



Cube Power Panel Astonished by Oracle MySQL HeatWave's Torrid Innovation Pace

by, Marc Staimer

2022

Oracle MySQL HeatWave Innovation

Oracle has been on an unprecedented innovation streak with their MySQL HeatWave database service. In a word the advancements have been breathtaking. This is what stunned the [Cube Power Panel](#) of distinguished analysts and experts led by **Dave Vellante**, Co-CEO Silicon Angle; **Matt Kimball**, Sr. Analyst Moor Insights & Strategy; **Holger Mueller**, VP and Principal Analyst Constellation Research; **Bob Evans**, Principal and Founder of Cloud Wars; **Ron Westfall**, Sr. Analyst and Research Director of Futurum; and this author **Marc Staimer**, Senior Analyst at Wikibon. To understand what has caused that astonishment requires a brief MySQL HeatWave overview and background over the past 15 months. It starts with the urgent MySQL database user issues.

MySQL open source databases are the most popular databases on the planet both on-premises and in the cloud with the greatest market share. Every major public cloud provider offers a MySQL cloud database service. But MySQL and all of the MySQL forks – MariaDB and PerconaDB – have well known meaningful limitations that cause users headaches. MySQL is a traditional relational transactional database. Nothing else is built-in. That means it has very limited reporting and analytics, essentially non-existent automation, and zero machine learning (ML) capabilities.

The database cloud service providers know this. The way they currently deal with this situation is to offer additional and separate cloud database services such as a data warehouse cloud service, a ML cloud service,

Cube Power Panel Astonished by Oracle HeatWave's Torrid Innovation Pace

and an ETL cloud service to integrate it all with their MySQL cloud database service. Each of these have their own monthly fees, usage metering, storage fees, interfaces, SLAs, as well as skills and training requirements. These costs add up rapidly.

Consider that few DBAs are knowledgeable and experienced in each technology of transactions, analytics, and ML. Determining the best ML model takes knowledge and experience. Teaching the model takes a lot of time. And understanding and explaining why a given model decides what it does is difficult at best and often impossible.

Making matters worse is the severe growing shortage of skilled DBAs and a greater scarcity of quality ML data scientists. Integrating and managing these technological wonders in a coherent manner is complicated, difficult, time-consuming, and costly. Not having enough qualified administrators causes frequent project failures, wasted time, and greatly increased time to useful results. That's just part of the problem.

Sharing data between these different databases is a thankless, laborious, cumbersome task. That's because the data has to be replicated, moved, massaged, and stored in distinct data stores, on a never-ending basis. Although commonly called an ETL or extract, transform, and load, ETL doesn't begin to express the complexity of the issue. It's not a one-time task. It has to be performed, tuned, tweaked, and managed all the time. All these processes add more time and more cost. Costs that rapidly reduce any and sometimes all potential net value. This is further exacerbated by the differences in database semantics.

Another major problem is elasticity. The public cloud is supposed to deliver real-time elasticity. For many services it does. But for analytics and ML, not so much. For most cloud database services, elasticity is disruptive and with coarse granularity. Customers can scale up clusters but only on fixed hardware sizes. If they need only a few more vCPUs that's not going to happen. They will be purchasing much more. Compounding this problem is the disruption or outage that occurs while the system reshuffles the data. No queries, DMLs, or load is permitted until the resizing is complete. Resizing time is variable, ranging from 10-to-30 minutes, depending on whether there is a prepared snapshot. Many cloud database services require customers to manually balance the data across the nodes of the cluster after the elasticity operation is completed.

These are all the problems Oracle sought to solve and has done so with their unprecedented and relentless series of MySQL HeatWave innovations. Innovations such as:

1. HeatWave tightly integrated built-in analytics engine with Oracle MySQL cloud service introduced in Dec. 2020.
 - No ETLs, ETL services, or connectors.
 - No data duplication.
 - No data massaging.
 - No data movement.
 - No separate data stores.
 - No additional skills, experience, or training.
 - No additional service fees.
 - No additional metering.
 - No changes to MySQL or MySQL compatible applications.
 - Analytics that range up to multiple orders of magnitude faster than competitive analytical database cloud services, based on standard replicable, fully transparent benchmarks. These benchmarks are conservative in that they do not penalize those other analytical database cloud services for the lag in results based on the ETLs, connectors they must have, or how much users have to pay additionally for the transactional database service and/or ML service. None of the data those analytical cloud database services use originates within that service. It originates from a transactional database on premise or in the cloud.
 - At a cost often less than half of one individual competitive cloud services, and much less for the combination of cloud services required. Customers report cost savings ranging from 67% to 83% over other cloud database services such as Amazon, Snowflake, Azure, and GCP.
2. HeatWave with Autopilot that added extensive automation introduced in August 2021.
 - Automation that eliminates scores of manual labor-intensive DBA tasks including provisioning, data-placement, query-optimization, and scale-out-data-management.

Cube Power Panel Astonished by Oracle HeatWave's Torrid Innovation Pace

- Greater scalability with near linear performance based on TPC-H and TCP-DS benchmarks.
 - A major increase in the already astounding performance and price performance compared to Amazon, Snowflake, Azure, and GCP.
 - Again, all of this at no additional cost.
 - These innovations plus the previous ones translate into a 13-17x cost/performance advantage over those same competitive cloud database services.
3. HeatWave with built-in machine learning capabilities, real automated elasticity, and enhanced scalability introduced in March 2022.
- Fully automated machine learning training.
 - No additional cost.
 - One pass iteration-free ML training pipeline that's exceedingly fast. Up to 45.2x faster than Amazon Redshift ML. That's because it is completely built into MySQL HeatWave. Just like analytics, it's not a bolt on. It does not handoff to other services. It is true real-time ML.
 - Not a set of manual tools requiring data scientist skills, guesstimation, and extensive trial and error. There's no need to figure out the best parameters or models. HeatWave ML does all of that based on imbalance-aware adaptive sampling. The adaptive sampling performs hyper-parameter self-tuning using highly parallel gradient-based search space reduction. Then automatically converges the search space in each stage of the ML pipeline. This makes HeatWave ML easy-to-use, fast, and accurate for both the novice and the expert.
 - Built-in native support for model and prediction explainers in the training pipeline. Model-agnostic techniques can explain any HeatWave ML model. Explanations that don't need a reference dataset when providing better-quality performance, quality, and repeatable explanations.
 - Compatible with market de facto data science products such as Jupyter or Zeppelin Notebook.
 - Results that are both accurate and fast. HeatWave ML's benchmarks using standard ML benchmarks showed better accuracy at a much higher performance running an average 25x faster than Amazon Redshift ML. It was also able to complete all of the vertical industry benchmarks. In contrast, Redshift ML was not able to complete 3 out of the 12 benchmarks- i.e., AWS failed to finish in the default time.
 - Lower costs at approximately 1% of Amazon Redshift ML in running these benchmarks. That is based solely on the Redshift ML service. It does not include the costs of the required other services to move data and databases or tools to run ML. None of the data in Redshift ML originates there. It originates in some other database such as Amazon Aurora, which is a separate fee. It requires an ETL service such as Amazon Glue to move the data into Redshift ML. And it needs duplicate S3 storage. In addition, Redshift ML makes calls out to Amazon SageMaker to actually perform the ML. These other service fees add up. The 99% savings is only comparing against Redshift ML. It's actually considerably greater than that when these other services are included.
 - Performance that scales near linearly as cluster size increases.
 - Real-time cluster size elasticity that admits queries to N+K node HeatWave clusters. It always admits queries on existing nodes without any disruption, downtime, or data shuffling between nodes. All queries, DMLs, and loads are permitted on the cluster during resizing. It's flexible in upsizing or downsizing to any number of nodes. Data across nodes is automatically balanced after the resize. The essential point is that there's minimal data movement during the resize with the data loaded at object storage bandwidth. And most importantly, the resize time is near constant and predictable: provisioning time + load time + DML propagation time.
 - Twice the amount of data processed per cluster node. This enables MySQL HeatWave implementations at half the cost but slower while maintaining price/performance ratios. Or the same size clusters as previously before the March 2022 release, that are markedly faster but at the same cost.

That's an amazing number of impactful and very useful innovations that solve real urgent problems in just 15 months. This is why the [Cube Power Panel](#) was astonished and had this to say:

Cube Power Panel Astonished by Oracle HeatWave's Torrid Innovation Pace

Bob Evans, Principal and Founder Cloud Wars:

"A single MySQL HeatWave database service is simpler than two from AWS plus all the associated ETL tools and data movers—that's a fact, not an opinion...I've been around the tech business for some time, and I don't recall seeing performance gaps as profound as the ones that Oracle has cited for its new MySQL HeatWave database. When technological breakthroughs of this magnitude happen—and please trust me when I say that breakthroughs of this scale do not happen often—the result is not just small-scale incremental moves by competitors trying to play catch-up in an outdated game offering outdated results. Instead, the competitors are forced to regroup and head back to the labs and attempt to compete against entirely new standards, new levels of performance, new levels of value, and ultimately new levels of customer expectations. Oracle jammed its MySQL HeatWave database with a slew of enhancements allowing it to blow away competitive services from AWS, Microsoft, Snowflake, and Google."

Ron Westfall, Sr. Analyst and Research Director Futurum:

"MySQL HeatWave represents the fiscally responsible approach to cloud databases while AWS Redshift and Snowflake represent the fiscally reckless approach. From my perspective, the question for developers is: "Are you looking to be more productive, or spend time haggling with ETL tools and shuffling data back and forth?"

"The in-database HeatWave ML puts Redshift ML and Snowflake on notice. Are these solutions more like yesterday's tech in terms of engineering, performance, and cost? Because they're slower and more expensive, chances are the answer is yes."

Holger Mueller, VP and Principal Analyst Constellation Research:

"Kudos to Oracle to being aggressive, differentiated and they seem to have impeccable benchmarks...this is the third MySQL HeatWave release in one year by the HeatWave team...and so everybody in the database open source market—and there's so many MySQL spinoffs, to a certain point they were put to shame by Nipun Agarwal's team. Oracle MySQL HeatWave ML puts the other MySQL cloud services to shame.

MySQL HeatWave is really a suite of capabilities. Suites have historically been jack-of-all trades and master of none versus best-of-breed. They're typically reduced functionality and performance at a lower cost. HeatWave is different with more functionality and more performance versus best-of-breed and still at a lower cost. When Oracle compared HeatWave performance benchmarks to its competitors, they had to use best-of-breed benchmarks because of the current nature of their competitors. And yet they still came away considerably faster."

Marc Staimer, Sr. Analyst Wikibon:

"Oracle is no longer just a database vendor. They have become primarily a cloud vendor with database, software, platform, and infrastructure services. The key to successfully penetrating a new market where you do not dominate is to offer compelling products or services. To be compelling, it generally means the offering has to minimally be twice as fast, with twice as much functionality, at half the cost. Oracle is doing this and more with their Autonomous Database, applications, and now their MySQL HeatWave ML. MySQL HeatWave ML is even more compelling than the definition. For a lot less cost than half it provides considerably more than twice the performance and functionality.

Keep in mind that the worst job in the data center is data movement. There are three things that happen when data is moved, and two of them are bad. MySQL HeatWave ML does away with all that data movement. If the need is to analyze the transactional data in MySQL, it's simple to just do it. The analytic engine and data's there with it. If the need is to utilize ML of the data, again it's simple to just do it because the models and the data are there. Any other cloud provider requires multiple services, ETLs, duplicate storage, and of course data movement. This makes them much slower, a lot less functional, and a lot more expensive vs. MySQL HeatWave ML.

There are few 'no-brainer' decisions in IT. For MySQL users, MySQL HeatWave ML is such a decision."

Cube Power Panel Astonished by Oracle HeatWave's Torrid Innovation Pace

Cube Power Panel Summary and Conclusion

Oracle released its third iteration of the MySQL HeatWave in March of 2022. Yet again, another significant major innovation release in just 15 months. Each release has dramatically and measurably pushed the cloud database service envelope by solving entrenched database problems which up until now, went unresolved.

SiliconANGLE put together a Cube Power Panel to discuss the development effort that Oracle has made with MySQL HeatWave to solve these serious database problems. The entire panel was astonished by the depth and relentless pace of these MySQL HeatWave innovations and how well each iteration of HeatWave release builds upon the previous one.

In the end, Oracle is providing a MySQL cloud database service with a scale out query accelerator which is faster than specialized analytic database services, easy-to-use, extensive automation, and very efficient machine learning, real elastic scalability, at a very low cost. This is for now and will continue be a very tough combination to beat. Especially when considering It took a very skilled, knowledgeable, and experienced MySQL Engineering team 10 years to develop HeatWave.

Competitors will find just catching up to the current HeatWave to be exceptionally challenging. It will be exceedingly problematic coming up with a viable answer in months or even years. It takes a lot of complex database engineering and time to meet, let alone keep up, with the pace of HeatWave innovations. As Oracle continues to raise the HeatWave bar at an unprecedented pace, that 10-year head start may well prove impossible to bridge.