H_2O.ai

Company Overview

Company

- Team: 35. Founded in 2012, Mountain View, CA
- Stanford Math & Systems Engineers

Product

- Open Source Leader in Machine & Deep learning
- Ease of Use and Smarter Applications
- R, Python, Spark & Hadoop Interfaces
- Expanding Predictions to Mass Analyst markets

H_2O.ai
Machine Intelligence

Executive Team

Sri Satish Ambati
CEO & Co-founder

DataStax

Cliff Click
CTO & Co-founder

Sun, Java Hotspot

Tom Kraljevic
VP of Engineering

Abrizio, Intel

Arno Candel
Chief Architect

Physicist, Deep Learning

Board of Directors

Jishnu Bhattacharjee // Nexus Ventures
Ash Bhardwaj // Flextronics

Scientific Advisory Council

Trevor Hastie
Stephen Boyd
Rob Tibshirani

H2O.ai Machine Intelligence

H2O.ai Team

H2O.ai
Machine Intelligence

Scientific Advisory Council

Dr. Trevor Hastie

- PhD in Statistics, Stanford University
- John A. Overdeck Professor of Mathematics, 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*
- 108,404 citations (via Google Scholar)

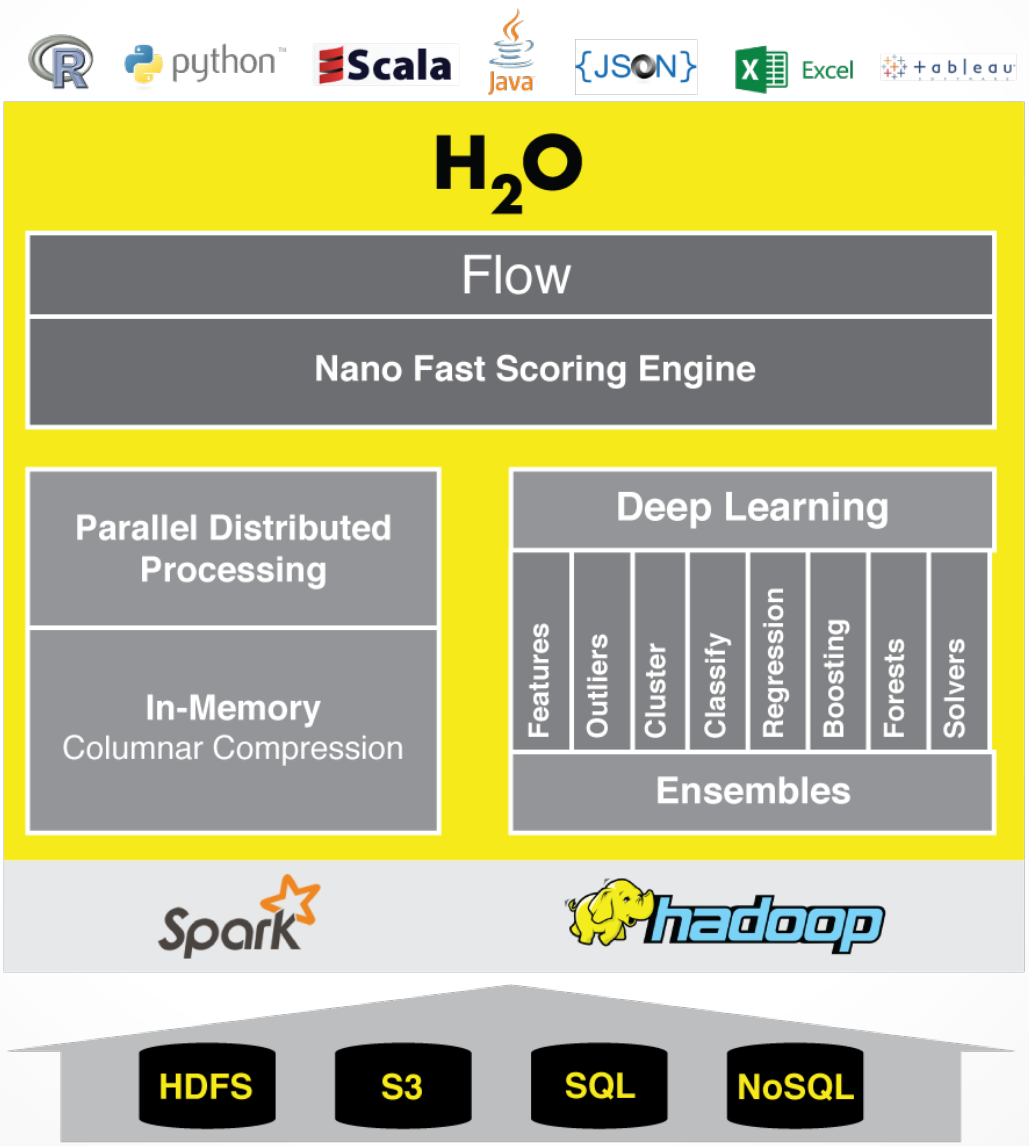
Dr. Rob Tibshirani

- PhD in Statistics, Stanford University
- Professor of Statistics and Health Research and Policy, Stanford University
- COPPS Presidents' Award recipient
- 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. Stephen Boyd

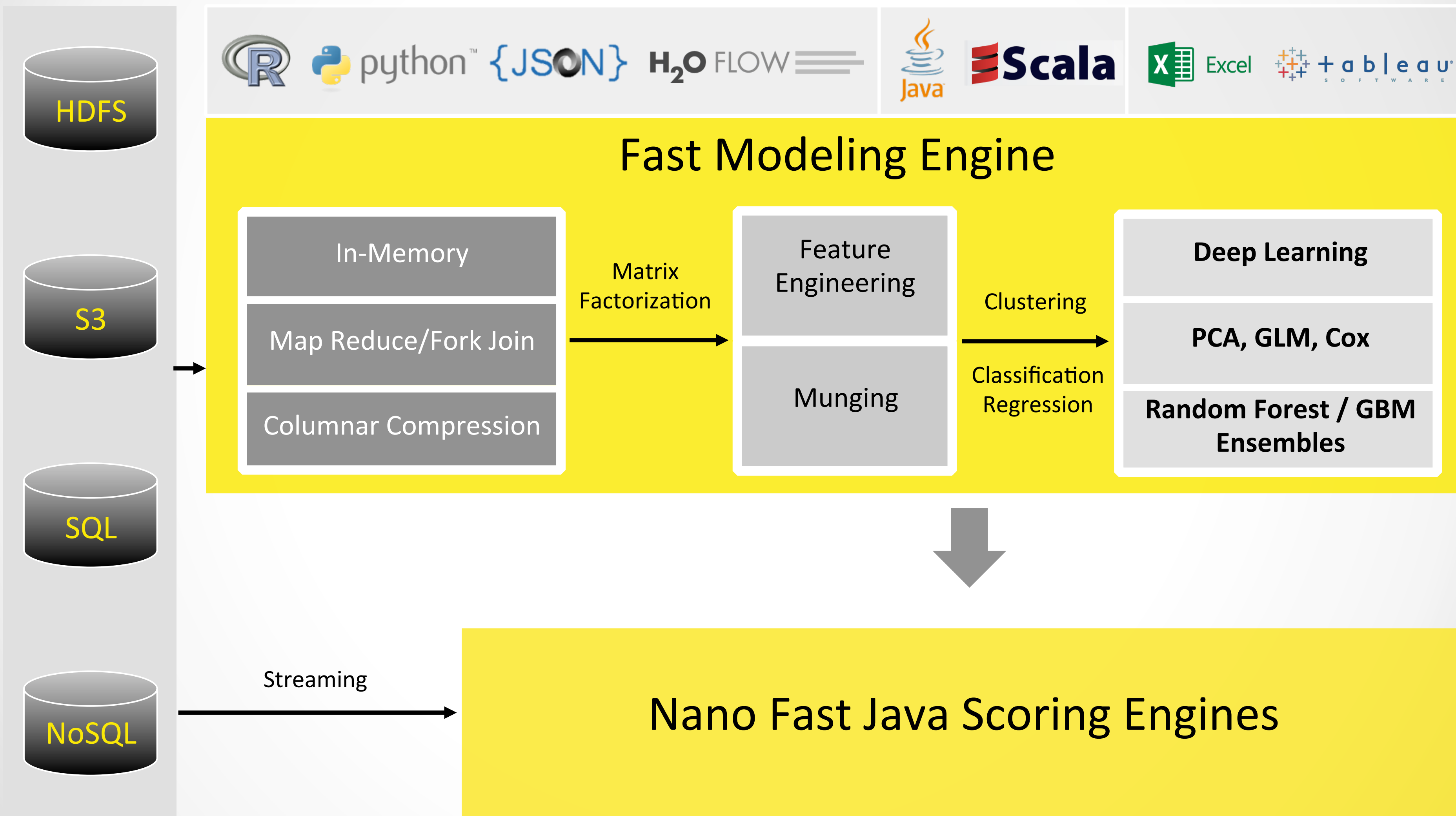
- PhD in Electrical Engineering and Computer Science, UC Berkeley
- Professor of Electrical Engineering and Computer Science, Stanford University
- Co-author, Convex Optimization
- Co-author, Linear Matrix Inequalities in System and Control Theory
- Co-author, *Distributed Optimization and Statistical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

Accuracy with Speed and Scale



H2O.ai
Machine Intelligence

Accuracy with Speed and Scale



Algorithms on H_2O

Supervised Learning

Statistical Analysis

- Generalized Linear Models : Binomial, Gaussian, Gamma, Poisson and Tweedie
- Cox Proportional Hazards Models
- 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

Algorithms on H_2O

Unsupervised Learning

Clustering	• K-means : Partitions observations into k clusters/ groups of the same spatial size
Dimensionality Reduction	• Principal Component Analysis : Linearly transforms correlated variables to independent components
Anomaly Detection	• Autoencoders: Find outliers using a nonlinear dimensionality reduction using deep learning

Reading Data from HDFS into H2O with R

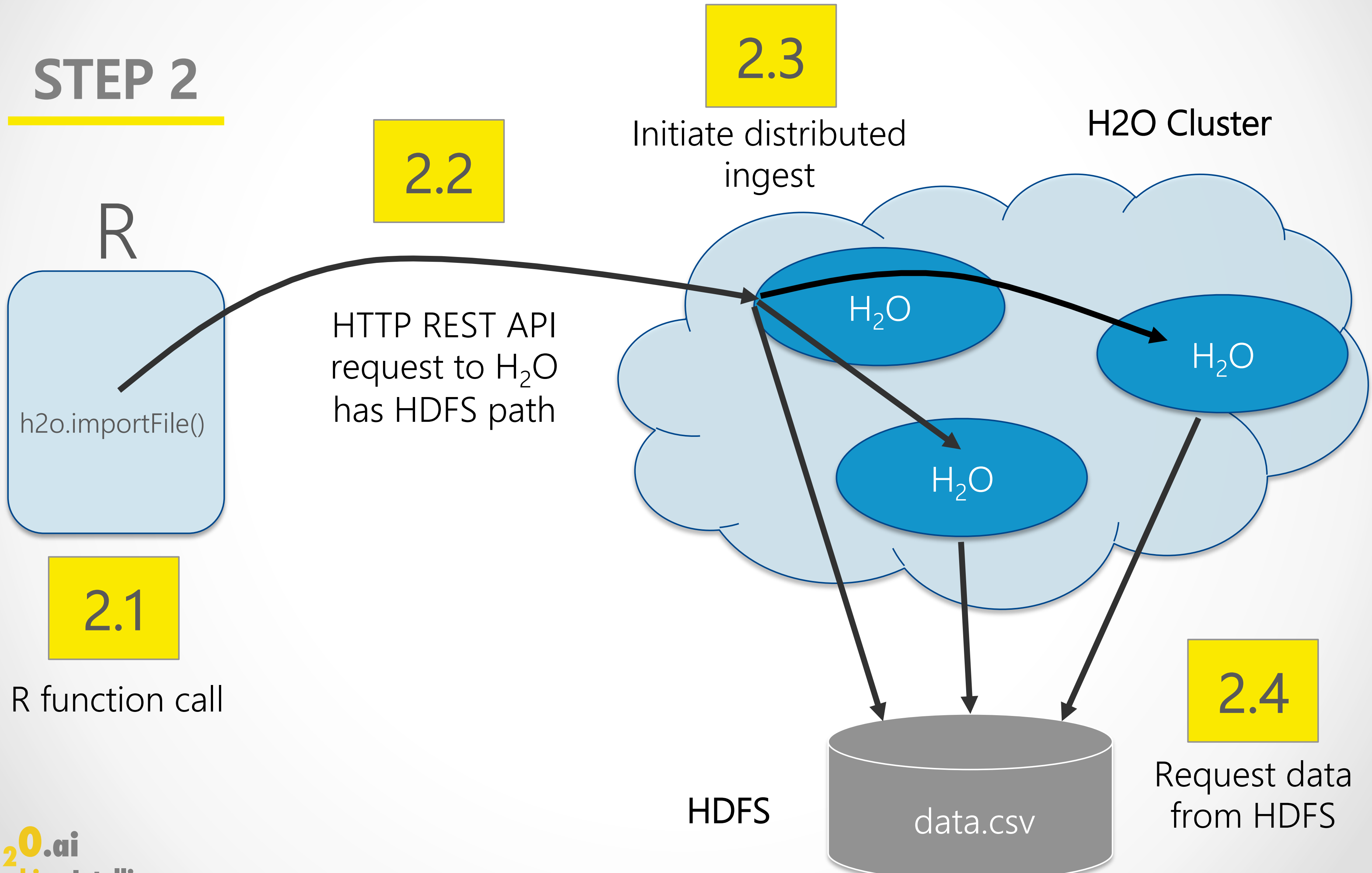
STEP 1

R user

h2o_df = h2o.importFile(“hdfs://path/to/data.csv”)

H2O.ai
Machine Intelligence

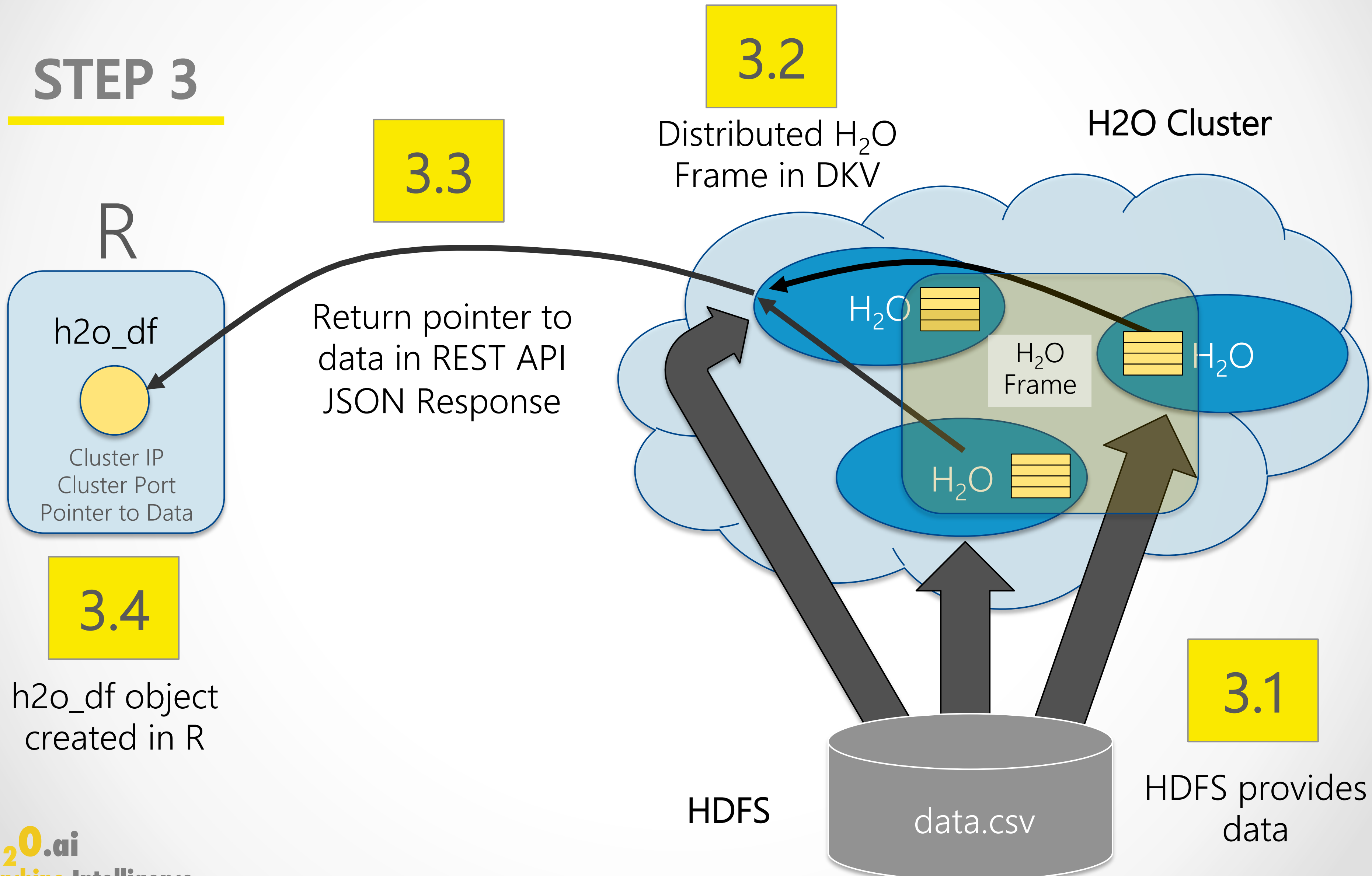
Reading Data from HDFS into H2O with R



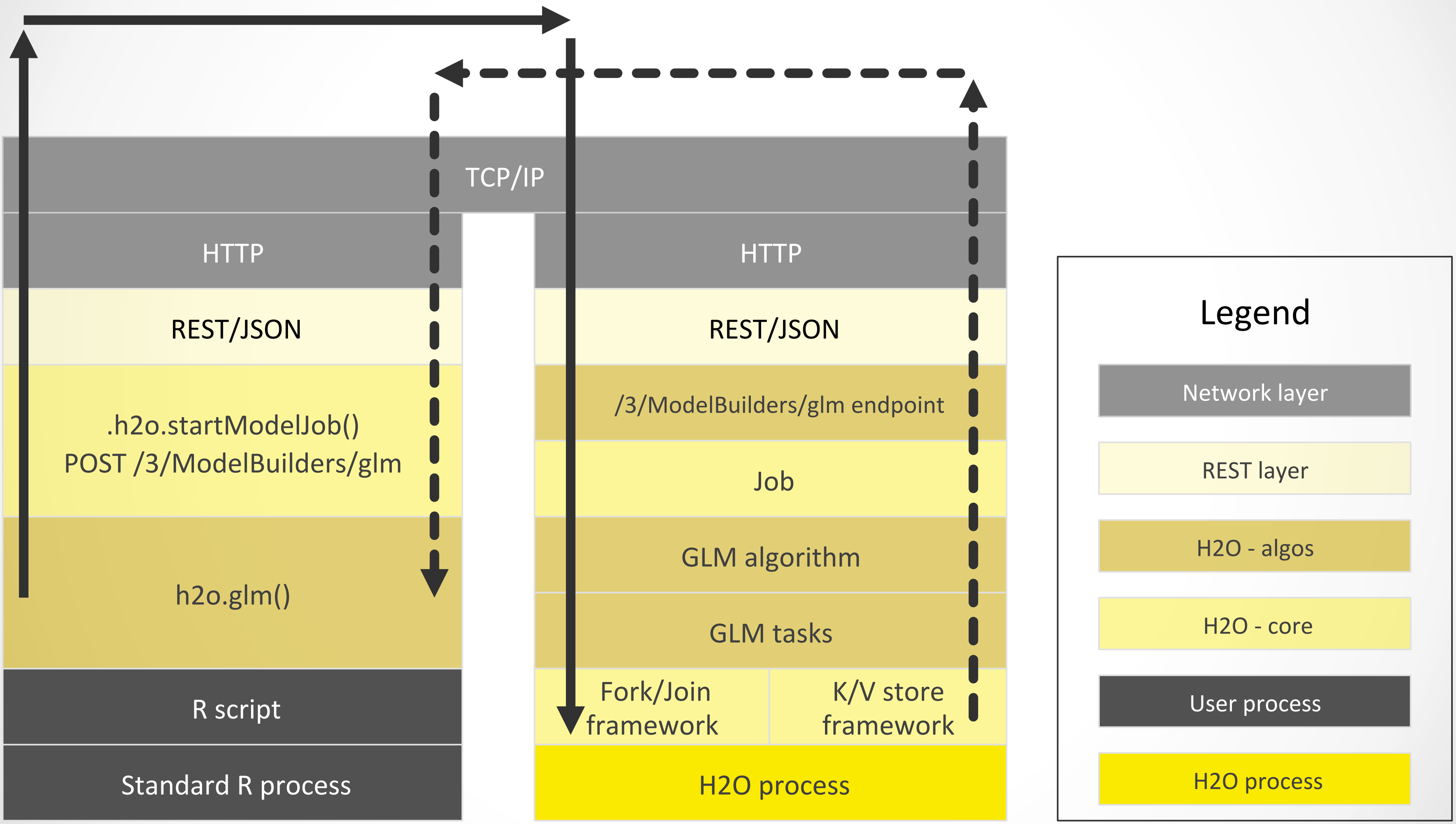
H2O.ai Machine Intelligence

Reading Data from HDFS into H2O with R

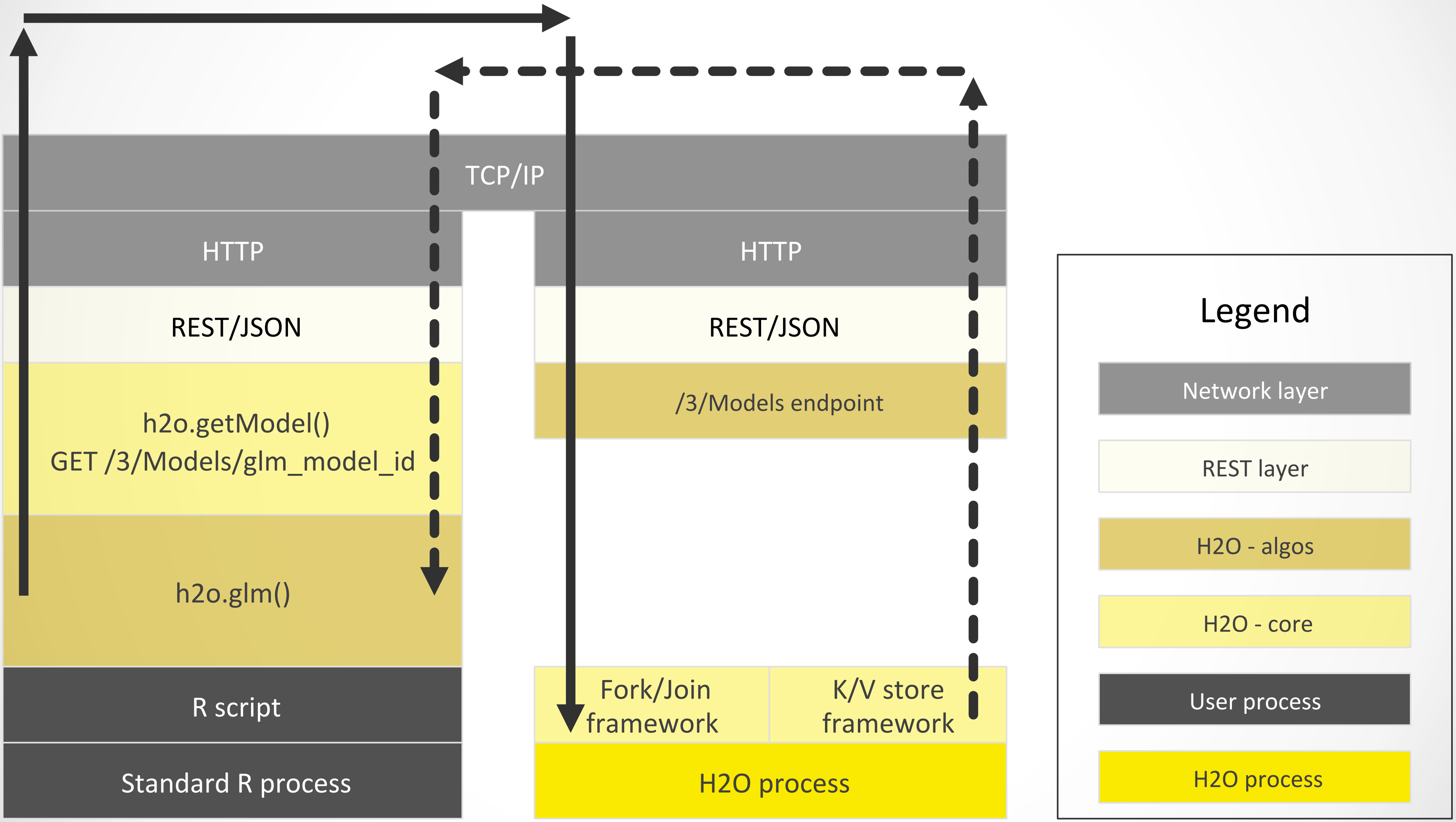
STEP 3



R Script Starting H2O GLM



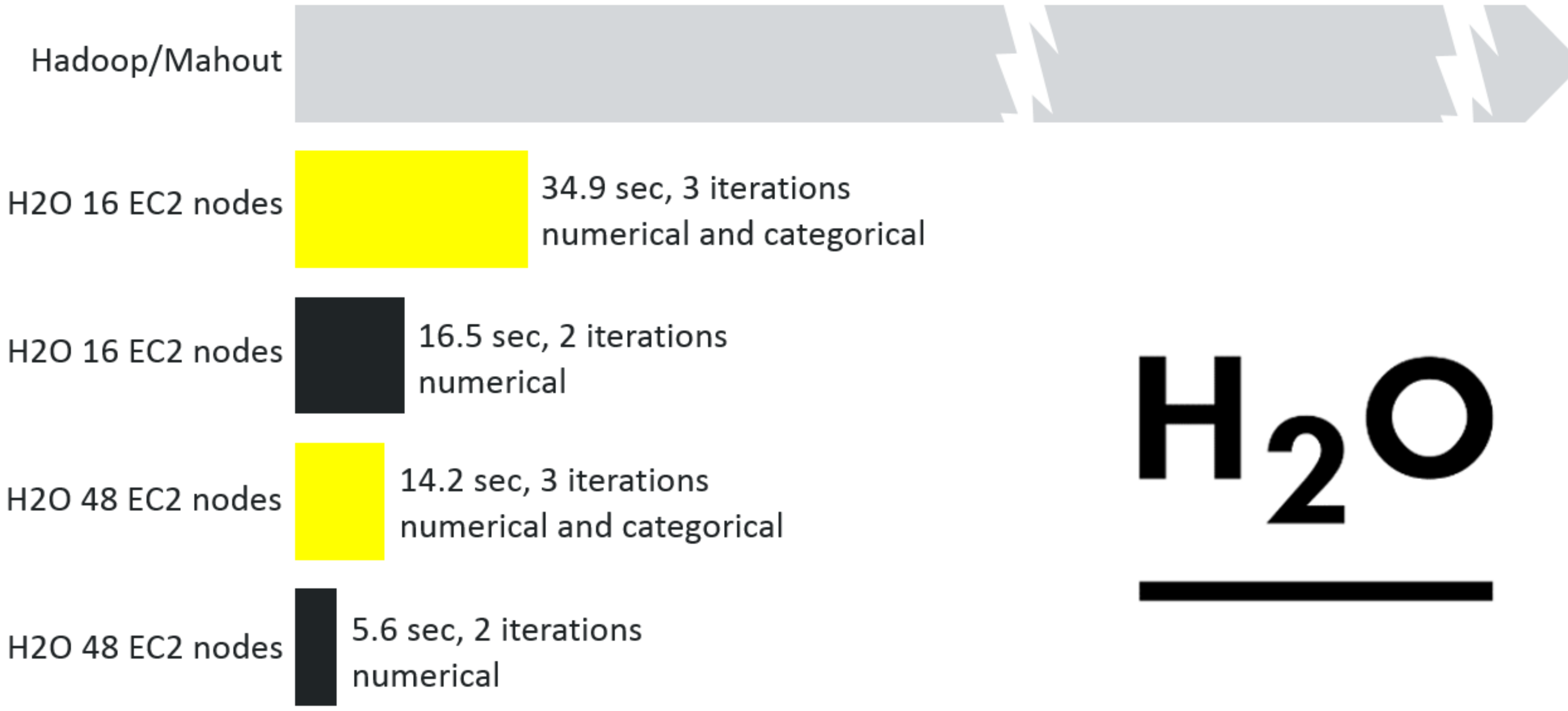
R Script Retrieving H2O GLM Result



H2O.ai
Machine Intelligence

H2O Billion Row Machine Learning Benchmark

GLM Logistic Regression



Compute Hardware: AWS EC2 c3.2xlarge - 8 cores and 15 GB per node, 1 GbE interconnect

Airline Dataset 1987-2013, 42 GB CSV, 1 billion rows, 12 input columns, 1 outcome column
9 numerical features, 3 categorical features with cardinalities 30, 376 and 380

Demo Time!

Community

tableau SOFTWARE

Spark

hadoop

DataRobot