

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

Scalable Machine Learning  
For Smarter Applications

# Who am I?

Hank Roark ([hank@h2oai.com](mailto:hank@h2oai.com), @hankroark)

Data Scientist & Hacker @ H2O.ai

Lecturer in Systems Thinking, UIUC

13 years at John Deere doing Research, New Product Development, New High Tech Ventures

Previously at startups and consulting

Physics Georgia Tech

Systems Design & Management MIT

# H2O.ai Overview

- Founded: 2011 venture-backed, debuted in 2012
- Product: H2O open source in-memory prediction engine
- Team: 37 - Distributed Systems Engineers doing ML
- HQ: Mountain View, CA



# Team Work @ H2O.ai



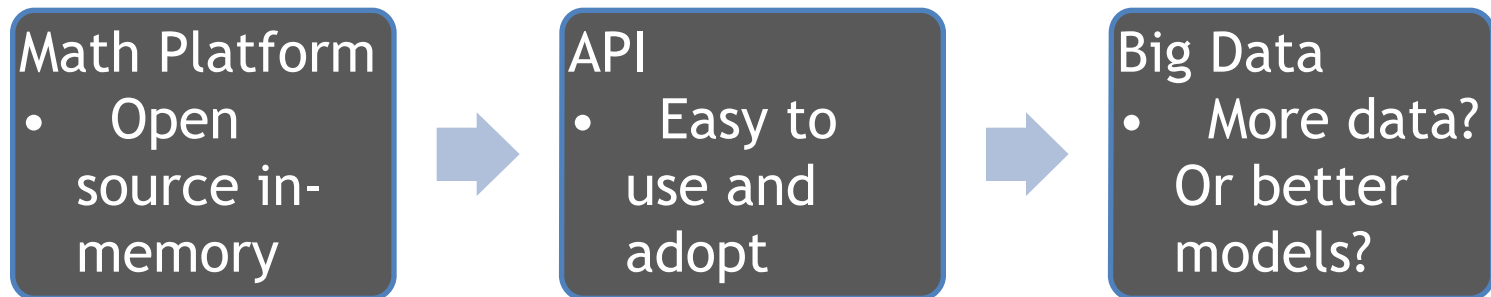
## H2O World Conference 2014



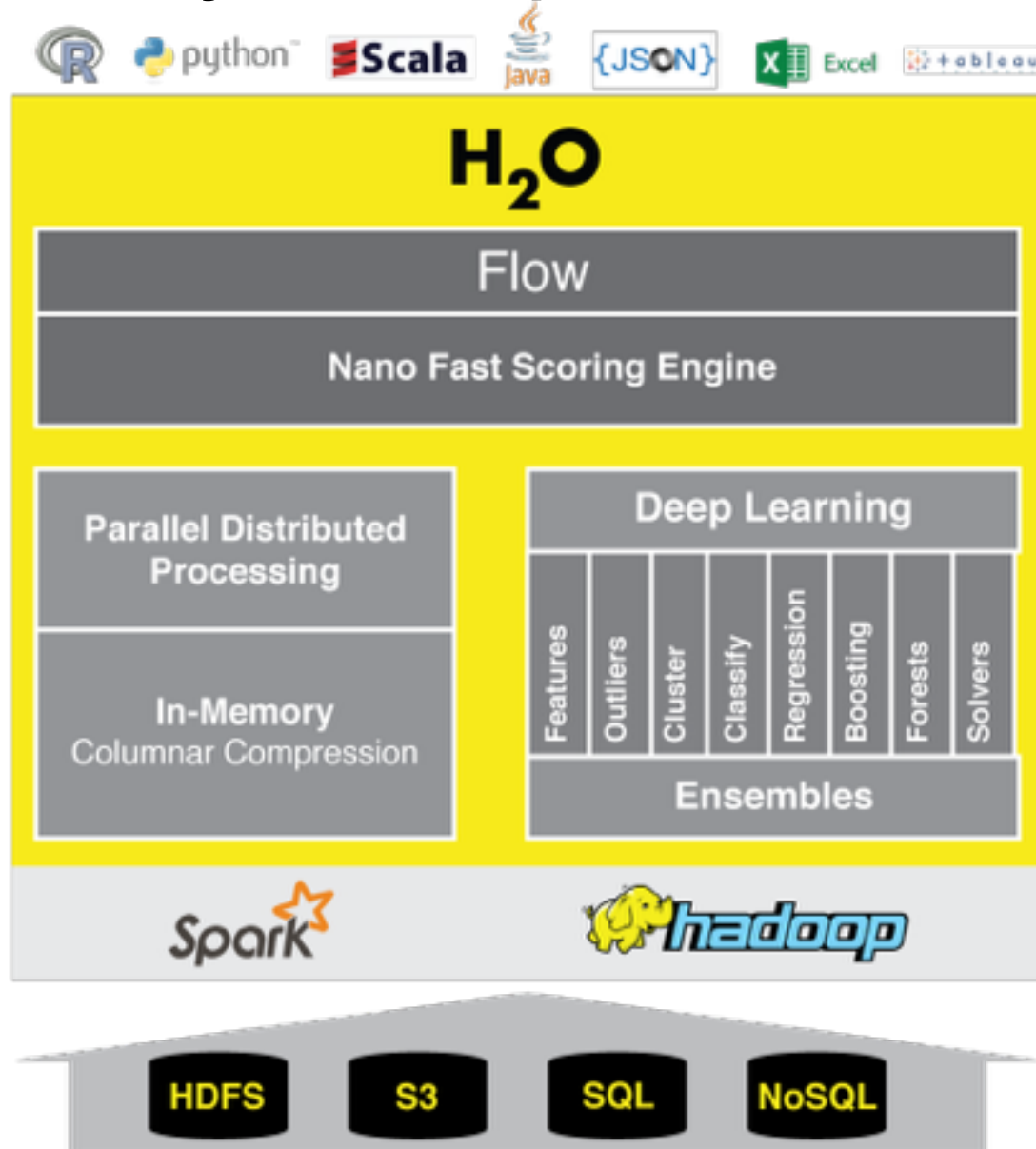
Join H2O World Nov 9-11 2015!



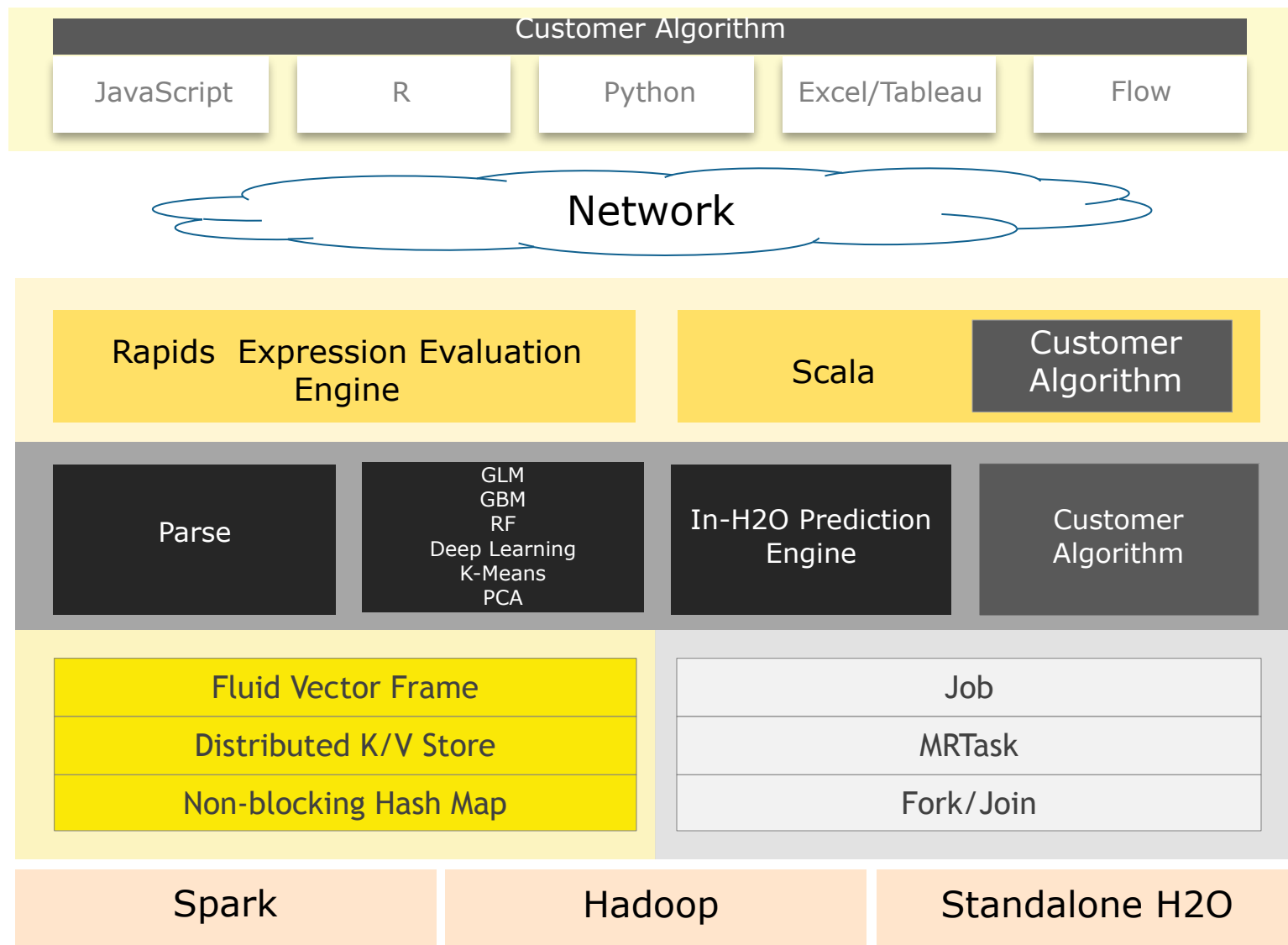
# What is H2O?



# Accuracy with Speed and Scale



# H2O Software Stack



# Reading Data from HDFS into H2O with R

## STEP 1



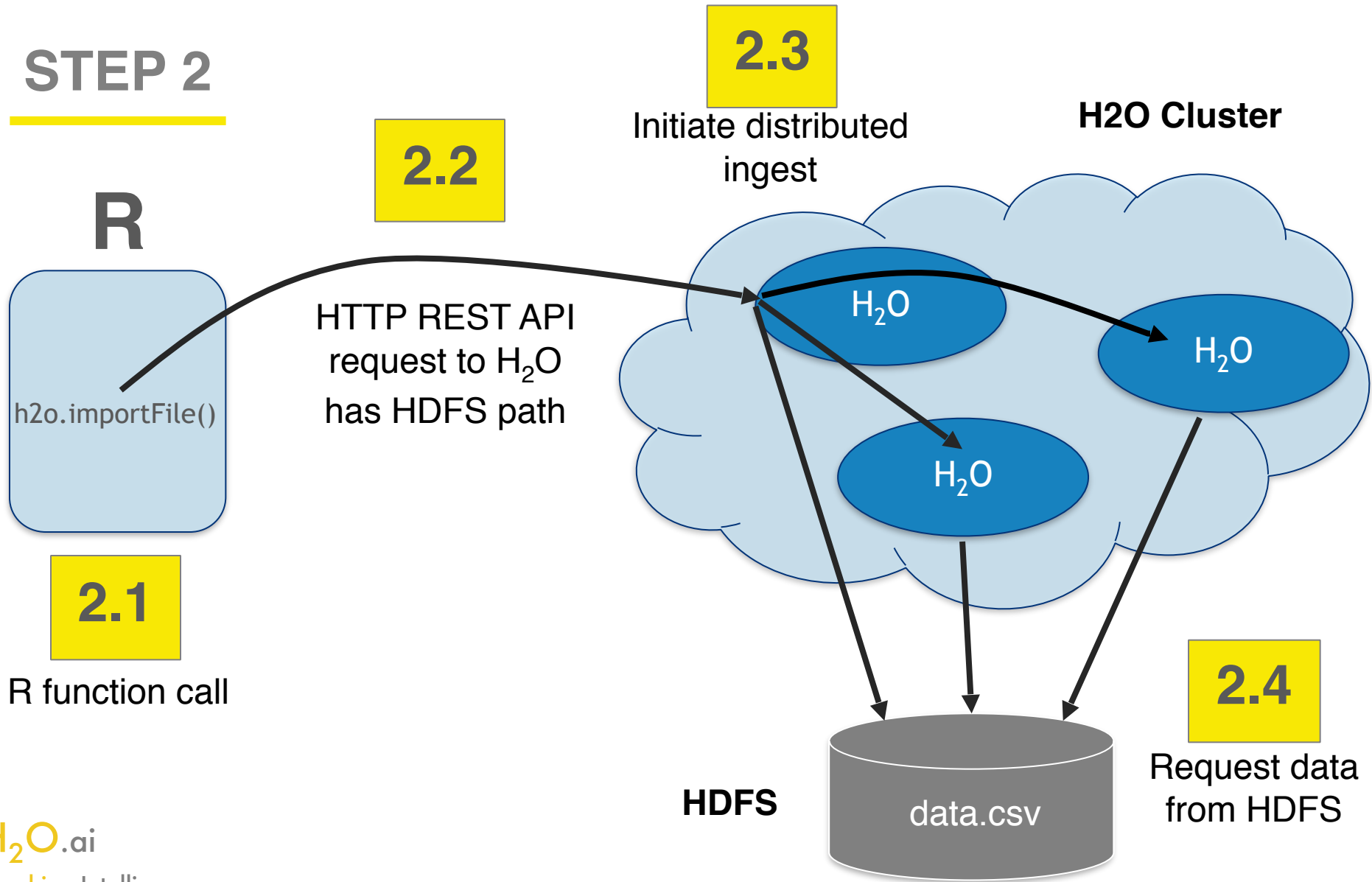
R user

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



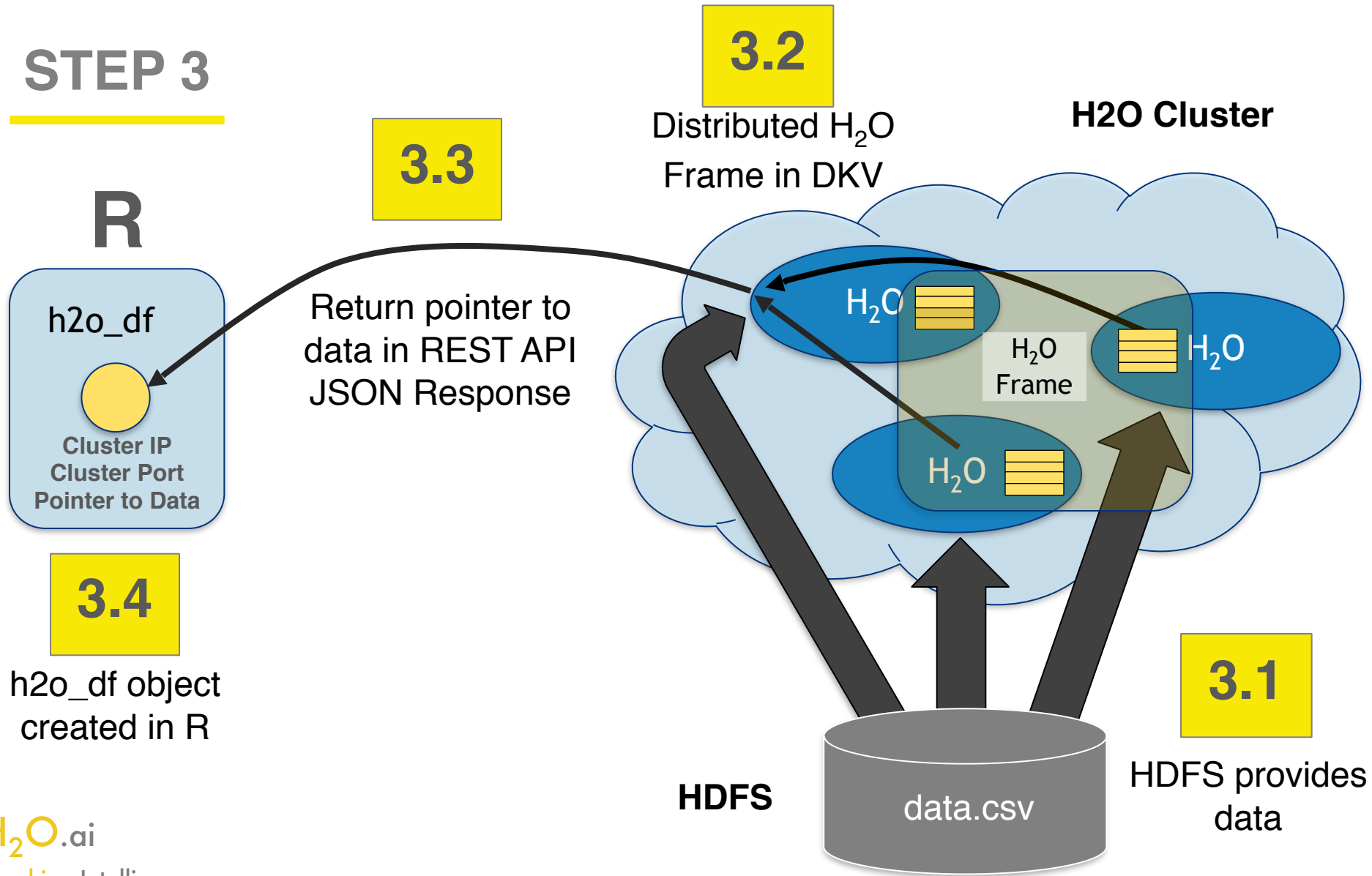
# Reading Data from HDFS into H2O with R

## STEP 2

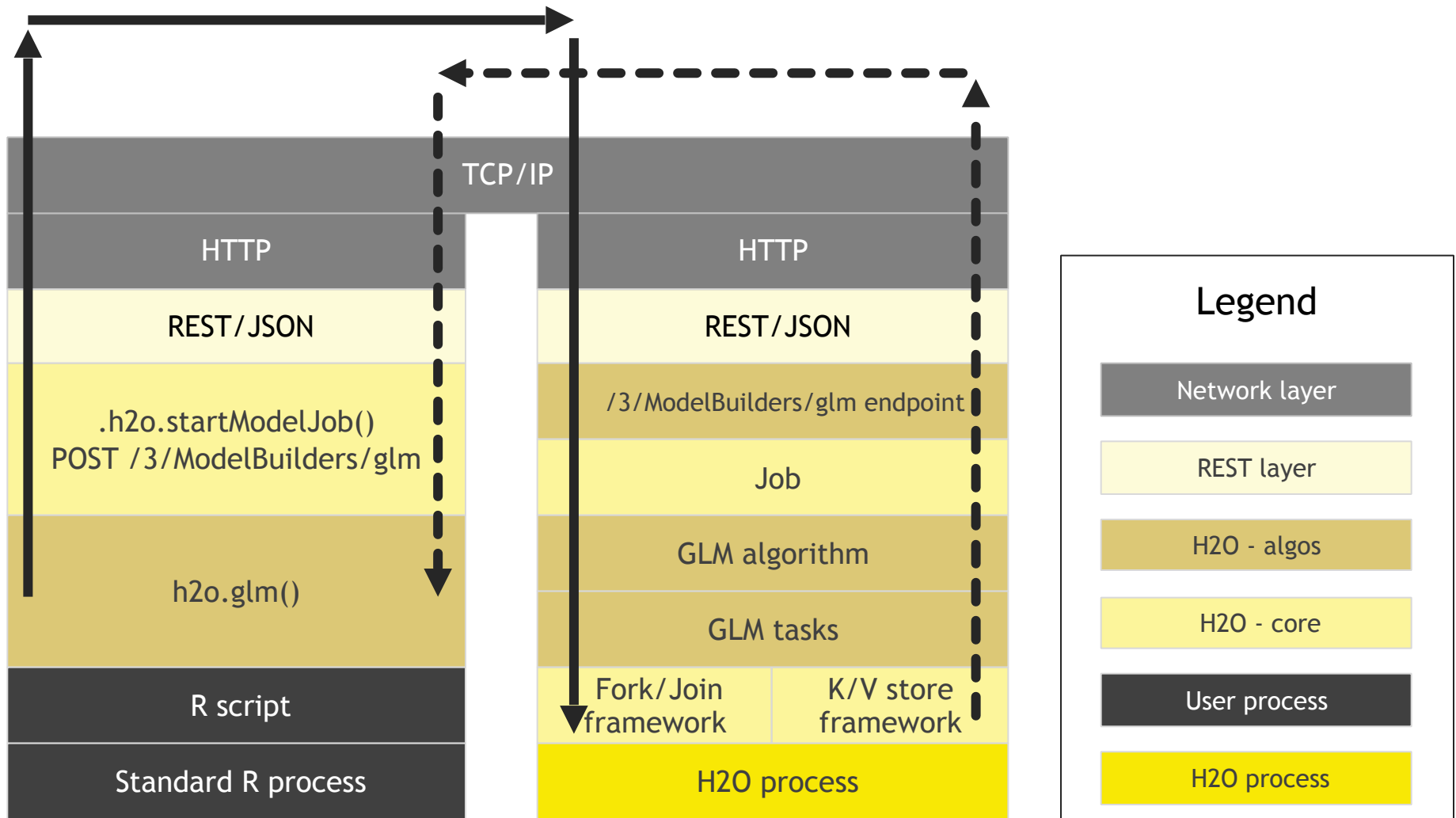


# Reading Data from HDFS into H2O with R

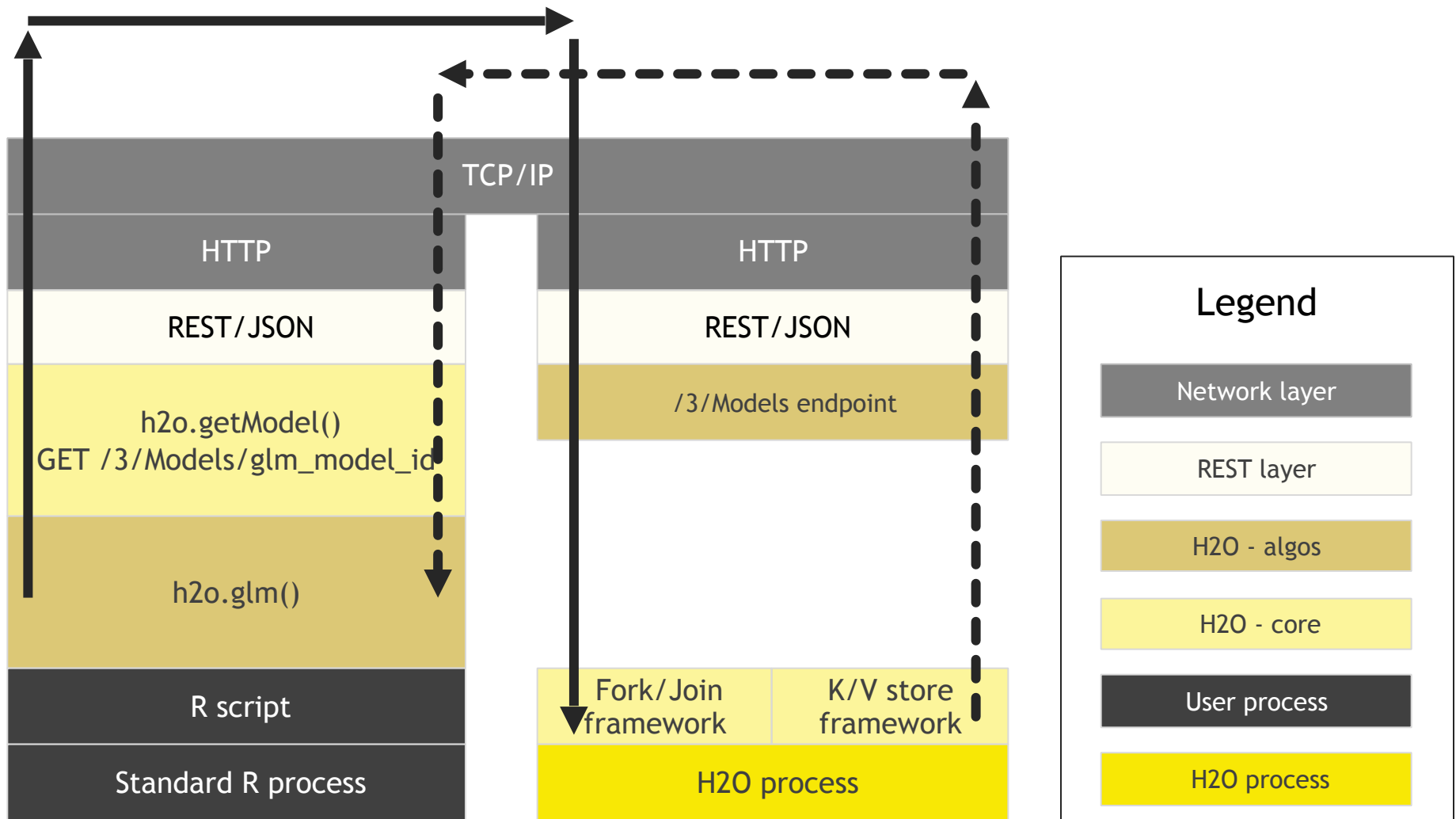
## STEP 3



# R Script Starting H2O GLM



# R Script Retrieving H2O GLM Result



# Step 1

- Download and install h2o: [h2o.ai](http://h2o.ai), hit download button
- SL Guest, password SerendipityOSW
- Only requirement is JDK 1.7+
- plus required packages if using R or Python
- Pick R, Python (2.7.x), or Standalone for tonight

# Step 2

- `http://bit.ly/1SNkmFa`
- Contains training and validation data, starter R and Python scripts
- Unzip `chicago.zip`

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

Thank You

(final holdout test set for Chicago Meetup is at <https://s3.amazonaws.com/0xdata-public/hank/mikeditka.csv>)

Thank you Chicago for a great time!