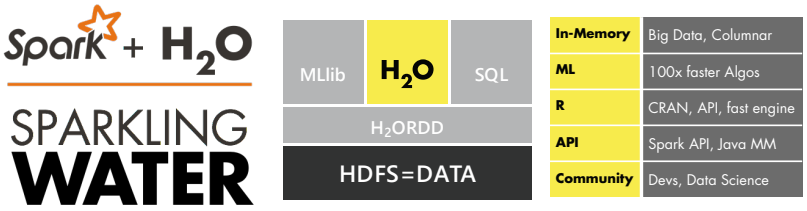
H2O.ai

DATASHEET

Fast, Scalable Machine & Deep Learning For Smarter Applications



H2O is the Killer App for Spark

Sparkling Water blends H2O's machine learning technology with Spark's fast and intuitive data-munging capabilities, creating an ideal solution that meets the demands of both the Spark data science and developer communities.

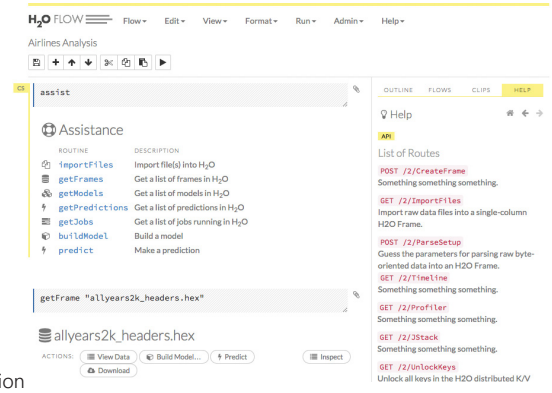
Sparkling Water can be used to transform an H2O data frame into a Schema RDD to run queries and join data frames in Spark. It can also convert a Schema RDD that has been parsed into Spark into an H2O data frame that can H2O algorithms.

http://h2o.ai/product/sparkling-water/

H2O FLOW python™

- Flow is H2O's new interactive web interface that seamlessly blends a command-line, text-based shell with a modern GUI. With Flow, H2O users can combine code execution, text, mathematics, plots and rich media into a single document without any programming experience.
- Flow offers mechanisms for capturing, replaying, annotating, sharing and presenting your analysis workflow. It also allows for model comparison and a mixed environment for:

- Coffeescript
- Text & Markdown
- Charts & Visualization
- R/Spark/Python code
- Mathematics Equations
- Video & Rich Media



Capabilities

Grid Search
- Create multiple models in a single job for hyperparameter optimization and selection of best model

Scoring Engine
- NanoFast predictions and decisions made on model output in POJO form

Open Source

Generalized Linear Model (GLM)
- Logistical regression to model a billion rows in 5 seconds

L-BFGS
- Approximation to handle millions of column

Distributed Random Forest
- Choose from two implementations that will generate your forest of decision trees optimally depending on data size

Gradient Boosting Machine (GBM)
- Produces an ensemble of decision trees with increasing refined approximations for better predictions
- Fast distributed, can be used for customer intelligence

K-Means
- Cluster like-elements for unsupervised learning use cases

Cox Proportional Hazard
- Survival models that can be used to predict customer rentention and length of patient care

Naïve Bayes
- Highly scalable linear learning algorithm

Exploratory Data Analysis with R and Python
- Quantiles, Summary, Aggregates
- Principal Component Analysis (PCA): Dimensionality Reduction
- Data Transformation: Bind and Splice
  Deep Features: Nonlinear Dimensionality Reduction

Deep Learning
- 99.1% accuracy nonlinear models with world-record performance on MNIST
- Applications in fraud and text/NLP
- Anomaly detection
- Easy to use and fast to production
- Automatic feature generation

Build smarter apps with H2O!

There are thousands of application developers who have brilliant ideas about what they want to build but are missing the data science experience necessary to make it reality. With Storm and Spark Streaming, Machine Learning can be delivered just in time through applications. Together, we can help the world build smarter apps!

Beat Bill Belichick - How do you beat the Patriots? This is the question NFL teams have been asking themselves for years. Using our GBM model, we can predict PASS / RUN given the game situation.

Find Better Bordeaux - Can we predict the quality of the current vintage of Bordeaux wine and beat wine snobs at their own game? Use our anomaly detector to see if this year's wine is worth the investment!

Churn Anaylsis - Accurately predict customer churn and create better retention campaigns. Save in customer acquisition costs and increase ROI on targeted marketing.

Key Benefits

Better Predictions - Ready-to-use, powerful algorithms that use all data. Fine-grain parallel distribution on data enables accurate computations across one or many nodes by moving the code to the data.

Speed - In-memory parallel processing provides real-time responsiveness, increases efficiency, and enables users to run more models, no sampling required.

Ease of Use - Easy to setup and adopt with intuitive H2O Flow webUI. Support for existing languages R, Java, Scala, and Python through H2O's REST API.

Extensibility - Seamless Hadoop integration with distributed data ingestion from HDFS and S3. Models are built and exportable in plain Java code.

Scalability - Easy to iterate, develop, and train models on all data sets.

Real-Time Scoring - Predict and score more accurately and 10x faster than the next best technology on the market.

Tel: +1.650.429.8337 | support@h2o.ai

http://github.com/h2oai @h2oai

Use Cases

Fraud Detection and Prevention

PayPal's Fraud Detection team uses Deep Learning to predict fraudulent activity and stop fraud before claims are paid by identifying repeat offenders and scoring incoming claims based on fraudulent history patterns. The team incorporates H2O Deep Learning for increased scalability, superior performance, and flexible deployment with high accuracy.

Predictive Modeling Factories

Cisco uses Propensity to Buy modeling factories for scoring to determine probable purchase windows of customers. By integrating H2O into the modeling process, Cisco is able to achieve real-time, immediate integration of changing trends for faster analysis and better-customized targeting without waiting for new models to be built.

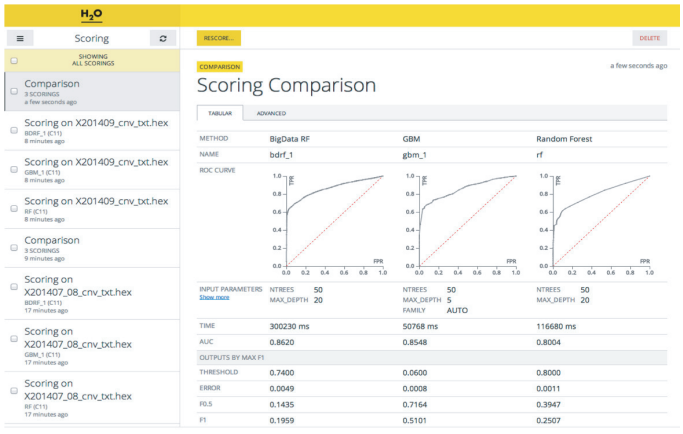
AdTech ROI Maximization

ShareThis utilizes H2O for faster model R&D and model comparisons to optimize advertising campaign placement and performance. This allows ShareThis to maximize advertising ROI.

Marketing Optimization

MarketShare's DecisionCloud software aims to improve efficiency in cross-channel attribution, revenue optimization, and forecasting in brand marketing. By using H2O, MarketShare is able to help marketers model in real-time, assess large volumes of data quickly, and customize solutions for a diverse customer base.

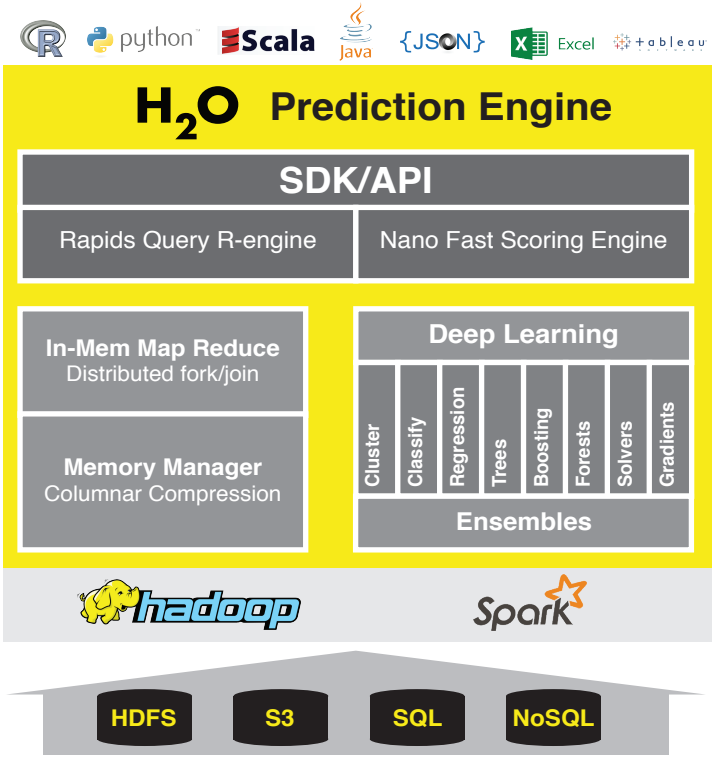
Multi-model scoring engine for model comparison



Community and Open Source

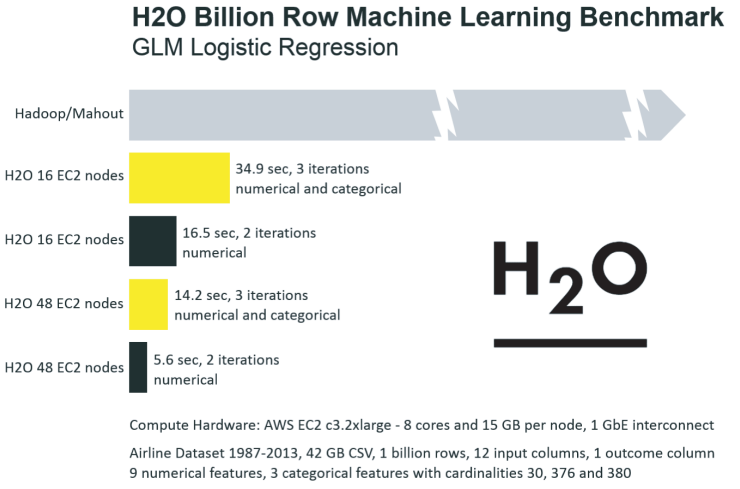
H2O is the fastest growing machine learning and data science project on Github with 20,000 commits and 10,000+ installations. H2O is Apache v2 Open Source and allows extensibility and customization by users. With 125+ meetups over 2-years, H2O is a word-of-mouth movement that brings mathematicians, engineers, and analysts together to learn and improve the product.

H2O WORLD H2O World, our first annual community conference, was sold out and brought 350+ attendees, 30+ sessions, Kaggle winners, customers and community over 2-days. Stay Tuned for H2O World 2015!



Work with familiar tools and intuitive interfaces

- H2O works seamlessly with R and R Studio, features nativ support for Java, Scala, Python, and Spark, and has a interface driven by JSON APIs, making it easy to plug into a organization's existing tools.
- H2O can be run as a standalone platform or within an existin Hadoop installation, bringing in-memory performance Hadoop. This solution also works with data in HDFS an Amazon Web Services and supports familiar programmin tools such as Hive and Pig.



Try H2O for yourself!

Download the latest version at http://www.h2o.ai/download

Join our community!

http://h2o.ai/events/
http://github.com/h2oai/h2o

cloudera databricks MAPR Hortonworks tableau