

H2O

# **H2O.ai** **демократизирует** **искусственный** **интеллект**

В этом нет ничего искусственного

## **Движение ИИ повышает ценность бизнеса**

H2O.ai является создателем ведущей платформы машинного обучения и искусственного интеллекта с открытым исходным кодом, которой доверяют сотни тысяч ученых, занимающихся разработкой данных, что повышает ценность более чем 14 000 предприятий по всему миру. Видение состоит в том, чтобы демократизировать искусственный интеллект для всех с помощью наших отмеченных наградами платформ обработки данных.



Платформа машинного обучения № 1 с  
открытым исходным кодом для  
предприятия

## **Open Source, распределенное машинное обучение для всех**

H2O - это платформа с полностью открытым исходным кодом, распределенная в памяти, с линейной масштабируемостью. H2O поддерживает наиболее широко используемые алгоритмы статистического и машинного обучения, включая алгоритмы градиентного бустинга, обобщенные линейные модели, глубокое обучение и многое другое. H2O также обладает лидирующей в отрасли функциональностью AutoML, которая автоматически проходит через все алгоритмы и их гиперпараметры, создавая список лучших моделей. Платформа H2O используется более чем 14 000 организаций по всему миру и чрезвычайно популярна в сообществах R & Python.

# Ключевые особенности **H2O**

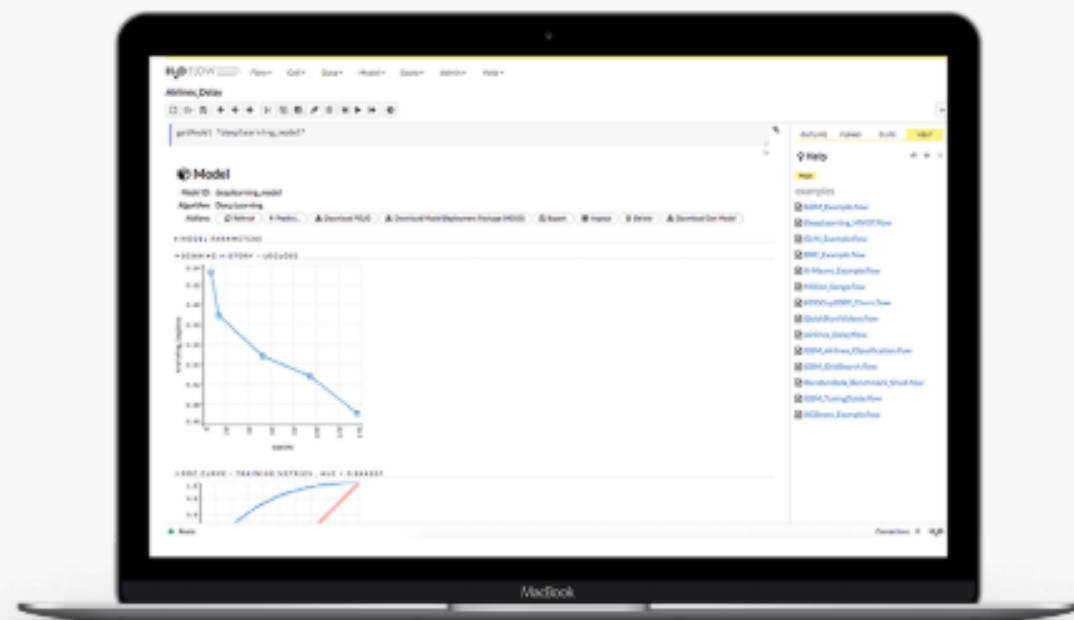
Ведущие алгоритмы

Доступ из R, Python, Flow и многое другое ...

AutoML (NEW)

Распределенная обработка в памяти

Простое развертывание



- Ведущие алгоритмы

Алгоритмы, разработанные с нуля для распределенных вычислений, а также для контролируемых и неконтролируемых подходов, включая Random Forest, GLM, GBM, XGBoost, GLRM, Word2Vec и многие другие.

- Доступ из R, Python, Flow и многое другое ...

Используйте язык программирования, который вы уже знаете, например, R, Python и другие, для построения моделей в H2O или используйте H2O Flow, графический интерактивный пользовательский интерфейс на основе ноутбука, который не требует никакого кодирования.

- AutoML (NEW)

AutoML от H2O можно использовать для автоматизации рабочего процесса машинного обучения, который включает автоматическое обучение и настройку многих моделей в течение установленного пользователем срока. Сложенные ансамбли будут автоматически обучаться на коллекциях отдельных моделей, чтобы создавать высокопроизводительные ансамблевые модели, которые в большинстве случаев будут самыми эффективными моделями в таблице лидеров AutoML.

- Распределенная обработка в памяти

Обработка в памяти с быстрой сериализацией между узлами и кластерами для поддержки больших наборов данных. Распределенная обработка больших данных обеспечивает скорость до 100 раз быстрее с мелкозернистым параллелизмом, обеспечивая оптимальную эффективность без ухудшения точности вычислений.

- Простое развертывание

Простота развертывания ROJO и MOJO для развертывания моделей для быстрого и точного скоринга в любой среде, включая очень большие модели.

# Как это устроено

## **Данные предприятия**

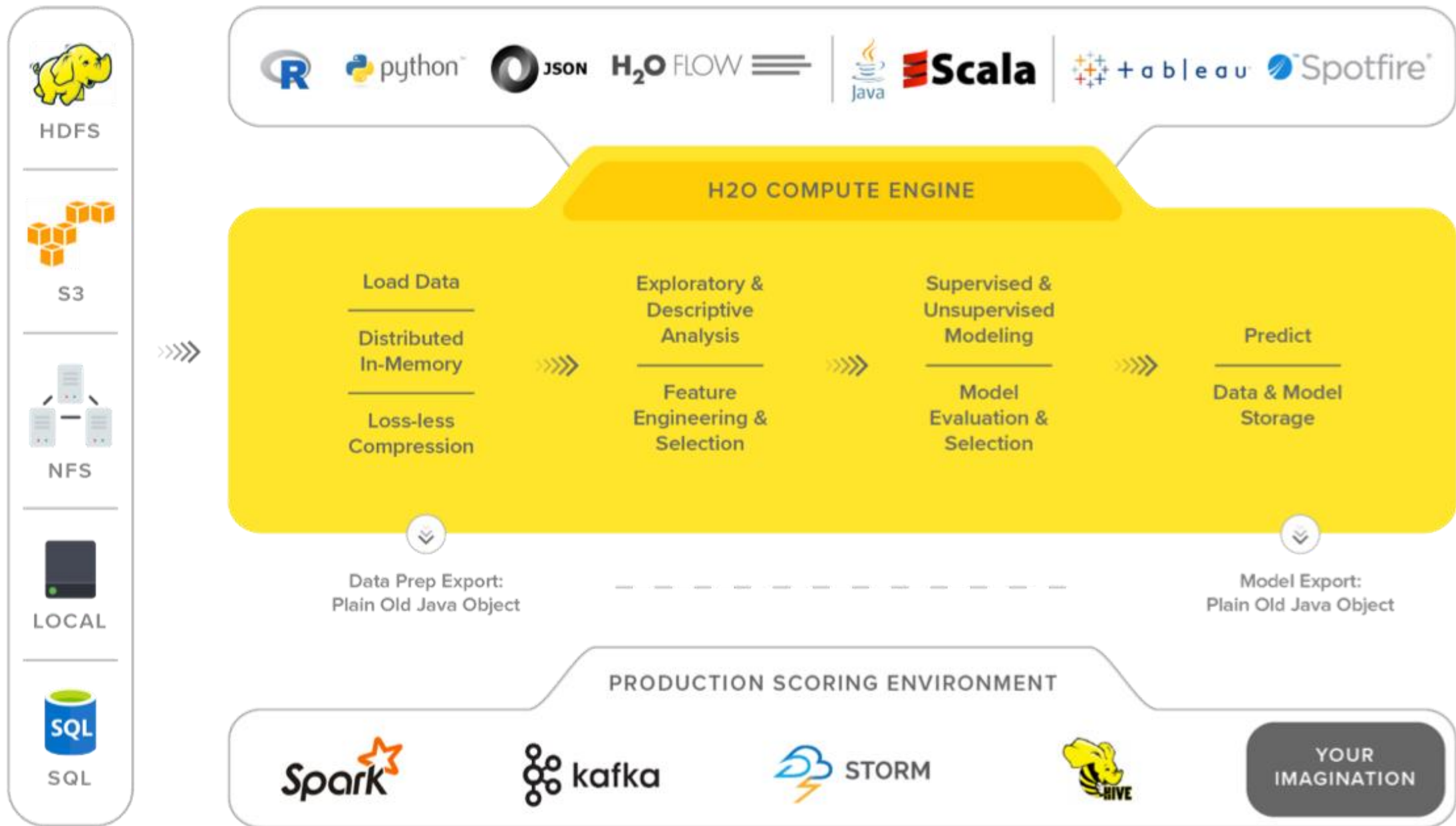
H2O работает на существующей инфраструктуре больших данных, на голом железе или поверх существующих кластеров Hadoop или Spark. Он может загружать данные напрямую из HDFS, Spark, S3, Azure Data Lake или любого другого источника данных в хранилище значений ключей в памяти.

## **Распределенное машинное обучение в памяти**

H2O использует преимущества вычислительных мощностей распределенных систем и вычислений в памяти для ускорения машинного обучения, используя параллельные отраслевые алгоритмы, которые используют преимущества мелкозернистого преобразования в памяти ..

## **Бесшовное развертывание**

Быстрое и простое внедрение моделей в производство с использованием Java (POJO) и двоичных форматов (MOJO). Кроме того, модели H2O могут производиться различными способами, как указано [здесь](#) .



ИТ-ландшафт H2O

# **H2O Sparkling Water**

The Best Machine Learning on Spark

## **Лучшее в обоих мирах с H2O и Spark**

Sparkling Water позволяет пользователям комбинировать быстрые, масштабируемые алгоритмы машинного обучения H2O с возможностями Spark. Spark - это элегантная и мощная универсальная платформа с открытым исходным кодом, находящаяся в оперативной памяти и обладающая огромной динамикой. H2O является платформой для машинного обучения в оперативной памяти, которая меняет то, как люди применяют математические и прогнозные аналитические решения для своих бизнес-задач. Интеграция этих двух сред с открытым исходным кодом обеспечивает беспроблемную работу для пользователей, которые хотят сделать запрос с использованием Spark SQL, передать результаты в H2O, чтобы построить модель и сделать прогнозы, а затем снова использовать результаты в Spark. Для любой данной проблемы лучшая совместимость между инструментами обеспечивает лучший опыт.



# Key Features of Sparkling Water

Доступ к алгоритмам H2O

Диск вычислений от Scala, R и многое другое  
...

Простое развертывание



- Access to H2O Algorithms

Доступ к алгоритмам H2O, разработанным с нуля для распределенных вычислений, а также для контролируемых и неконтролируемых подходов, включая Random Forest, GLM, GBM, XGBoost, GLRM, Word2Vec и многие другие.

- Drive Computation from Scala, R and more...

Управляйте вычислениями из Scala, R или Python и используйте пользовательский интерфейс H2O Flow, предоставляя идеальную платформу машинного обучения для разработчиков приложений.

- Simple Deployment

Простота развертывания POJO и MOJO для развертывания моделей для быстрого и точного скоринга в любой среде, включая очень большие модели.

# Как это устроено

## **Распределенное машинное обучение в памяти**

Sparkling Water предназначена для использования в качестве обычного приложения Spark.

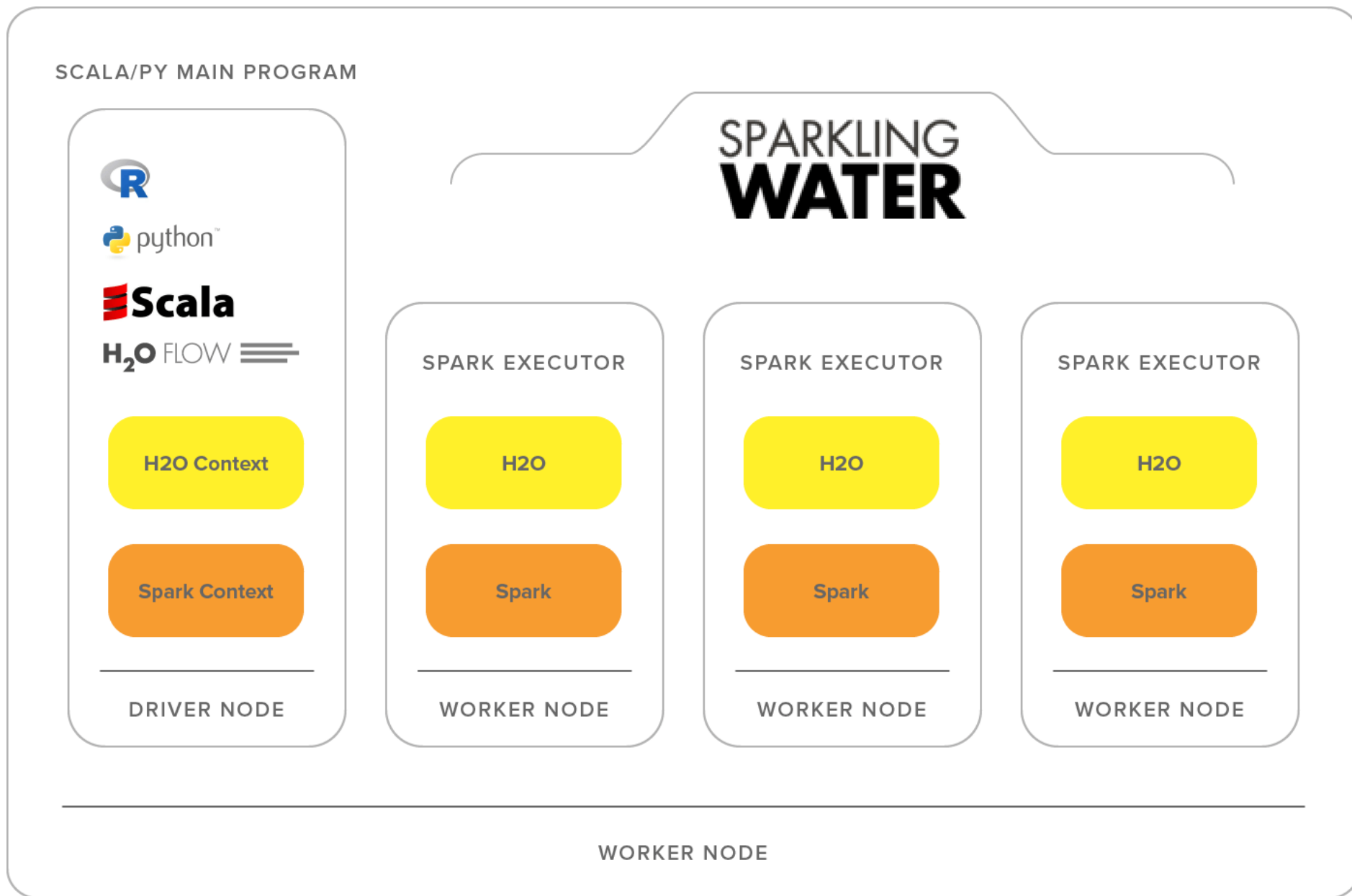
Она предоставляет способ инициализации сервисов H2O в Spark и доступа к данным, хранящихся в структурах данных Spark и H2O.

## **Усовершенствованное машинное обучение для Spark**

Используйте лучшие алгоритмы для распределенных вычислений в памяти с вашей существующей реализацией Spark.

## **Развертывание результатов в Spark**

Результаты от H2O могут быть легко развернуты с использованием конвейера машинного обучения с низкой задержкой H2O или в Spark для оценки.



ИТ-ландшафт Sparkling Water

# **H2O Driverless AI**

## **Ускорьте свою стратегию ИИ**

H2O Driverless AI использует методы людей экспертов-исследователей данных в простом в использовании приложении, которое помогает масштабировать ваши усилия в области обработки данных. Driverless AI дает ученым, работающим с данными, возможность быстрее работать над проектами, используя автоматизацию и самые современные вычислительные возможности графических процессоров, для выполнения задач за считанные минуты, которые раньше занимали месяцы.

С помощью Driverless AI каждый (включая экспертов и младших исследователей данных, специалистов в области предметной области и инженеров данных) может разработать надежные модели машинного обучения. Эта платформа автоматического машинного обучения следующего поколения предоставляет уникальные и расширенные функциональные возможности для визуализации данных, разработки функций, интерпретации моделей и развертывания с малыми задержками.

# Ключевые особенности H2O Driverless AI

Автоматическое проектирование

Интерпретируемость машинного обучения (MLI)

НЛП с TensorFlow

Автоматический скоринг трубопроводов

Временные ряды

Автоматическая визуализация

Гибкость данных и развертывания

NVIDIA GPU Ускорение



## Автоматическое проектирование

Разработка (инженерия) функций (фичей) - это секретное оружие, которое продвинутые ученые используют для получения наиболее точных результатов из алгоритмов. H2O Driverless AI использует библиотеку алгоритмов и преобразований функций, чтобы автоматически создавать новые, ценные функции для данного набора данных.

## Интерпретируемость машинного обучения (MLI)

H2O Driverless AI обеспечивает надежную интерпретацию моделей машинного обучения для объяснения результатов машинного обучения. В представлении MLI H2O Driverless AI использует множество различных методов и методологий для интерпретации и объяснения результатов своих моделей, автоматически генерируются четыре диаграммы, в том числе: LIME, Shapley, Variable Importance, Decision Tree, Partial Dependence и многое другое.

## NLP (Natural language processing) с TensorFlow

Текстовые данные могут содержать критически важную информацию для более точных прогнозов. Driverless AI автоматически преобразует короткие текстовые строки в функции, используя мощные методы, такие как TF-IDF (***TF*** — *term frequency*, ***IDF*** — *inverse document frequency*). С TensorFlow Driverless AI может также обрабатывать большие текстовые блоки и строить модели, используя все доступные данные, и решать бизнес-задачи, такие как анализ настроений, классификация документов и тегирование контента.

## Automatic Scoring Pipelines

Автоматически генерирует как конвейеры оценки Python, так и новые конвейеры автоматической оценки со сверхнизкими задержками. Новый конвейер автоматического скоринга - это уникальная технология, в которой используются все возможности проектирования и выигрышная модель машинного обучения в высокооптимизированном готовом Java-коде с минимальными задержками, который можно развернуть где угодно.

## **Time Series**

H2O Driverless AI обеспечивает превосходные возможности временных рядов для оптимизации практически для любого временного окна прогнозирования, включает в себя данные от многочисленных предикторов, обрабатывает структурированные символьные данные и категориальные переменные с высокой степенью кардинальности, а также обрабатывает пробелы в данных временных рядов и других пропущенных значениях.

## **Automatic Visualization**

H2O Driverless AI автоматически генерирует визуализации и создает графики данных, которые наиболее актуальны с точки зрения статистики на основе наиболее релевантной статистики данных, чтобы помочь пользователям быстро понять свои данные до начала процесса построения модели.

## **Гибкость данных и развертывания**

H2O Driveless AI работает с различными наборами данных, включая Hadoop HDFS, Amazon S3 и многие другие. H2O Driver AI можно развернуть повсюду, включая все облака (Microsoft Azure, AWS, Google Cloud) и локально в любой системе, но он идеально подходит для систем с графическими процессорами, включая IBM Power 9 со встроенными графическими процессорами.

H2O Driverless AI

## **NVIDIA GPU Acceleration**

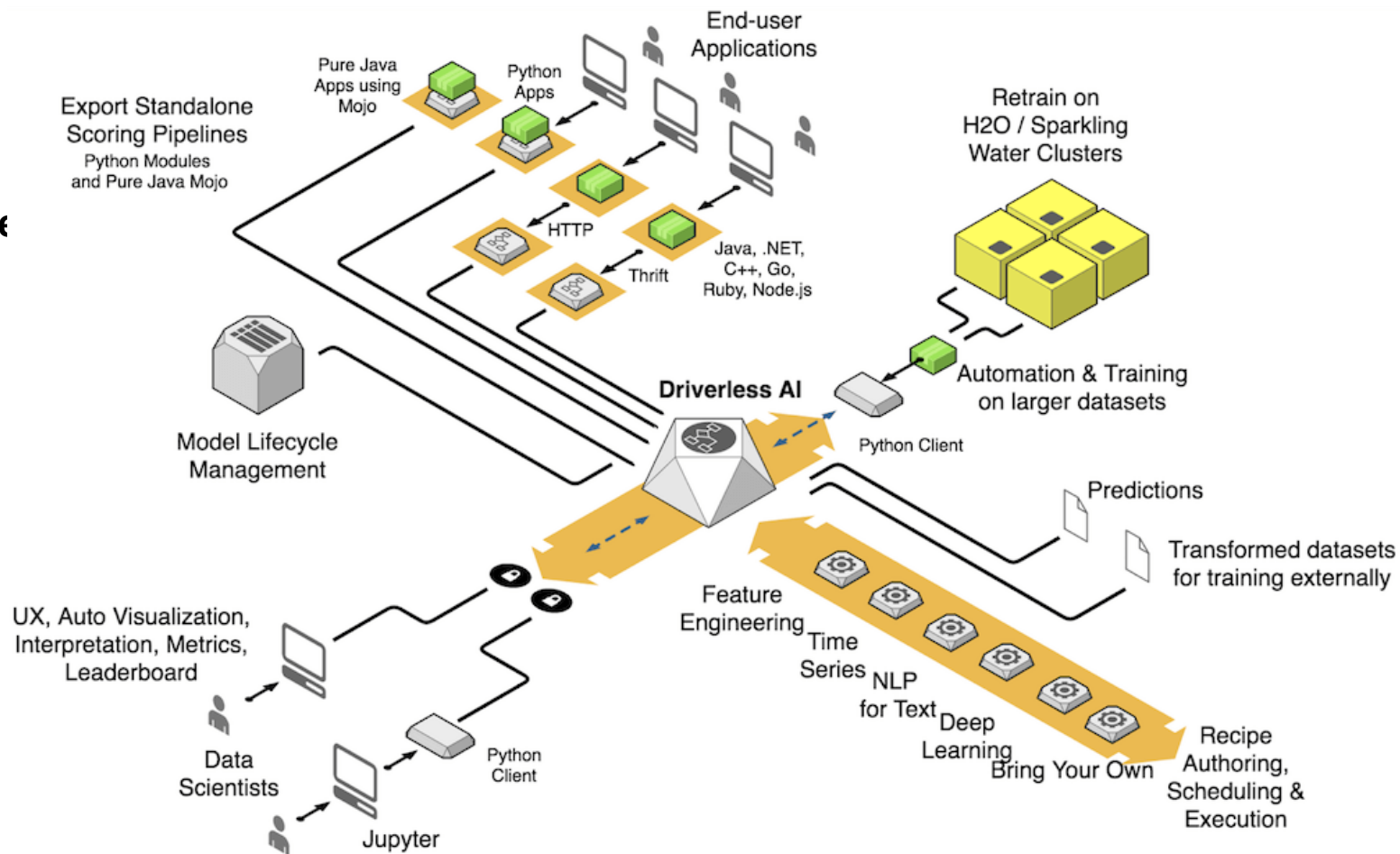
H2O Driverless AI оптимизирован, чтобы использовать преимущества ускорения графического процессора для достижения ускорения до 40X для автоматического машинного обучения. Включая алгоритмы с несколькими графическими процессорами для XGBoost, GLM, K-Means и других. Графические процессоры позволяют выполнять тысячи итераций функций и оптимизаций модели.



# Как это устроено

## Ускорьте проекты машинного обучения, чтобы получить представление о бизнесе

H2O Driverless AI упрощает применение передового машинного обучения для решения современных бизнес-задач.



## Выживание сильнейшего

- Используйте высокопроизводительные вычисления для создания и тестирования тысяч итераций модели за считанные минуты.
- Избегайте распространенных ошибок в науке о данных
- Используйте лучшие практики экспертов в области данных, чтобы избежать ошибок, таких как утечка данных, переобучение и многое другое.

## **Внедрение инноваций в бизнесе с машинным обучением с ускорением на GPU**

H2O4GPU - это пакет машинного обучения с ускорением на GPU с открытым исходным кодом и API-интерфейсами на Python и R, который позволяет любому использовать преимущества графических процессоров для создания современных моделей машинного обучения. Доступно множество популярных алгоритмов, включая машины повышения градиента (GBM), обобщенные линейные модели (GLM) и кластеризацию K-means. Тесты показали, что модели машинного обучения на графических процессорах были в 40 раз быстрее, чем на CPU.

# Основные характеристики H2O4GPU

Оптимизирован для производительности GPU

Широкий выбор алгоритмов с поддержкой графического процессора

Основывается на Scikit-Learn Python API

Доступный R API



### **Оптимизирован для производительности GPU**

Полностью оптимизирован для работы на графических процессорах NVIDIA® Volta последнего поколения, программном обеспечении NVIDIA Tesla® V100 и CUDA 9.

### **Широкий выбор алгоритмов с поддержкой графического процессора**

Доступные алгоритмы включают GBM (Gradient Boosting Machine), обобщенные линейные модели (GLM), кластеризацию K-means, SVD, PCA, K-means и XGBoost для более быстрой доставки моделей и результатов в бизнес.

### **Основывается на Scikit-Learn Python**

API H2O4GPU - это коллекция вычислителей графических процессоров с открытым исходным кодом, созданная H2O.ai. Она основана на простом в использовании Python API-интерфейсе scikit-learn и хорошо протестированных алгоритмах на базе CPU. Она может быть использована в качестве замены для scikit-learn с поддержкой графических процессоров на выбранных (и постоянно растущих) алгоритмах.

### **Доступный R API**

Новый R API предоставляет преимущества машинного обучения с GPU-ускорением сообществу пользователей R. Пакет R является оболочкой для пакета Python H2O4GPU, а интерфейс соответствует стандартным соглашениям R для моделирования.

## Поддержка предприятия

Когда ИИ становится критически важным для успеха предприятия, H2O.ai должен помочь. Служба поддержки H2O Enterprise Services предоставляет услуги, необходимые для оптимизации ваших инвестиций в людей и технологии для реализации вашего видения ИИ. Поддержка H2O Enterprise включает в себя обучение, специальный менеджер по работе с клиентами, круглосуточную поддержку, ускоренное решение проблем и прямые запросы на улучшение. Корпоративная поддержка также предоставляет вам доступ к экспертам H2O по науке о данных, платформе H2O и разворачиванию DevOps / production, чтобы ускорить и расширить использование AI.

