Overview of Machine Learning and H2O.ai

Patrick Hall
22 Jan 2017
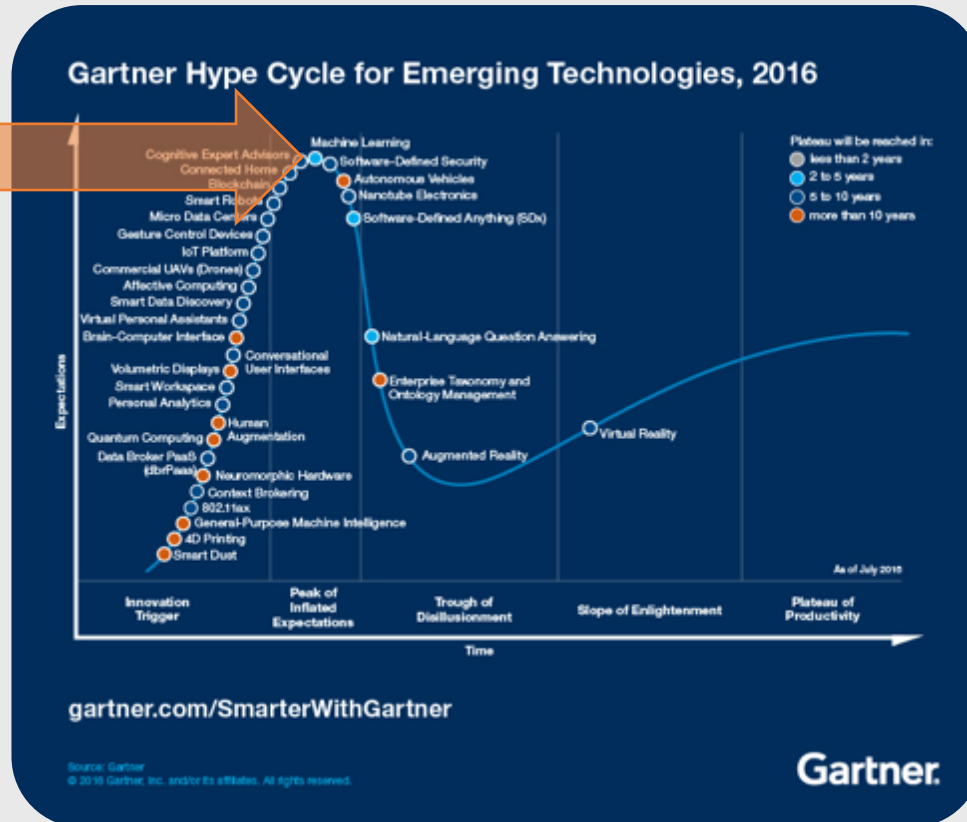
H₂O.ai

Machine Learning Overview

What is machine learning?

A field of study that gives computers the ability to learn without being explicitly programmed.

-- Arthur Samuel, 1959

H2O.ai

Gartner Hype Cycle for Emerging Technologies, 2016

Why now?

- Data, computers, and algorithms are commodities

- Unstructured data

- Increasing competition in business

H₂O.ai

Estimating a model for inference	Training a model for prediction
What happened? Why?	What will happen?
Assumptions, parsimony, interpretation	Predictive accuracy, production deployment
Linear models, statistics	Machine learning
Models tend to be static	Many models evolve elegantly

H$_2$O.ai
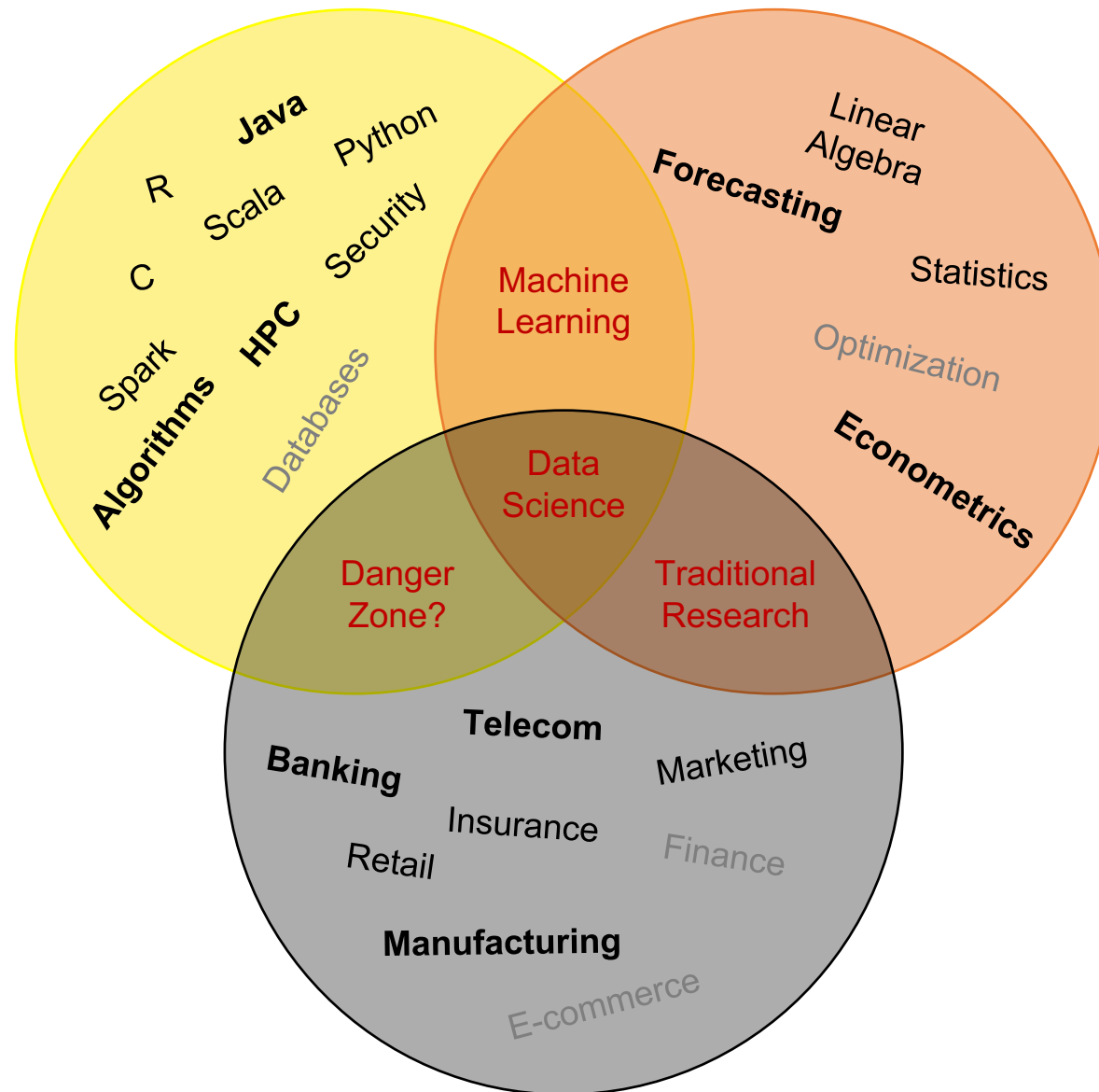
1. There is no perfect language.

If someone claims to have the perfect programming language, he is either a fool or a salesman or both.

-- Bjarne Stroustrup

2. There is no perfect algorithm.

Algorithms that search for an extremum of a cost function perform exactly the same when averaged over all possible cost functions.

-- D.H. Wolpert

3. Doing things right is always hard.

Developing and deploying ML systems is relatively fast and cheap, but maintaining them over time is difficult and expensive.

-- Google,
Hidden Technical Debt in Machine Learning Systems
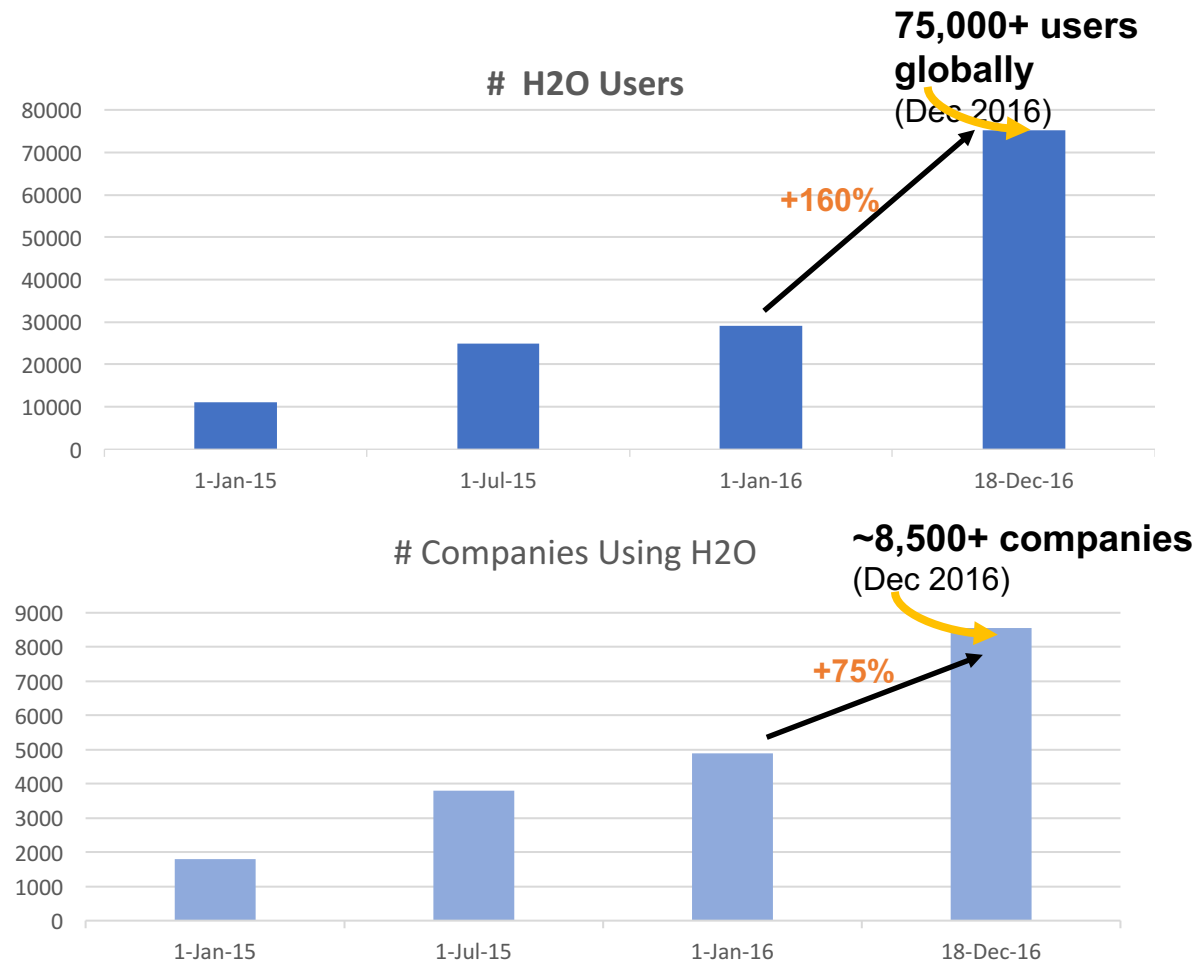
H$_2$O.ai Overview

Company Overview

Founded	2011 Venture-backed, debuted in 2012
Products	**H2O**: In-Memory AI Prediction Engine **Sparkling Water**: Spark Integration **Steam**: Deployment engine **Deep Water**: Deep Learning
Mission	Operationalize Data Science, and provide a platform for users to build beautiful data products
Team	70 employees - Distributed Systems Engineers doing Machine Learning - World-class visualization designers
Headquarters	Mountain View, CA

H2O.ai

H2O Community Growth
Tremendous Momentum Globally

 # H2O Users

75,000+ users globally
(Dec 2016)

+160%

80000			
70000			
60000			
50000			
40000			
30000			
20000			
10000			
0	1-Jan-15	1-Jul-15	1-Jan-16	18-Dec-16

 # Companies Using H2O

~8,500+ companies
(Dec 2016)

+75%

9000			
8000			
7000			
6000			
5000			
4000			
3000			
2000			
1000			
0	1-Jan-15	1-Jul-15	1-Jan-16	18-Dec-16

* DATA FROM GOOGLE ANALYTICS EMBEDDED IN THE END USER PRODUCT

Large User Circle

- 75,000+ users from 8,500+ companies in 142 countries. Top 5 from:

1. 🇺🇸 United States
2. 🇮🇳 India
3. 🇯🇵 Japan
4. 🇩🇪 Germany
5. 🇬🇧 United Kingdom

H₂O.ai

H2O.ai Offers AI Open Source Platform
Product Suite to Operationalize Data Science

100% Open Source

H2O.ai

In-Memory, Distributed Machine Learning Algorithms with Speed and Accuracy

Deep Water

State-of-the-art Deep Learning on GPUs with TensorFlow, MXNet or Caffe with the ease of use of H2O

Spark + H2O SPARKLING WATER

H2O Integration with Spark. Best Machine Learning on Spark.
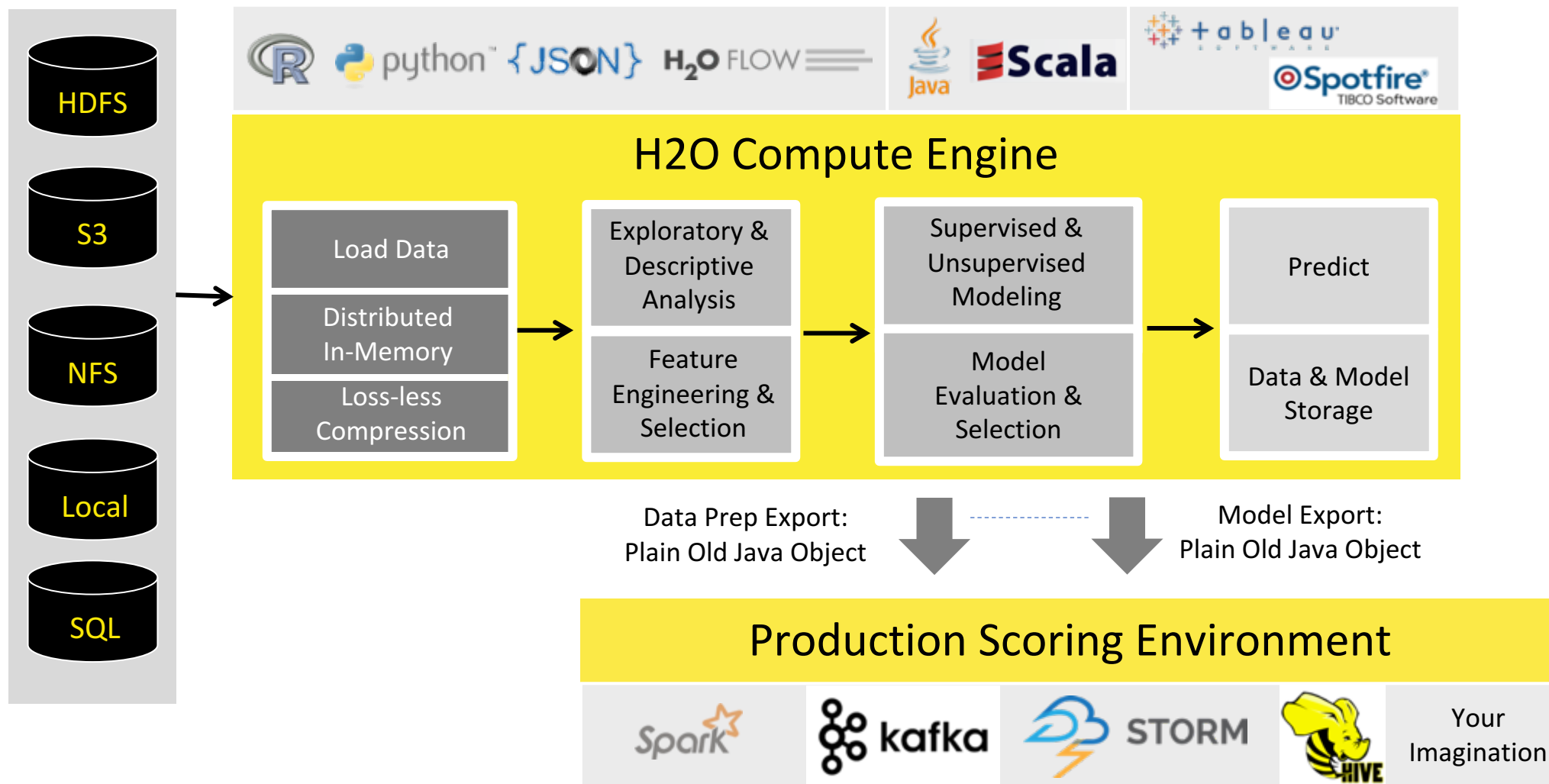
Steam

Operationalize and Streamline Model Building, Training and Deployment Automatically and Elastically

H$_2$O.ai Now Focused On **Experience**
Beyond Algorithms and Data

- **H$_2$O Flow**
  Single web-based Document for code execution, text, mathematics, plots and rich media

- **R, Python, Spark APIs**
  Advanced, scalable ML in the language of your choice

- **H$_2$O Steam**
  Elastic ML & Auto ML
  Operationalize Data Science

STEAM
AI ENGINE

VERTICALS

EXPERIENCE

ALGORITHMS

H$_2$O

H$_2$O.ai

Deep **Water**

Spark + H$_2$O
SPARKLING **WATER**

DATA

H$_2$O.ai
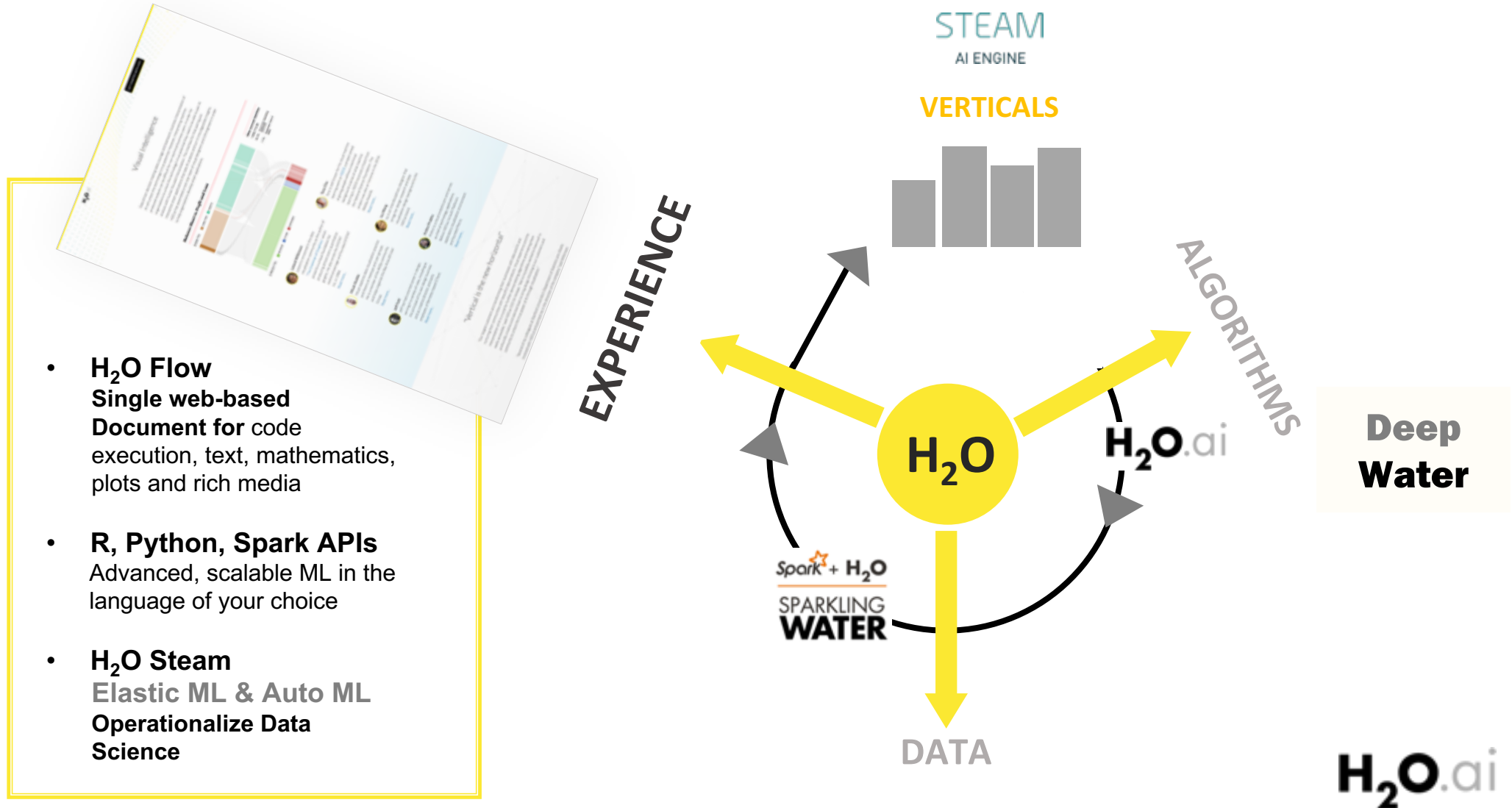
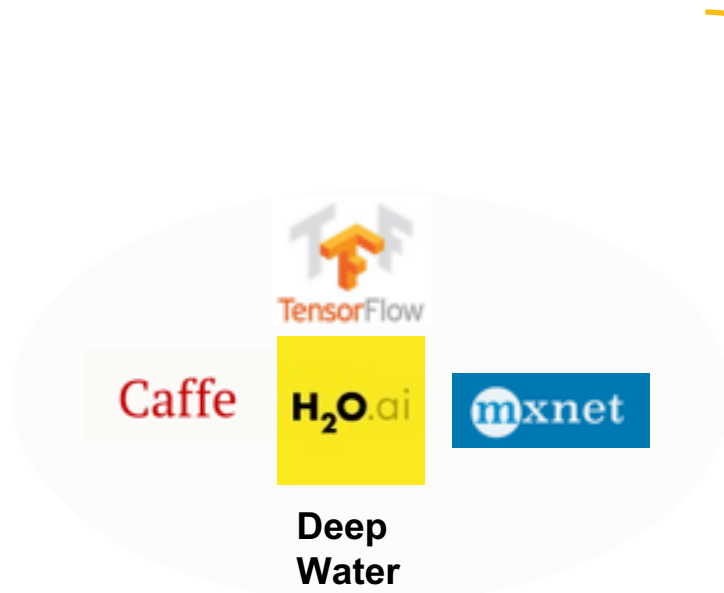# High Level Architecture

Deep Water

Next-Gen Distributed Deep Learning with H_2O

One Interface - GPU Enabled - Significant Performance Gains

Inherits All H_2O Properties in Scalability, Ease of Use and Deployment

**Deep
Water**

H_2O integrates with existing GPU backends for significant performance gains

Convolutional Neural Networks enabling **Image, video, speech recognition**

Hybrid Neural Network Architectures enabling **speech to text translation, image captioning, scene parsing** and more

Recurrent Neural Networks enabling **natural language processing, sequences, time series,** and more

H_2O.ai

Bring the Greatest Possible Machine Learning Functionality and Choice to Spark

Sparkling Water 2.0

Spark MLlib

- Elegant APIs
- RDD's
- Multi-tenant Context

H2O.ai

- Speed and scalability
- Distributed
- Columnar compression
- Fully-featured ML algorithms
- Interactive GUI – Flow
- R, Python, Java, Scala APIs

- Spark 2.0 API compatibility
- Use H2O algorithms in conjunction with, *or instead of*, MLlib algorithms on Spark
- Build Ensembles using H2O and MLlib Algorithms
- Run Spark, MLlib, Scala in Flow
- Export Spark models as POJOs
- Toolchain for ML pipelines and debugging support

H2O.ai

H$_2$O.ai Makes a Difference as an AI Platform

Open Source	Big Data Ecosystem	Flexible Interface	Smart and Fast Algorithms
• 100% open source		R Scala python Java H$_2$O Flow	
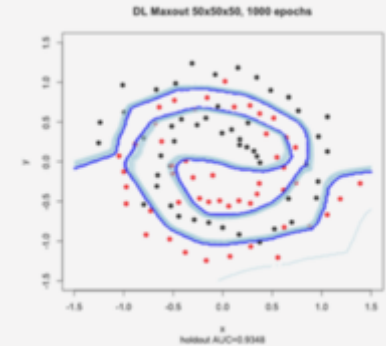
Scalability and Performance	Rapid Model Deployment	GPU Enablement	Cloud Integration
• Distributed In-Memory Computing Platform • Distributed Algorithms • Fine-Grain MapReduce	• Highly portable models deployed in Java (POJO) and Model Object Optimized (MOJO) • Automated and streamlined scoring service deployment with Rest API	GPU	
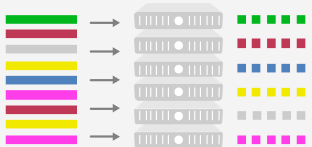
Questions?

H2O.ai