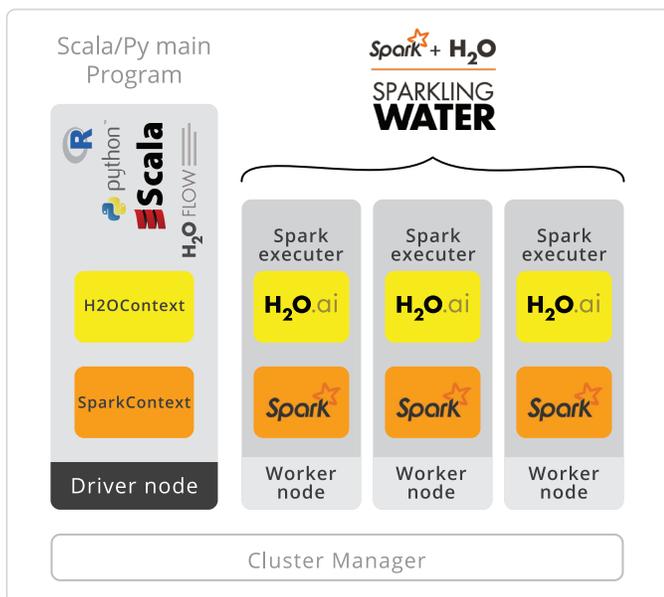


Sparkling Water

Sparkling Water blends data science workflows into developers' applications using H2O's machine algorithms and Spark's fast data munging capabilities. Sparkling Water enables usage of H2O algorithms with Spark Data Frames by providing a transparent API to exchange data between H2O Frames and Spark Data Frames.

Why Sparkling Water?



Sparkling Water was designed to allow users to get the best of Apache Spark - its elegant APIs, SQL-based data munging, machine learning pipelines - along with H2O's computation speed of fully-featured machine learning algorithms. Sparkling Water **also allows for greater flexibility when it comes to finding** the best algorithm for a given use case. Apache Spark's **MLib offers a library of popular algorithms directly built using Spark**. Sparkling Water empowers enterprise customers to use H2O algorithms in conjunction with, or instead of, MLib algorithms on Apache Spark.

1. **Parallelized data processing:** H2O is designed to quickly process huge amounts of data in a distributed and fully parallelized fashion.

Benefits

- Seamlessly transition back and forth between Spark and H2O
- Use Scala, Python or R to build models
- Power of Spark SQL-based data munging combined with the speed of H2O
- All the features of H2O included (Flow - UI, model export)

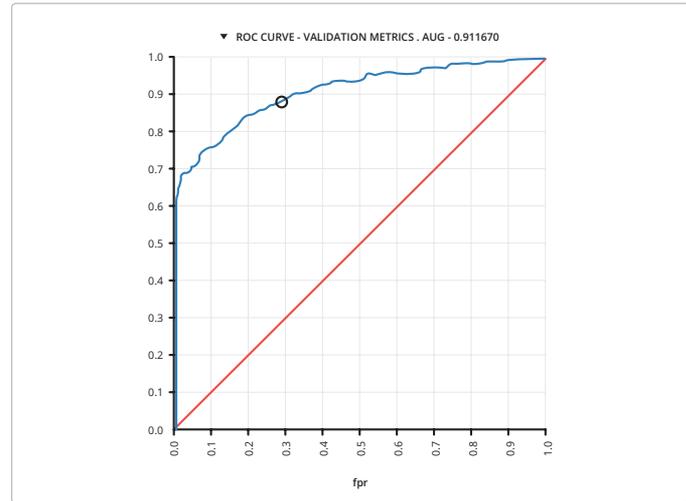
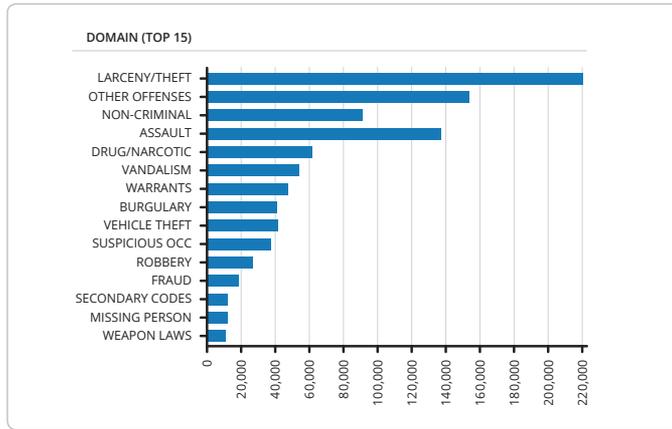
Highlights

- **Accuracy:** AutoML, Ensembles, GBM, GLM, DRF, Deep Learning
- **Speed:** In Memory, Distributed Computation
- **Interface:** R, Python, Flow
- **Developers:** Spark API, PySpark, Sparklyr
- **Community:** Expert Data Scientists, Developers, Data Engineers
- **Cloud:** Databricks Cloud, AWS, Azure

Features

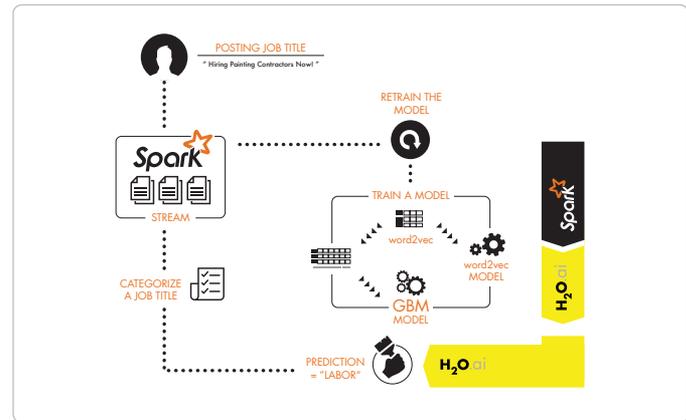
- Seamless integration with Spark API.
- Run Scala code in Flow.
- H2O algorithms are exposed as Spark estimator enabling transparent integration with Spark machine learning pipeline
- Bringing H2O's Visual Intelligence to MLib algorithms.
- Support of Driverless AI MOJO pipelines

2. **Streamline model training, evaluation & comparison and scoring:** H2O operationalizes this process by:
- a. Providing a library of ML algorithms **supporting advanced, algorithm-specific features**. Moreover,



H2O allows combining models into ensembles (super-learners) or finding the best model with AutoML.

- b. Performing **fast exploration of hyper-space of parameters** (a.k.a. grid search).
- c. Providing the ability to **specify various criteria** that identify and select the best model, e.g. accuracy, building time, scoring time, etc.
- d. **Ability to continue model training** with modified parameters and additional relevant input data.
- e. Continuous modeling feedback: **Visualization of various model characteristics** on-the-fly during training as well as of the final model.



Sparkling Water use-case example architecture

different execution environments and allow to manage H2O cluster as part of Spark cluster or separately.

3. **Deployment of optimized models:** Model deployment is one of the most critical elements of the machine learning process. H2O and Driverless AI allows for the export of trained models as an optimized code for deployment into target systems (i.e., web services, applications, etc.) The exported models can be also used as part of Spark machine learning pipelines.
4. **Sparkling Water deployment:** Easy use of Sparkling Water with existing Spark distribution with help of published Sparkling Water package. Moreover, Sparkling Water provides two operation modes (internal and external) which reflect demand of

H2O.ai is focused on bringing AI to businesses through software. Its flagship product is **H2O**, the leading open source platform that makes it easy for financial services, insurance and healthcare companies to deploy machine learning and predictive analytics to solve complex problems. More than **13,000 organizations** and 130,000+ data scientists depend on H2O for critical applications like predictive maintenance and operational intelligence. The company accelerates business transformation for 222 Fortune 500 enterprises, 8 of the world's 12 largest banks, 7 of the 10 largest auto insurance companies and all 5 major telecommunications providers.

Follow us on Twitter [@h2oai](#). To learn more about H2O customer use cases, please visit <http://www.h2o.ai/customers/>. [Join the Movement.](#)