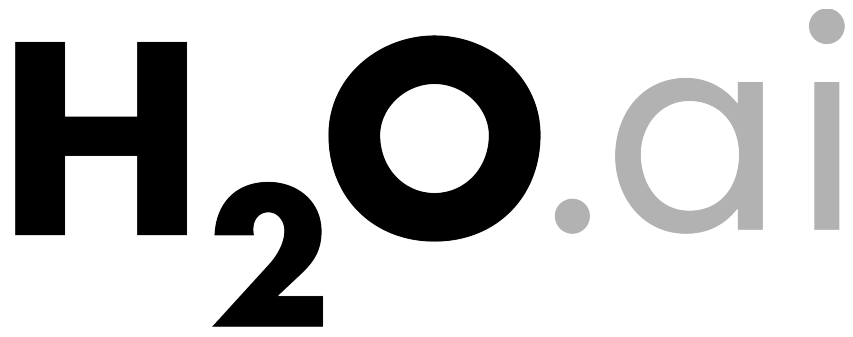
Introduction to Scalable Machine Learning with H_2O



Jo-fai (Joe) Chow

Data Scientist

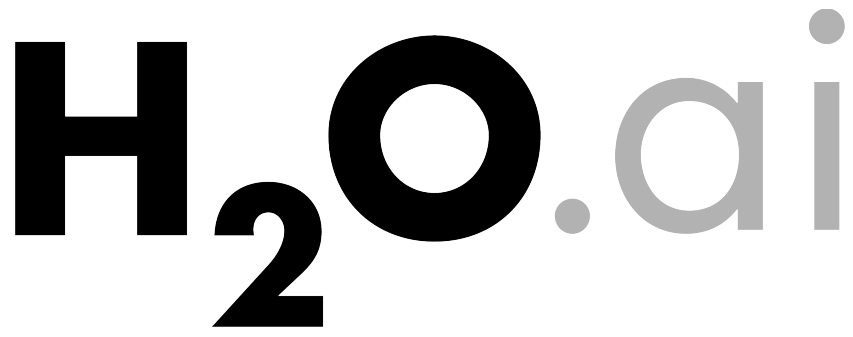
joe@h2o.ai

@matlabulous

Scotland Data Science and Technology Meetup

30th November, 2017

Introduction to Scalable Machine Learning with H_2O



Jo-fai (Joe) Chow

Data Scientist

joe@h2o.ai

@matlabulous

All slides, data and code examples

http://bit.ly/h2o_meetups

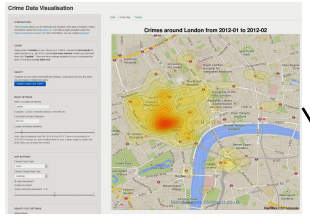
Agenda

- Overview
  - Company / People
- Open Source H_2O Machine Learning Platform
  - Overview
  - Demos
    - H_2O on Hadoop
    - Automatic Machine Learning
- Q&A

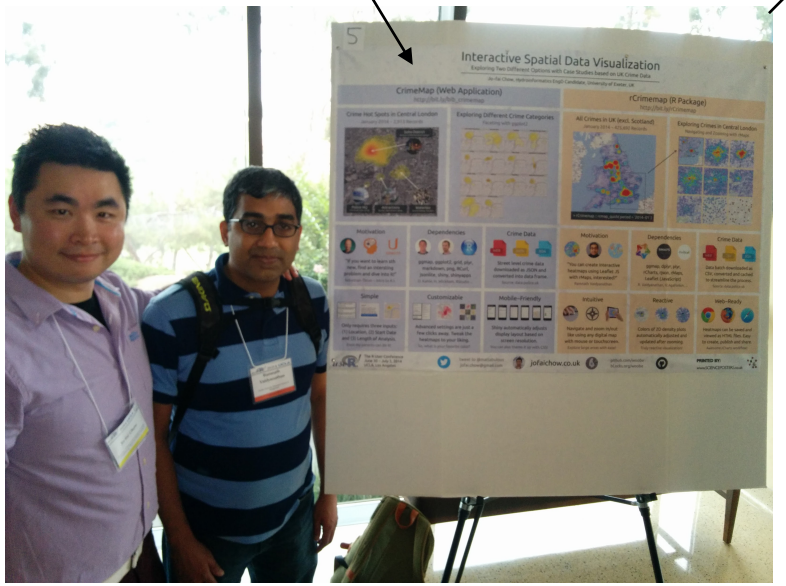
About Me

- Civil (Water) Engineer
  - 2010 – 2015
    - Consultant (UK)
      - Utilities
      - Asset Management
      - Constrained Optimization
    - EngD (Industrial PhD) (UK)
      - Infrastructure Design Optimization
      - Machine Learning + Water Engineering
      - Discovered H_2O in 2014
- Data Scientist
  - 2015 – 2016
    - Virgin Media (UK)
    - Domino Data Lab (Silicon Valley)
  - 2016 – Present
    - H_2O.ai (Silicon Valley)
  - How?
    - bit.ly/joe_kaggle_story

useR! 2014 Conference in US



CrimeMap -> poster



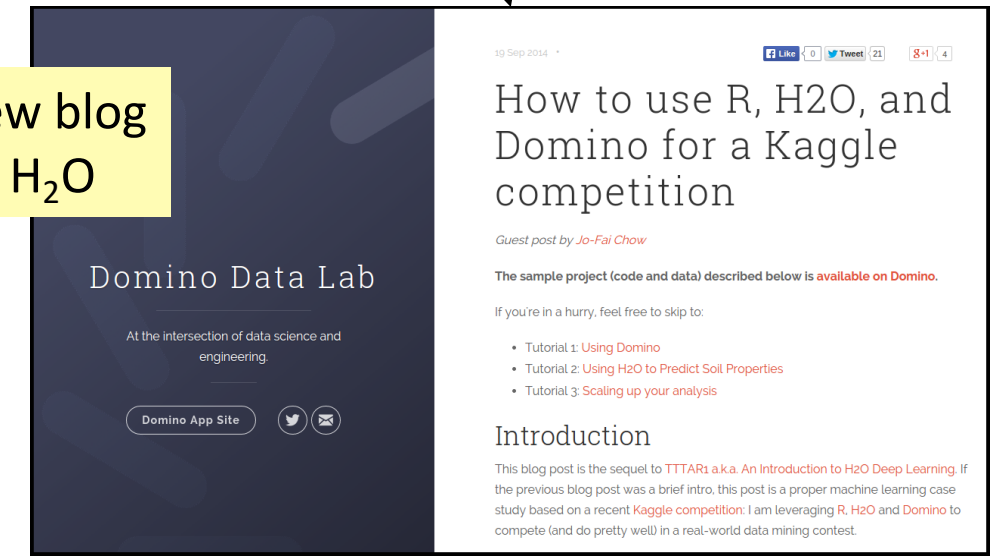
H2O.ai
@h2oai
Following

#linus for #rstat #user2014 #JohnChambers

John Chambers mentioned H_2O

… so I gave it a try

and wrote a few blog posts about H_2O



The Long Story:
bit.ly/joe_kaggle_story

H2O.ai

Company Overview

Founded	2011 Venture-backed, debuted in 2012
Products	• **H_2O Open Source In-Memory AI Prediction Engine** • Sparkling Water (H_2O + Spark) • Enterprise Steam • Driverless AI
Mission	Operationalize Data Science, and provide a platform for users to build beautiful data products
Team	75+ employees • Distributed Systems Engineers doing Machine Learning • World-class visualization designers
Headquarters	Mountain View, CA





H2O.ai

Scientific Advisory Council

Dr. Trevor Hastie

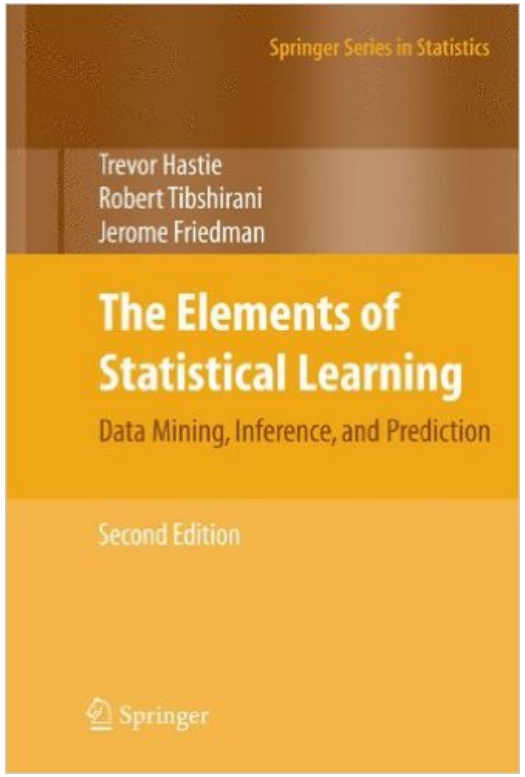
- John A. Overdeck Professor of Mathematics, Stanford University
- PhD in Statistics, Stanford University
- Co-author, *The Elements of Statistical Learning: Prediction, Inference and Data Mining*
- Co-author with John Chambers, *Statistical Models in S*
- Co-author, *Generalized Additive Models*

Dr. Robert Tibshirani

- Professor of Statistics and Health Research and Policy, Stanford University
- PhD in Statistics, Stanford University
- Co-author, *The Elements of Statistical Learning: Prediction, Inference and Data Mining*
- Author, *Regression Shrinkage and Selection via the Lasso*
- Co-author, *An Introduction to the Bootstrap*

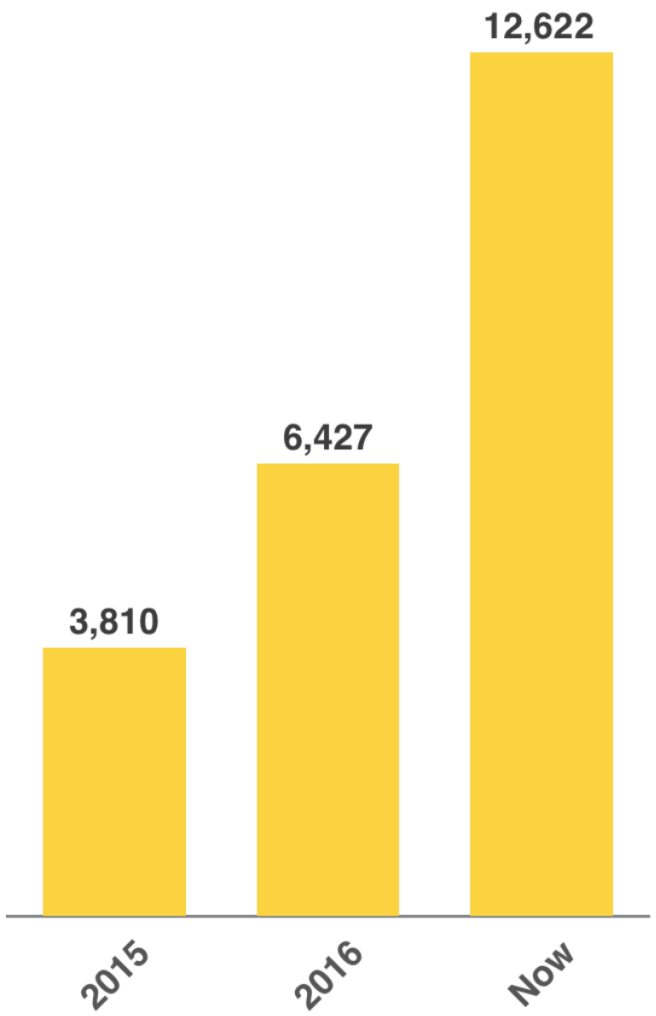
Dr. Steven Boyd

- Professor of Electrical Engineering and Computer Science, Stanford University
- PhD in Electrical Engineering and Computer Science, UC Berkeley
- Co-author, *Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers*
- Co-author, *Linear Matrix Inequalities in System and Control Theory*
- Co-author, *Convex Optimization*



H2O.ai

Companies Using H2O.ai



222 OF THE FORTUNE 500

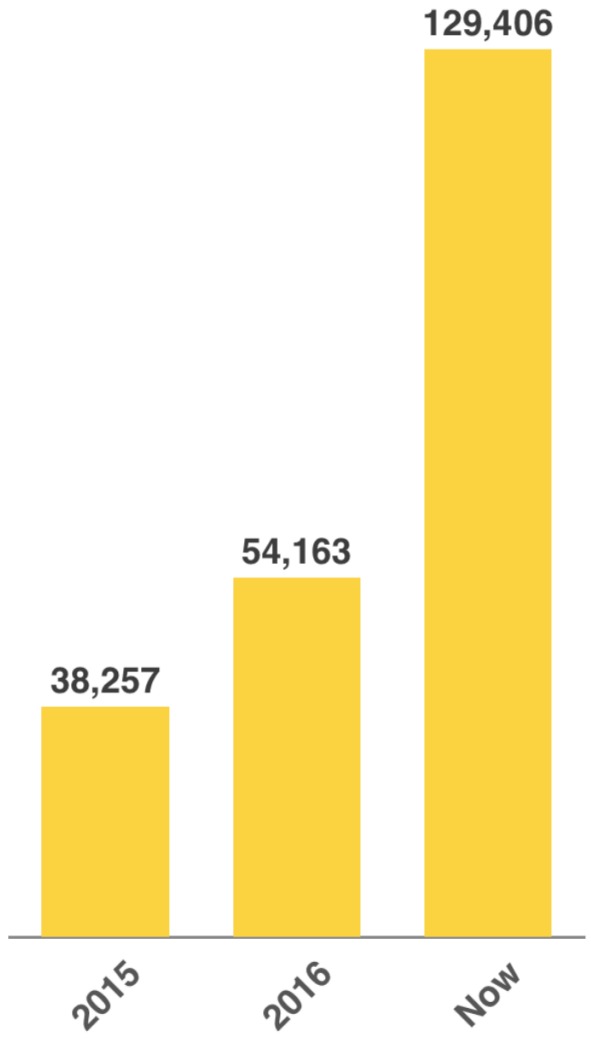
H2O

8 OF TOP 10 BANKS

7 OF TOP 10 INSURANCE COMPANIES

4 OF TOP 10 HEALTHCARE COMPANIES

H2O.ai Users



H2O.ai

Community Expansion

15 Meetups a month

66,843	33	50	45	18
members	interested	Meetups	cities	countries

Find out more: **www.h2o.ai/community/**

Select Customers

"Overall customer satisfaction is very high." - Gartner

Award-Winning Application (PwC + H_2O.ai)

pwc | Global

PwC Global › About us › Global Annual Review 2017 › Stories about our people, clients and communities

Harnessing the power of AI to transform the detection of fraud and error

Setting the scene

PwC has made a significant investment in pioneering artificial intelligence (AI) for the audit. For the past 18 months, we have partnered with H2O.ai – a leading Silicon Valley company – to build a revolutionary bot that uses AI and machine learning to 'x-ray' a business, analysing billions of data points in milliseconds, seeing what humans can't, and applying judgement to detect anomalies in the general ledger. Called GL.ai, it is the first module of PwC's Audit.ai.

GL.ai was named the 'Audit Innovation of the Year' by the International Accounting Bulletin in October 2017.

The Accountant International Accounting Bulletin
Digital Accountancy Forum and Awards
Audit Innovation of the Year
PwC's GL.ai
Winner 2017

"The reason this is such a brilliant tool is its ability to look at different risks, in context, at the same time. For example, it would be uneconomical for an auditor to look at every single user's pattern of activity to decide what's unusual. With GL.ai, the algorithms do it for us."

Laura Needham partner, PwC UK

PwC
@PwC
Follow

Exciting night at this year's @WAI_News Awards: PwC wins 2017 Audit Innovation of the Year! pwc.to/Glai17 #taandiab17

10:15 PM - 4 Oct 2017

http://www.pwc.com/gx/en/about/stories-from-across-the-world/harnessing-the-power-of-ai-to-transform-the-detection-of-fraud-and-error.html

H_2O.ai

Partners

NVIDIA's (NASDAQ: NVDA) invention of the GPU in 1999 sparked the growth of the PC gaming market, redefined modern computer graphics, and revolutionized parallel computing. More recently, GPU deep learning ignited modern AI -- the next era of computing -- with the GPU acting as the brain of computers, robots, and self-driving cars that can perceive and understand the world.
Website

Founded in 1975, Microsoft (Nasdaq "MSFT") is the worldwide leader in software, services, devices and solutions that help people and businesses realize their full potential. You can launch Sparkling Water on Azure HDInsights with just a few clicks to build your data science pipeline on the cloud.
Website | Documentation

Watson Data Platform enables AI-powered decision making by simplifying and automating the development and operationalization of new insights. It enables unprecedented levels of collaboration amongst data savvy professionals.

Kensu's mission is to lift Data Science to the Enterprise level focusing on the production environment and the maintenance across time.
Website

SigOpt is the optimization platform that amplifies your research. SigOpt takes any research pipeline and tunes it, right in place.
Website

Nimbix is the leading provider of purpose-built cloud computing for machine learning, AI and HPC applications. Powered by JARVICE™, the Nimbix Cloud provides high-performance software as a service, dramatically speeding up data processing for Energy, Life Sciences, Manufacturing, Media and Analytics applications.
Website

Cloudera delivers a modern platform for data management and analytics, helping businesses solve their most challenging problems with data.

Anaconda is the leading open data science platform powered by Python. Continuum Analytics is the creator and driving force behind Anaconda. We put superpowers into the hands of people who are changing the world.

The MapR Converged Data Platform integrates enormous power of Hadoop and Spark with global event streaming, real-time database capabilities, and enterprise storage.

Splunk Inc. provides the leading platform for Operational Intelligence. Customers use Splunk to search, monitor, analyze and visualize machine data.
Website

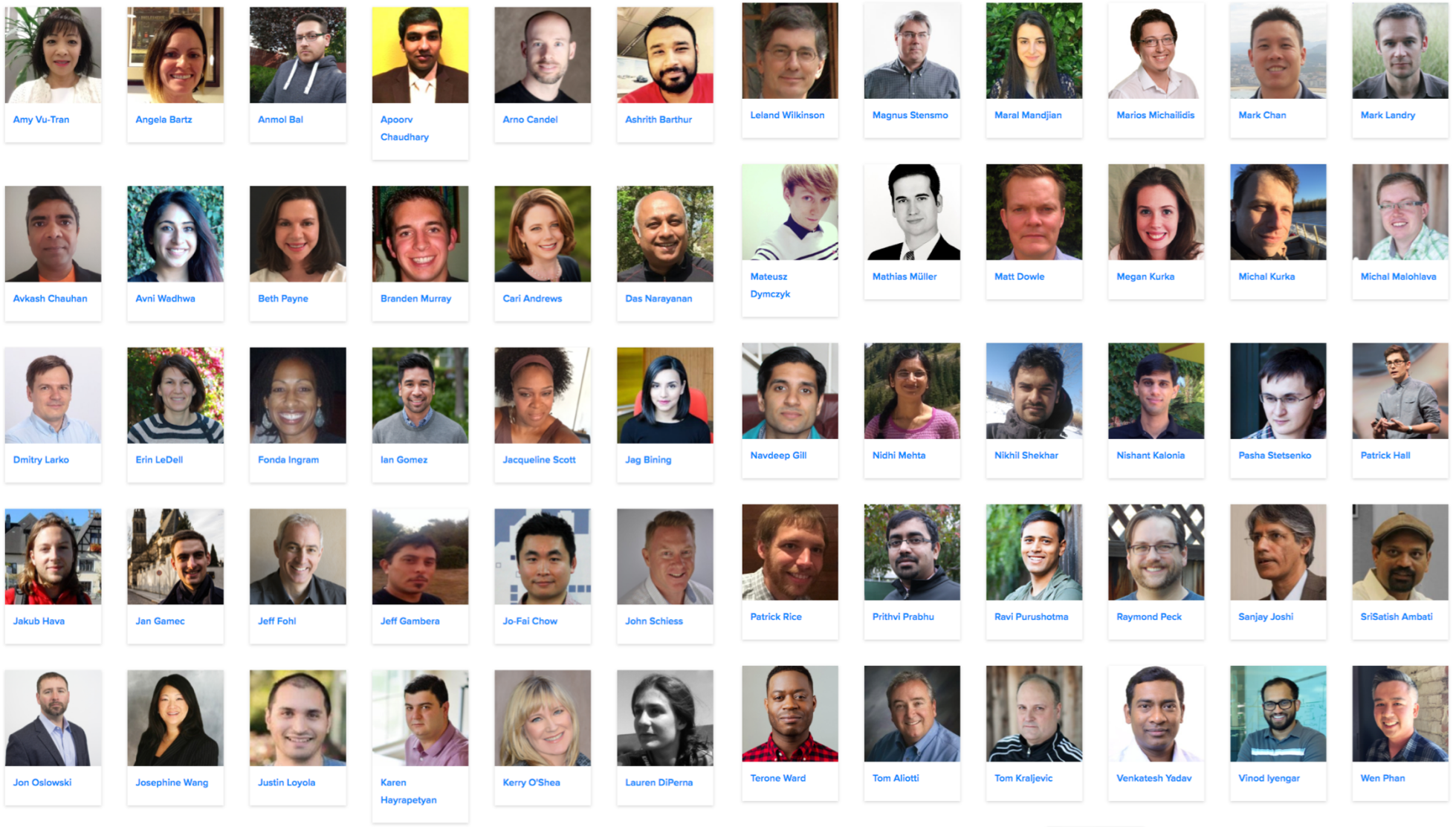
DataScience.com pairs data expertise with powerful tools to help businesses unlock the value in their data. Enabling data science for every business.
Website

MapD makes queries faster and creates a fluid and immersive data exploration experience that removes the disconnect between an analyst and their data. Making extracting insight from data effortless and lightning fast.
Website

Databricks provides a just-in-time data platform, to simplify data integration, real-time experimentation, and robust deployment of production applications.

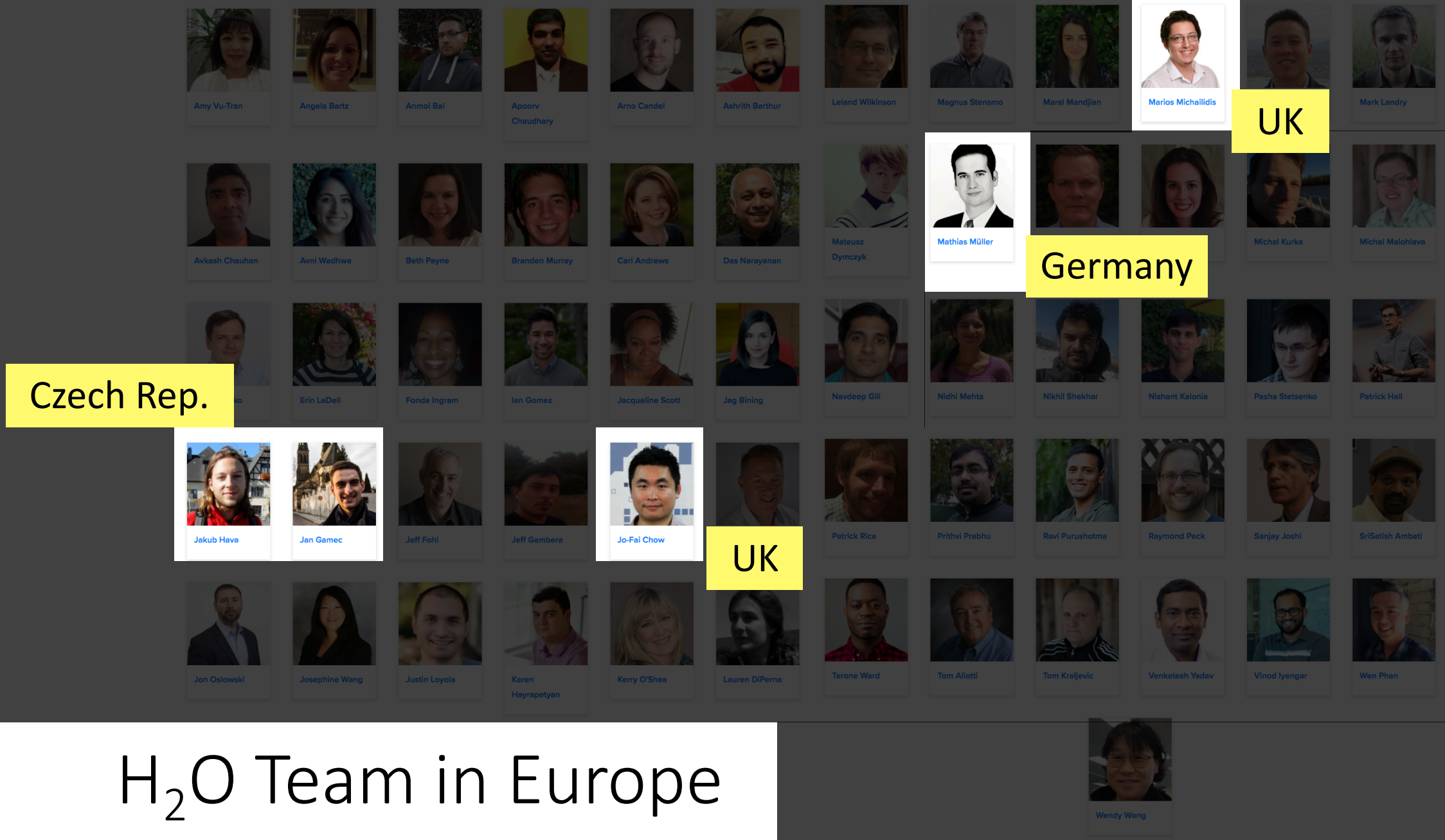
Hortonworks drives actionable intelligence with Connected Data Platforms that maximize the value of all data—data-in-motion and data-at-rest.

Minio is an object storage server built for cloud application developers and devops.



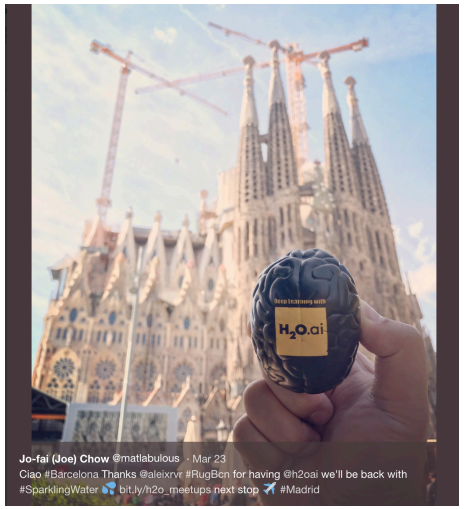
H_2O Team

H_2O.ai



H_2O Team in Europe

What is Joe's role at H_2O.ai?



- Data Scientist / Sales Engineer / Speaker / Meetup Organiser / Community Growth Hacker (on paper)
- Unofficial Photographer of H_2O.ai SWAG (the travelling data scientist)
- H_2O.ai SWAG EMEA Distributor (please help yourself)

H_2O.ai

#AroundTheWorldWithH2Oai

Month	Joe's H_2O Events in 2017
Jan	London
Feb	London, Warsaw, Oxford
Mar	Bay Area, London, Cologne, Barcelona, Madrid, Vienna
Apr	Amsterdam, Rotterdam, Poznan, London
May	Belgrade, Hamburg, Berlin
Jun	Amsterdam, Stockholm, Budapest, London, Munich, Prague
Jul	Berlin, Brussels
Aug	🏖
Sep	London, Dublin
Oct	Exeter, Munich, Dublin, The Hague, Amsterdam, Frankfurt
Nov	Munich, Zurich, London, **Glasgow (TODAY)**
Dec	**Bay Area (Next Week)**, London, 🎄

25+ Cities in 2017

H2O.ai

Meetup

Create a Meetup | Explore | Messages | Notifications

London Artificial Intelligence & Deep Learning

PRO

H2O Artificial Intelligence and Machine Learning -
39 groups

Location
London, United Kingdom

Members
4,354

4000+ Members

Organizers
Ian Gomez and **2 others**

Schedule

Our group | Meetups | Members | Photos | Discussions | More

Next Meetup

Next London Meetup: **12 Dec**

See all

12 DEC

Tuesday, December 12, 2017, 6:00 PM

Interpretable Machine Learning, Tweet Classifier, H2O World Highlights and More

Hosted by Jo-fai Chow

17

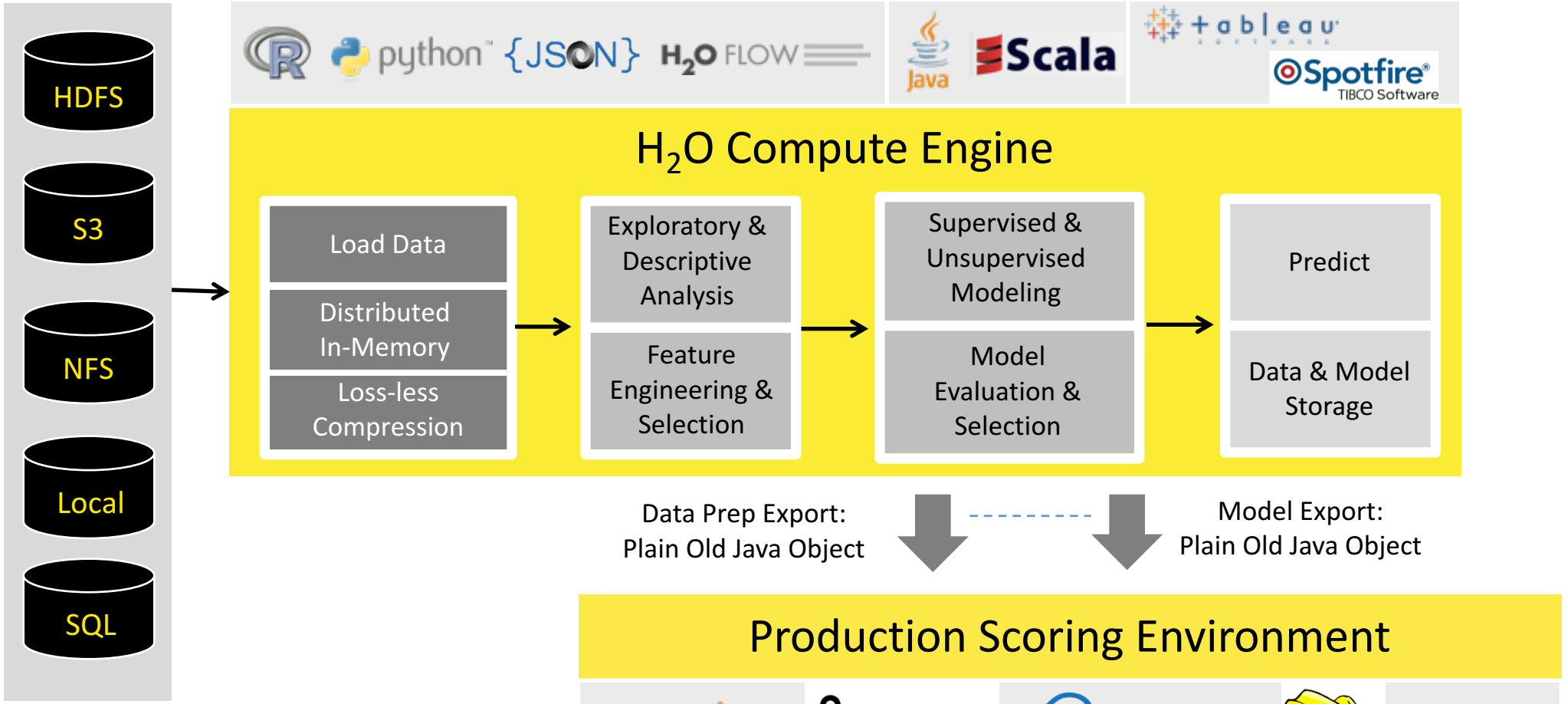
Edit

Moody's Analytics
1 Canada Square, Canary Wharf, E14 5AB · London

Dear All, We are thrilled to see the growth of this meetup group (4100+ members right now).

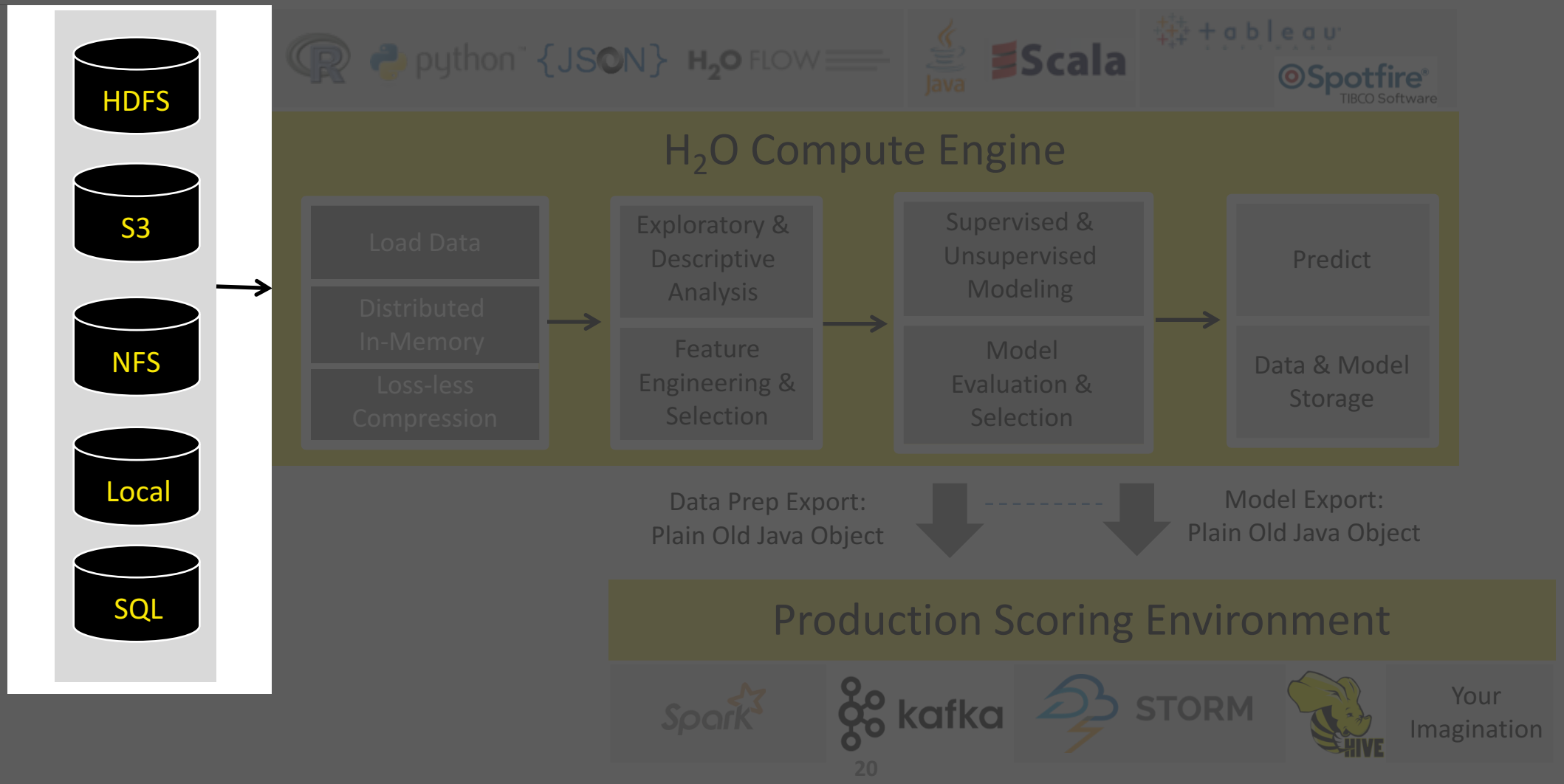
H_2O Machine Learning Platform

High Level Architecture



High Level Architecture

Import Data from Multiple Sources

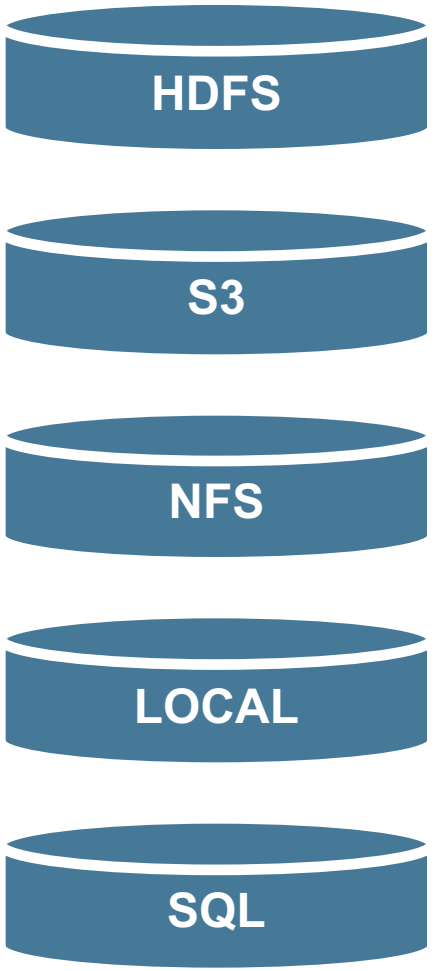


Supported Formats & Data Sources

9 Formats

File type or Folder of Files

- CSV
- XLS
- XLSX
- ORC*
- Hive*
- SVMLight
- ARFF
- Parquet
- Avro 1.8.0*



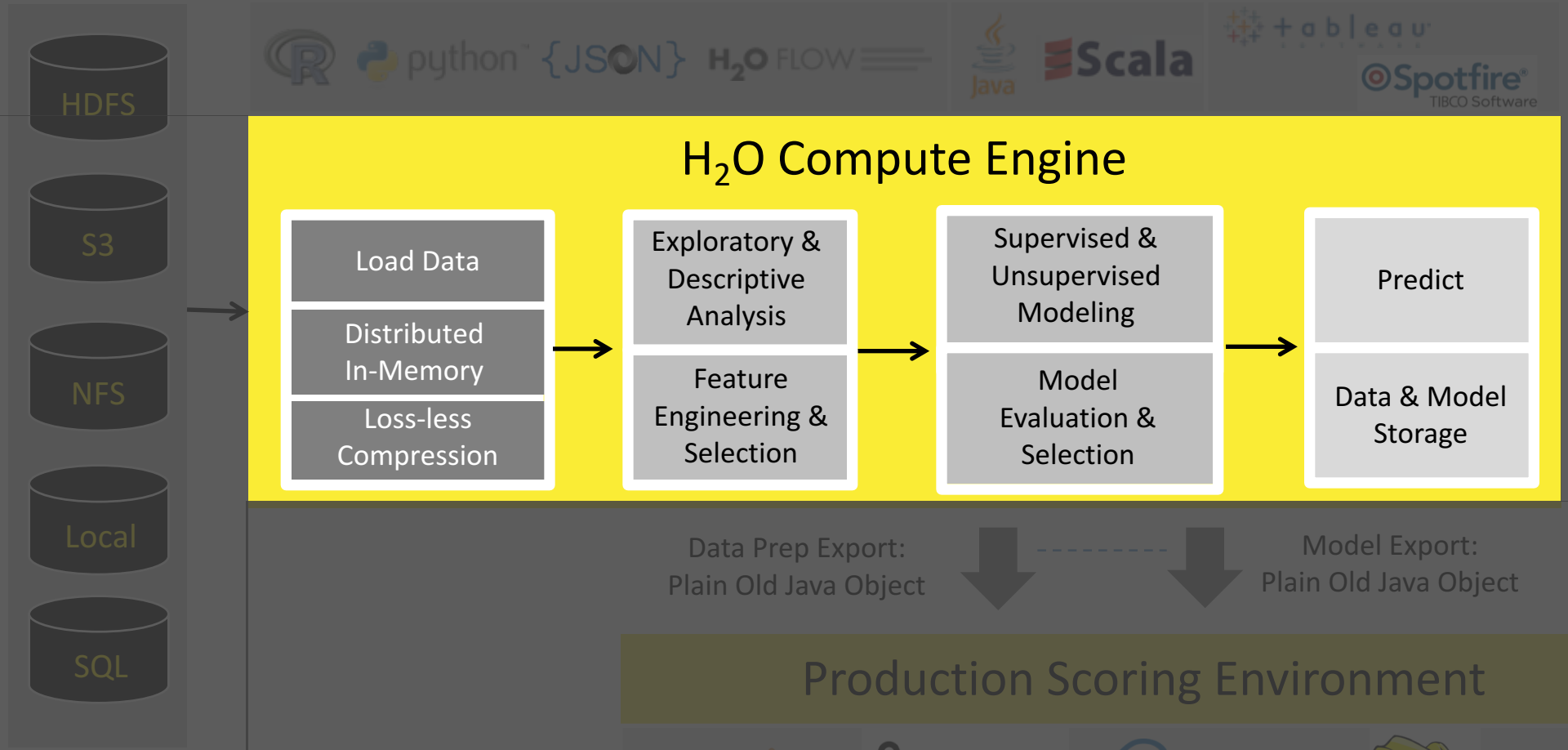
5 Sources

* 1. only if H2O is running as a Hadoop job
* 2. Hive files that are saved in ORC format
* 3. without multi-file parsing or column type modification

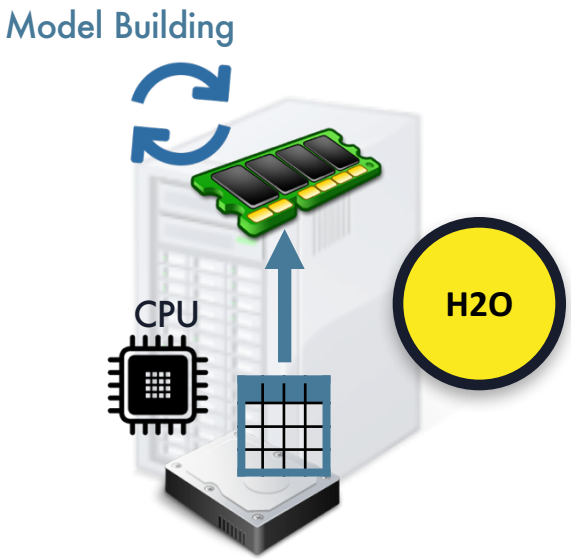
H2O.ai

High Level Architecture

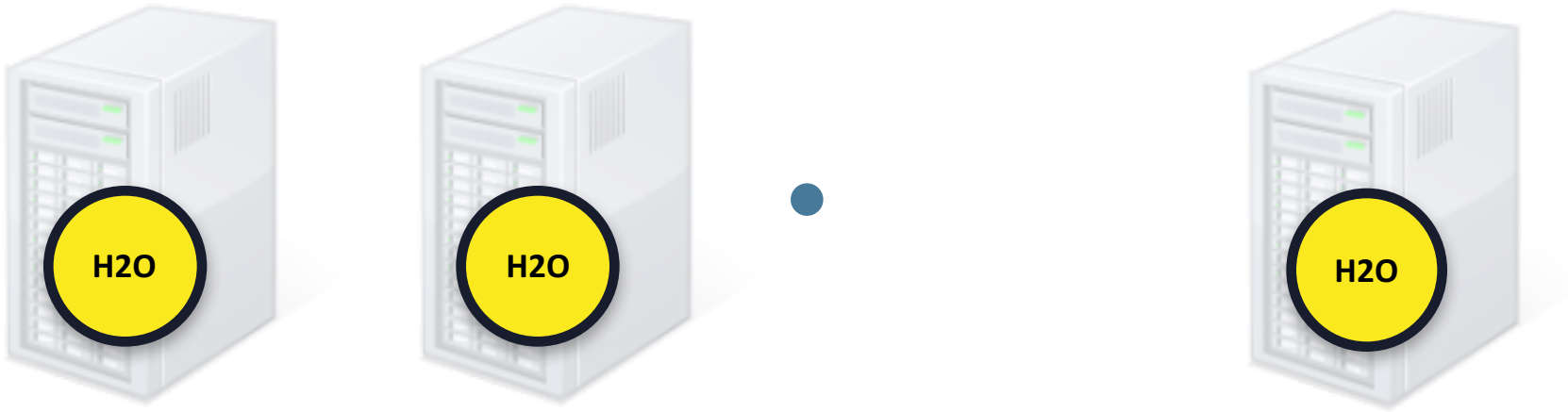
Fast, Scalable & Distributed Compute Engine Written in Java



H_2O Core



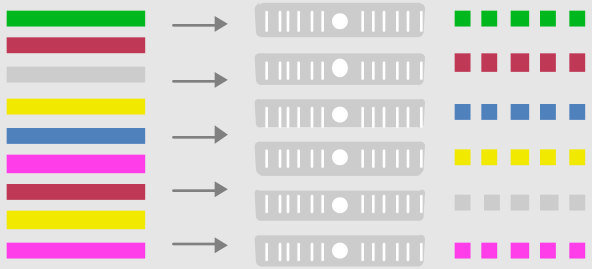
H_2O Core



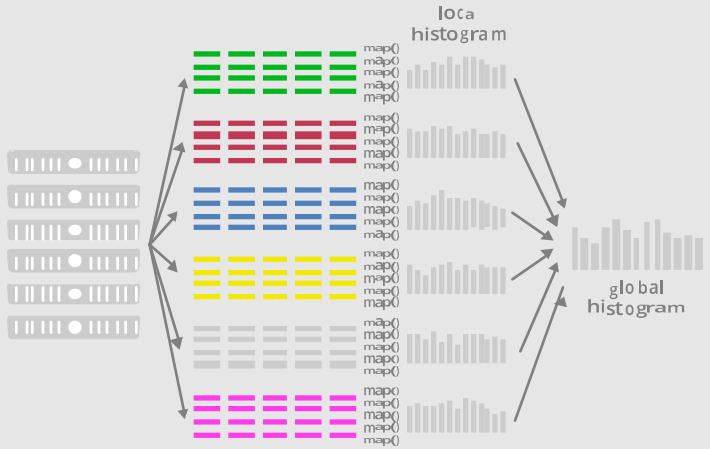
H_2O Core

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_2O: 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**

Algorithms Overview

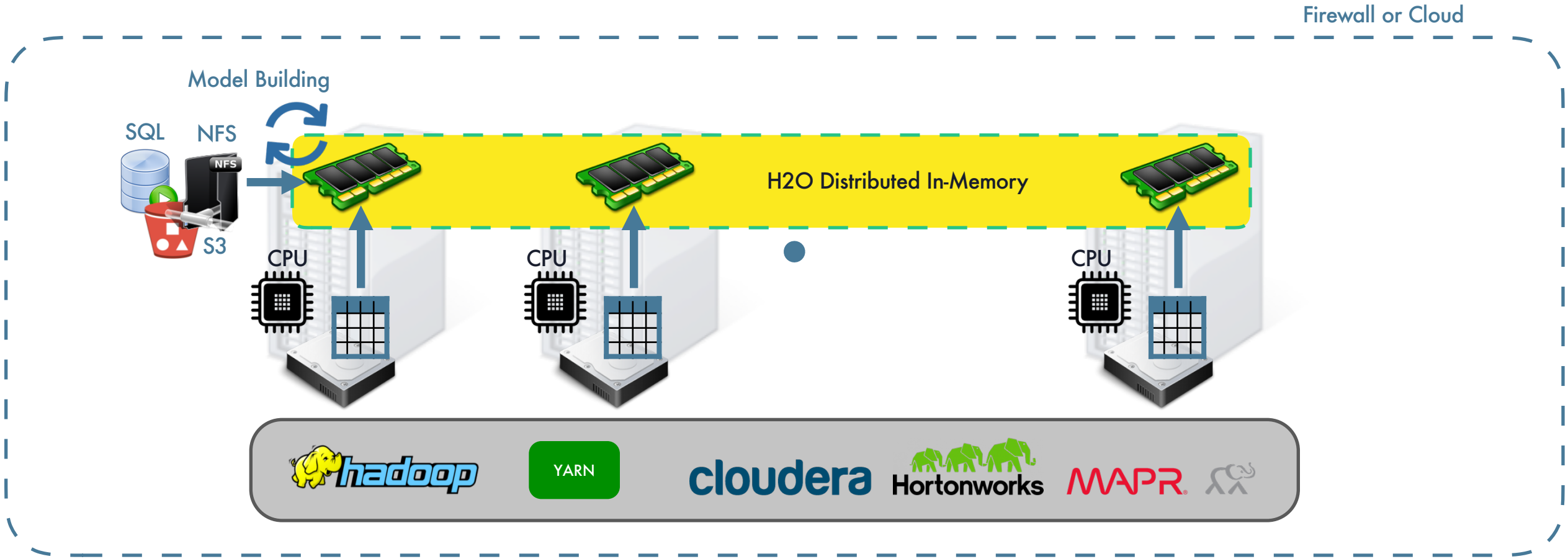
Supervised Learning

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

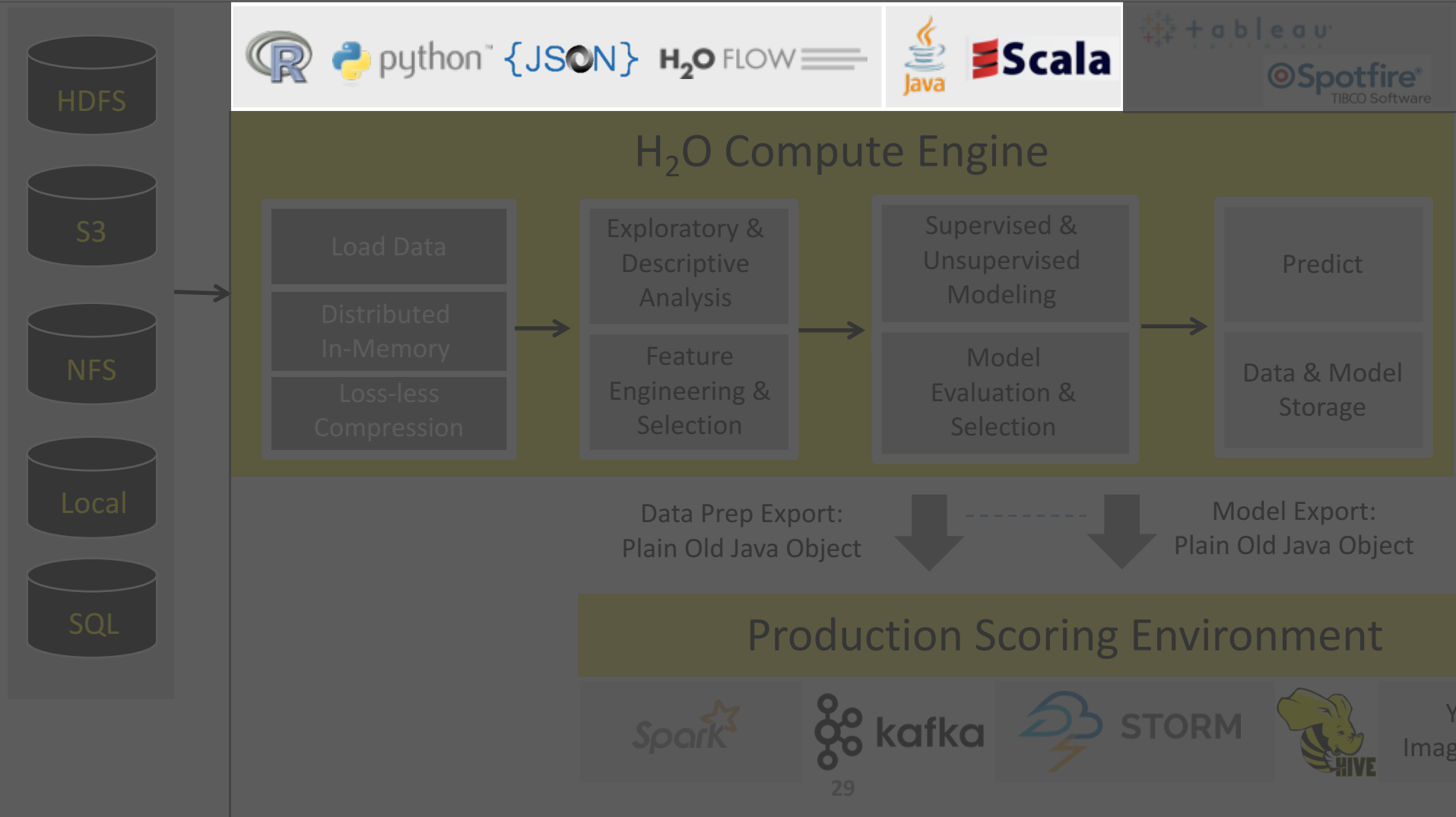
Clustering	• **K-means**: Partitions observations into k clusters/groups of the same spatial size. Automatically detect optimal k
Dimensionality Reduction	• **Principal Component Analysis**: Linearly transforms correlated variables to independent components • **Generalized Low Rank Models:** extend the 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

H_2O Core

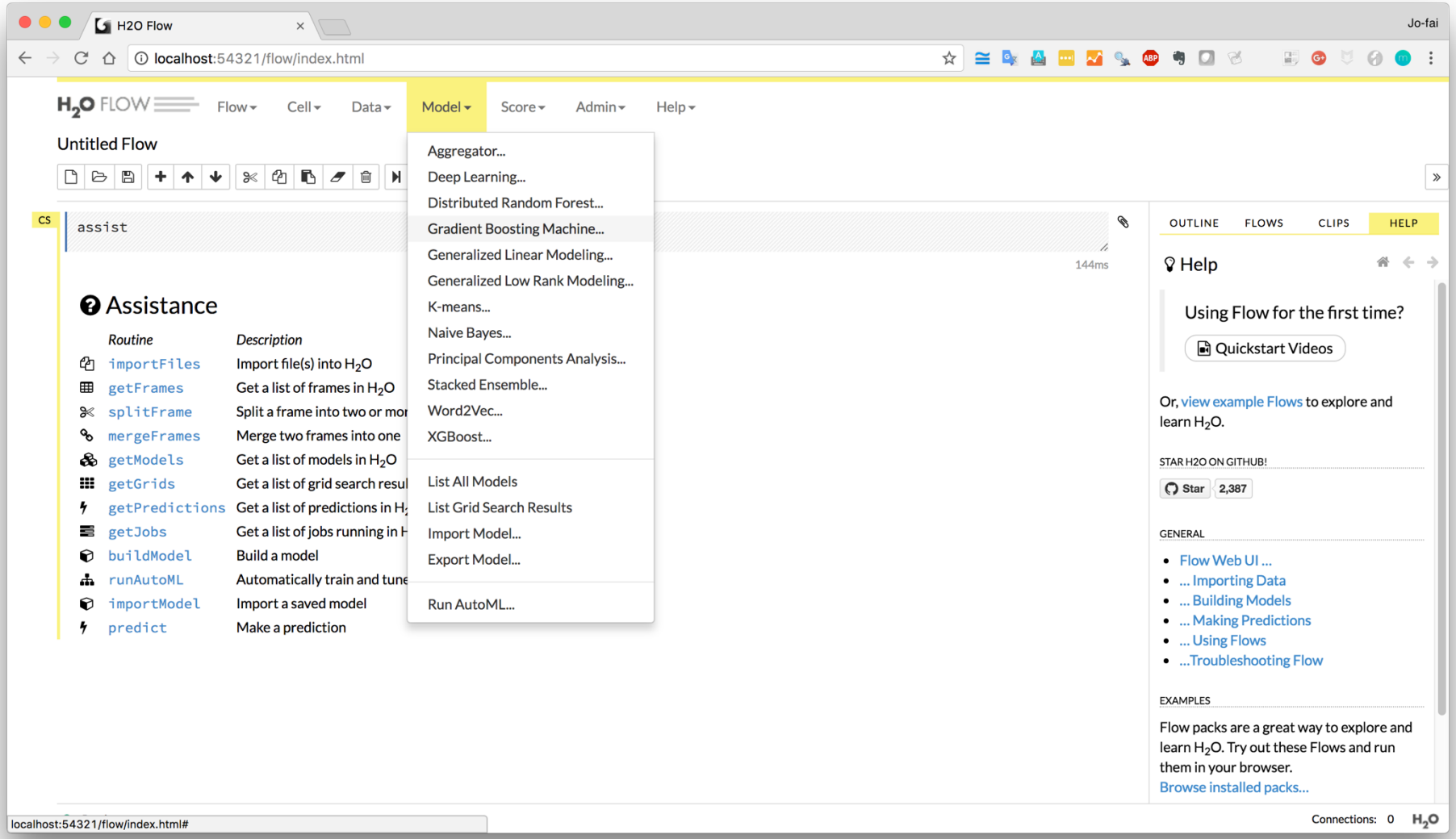

Firewall or Cloud
Model Building
SQL
NFS
NFS
S3
H2O Distributed In-Memory
CPU
CPU
CPU
hadoop
YARN
cloudera
Hortonworks
MAPR

High Level Architecture

Multiple Interfaces



H_2O Flow (Web) – Today's Demos



H2O.ai

Using H_2O with R and Python

~/Documents/repo_h2o/sales-engineering - master - RStudio Source Editor

credit_card_example.R

```r
# Credit Card Example

# Datasets:
# https://s3.amazonaws.com/h2o-training/credit_card/credit_card_train.csv
# https://s3.amazonaws.com/h2o-training/credit_card/credit_card_test.csv

# Start and connect to a local H2O cluster
library(h2o)
h2o.init(nthreads = -1)

# Import datasets from s3
df_train = h2o.importFile("https://s3.amazonaws.com/h2o-training/credit_card/credit_card_train.csv")
df_test = h2o.importFile("https://s3.amazonaws.com/h2o-training/credit_card/credit_card_test.csv")

# Look at datasets
summary(df_train)
summary(df_test)

# Define features and target
features = colnames(df_test)
target = "DEFAULT_PAYMENT_NEXT_MONTH"

# Train a GBM model
model_gbm = h2o.gbm(x = features,
                    y = target,
                    training_frame = df_train,
                    seed = 1234)
print(model_gbm)

# Use GBM model for making predictions
yhat_test = h2o.predict(model_gbm, newdata = df_test)
head(yhat_test)

# (Extra) Use H2O's AutoML
aml = h2o.automl(x = features,
                 y = target,
                 training_frame = df_train,
                 max_runtime_secs = 60,
                 seed = 1234)

# Print leaderboard
print(aml@leaderboard)

# Use best model for making predictions
best_model = aml@leader
yhat_test = h2o.predict(best_model, newdata = df_test)
head(yhat_test)
```

49:1 (Top Level) R Script

credit_card_example

localhost:8888/notebooks/credit_card_example.ipynb

jupyter credit_card_example Last Checkpoint: 5 minutes ago (unsaved changes)

File Edit View Insert Cell Kernel Help Python 2

Code CellToolbar

In [2]:
```r
# Start and connect to a local H2O cluster
import h2o
h2o.init(nthreads = -1)
```

```
Checking whether there is an H2O instance running at http://localhost:54321..... not found.
Attempting to start a local H2O server...
  Java Version: java version "1.8.0_72"; Java(TM) SE Runtime Environment (build 1.8.0_72-b15); Java HotSpot(TM) 64-Bit Server VM (build 25.72-b15, mixed mode)
  Starting server from /Users/jofaichow/anaconda/lib/python2.7/site-packages/h2o/backend/bin/h2o.jar
  Ice root: /var/folders/4z/p7yt7_4n4fj1jlyq6g4qhfbw0000gn/T/tmpFdP3Av
  JVM stdout: /var/folders/4z/p7yt7_4n4fj1jlyq6g4qhfbw0000gn/T/tmpFdP3Av/h2o_jofaichow_started_from_python.out
  JVM stderr: /var/folders/4z/p7yt7_4n4fj1jlyq6g4qhfbw0000gn/T/tmpFdP3Av/h2o_jofaichow_started_from_python.err
  Server is running at http://127.0.0.1:54321
Connecting to H2O server at http://127.0.0.1:54321... successful.
```

H2O cluster uptime:	02 secs
H2O cluster version:	3.13.0.3981
H2O cluster version age:	29 days
H2O cluster name:	H2O_from_python_jofaichow_id7iqa
H2O cluster total nodes:	1

In [3]:
```r
# Import datasets from s3
df_train = h2o.import_file("https://s3.amazonaws.com/h2o-training/credit_card/credit_card_train.csv")
df_test = h2o.import_file("https://s3.amazonaws.com/h2o-training/credit_card/credit_card_test.csv")
```

```
Parse progress: |████████████████████████████████████████████████| 100%
Parse progress: |████████████████████████████████████████████████| 100%
```

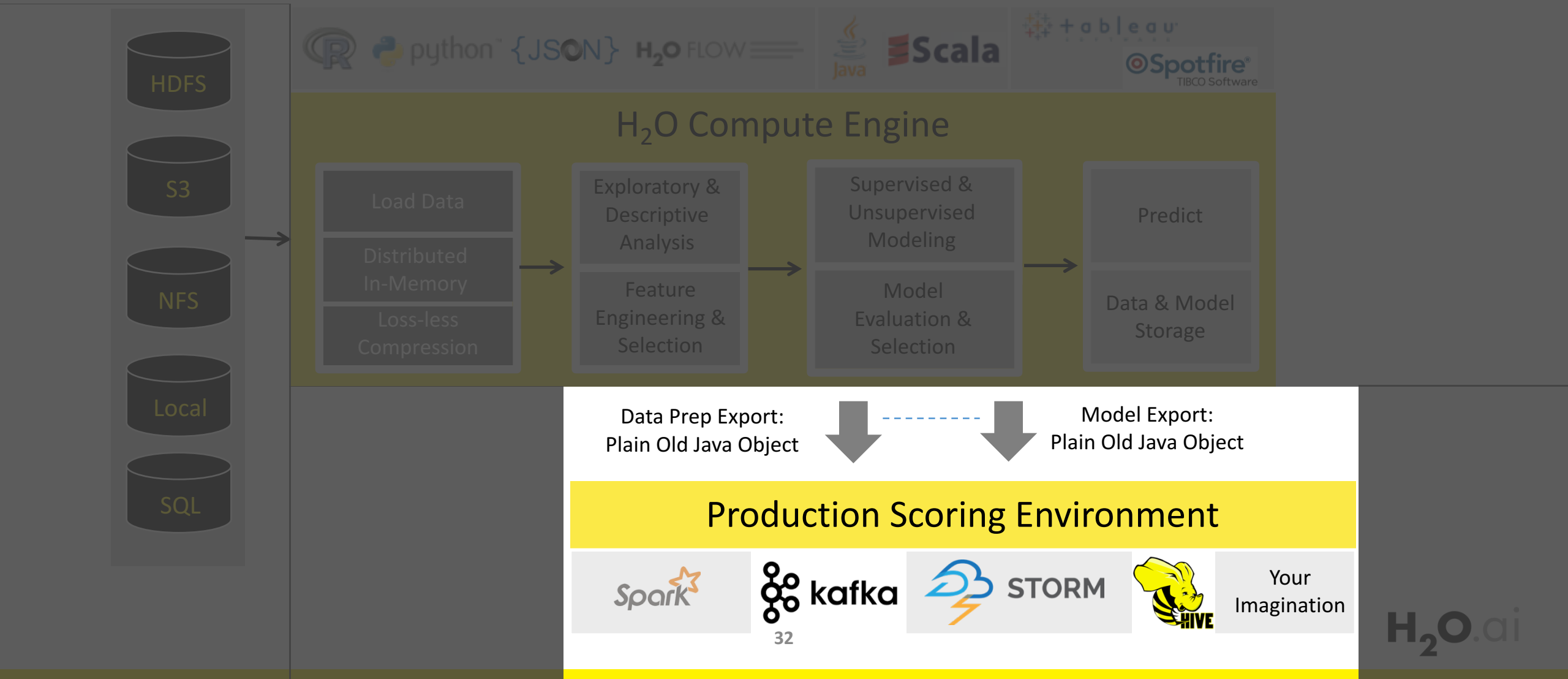
In [4]:
```r
# Look at datasets
df_train.summary()
df_test.summary()
```

	LIMIT_BAL	SEX	EDUCATION	MARRIAGE	AGE	PAY_0	PAY_2	PAY_3	PAY_4
type	int	enum	int	int	int	int	int	int	int
mins	10000.0		0.0	0.0	21.0	-2.0	-2.0	-2.0	-2.0
mean	165471.466667		1.85	1.55578703704	35.4053240741	-0.00523148148148	-0.122361111111	-0.15537037037	-0.21060
maxs	1000000.0		6.0	3.0	79.0	8.0	8.0	8.0	8.0
sigma	128853.314839		0.779559696278	0.522505078476	9.27675421641	1.12668964211	1.20086854503	1.20727030901	1.17217
zeros	0		9	37	0	10563	11284	11309	11905
missing	0	0	0	0	0	0	0	0	0

H2O.ai

High Level Architecture

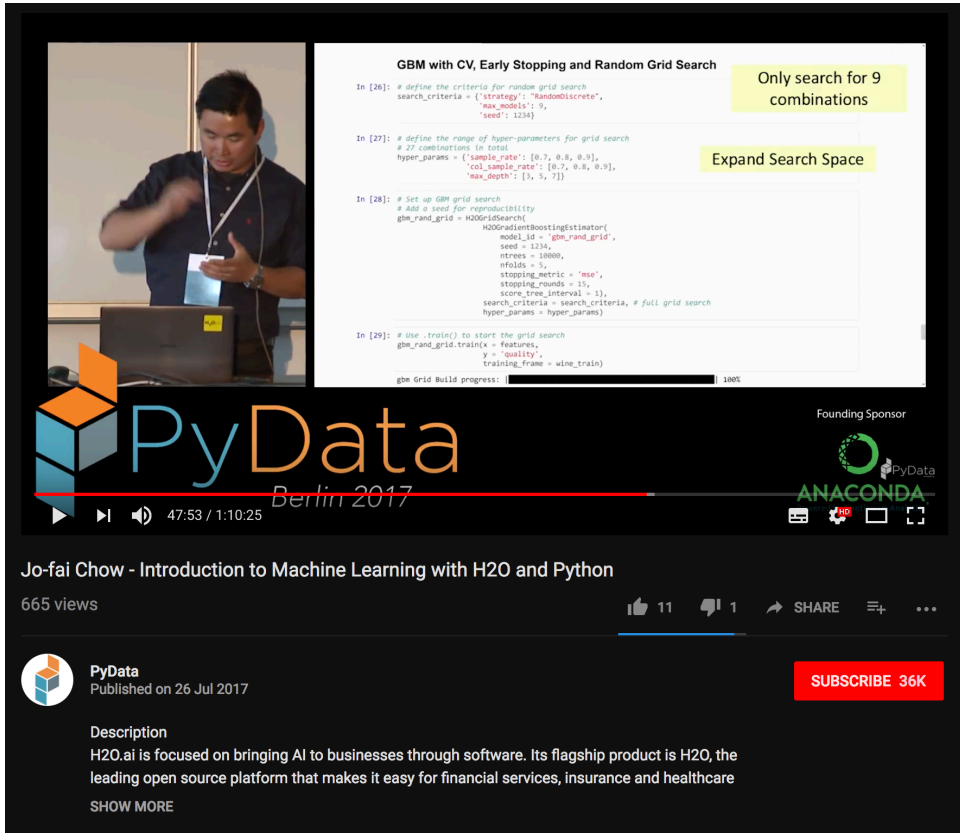
Export Standalone Models for Production



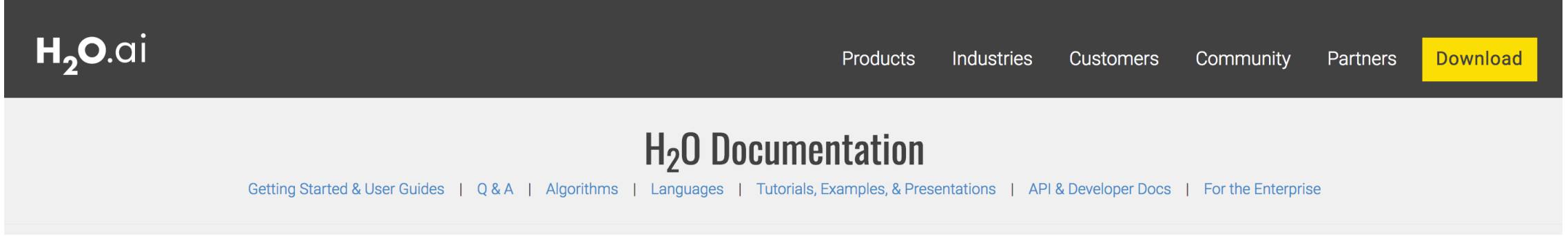
My Training Materials

- H_2O + Python
  - Workshop at PyData Conferences
  - bit.ly/joe_h2o_tutorials

- H_2O + R
  - github.com/woobe/h2o_training_2017_10



H2O.ai



Getting Started & User Guides

Open Source | Commercial

H2O

What is H2O?
H2O User Guide (Main docs)
H2O Book (O'Reilly)
Recent Changes
Open Source License (Apache V2)

Quick Start Video - Flow Web UI
Quick Start Video - R
Quick Start Video - Python

Download H2O

Sparkling Water

What is Sparkling Water?
Sparkling Water Booklet
PySparkling Readme 2.0 | 2.1 | 2.2
RSparkling Readme
Open Source License (Apache V2)

Quick Start Video - Scala

Download Sparkling Water

Driverless AI

What is Driverless AI?
Driverless AI User Guide HTML PDF
Driverless AI Booklet
MLI with Driverless AI Booklet

Driverless AI Webinars

Download Driverless AI

H2O4GPU (alpha)

H2O4GPU Readme
Open Source License (Apache V2)

Download H2O4GPU

URL: **docs.h2o.ai**

H2O.ai

Demo: Running H_2O on Hadoop

Higgs Boson Machine Learning Challenge

Use the ATLAS experiment to identify the Higgs boson

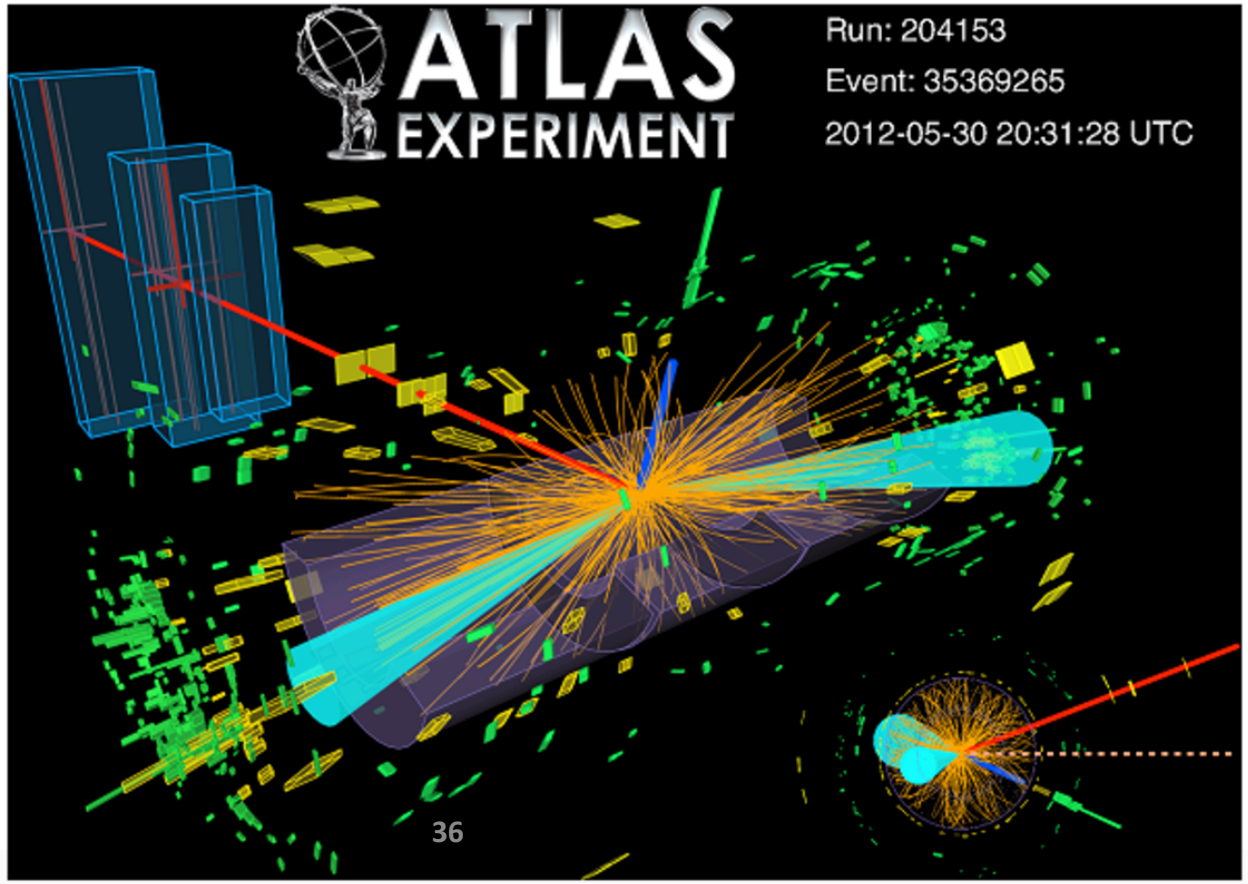
$13,000 · 1,785 teams · 3 years ago

Overview | Data | Discussion | Leaderboard | Rules | Team | My Submissions | Late Submission

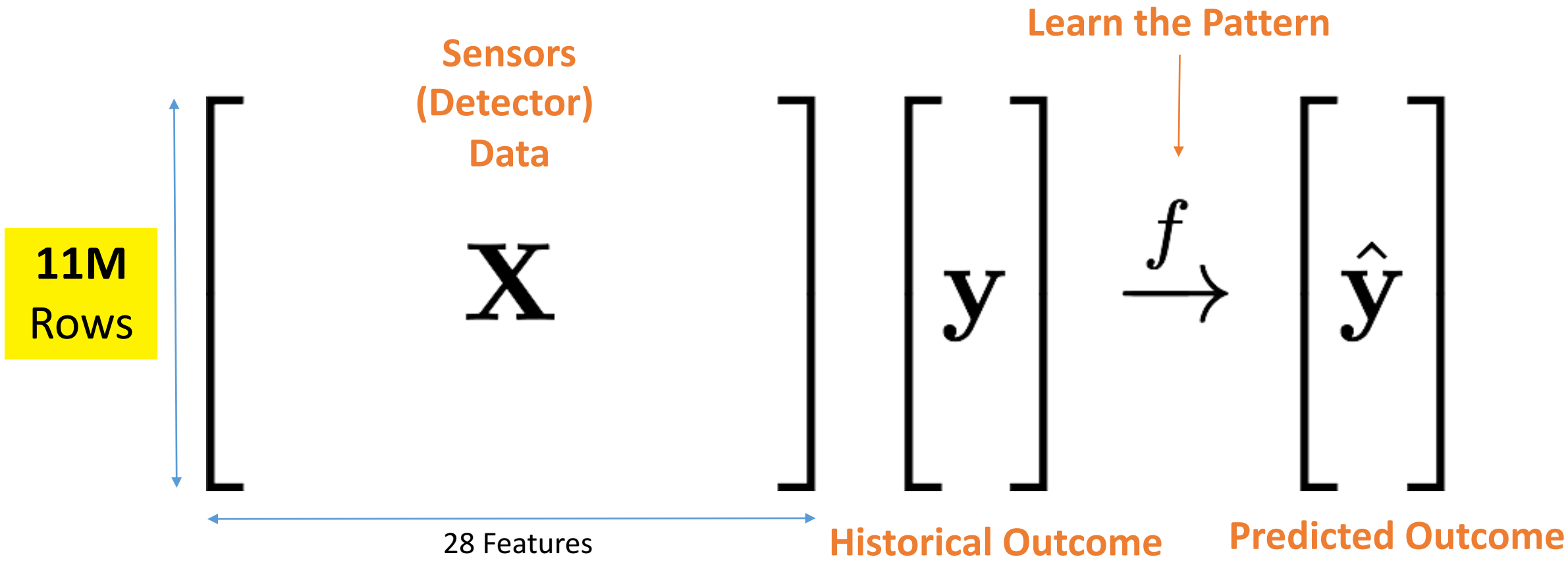
https://www.kaggle.com/c/higgs-boson

Overview

- Description
- Evaluation
- Prizes
- About The Sponsors
- Timeline
- Winners



Learning from Higgs Boson Machine Data



Raw Data Size: **7.48 GB**

11M Rows

Size (Raw): **7.48 GB**
Compressed: **2.00 GB** (≈ 27% of Raw)

HIGGS.hex

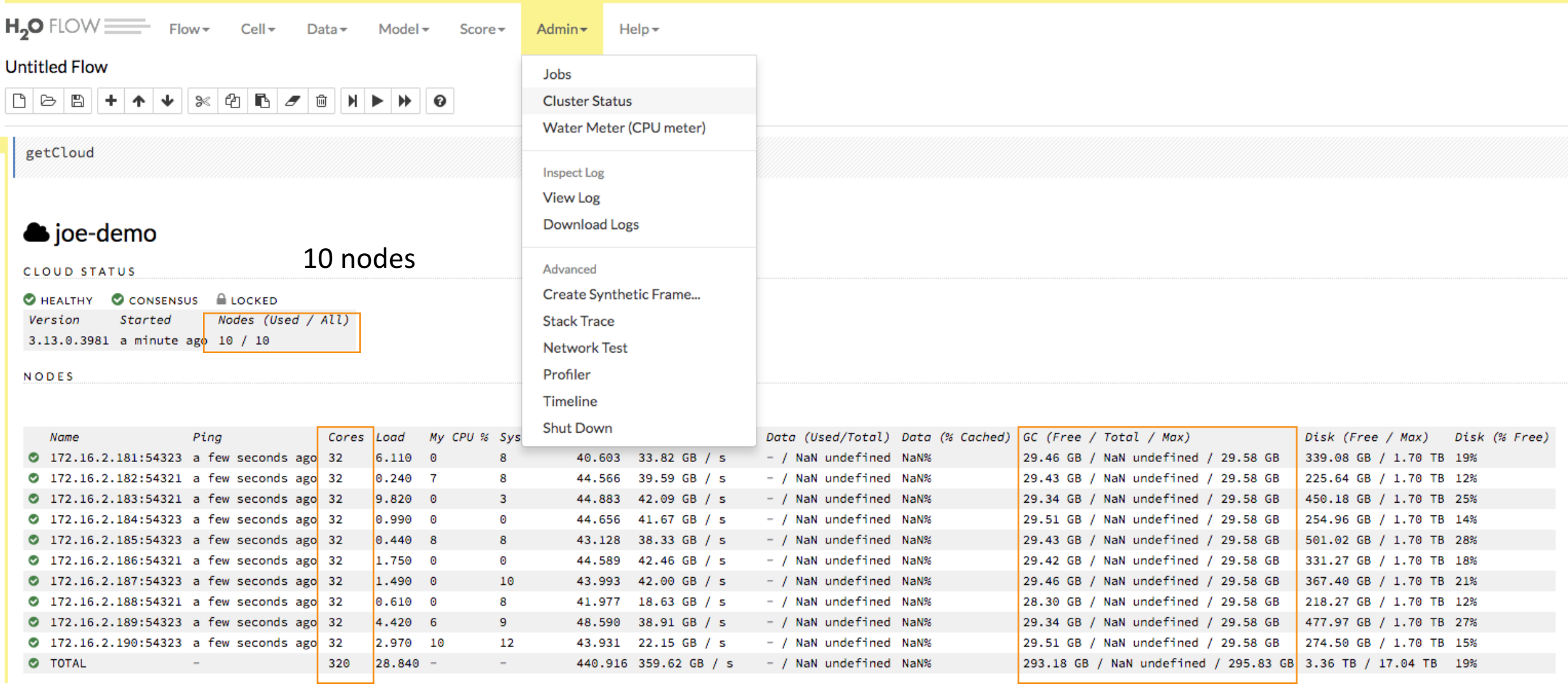
Actions: View Data | Split... | Build Model... | Predict | Download | Export

Rows	Columns	Compressed Size
11000000	29	2GB

COLUMN SUMMARIES

label	type	Missing	Zeros	+Inf	-Inf	min	max	mean	sigma	cardinality	Actions
C1	enum	0	5170877	0	0	0	1.0	0.5299	0.4991	2	Convert to numeric
C2	real	0	0	0	0	0.2747	12.0989	0.9915	0.5654	·	·
C3	real	0	0	0	0	-2.4350	2.4349	-0.0	1.0088	·	·
C4	real	0	0	0	0	-1.7425	1.7432	-0.0	1.0063	·	·
C5	real	0	0	0	0	0.0002	15.3968	0.9985	0.6000	·	·
C6	real	0	0	0	0	-1.7439	1.7433	0.0	1.0063	·	·
C7	real	0	0	0	0	0.1375	9.9404	0.9909	0.4750	·	·
C8	real	0	0	0	0	-2.9697	2.9697	-0.0	1.0093	·	·
C9	real	0	0	0	0	-1.7412	1.7415	0.0	1.0059	·	·
C10	real	0	5394611	0	0	0	2.1731	1.0	1.0278	·	·
C11	real	0	0	0	0	0.1890	11.6471	0.9927	0.5000	·	·
C12	real	0	0	0	0	-2.9131	2.9132	-0.0	1.0093	·	·
C13	real	0	0	0	0	-1.7424	1.7432	-0.0	1.0062	·	·
C14	real	0	5523912	0	0	0	2.2149	1.0	1.0494	·	·
C15	real	0	0	0	0	0.2636	14.7090	0.9923	0.4877	·	·
C16	real	0	0	0	0	-2.7297	2.7300	0.0	1.0087	·	·
C17	real	0	0	0	0	-1.7421	1.7429	0.0	1.0063	·	·
C18	real	0	6265240	0	0	0	2.5482	1.0	1.1937	·	·
C19	real	0	0	0	0	0.3654	12.8826	0.9861	0.5058	·	·
C20	real	0	0	0	0	-2.4973	2.4980	-0.0	1.0077	·	·

H2O.ai

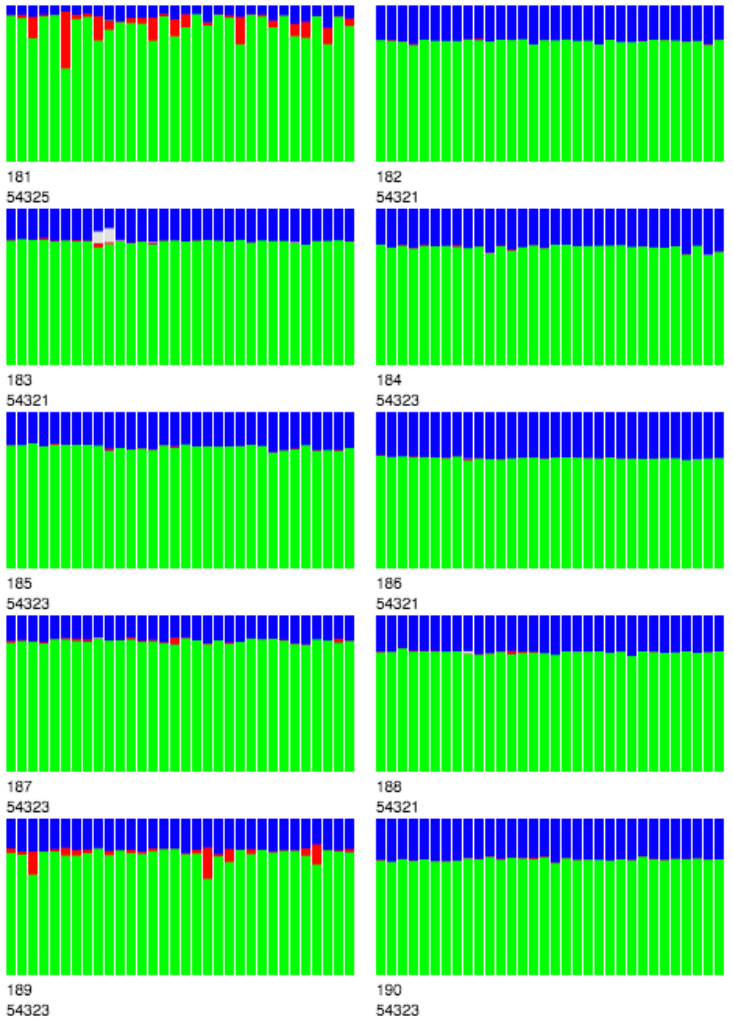


10 x 32 =
320 Cores

10 x 29.6 = 296
GB Memory

H_2O Water Meter (CPU Monitor)

10 x 32 = 320 Cores



Demo: AutoML

Automatic Machine Learning with H_2O

AutoML: Automatic Machine Learning

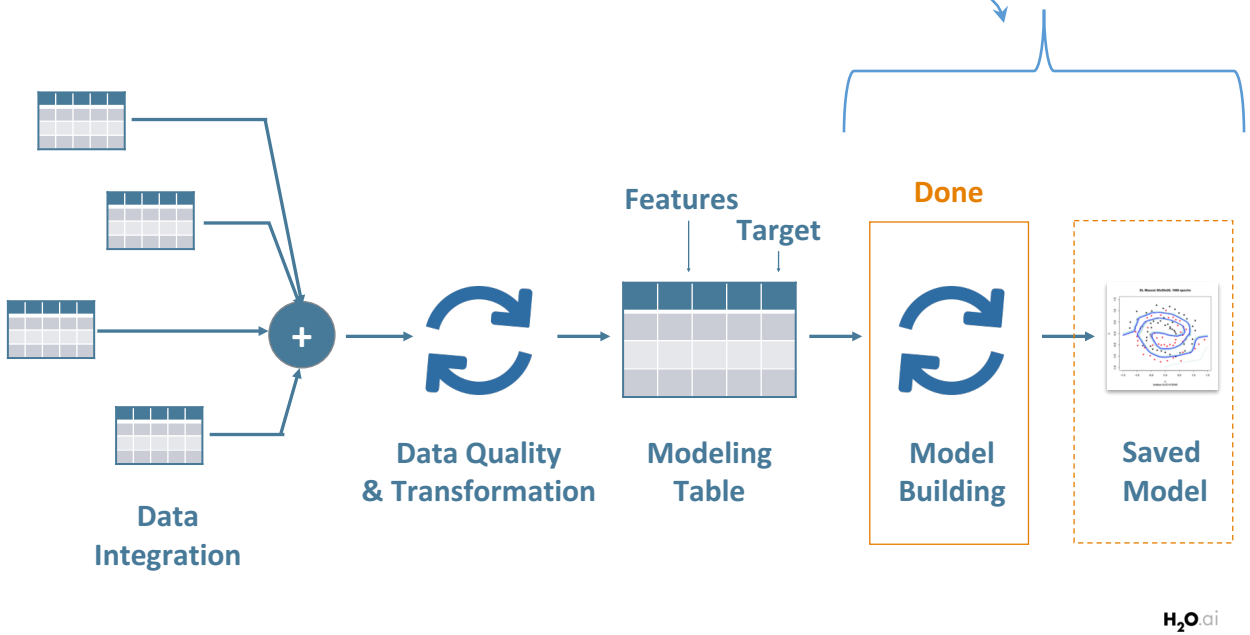
In recent years, the demand for machine learning experts has outpaced the supply, despite the surge of people entering the field. To address this gap, there have been big strides in the development of user-friendly machine learning software that can be used by non-experts. The first steps toward simplifying machine learning involved developing simple, unified interfaces to a variety of machine learning algorithms (e.g. H2O).

Although H2O has made it easy for non-experts to experiment with machine learning, there is still a fair bit of knowledge and background in data science that is required to produce high-performing machine learning models. Deep Neural Networks in particular are notoriously difficult for a non-expert to tune properly. In order for machine learning software to truly be accessible to non-experts, we have designed an easy-to-use interface which automates the process of training a large selection of candidate models. H2O's AutoML can also be a helpful tool for the advanced user, by providing a simple wrapper function that performs a large number of modeling-related tasks that would typically require many lines of code, and by freeing up their time to focus on other aspects of the data science pipeline tasks such as data-preprocessing, feature engineering and model deployment.

H2O's AutoML can be used for automating the machine learning workflow, which includes automatic training and tuning of many models within a user-specified time-limit. The user can also use a performance metric-based stopping criterion for the AutoML process rather than a specific time constraint. Stacked Ensembles will be automatically trained on the collection individual models to produce a highly predictive ensemble model which, in most cases, will be the top performing model in the AutoML Leaderboard. Stacked ensembles are not yet available for multiclass classification problems, so in that case, only singleton models will be trained.

AutoML Interface

The H2O AutoML interface is designed to have as few parameters as possible so that all the user needs to do is point to their dataset, identify the response column and optionally specify a time constraint, a maximum number of models constraint, and early stopping parameters.



H2O.ai

AutoML Output

The AutoML object includes a "leaderboard" of models that were trained in the process, ranked by a default metric based on the problem type (the second column of the leaderboard). In binary classification problems, that metric is AUC, and in multiclass classification problems, the metric is mean per-class error. In regression problems, the default sort metric is deviance. Some additional metrics are also provided, for convenience.

Here is an example leaderboard for a binary classification task:

model_id	auc	logloss
StackedEnsemble_0_AutoML_20170605_212658	0.776164	0.564872
GBM_grid_0_AutoML_20170605_212658_model_2	0.75355	0.587546
DRF_0_AutoML_20170605_212658	0.738885	0.611997
GBM_grid_0_AutoML_20170605_212658_model_0	0.735078	0.630062
GBM_grid_0_AutoML_20170605_212658_model_1	0.730645	0.67458
XRT_0_AutoML_20170605_212658	0.728358	0.629296
GLM_grid_0_AutoML_20170605_212658_model_1	0.685216	0.635137
GLM_grid_0_AutoML_20170605_212658_model_0	0.685216	0.635137

Demo Introduction

/ **Use Case**: Probability of Default for Credit Card Loans

/ **Features**

- **default.payment.next.month**: Did the next loan payment default (1=True, 0=False)
- **LIMIT_BAL**: Credit limit in (NT) dollars
- **SEX**, **EDUCATION**, **MARRIAGE**, **AGE**
- **PAY_0**: Was a payment received in the current month?
- **PAY_2**: Was a payment received in the 2 months ago?

…

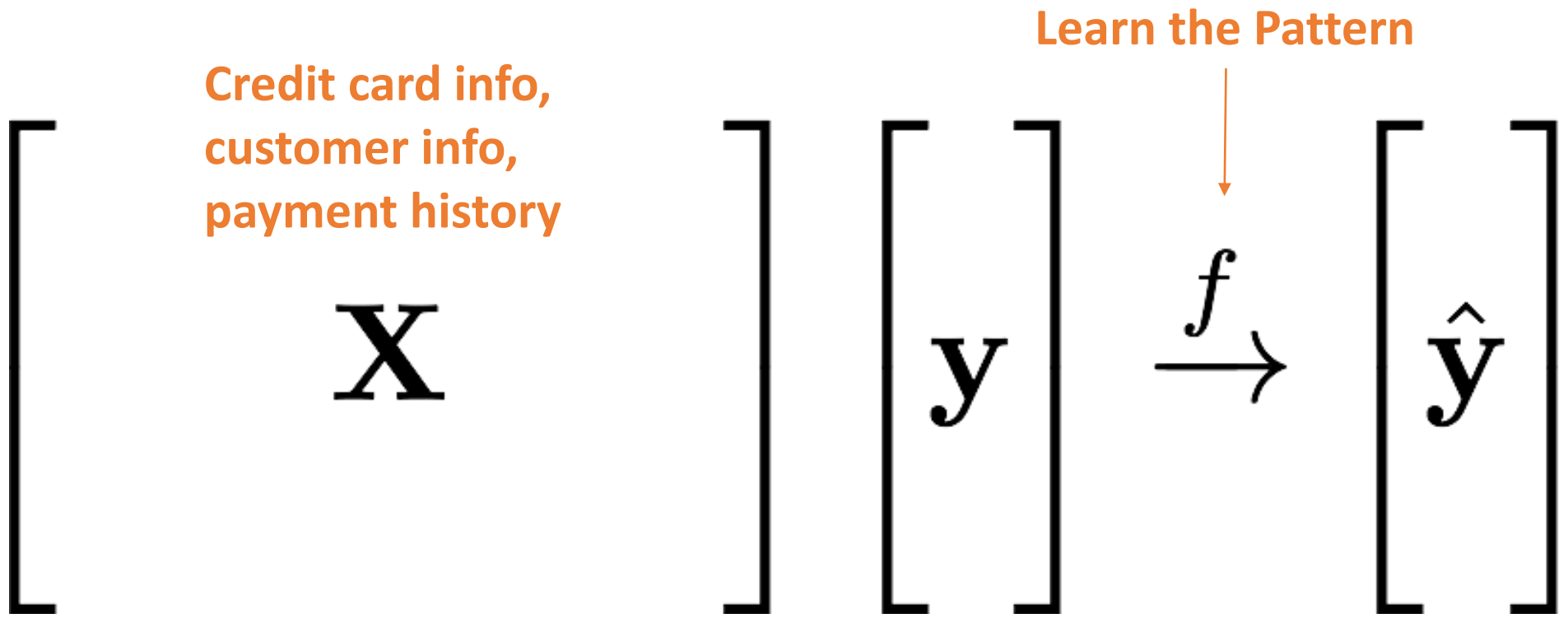
- **BILL_AMT1**: Amount of bill statement in 1 month ago
- **BILL_AMT2**: Amount of bill statement in 2 months ago

…

- **PAY_AMT1**: Amount of previous payment 1 month ago
- **PAY_AMT2**: Amount of previous payment 2 months ago

…

Learning from Credit Card Data



H2O is back with its flagship event, H2O World 2017.

Whether you're just getting started with H2O or you're a power user looking to expand your skill set even more, join

Thank you!

- Organisers & Sponsors
  - Michael Young
  - Scotland Data Science and Technology Meetup

- Code, Slides & Documents
  - bit.ly/h2o_meetups
  - docs.h2o.ai
- Contact
  - joe@h2o.ai
  - @matlabulous
  - github.com/woobe
- Please search/ask questions on **Stack Overflow**
  - Use the tag `**h2o**` (not h2 zero)

H2O.ai