



# MySQL HeatWave Delivers Round Three of Knockout Benchmarks and Customer Validation

*Third Release of MySQL HeatWave Brings the Pain to Other Cloud Providers with More Transparent Benchmarks for Machine Learning and Analytics*

**Q2 2022**

**DANIEL NEWMAN**  
Principal Analyst + Founding Partner

**SHELLY KRAMER**  
Principal Analyst + Founding Partner

**RON WESTFALL**  
Senior Analyst + Research Director

IN PARTNERSHIP WITH

**ORACLE**

**Published: April 2022**

## Introduction

### Oracle Debuts MySQL HeatWave ML: Unleashing a Host of Advantageous Database Capabilities for Customers

MySQL is the foremost popular SQL database used in the world and is widely used across the world's leading web applications such as Uber and Facebook. MySQL's open-source relational database (DB) management system design enables multiple users to manage and create DBs according to their rapidly evolving business and technical requirements. The latest announcement from Oracle represents the third major release of MySQL HeatWave in just 15 months.

We are basing our assessments here on the current information and data presently made available to us. In this case, we are utilizing data provided by Oracle. All ML benchmarks and all the TPC-DS benchmarks are available on Github, so they can be replicated by anyone wanting to do their own assessments. The TPC-DS benchmarks showcased here were performed by the independent third-party organization and all the ML benchmarks were performed by MySQL HeatWave Engineering. We would note that to date, no competitor has challenged round one, round two, or the current round three benchmarks delivered by the MySQL HeatWave Engineering team. In the absence of a response, silence says a lot.

Based on that current data and information, we see the new Oracle MySQL HeatWave ML offering as compelling, fulfilling a wide range of top-priority customer demands focused on machine learning. The MySQL Heatwave launch in summer 2021 delivered MySQL Autopilot, 64 nodes scalability, and improved performance advances that have resulted in customer migrations from AWS Aurora, AWS RDS, AWS Redshift, Teradata, SAP HANA, and Google BigQuery platforms. Throughout these successful migrations, Oracle prioritized MySQL customer feedback on further improving the overall MySQL experience. As a result, four major areas of prioritization were identified.

First, with more data than ever being stored in MySQL HeatWave, customers were required to use ETL to run machine learning. Second, although HeatWave is highly scalable, customers still needed downtime to manually resize workloads. Third, customers showed strong preference for an even lower entry point for cost per cluster size. Fourth, they sought new benchmarks for other kinds of query-intensive workloads. Now Oracle has responded in full with in-database machine learning; real-time elasticity; twofold more data per node; a pause/resume feature to lower costs; and TPC-DS benchmark validation.

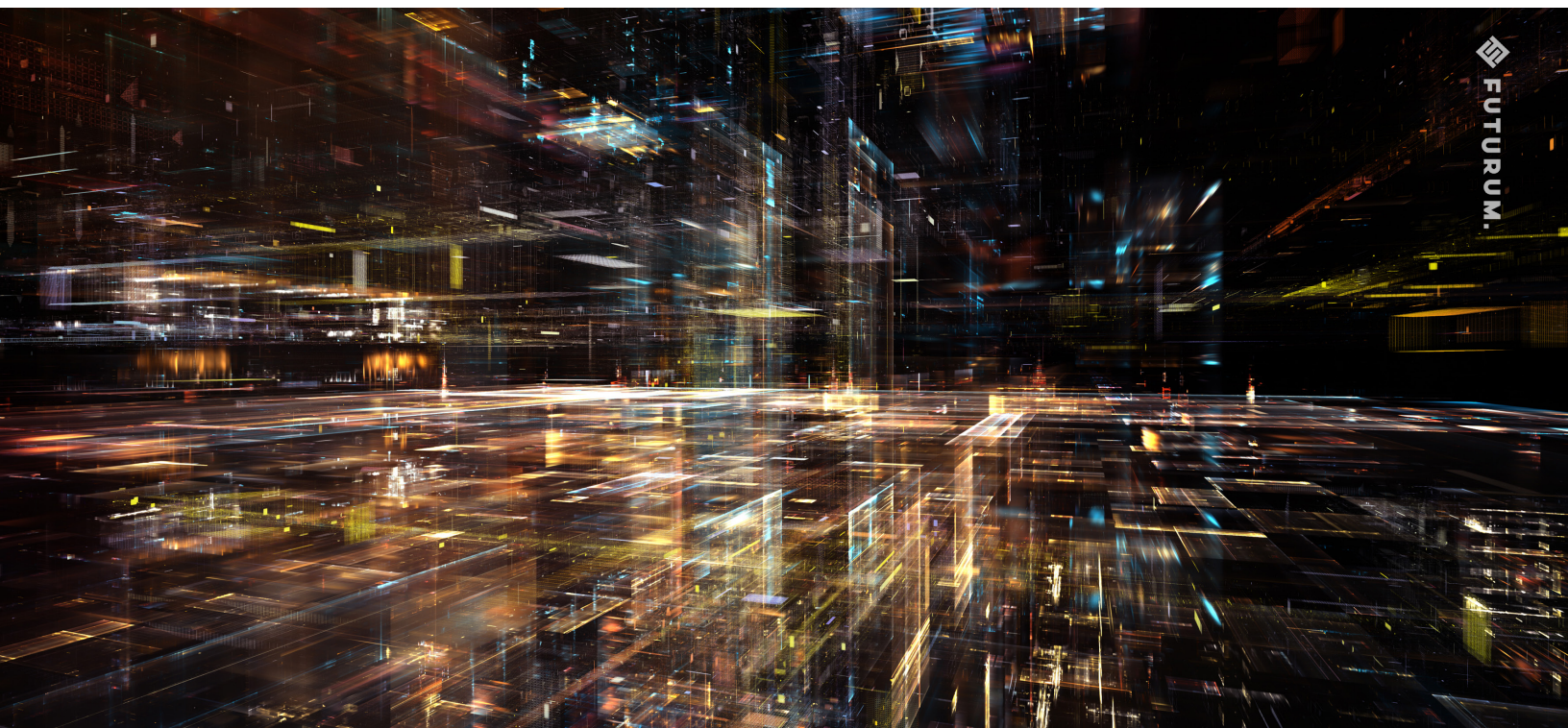
In examining the new MySQL HeatWave ML solution, we analyze the following elements:

- Native support for in database Machine Learning
- Full automation of training with HeatWave ML
- All models in HeatWave ML can be explained
- ML explanations in HeatWave are differentiated
- Improved performance, quality, and repeatability of explanations
- HeatWave ML scales with cluster size



## Executive Summary

- The third release of MySQL HeatWave delivers more transparent benchmarks for Machine Learning and Analytics.
- ML explanation is critical for enterprise workloads, particularly when explanations are integrated with the training pipeline along with model-agnostic techniques, which can explain any HeatWave ML model.
- We see Oracle demonstrating clear competitive differentiation against key rivals—Amazon Redshift ML and Snowflake ML—across eight key features consisting of AutoML, model and prediction explanations, user-facing API, user control of ML functionality, data/model locality, data available, supported number of nodes for training, and data sampling.
- The cost and performance differentiators make HeatWave ML an economically compelling offering for customers of all sizes and all industries.
- My SQL HeatWave’s support of real-time elasticity is a key feature that cloud DB users are increasingly demanding when aligning their key scalability requirements.
- Integral to the differentiation of HeatWave is the doubling of the data that can be processed per node.
- HeatWave ensures that users experience no impact on load time when compression is in use, including reading into HeatWave memory and writing to object store processes.
- Other MySQL HeatWave portfolio enhancements that warrant consideration include new query constructs such as support for views, and data, time, cast functions that accelerate HeatWave.



FUTURUM.

## MySQL HeatWave ML Overview: Meeting Customer Priorities

In the tech industry, some companies listen to customer requests, some ignore them, and some just create whatever they think will sell. We believe Oracle is clearly listening to its customer base and its partners as these public testimonials demonstrate:

“MySQL HeatWave improved our complex query performance by 300X for responses in seconds and at 85% of the cost compared to Google BigQuery with no code changes.” — Estuda.com

“VRGlass migrated all application data to MySQL HeatWave from AWS EC2. Within three hours, we achieved a 5X increase in database performance for a virtual event that accommodated more than 1 million visitors and 1.7 million sessions with greater security and at half the cost.” — VRGlass

“We found MySQL HeatWave improved performance by 90X which solved all our challenges and concerns we had in moving data to realize real-time analysis.” — Genius Sonority

“We found HeatWave ML very innovative, easy-to-use, very fast and most important, it is secure since the data or the model don't leave the database.” — Astute Business Solutions

“MySQL HeatWave on OCI increased our query performance by 300X with an 80% TCO reduction...now we can get real-time analytical reporting within our OLTP database to accelerate enhancing our security application.” — Neovera

An overview of Oracle's MySQL HeatWave value proposition is warranted to fully understand why the solution delivers some clear-cut price point differentiators. For starters, MySQL HeatWave is designed and optimized from the inception for the cloud, enabling the attainment of breakthrough capabilities such as scale-out processing and the ability to use commodity cloud services that offer the least expensive options in computing, storage, networking, and Virtual Machines (VMs). In addition, the solution makes pervasive use of machine learning (ML) and is optimized for Oracle Cloud Infrastructure (OCI) to assure such superior price/performance outcomes.

We see MySQL customers putting topmost selection emphasis on automation throughout the life cycle of their MySQL implementations, especially ML, capitalizing on performance advances required to manage massive data growth, and using HeatWave capabilities to accelerate more constructs. Moreover, they are requiring support for fully automated ML capabilities that dramatically improve overall efficiency and swiftly provide built-in explanations.

Through the cloud-first development strategy, the new MySQL HeatWave ML solution delivers the pervasive automation, unparalleled performance, online elastic scalability, lower costs, blended workloads, and security that MySQL customers are prioritizing.

## MySQL HeatWave ML Overview: Training with Full Automation

Oracle's strategic commitment to machine learning (ML) centers on its philosophy of bringing the algorithms to the data and running ML within the DB where data resides. Based on data provided and information available to us, we view Oracle's approach as minimizing or eliminating data movement, fully automating ML training, advancing scalability, preserving data security, and accelerating time-to-market. Contrast this approach with AWS Redshift or Snowflake, which require external tools and services that force both the data and the models to be moved out of the database and back, further increasing costs and making machine learning much more tedious than it needs to be.

The training process with HeatWave is fully automated and very efficient. Specifically, meta-learned proxy models can make accurate one-pass decisions at every pipeline stage, enabling iteration-free pipeline flows and early algorithm selection to enable accurate sampling and feature selection. We identify key capabilities as including imbalance-aware adaptive sampling, highly parallel gradient-based search space reduction for hyperparameter tuning, automatic convergence through reduction in search space at each stage of the pipeline, and native support for model and prediction explanation in the training pipeline.

## MySQL HeatWave ML: Explaining All Models Sets Stage for Sharp Differentiation

ML explanation is critical for enterprise workloads, particularly when explanations are integrated with the training pipeline along with model-agnostic techniques, which can explain any HeatWave ML model. The ML explanation of any and all models in HeatWave ML aids enterprises in fulfilling key organization-wide objectives such as:

- **Regulatory compliance:** May imply the right to an explanation for algorithms affecting users
- **Fairness:** By allowing validation that predictions are unbiased
- **Repeatability:** Ensures that small changes in input do not lead to large changes or disruptions in the explanation
- **Causality:** Allows verification that only causal correlations between features and predictions are selected
- **Trust:** Interpretable explanations encourage usage of ML-based predictions

We see these ML explanations in HeatWave ML as providing clear differentiation across usability and interpretability, quality, along with performance and scalability criteria. For usability and interpretability, the model-agnostic techniques enable explanation of all HeatWave models while not requiring a reference dataset when performing explanations. As such, intuitive explanations assist users in determining which factors contribute most to a prediction. Quality is improved by leveraging dataset characteristics, which provide repeatable and stable explanations and support novel improvements that more accurately explain the model's behavior. Improved repeatability means less variance between explanations and lower error nearby to the sample to explain. Performance and scalability are augmented, since explanations scale linearly, and not exponentially, with the number of features. This results in real-time explanations that are possible due to distribution of work among workers and cores.

## MySQL HeatWave ML: Competitive Advantages Against Rivals

In direct comparison to rival Cloud DB solutions, we took a look at the competitive differentiators based on data provided by and testing done by Oracle. We are impressed that Oracle demonstrates clear competitive differentiation against key rivals—Amazon Redshift ML and Snowflake ML—across eight key features. These features include AutoML, model and prediction explanations, user-facing API, user control of ML functionality, data/model locality, data available, supported number of nodes for training, and data sampling, as indicated in the chart below:

Qualitative comparison with Cloud DB ML solutions			
Feature	HeatWave ML	Redshift ML	Snowflake ML
User API	SQL	SQL	User writes code using Java, Scala, Python
ML inside database	Yes	No	No
Data & model secure	Yes	No	No
Explanation for all models	Yes	No	No
Fully automated	Yes	No	Not Applicable
Scales to more nodes	Yes	No	Not Applicable

Source: Oracle

For example, HeatWave ML natively supports fully automated training of ML models and is convergence-based. In contrast, Redshift ML requires users to provide input and details during the training process, for example, the amount of time it will take to train the model. Snowflake ML does not provide native support for machine learning at all, requiring users to rely on third-party libraries. In that respect, Snowflake is bringing customers back to yesterday's approach to ML, which leads to organizations racking up huge bills in the process.

For data sampling, HeatWave ML brings intelligent sampling, whereas Redshift ML uses random sampling and Snowflake again compels users to resort to third party libraries. From our perspective, demonstrating such differentiators can broaden HeatWave's addressable market, win more mindshare across the cloud DB ecosystem, and accelerate Oracle sales cycles.



## MySQL HeatWave ML: Let the Competition Begin

We see HeatWave ML achieving solid differentiation across a wide range of top cloud DB selection criteria that augment the competitive advantages across the eight key features.

In the area of tuning, we are impressed that HeatWave ML is self-tuning, relieving users of the uncertainties of pre-selecting parameters. Conversely, Redshift ML requires the user to specify the desired run time and maximum number of cells, including the headache of using larger values which typically trigger higher costs. Moreover, users face the added frustration of not getting a better model by simply spending more.

Based on testing by Oracle, it appears as though HeatWave ML is 25X faster and more accurate than Redshift ML at 1% of the cost, based on the default training time that the Amazon solution budgets as indicated in the chart below:

<b>HeatWave ML is <i>faster</i> and more accurate than Redshift ML</b>					
<b>Default training time of Redshift ML</b>					
Dataset	Accuracy		Training Time (minutes)		Speedup
	Redshift ML	HeatWave ML	Redshift ML	HeatWave ML (2 nodes)	
Airlines	0.5	0.6524	90.00	2.71	33.21
Bank	0.8378	0.7115	90.00	3.72	24.19
CNAE-9	FAILED	0.9167	FAILED	5.91	N/A
Connect-4	0.6752	0.6970	90.00	7.13	12.62
Fashion MNIST	FAILED	0.9073	FAILED	181.85	N/A
Nomao	0.9512	0.9602	90.00	3.30	27.27
Numerai	0.5	0.5184	90.00	0.34	264.71
Higgs	0.5	0.758	90.00	68.58	1.31
Census	0.7985	0.7946	90.00	1.22	73.77
Titanic	0.9571	0.7660	90.00	0.47	191.49
CC Fraud	0.9154	0.9256	90.00	29.06	3.10
KDD Cup	FAILED	0.5	FAILED	3.55	N/A
<b>GEOMEAN</b>	<b>0.712</b>	<b>0.754</b>	<b>90.00</b>	<b>3.561</b>	<b>25.271</b>

Source: Oracle

Based on the default training time budget by Redshift, it appears as though HeatWave ML is also exponentially more affordable than Redshift ML:

## HeatWave ML is *cheaper* than Redshift ML

Default training time of Redshift ML

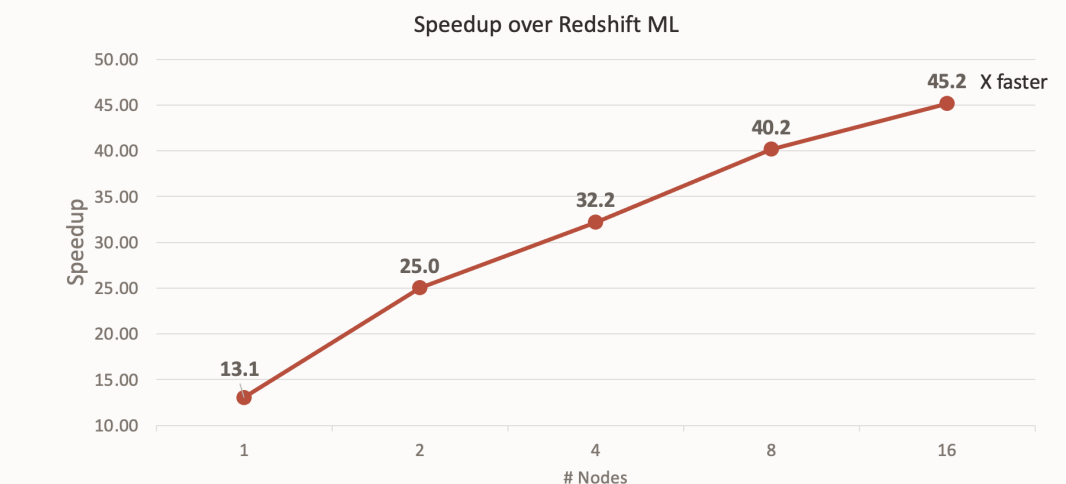
Dataset	Cost		Cheaper
	Redshift ML Cost with 1 year plan	HeatWave ML Cost \$ (2 nodes)	
Airlines	6.23	0.0479	130.03
Bank	5.68	0.0658	86.30
CNAE-9	FAILED	0.1045	N/A
Connect-4	6.18	0.1261	49.05
Fashion MNIST	X	3.2151	N/A
<u>Nomao</u>	5.96	0.0583	102.14
<u>Numerai</u>	5.49	0.0060	913.49
Higgs	7.27	1.2125	5.99
Census	6.12	0.0216	283.95
Titanic	5.60	0.0083	674.32
CC Fraud	6.70	0.5138	13.03
KDD Cup	FAILED	0.0628	N/A
<b>GEOMEAN</b>	6.115	0.063	<b>97.129</b>

Source: Oracle

The HeatWave ML differentiation cause is furthered by making it scalable. Each stage of the training pipeline has unique considerations for parallelism, which entails tradeoffs between runtime, accuracy, and scalability. For instance, iteration-free pipelines with fewer trials improve runtime but trade off parallelism, while reduced synchronization points in AutoML pipeline trades off accuracy for improved scalability. Likewise, convergence of the pipeline creates fewer trials at the end of the query, trading off scalability for faster runtime. In sum, we credit HeatWave ML's better performance as cluster size increases validate its value in handling AutoML scalability compared to Redshift ML's flatlining scalability as shown in the following chart provided by Oracle:

## Better performance with HeatWave ML with larger cluster size

Default settings for Redshift ML



Source: Oracle



These cost and performance differentiators make HeatWave ML an economically compelling offering for customers of all sizes and all industries. The pain of managing two databases, separate ML tools and a stack of ETL services is ubiquitous across all industries, and as such we see MySQL HeatWave as the antidote. Everyone is seeking to relief their excessive costs, management headaches and the suffering related to trying to analyze and cope with stale data. When it comes to cloud databases, HeatWave can be viewed as the Exedrin, Advil, Tylenol, Azithromycin, Benadryl and Pepto-Bismol all in one. Given the choice to pick the red pill or the blue pill, in this case we believe developers should pick red (Oracle).

## MySQL HeatWave: Real-Time Elasticity Further Fuels the Differentiation Engine

From our perspective, supporting real-time elasticity is a key feature that cloud DB users are increasingly demanding when aligning their key scalability requirements. We identify the most important real-time elasticity features as encompassing:

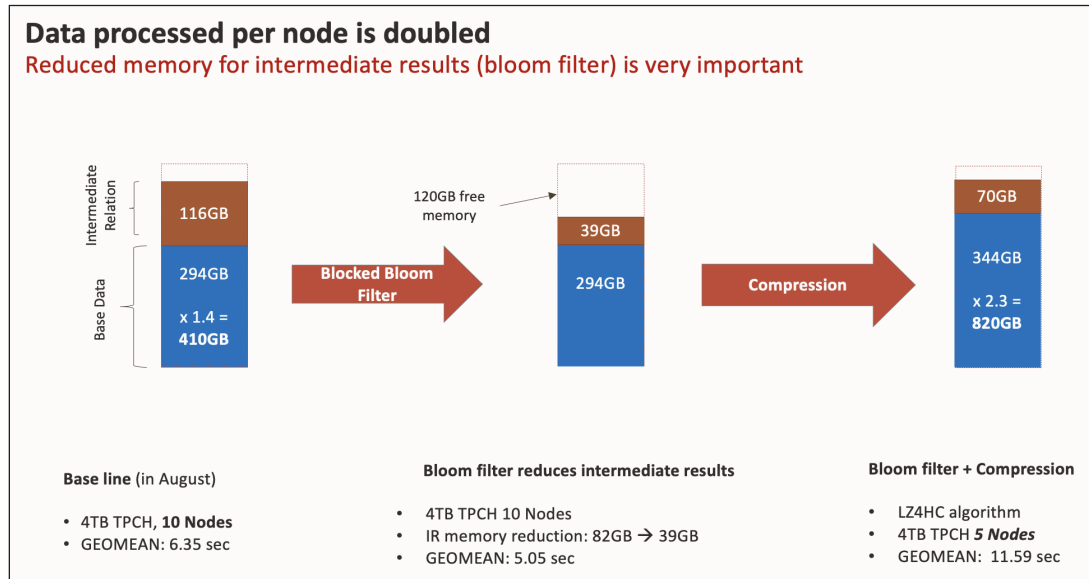
- **Availability:** All operations (queries, DMLs, load) are permitted on-cluster during resizing
- **Flexibility:** Upsize or downsize to any number of nodes
- **Balanced:** Data across nodes is balanced after resize
- **Performant:** Minimal data movement during resize. Data loaded at object store bandwidth
- **Predictable:** Resize time is both constant and predictable

By delivering these features across the board, we regard HeatWave as providing the fully available up/down scaling to any size cluster needed to stand out from the competition. Snowflake, for example, provides instant elasticity but with delayed loading of data on first query. Of equal concern, Snowflake only allows scale up or down in cluster sizes such as small, medium, large (i.e., 32 to 64 to 128 CPUs). A customer who needs 33 nodes to run a workload is obligated to move to the next size, 64 nodes, which does not fully utilize the large cluster and is paying for much more than needed. As we've noted before, this is like being forced to pay, drive and maintain a Greyhound bus when you're a family of two and could fit just fine in a compact Tesla Model 3 sedan. One is economical and right-sized and the other consumes enough resources to transport an entire neighborhood block.

Redshift resizing can take 10-15 minutes during which time writes are not allowed, which constrains user flexibility. In addition, during resizing, queries can be held back or timed out and if the user is not prepared with a snapshot, resizing can take even longer to implement. Adding to the potential for additional user frustration, any resulting data-skew must be identified manually and resolved with a classic resize, which can take hours or days. It's fair to say that these are difficulties any cloud DB service user would prefer be avoided altogether.

## MySQL HeatWave: Doubling Data per Node and pause/resume to lower costs

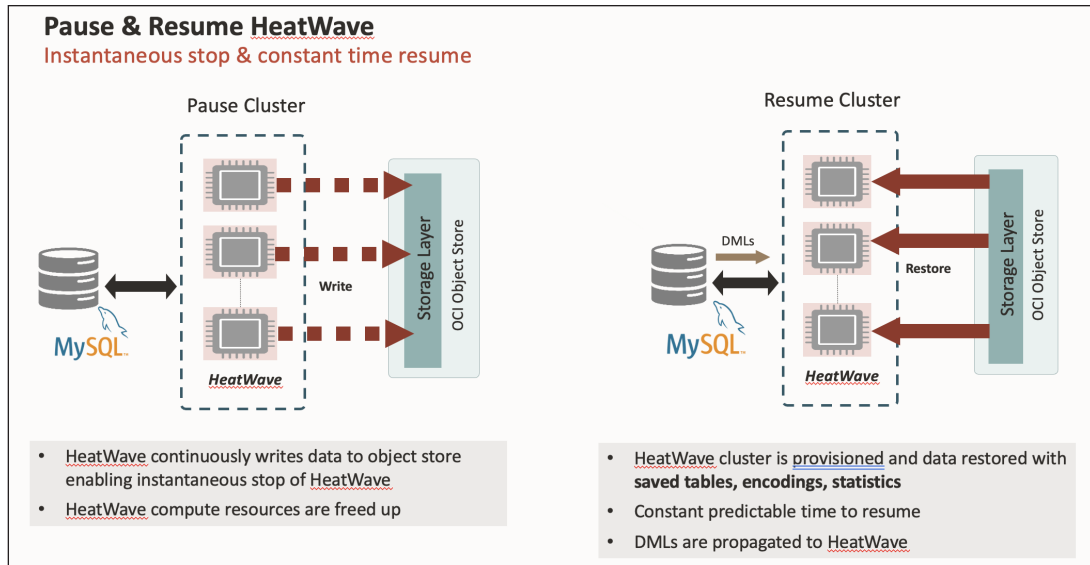
Integral to the differentiation of HeatWave is the doubling of the data that can be processed per node as shown in the chart below lowering costs by nearly 50 percent, while maintaining the same price performance ratio.



Source: Oracle

HeatWave ensures that users experience no impact on load time when compression is in use, including reading into HeatWave memory and writing to object store processes. This is achieved by offloading compression to the HeatWave cluster.

In addition, pause and resume functions now include instantaneous stop and constant time resume. To achieve this, HeatWave continuously writes data to object store, which enables instantaneous stop on HeatWave, also freeing up HeatWave compute resources. The HeatWave cluster also provisions and restores data through saved tables, encodings, and statistics that provide predictable time to resume capabilities. Upon resuming, both the data and the statistics needed for MySQL Autopilot are automatically reloaded into HeatWave. With pause and resume, customers can turn off HeatWave when not using it, such as when developers are off nights and weekends, reducing costs.



Source: Oracle

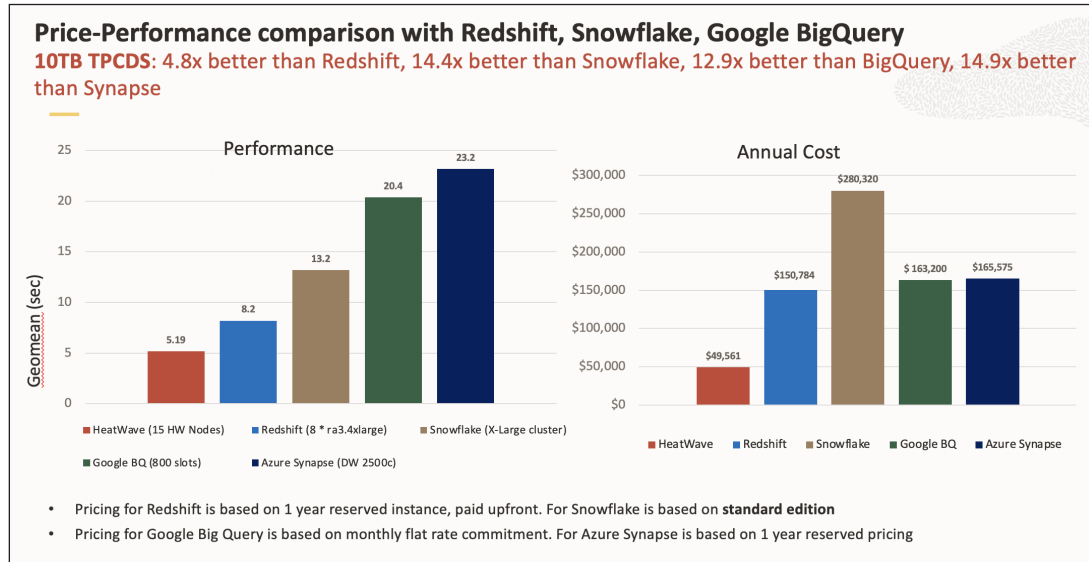
## MySQL HeatWave: Other Enhancements that Warrant Consideration

Other MySQL HeatWave portfolio enhancements that warrant consideration include new query constructs such as support for views, and data, time, cast functions that accelerate HeatWave, as well as improved monitoring such as HeatWave runtime statistics in performance schemes and runtime status to indicate the execution engine for running queries, which bolster the overall user experience.

The supported number of columns increased from 450 to 1K, and size from 8KB to 64KB. Plus, MySQL DB supports data masking, further boosting security credentials.

## MySQL HeatWave: Let the Price/Performance Competition Begin

As a capstone, we believe HeatWave competitive differentiation benefits substantially from demonstrating price/performance advantages in 10 TB TPC-DS tests over Redshift, Snowflake, Google BigQuery, and Azure Synapse. TPC-DS was chosen to complement previous TPC-H benchmarks to demonstrate HeatWave's performance with larger datasets and more complex queries. Specifically, HeatWave provides better price-performance of 4.8x in relation to Redshift, 14.4x better than Snowflake, 12.9x better than BigQuery, and 14.9x better than Synapse, as shown below.



Source: Oracle

Why are other cloud providers not competing in these demolition derby style benchmarks? Is there something they are afraid to show? Do they feel that the fact that they would need to benchmark two of their databases just to compete with MySQL HeatWave put them at a fundamental architectural disadvantage? Whatever their excuse may be, they are completely silent on the matter. Meanwhile, we see the MySQL HeatWave Engineering team as delivering consistent innovation, and dropping new release after new release, with customer references with each announcement further validating their claims. We believe that the rarefied combination of real-world customer references, fully transparent and repeatable benchmarks and a continuous onslaught of innovation is the proven formula that sets MySQL HeatWave apart from anything else in the cloud database market today. It will force an architectural shift across the cloud database landscape, as doing the same exact thing as you've always done only means that you quickly become outdated; passe; and eventually a relic. A dinosaur among a tech utopia. A diesel truck in a sea of EVs. An architectural mess in a cloud of intuitive design engineering. Which is why we see the very real possibility of the HeatWave proposition gaining broader consideration and making further inroads across the cloud DB ecosystem due to HeatWave's demonstrable price-performance competitive edge.



## Conclusion and Considerations

In sum, based on the data made available to Futurum Research at this time, we believe that the new MySQL HeatWave ML is well-positioned to meet the most demanding, extreme machine learning applications and expectations of enterprise users. From our perspective, the new innovations that Oracle has built into MySQL HeatWave ML enable the solution to stand out with a compelling feature set such as AutoML scalability, functionality, performance, and cost advantages, over offerings from other cloud database services. Moreover, new ML explanations in HeatWave ML deliver differentiation across usability and interpretability, and quality with scalable performance. Together with additional HeatWave innovations, including real-time elasticity, 2X data per node, pause and resume, plus price/performance differentiation proven in benchmarks will probably make the competition re-think strategies.

Of key consideration, the latest MySQL HeatWave TPC-DS benchmarks made available to us at the time of writing this report demonstrate that Amazon Redshift, Snowflake, Azure Synapse, and Google BigQuery are all both slower and more expensive. This gives MySQL customers further warrant to fast track the evaluation of HeatWave and validate for themselves that the dramatic price/performance differentiators are suited for their evolving and expanding cloud DB demands.

**MySQL Decision Makers Should Consider HeatWave ML.** MySQL decision makers should add the new Oracle MySQL HeatWave ML solution to their consideration list. It delivers compelling performance and cost advantages versus Amazon Redshift ML, key qualitative feature set, native AutoML, security, explainability, self-tuning, cluster-aligned scaling, and predictable cost, providing what are clearly major differentiators for Oracle.

**Machine Learning Assures MySQL HeatWave Automation Advantages.** MySQL decision makers should consider the ML capabilities of the MySQL Autopilot solution given ML's integral role in ensuring that HeatWave gains more intelligence over time.

**Swiftly Evaluate the MySQL HeatWave ML Proposition.** MySQL decision makers should prioritize the HeatWave ML solution as they evaluate cloud database offerings and explore the value proposition of taking advantage of automation innovations across AutoML, model & prediction explanations, user-facing API, user control of ML functionality, data/model locality, data availability, a higher number of nodes for training, and data sampling features.

## IMPORTANT INFORMATION ABOUT THIS REPORT

### CONTRIBUTORS

Daniel Newman  
Founding Partner + Principal Analyst  
*Futurum Research*

Shelly Kramer  
Founding Partner + Principal Analyst  
*Futurum Research*

Ron Westfall  
Senior Analyst + Research Director  
*Futurum Research*

### PUBLISHERS

Daniel Newman  
Founding Partner + Principal Analyst  
*Futurum Research*

Shelly Kramer  
Founding Partner + Principal Analyst  
*Futurum Research*

### INQUIRIES

Contact us if you would like to discuss this report and Futurum Research will respond promptly.

### CITATIONS

This paper can be cited by accredited press and analysts, but must be cited in-context, displaying author's name, author's title, and "Futurum Research." Non-press and non-analysts must receive prior written permission by Futurum Research for any citations.

### LICENSING

This document, including any supporting materials, is owned by Futurum Research. This publication may not be reproduced, distributed, or shared in any form without the prior written permission of Futurum Research.

### DISCLAIMER

This report is sponsored by Oracle. Technology is constantly being updated and any claims herein are based upon current data available to us as well as data available to evaluate current market offerings. Our recommendations are based upon assessing the specifications against specific workload demands and competitive data where available. In no way is our assessment a buying recommendation, we recommend potential buyers of this, or any database technology, do their own due diligence and make assessments thereon before making any purchase decisions.

### DISCLOSURES

Futurum Research provides research, analysis, advising, and consulting to many high-tech companies, including those mentioned in this paper. No employees at the firm hold any equity positions with any companies cited in this document.

### ABOUT ORACLE

Oracle offers integrated suites of applications plus secure, autonomous infrastructure in the Oracle Cloud. For more information about Oracle (NYSE: ORCL), please visit us at [oracle.com](https://oracle.com).

### ABOUT FUTURUM RESEARCH

Futurum is an independent research, analysis, and advisory firm, focused on digital innovation and market-disrupting technologies and trends. Every day our analysts, researchers, and advisors help business leaders from around the world anticipate tectonic shifts in their industries and leverage disruptive innovation to either gain or maintain a competitive advantage in their markets. [Read our disclaimer statement here.](#)

## CONTACT INFORMATION

Futurum Research, LLC | [futurumresearch.com](https://futurumresearch.com) | 817-480-3038 | [info@futurumresearch.com](mailto:info@futurumresearch.com)

Twitter: [@FuturumResearch](https://twitter.com/FuturumResearch)

©2022 Futurum Research. Company and product names are used for informational purposes only and may be trademarks of their respective owners.