

White Paper

MySQL HeatWave Goes on Autopilot Delivering Automated Real-Time Query Analytics Faster, Cheaper

Sponsored by: Oracle Corp.

Carl W. Olofson
November 2021

IN THIS WHITE PAPER

The database cloud service market is crowded with offerings that claim scalable, high-performance query capability. Most of these offerings have limitations that ultimately frustrate their users. A key limitation is their inability to analyze online transactions, forcing users to migrate data from OLTP databases to databases that are specialized for analytics, thereby increasing cost and complexity and opening up potential security threats and surface area exposures during the extract, transform, and load (ETL) process. This white paper examines the phenomenon of cloud database services offering powerful analytics at an affordable price and the advantage to be gained from such services when they also support OLTP and enable query processing against live transactional data. It considers the open source dimension also, in this regard. It then presents the Oracle MySQL HeatWave service as a premier analytic database capability, based on open source technology, which complements Oracle MySQL Database Service's OLTP functionality with a scalable, high-performance analytic query accelerator that most users can afford – at a fraction of the cost of other cloud database services. This white paper also discusses MySQL Autopilot, a new addition to the MySQL Database Service with HeatWave, providing machine learning (ML)-based automation capabilities that eliminate many traditionally manual tasks.

SITUATION OVERVIEW

Not long ago, enterprises were forced to invest in expensive database systems for reliable, high-performance transaction processing and in other expensive database systems for data analytics. They would shuttle data from the transactional systems to the analytic systems by means of extract, transform, and load operations. It was slow, it was error prone, and it was a manual, resource-intensive process that led to analysis of stale data – not optimal for organizations striving to be data driven. In fact, many enterprises still find themselves in this position.

More recently, several technologies have arisen that are much more manageable from a cost perspective because they are cloud native and, in many cases, open source. A range of open source OLTP database management systems (DBMSs) are available to handle transactional needs, and a number of powerful cloud-native analytic database services are also there for the taking. But both are limited. The open source transactional databases, mostly relational, are not so good at complex query processing. The analytic databases don't handle transaction processing at scale. And there's still the need for ETL.

The cloud offers the promise of much more than lower operational costs. It offers more efficient management. It offers better security. It also represents an approach to IT that is characterized by a high degree of deployment flexibility and ease of management. These benefits were all called out by database cloud adopters in a recent IDC survey (see Figure 1).

But how flexible and manageable is a system that still depends on ETL-driven data movement, and the coordination of schemas between OLTP and analytic relational databases? The cloud is enabling a rapid rise in data volumes from a variety of sources, and greater data volumes are exposing the fragility of a two-database system joined by a manually managed link. Users want their open source OLTP. They want their high-performance analytics. Can't the two be combined somehow?

FIGURE 1

Benefits of Cloud Migration

Q. *What benefits did you achieve after migrating your on-premises databases to the cloud?*



n = 406

Source: IDC's *U.S. Cloud Database Migration and Architecture Survey*, December 2020

The Emergence of Analytic Transaction Processing

Over the past 10 years, we have seen the emergence of a variety of technologies that combine transactional and analytic functions on one database. In the early days, although both types of data were in one database, they were processed in different formats (rows for OLTP, columns for analytics), and analytics could not be performed on live transactional data, but only on data that was a little old, which made combinational processing – in which immediately after a transaction any analytic query would be driven by the new database situation – impossible. Eventually, most of these systems were enhanced to allow seamless combinations of transactions and analytics that are driven by each other. IDC calls this type of processing not OLTP or analytic but analytic transaction processing (ATP, not to be confused with Oracle's Autonomous Transaction Processing). The problem has been that this kind of functionality was only available in high-end product platforms.

One of the most popular open source relational DBMSs (RDBMSs) available is MySQL, which is owned by Oracle Corp. It is a system designed for OLTP that is mature and well established. Oracle also offers a cloud service called MySQL Database Service centered on the latest version of MySQL, 8.x. On this platform, Oracle has made available to users an ATP capability formerly only available on expensive proprietary systems. Oracle calls this capability MySQL HeatWave.

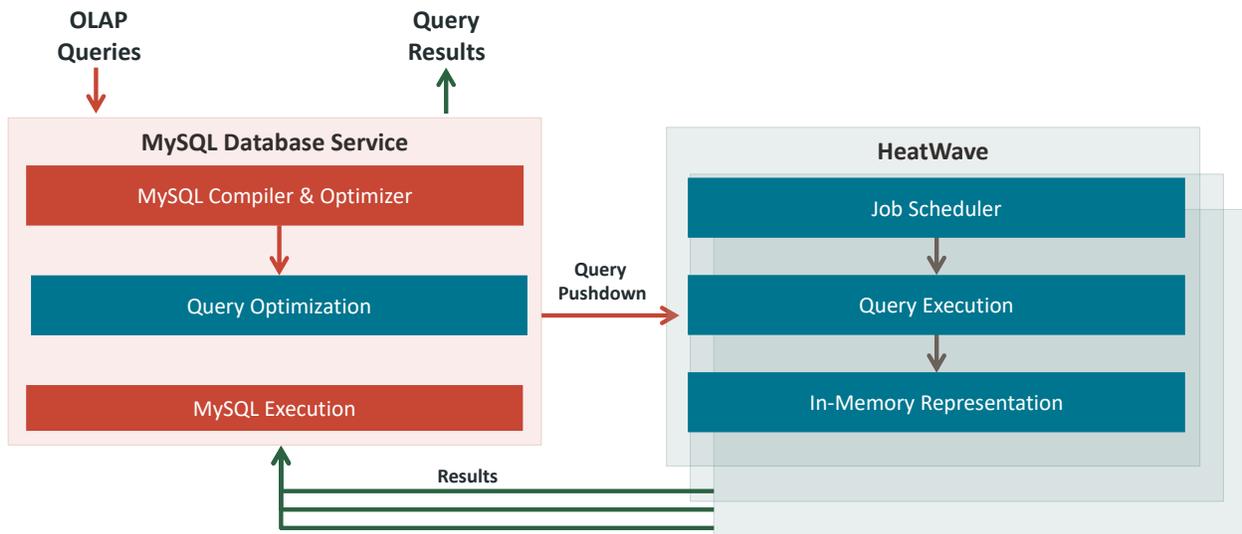
MySQL HeatWave

MySQL Database Service with HeatWave processes transactions as MySQL always has, so existing MySQL applications work with no modifications required. MySQL HeatWave has been designed for a cloud architecture and optimized for commodity hardware and has implemented novel algorithms for distributed query processing to achieve extremely compelling price performance. The data is represented in a compressed hybrid in-memory format, and cache-conscious SIMD (single instruction, multiple data) processing is leveraged. This functionality delivers complex query performance that rivals top data warehouse solutions available anywhere in the cloud and at a fraction of the cost of competing systems while offering compelling performance and price performance. The queries can be performed on data that includes the latest transactional updates for true ATP functionality. Figure 2 illustrates how HeatWave OLAP functions are organized and interoperate.

As a full cloud database service, MySQL HeatWave offers ease of administration and full professional support at both the infrastructure and database server levels. On top of that, the open source nature of the core product, MySQL, and its widespread popularity make MySQL HeatWave a compelling offering, especially for organizations that have difficulty in staffing up highly technical database professionals. Since MySQL is perhaps the best known and understood SQL interface out there, finding database team members won't be a problem. Of course, eliminating the need for ETL to get transactional data into a form usable for complex, high-performance queries not only simplifies operations and eliminates a key source of human error but also eliminates the latency caused by ETL, which means decisions can now be made in real time.

FIGURE 2

HeatWave OLAP Functions and Operations



Source: Oracle, 2021

Newest Enhancements to MySQL HeatWave

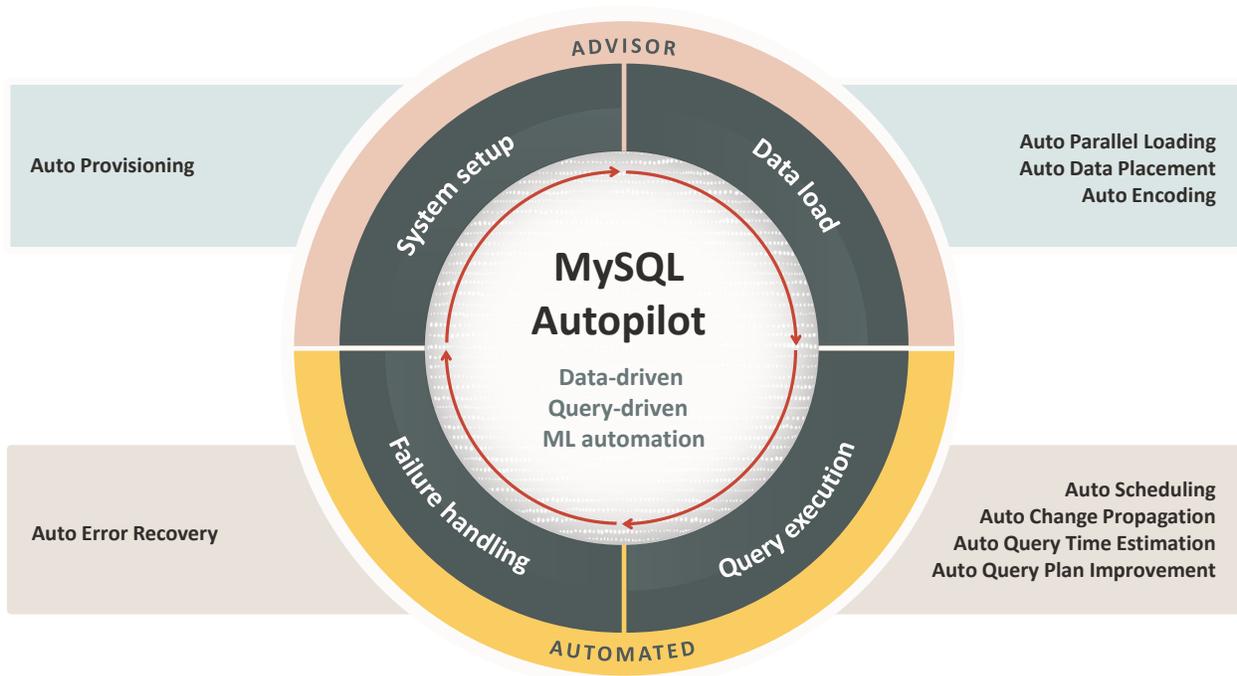
MySQL Autopilot

Oracle has introduced machine learning-driven automated capabilities to HeatWave, called MySQL Autopilot. With it, the database instance becomes more intelligent over time as more queries are executed, governed by instance-specific ML models. When conditions change within the database, the system adapts automatically, so manually setting and changing rules or settings are not necessary. Autopilot also includes a predictive element so that users can anticipate performance improvements over time as the system adapts to the data and usage patterns. Figure 3 provides a summary of AutoPilot features. MySQL Autopilot is offered to HeatWave customers at no additional charge.

MySQL Autopilot offers auto provisioning at system start-up, auto parallel loading and auto encoding when data loading is required, auto error recovery in cases of failure, and auto scheduling, auto change propagation, auto query time estimation, and auto query plan improvement for query optimization. The auto provisioning feature uses a history repository that shows database performance by size, structure, and other factors to provision just the right size cluster for the database. An auto data placement feature results in an automated partitioning of the database and optimal placement of data in the resulting partitions. Auto scheduling optimizes overall wait time for mixed workloads by prioritizing short queries over long queries, ensuring that queries from customer-facing online applications complete significantly faster while reports get minimally delayed.

FIGURE 3

MySQL Autopilot



Source: Oracle, 2021

Other Improvements

MySQL HeatWave provides the ability to load data into the HeatWave cluster in a scale-out manner. As a result, the time it takes to load a cluster during a reload operation like recovery from a node failure is independent of the data size. This is a 100x improvement from the current performance for a 10TB dataset. In addition, HeatWave now scales to 64 nodes and can process up to 32TB of data at a given time, and scalability across nodes improves by 20% to 0.90 (1.0 is the best, but, in practice, unreachable by any system to date). MySQL HeatWave also encrypts data at every stage: in storage, in network transit, and in memory.

MySQL Database Service with back-to-back innovations in the form of HeatWave and now Autopilot brings a big boost to the MySQL user community, and wherever there's increased competition, customers will benefit. MySQL Database Service may drive the open source database industry to accelerate engineering efforts. Combined with HeatWave and Autopilot, this may very well be the single greatest innovation in open source cloud databases in the past 20 years.

The MySQL Engineering Team makes quarterly code contributions to the open source community for MySQL Database Service. This means that the same software can be run on premises, in a Docker container, on a Mac, or a virtual machine.

According to Oracle, new customers that had moved from Amazon Aurora to HeatWave, including Tetris.co, Red3i, and FANCOM, provided testimonies at the MySQL Autopilot announcement citing vastly improved performance (as much as 85%), 60% lower costs, and no need for application changes.

Note: The reader is invited to visit the Oracle website to see both prices and price/performance comparisons with other leading cloud database services.

FUTURE OUTLOOK

There is a race underway in the world of cloud database services, as vendors seek to outdo each other in price, price/performance, scalability, ease of use, automated database tuning and administration, and other factors. Oracle MySQL Database Service with HeatWave with the new MySQL Autopilot addition stands out in all these areas.

The world of database technology is constantly evolving, and as this service moves forward, we should expect to see the latest innovations reflected in the offering. One benefit that any enterprise using MySQL can count on, however, is that the skill set for MySQL is likely to remain stable, and so there should be a ready supply of talent to build MySQL applications into the indefinite future. With MySQL Autopilot, this talent will be spending less time on manual, time-consuming tasks and more time developing hot new applications that propel business growth.

CHALLENGES/OPPORTUNITIES

This is a very competitive market, and so as Oracle's MySQL team continues to innovate, the company needs to stay well ahead of other technologies in the market, some of which are at least partially based on the MySQL interface as well. The requirements going forward will include ever easier administration, better performance and scalability, and support for rapidly evolving applications. Others will be working on these issues, and Oracle must continue to strive for leadership in technological innovation.

CONCLUSION

Many enterprises are committed to an open source approach to database management. Many of these also recognize that the relational data model offers the broadest range of data management capabilities, including the assurance of data consistency and comprehensive query support. Yet open source relational database management systems have lacked some of the functionality that enables enterprises to take their operations to the next level in terms of comprehensive data management, near-real-time query support, and automatic scaling.

MySQL HeatWave and AutoPilot represent a quantum leap for those enterprises, offering an RDBMS optimized with ML, topflight query performance, and superb transaction support – all without any changes to existing MySQL applications and at an affordable price. This development pushes the outside edge of the envelope not only for open source RDBMS but also for RDBMS in general and is a clear testament to Oracle's ongoing commitment to MySQL and the MySQL user community.

About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

Global Headquarters

140 Kendrick Street
Building B
Needham, MA 02494
USA
508.872.8200
Twitter: @IDC
blogs.idc.com
www.idc.com

Copyright Notice

External Publication of IDC Information and Data – Any IDC information that is to be used in advertising, press releases, or promotional materials requires prior written approval from the appropriate IDC Vice President or Country Manager. A draft of the proposed document should accompany any such request. IDC reserves the right to deny approval of external usage for any reason.

Copyright 2021 IDC. Reproduction without written permission is completely forbidden.

