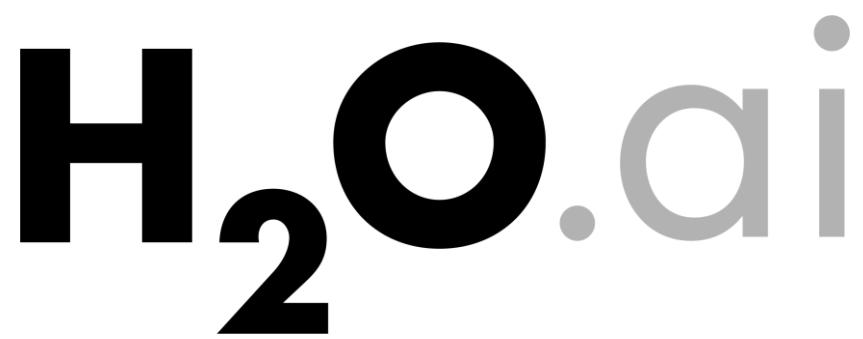


# Project “Deep Water”

H<sub>2</sub>O’s Integration with TensorFlow



Jo-fai (Joe) Chow

Data Scientist

joe@h2o.ai

@matlabulous

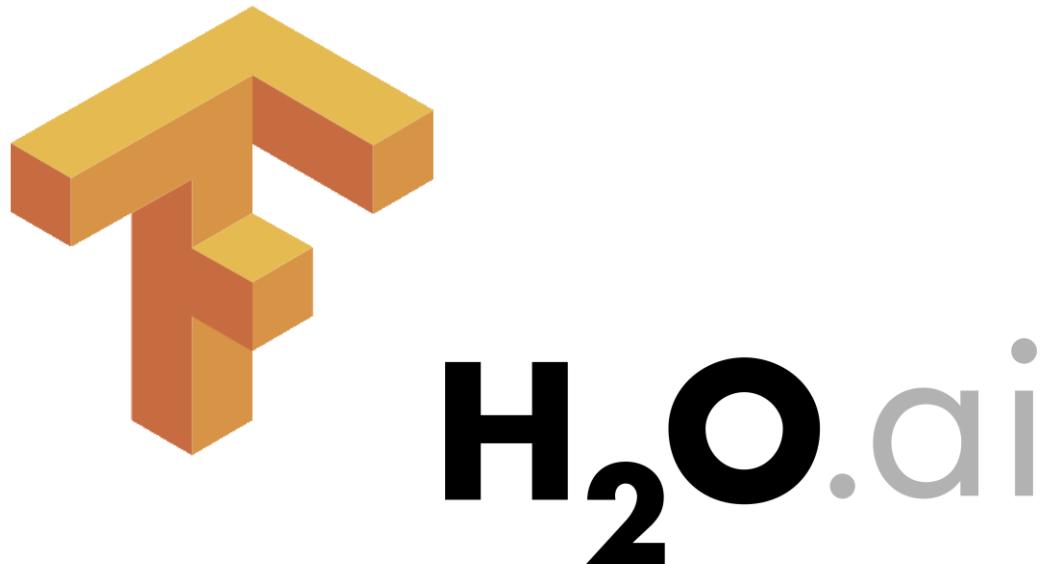
TensorFlow Paris Meetup  
30<sup>th</sup> November, 2016

# About Me

- Civil (Water) Engineer
  - 2010 – 2015
  - Consultant (UK)
    - Utilities
    - Asset Management
    - Constrained Optimization
  - Industrial PhD (UK)
    - Infrastructure Design Optimization
    - Machine Learning + Water Engineering
    - Discovered H2O in 2014
- Data Scientist
  - From 2015
  - Virgin Media (UK)
  - Domino Data Lab (Silicon Valley)
  - H<sub>2</sub>O.ai (Silicon Valley)

# Agenda

- Introduction
  - About TensorFlow
  - TensorFlow Use Cases
  - About H<sub>2</sub>O.ai
- Project Deep Water
  - Motivation
  - Benefits
  - H<sub>2</sub>O + TensorFlow Live Demo
- Conclusions

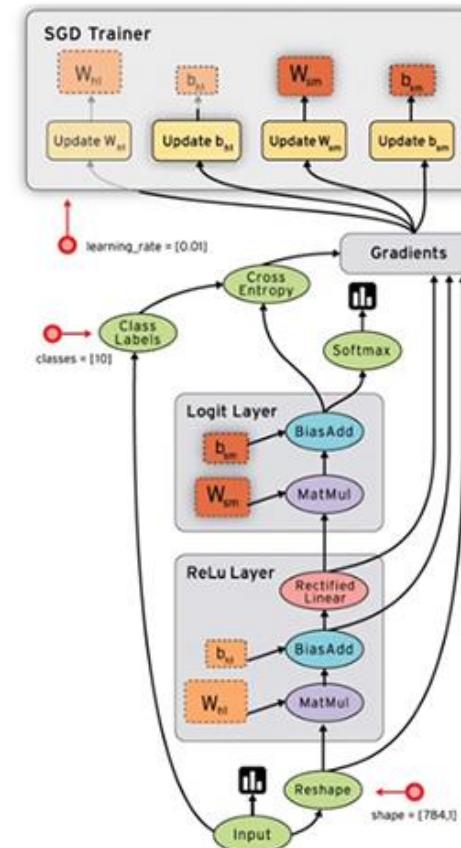


# About TensorFlow



# About TensorFlow

- Open source machine learning framework by Google
- Python / C++ API
- TensorBoard
  - Data Flow Graph Visualization
- Multi CPU / GPU
  - v0.8+ distributed machines support
- Multi devices support
  - desktop, server and Android devices
- Spark support
- Image, audio and NLP applications
- HUGE Community



# TensorFlow Wrappers

- [Scikit Flow](#) – Simplified interface
- [keras](#) – TensorFlow + Theano
- [tensorflow.rb](#) – Ruby wrapper
- [TensorFlow.jl](#) – Julia wrapper
- ... and many more!
- See: [github.com/jtoy/awesome-tensorflow](#)



# ~~TensorFlow Use Cases~~

## Very Cool TensorFlow Use Cases

Some of the cool things you can do with TensorFlow



# Tinker With a Neural Network Right Here in Your Browser.

## Don't Worry, You Can't Break It. We Promise.

Iterations  
000,084Learning rate  
0.03Activation  
TanhRegularization  
NoneRegularization rate  
0Problem type  
Classification

## DATA

Which dataset do you want to use?



Ratio of training to test data: 50%

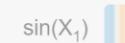
Noise: 0

Batch size: 10

REGENERATE

## FEATURES

Which properties do you want to feed in?



+ - 2 HIDDEN LAYERS

+

-

4 neurons

+

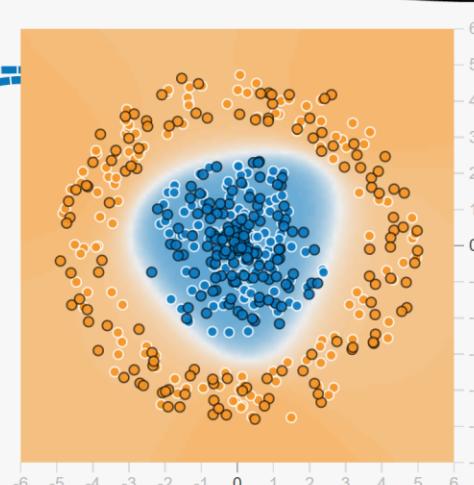
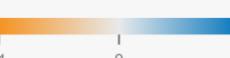
-

2 neurons

The outputs are mixed with varying weights, shown by the thickness of the lines.

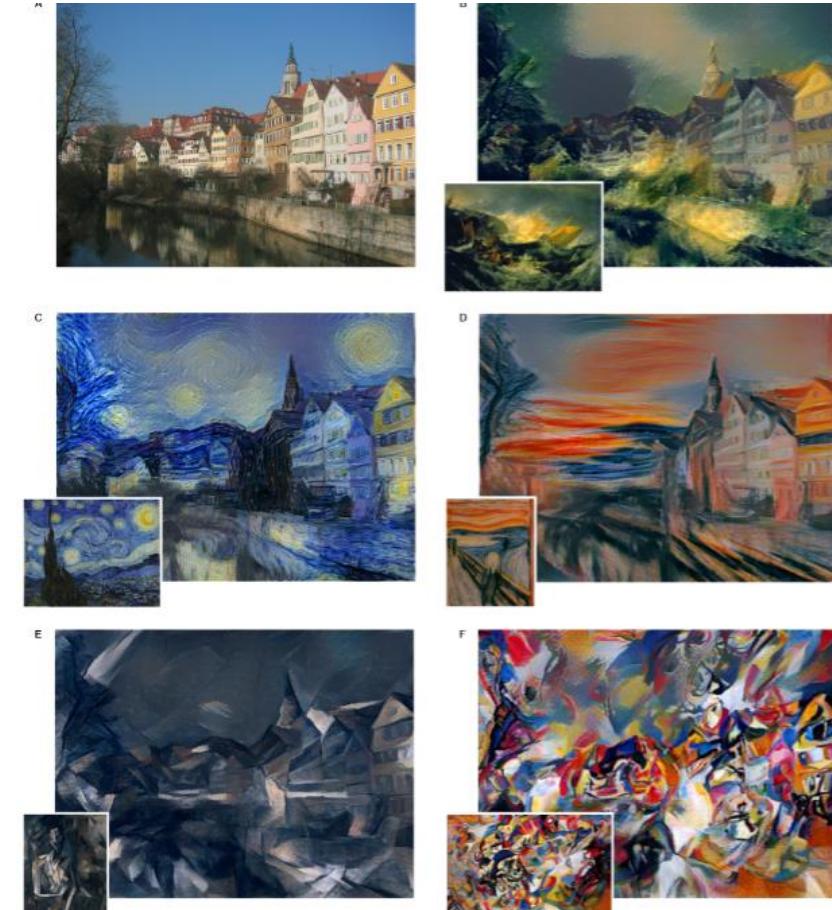
This is the output from one neuron. Hover to see it larger.

## OUTPUT

Test loss 0.020  
Training loss 0.023[playground.tensorflow.org](http://playground.tensorflow.org)Colors shows  
data, neuron and  
weight values. Show test data Discretize output

# Neural Style Transfer in TensorFlow

- Neural Style
  - “... a technique to train a deep neural network to separate artistic style from image structure, and combine the style of one image with the structure of another”
- Original Paper
  - [A Neural Algorithm of Artistic Style](#)
- TensorFlow Implementation
  - [\[Link\]](#)



# Sorting Cucumbers



- **Problem**

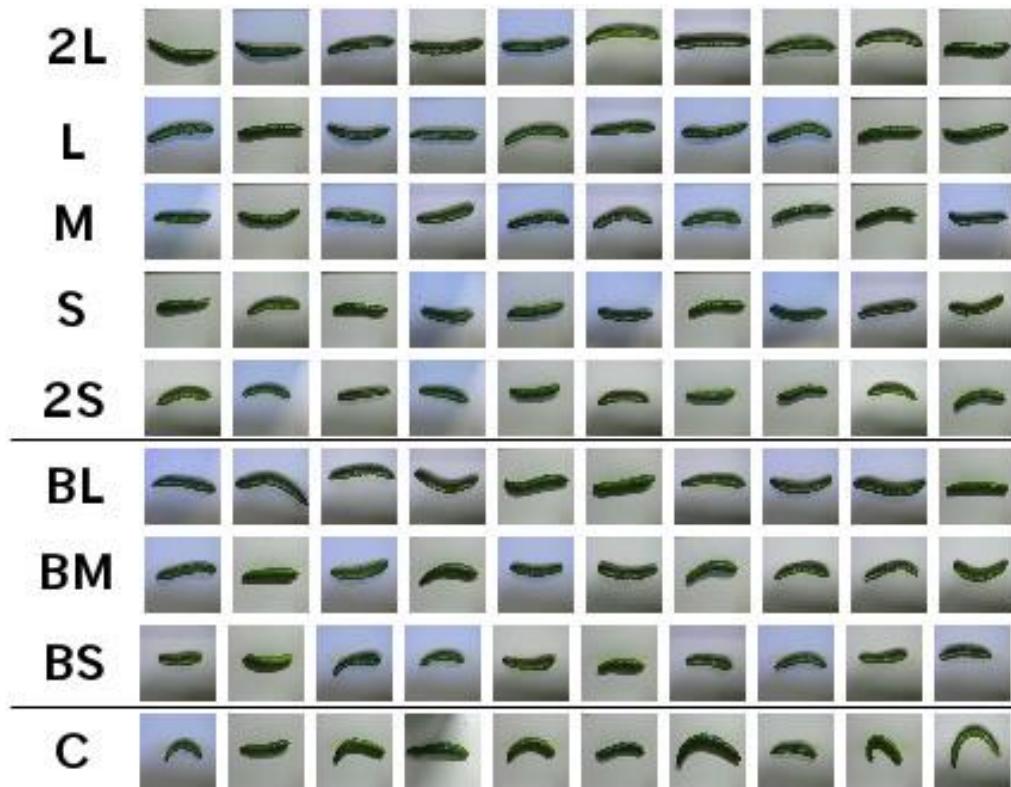
- Sorting cucumbers is a laborious process.
- In a Japanese farm, the farmer's wife can spend up to **eight hours a day** sorting cucumbers during peak harvesting period.

- **Solution**

- Farmer's son (Makoto Koike) used TensorFlow, Arduino and Raspberry Pi to create an automatic cucumber sorting system.

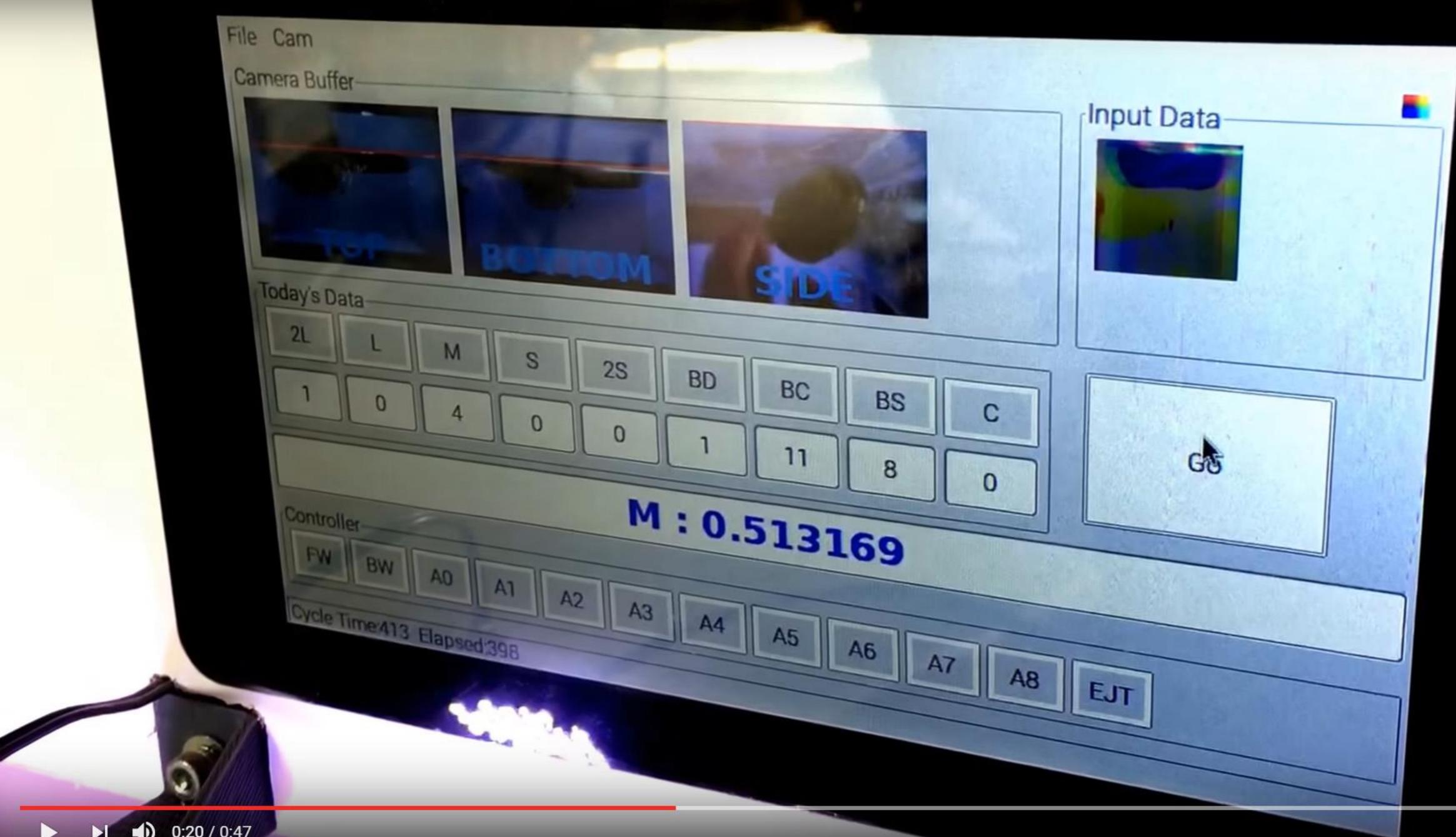
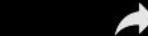


# Sorting Cucumbers



- Classification Problem
  - Input: cucumber photos (side, top, bottom)
  - Output: one of nine classes
- Google's Blog Post [[Link](#)]
- YouTube Video [[Link](#)]







Of course  
there are more TensorFlow use cases

The key message here is ...



TensorFlow  
**democratizes**  
the power of deep learning



# About H<sub>2</sub>O.ai

What exactly is H<sub>2</sub>O?

# Company Overview

<b>Founded</b>	2011 Venture-backed, debuted in 2012
<b>Products</b>	<ul style="list-style-type: none"><li>• H2O Open Source In-Memory AI Prediction Engine</li><li>• Sparkling Water</li><li>• Steam</li></ul>
<b>Mission</b>	Operationalize Data Science, and provide a platform for users to build beautiful data products
<b>Team</b>	70 employees <ul style="list-style-type: none"><li>• Distributed Systems Engineers doing Machine Learning</li><li>• World-class visualization designers</li></ul>
<b>Headquarters</b>	Mountain View, CA



H<sub>2</sub>O.ai

A large, semi-transparent image of an underwater scene with bright yellow sunlight rays filtering down through dark blue water.

**H<sub>2</sub>O** is an open source platform  
empowering business transformation

# Bring AI To Business Empower Transformation

## Financial Services, Insurance and Healthcare as Our Vertical Focus



## Community as Our Foundation

# Users In Various Verticals Adore H<sub>2</sub>O



# H2O In Action

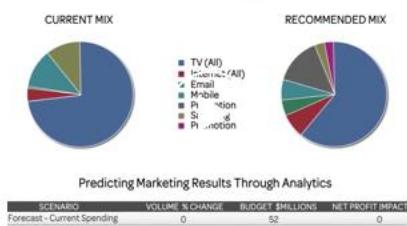
[www.h2o.ai/customers](http://www.h2o.ai/customers)

## Capital One



Capital One uses H2O open source machine learning for various use cases.

## MarketShare



H2O predictive analytics helps boost the impact and results of digital marketing.

## Kaiser



Kaiser uses H2O machine learning to save lives.

## Zurich Insurance



Zurich turned to H2O as a strategic differentiator for commercial insurance.

## Progressive



Progressive uses H2O predictive analytics for user-based insurance.

## Comcast



Comcast uses H2O to improve customer experience.

## Hospital Corporation of America



HCA uses H2O to predict patient outcomes in real-time.

## McKesson



McKesson discusses the adoption of artificial intelligence in healthcare.

## Macy's



Macy's uses H2O for personalized site recommendations.

## Transamerica



Transamerica turns to H2O to develop a product recommendation platform for insurance.

## Paypal



Paypal turned to H2O Deep Learning for fraud detection and customer churn.

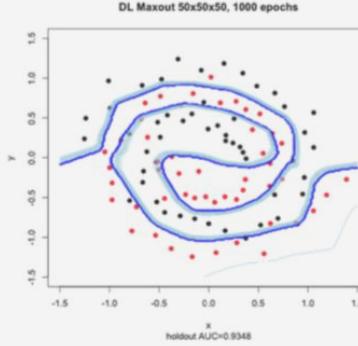
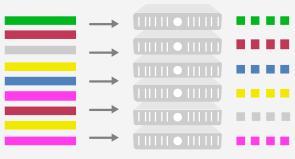
## eBay



eBay chose H2O for open source machine learning.

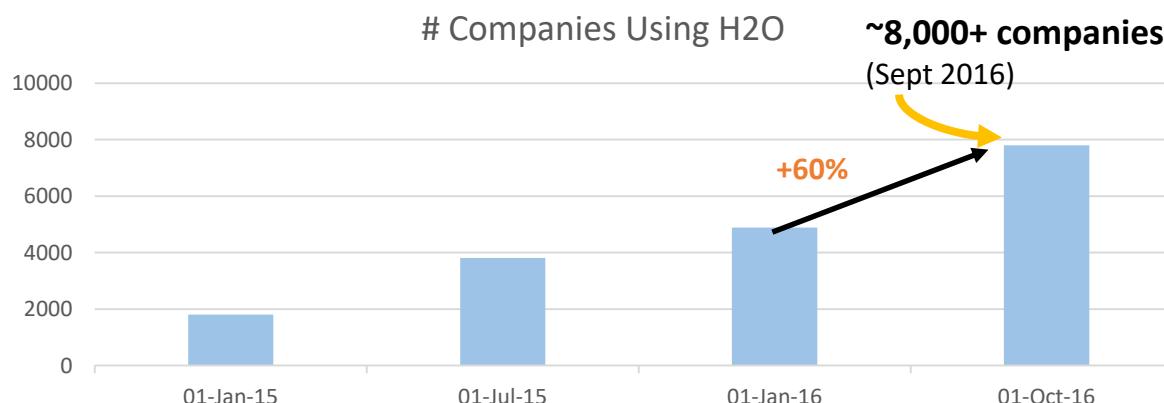
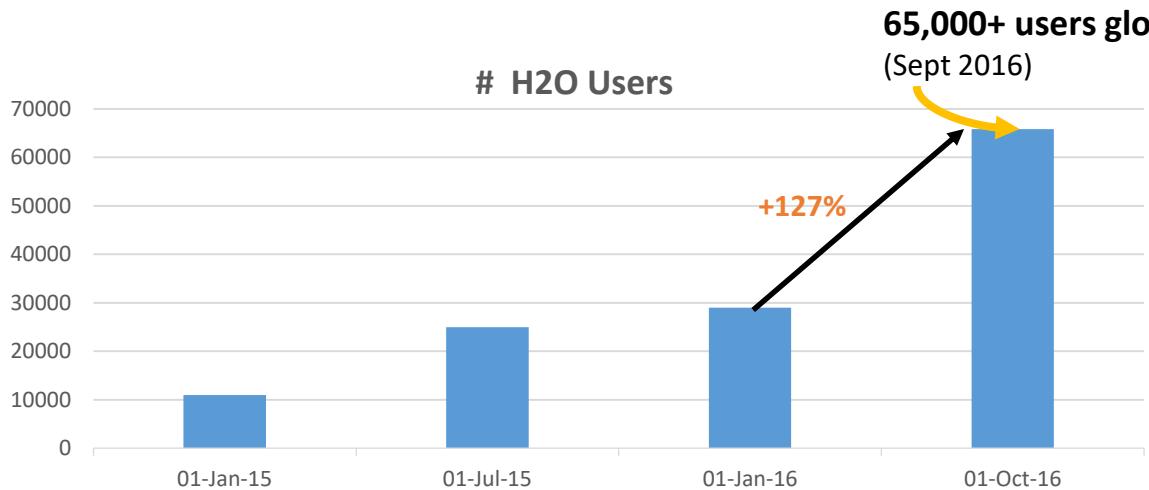
**H<sub>2</sub>O.ai**

# H<sub>2</sub>O.ai Makes A Difference as an AI Platform

Open Source	Big Data Ecosystem	Flexible Interface	Smart and Fast Algorithms
 <ul style="list-style-type: none"><li>• 100% open source</li></ul>	 	    <b>H<sub>2</sub>O Flow</b>	
Scalability and Performance	Rapid Model Deployment	GPU Enablement	Cloud Integration
 <ul style="list-style-type: none"><li>• Distributed In-Memory Computing Platform</li><li>• Distributed Algorithms</li><li>• Fine-Grain MapReduce</li></ul>	<ul style="list-style-type: none"><li>• Highly portable models deployed in Java (POJO) and Model Object Optimized (MOJO)</li><li>• Automated and streamlined scoring service deployment with Rest API</li></ul> 		  

# H<sub>2</sub>O Community Growth

## Tremendous Momentum Globally



\* DATA FROM GOOGLE ANALYTICS EMBEDDED IN THE END USER PRODUCT

23

### Large User Circle

- 65,000+ users from ~8,000 companies in 140 countries. Top 5 from:

1. United States
2. India
3. Japan
4. Germany
5. United Kingdom

# H<sub>2</sub>O Community Support

## Google forum – h2osteam

The screenshot shows the Google forum interface for the group "h2osteam". The sidebar on the left includes sections for Groups, My groups, Home, Starred, Favourites, Recently viewed, Recent searches, and Recently posted to. A yellow callout box highlights the "Favourites" section with the text "Click on a group's star icon to add it to your favourites". The main content area displays a list of topics under the heading "H2O Open Source Scalable Machine Learning - h2osteam Shared publicly". Topics include "When is Steam going to be released?", "H2O Python Modules", "H2O Installation", "PySparkling launch problem with Python 2.6 or older", "Predicted Values", and "Combining holdout predictions, while keep\_cross\_validation\_predictions parameter is active in Python". A note at the bottom encourages users to shift their energy toward building the new community website.

community.h2o.ai

Please try

The screenshot shows the H2O community website at <https://community.h2o.ai/index.html>. The sidebar on the right lists categories such as Algorithms, Announcements, Artificial Intelligence, Deep Water, Demos, H2O, Java, Machine Learning, Python, R, Source Code, Sparkling Water, Steam, Tools, and Troubleshooting. A prominent yellow box on the right announces the "Sparkling Water Release 0.8/30" with the text: "We are happy to announce that the Sparkling Water 2.0 release is almost here. On September 1, 2016 we will release Sparkling Water 2.0. Download info is coming soon." The main content area shows a list of posts under "All Posts", including discussions about Steam release, H2O Python Modules, H2O Installation, PySparkling launch problems, Predicted Values, and holdout predictions in Python.

# #AroundTheWorldWithH2Oai

London Kaggle Meetup



Strata Hadoop London



Chelsea FC



Big Data London



PyData Amsterdam



useR! 2016 Stanford



satRdays Budapest



Paris ML Meetup



Data Science Milan

**H<sub>2</sub>O.ai**

# H<sub>2</sub>O for Kaggle Competitions

**CIFAR-10 Competition**  
**Winners: Interviews with Dr.**  
**Ben Graham, Phil Culliton, &**  
**Zygmunt Zajac**

Triskelion | 01.02.2015

[READ MORE](#)

“I did really like H2O’s deep learning implementation in R, though - the interface was great, the back end extremely easy to understand, and it was scalable and flexible. Definitely a tool I’ll be going back to.”

**Kaggle challenge**  
**2nd place winner**  
**Colin Priest**

for creating this corpus. , do not contain Spanish sent. is a widespread major langu. reason was to create a corp. tasks. These tasks are com

Completed • Knowledge • 161 teams

**Denoising Dirty Documents**

Mon 1 Jun 2015 – Mon 5 Oct 2015 (3 months ago)

[READ MORE](#)

“For my final competition submission I used an ensemble of models, including 3 deep learning models built with R and h2o.”

**H<sub>2</sub>O.ai**

# H<sub>2</sub>O for Academic Research

European Journal of Operational Research

Available online 22 October 2016

In Press, Accepted Manuscript — Note to users



Innovative Applications of O.R.

Deep neural networks, gradient-boosted trees, random forests:  
Statistical arbitrage on the S&P 500

Christopher Krauss<sup>1,a</sup>, Xuan Anh Do<sup>1,a</sup>, Nicolas Huck<sup>1,b</sup>.

Received 15 April 2016, Revised 22 August 2016, Accepted 18 October 2016, Available online 22 October 2016

**Highlights**

- Latest machine learning techniques are deployed in a statistical arbitrage context.
- Deep neural networks, gradient-boosted trees, and random forests are considered.
- An equal-weighted ensemble of these techniques produces the best performance.
- Daily returns are substantial though declining over time.
- The system is especially effective at times of financial turmoil.

<http://www.sciencedirect.com/science/article/pii/S0377221716308657>

Cornell University Library

We gratefully acknowledge support from the Simons Foundation and member institutions

arXiv.org > physics > arXiv:1509.01199

Search or Article-id (Help | Advanced search) All papers ▾ Go!

Physics > Physics and Society

**Inferring Passenger Type from Commuter Eigentravel Matrices**

Erika Fille Legara, Christopher Monterola

(Submitted on 25 Aug 2015)

A sufficient knowledge of the demographics of a commuting public is essential in formulating and implementing more targeted transportation policies, as commuters exhibit different ways of traveling. With the advent of the Automated Fare Collection system (AFC), probing the travel patterns of commuters has become less invasive and more accessible. Consequently, numerous transport studies related to human mobility have shown that these observed patterns allow one to pair individuals with locations and/or activities at certain times of the day. However, classifying commuters using their travel signatures is yet to be thoroughly examined. Here, we contribute to the literature by demonstrating a procedure to characterize passenger types (Adult, Child/Student, and Senior Citizen) based on their three-month travel patterns taken from a smart fare card system. We first establish a method to construct distinct commuter matrices, which we refer to as eigentravel matrices, that capture the characteristic travel routines of individuals. From the eigentravel matrices, we build classification models that predict the type of passengers traveling. Among the models explored, the gradient boosting method (GBM) gives the best prediction accuracy at 76%, which is 84% better than the minimum model accuracy (41%) required vis-à-vis the proportional

**Download:**

- PDF
- Other formats (license)

Current browse context: physics.soc-ph  
< prev | next >  
new | recent | 1509

Change to browse by: cs cs.CY physics physics.data-an stat stat.AP stat.ML

References & Citations

- INSPIRE HEP (refers to | cited by )
- NASA ADS

Bookmark (what is this?)



<https://arxiv.org/abs/1509.01199>

$H_2O$   
**democratizes**  
artificial intelligence & big data science

# Our Open Source Products

100% Open Source. Big Data Science for Everyone!

# H<sub>2</sub>O.ai Offers AI Open Source Platform Product Suite to Operationalize Data Science with Visual Intelligence



Visual Intelligence and UX Framework For Data Interpretation and Story Telling on top of Beautiful Data Products

**100% Open Source**



**Deep  
Water**

---

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

---

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

**Spark + H<sub>2</sub>O**  
SPARKLING  
**WATER**

---

H2O Integration with Spark.  
Best Machine Learning on  
Spark.

**Steam**

---

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

# H<sub>2</sub>O.ai Offers AI Open Source Platform Product Suite to Operationalize Data Science with Visual Intelligence



Visual Intelligence and UX Framework For Data Interpretation and Story Telling on top of Beautiful Data Products

100% Open Source



---

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



---

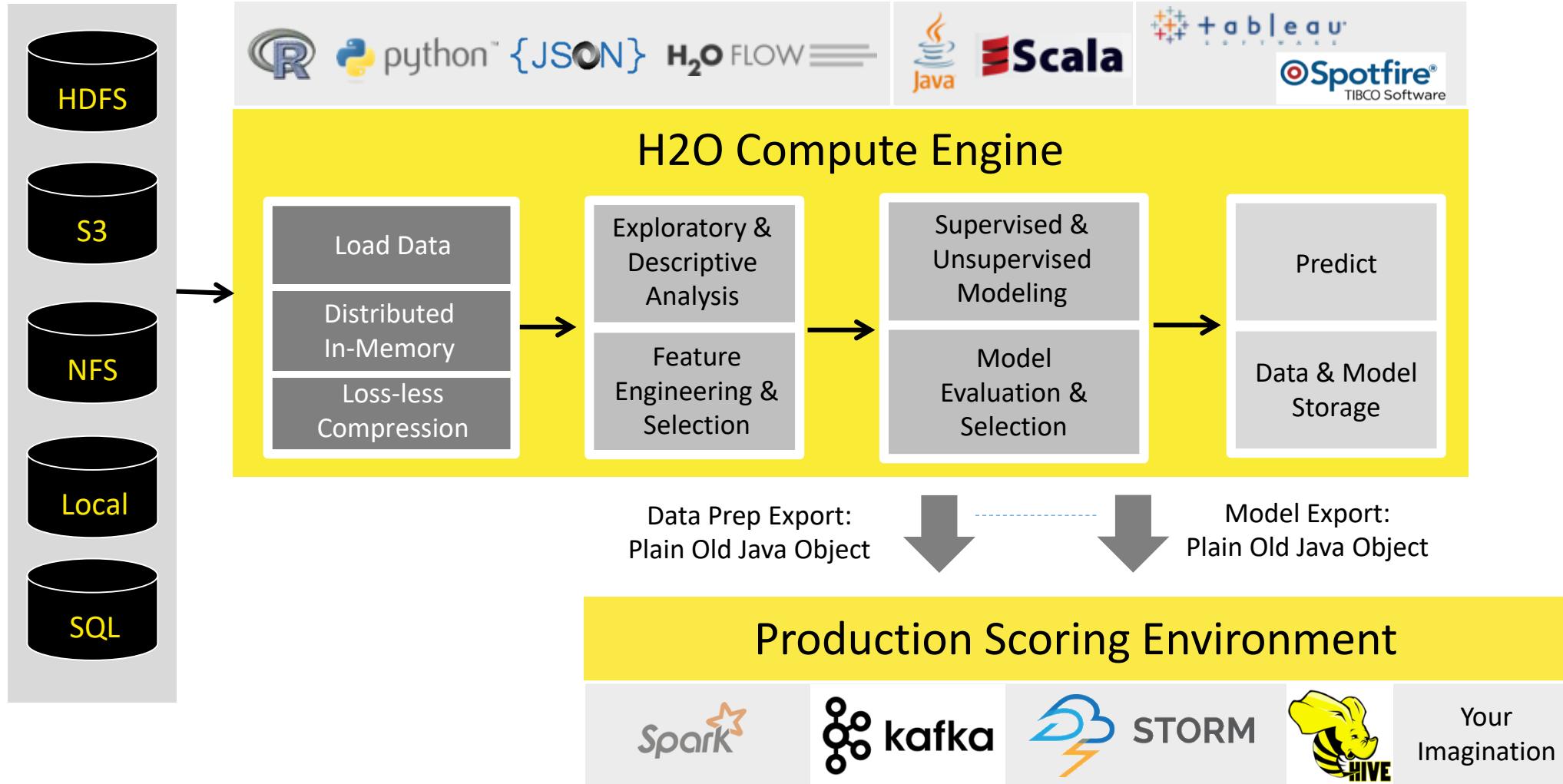
H2O Integration with Spark.  
Best Machine Learning on  
Spark.

## Steam

---

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

# High Level Architecture



# Algorithms Overview

## Supervised Learning

### Statistical Analysis

- **Generalized Linear Models:** Binomial, Gaussian, Gamma, Poisson and Tweedie
- **Naïve Bayes**

### Ensembles

- **Distributed Random Forest:** Classification or regression models
- **Gradient Boosting Machine:** Produces an ensemble of decision trees with increasing refined approximations

### Deep Neural Networks

- **Deep learning:** Create multi-layer feed forward neural networks starting with an input layer followed by multiple layers of nonlinear transformations

## Unsupervised Learning

### Clustering

- **K-means:** Partitions observations into k clusters/groups of the same spatial size. Automatically detect optimal k

### Dimensionality Reduction

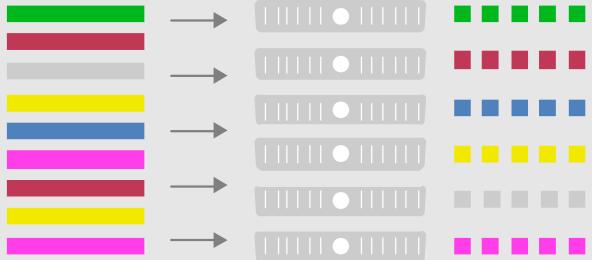
- **Principal Component Analysis:** Linearly transforms correlated variables to independent components
- **Generalized Low Rank Models:** extend the idea of PCA to handle arbitrary data consisting of numerical, Boolean, categorical, and missing data

### Anomaly Detection

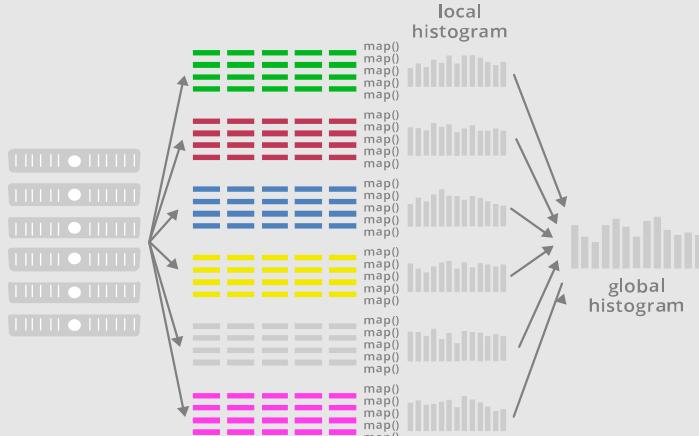
- **Autoencoders:** Find outliers using a nonlinear dimensionality reduction using deep learning

# Distributed Algorithms

## Foundation for Distributed Algorithms



Parallel Parse into **Distributed Rows**



**Fine Grain Map Reduce Illustration:** Scalable  
Distributed Histogram Calculation for GBM

## Advantageous Foundation

- Foundation for In-Memory Distributed Algorithm Calculation - **Distributed Data Frames** and **columnar compression**
- All algorithms are distributed in H<sub>2</sub>O: GBM, GLM, DRF, Deep Learning and more. Fine-grained map-reduce iterations.
- **Only enterprise-grade, open-source distributed algorithms in the market**

## User Benefits

- “Out-of-box” functionalities for all algorithms (**NO MORE SCRIPTING**) and uniform interface across all languages: R, Python, Java
- **Designed for all sizes of data sets, especially large data**
- **Highly optimized Java code for model exports**
- **In-house expertise for all algorithms**

# H<sub>2</sub>O Deep Learning in Action

116M rows, 6GB CSV file  
800+ predictors (numeric + categorical)

airlines\_all\_selected\_cols.hex

Actions: View Data, Split..., Build Model..., Predict, Download, Export

Rows	Columns	Compressed Size
116695259	12	2GB



Job

Run Time 00:00:36.712

Remaining Time 00:00:17.188

Type Model

Key Q deeplearning-dd2f42f7-81f7-42e8-9d98-e34437309828

Description DeepLearning

Status RUNNING

Progress 69%

Iterations: 12. Epochs: 0.628821. Speed: 2,243,735 samples/sec. Estimated time left: 21.849 sec

Actions View, Cancel Job

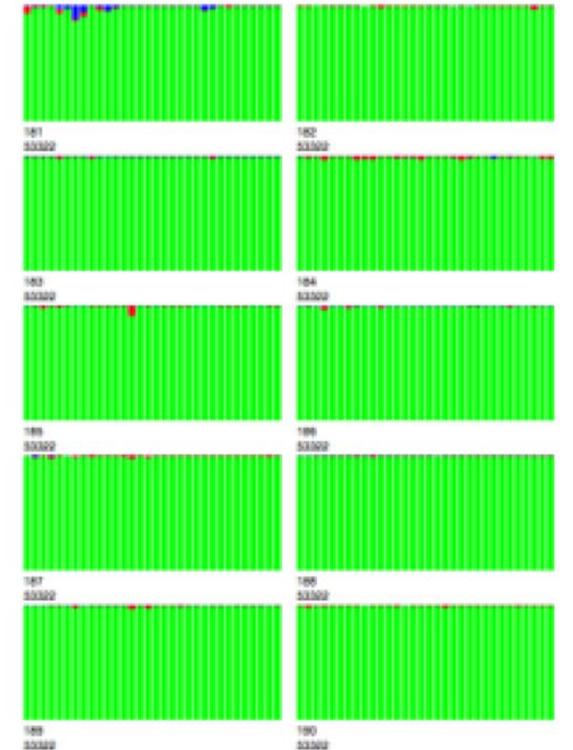
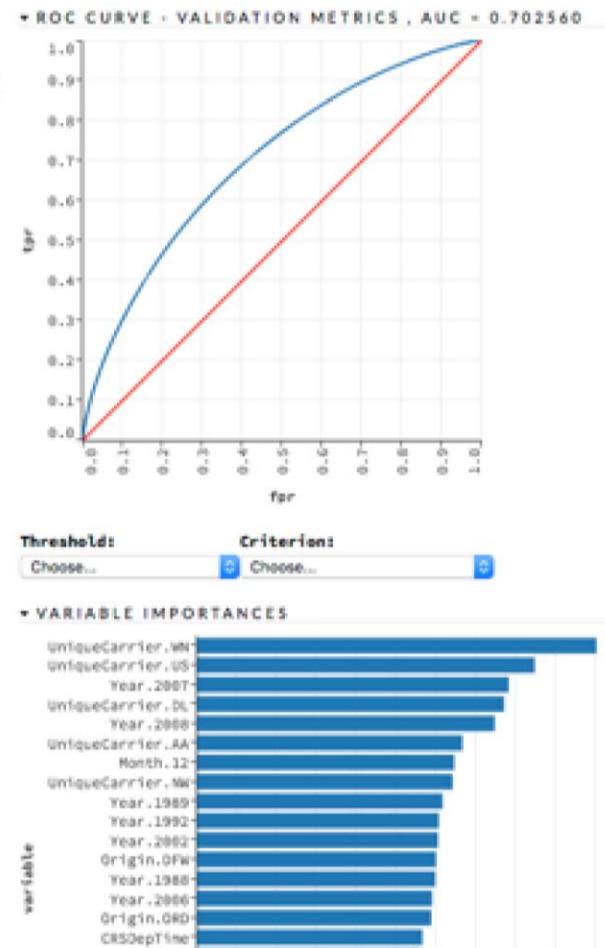
\* OUTPUT - STATUS OF NEURON LAYERS (PREDICTING ISDELAYED, 2-CLASS CLASSIFICATION, BERNoulli DISTRIBUTION, CROSSENTROPY LOSS, 17,462 WEIGHTS/BIASES, 221.3 KB, 106,585,385 TRAINING SAMPLES, MINI-BATCH SIZE 1)

layer	units	type	dropout	l1	l2	mean_rate	rate_RMS	momentum	weight_RMS	mean_weight	weight_RMS	mean_bias	bias_RMS
1	887	Input	0										
2	20	Rectifier	0	0	0	0.0493	0.2020	0	-0.0021	0.2111	-0.9139	1.0036	
3	20	Rectifier	0	0	0	0.0157	0.0227	0	-0.1833	0.5362	-1.3988	1.5259	
4	20	Rectifier	0	0	0	0.0517	0.0446	0	-0.1575	0.3068	-0.8846	0.6046	
5	20	Rectifier	0	0	0	0.0761	0.0844	0	-0.0374	0.2275	-0.2647	0.2481	
6	2	Softmax	0	0	0	0.0161	0.0083	0	0.0741	0.7268	0.4269	0.2056	

H<sub>2</sub>O.ai

Deep Learning Model

real-time, interactive  
model inspection in Flow



10 nodes: all  
320 cores busy



# H<sub>2</sub>O in Action

Quick Demo (5 minutes)

# Key Learning Resources

- Help Documentations
  - [docs.h2o.ai](https://docs.h2o.ai)
- Meetups
  - [bit.ly/h2o\\_meetup](https://bit.ly/h2o_meetup)
- YouTube Channel
  - [bit.ly/h2o\\_youtube](https://bit.ly/h2o_youtube)



## H2O, Sparkling Water, and Steam Documentation

[Getting Started](#) [Data Science Algorithms](#) [Languages](#) [Tutorials, Examples, & Presentations](#) [For Developers](#) [For the Enterprise](#)

### Getting Started

H2O
<a href="#">What is H2O?</a>
<a href="#">H2O User Guide</a>
<a href="#">Recent Changes</a>
<a href="#">Open Source License (Apache V2)</a>
<a href="#">Quick Start Video - Flow Web UI</a>
<a href="#">Quick Start Video - R</a>
<a href="#">Quick Start Video - Python</a>
<a href="#">Download H2O</a>

Sparkling Water
<a href="#">What is Sparkling Water?</a>
<a href="#">Sparkling Water Booklet</a>
<a href="#">PySparkling Readme</a>
<a href="#">RSparkling Readme</a>
<a href="#">Open Source License (Apache V2)</a>
<a href="#">Quick Start Video - Scala</a>
<a href="#">Quick Start Video - Python</a>
<a href="#">Download Sparkling Water</a>

Steam
<a href="#">What is Steam?</a>
<a href="#">Steam User Guide</a>
<a href="#">Recent Changes</a>
<a href="#">Open Source License (AGPL)</a>
<a href="#">Download Steam</a>

Questions and Answers
<a href="#">FAQ</a>
<a href="#">Community Forum</a>
<a href="#">h2ostream Google Group</a>
<a href="#">Issue Tracking (JIRA)</a>
<a href="#">Gitter</a>
<a href="#">Stack Overflow</a>
<a href="#">Cross Validated</a>
<a href="#">For Supported Enterprise Customers</a>
<a href="#">Enterprise Support via Web   Email</a>

### Data Science Algorithms

#### Supervised Learning

Generalized Linear Modeling (GLM)	<a href="#">Tutorial</a>	<a href="#">Booklet</a>	<a href="#">Reference</a>	<a href="#">Tuning</a>
Gradient Boosting Machine (GBM)	<a href="#">Tutorial</a>	<a href="#">Booklet</a>	<a href="#">Reference</a>	<a href="#">Tuning</a>
Deep Learning	<a href="#">Tutorial</a>	<a href="#">Booklet</a>	<a href="#">Reference</a>	<a href="#">Tuning</a>
Distributed Random Forest	<a href="#">Tutorial</a>	<a href="#">Booklet</a>	<a href="#">Reference</a>	<a href="#">Tuning</a>
Naive Bayes	<a href="#">Tutorial</a>	<a href="#">Booklet</a>	<a href="#">Reference</a>	<a href="#">Tuning</a>
Ensembles (Stacking)	<a href="#">Tutorial</a>	<a href="#">Booklet</a>	<a href="#">Reference</a>	<a href="#">Tuning</a>

#### Unsupervised Learning

Generalized Low Rank Models (GLRM)	<a href="#">Tutorial</a>	<a href="#">Reference</a>
K-Means Clustering	<a href="#">Tutorial</a>	<a href="#">Reference</a>
Principal Components Analysis (PCA)	<a href="#">Tutorial</a>	<a href="#">Reference</a>

# H<sub>2</sub>O.ai Offers AI Open Source Platform Product Suite to Operationalize Data Science with Visual Intelligence



Visual Intelligence and UX Framework For Data Interpretation and Story Telling on top of Beautiful Data Products

100% Open Source



---

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



---

H2O Integration with Spark.  
Best Machine Learning on  
Spark.

## Steam

---

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

# Both TensorFlow and H<sub>2</sub>O are widely used

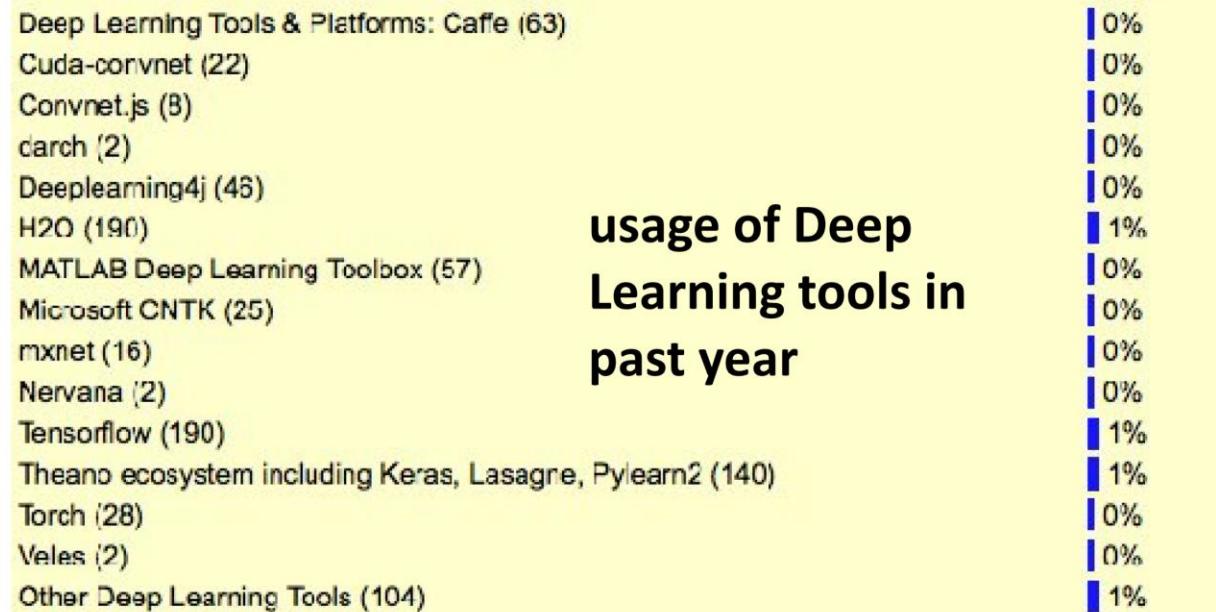
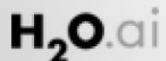
The usage of Hadoop/Big Data tools grew to 39%, up from 29% in 2015 (and 17% in 2014), driven by Apache Spark, MLlib (Spark Machine Learning Library) and H2O.

See also

- KDnuggets interview with Spark Creator Matei Zaharia
- KDnuggets interview with Arno Candel, H2O.ai on How to Quick Start Deep Learning with H2O

<http://www.kdnuggets.com>

H2O and TensorFlow are tied



**TensorFlow** democratizes the power of deep learning.

**H2O** democratizes artificial intelligence & big data science.

There are other open source libraries like MXNet and Caffe too.  
Let's have a party, this will be fun!

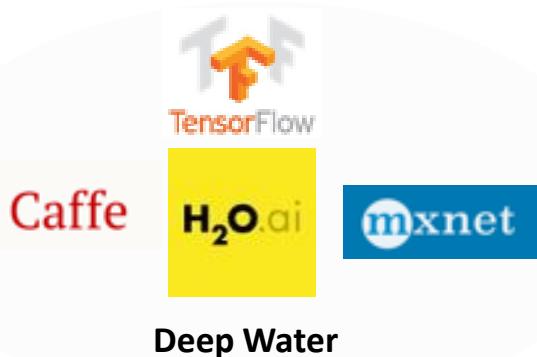


# Deep Water

Next-Gen Distributed Deep Learning with H<sub>2</sub>O

**One Interface - GPU Enabled - Significant Performance Gains**

Inherits All H<sub>2</sub>O Properties in Scalability, Ease of Use and Deployment



H<sub>2</sub>O integrates with existing **GPU** backends  
for **significant performance gains**



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



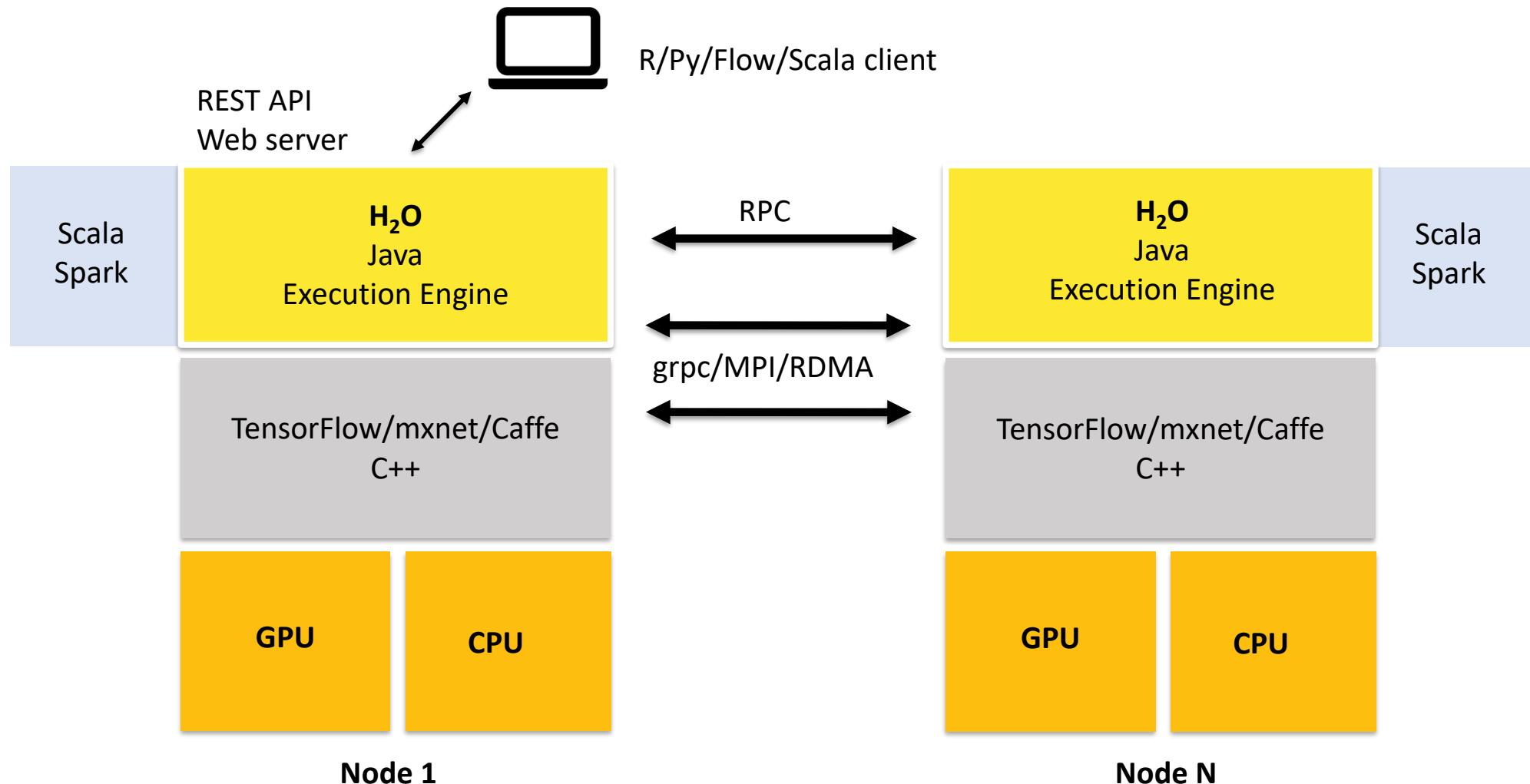
Hybrid Neural Network Architectures  
enabling **speech to text translation, image  
captioning, scene parsing** and more



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

**H<sub>2</sub>O.ai**

# Deep Water Architecture





Flow ▾

Cell ▾

Data ▾

Model ▾

Score ▾

Admin ▾

Help ▾

## Untitled Flow



CS

Expression...

Using H<sub>2</sub>O Flow to train Deep Water Model

Deep Learning...

Deep Water...

Distributed Random Forest...

Gradient Boosting Method...

Generalized Linear Modeling...

Generalized Low Rank Modeling...

K-means...

Naive Bayes...

Principal Components Analysis...

List All Models

List Grid Search Results

Import Model...

Export Model...



Ready

# Same H2O R/Python Interface

To build a LeNet image classification model in H2O, simply specify network = "lenet":

```
model <- h2o.deepwater(x=path, y=response,
                        training_frame=df, epochs=50,
                        learning_rate=1e-3, network = "lenet")
model
|=====
Model Details:
=====

H2OMultinomialModel: deepwater
Model ID: DeepWater_model_R_1477378862430_2
Status of Deep Learning Model: lenet, 1.6 MB, predicting C2, 3-class classification, 14,336 training samples, mini-batch size 32
  input_neurons    rate momentum
  1           2352  0.000986  0.990000

H2OMultinomialMetrics: deepwater
** Reported on training data. **
** Metrics reported on full training frame **

Training Set Metrics:
=====

Extract training frame with `h2o.getFrame("cat_dog_mouse.hex_sid_95f8_1")`
MSE: (Extract with `h2o.mse`) 0.131072
RMSE: (Extract with `h2o.rmse`) 0.3620386
Logloss: (Extract with `h2o.logloss`) 0.4176429
```

# Deep Water Roadmap (Q4 2016)



**Finish TensorFlow integration (C++/Python/Java):  
Package Python on the backend to create trainable graphs**



**Finish Caffe integration (pure C++/Java):  
Optimized Multi-GPU training (NVIDIA NCCL)**



**Add multi-GPU support for mxnet  
Add more capabilities to H2O Deep Water:  
Text/NLP, Time Series, LSTM, AutoEncoder,  
Feature Extraction, Input/Output shape mapping, etc.**

# H<sub>2</sub>O + TensorFlow Live Demo

# Deep Water H<sub>2</sub>O + TensorFlow Demo

- H<sub>2</sub>O + TensorFlow
  - Dataset – Cat/Dog/Mouse
  - TensorFlow as GPU backend
  - Train a LeNet (CNN) model
  - Interfaces
    - Python (Jupyter Notebook)
    - Web (H<sub>2</sub>O Flow)
- Code and Data
  - [github.com/h2oai/deepwater](https://github.com/h2oai/deepwater)

# Code and References

## Python/R Jupyter Notebooks

Check out a sample of cool Deep Learning [Jupyter notebooks!](#)

## PreRelease Downloads

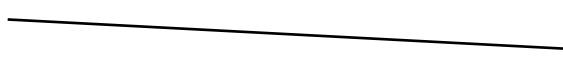
For the following system dependencies, we provide recent builds for your convenience.

- Ubuntu 16.04 LTS
- Latest NVIDIA Display driver
- CUDA 8 (latest available) in /usr/local/cuda
- CUDNN 5 (inside of lib and include directories in /usr/local/cuda/)

In the future, we'll have more pre-built jars for more OS/CUDA combinations.

- Required to run Jupyter notebook: [H2O Deep Water enabled Python module](#) -- install via `pip install <file>`
- To build custom networks: [Matching MXNet Python egg](#) -- install via `easy_install <file>`
- To run from Flow only: [H2O Standalone h2o.jar](#) -- launch via `java -jar h2o.jar`

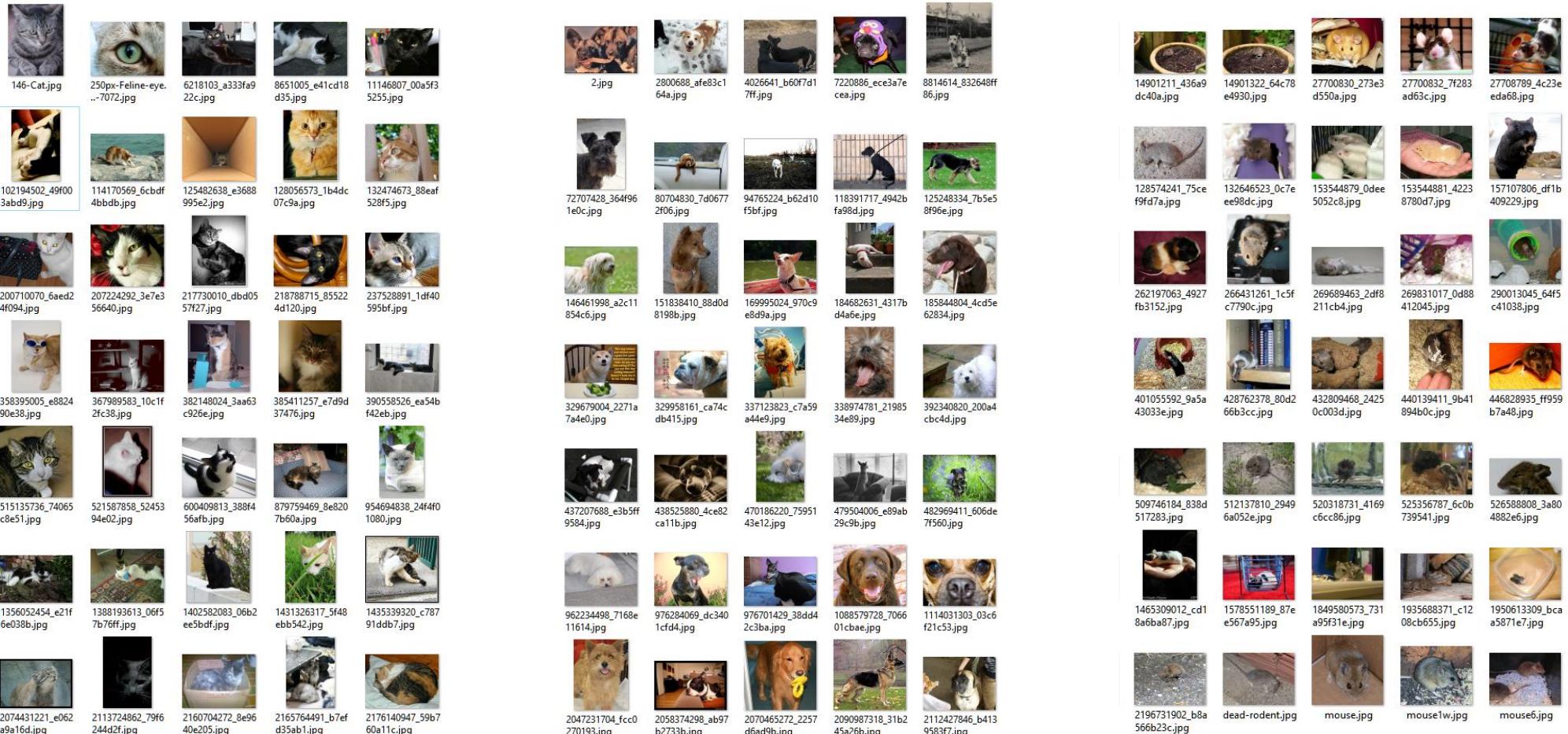
If you are interested in running H2O Deep Water on a different infrastructure, see the DIY build instructions below



h2oai / h2o-3		Unwatch ▾ 246	★ Unstar 1,475	Fork 655
Code Pull requests 26 Projects 0 Pulse Graphs				
Branch: master h2o-3 / examples / deeplearning / notebooks /		Create new file	Upload files	Find file
arnocandel committed on GitHub	Update README.md	Latest commit 7f61d39 7 days ago		
..				
<a href="#">images</a>	Add cat/dog/mouse lenet example.	11 days ago		
<a href="#">README.md</a>	Update README.md	7 days ago		
<a href="#">deeplearning_anomaly_detection.ipynb</a>	Update notebooks, introduce local paths to ~/h2o-3/	11 days ago		
<a href="#">deeplearning_benchmark_mnist.ipynb</a>	Update lenet test to remove all. Update MNIST benchmark with comments.	8 days ago		
<a href="#">deeplearning_cat_dog_mouse_incep...</a>	Add credit card default risk model, update other notebooks.	11 days ago		
<a href="#">deeplearning_cat_dog_mouse_lenet...</a>	Add credit card default risk model, update other notebooks.	11 days ago		
<a href="#">deeplearning_cat_dog_mouse_lenet...</a>	Add back model.plot() and scoring history.	9 days ago		
<a href="#">deeplearning_cifar10_vgg.ipynb</a>	Rename notebooks.	12 days ago		
<a href="#">deeplearning_credit_card_default_ri...</a>	Update notebooks, introduce local paths to ~/h2o-3/	11 days ago		
<a href="#">deeplearning_grid_iris.ipynb</a>	Add two new notebooks: Lenet for R and iris grid for python	10 days ago		
<a href="#">deeplearning_grid_iris.Ripynb</a>	Update R py notebook.	10 days ago		
<a href="#">deeplearning_image_reconstruction...</a>	Update notebooks, introduce local paths to ~/h2o-3/	11 days ago		
<a href="#">deeplearning_mnist_convnet.ipynb</a>	Update notebooks, introduce local paths to ~/h2o-3/	11 days ago		
<a href="#">deeplearning_mnist_introduction.ip...</a>	Add missing file.	10 days ago		

github.com/h2oai/deepwater

# Data – Cat/Dog/Mouse Images

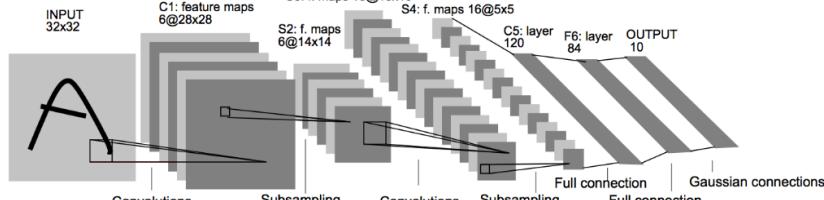


# Data – CSV

	A	B
1	bigdata/laptop/deepwater/imagenet/cat/102194502_49f003abd9.jpg	cat
2	bigdata/laptop/deepwater/imagenet/cat/11146807_00a5f35255.jpg	cat
3	bigdata/laptop/deepwater/imagenet/cat/1140846215_70e326f868.jpg	cat
4	bigdata/laptop/deepwater/imagenet/cat/114170569_6cbdf4bbdb.jpg	cat
5	bigdata/laptop/deepwater/imagenet/cat/1217664848_de4c7fc296.jpg	cat
6	bigdata/laptop/deepwater/imagenet/cat/1241603780_5e8c8f1ced.jpg	cat
7	bigdata/laptop/deepwater/imagenet/cat/1241612072_27ececbdef.jpg	cat
8	bigdata/laptop/deepwater/imagenet/cat/1241613138_ef1d82973f.jpg	cat
9	bigdata/laptop/deepwater/imagenet/cat/1244562192_35becd66bd.jpg	cat
10	bigdata/laptop/deepwater/imagenet/cat/125482638_e3688995e2.jpg	cat
11	bigdata/laptop/deepwater/imagenet/cat/128056573_1b4dc07c9a.jpg	cat
12	bigdata/laptop/deepwater/imagenet/cat/12945197_75e607e355.jpg	cat
13	bigdata/laptop/deepwater/imagenet/cat/132474673_88eaf528f5.jpg	cat
14	bigdata/laptop/deepwater/imagenet/cat/1350530984_ecf3039cf0.jpg	cat
15	bigdata/laptop/deepwater/imagenet/cat/1351606235_c9fbef634.jpg	cat
16	bigdata/laptop/deepwater/imagenet/cat/1356052454_e21f6e038b.jpg	cat
17	bigdata/laptop/deepwater/imagenet/cat/1388193613_06f57b76ff.jpg	cat

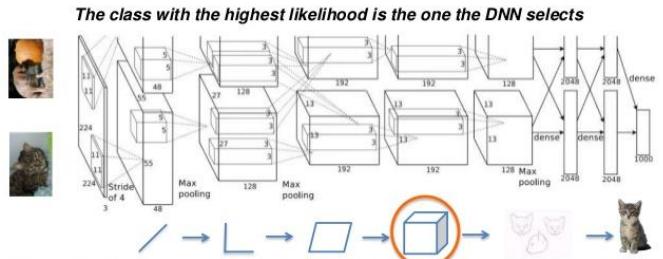
# Available Networks in Deep Water

- LeNet
- AlexNet
- VGGNet
- Inception (GoogLeNet)
- ResNet (Deep Residual Learning)
- Build Your Own



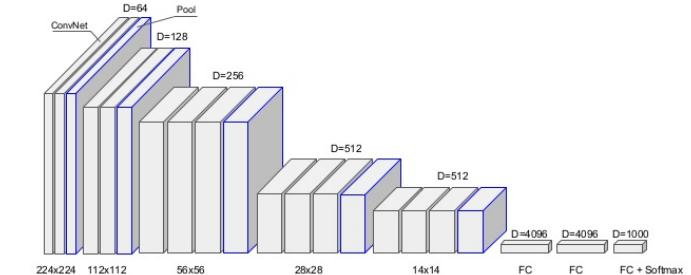
CNN called LeNet by Yann LeCun (1998)

AlexNet (Krizhevsky et al. 2012)

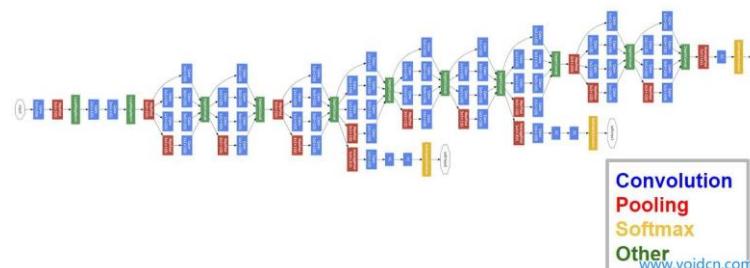


When AlexNet is processing an image, this is what is happening at each layer.

Classical CNN topology - VGGNet (2013)

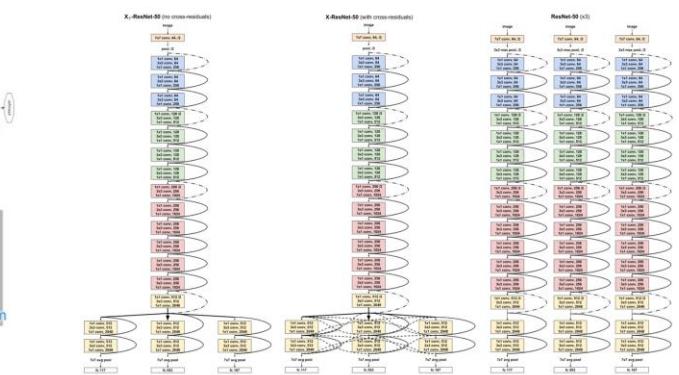


GoogLeNet



51

ResNet



# Want to try Deep Water?

- Build it
  - [bit.ly/h2o\\_deepwater](http://bit.ly/h2o_deepwater)
  - Ubuntu 16.04
  - CUDA 8
  - cuDNN 5
  - ...
- Pre-built Amazon Machine Images (AMIs)
  - Info to be confirmed

# Deep Water in Action

Quick Demo (5 minutes)

# H<sub>2</sub>O.ai Offers AI Open Source Platform Product Suite to Operationalize Data Science with Visual Intelligence



Visual Intelligence and UX Framework For Data Interpretation and Story Telling on top of Beautiful Data Products

100% Open Source



## Deep Water

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

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



H2O Integration with Spark. Best Machine Learning on Spark.

## Steam

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

# Want to find out more about Sparkling Water and Steam?

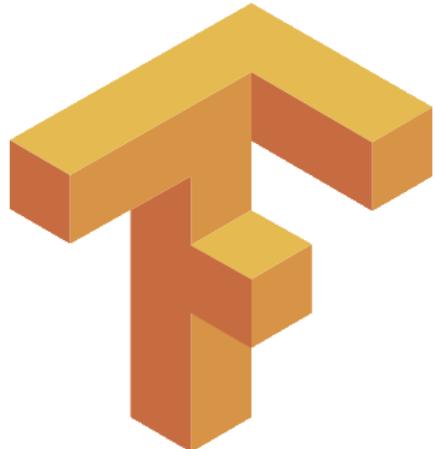
- R Addicts Paris
- Tomorrow 6:30pm
- Three H<sub>2</sub>O Talks:
  - Introduction to H<sub>2</sub>O
    - Demos: H<sub>2</sub>O + R + Steam
  - Auto Machine Learning using H<sub>2</sub>O
  - Sparkling Water 2.0

The screenshot shows the R Addicts Paris Meetup group page. The header reads "R Addicts Paris". The main navigation menu includes Home, Members, Sponsors, Photos, Discussions, and More. A "My profile" button is in the top right. On the left, there's a sidebar with the R Addicts logo, group stats (1,080 members), and links for About us..., Invite friends, and a calendar. The main content area features an event titled "Hors série: H2O" on Thursday, December 1, 2016, at 6:30 PM at NUMA PARIS (39, rue du Caire 75002, Paris). It notes that English speakers are welcome. To the right, there's a sidebar for RSVP status ("Your RSVP: Yes") and a list of attendees (150 going) with their profiles: Jo-fai Chow (Data Scientist at H2O.ai), François Guillemin (Organizer, Event Host), and Diane BELDAME (Co-Organizer, Event Host). At the bottom right is the H<sub>2</sub>O.ai logo.

# Conclusions

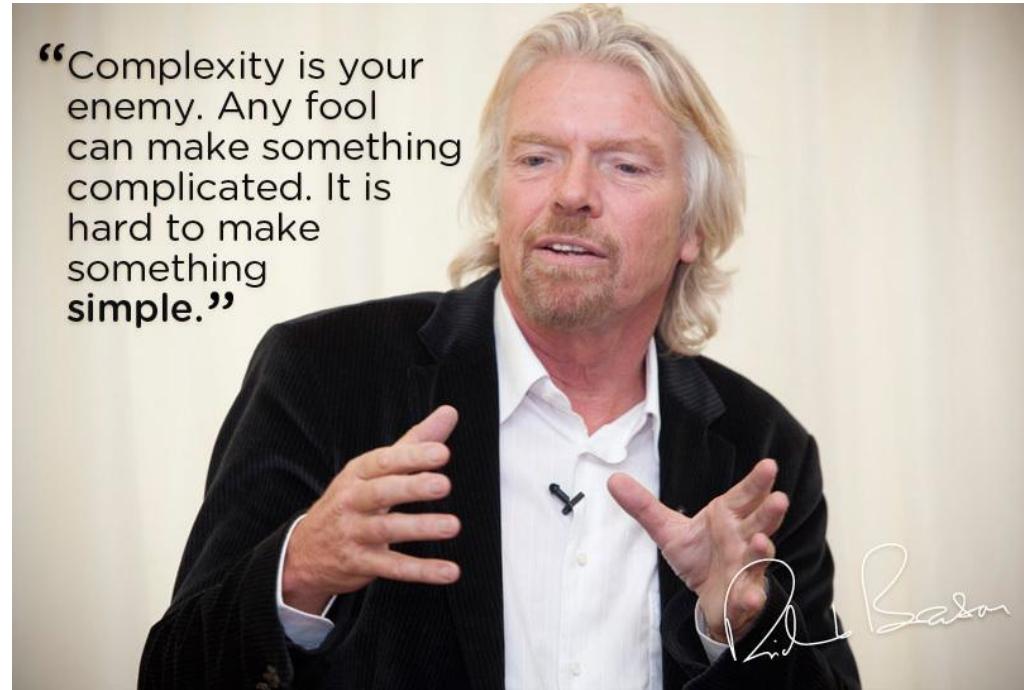
# Project “Deep Water”

- H<sub>2</sub>O + TensorFlow
  - a powerful combination of two widely used machine learning libraries.
- All Goodies from H<sub>2</sub>O
  - inherits all H<sub>2</sub>O properties in scalability, ease of use and deployment.
- Unified Interface
  - allows users to build, stack and deploy deep learning models from different DL libraries efficiently.
- 100% Open Source
  - the party will get bigger!



H<sub>2</sub>O.ai

# H<sub>2</sub>O's Mission



“Complexity is your enemy. Any fool can make something complicated. It is hard to make something simple.”

## Making Machine Learning Accessible to Everyone

*Photo credit: Virgin Media*

H<sub>2</sub>O.ai

# Deep Water – Current Contributors



Fabrizio Milo



Cyprien Noel



Qiang Kou



Arno Candel



Caffe



H<sub>2</sub>O.ai

H<sub>2</sub>O.ai

# Merci beaucoup!

- Organizers & Sponsors
  - Jiqiong, Natalia & Renat
  - Dailymotion
- Code, Slides & Documents
  - [bit.ly/h2o\\_meetups](http://bit.ly/h2o_meetups)
  - [bit.ly/h2o\\_deepwater](http://bit.ly/h2o_deepwater)
  - [docs.h2o.ai](http://docs.h2o.ai)
- Contact
  - [joe@h2o.ai](mailto:joe@h2o.ai)
  - [@matlabulous](https://twitter.com/matlabulous)
  - [github.com/woobe](https://github.com/woobe)



Haven't seen [this](#) before?

**H<sub>2</sub>O.ai**