

# Pentaho Machine Learning Orchestration

## DATASHEET

Pentaho from Hitachi Vantara streamlines the entire machine learning workflow and enables teams of data scientists, engineers and analysts to train, tune, test and deploy predictive models.

Pentaho Data Integration and analytics platform ends the 'gridlock' associated with machine learning by enabling smooth team collaboration, maximizing limited data science resources and putting predictive models to work on big data faster — regardless of use case, industry, or language — whether models were built in R, Python, Scala or Weka.

### Streamline Four Areas of the Machine Learning Workflow

Most enterprises struggle to put models to work because data professionals often operate in silos and create bottlenecks in the data preparation to model updates workflow. The Pentaho platform enables collaboration and removes bottlenecks in four key areas:

#### 1 Prepare Data and Engineer New Features

Pentaho makes it easy to prepare and blend traditional sources like ERP and CRM with big data sources like sensors and social media. Pentaho also accelerates notoriously difficult and costly tasks of feature engineering, automating data onboarding, data transformation and data validation in an easy-to-use drag and drop environment.

#### 2 Train, Tune and Test Models

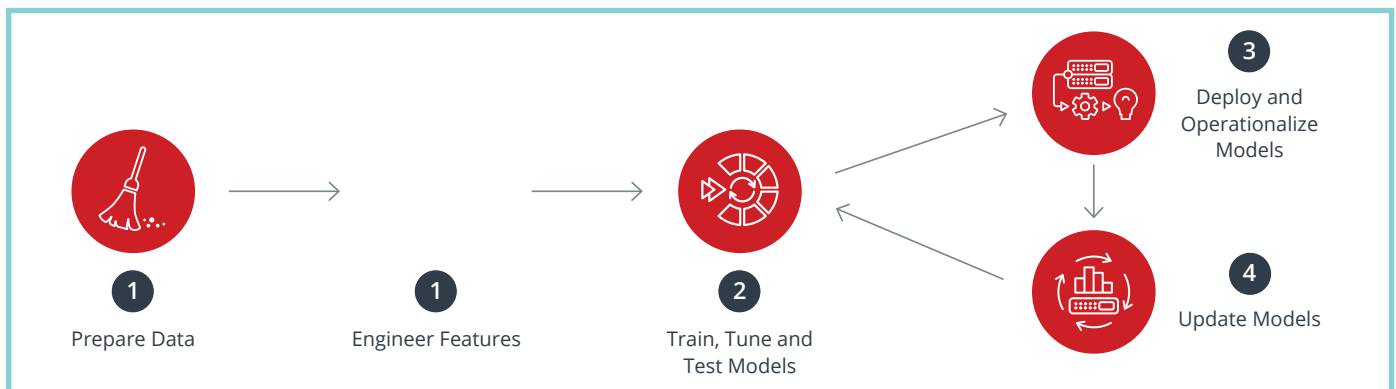
Data scientists often apply trial and error to strike the right balance of complexity, performance and accuracy in their models. With integrations for languages like R and Python, and for machine learning libraries like Spark MLlib and Weka, Pentaho allows data scientists to seamlessly train, tune, build and test models faster.

#### 3 Deploy and Operationalize Models

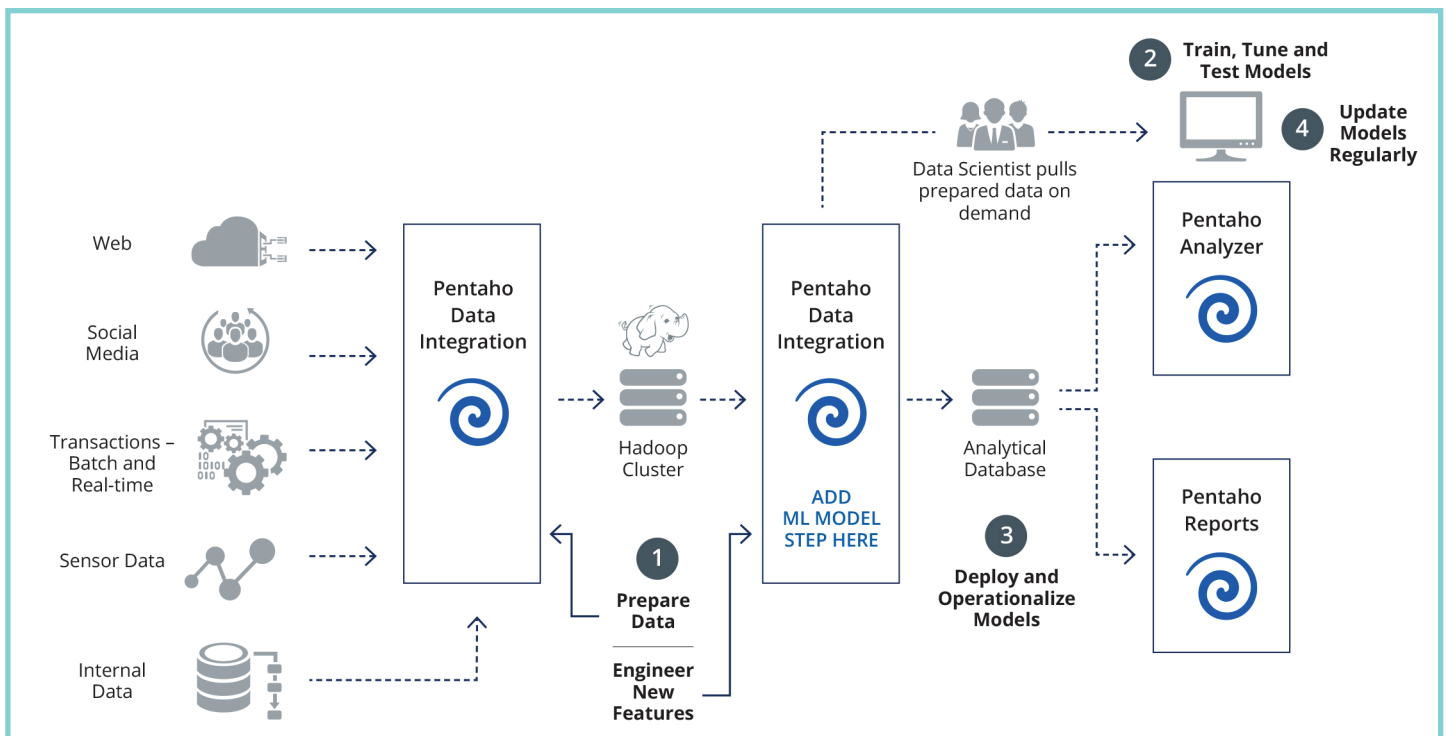
Pentaho allows data professionals to easily embed models developed by a data scientist directly in an operational workflow. They can leverage existing data and feature engineering efforts, significantly reducing time-to-deployment. With embeddable APIs, organizations can also include the full power of Pentaho within existing applications.

#### 4 Update Models Regularly

Ventana Research finds that less than a third (31%) of organizations use an automated process to update their models. With Pentaho, data engineers and scientists can re-train existing models with new data sets or make feature updates using custom execution steps for R, Python, Spark MLlib and Weka. Pre-built workflows can automatically update models and archive existing ones.



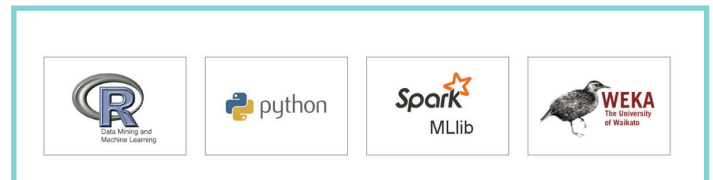
Pentaho addresses the four most important steps in the data science workflow



Deploy machine learning models using Pentaho in a complex data environment

## End-to-End Architecture

Pentaho makes it easy to onboard a wide variety of data sources into your data management environment. Using our drag and drop user interface, you can then blend, cleanse, and standardize data quickly. A data scientist can then engineer new features and pull this prepared data on-demand to train, tune and test machine learning models. The data engineer can then deploy these models into a production environment and transform your business. Finally, to update models, the data scientist can regularly use new training data with the transformations already built in Pentaho.



Integrated machine learning languages and packages

“Pentaho fills a gap to operationalize the data integration process for advanced and predictive analytics. We have embedded Pentaho for over seven years to provide remote and onboard analytics for maritime fleets and ships and have several years’ experience using Pentaho Data Integration. With Weka and R integration, we are now helping clients blend a 360-degree view of all equipment data sources to enable early prediction of potential machinery failure.”

– Ken Krooner, President, CAT Marine Asset Intelligence (on deploying R and Weka algorithms in Pentaho for predictive maintenance)

## Hitachi Vantara



Corporate Headquarters  
2845 Lafayette Street  
Santa Clara, CA 95050-2639 USA  
www.HitachiVantara.com | community.HitachiVantara.com

Regional Contact Information  
Americas: +1 866 374 5822 or info@hitachivantara.com  
Europe, Middle East and Africa: +44 (0) 1753 618000 or info.emea@hitachivantara.com  
Asia Pacific: +852 3189 7900 or info.marketing.apac@hitachivantara.com

HITACHI is a registered trademark of Hitachi, Ltd. All other trademarks, service marks, and company names are properties of their respective owners.

P-021-A DG February 2018