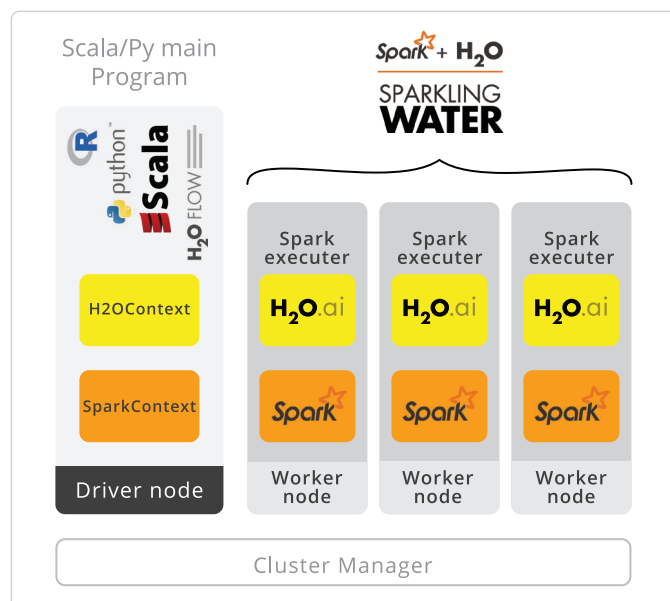


Sparkling Water

Sparkling Water blends data science workflows into developers' applications using H2O's machine learning technology and Spark's fast data munging capabilities. Sparkling Water provides the API calls to transform an H2O Frame to a Spark Data Frame, allowing access to Spark's SQL engine and Sparkling Water conversely transforms Spark Data Frames to H2O Frames for access to H2O's algorithms.

Why Sparkling Water?



Sparkling Water was designed to allow users to get the best of Apache Spark - its elegant APIs, RDDs, multi-tenant Context - along with H2O's speed, columnar-compression and 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 MLlib 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, MLlib algorithms on Apache Spark.

1. **Parallelized data processing:** H2O is designed to 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 combined with the speed of H2O
- All the features of H2O included (Flow - UI, model export)

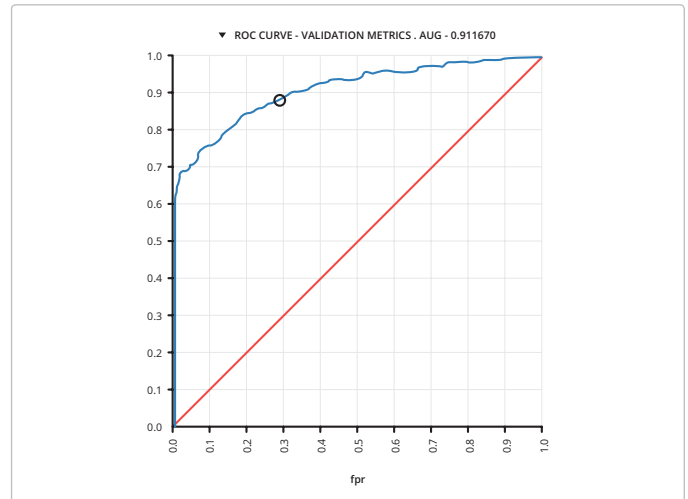
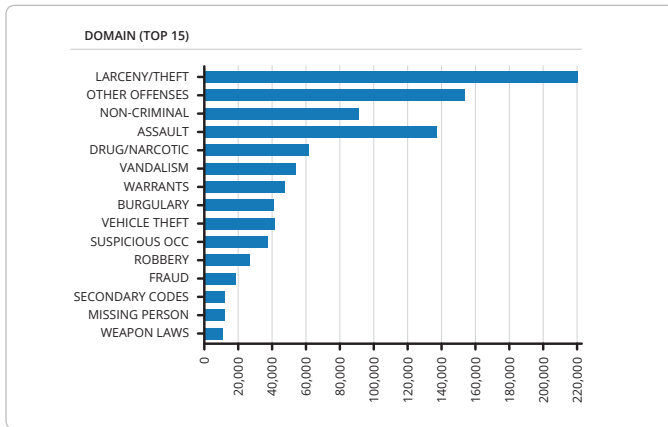
Highlights

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

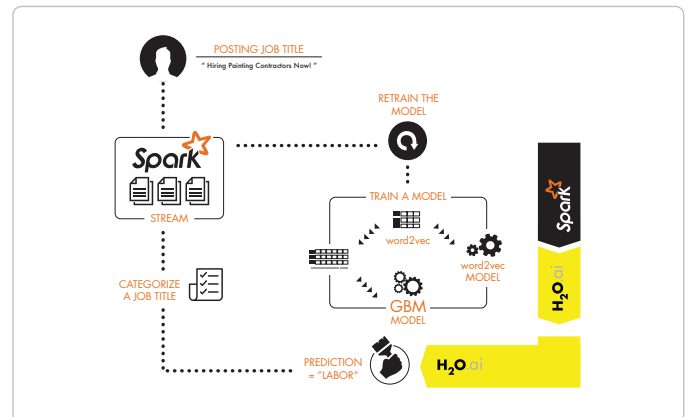
Features

- MLlib support in H2O Flow.
- Run Scala code in Flow.
- Export pipelines as executable java code for easy deployment (Spark logic + H2O models)
- Bringing H2O's Visual Intelligence for MLlib algorithms.

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).



- 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 preparation with modified parameters and additional relevant training data.
 - e. Visualization of various model characteristics on-the-fly during training as well as of the final model.
3. **Deployment of optimized models:** Model deployment is one of the most critical elements of the machine learning process. H2O allows for the export of trained models as an optimized code for deployment into target systems (i.e., web services, applications, etc.)



Sparkling use-case example architecture

About H2O.ai

At H2O.ai we see a world where all software will incorporate AI, and we're focused on bringing AI to business through software. H2O.ai is the maker behind H2O, the leading open source machine learning platform for smarter applications and data products. H2O operationalizes data science by developing and deploying algorithms and models for R, Python and the Sparkling Water API for Spark. Some of H2O's mission critical applications include predictive maintenance, operational intelligence, security, fraud, auditing, churn, credit scoring, user based insurance, predicting sepsis, ICU monitoring and more in over 5,000 organizations. H2O is brewing a grassroots culture of data transformation in its customer communities. Customers include Capital One, Progressive Insurance, Zurich North America, Transamerica, PricewaterhouseCoopers, Comcast, Nielsen Catalina Solutions, Neustar, Macy's, Walgreens, Kaiser Permanente and Aetna.

