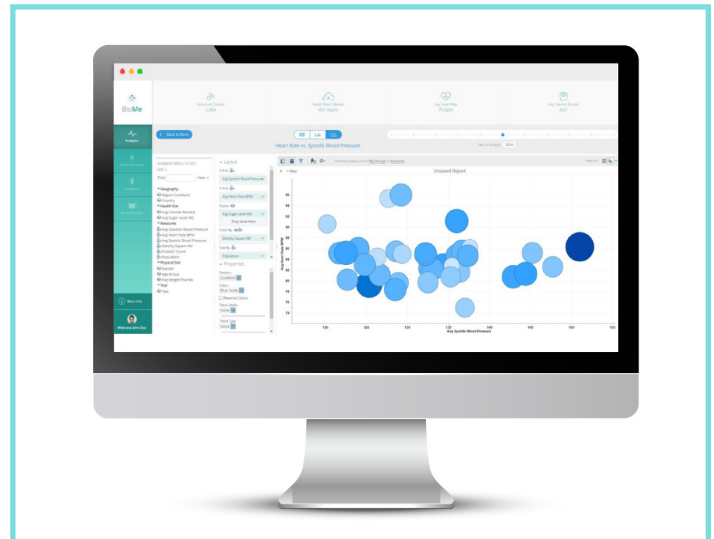**Figure 5. Intuitive Dashboard in Pentaho Business Analytics**

Figure 6. Pentaho Analytics Capability Embedded Into a Web Application

Hitachi Vantara