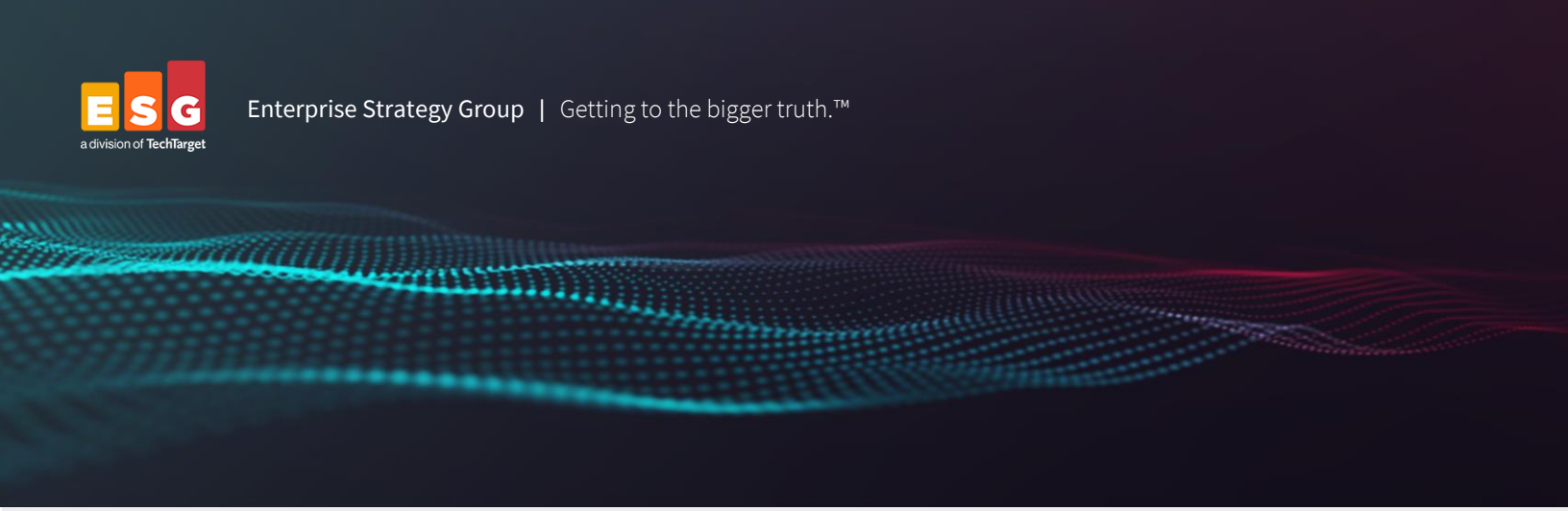

ESG WHITE PAPER

Accelerate the Use of Machine Learning with MySQL HeatWave

By Mike Leone, ESG Senior Analyst

March 2022

This ESG White Paper was commissioned by Oracle and is distributed under license from TechTarget, Inc.

Contents

Introduction 3
Machine Learning Adoption on the Rise 3
Focusing on Time to Value 3
What is Preventing More Pervasive Use of ML? 3
Skills Gaps 4
Infrastructure Stack 4
Accelerating ML Adoption with Oracle MySQL HeatWave ML 4
Native Support for Machine Learning in MySQL HeatWave 5
Automating the Machine Learning Lifecycle 5
Explainability as a Core Feature 6
Scaling HeatWave ML 6
Competitive Landscape 6
Additional HeatWave Enhancements 7
The Bigger Truth 7

Introduction

As organizations continue to embrace machine learning (ML), the focus has shifted to democratization. Organizations are looking to better empower stakeholders to ramp up and scale ML regardless of their skill level. It's not just the data scientists asking for help. It's IT. It's data engineers. It's developers. It's the line of business. How can organizations simplify the operational complexities that come with managing a dynamic and diverse data ecosystem? How can organizations simplify the process of fueling machine learning models with trusted data at scale? How can organizations do this as reliably and cost-effectively as possible without compromising agility and performance? The latest MySQL HeatWave announcements from Oracle are looking to be the answer.

Machine Learning Adoption on the Rise

As organizations turn to machine learning, they hope for a future rich in timely data insights and fast time to value. Whether they turn to ML to provide better predictive insights into the future of the business or look to develop products and services infused with machine learning to capture new opportunities in existing or emerging markets, businesses continue placing massive bets on the transformational technology. In fact, ESG research shows that 62% of organizations plan to increase their YoY spending on machine learning.[1] And they're looking to make those investments in people, processes, and technology with a goal of increasing the pervasiveness of machine learning and improving the time to value.

62% of organizations plan to increase their year-over-year spending on machine learning.

While business objectives like improving the customer experience, improving operational efficiency, and reducing risk around business decision and strategy are just a few of the many areas where ML can help improve businesses, organizations continue to scrutinize time to value as an important area to improve. With ESG research showing that 55% of organizations have yet to operationalize ML,[2] opportunities to reduce time to value continue to pave the way for key technology vendors to help simplify the adoption and increase the use of ML within organizations.

Focusing on Time to Value

Due to the diverse mix of business objectives that organizations are looking to solve with machine learning, the increasing level of customization that must be accounted for when exploring the use of machine learning to enable a smarter business is impacting when organizations see real value from their investments. In fact, it still takes a relatively long time to see real value from custom ML. While this may be a deterrent for organizations looking to add immediate value to the business, 82% of ESG research respondents said that it took 6 months or less for their organizations to start seeing value from the ML initiatives.[3] With advancements in technology solutions that can quickly ramp up an organization's use of ML to address a popular use case, ESG expects that time to value will shrink over time. And the result will be a reshaping of the business fueled by real-time intelligence, with ML infused in most, if not all, aspects of the business.

What is Preventing More Pervasive Use of ML?

[1] Source: ESG Research Report, *2022 Technology Spending Intentions*, November 2021.
[2] Source: ESG Survey Results, *Supporting AI/ML Initiatives with a Modern Infrastructure Stack*, May 2021.
[3] Ibid.

Between skills gaps throughout the ML lifecycle, weak links throughout the infrastructure stack, aggressive timelines, and tight budgets, organizations need help in not only ramping up the use of machine learning but also effectively scaling its use across the business.

Skills Gaps

While data scientist shortages continue to grab the headlines, whether that shortage is due to the lack of someone on staff in that role or the overburdening of an existing data scientist on staff with tasks outside of their core skillset, IT is increasingly being viewed as a problematic area. In fact, ESG research shows that 1 in 3 organizations have a problematic IT skills shortage in machine learning.[4] Despite these skills gaps, organizations cannot afford to delay adoption and are quickly overcome by higher CapEx and OpEx costs, infrastructure bottlenecks, an inability to scale, and significant delays in the ability to access the right resources for a particular use case. To address the skills gaps, organizations are looking for help to simplify adoption through automation within scalable environments that can enable fast ramp-up and effective operationalization of machine learning.

Infrastructure Stack

The right infrastructure to support machine learning development consists of several optimized and tightly integrated components across both software and hardware. The fact of the matter is that, today, 98% of organizations of recently surveyed machine learning adopters identified or anticipated a weak component somewhere in their supporting infrastructure stack. In fact, when ESG asked organizations about the parts of their existing infrastructure stacks that they believe to be their organization's weakest links in delivering an effective machine learning environment, the top response was resource sharing (26%), followed by an integrated development environment (25%), processing (both GPU and CPU; 25% for each), and storage (22%).[5] In other words, performance, scale, and reliability are critical. This highlights the increasing need for an approach that delivers on the requirements of all stakeholders, including business users, data science teams, developers, and IT.

98% of organizations identified or anticipated a weak component somewhere in their machine learning infrastructure stack, including resource sharing and an integrated development environment.

Accelerating ML Adoption with Oracle MySQL HeatWave ML

Oracle MySQL HeatWave is a fully managed database service that provides an in-memory, massively parallel, hybrid columnar query-processing engine that distributes query processing for ultra-high performance through a highly partitioned architecture that enables inter- and intra-node parallelism. An intelligent query scheduler overlaps computation with network communication tasks to achieve very high scalability across thousands of cores for real-world applications. HeatWave enables organizations to embrace real-time analytics by having modifications made by OLTP transactions propagated in real time to HeatWave and then immediately made available and visible for analytics queries.

HeatWave is the only service that enables database administrators and application developers to run OLTP and OLAP workloads directly from their MySQL databases. The database service is designed to enable customers to run analytics on data that is stored in MySQL databases, eliminating the need for complex, time-consuming, and expensive data movement, integration, and ETL processes across separate OLTP and OLAP databases. Since HeatWave is a native MySQL implementation, all existing MySQL applications run on HeatWave without changes. MySQL HeatWave utilizes MySQL

[4] Source: ESG Research Report, *2022 Technology Spending Intentions*, November 2021.
[5] Source: ESG Survey Results, *Supporting AI/ML Initiatives with a Modern Infrastructure Stack*, May 2021.

Autopilot, which uses advanced ML techniques to automate everything from provisioning and data loading to query execution and failure handling. By sampling data, collecting statistics on data and queries, and building machine learning models to model memory usage, network load, and execution time, organizations gain an increasingly intelligent query optimizer that makes it easier to use and further improves performance and scalability.

Understanding the criticality of security, Oracle ensures that all data at rest and in transit between MySQL database and the nodes of the HeatWave cluster is encrypted by default, reducing risk of compromise during any ETL processes. Further, by relying on a single database for OLTP and OLAP, organizations virtually eliminate the need for different identity management software.

Native Support for Machine Learning in MySQL HeatWave

For MySQL users to leverage machine learning today, they must utilize disparate tools, services, and processes. For example, to leverage data stored in a MySQL database for machine learning, customers must export that data via ETL outside of the core database to a different service or environment. In that environment, data science notebooks like Jupyter or Zeppelin that enable organizations to effectively traverse the ML lifecycle, including building and training models, inference, and explanation are then made available. This process takes time due to data extraction and movement; it adds complexity and takes additional effort, as customers need to understand yet another environment; it increases costs due to the need to run yet another data service; and it jeopardizes security because the data and model are outside of the core database. Regardless of how seamless this may appear today to organizations that may be leveraging technology from certain cloud technology providers, this ETL process and the need for yet another service is a reality for all customers.

To address these challenges, complexities, and risks, Oracle is introducing HeatWave ML with native, in-database support for machine learning. All model training, inference, explanation, and storage of trained models is done directly within the MySQL database. This eliminates any need for an extract, transform, load (ETL) process or movement of data to a different environment. All data in the model is secured and protected by the same access control policy as the underlying data itself. This reduces overall effort since it does not require data scientist expertise. Anyone can invoke a single command of a stored procedure or SQL command. Performance is lightning fast, and the process is fully automated. It also drastically reduces cost because, contrary to the competition, there is no additional charge for invoking ML, utilizing any of these capabilities, and, of course, none of the extra cost for ETL and using a separate machine learning tool or service.

Automating the Machine Learning Lifecycle

The machine learning lifecycle consists of several stages, including preprocessing, algorithm selection, feature engineering sampling, hyperparameter optimization, training, A/B testing, inference, and explainability. Of these phases, training is notoriously the most time-consuming and often dictates the quality of the model. The higher the model quality, the higher the inferencing quality and accuracy. Most times, training requires an expert like a data scientist to ensure that the right algorithm, features, and hyperparameters are selected and optimized for training to yield a high-quality model. HeatWave ML completely automates this entire process. This will help democratize access to machine learning and empower more stakeholders to quickly traverse the machine learning lifecycle. In some cases, Oracle has achieved speeds as much as 25x faster at 1% of the cost of other cloud services. By training (and retraining) faster, organizations can keep models up to date with the latest data and ultimately ensure that models are more accurate and of high quality and accuracy.

So how is this achieved? HeatWave ML leverages meta-learned proxy models that make accurate one-pass decisions at every pipeline stage, creating an iteration-free machine learning pipeline. This includes early algorithm selection to enable more accurate sampling with imbalance-aware adaptive sampling and feature selection. HeatWave ML is able to improve hyperparameter tuning with highly parallel gradient-based search space reduction and also automatically reduce the

search space in each stage of the ML pipeline. Included in HeatWave ML is native support for model and prediction explainability in the training and ML pipeline.

Explainability as a Core Feature

HeatWave ML provides fully integrated training with explanations. All models generated by HeatWave ML are completely explainable. To ensure usability and interpretability, model-agnostic explanation techniques are utilized to ensure organizations gain access to intuitive explanations that assist stakeholders in determining which factors matter most to a prediction. References to a training data set are not required for local explanations, even though high-quality explanations are delivered in a repeatable way by leveraging characteristics of the underlying data set to explain the model's behavior more accurately. As organizations ramp up model training, increasing the need for explanations, whether it be the number of models trained or the increasing complexity of an individual model, HeatWave delivers scalable performance, especially as the number of features in a model increases over time. This can all be done in real time due to the architecture's distribution of workers and cores.

Explanations provided by HeatWave ML can help ensure:

- Regulatory compliance, by implying "right to an explanation" for algorithms affecting users.
- Fairness, by validating that predictions are unbiased.
- Repeatability, by ensuring that small changes to input do not lead to large changes in the explanation.
- Causality, by verifying that only causal correlation between features and predictions are selected.
- Trust, by delivering interpretable explanations that provide users with confidence in machine-learning-based predictions.

Scaling HeatWave ML

Scaling and parallelizing machine learning is difficult because each stage of the pipeline has different characteristics and parameters, whether they are feature selection or hyperparameter tuning. HeatWave ML provides automated tuning and training of models and considers unique requirements for each stage of the pipeline to ensure effective parallelism across multiple nodes of a HeatWave cluster. Put it all together, and customers no longer need to make tradeoffs between runtime, accuracy, and scalability.

Competitive Landscape

As organizations look at the competitive landscape for help in democratizing machine learning, several features/capabilities must be considered. First and foremost, as organizations look to simplify the machine learning on-ramping, several cloud database solutions rely mostly, if not entirely, on third-party libraries, partnerships, and manual coding in Java, Scala, and Python. This instantly discounts their ability to effectively deliver a seamless machine learning experience, as operational complexities, cost, security, and the overarching end-user experience can be disruptive. For those vendors with more tightly integrated machine learning offerings, the devil is in the details. Does data need to be exported out of the database to a different tool or service? What type of data is supported (OLTP versus OLAP)? Is it real-time data? What expertise level is best served and/or underserved? To what degree is explainability available? How is data sampling conducted to feed model training?

Additionally, when it comes to performance in machine learning, metrics like accuracy and training time are essential and can vary widely depending on the data set and use case. ESG suggests pressing vendors for validated benchmark results to best understand how each solution can perform. This should be extended to evaluate cost via price/performance. To date, Oracle is the only vendor to publish fully-transparent, repeatable ML benchmarks on GitHub.

Additional HeatWave Enhancements

Prior to the recent MySQL HeatWave release on 3/29/22, resizing a HeatWave cluster, as with other cloud database services, is a manual task that comes with downtime. The recent HeatWave update enables users to gain access to real-time elasticity with automated resizing to any number of nodes and no downtime. Customers gain improved availability and flexibility with support for all operations *during* the resize process, including queries and loads, whether scaling up or down. After resizing, data is balanced across the remaining nodes in the cluster and HeatWave ensures minimal data movement during the cluster resize with data loads still supported at object store bandwidth speed. The time for a resize to take place is constant and predictable based on the provisioning time, load time, and data manipulation language (DML) propagation time.

Additionally, HeatWave now supports blocked bloom filters that, when parallelized with AVX instructions, will yield 3x more efficiency compared to standard bloom filters. This allows for pervasive use of bloom filters in HeatWave. In addition, the data is compressed in HeatWave memory with no impact to load performance. As a result, the amount of data that can be processed by HeatWave is doubled. This enables customers to lower costs by nearly 50 percent, while maintaining the same price performance ratio. In addition, customers can now pause and resume HeatWave clusters instantaneously, further reducing costs.

The Bigger Truth

Organizations continue to emphasize the importance of machine learning but have done little to democratize access to the right tools and processes. AutoML is a valuable tool for organizations looking to ramp up machine learning usage, but several challenges remain associated with integration, data movement, security, scale, customization, and cost. And while top technology vendors are looking to best empower more stakeholders with solutions that simplify machine learning adoption, many are either geared toward an expert, technical stakeholder or lack key capabilities that are essentially required to better appeal to generalists. So how can organizations empower more stakeholders to leverage machine learning on their terms without compromise? What solution is available today that can enable organizations to not have to make tradeoffs between performance, scale, or cost? Simply put, they couldn't. Until now.

The latest enhancements to Oracle MySQL HeatWave continue to enable customers to embrace data and now machine learning on their terms. The flagship announcement is HeatWave ML, which provides native support for machine learning directly in the database. No more ETLs. No more security gaps due to a mix of environments to satisfy the machine learning lifecycle. No more confusion about what a model is doing or why. No more surprise bills. Customers gain access to a self-tuning database service that is fast, scalable, and predictable from both a performance and cost standpoint. Generalists and experts alike gain access to a powerful architecture that delivers right-sized machine learning in a flexible, automated, and intelligent way. And to satisfy the growing need for explainability of ML models and outcomes, HeatWave ML delivers robust and comprehensive explanation capabilities focused on usability, interpretability, quality, performance, and repeatability at scale. Paired with additional enhancements like real-time elasticity for easy scale-up/scale-down and blocked bloom filter and compression to enable 2x the amount of data per node, it's no wonder that enterprises continue to look to HeatWave to set themselves up for transformational data success.

Simply put, there are no more excuses for ML projects failing due to a lack of data scientists, taking too long to execute, using old data, or costing too much. With HeatWave ML, machine learning is democratized, it's fast, uses up-to-date data, and costs less than other cloud database services. Choosing not to invest in MySQL HeatWave ML and continuing to use 2-3 ETL tools plus two databases comes down to deciding whether you want to embrace the future or fight the future. Inevitably, all cloud database services are headed toward higher degrees of convergence and automation. The question is whether you want to explain to management that you could be saving money and time but instead are stuck in the past. And as long as all requirements are met, managing one database is always better than two. Remember that.

Enterprise Strategy Group is an integrated technology analysis, research, and strategy firm that provides market intelligence, actionable insight, and go-to-market content services to the global IT community.

www.esg-global.com contact@esg-global.com 508.482.0188