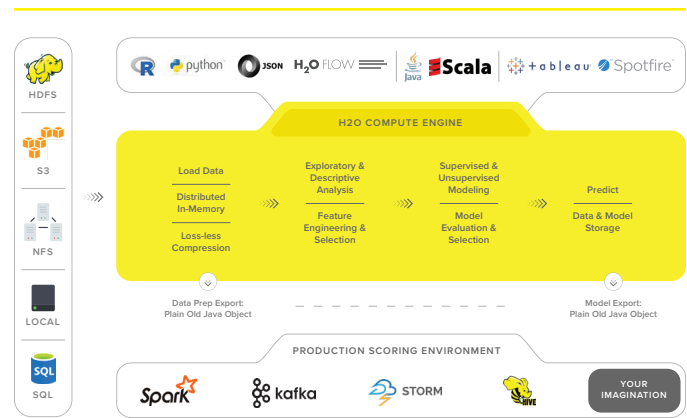# H2O Platform

100% open source, fully distributed in-memory machine learning platform with linear scalability

H2O makes it possible for anyone to easily apply machine learning and predictive analytics to solve today's most challenging business problems. It intelligently combines the following features:

- **Best of Breed Open Source Technology –** Enjoy the freedom that comes with big data science powered by open source technology. H2O was written from scratch in Java and seamlessly integrates with the most popular open source products like Apache Hadoop® and Spark™.

- **Easy-to-use WebUI and Familiar Interfaces –** Set up and get started quickly using either H2O's intuitive web-based Flow GUI or familiar programming environments like R, Python, Java, Scala, JSON, and through our powerful APIs. Models can be visually inspected during training.

- **Data Agnostic Support for all Common Database and File Types –** Easily explore and model big data from within Microsoft Excel, R Studio, Tableau and more. Connect to data from HDFS, S3, SQL and NoSQL data sources. Install and deploy anywhere, in the cloud, on premise, on workstations, servers or clusters.

- **Massively Scalable Big Data Munging and Analysis** **–** H2O Big Joins performs 7x faster than R data.table in a benchmark, and linearly scales to 10 billion x 10 billion row joins. Train a model on complete data sets, not just small samples, and iterate and develop models in real-time with H2O's rapid in-memory distributed parallel processing.

- **Real-time Data Scoring –** Rapidly deploy models to production via plain-old Java objects (POJO), model-optimized Java objects (MOJO) or REST API. Score new data against models for accurate and fast predictions in any environment.

H2O supports the following distributed algorithms:

Supervised Learning

- Statistical Analysis
  - Generalized Linear Models
  - Naïve Bayes
- Ensembles
  - Distributed Random Forest
  - Gradient Boosting Machine
- Deep Neural Networks
  - Deep Learning

Unsupervised Learning

- Clustering
  - K-means
- Dimensionality Reduction
  - Principal Component Analysis
  - Generalized Low Rank Models
- Anomaly Detection
  - Autoencoders
- Additional
  - AutoML
  - Word2Vec

Advanced Features for Data Scientists

- Automatic standardization of the predictors
- Automatic initialization of model parameters (weights & biases)
- Automatic adaptive learning rates
- Ability to specify manual learning rate with annealing and momentum
- Automatic handling of categorical and missing data
- Regularization techniques to manage complexity
- Automatic parameter optimization with grid search
- Early stopping based on validation datasets
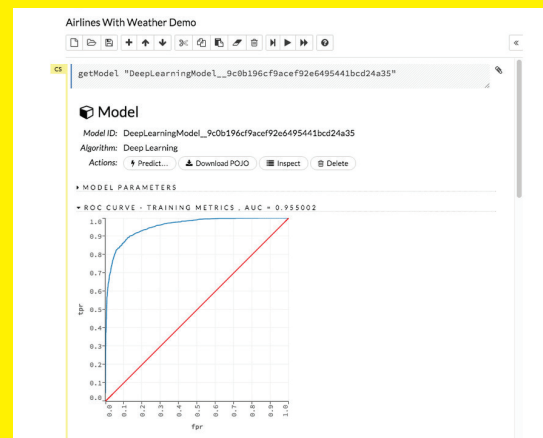- Automatic cross-validation

Quick Start

- Go to **h2o.ai/downloads**. Download H2O. This is a zip file that contains everything you need to get started.
- From your terminal, run:

```
cd ~/Downloads
unzip h2o-3.16.0.1.zip
cd h2o-3.16.0.1
java -jar h2o.jar
```

- Point your browser to **http://localhost:54321**

H₂O FLOW

- H2O provides an interactive web interface called H2O Flow that seamlessly blends a command-line text-based shell with a modern notebook-style GUI.

- With Flow, data scientists can import data, inspect it, join with another frame, split into train & test, build models with grid search, evaluate results and export a POJO or MOJO, all without writing a single line of code.

About H2O.ai

H2O.ai is the open source leader in AI and automatic machine learning with a mission to democratize AI for everyone. H2O.ai is transforming the use of AI to empower every company to be an AI company in financial services, insurance, healthcare, telco, retail, pharmaceuticals and marketing. H2O.ai is driving an open AI movement with H2O, which is used by more than 18,000 companies and hundreds of thousands of data scientists. H2O Driverless AI, an award winning and industry leading automatic machine learning platform for the enterprise, is helping data scientists across the world in every industry be more productive and deploy models in a faster, easier and cheaper way. H2O.ai partners with leading technology companies such as NVIDIA, IBM, AWS, Intel, Microsoft Azure and Google Cloud Platform and is proud of its growing customer base which includes Capital One, Nationwide Insurance, Walgreens and MarketAxess. H2O.ai believes in AI4Good with support for wildlife conservation and AI for academics. Learn more at www.H2O.ai