H_2O.ai

H_2O
The Open Source In-Memory Prediction Engine

Key Benefits

BETTER PREDICTIONS

- Ready-to-use, powerful algorithms for regression, classification, clustering, and deep learning—along with advanced capabilities for churn prediction, recommendations, fraud prediction, and more.

SPEED

- In-memory processing provides real-time responsiveness and enables you to run more models.
- Fine-grain parallel distribution on big data—enabling accurate computations across one or many nodes by moving the code to the data.

EASE OF USE

- Easy set up and use, either through an intuitive Web interface or your existing tools, including R, Java, Scala, and Python.
- Model export in plain Java code for real-time scoring in any environment.

EXTENSIBILITY

- Seamless Hadoop integration with distributed data ingestion from HDFS and S3.

Algorithms

EXPLORATORY DATA ANALYTICS (EDA)

- Summary*
- K-Means*
- PCA*
- Data Munging / Transformation*

* Supported in R

ADVANCED ALGORITHMS

- Generalized Linear Model (GLM)—Poisson, Gamma Tweedie, binomial (logit), Gaussian*
- Random Forest*
- Gradient Boosted Regression*
- Gradient Boosted Classification*

* Low Latency Java Scoring

SCORING AND PREDICTION ENGINES

- GLM
- Random Forest
- Gradient Boosted Regression
- Gradient Boosted Classification
- K-Means

DEEP LEARNING

- Neural Networks

H_2O is the world's fastest in-memory platform for machine learning and predictive analytics on big data. It is the only alternative to combine the power of highly advanced algorithms, the freedom of open source, and the capacity of truly scalable in-memory processing for big data on one or many nodes. Combined, these capabilities make it faster, easier, and more cost effective to harness big data to maximum benefit for the business.

With H_2O, you can:

- **Make better predictions**. Harness sophisticated, ready-to-use algorithms and the processing power you need to analyze bigger data sets, more models, and more variables.
- **Get started with minimal effort and investment**. H_2O is an extensible open source platform that offers the most pragmatic way to put big data to work for your business. With H_2O, you can work with your existing languages and tools. Further, you can extend the platform seamlessly into your Hadoop environments.

Churn Prediction

- **Banking**. What are the profiles and usage patterns of customers who are most likely to defect?
- **Online retail**. What are the leading indicators and patterns of behavior to predict customer churn? Predict the segment of customers most likely to churn, and when, in order to intercept it and change their behavior.

Fraud Prediction

- **Payment processor**. Predict fraudulent activity using anomaly detection methods.
- **Insurance**. Stop fraud before claims are paid using real time scoring. Identify repeat offenders and score incoming claims based on fraudulent history patterns.

What can you do with better predictions? Expect more from your data.

Scoring Engine

- Score customers based on purchase history and analyze the lifetime value of key accounts to discover upsell and cross-sell opportunities.

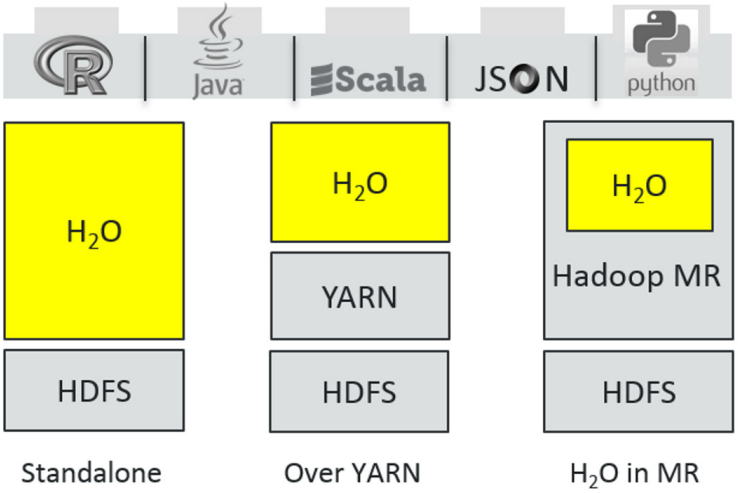
Pricing Engine

- **Travel**. Analyze different cost and promotional packages to create the most competitive combination of services.
- **Healthcare**. Discover new insights and create competitive services and healthcare programs by analyzing patient attributes, including environment, lifestyle and medical history.

Forecast

- **Real estate**. Predict property value and forecast sales by neighborhoods and regional variables. Analyze larger nationwide datasets vs. smaller sample sets to realize greater accuracy and find previously unnoticed patterns.

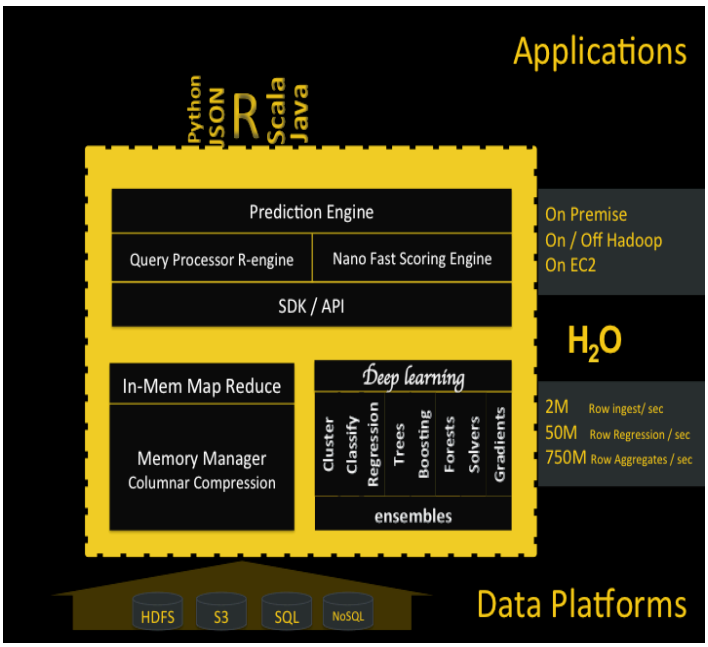
H_2O on Hadoop



Customers

vendavo Sell. More. Profitably.™ trulia® Prepaid Visa® RUSHCARD

1185 Terra Bella Ave, Mountain View, CA 94043 | 650-429-8337 | h2o.ai



Work with R, Familiar Tools and Intuitive Interfaces

Through its intuitive Web interface and integration with common tools, H_2O makes it fast and easy to get started with big data analytics. The solution works seamlessly with R and R Studio. For example, using the R interface, you can forward workflows to H_2O for big data processing, and work in a familiar interface while running algorithms on data sets that are hundreds of times larger than what would be possible on a user machine. H_2O also features native support for Java, Scala, and Python. The solution's interface is driven by JSON APIs, which makes it easy to plug into your organization's existing tools and processes to train your data and continuously improve your models and predictive accuracy.

In-Memory Processing Responsiveness

With H_2O, your organization can harness the responsiveness of highly optimized in-memory processing, so you can operationalize many more models and gain real-time intelligence in business transactions and interactions. With model export as plain Java code, you gain lightning fast real-time scoring in any environment. In addition, the solution enables data scientists to view partial query results while longer processes are running, so they can immediately spot a job that should be stopped and more quickly iterate to find the optimal approach.

Native R and Seamless Hadoop Integration

H_2O can run as a standalone platform or within an existing Hadoop installation, bringing in-memory performance to Hadoop. H_2O works with data in HDFS and supports familiar programming tools, such as Hive and Pig. In addition, the solution can be efficiently run in Amazon Web Services environments.

Fine-Grain Distributed Processing on Big Data at Speeds Up to 100x Faster

Faster H_2O lets you model interactively using in-memory processing, and delivers parallel distributed scalability required to support your big data production environments. The solution combines the responsiveness of in-memory processing with the ability to run fast serialization between nodes and clusters—so you can support the size requirements of your large data sets. Further, H_2O does this distributed processing with fine-grain parallelism, which enables optimal efficiency, without introducing degradation in computational accuracy.

Join the H_2O Movement

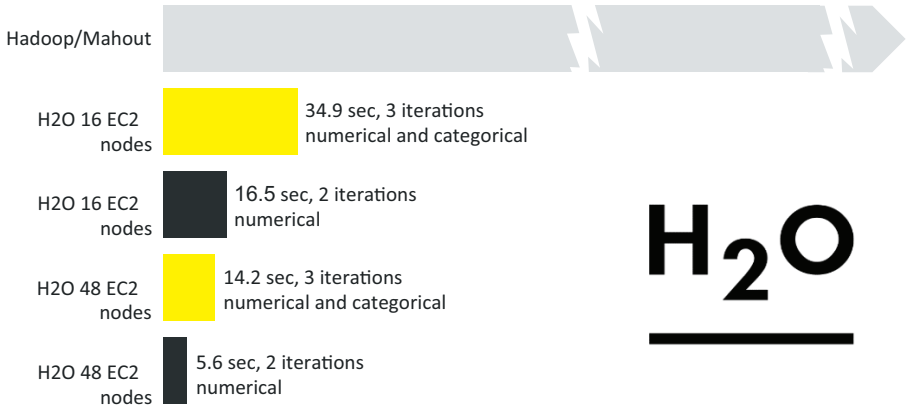
H_2O brings better algorithms to big data. H_2O is a fast open source in-memory prediction engine and machine learning platform. With H_2O enterprises can use all of their data (instead of sampling) in real-time for better predictions. Users can model data quickly and make better data-driven decisions faster by running advanced algorithms such as Deep Learning, Classification, Regression, Decision Trees, Forests, Gradient Boosting, GLM, PCA and more. Data Scientists can take both simple & sophisticated models to production from the same interactive platform used for modeling within R and JSON.

Our earliest customers have built powerful domain specific predictive engines for Recommendations, Pricing, Outlier Detection and Fraud Prediction for Insurance and Ad Platforms. H_2O is nurturing a grassroots movement of math, systems and data scientists to herald the new wave of Discovery with Big Data Science. H_2O is on CRN's 10 Coolest Big Data Products of 2013. www.h2o.ai

For latest features and updates, go to H_2O Open Source Github Repository http://0xdata.github.io/h2o/

H2O Billion Row Machine Learning Benchmark

GLM Logistic Regression



Compute Hardware: AWS EC2 c3.2xlarge - 8 cores and 15 GB per node, 1 GbE interconnect

Airline Dataset 1987-2013, 42 GB CSV, 1 billion rows, 12 input columns, 1 outcome column
9 numerical features, 3 categorical features with cardinalities 30, 376 and 380

H_2O.ai

1185 Terra Bella Ave, Mountain View, CA 94043 | 650-429-8337 | h2o.ai