

Publication date:

14 Nov 2022

Author:

Bradley Shimmin, Chief Analyst, AI & Data Analytics

Oracle CloudWorld 2022: Oracle redefines home field advantage with MySQL HeatWave

Omdia view

Summary

Oracle brought its long-standing but COVID-19-interrupted premier user conference back live in Las Vegas, Nevada, in October 2022. With more than 1,000 learning workshops (100 of which were hands-on) and free technology certification for attendees on tap at Oracle CloudWorld (OCW) 2022, the vendor reentered the live trade show market with a reminder that it has not lost its technical prowess. This was particularly evident in many of the announcements made by the vendor that espoused a unique approach to the challenge of integrating artificial intelligence (AI), data, and analytics both across business software and between premises, cloud, and multicloud.

Oracle shifts into overdrive at CloudWorld 2022

Seemingly in response to the constrained worldview that has permeated the global technology market since the onset of the COVID-19 pandemic in February 2020, in bringing its preeminent user conference back to the live stage in Las Vegas, Oracle pulled out all of the proverbial stops in both the breadth of its product announcements and the scope of its ambition. With regards to AI, data, and analytics-related announcements, Oracle rolled out several noteworthy updates across several key products.

Oracle Analytics Cloud (OAC): Highlighting its desire to embed analytics into every line-of-business process, Oracle showcased several new capabilities including a semantic modeler, which democratizes data access across the business by masking all underlying data integration/repository complexities. Another key addition to OAC, composite visualizations, enables users to add metrics to existing charts in order to rapidly surface insights. Also, a new, proactive automated insights feature delivers recommended visualizations based on current user context. And continuing its earlier work to bring AI directly into OAC, Oracle now lets users analyze image-based data using Oracle Cloud Infrastructure (OCI) Vision.

Oracle Database 23c (beta): Combining 300+ new features, this long-term support (LTS) release amps up the converged, multi-model capabilities of its flagship database. For example, the database now provides a richer MongoDB API support that will let developers move their apps to Oracle with zero code changes. More profoundly, this release includes a new capability promoted as JSON relational duality, which allows users to use the Oracle relational database as though it were a native JSON document store. And relatedly, 23c now incorporates property graph capabilities built on top of the emerging SQL/PGQ (graph query language) open standard.

Oracle Autonomous Data Warehouse (ADW): Looking to improve ADW's stance as an open platform, Oracle introduced the capacity to exchange data using Databricks' popular open-source Delta Sharing protocol. Relatedly, seeking to work more directly with a broader spectrum of analytics tools, Oracle added the ability for users to work directly with Microsoft Excel on top of ADW using a standard plug-in architecture.

Taken individually, these updates and innovations serve to sharpen Oracle's alignment with a market trending away from monolithic software and toward a unified but also flexible, democratized, open, and cloud-savvy tool set for the AI, data, and analytics toolchain. Naturally, such announcements are quite common at major events such as Oracle CloudWorld, particularly in supporting the market's inexorable march to the cloud. On the surface, Oracle is no different than its rivals, introducing functionality that positions Oracle Cloud Platform as a top-tier hyperscale destination for enterprise data and apps, right alongside Google Cloud Platform, Microsoft Azure, and Amazon Web Services (AWS).

However, one announcement not mentioned above warrants a closer view and paints Oracle in a very different light opposite its competitors. That announcement, Oracle MySQL HeatWave Lakehouse (beta), is important for two reasons. First, this release enables the company's implementation of MySQL to directly process data stored in object stores across multiple file formats at scale. Second, this capability adds fuel to the company's multicloud architecture for MySQL, which uniquely leverages Oracle OCI while running on competing cloud platforms, including AWS and Azure. Together, these make for a unique approach to the task of modernizing the AI, data, and analytics toolchain—not just on Oracle Cloud but across competing cloud platforms.

The problem with multicloud data and analytics workloads

During his customary afternoon keynote address to the Oracle CloudWorld audience in Las Vegas, Oracle co-founder and chief technology officer (CTO) Larry Ellison spoke about his desire to automate the entire value chain for specific outcomes, a goal evident in the company's continued push to vertically integrate its line-of-business apps, moving from storage up through the database and encompassing AI and analytics workloads—all seamlessly interwoven into the very fabric of daily business processes.

During his discourse, Mr. Ellison spoke at length about how this might play out across the healthcare marketplace with Oracle working closely with partners such as University of Oxford, Ronin (a consultancy), and MD Anderson Cancer Center to create a unified patient database leveraging its purchase of Cerner for the US market as a starting point.

Along the way, he emphasized the importance of a truly open cloud platform, not one that enabled traditional data/app integration but one that made such integration a transparent, native capability. According to Mr. Ellison, customers must be free to choose the service that best matches their requirements without compromise, even if those services live on a different cloud platform. In his own words, "The garden walls [must] come tumbling down!"

What does this mean in the world of data and analytics? As espoused by current efforts, this would mean providing a seamless cross-cloud software experience to enterprise practitioners by allowing them to provision and run two copies of the same software on different cloud platforms. Depending on maturity, those two copies might share the same metadata or enable a more direct connection for the transference of data as with solutions like MongoDB Atlas and Google BigQuery Omni (and Google Anthos). Other solutions such as Azure Arc-enabled SQL Server allow users to extend their Azure SQL Server instances down to the data center, edge, or across external clouds.

At the end of the day, while these solutions do provide a unified control plane across disparate cloud platforms, they remain deeply anchored to their home platform and depend on the target platform for top-level provisioning as well as underlying storage and compute services. It also means that whether migrating or running concurrently across two cloud platforms, customers must pay for data egress charges—an issue that is rapidly turning into a hard “no” for many companies looking to leverage multicloud deployments. However, it should be noted that Oracle and Microsoft do not charge egress fees in the Oracle Database Service for Microsoft Azure.

Oracle’s approach to multicloud deployments

In September 2022, Oracle announced that MySQL HeatWave was available on AWS, a move that on the surface did not appear very earth-shattering. Already companies could stand up the open source software (OSS) online transaction processing (OLTP) implementation of the MySQL database on AWS directly, or they could simply use Amazon Relational Database Service (RDS), which supports several database formats, including MySQL Community Edition. Or they could provision AWS’ own specialized implementation of MySQL, Amazon Aurora, to gain an extra kick in performance and scale. Both of these are specialized OLTP databases that require ETLs to separate database services for analytics or machine learning—neither of these is free. In contrast, MySQL HeatWave features in-database OLTP, analytics, and machine learning—no ETLs required.

What is one more implementation of MySQL then? As with the housing market, location is everything. As mentioned above, most multicloud software implementations run directly on top of the underlying infrastructure of the host platform. But Oracle has taken a different approach. MySQL HeatWave is a native implementation on AWS that includes data plane, control plane, and an interactive console. To the AWS user running MySQL HeatWave on AWS, this database runs as any AWS-hosted service. Via the interactive console, users can manage schemas, execute queries, and monitor the performance of their queries and the utilization of the provisioned resources. MySQL Autopilot is also integrated with the interactive console, making it easier to use, and has access to AWS services. With MySQL HeatWave on AWS, there is no need to migrate data or ETL data across databases, which means no egress fees.

Going deeper with Microsoft Azure

Oracle’s unique approach to MySQL HeatWave deployment takes an interesting and compelling turn with the company’s November announcement that Microsoft Azure users could provision MySQL HeatWave running back on OCI as if it was any other Azure resource. Stated a different way, on AWS, MySQL HeatWave runs as a platform-native service with data plane, control plane, and console all inside AWS. On Azure, however, MySQL HeatWave backend services run back on Oracle’s own OCI infrastructure where they can benefit from OCI’s underlying hardware capabilities. Data will flow between the two cloud platforms via Oracle’s network of high-speed cloud interconnections built with a partner, Equinix. However, Oracle will not charge customers for any of this data movement between the two clouds. Customers will simply pay for the OCI and Azure resources consumed locally on each platform, respectively.

This approach to multicloud deployment on Microsoft Azure may sound somewhat counter-intuitive. Why try to split both storage and compute over two clouds just to present users with a fully native user experience? As it turns out, Oracle has a very good reason to do this, and that reason centers upon Oracle's desire to deliver an optimal price/performance ratio to users. Like many cloud platform providers, Oracle has modified the core MySQL database, creating a series of performance-related architectural changes:

- A built-in, columnar, in-memory analytics engine centered on overall performance as well as query acceleration
- A scale-out data management layer built on top of OCI's object storage capable of rapid restarts and error recovery
- An AI-influenced automation layer (MySQL Autopilot) that improves database performance and automates many database operations such as provisioning, data loading, and query processing
- In-database machine learning (ML) training and inference facilities, capable of running against data in near-real-time without extract transform and load (ETL) requirements
- In-database security measures such as server-side data masking and de-identification, as well as firewalling, and asymmetric data encryption

Whether natively on AWS or remotely on Microsoft Azure, in building MySQL to run in a highly optimized, Oracle claims to deliver an optimized OLTP cost model that outperforms native MySQL offerings, Amazon RDS, Microsoft Azure MySQL, and Google CloudSQL at one-third the cost. The company is confident in these claims, so confident that it has posted its findings, code, scripts, and all configuration data on GitHub, allowing customers and competitors alike to test and replicate its findings.

Of converged databases and data lakehouses

Price/performance optimization for running queries is indeed crucial. Yet, there is a larger specter that haunts multicloud AI, data, and analytics workloads. That specter is data movement, a seemingly ubiquitous malady that arises any time a database user extracts, loads, duplicates, exports, and queries data across more than one data store. In chasing AI and analytics workloads in particular, this kind of data movement can prove costly in both time and money, especially if data needs to move from one cloud platform to another—a common task that typically results in heavy data egress and sometimes ingress charges. Even within the same cloud platform, the cost of moving data in and out between OLTP, analytical, and AI solutions can prove cost- and time-prohibitive.

To solve this problem, at least within the confines of a single public or private cloud, many enterprises are turning to multi-model or “converged” databases (to use Oracle's parlance). As illustrated in Omdia's recent look at modern data analytics platforms (See *Omdia Universe: Modern Data Analytics Platform, 2022*), all major data warehousing solutions are embracing both structured and unstructured data in an effort to mitigate the cost of managing and integrating numerous data silos. As with any impactful idea, this notion has been central to Oracle Database and Autonomous Database approach for years.

It is now beginning to catch on with newer transactional databases as demonstrated by MongoDB Atlas and Oracle MySQL HeatWave—both of which enable application developers to conduct near real-time analytics without having to first move data to an external data warehouse or data lake. In contrast, other hyperscalers, notably AWS, promote the use of multiple single-purpose cloud database services, charging separately for each as well as the tools and storage for moving data. With MySQL HeatWave, Oracle is

promoting a more converged ideal, one that adds analytics, machine learning, and Autopilot on top of MySQL without additional cost.

Now Oracle is pushing the edges of this notion even further by announcing at Oracle CloudWorld a new edition of MySQL HeatWave, Oracle MySQL HeatWave Lakehouse. Currently in beta and expected to reach general availability in the first half of 2023, running on OCI initially, this new offering will enable users to pull in and then query semi-structured data residing within common data lake cloud object stores. At the outset, this object storage will include native MySQL syntax access to data exported from both Amazon Redshift and Aurora. Users will also be able to load data automatically from CSV and the popular Parquet format.

Leveraging MySQL HeatWave's massively parallel scale-out architecture and AI-fueled automation and query optimization, this new offering is built for speed and scale, accommodating queries reaching up to 400 terabytes of data across 512 nodes (note that the solution starts at 16 gigabytes and 64 nodes). Here too, Oracle intends to demonstrate a superior price/performance ratio, so far citing a 6x advantage over Amazon Redshift and 17x over Snowflake in query performance, and advantages of 8x over Redshift and 2.7x over Snowflake in data loading.

A good part of this difference is due to MySQL HeatWave's Autopilot, which uses techniques like automatic schema inference, adaptive sampling, auto provisioning, and load, as well as automatic query plan optimization to both speed performance and reduce manual, hands-on administrative work—a key idea Oracle has promoted within Oracle Autonomous Database and Autonomous Data Warehouse running natively on Oracle Cloud and OCI infrastructure. These offerings greatly cut down on the complexity and propensity for human errors in managing complex database workloads.

A tale of home fields and data lakehouses

For enterprises seeking to build responsive, cloud-savvy applications using MySQL that can be enriched through real-time analytics and AI, Oracle's new data lakehouse presents an interesting conundrum. Should they stick with a pure OSS implementation that is self-managed, should they utilize the host platform provider's managed implementation MySQL, or should they opt for Oracle's highly automated and managed MySQL HeatWave (and HeatWave Lakehouse) database for potentially less money, greater scale, higher performance, and availability in multiclouds—OCI, AWS, and Azure?

Ultimately, Oracle's goal is to encourage users to use data from Amazon RDS, Aurora, and Redshift in HeatWave without having to migrate data. The same goes for other multicloud analytical databases like Snowflake and Databricks Delta Lake. Overall, Omdia views the announcements made at Oracle CloudWorld 2022 bring Oracle closer to this goal. For users seeking to natively blend transaction processing, analytics, ML, and data lakehouse functionality within a single database, Oracle's automated approach to database management and optimization will hold great appeal, especially as that database can run faster, more efficiently, and at lower cost than its rivals.

Even so, as of this writing, Oracle still has some work to do to completely overcome the home field advantage on third party cloud platforms. On a technical front, Oracle MySQL HeatWave Lakehouse is still incomplete as its in-database AutoML feature can only work on tabular data and cannot yet incorporate unstructured object store data. Further, Oracle will need to incorporate a wider array of data load file formats such as AVRO and ORC.

Regardless, between Oracle's unique approach to multicloud deployments and the company's approach to converged database workloads, the Oracle MySQL HeatWave family may just give Oracle the boost it is looking for, a boost that enables them to truly rival hyperscalers. After all, with MySQL HeatWave and

HeatWave Lakehouse, Oracle is presenting MySQL customers on AWS and Microsoft Azure with a proposition they may not be able to refuse.

Appendix

Further reading

[“HeatWave TPC-H,”](#) GitHub (retrieved October 2022)

[Omdia Universe: Modern Data Analytics Platform](#) (July 2022)

Author

Bradley F. Shimmin, Chief Analyst, AI Platforms, Analytics, and Data Management

askananalyst@omdia.com

Citation policy

Request external citation and usage of Omdia research and data via citations@omdia.com.

Omdia consulting

We hope that this analysis will help you make informed and imaginative business decisions. If you have further requirements, Omdia's consulting team may be able to help you. For more information about Omdia's consulting capabilities, please contact us directly at consulting@omdia.com.

Copyright notice and disclaimer

The Omdia research, data and information referenced herein (the "Omdia Materials") are the copyrighted property of Informa Tech and its subsidiaries or affiliates (together "Informa Tech") or its third party data providers and represent data, research, opinions, or viewpoints published by Informa Tech, and are not representations of fact.

The Omdia Materials reflect information and opinions from the original publication date and not from the date of this document. The information and opinions expressed in the Omdia Materials are subject to change without notice and Informa Tech does not have any duty or responsibility to update the Omdia Materials or this publication as a result.

Omdia Materials are delivered on an "as-is" and "as-available" basis. No representation or warranty, express or implied, is made as to the fairness, accuracy, completeness, or correctness of the information, opinions, and conclusions contained in Omdia Materials.

To the maximum extent permitted by law, Informa Tech and its affiliates, officers, directors, employees, agents, and third party data providers disclaim any liability (including, without limitation, any liability arising from fault or negligence) as to the accuracy or completeness or use of the Omdia Materials. Informa Tech will not, under any circumstance whatsoever, be liable for any trading, investment, commercial, or other decisions based on or made in reliance of the Omdia Materials.

CONTACT US

omdia.com

askananalyst@omdia.com

