

Universal PMML Plug-in (UPPI) for Hadoop/Hive

Zementis Delivers Standards-Based Execution of Predictive Analytics on Hadoop/Hive

Zementis is committed to bring its standards-based predictive scoring engine into a variety of Big Data platforms, including the cloud as well as in-database. With the Universal PMML Plug-in (UPPI) for Hadoop/Hive, Zementis takes a big step in making its technology available for companies around the globe to easily deploy, execute, and integrate scalable standards-based predictive analytics on a massive parallel scale through the use of Hive, a data warehouse system for Hadoop.

- Linear and Logistic Regression (binary and multinomial)
- Naïve Bayes Classifiers
- General and Generalized Linear Models
- Cox Regression Models
- Rule Set Models (flat decision trees)
- Clustering Models: Distribution-Based, Center-Based, and 2-Step Clustering
- Scorecards (including reason codes)
- Multiple Models: Model segmentation, chaining, composition, cascading and ensemble (including Random Forest Models and Boosted Trees)

It also implements the definition of a data dictionary, missing and invalid values handling, outlier treatment, as well as a myriad of functions for data pre- and post-processing, including: value mapping, discretization, normalization, scaling, logical and arithmetic operators, conditional logic, built-in functions, and business decisions and thresholds.

UPPI shortens time to market for predictive insights and empowers users through instant deployment and execution of predictive models.

UPPI for Hadoop/Hive

Hive makes it possible for large datasets stored in Hadoop compatible systems to be easily analyzed. Since it provides a mechanism to project structure onto the data, Hive allows for queries to be

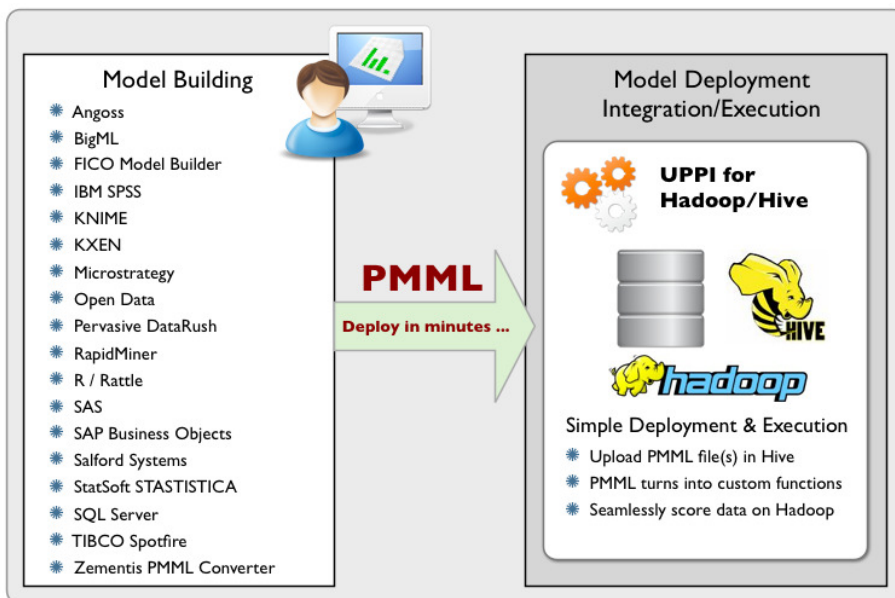


made using a SQL-like language called HiveQL.

Once deployed in UPPI, predictive models turn into UDFs (User-defined Functions). These can then be invoked directly in HiveQL. In this way, UPPI offers Hadoop users the best combination of open standards and scalability for the application of predictive analytics.

UPPI for Hadoop/Hive delivers instant and scalable scoring for Big Data while retaining compatibility with most major data mining tools through the PMML Standard. It also brings the scalability of Hadoop to the execution of predictive analytics.

Please [contact us](#) today to find out more!



UPPI Features

UPPI fully supports the Predictive Model Markup Language (PMML), the de-facto standard for data mining applications. PMML, developed by the Data Mining Group, provides a standard way for an application to define statistical and data mining models so that they can be easily shared with any other application that supports PMML.

With PMML, UPPI delivers a wide range of predictive analytics for high performance scoring, including:

- Decision Trees for classification and regression
- Neural Network Models: Back-Propagation, Radial-Basis Function, and Neural-Gas
- Support Vector Machines for regression, binary and multi-class classification