Ondřej Bílek
ondrej@h2o.ai

Introduction to Enterprise Steam with Sparkling Water

H2O.ai, Prague
Oct 16, 2018

About me

- System programming @ CTU in Prague
- Software Engineer @ H2O.ai
- Codebase owner of Enterprise Steam

The Hadoop problem

- Hadoop configuration is hard!
- Existing tools focus on admins, not users
- Multi-user environments are tricky
- Limited control over resources
- Security over convenience

H_2O
WORLD

H₂O Products

In-Memory, Distributed Machine Learning Algorithms with H2O Flow GUI

H2O AI Open Source Engine Integration with Spark

Lightning Fast machine learning on GPUs

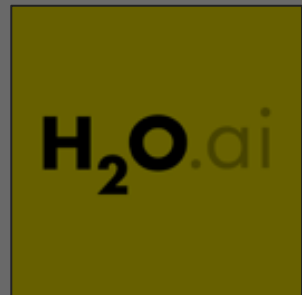Automatic feature engineering, machine learning and interpretability

Enterprise Steam

Secure multi-tenant H2O clusters

Enterprise Steam

- Multi-tenant system for managing H2O clusters on Hadoop
- Powerful admin capabilities
- Simplicity in mind
- Web frontend, Python & R API
- Enterprise grade security

Steam + Sparkling Water

H₂O.ai	**Spark + H₂O** SPARKLING WATER	**H2O4GPU**	DRIVERLESS AI
In-Memory, Distributed Machine Learning Algorithms with H2O Flow GUI	H2O AI Open Source Engine Integration with Spark	Lightning Fast machine learning on GPUs	Automatic feature engineering, machine learning and interpretability

Enterprise Steam

Secure multi-tenant H2O clusters

Why?

- Setting-up SW on Hadoop is not trivial
- Restricted to Hadoop node
- Reusable Spark sessions
- Pre-made configurations / templates
- Notebook integration (Jupyter)
- Security

Demo Time