

H2O Machine Learning Platform



R-Ladies Helsinki
June 2020

Erin LeDell Ph.D.
@ledell

What is H2O?

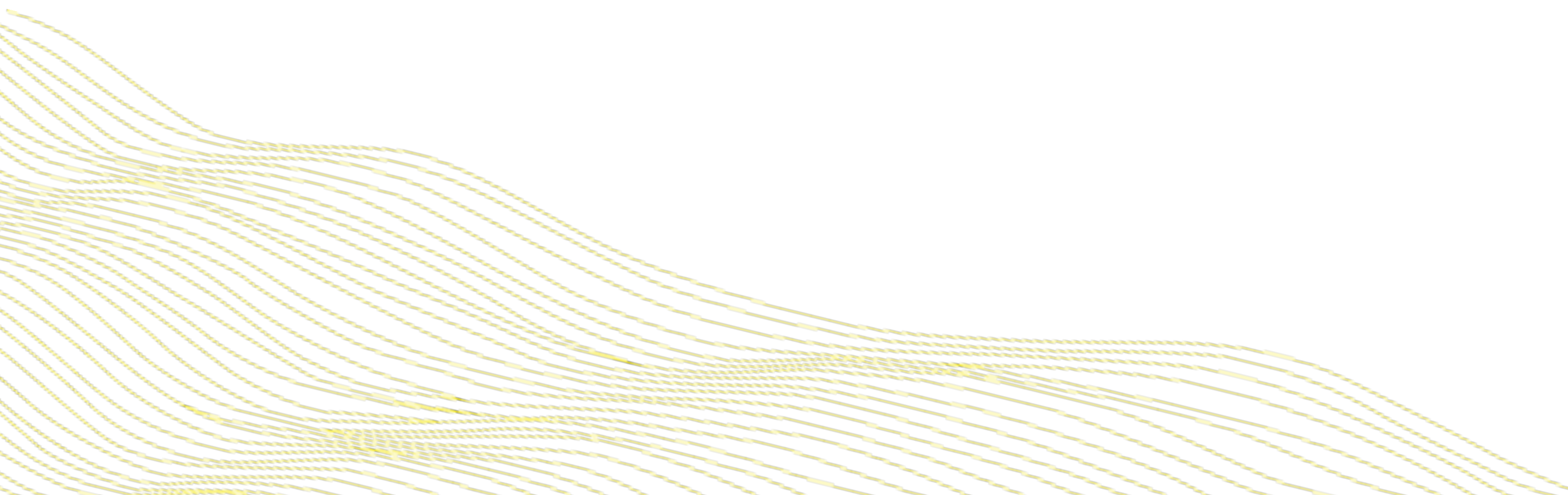


H2O.ai, the
company

H2O, the
platform

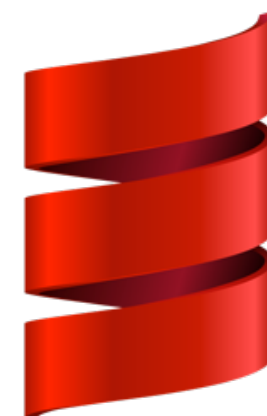
- Founded in 2012
 - Advised by Stanford Professors Hastie, Tibshirani & Boyd
 - Headquarters: Mountain View, California, USA
-
- Open Source Software (Apache 2.0 Licensed)
 - R, Python, Scala, Java and Web Interfaces
 - Distributed Machine Learning Algorithms for Big Data

H2O Platform



H2O Machine Learning Platform

- Distributed (multi-core + multi-node) implementations of cutting edge ML algorithms.
- Core algorithms written in high performance Java.
- APIs available in R, Python, Scala; web GUI.
- Easily deploy models to production as pure Java code.
- Works on Hadoop, Spark, EC2, your laptop, etc.



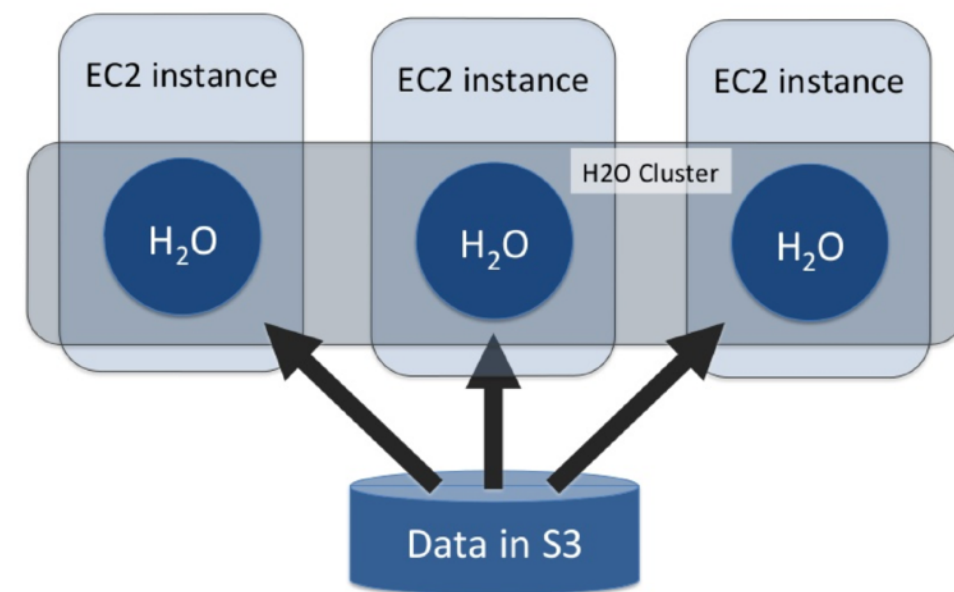
H2O Machine Learning Features



- Supervised & unsupervised machine learning algos (GBM, RF, DNN, GLM, Stacked Ensembles, etc.)
- Imputation, normalization & auto one-hot-encoding
- Automatic early stopping
- Cross-validation, grid search & random search
- Variable importance, model evaluation metrics, plots

H2O Distributed Computing

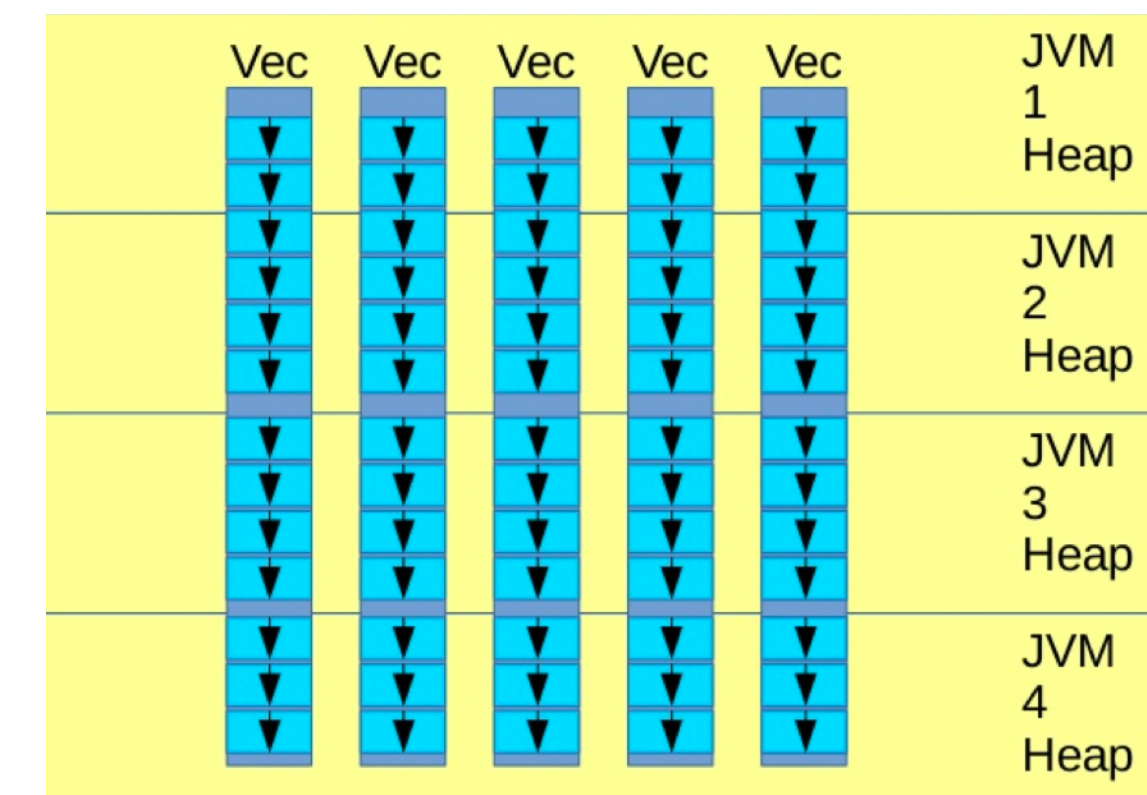
H2O Cluster



- Multi-node cluster with shared memory model.
- All computations in memory.
- Each node sees only some rows of the data.
- No limit on cluster size.

- Distributed data frames (collection of vectors).
- Columns are distributed (across nodes) arrays.
- Works just like R's data.frame or Python Pandas DataFrame

H2O Frame



H2O Machine Learning Tutorial



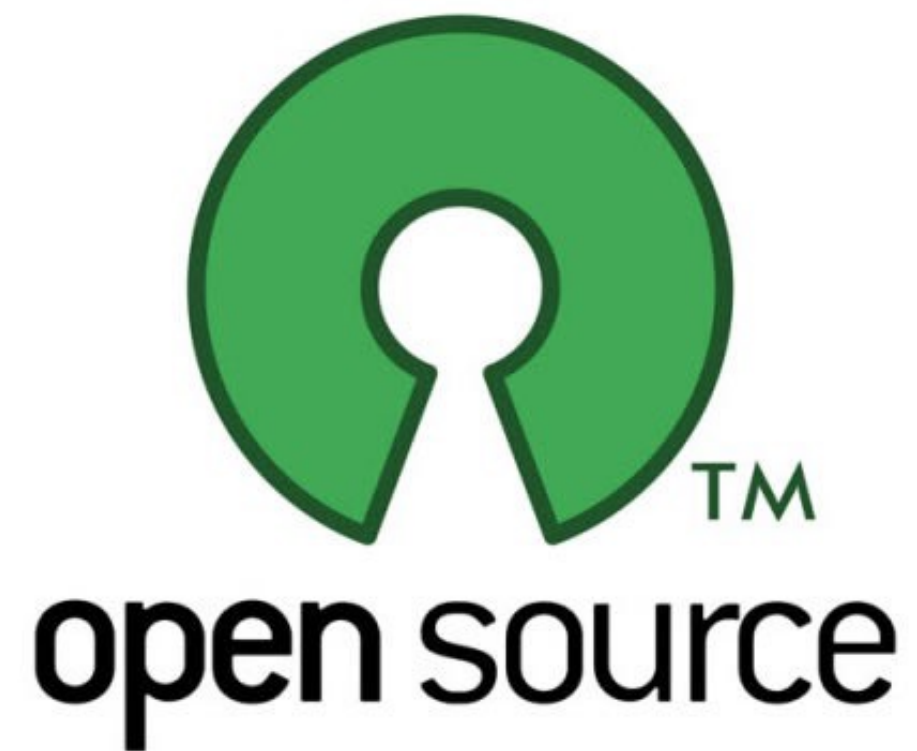
<https://tinyurl.com/latinr-h2o>

H2O Resources

- Documentation: <http://docs.h2o.ai>
- Tutorials: <https://github.com/h2oai/h2o-tutorials>
- Slidedecks: <https://github.com/h2oai/h2o-meetups>
- Videos: <https://www.youtube.com/user/0xdata>
- Stack Overflow: <https://stackoverflow.com/tags/h2o>
- Google Group: <https://tinyurl.com/h2ostream>
- Gitter: <http://gitter.im/h2oai/h2o-3>
- Events & Meetups: <http://h2o.ai/events>



Contribute to H2O!



Get in touch over email, Gitter or JIRA.

<https://github.com/h2oai/h2o-3/blob/master/CONTRIBUTING.md>

Thank you!

@ledell on Github, Twitter
erin@h2o.ai

