ZEMENTIS
Empowering Big Data Insights

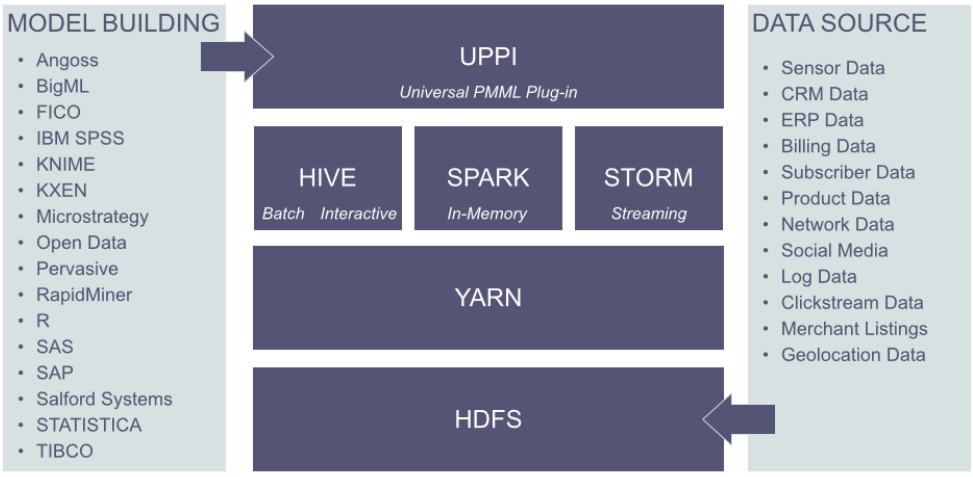
Universal PMML Plug-in (UPPI) for Hadoop

Standards-Based Predictive Analytics Execution on Hadoop - Hive, Spark & Storm

Zementis is committed to help organizations easily deploy, execute and integrate scalable, standards-based predictive analytics on a variety of Big Data platforms.

The Zementis Universal PMML Plug-in (UPPI) for Hadoop delivers a common predictive analytics strategy across the entire Hadoop ecosystem. UPPI cements an efficient process for instant operational deployment, addressing the highest execution requirements for batch processing, in-memory computation and streaming data. It enables customers to implement predictive solutions that are:

- **Highly Scalable:** Score advanced predictive analytics models for big data, in batch or real-time, and at scale
- **Plug & Play:** Use standard components of Hadoop, including Hive, Spark and Storm, without complex customization or optimization
- **100% Standards-based & Vendor-neutral:** Deploy any predictive model from virtually any data mining tool
- **Partner-certified:** Rely on Zementis' partner ecosystem for certified solutions and enterprise-grade support



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 for Hadoop 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
- 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 is a data warehouse system for Hadoop, and is optimized for querying and managing large distributed data sets. It enables access to files stored either directly in Apache HDFS (Hadoop Distributed File System) or in other compatible data storage systems. With Hive, organizations can readily analyze large data sets stored in Hadoop-compatible systems. Since it provides a mechanism

United States: Headquarters
3570 Carmel Mountain Road, Suite #300
San Diego, CA 92130
USA
Tel: +1 (619) 330-0780

United States: Silicon Valley
123 10th Street
San Francisco, CA 94103
USA
Tel: +1 (415) 800-5461
e-mail: info@zementis.com

Asia
19/F, Unit A Ho Lee Commercial Bldg.
38-44 D'Aguilar Street
Hong Kong
Tel: +852 2868 0878

ZEMENTIS
Empowering Big Data Insights

Universal PMML Plug-in (UPPI) for Hadoop

to project structure onto the data, Hive allows users to make queries using a SQL-like language called HiveQL.

Once deployed in the Zementis Universal PMML Plug-in (UPPI) for Hive, 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 Hive delivers instant and scalable scoring for big data while retaining compatibility with most major data mining tools through the PMML Standard. It brings the scalability of Hadoop to the execution of predictive analytics with the option to use Map Reduce, Tez or Spark as the underlying execution engine.

UPPI for Hadoop/Spark

Spark™ is a general cluster computing engine for large-scale data processing. For certain applications, its in-memory processing capabilities provide a much higher performance than the traditional, disk-based Map Reduce paradigm. Spark is highly applicable for machine learning and advanced predictive analytics.

With the Zementis Universal PMML Plug-in (UPPI) for Spark, PMML-based predictive models can easily be integrated into Spark Streaming. It ingests data in mini-batches and applies predictive models, which were originally built in other data mining tools, e.g., R, SPSS or SAS, on those mini-batches. This design enables the same set of application code written for batch analytics to be used in streaming analytics, on a single engine. UPPI for Hive is also able to leverage the Spark execution engine as an third option in addition to Map Reduce and Tez.

UPPI for Hadoop/Storm

Storm is a distributed real-time computation system and designed for reliably processing streaming data. Its vision is to do for real-time processing what Hadoop Map Reduce did for batch processing. Apache Storm is highly scalable and applies to a broad set of use cases, but it especially shines in real-time processing for advanced predictive analytics and machine learning models.

The Zementis Universal PMML Plug-in (UPPI) for Storm offers users a unique combination of open standards and scalability for the application of predictive analytics. With the Predictive Model Markup Language (PMML) industry standard as the bridge between the model development environment and a distributed realtime computation infrastructure, UPPI for Storm offers standards-based deployment of predictive models and execution on a highly scalable platform. UPPI seamlessly incorporates the power of advanced predictive models into the Storm infrastructure to deliver superior performance for mission-critical analytics solutions. Practically, PMML becomes a storm bolt/trident function offering execution performance that can meet the volume and performance requirements of the most demanding environments.

About Zementis

Zementis, Inc. provides software solutions for predictive analytics.

The company was founded on the principle that data science teams and IT departments can collaborate seamlessly and efficiently, allowing predictive models to rapidly move from development to deployment, so that businesses and other data-centric organizations can easily incorporate predictive analytics into their routine operations. Agile deployment of predictive solutions is the cornerstone of the Zementis philosophy.

Core solutions include ADAPA®, a decision engine for predictive analytics, and UPPI™, a universal plug-in utility for industry-leading analytics and data warehouse platforms. Zementis customers can deploy these solutions on-premise or in the cloud, with access via an intuitive Web-based console, via one of multiple industry-leading analytics platforms or as a simplified Hadoop interface.

To learn more about how your organization can benefit from Zementis, go to http://www.zementis.com, call one of our office locations or e-mail us at info@zementis.com.

Gartner 2014 Cool Vendor

CIOReview CIOBigData20 2013 20 Most Promising Big Data Companies

United States: Headquarters
3570 Carmel Mountain Road, Suite #300
San Diego, CA 92130
USA
Tel: +1 (619) 330-0780

United States: Silicon Valley
123 10th Street
San Francisco, CA 94103
USA
Tel: +1 (415) 800-5461
e-mail: info@zementis.com

Asia
19/F, Unit A Ho Lee Commercial Bldg.
38-44 D'Aguilar Street
Hong Kong
Tel: +852 2868 0878