H2O Around the World

London Kaggle Meetup

Strata Hadoop

Chelsea Football Club

PyData
Amsterdam

useR! 2016
Stanford

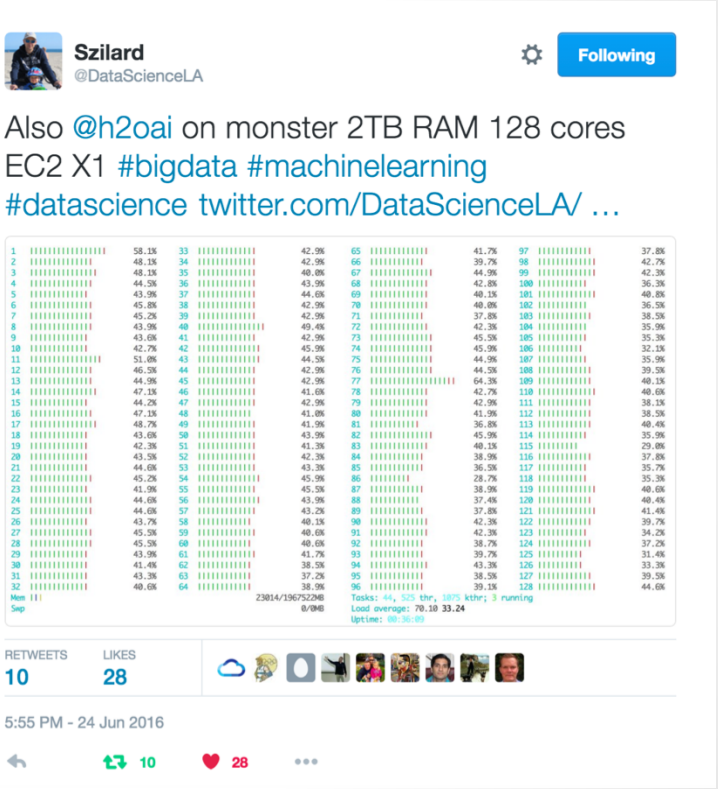
satRdays
Budapest

H2O Around the World

Paris Machine Learning Meetup (last week)

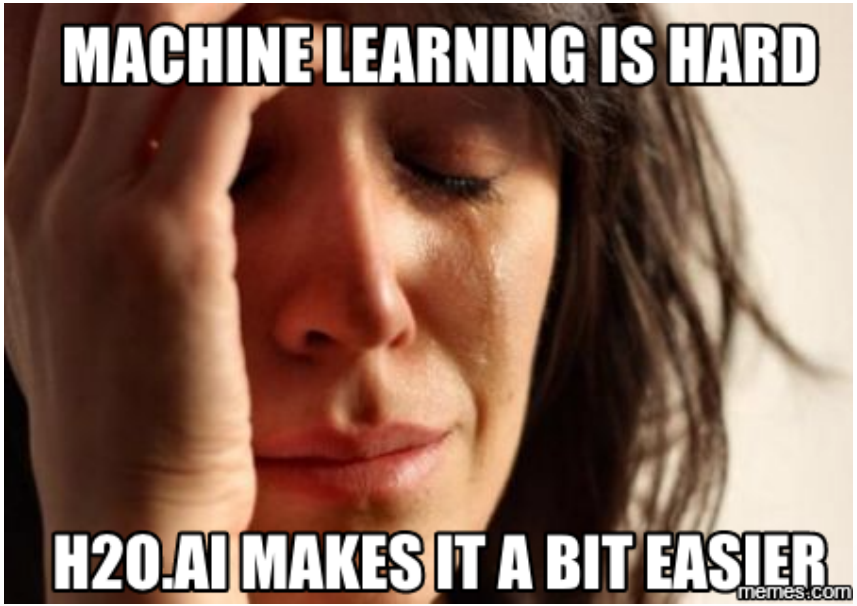
Szilard Pafka – Chief Data Scientist at Epoch

- Sziland's talks / blog posts about H2O:
  - ML Benchmark
  - Intro to ML with H2O
  - H2O Scoring
  - Tweets

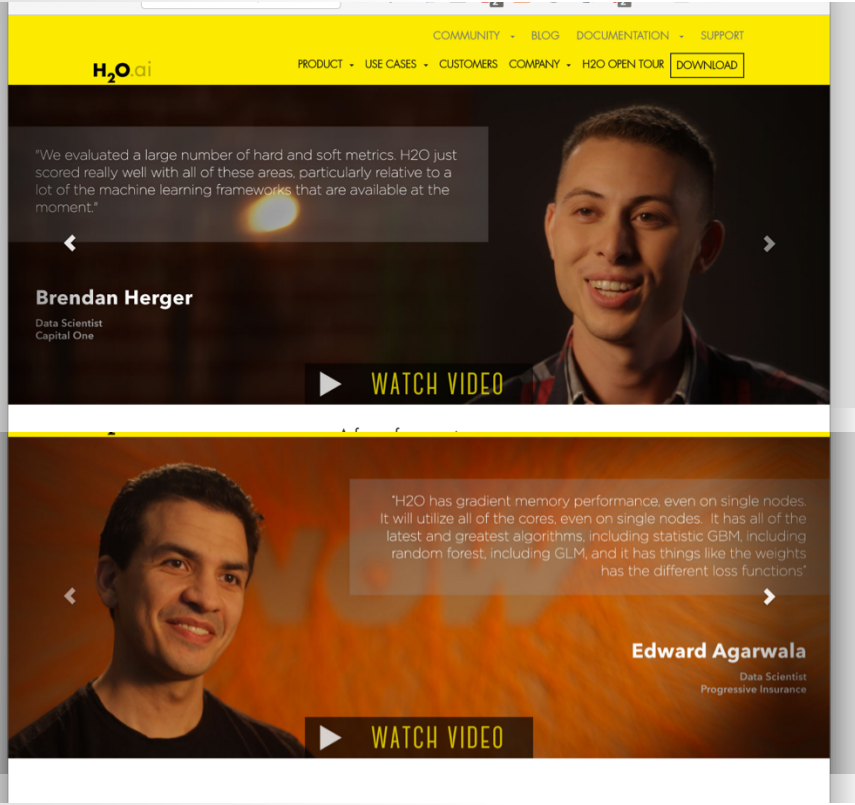


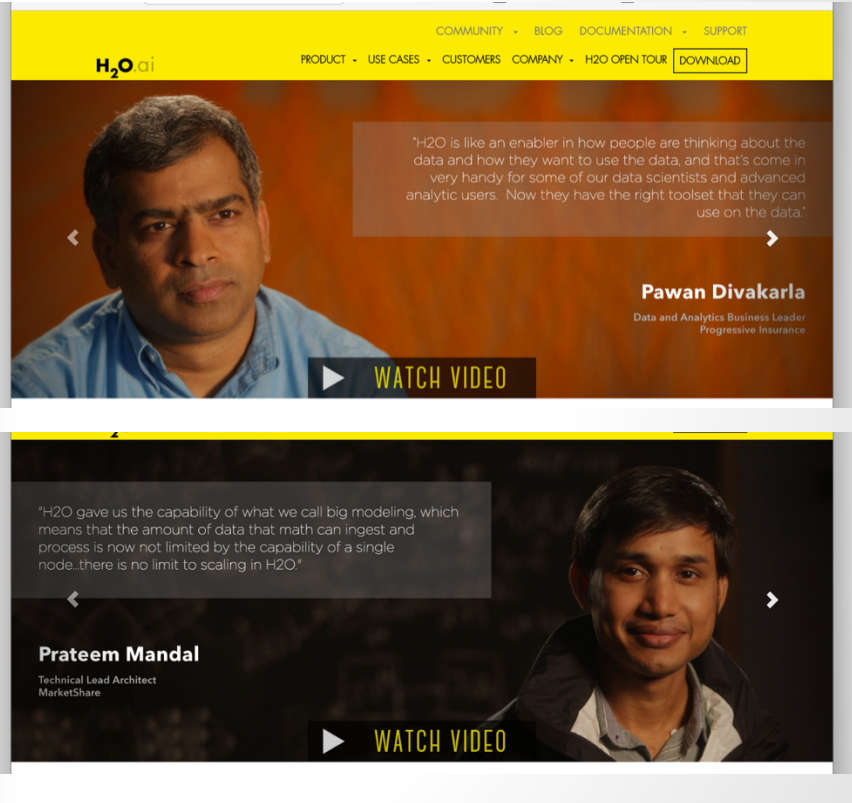
Szilard Pafka – Why H2O?

- Szilard's Summary Slide



H2O Customers

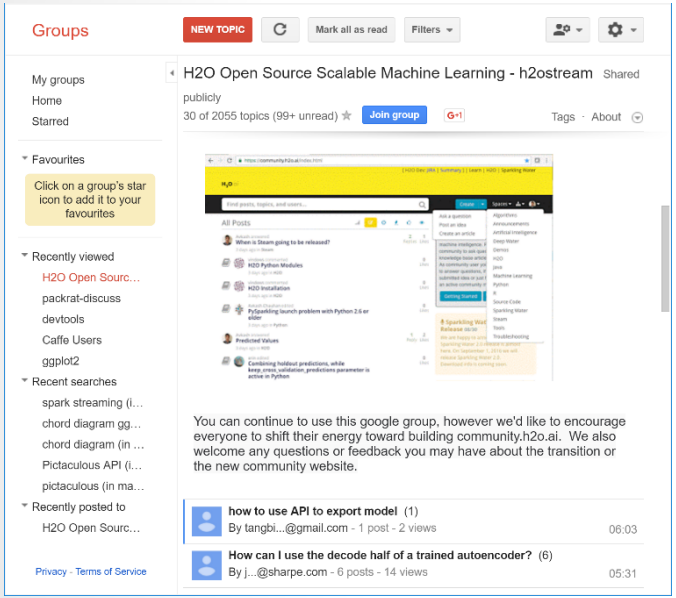




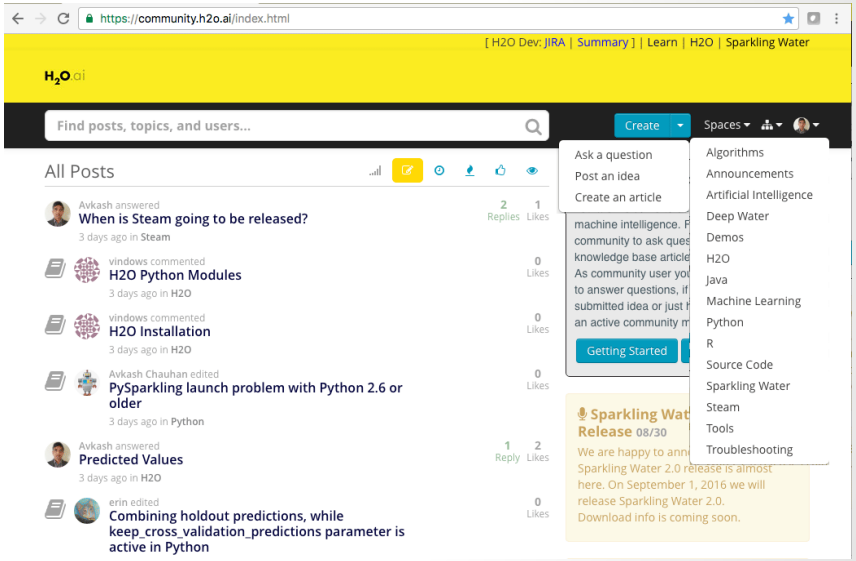
Check out the videos - **www.h2o.ai**

H2O Community Support

Google forum – **h2osteam**

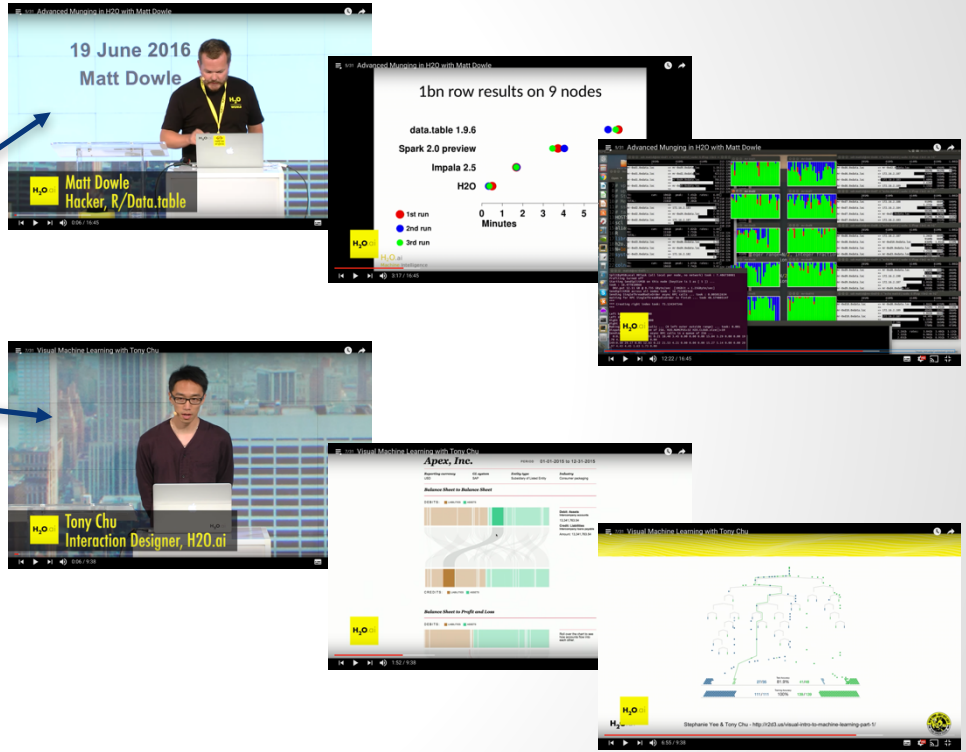


community.h2o.ai



H2O is Evolving

- H2O Open Tour NYC YouTube Playlist
  - Advanced data munging
  - Visual ML
  - Deep Water (2nd talk)
  - Sparkling Water (3rd talk)
  - Steam
    - New data science platform



Why Deep Water?

H2O 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 deeplearning-dd2f42f7-81f7-42e8-9d98-e34437309828
Description DeepLearning
Status RUNNING
Progress 69%

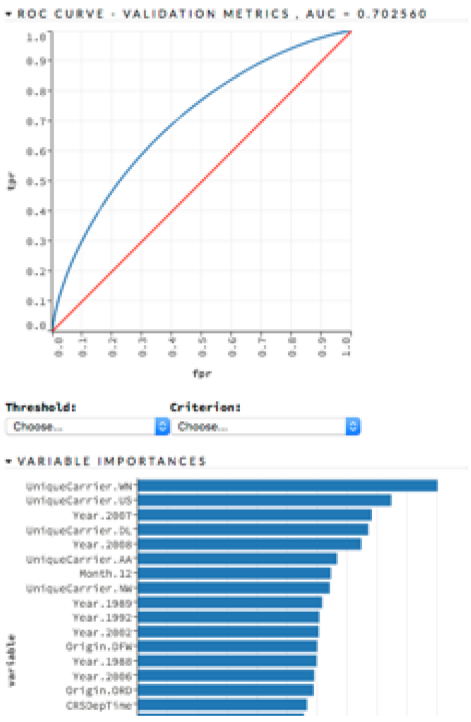
Iterations: 12. Epochs: 0.628821. Speed: 2,243,735 samples/sec. Estimated time left: 21.849 sec

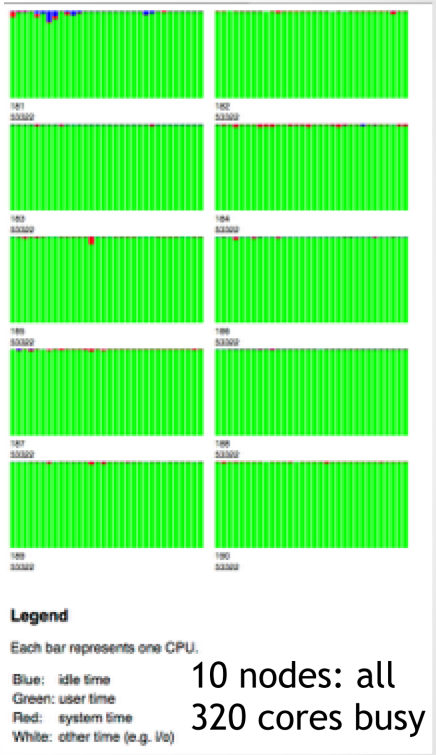
Actions: View | Cancel Job

model trained in <1 min:
2M+ samples/second

▾ OUTPUT - STATUS OF NEURON LAYERS (PREDICTING ISDEPDELAYED, 2-CLASS CLASSIFICATION, BERNOULLI DISTRIBUTION, CROSSENTROPY LOSS, 17,462 WEIGHTS/BIASES, 221.3 KB, 106,585,365 TRAINING SAMPLES, MINI-BATCH SIZE 1)

layer	units	type	dropout	l1	l2	mean_rate	rate_RMS	momentum	mean_weight	weight_RMS	mean_bias	bias_RMS
1	807	Input	0									
2	20	Rectifier	0	0	0	0.0403	0.2020	0	-0.0021	0.2111	-0.9139	1.0036
3	20	Rectifier	0	0	0	0.0197	0.0227	0	-0.1033	0.5362	-1.3988	1.5259
4	20	Rectifier	0	0	0	0.0517	0.0446	0	-0.1575	0.3068	-0.0846	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	0.0161	0.0083	0	0.0741	0.7260	0.4269	0.2056





10 nodes: all
320 cores busy

real-time, interactive
model inspection in Flow

H2O.ai

Deep Learning Model

H2O Deep Learning Community Quotes

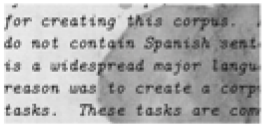
CIFAR-10 Competition Winners: Interviews with Dr. Ben Graham, Phil Culliton, & Zygmunt Zając

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



Completed • Knowledge • 161 teams

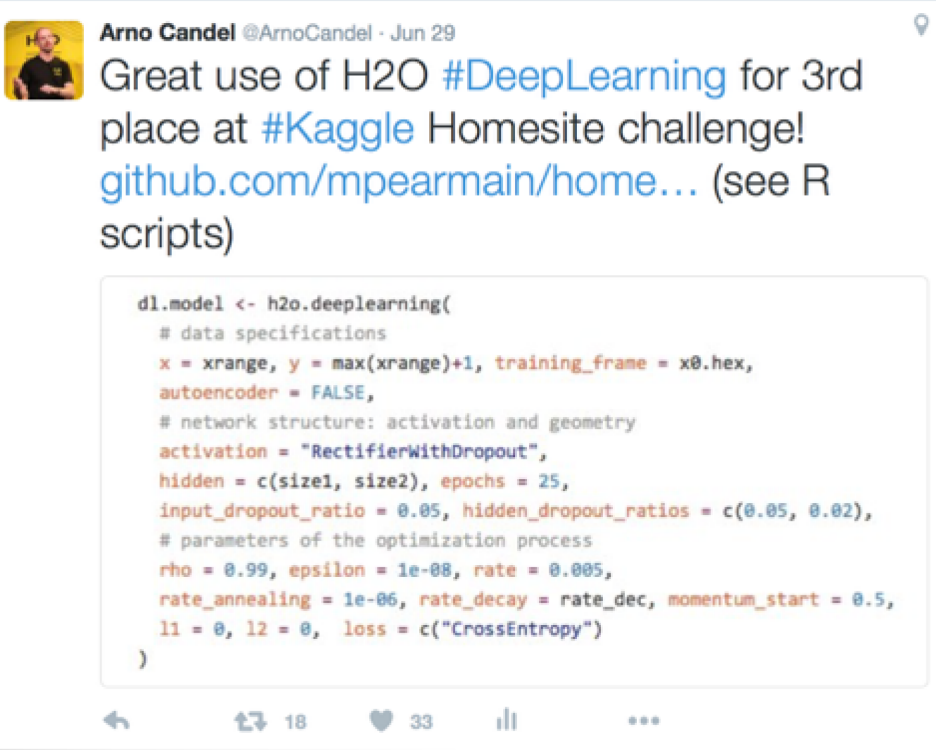
Denoising Dirty Documents

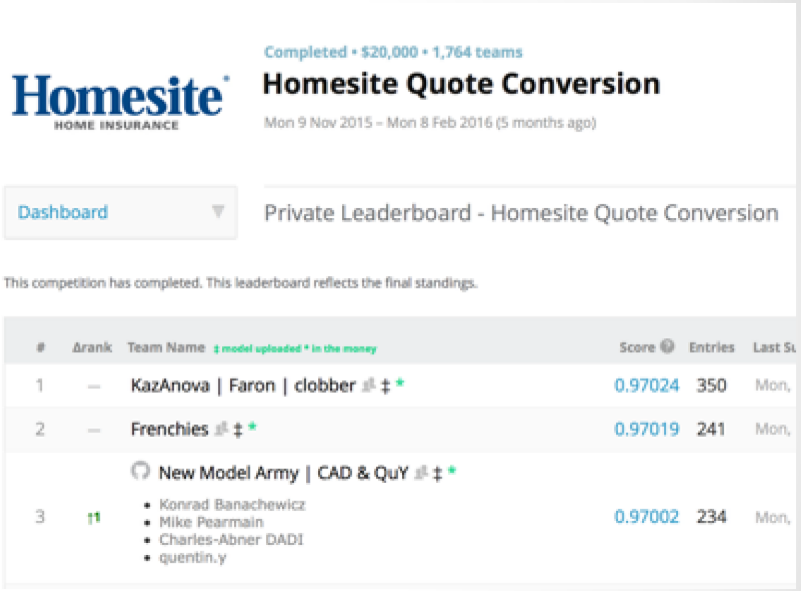
Mon 1 Jun 2015 – Mon 5 Oct 2015 (3 months ago)

READ MORE

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

H2O Deep Learning Community Quotes





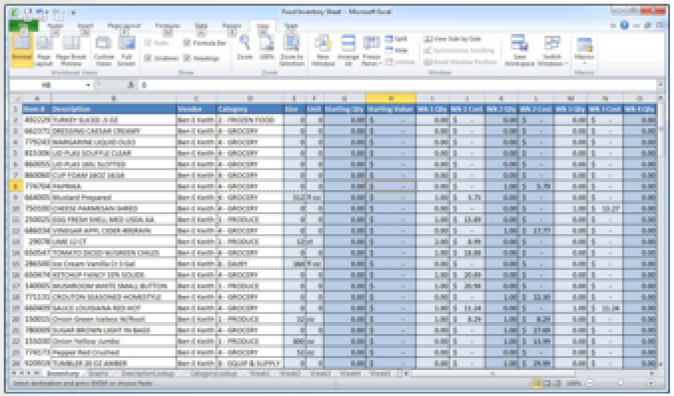
H2O.ai

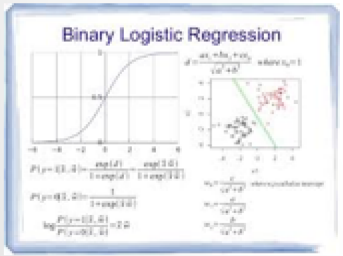
Deep Water: Best Open-Source Deep Learning

Enterprise Deep Learning for Business Transformation

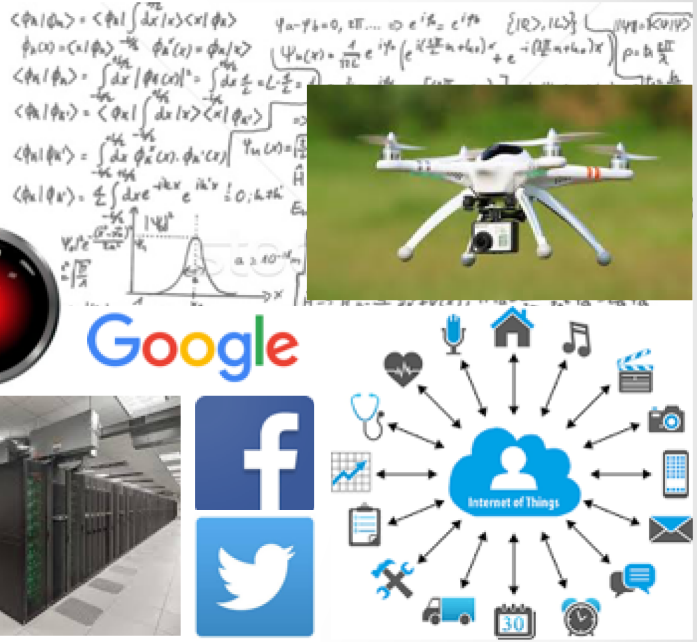
Deep Water = THE Deep Learning Platform	H2O integrates the top open-source DL tools 
Native GPU support	 is up to 100x faster than 
Enterprise Ready	Easy to train and deploy, interactive, scalable, etc. Flow, R, Python, Spark/Scala, Java, REST, POJO, **Steam**
New Big Data Use Cases (previously impossible or difficult in H2O)	**Image** - social media, manufacturing, healthcare, … **Video** - UX/UI, security, automotive, social media, … **Sound** - automotive, security, call centers, healthcare, … **Text** - NLP, sentiment, security, finance, fraud, … **Time Series** - security, IoT, finance, e-commerce, …

More Data is Better! — Images, Video, Text, Logs, Streams, …





Example:
Fraud Prediction



Today → Tomorrow

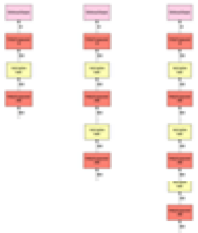
Better Data — Better Models — Better Results

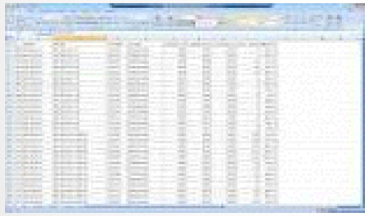
H2O.ai

Deep Water opens the Floodgates for state-of-the-art Deep Learning

H2O Deep Learning:
simple multi-layer neural networks

1-5 layers
MBs/GBs of data



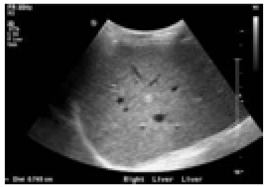


AMD
intel

Limited to business analytics, statistical models (CSV data)

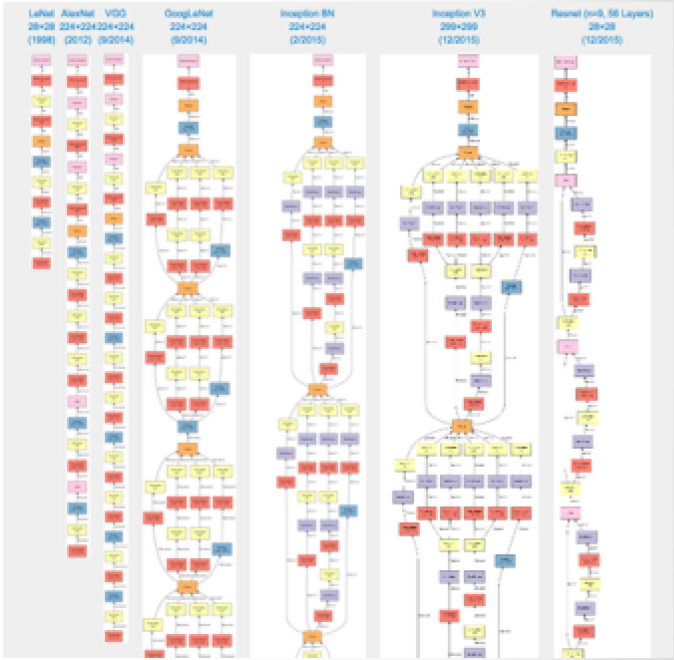
Deep Water: deep complex networks

5-1000 layers
GBs/TBs of data



AMD
intel

NVIDIA
CUDA



Large networks for big data
(e.g. image 1000x1000x3 -> 3m inputs)

H2O.ai

Current Contributors (more H2O.ai folks joining soon)

Fabrizio Milo

Cyprien Noel

Qiang Kou

Arno Candel

TensorFlow

Caffe

mxnet

H2O.ai

This repository Search

h2oai / deepwater

Deep Water Demos

- **H2O + mxnet**
  - Dataset:
    - Cat / Dog / Mouse
  - H2O Python interface
  - mxnet GPU backend
  - Train a LeNet (CNN) model
  - Explore model in Flow
- **H2O + TensorFlow**
  - Dataset:
    - MNIST hand-written digits
  - Sparkling Water + Jupyter Notebook
  - Convert TensorFlow model into H2O
  - Explore models in Flow

For Online Audience

- **H2O + mxnet**
  - bit.ly/h2o_paris_1
  - **demo_03_mxnet**
- **H2O + TensorFlow**
  - bit.ly/h2o_paris_1
  - **demo_04_tensorflow**