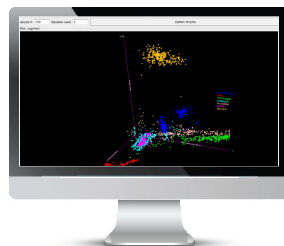
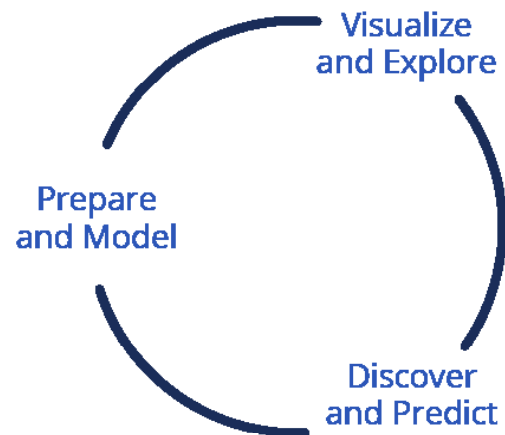


# Pentaho Predictive Analytics

Big Data is only as valuable as the insights organizations glean from it to make better decisions. Pentaho understands that looking into the future can be more compelling than reporting on the past. Pentaho Predictive Analytics is a logical next step in leveraging big data investments to make better future decisions. Pentaho Predictive Analytics provides capabilities to:

## PREDICT

- Determine the probability of something occurring for example, customer attrition or consumer demand for goods and services
- Forecast the upper and lower boundaries of a future value, for example the likely performance of a business application over the next 90 days



## OPTIMIZE

- Identify and manage risk for example, predicting consumer behavior by monitoring twitter feeds to identify conversations that indicate an intention to commit fraud
- Reduce the complexity of questions that exceed human intuition with too many variables where dozens or hundreds of factors affect an outcome

# Extending Big Data Analytics

From the point of data origin through analysis and predictive analytics, Pentaho tightly couples data integration with business analytics in a continuous big data solution to remove complexity and reduce the time to realize value from big data.

Beyond interactive visualization and exploration of data, Pentaho provides powerful, state-of-the-art machine learning algorithms and data processing tools. Data scientists and analysts can uncover meaningful patterns and correlations otherwise hidden with standard analysis and reporting. Sophisticated, advanced analytics such as time series forecasting help plan for future outcomes based on a better understanding of prior history.

Pentaho Predictive Analytics supports the whole process of predictive analytics including:

- Preparation of input data
- Statistical evaluation of learning schemes
- Visualization of input data and the result of learning
- Dozens of powerful algorithms such as classification, regression, clustering and association

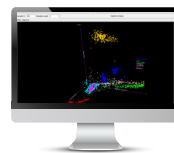
## PENTAHO'S MODERN UNIFIED BUSINESS ANALYTICS PLATFORM



DBA/ETL/BI DEVELOPER



BUSINESS USERS



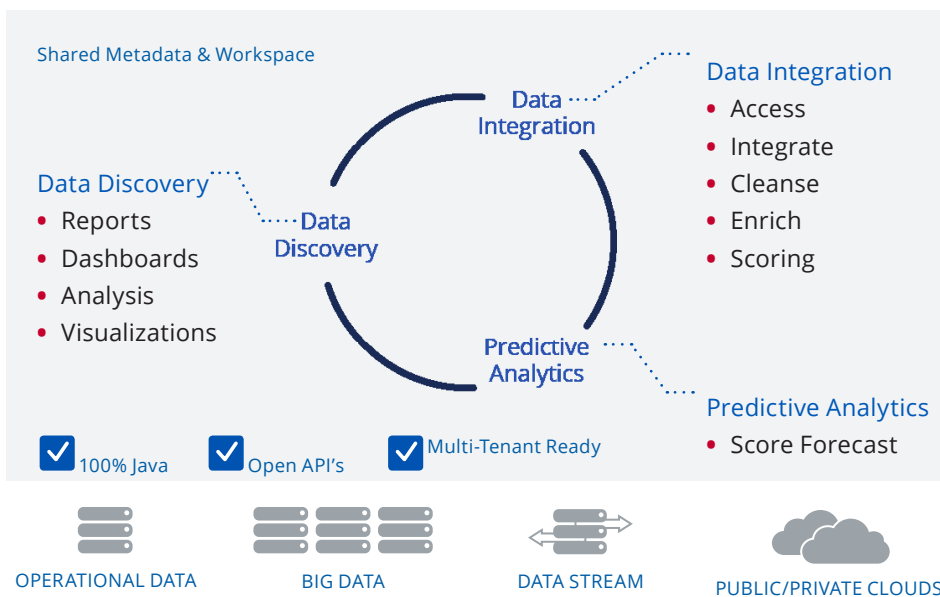
DATA ANALYSTS

→ Allows import of 3rd-party models using Predictive Modeling Markup Language (PMML 3.0)

→ Allows storing and versioning of models using the Pentaho enterprise repository Collaboration

→ Uses Pentaho Data Integration (PDI) to operationalize models by scoring records inside or outside of a Hadoop Cluster

→ Incorporates algorithms into Pentaho's visual interface



## Pentaho Time Series Forecasting

Pentaho provides practical applications for predictive analytics such as time series forecasting.

Time series forecasting is the process of using statistical techniques to model and explain a time-dependent series of data points, especially valuable when working with big data. Using a model, predictions or forecasts can be generated for future events based on known past events. For example, a large data center must maintain high availability of systems and software. With time series forecasting, collected machine generated data can be used to determine the probability of a system having performance issues. With this predicted information, budgets can be spent more efficiently reducing expenditures on back up and redundant software and systems. Forecasts like this based on large volumes of data allow organizations to reduce costs and improve efficiencies in:

- Capacity planning
- Inventory replenishment
- Future staffing level forecasts

Time series data has a natural temporal ordering which differs from typical data mining/machine learning applications where each data point is an independent example of the concept to be learned and the ordering of data points within a data set does not matter. Pentaho's time series analysis capabilities allow Data Scientists to develop, evaluate and visualize forecasting models.

## Pentaho Innovates Through Weka

Pentaho is a major sponsor of the open source project Weka, a popular suite of machine learning software written in Java. Pentaho provides time series forecasting and other capabilities through Weka as part of Pentaho Data Integration. Weka components are incorporated as tools for data pre-processing, classification, regression, clustering, association rules and visualization as well as developing new machine learning schemes.

### PENTAHO PREDICTIVE ANALYTICS - WEKA COMPONENTS

In addition to time series forecasting, other Weka components are available in the Enterprise Edition of Pentaho Data Integration as part of Pentaho Predictive Analytics.

### SCORING

Scoring allows classification and clustering models created with Weka to be used to generate probabilities and assign categories within a Pentaho Data Integration transformation. Scoring attaches a prediction to an incoming row of data. The scoring component handles all types of classifiers and clusterers that can be constructed in Weka and can also handle many types of models expressed in PMML.

"As the use of BI and Analytics have spread, innovators looking for the next competitive edge have begun to work with predictive analytics, which shifts the focus from backward-looking historical analysis to forecasting the future and providing a range of potential courses of action."

- VENTANA RESEARCH

*Benchmark Research Predictive Analytics,  
Improving Performance by Making the Future More Visible*

Be social  
with Pentaho:

