SOLUTION BRIEF

Artificial Intelligence
Produce and Food Services

Hortifrut Uses Accelerated Machine Learning to Predict Produce Quality and Faster Farm to Fork Distribution

Leading Global Producer and Distributor of Blueberries Solves Data Science Business Challenges with H2O Driverless AI* on Intel® Architecture

INTEL® AI BUILDERS MEMBER

H2O.ai

When it comes to perishable groceries, fruit has a short life. Optimizing the time from farm to fork is an objective that benefits the grower, grocer, and customer. But a lot of factors can go into predicting fruit quality so it can be brought to the table before its spoilage date. Today, companies can leverage artificial intelligence (AI) and machine learning to get product to consumers quickly.

Based in Chile, Hortifrut is the largest producer of blueberries in the world. They also grow strawberries and raspberries. The company operates farms in Peru, Chile, Mexico, Argentina, the United States, Spain, Morocco, and China, with distribution of fruit across 37 countries.

Optimizing distribution processes to maximize crop yields while having enough time to get produce to grocers—and consumers—requires many factors. Hortifrut uses H2O.ai Driverless AI platform on Intel® Architecture to make distribution decisions across their expansive operations. They are able to predict the quality of the blueberries from origin to final destination, improving the produce freshness delivered to the table and increasing revenue.

H2O.ai—AI To Do AI

H2O.ai is an open source leader in artificial intelligence (AI) and machine learning (ML). Its commercial automatic machine learning platform, H2O Driverless AI, provides critical machine learning capabilities that accelerate model building. H2O Driverless AI uses automation and state-of-the-art computing power to allow data scientists or data analysts to accomplish tasks in minutes or hours compared to what takes humans months. The platform delivers unique and advanced functionality for data visualization, model interpretability, and low-latency deployment. Capabilities include:

- **Feature engineering.** Feature engineering allows data scientists to extract the most accurate results from algorithms. Driverless AI employs a library of algorithms and feature transformations to automatically engineer new, high-value features for a given data set.
- **Explainability.** Driverless AI provides robust interpretability of machine-learning models to explain modeling results. The system generates charts, including K-LIME,* Shapley,* Variable Importance, Decision Tree, Partial Dependence, and more.
- **Natural Language Processing (NLP).** Text data can contain critical information to inform better predictions. H2O Driverless AI automatically uses powerful NLP techniques to convert short text strings into features.

- **Scoring pipelines.** Driverless AI automatically generates Python* scoring pipelines and new ultra-low latency automatic scoring pipelines. This allows for highly optimized, low-latency, production-ready Java* code that can be deployed anywhere.
- **Time series.** Driverless AI delivers superior time-series capabilities to optimize for almost any prediction time window. The solution also incorporates data from numerous predictors; handles structured character data and high-cardinality categorical variables; and manages gaps in time-series data and other missing values.
- **Visualization.** To help users get a quick understanding of their data before model building begins, Driverless AI automatically generates visualizations and creates relevant data plots.
- **Data sets.** Driverless AI works across varied data sets, including Apache Hadoop* Distributed File System (HDFS) and Amazon S3.*

These powerful features deliver a data science platform that addresses the needs of a variety of use cases for enterprises across industries.

Getting Better Blueberries Farm to Fork

Transporting fruit from the farm may take weeks, so Hortifrut must be able to predict quality upon arrival. With Driverless AI, the company has better predictive insights into the quality of their blueberries and can predict freshness of fruit on arrival at the retail outlet. Data analysts can use that information to determine the best method (fastest and least expensive) to ship the fruit that helps ensure quality upon arrival. This saves the company time and money and reduces spoilage claims from customers.

"We are getting great results with H2O Driverless AI," said Gonzalo Bustos, Head of Data Analytics at Hortifrut. "We are building hundreds of models to ensure a quality product arrives at its destination. What once took us three to five months using Python* and AutoML* can now be done in three to five weeks with Driverless AI without having to add any additional data scientists to the team."

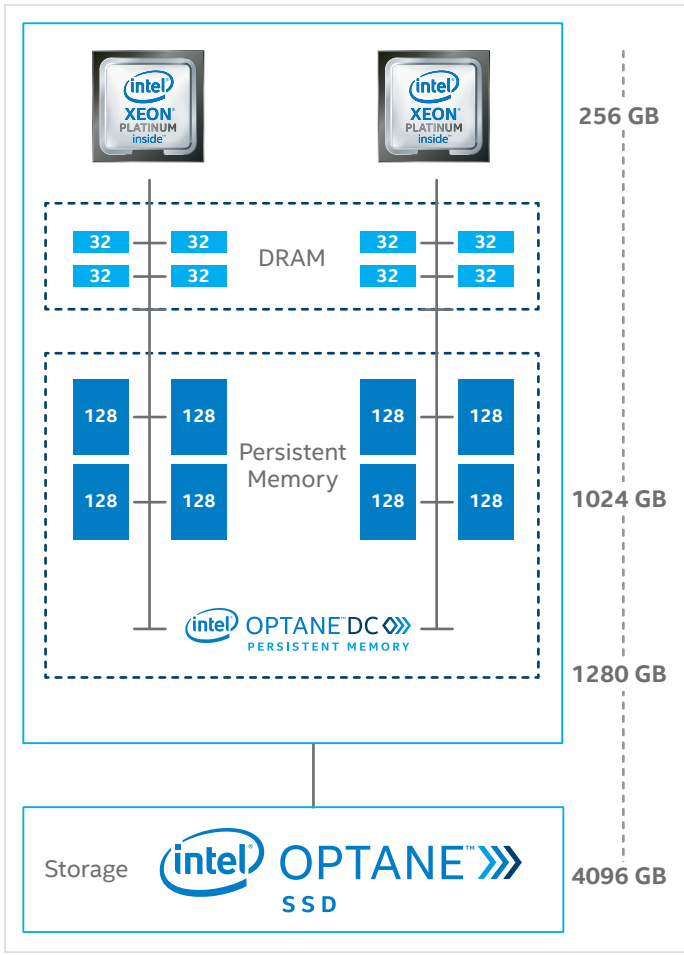
Hortifrut runs Driverless AI on servers based on Intel® Xeon® Scalable processors. By using Driverless AI on Intel Architecture, Hortifrut is able to scale their data science efforts to incorporate multiple factors, including varietal, farm origin, weather and climate, shipping time, vessel, and packaging. The company also uses Driverless AI to predict future production of their blueberries, strawberries, and raspberries depending on the origin, weather, varietal, and other factors.

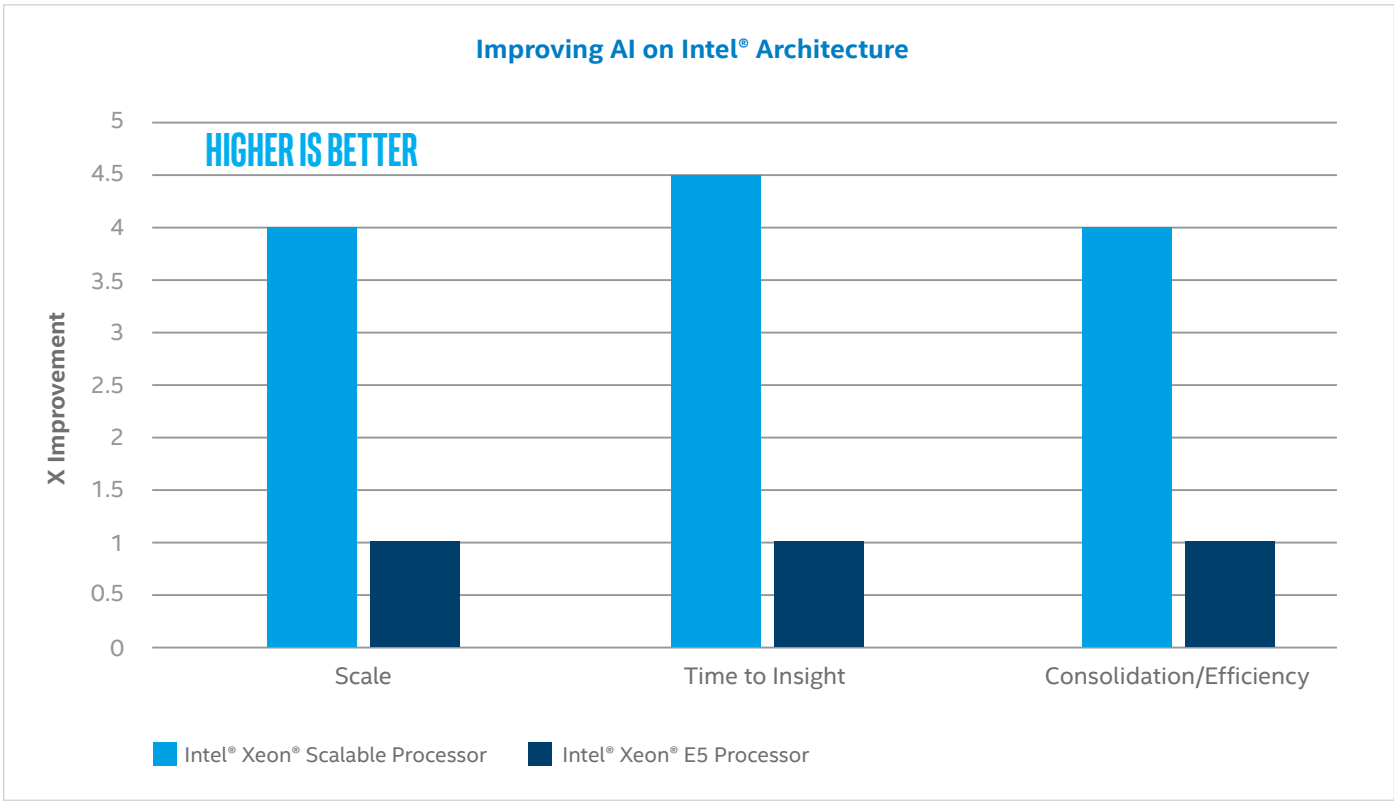
"H2O Driverless is and will be instrumental in saving Hortifrut time and costs associated with growing and shipping quality blueberries to all points of the world," Bustos added.

Intel® Xeon® Scalable Processors, Intel® Optane™ DC Persistent Memory Accelerate Driverless AI

H2O.ai and Intel are working together to democratize AI and accelerate machine learning algorithms and libraries on platforms built with 2nd Generation Intel Xeon Scalable processors and Intel® Optane™ DC persistent memory. Enterprise organizations beginning their AI journey can leverage a highly scalable, cost effective, and fast path to insights by combining H2O.ai's technology with Intel's most advanced processor and memory architecture in order to gain a competitive edge.

Recent testing by Intel and H2O.ai illustrate the scalability and time to insight achievable with H2O on 2nd Gen Intel Xeon Scalable processors and Intel Optane DC persistent memory.





- Scale up and out—Capable of handling 4X the data set size than traditional memory systems with improved TCO and superfast adaptability (no code changes).
- Faster time to insights—H2O with optimized XGBoost* delivers 4.5X improvement in training time.
- Server consolidation and efficiency—Run 100GB on a single machine (2nd Gen Intel® Xeon® Platinum 8200 processor plus Intel® Optane™ DC persistent memory) vs a 4-node cluster (Intel® Xeon® processor E5-2600).

About H2O.ai

H2O.ai is an open source leader in AI with a mission to democratize AI for everyone. H2O.ai is transforming the use of AI with its open source machine learning platform, H2O. More than 18,000 companies use open-source H2O in mission-critical use cases for Finance, Insurance, Healthcare, Retail, Telco, Sales, and Marketing. H2O Driverless AI uses AI to do AI in order to provide an easy, fast, and effective means of implementing data science. For more information and to learn more about H2O.ai, visit **www.h2o.ai**.

INTEL® AI BUILDERS MEMBER

H2O.ai is a member of the **Intel® AI Builders Program**, an ecosystem of industry-leading independent software vendors (ISVs), system integrators (SIs), original equipment manufacturers (OEMs), and enterprise end users, which have a shared mission to accelerate the adoption of artificial intelligence across Intel® platforms.

[1] Configuration: 2x 2nd Generation Intel® Xeon® Platinum 8280L processor @ 2.60GHz; 384 GB DRAM; Storage: Capacity tier: 6x 2TB Intel® SSD DC P4610; Cache tier: 2x 375 GB Intel® Optane SSD P4800x; 1.2T Intel® Optane™ DC Persistent memory; 1.45 TB Intel® Optane™ SSD. Dataset: 116695259. GBM training time: 1:18:01. 8-node Hadoop*: 274GB; 2x Intel® Xeon® processor E5-2650 v2 @ 2.60GHz (8 machine); 400 GB SDRAM (50 GB per node); HDFS. Dataset: 116695259. 00:35:03. 4-node Hadoop 274GB; 2x Intel® Xeon® processor E5-2650 v2 @ 2.60GHz (4 machine); 400 GB SDRAM (100 GB per node); HDFS. 116695259. 00:55:23.

2x 2nd Gen Intel® Xeon® Platinum 8280L processor @ 2.60GHz; 1.2T Intel Optane DC Persistent memory; 98 GB DRAM; 1.45 TB Intel Optane DC SSD. 41943039. 00:21:29. 4-node Hadoop 98 GB; 2x Intel® Xeon® processor E5-2650 v2 @ 2.60GHz (4 machine); 400 GB SDRAM (100 GB per node); HDFS. 41943039. 00:17:44: 8-node Hadoop 98 GB; 2x Intel® Xeon® processor E5-2650 v2 @ 2.60GHz (8 machine); 400 GB SDRAM (50 GB per node); HDFS. 41943039. 00:10:18

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations, and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of the product when combined with other products. For more complete information, visit http://www.intel.com/benchmarks.

Performance results are based on testing as of March 28th, 2019 and may not reflect all publicly available security updates. See configuration disclosure for details.

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Performance varies depending on system configuration. No computer system can be absolutely secure. Check with our system manufacturer or retailer or learn more at https://www.intel.com.

© Intel Corporation. Intel, the Intel logo, Intel Inside, the Intel Inside logo, Intel Optane, and Xeon are trademarks of Intel Corporation or its subsidiaries in the U.S. and/or other countries. *Other names and brands may be claimed as the property of others.

0819/BA/HBD/PDF ♻ Please Recycle 341007-001US