Hands-on Lab on Sparkling Water

AI Ukraine
Kyiv, October 13-14 2018

Jakub Háva,
jakub@h2o.ai
https://github.com/jakubhava
https://www.linkedin.com/in/havaj/

Who We Are

- **Kuba**
  - Senior Software engineer at H2O.ai - Core Sparkling Water
  - Master's at Charles University (CZ)
  - Implemented high-performance cluster monitoring tool for JVM based languages (JNI, JVMTI, instrumentation)
- **Michal**
  - VP of Engineering at H2O.ai
  - Creator of Sparkling Water
  - Ph.D at Charles University (CZ), PostDoc at Purdue

H2O+Spark = Sparkling Water

H2O + Spark

- H2O
  - Machine Learning Library
  - Distributed Algorithms
  - For ML experts

- Sparkling Water
  - Integrates H2O & Spark Ecosystems
  - Transparent for Spark users
  - Based on Spark pipelines & H2O

Benefits

Spark

- **Additional algorithms**
  - **NLP**
- **Powerful data munging**
- **ML Pipelines**

H$_2$O.ai

- Advanced algorithms
  - speed v. accuracy
  - advanced parameters
- Fully distributed and parallelised
- Graphical environment
- R/Python interface

Prepare the Environment

Prepare Spark Environment

- Download Spark 2.3.2 from https://www.apache.org/dyn/closer.lua/spark/spark-2.3.2/spark-2.3.2-bin-hadoop2.7.tgz
- Extract the downloaded file
- In your shell
  - `export SPARK_HOME=/path/to/extracted/spark`
  - `export MASTER=local`

Prepare Python Environment

- virtualenv -p /<path>/<to>/<python2.7+> sw_venv
- source sw_venv/bin/activate
- pip install -r requirements.txt

Prepare Sparkling Water Environment

- Download Sparkling Water from https://s3.amazonaws.com/h2o-release/sparkling-water/rel-2.3/15/index.html to the cloned repo
- Extract the downloaded zip file & cd inside

Open the Jupiter Notebook

- First, start Jupiter with Sparkling Water environment enabled

```
PYSPARK_DRIVER_PYTHON="ipython" PYSPARK_DRIVER_PYTHON_OPTS="notebook " sparkling-
water-2.3.15/bin/pysparkling --driver-memory 4G
```

- Open the AmazonFineFoodPipeline.ipynb notebook

Let's Give it a Try

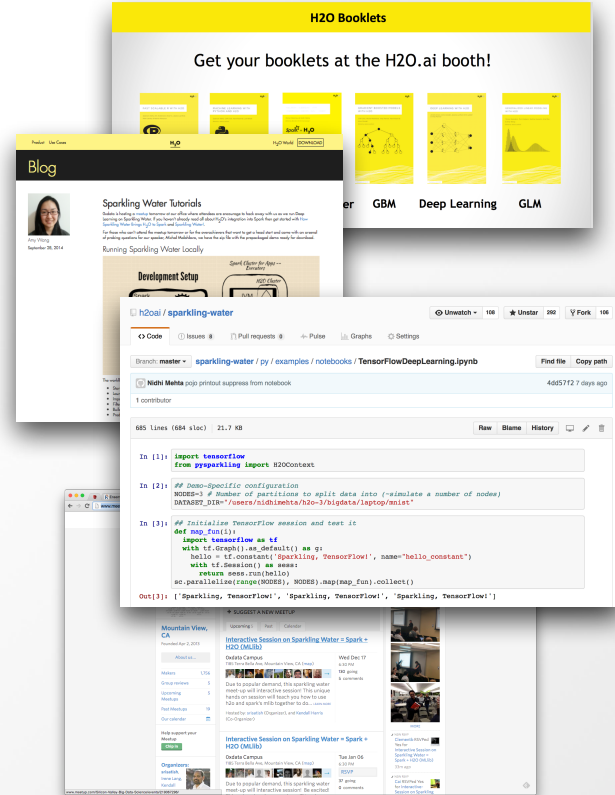# More Info

Checkout **H2O.ai** Training Books

http://h2o.ai/resources

Checkout **H2O.ai** Blog

http://h2o.ai/blog/

Checkout **H2O.ai** Youtube Channel

https://www.youtube.com/user/0xdata

Checkout GitHub

https://github.com/h2oai/sparkling-water

Thank you!

H2O.ai

Sparkling Water is
open-source
ML application platform
combining
power of Spark and H2O

I WANT YOU TO TRY SPARKLING WATER

www.says-it.com/unclesam/

Learn more at **h2o.ai**
Follow us at **@h2oai**

**H2O AI World London
29 & 30 October**

50 Free Tickets For AI Ukraine!

- **H2O.ai** is a **Leader** in 2018 Gartner Magic Quadrant for Data Science & Machine Learning Platform.
- **H2O AI World** is our **Community Event**.
  - 29th October: Hands-on H_2O Training
  - 30th October: Tech Talks, Real-World Use Cases, Kaggle Grandmasters Panel & More

Register at: **world.h2o.ai**
Promotional Code (First Come, First Served): **AIUKRAINE50**