ORACLE

HeatWave

Real-time Analytics for MySQL Database Service

Purpose statement

This document provides an overview of features and enhancements included in HeatWave. It is intended solely to help you assess the benefits of HeatWave and to plan your I.T. projects.

Disclaimer

This document in any form, software or printed matter, contains proprietary information that is the exclusive property of Oracle. Your access to and use of this confidential material is subject to the terms and conditions of your Oracle software license and service agreement, which has been executed and with which you agree to comply. This document and information contained herein may not be disclosed, copied, reproduced or distributed to anyone outside Oracle without prior written consent of Oracle. This document is not part of your license agreement nor can it be incorporated into any contractual agreement with Oracle or its subsidiaries or affiliates.

This document is for informational purposes only and is intended solely to assist you in planning for the implementation and upgrade of the product features described. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described in this document remains at the sole discretion of Oracle. Due to the nature of the product architecture, it may not be possible to safely include all features described in this document without risking significant destabilization of the code.

Benchmark queries are derived from the TPC-H benchmark, but results are not comparable to published TPC-H benchmark results since they do not comply with the TPC-H specification.

ORACLE

Table of contents

ORACLE

Executive Summary

MySQL has become the world's most popular open source database because of its reliability, high-performance, and ease of use. It powers the world's most trafficked web sites including Facebook, Twitter, LinkedIn and Booking.com.

The amount of data that organizations will need to manage is growing significantly. According to IDC[1], the sum of the world's data – the DataSphere — will grow from 59 zettabytes in 2020 to a mind-boggling 175ZB by 2025. By 2025, 49% of all data will be stored in public cloud environments and nearly 30% of the data generated will be consumed in real-time.

The companies that will thrive in the evolving digital landscape, will be those that make data and analytics a core part of their strategy. According to McKinsey[2], 92% of company leaders surveyed believed that their business model would not remain viable at the current rate of digitization. This fear of disruption is the leading driver behind the investment in modern data and analytics platforms. Fortune 1000 companies are recognizing that they must leverage their data assets if they are to compete successfully against their highly agile, data-driven competitors.

Oracle MySQL Database Service is a fully managed database service that lets developers quickly develop and deploy secure cloud native applications using the world's most popular open source database. MySQL Database Service is the only MySQL service with a massively-scalable integrated real-time analytics engine: HeatWave. This service, only available in Oracle Cloud Infrastructure, overcomes the limitations of traditional data warehouse and analytics environments that use periodic long-running ETL batch jobs required to refresh the data.

Challenges of Existing Data & Analytics Solutions

MySQL is optimized for OLTP, but it is not designed for analytic processing (OLAP). As a result, organizations which need to efficiently run analytics need to move their data to another database.

This approach of moving data to another database introduces complexity and additional cost to customers in multiple ways:

1. **Applications need to define complex logic** for extracting relevant data from MySQL.

2. **The extracted data needs to be transported to another database** across networks securely, consuming network bandwidth and incurring latency.

3. **Data in the other database needs to be manually kept in sync** with the MySQL database and as a result the data on which analytics is performed can become stale.

ORACLE

4. **Additional cost and overhead of managing multiple databases** for running OLTP and analytics applications.

Performance: Real Time Analytics

HeatWave is designed to enable customers to run analytics on data which is stored in MySQL databases without the need for ETL. This service is built on an innovative, in-memory analytics engine which is architected for scalability and performance and is optimized for Oracle Cloud Infrastructure (OCI) Gen 2 hardware. This results in a very performant solution for SQL analytics at a fraction of the cost compared to other industry solutions. For example, a customer running TPC-H Benchmark workloads with HeatWave will experience a 400x speedup over MySQL 8.0.

> "We ran the analytic workload of the Social Game Infrastructure group with HeatWave without requiring ETL or changing our application. HeatWave is up to 500x faster than our current on-premises MySQL 5.7 instance."
>
> **Masashi Hamahira**
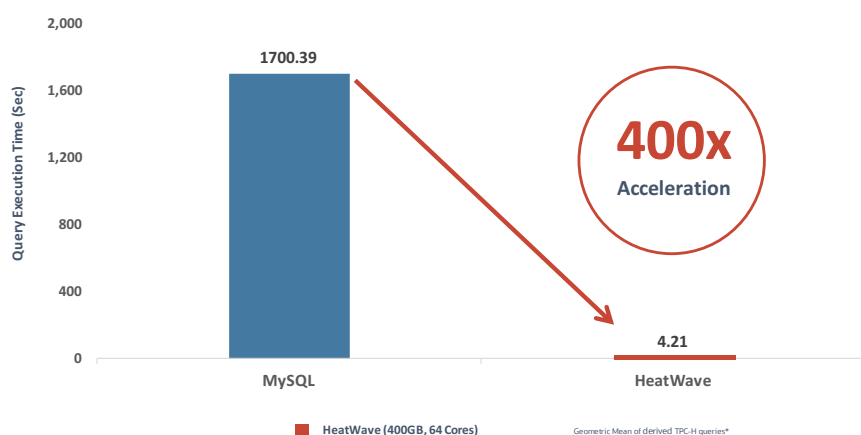> Senior Manager, Social Game Infrastructure Group
> Square Enix Co. Ltd.

Figure 1. HeatWave accelerates MySQL queries by 400x

Compared to Amazon Aurora, MySQL Database Service with HeatWave offers dramatic improvement in performance for complex and analytic queries. For example, on a 4TB TPC-H workload, MySQL Database Service with HeatWave is 1100x faster than Amazon Aurora. Furthermore, with HeatWave there is no need to create indexes on the base table which takes days with Amazon Aurora. As a result, the data is available to query much sooner with HeatWave than with Aurora.
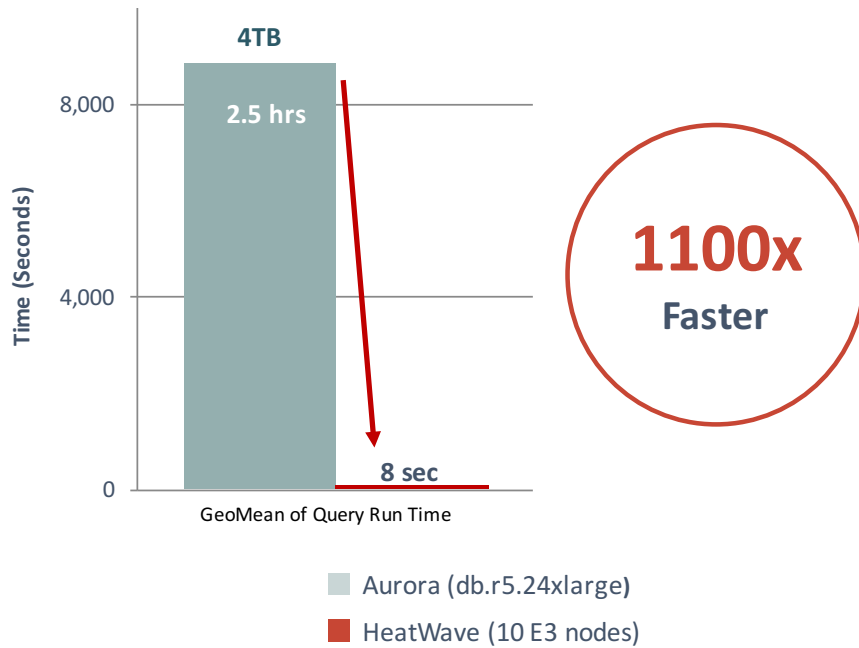
ORACLE

Figure 2. HeatWave is 1100x faster than Amazon Aurora

The performance improvement of MySQL Database Service with HeatWave over Aurora increases with the size of data.
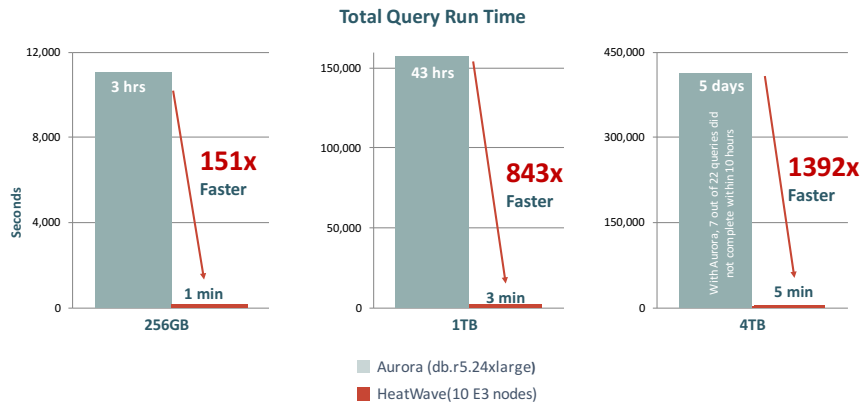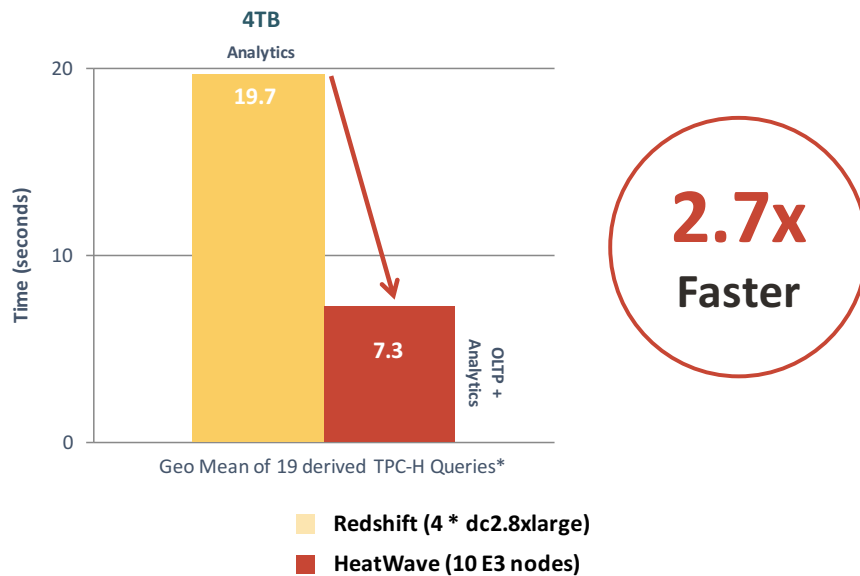

Figure 3. The performance advantage of HeatWave increases with data size vs. Amazon Aurora

Next, we compare with Amazon Redshift which is designed for analytics and is offered in multiple shapes. Compared to the fastest shape (dc2.8xlarge), HeatWave is up to 3x faster. Compared to the cheaper shape (RA3.4xLarge), HeatWave is up to 18x faster.

Also, unlike Amazon Redshift, HeatWave is capable of both OLTP and OLAP, without the need for ETL.

"We discovered astonishing performance improvements during our tests of the new HeatWave, reducing query times from hours to seconds in cases of non-index based queries. This enables us to expand our service and open the door to many more analytics-based capabilities while maintaining a single, seamless database platform."

Dr. Jens Uecker
Senior Key Expert
Siemens Healthineers

ORACLE

Figure 4. HeatWave is 2.7x faster than Amazon Redshift's fastest shape.
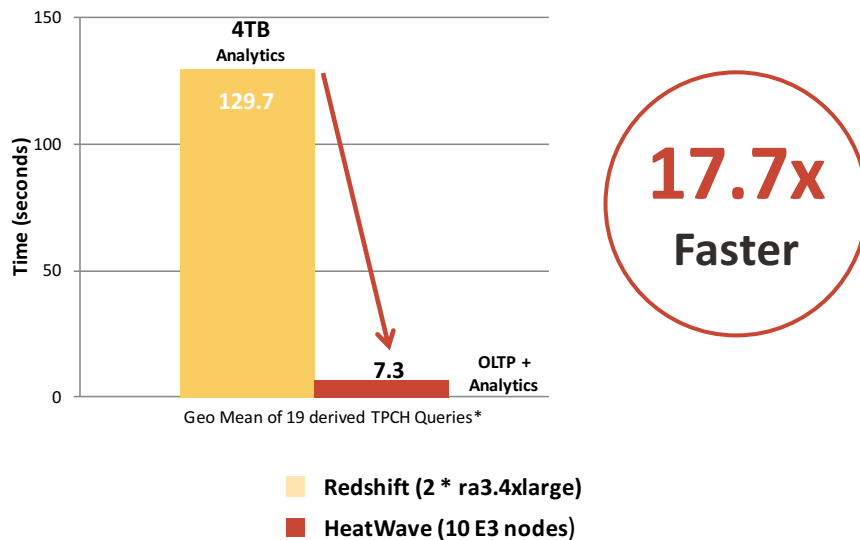
Figure 5. HeatWave is 17.7x faster than Amazon Redshift's cheaper shape

> "The performance of the MySQL database service analytics engine is over 20x faster than the analytics service of one of the largest cloud vendors, and accelerates our analytics workload by 45,000x compared to MySQL on-premises."
>
> **Kenichi Sasaki**
> SRE Team of Japan
> Mercari, Inc

Customers who use HeatWave will benefit from significantly better performance, eliminating the need for ETL, support for real-time analytics, reduced monthly cost and a single database for OLTP and OLAP.

Deployment Scenarios

HeatWave is a fully managed service and is available exclusively in the Oracle Cloud Infrastructure. The MySQL database has been enhanced to natively integrate this service, as a result, customers who store data in MySQL can seamlessly run analytics by enabling this service.

A HeatWave instance is a cluster composed of a MySQL Database Service (MDS) instance and multiple analytics nodes. When HeatWave is enabled, HeatWave server is installed on the MDS node. It is responsible for cluster management, loading data into the memory of the HeatWave nodes, query

ORACLE

scheduling and query execution. MySQL applications written in Java, PHP, Ruby, etc. work seamlessly with HeatWave using standard MySQL ODBC/JDBC connectors. HeatWave is an in-memory accelerator and supports all MySQL syntax. Hence, all existing tools and applications built using standard SQL will work without requiring any modification to queries (Figure 6).
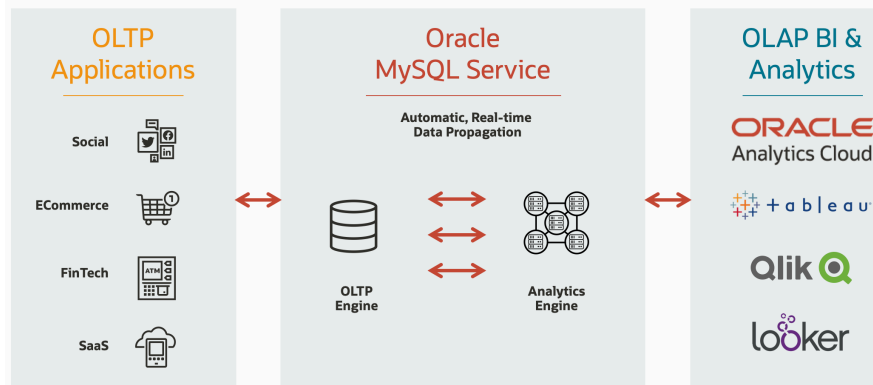

Figure 6. Existing tools and applications work seamlessly with HeatWave

Data which is needed for analytic processing is stored in HeatWave nodes in memory, in a hybrid columnar compressed format. The number of nodes needed to run for a workload depends on the amount of data present for analytic processing, the compression factor which is achieved on the dataset, and the query characteristics. The number of nodes needed can be automatically derived by using the Auto Provisioning advisor which is available with HeatWave.

HeatWave currently supports up to 24 nodes in one cluster and with a processing capacity of approximately 10 TB of analytics data. 10TB is the approximate amount of data which can be populated in the memory of the HeatWave nodes at a given moment. There is no limit to the amount of data which can be stored in the MySQL database and customers can choose which tables or columns from MySQL database schema to load into the memory of HeatWave nodes. If the tables are no longer needed by queries, user can remove the tables from the memory to make room for other data.

HeatWave provides a great solution for customers who need to run both transactional and analytical workloads. While transactional queries are run in the MySQL node, data updated in MySQL InnoDB is transparently propagated to the HeatWave cluster for accelerated analytical processing. This enables customers to run both OLTP and real-time analytics workloads simultaneously within a single database platform (Figure 7).

"HeatWave is 10 times faster than the analytics service of another major cloud vendor. Now there is no need for ETL. Compared to MySQL on-premises, the HeatWave is 4000 times faster."

Tetsuro Ikeda
Manager of Cloud IT Architecture Service Department
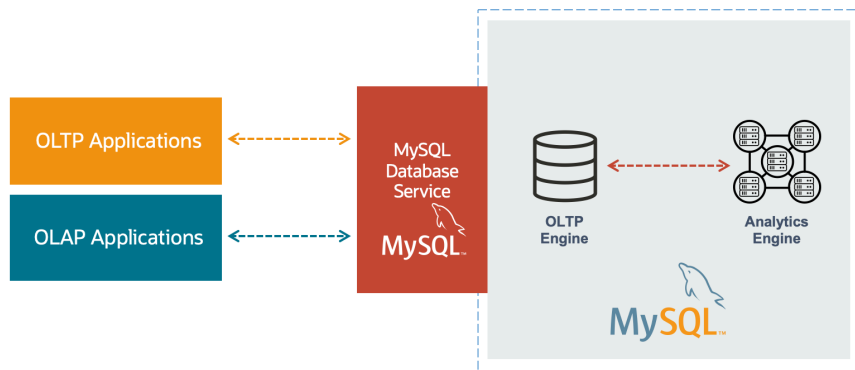SCSK Corporation

ORACLE

Figure 7. Single MySQL database for OLTP & OLAP applications, no ETL required.

On-premises customers who cannot move their MySQL deployment to a Cloud due to compliance or regulatory requirements, can still leverage HeatWave by using the hybrid deployment model as shown in Figure 8. In such a hybrid deployment, customers can leverage MySQL replication to replicate on-premise MySQL data to HeatWave without the need for ETL.
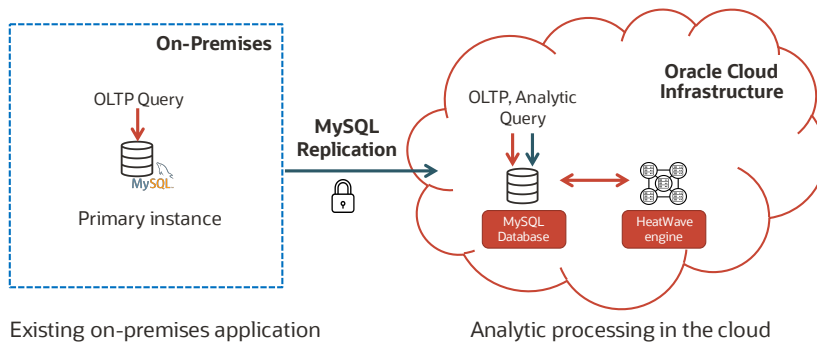

Figure 8. Hybrid deployment for enabling analytics on data stored on premise

HeatWave Architecture

There are three key architecture choices which lead to a compelling performance and cost advantage with HeatWave:

1. Innovative in-memory columnar analytics engine designed for scalability and performance which implements state of the art algorithms

2. Optimized for Oracle Cloud Infrastructure

3. Automation of various capabilities

The Heatwave engine uses a columnar in-memory representation that facilitates vectorized processing leading to very good query performance (Figure 9). The data is encoded and compressed prior to being loaded in memory. This compressed and optimized in memory representation is used for both numeric and string data. This results in significant performance speed up and reduced memory footprint which translates to reduced cost for customer.

"HeatWave surprised us by improving query performance significantly without modifying any queries or using any ETL tools. Now we can focus on innovating e-commerce business services."

Yoshihiko Goto
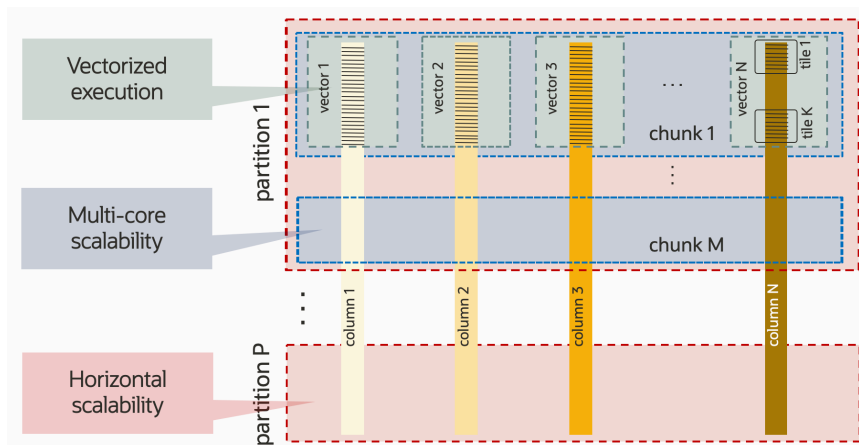Service Front Team, Server Infrastructure Group
DMM.com LLC

ORACLE

Figure 9. Vectorized in-memory columnar representation for analytic processing

One of the key design points of the HeatWave engine is to massively partition data across a cluster of HeatWave nodes, which can be operated upon in parallel in each node. This enables high cache hits for analytic operations and provides good inter-node scalability. Each HeatWave node within a cluster and each core within a node can process partitioned data in parallel, including parallel scans, joins, group-by, aggregation and top-k processing.

HeatWave has implemented state of art algorithms for distributed in-memory analytic processing. Joins within a partition are processed fast by using vectorized build and probe join kernels. The highly-optimized network communication between analytics nodes is achieved by using asynchronous batch I/Os. The algorithms are designed to overlap compute time with communication of data across nodes which helps achieve good scalability (Figure 10).
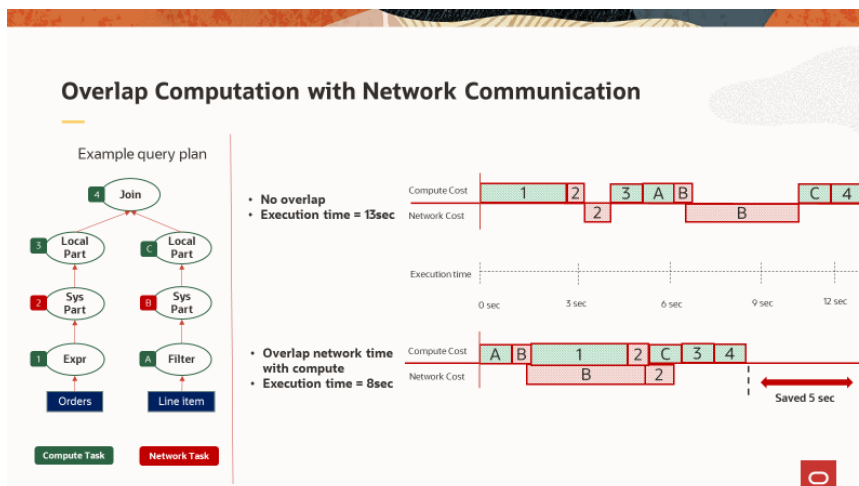

Figure 10. HeatWave implements state of art algorithms for distributed in-memory analytic processing.

ORACLE

Native MySQL Analytics

Integration of HeatWave with MySQL Database Service provides a single data management platform for all OLTP and analytics needs of an enterprise. HeatWave is designed as a MySQL pluggable storage engine, which completely shields all the low-level implementation details at the storage level from the end users. As a result, users can manage both HeatWave and MySQL Database Service with the same management tools including OCI console, REST API, and command line interface.

Since HeatWave is an in-memory processing engine, data is persisted in MySQL InnoDB storage engine. This allows users to manage analytics data the same way they manage transactional data in MySQL.

Users and applications interact with HeatWave through the MySQL database node in the cluster. Users connects to HeatWave through standard tools and standard-based ODBC/JDBC connectors. HeatWave supports the same ANSI SQL standard and ACID properties as MySQL and supports diverse data types. This enables existing applications to take advantage of HeatWave without any changes to their application, allowing easy and quick integration.

Once users submit a query to the MySQL database, the MySQL query optimizer transparently decides if the query should be offloaded to HeatWave cluster for accelerated execution. This is based on whether all operators and functions referenced in the query are supported by HeatWave and if the estimated time to process the query with HeatWave engine is less than with MySQL. If both conditions are met, the query is pushed to HeatWave nodes for processing. Once processed, the results are sent back to the MySQL database node and returned to users (Figure 11).
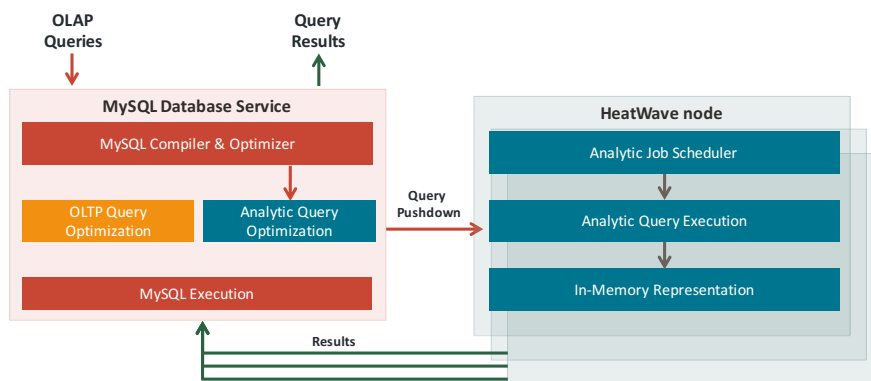

Figure 11. MySQL Integration for Query Processing

Data of HeatWave is persisted in MySQL InnoDB. Any updates to the tables are automatically propagated to the memory of the HeatWave nodes in real time. This allows subsequent queries to always have access to the latest data, as shown in Figure 12. This is done behind the scene by a light-weight change propagation algorithm that can keep up with MySQL data update rates.
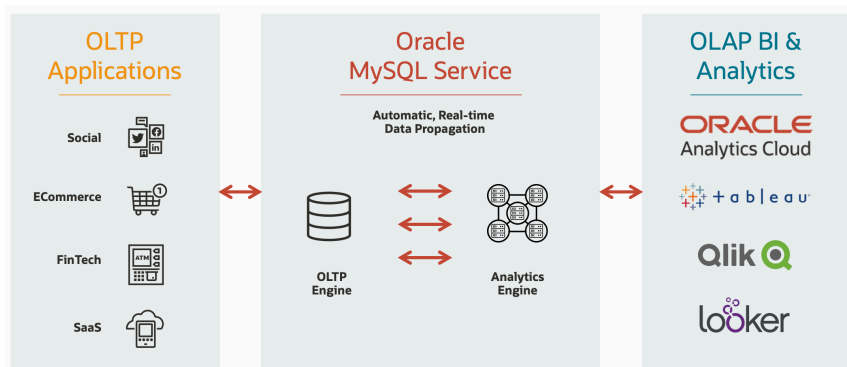
ORACLE

Figure 12. HeatWave integration with MySQL Database Service

Auto Provisioning

HeatWave leverages auto machine learning (AutoML) capabilities developed in Oracle Labs to automate various aspects of the service. This automation is helpful for customers and engineers managing the service. Since this automation is based on machine learning, the system can intelligently predict and act upon various scenarios.

Auto Provisioning provides recommendation on how many analytics nodes are needed to run a workload. When the service is started, database tables on which analytics queries are run need to be loaded to HeatWave cluster memory. The size of the cluster needed depends on tables and columns required to load, and the compression achieved in memory for this data. Figure 13 compares the traditional (i.e., manual) approach to estimating the cluster size with Auto Provisioning. In traditional provisioning, the user needs to guess a cluster size. Underestimation results in data load or query execution failure due to space limitations. Overestimation results in additional costs for unneeded resources. As a result, users iterate until they determine the right cluster size and this size estimate becomes inaccurate when tables are updated.
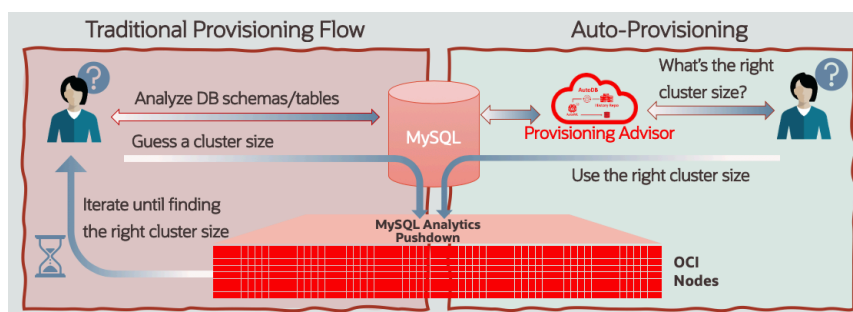

Figure 13. Comparison of a manual provisioning vs Auto provisioning

The right side of figure 13 shows how we solve this problem with Auto Provisioning, which is a ML-based cluster size estimation advisor. By leveraging well trained and accurate ML models, the user consults the Auto Provisioning model to obtain the right cluster size for their dataset. As a result, users do not need to guess the cluster size. Later, if the customer data grows or additional tables are added, the users can again take advantage of the Auto Provisioning advisor.

ORACLE

Integration with Oracle Cloud Services

OCI offers a wide range of services for data analytics, machine learning, and data lake. Native integration with these services makes it easier for existing applications to use HeatWave (Figure 14).

Oracle Analytics Cloud (OAC) provides the industry's most comprehensive cloud analytics in a single unified platform, including self-service visualization and inline data preparation to enterprise reporting, advanced analytics, and self-learning analytics that deliver proactive insights. Integration with OAC provides BI visualization platform for users to analyze their MySQL data.

OCI Data Integration Service provides extract, transform and load (ETL) capabilities to target data warehousing scenarios on the OCI platform. It supports various data sources, starting with relational, cloud and Hadoop. Integration with OCI Data Integration allows users to easily transform and import data from data sources other than MySQL to HeatWave, expanding the scope of data to take advantage of HeatWave.
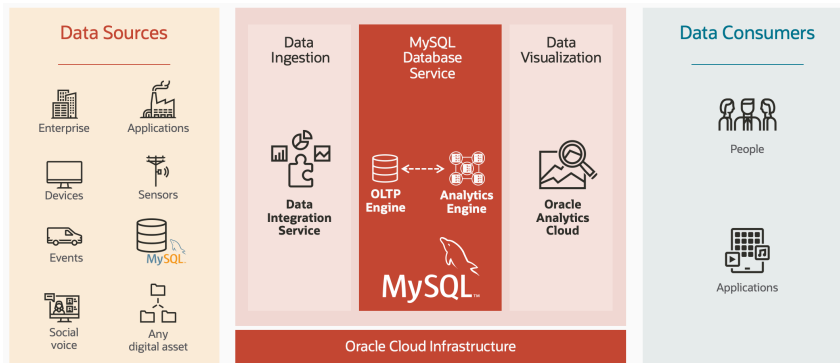

Figure 14. End to end integration with other Oracle Cloud Services from Data Ingestion to Data Visualization

1/3 the Cost of Amazon Aurora and Redshift

The cost of using HeatWave depends upon the number of HeatWave nodes provisioned. The size of an Heatwave cluster depends on the size of the dataset, and the characteristics of the workload. A single HeatWave node can hold approximately 400GB of data. Customers can expect to see a significant reduction in their costs when they migrate to HeatWave. Compared to Amazon Aurora and Redshift, HeatWave is 1/3 the cost (Figure 15 & 16).

Customers are likely to find that the cost benefit with HeatWave is higher because of the following:

- Aurora charges additional for Storage IO cost which can be substantial. There is no such cost with HeatWave

- When using HeatWave, the cost includes both OLTP and OLAP capabilities. With Amazon Redshift the cost is only for OLAP, and additional costs are needed for OLTP.

ORACLE

- Customers need to pay as they move their data around from one database to another with Amazon Aurora or Amazon Redshift.
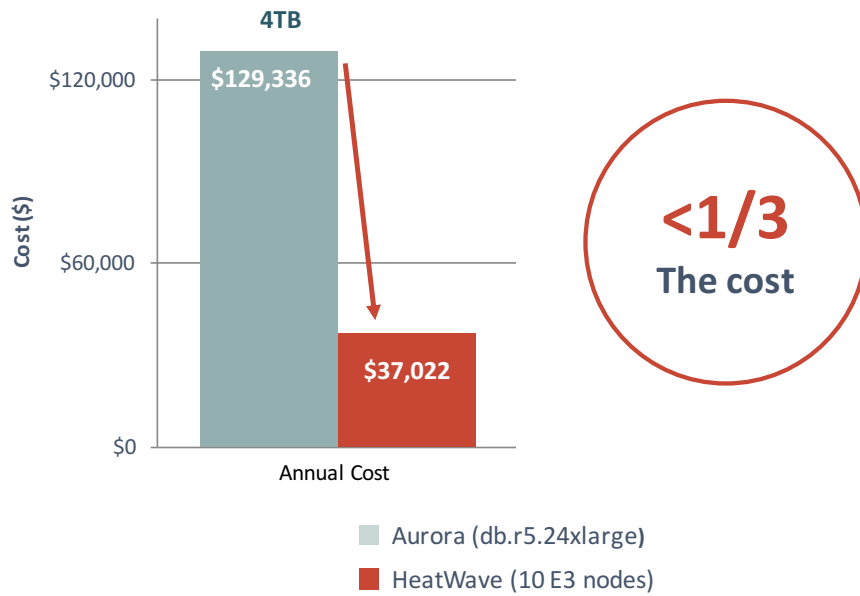
4TB
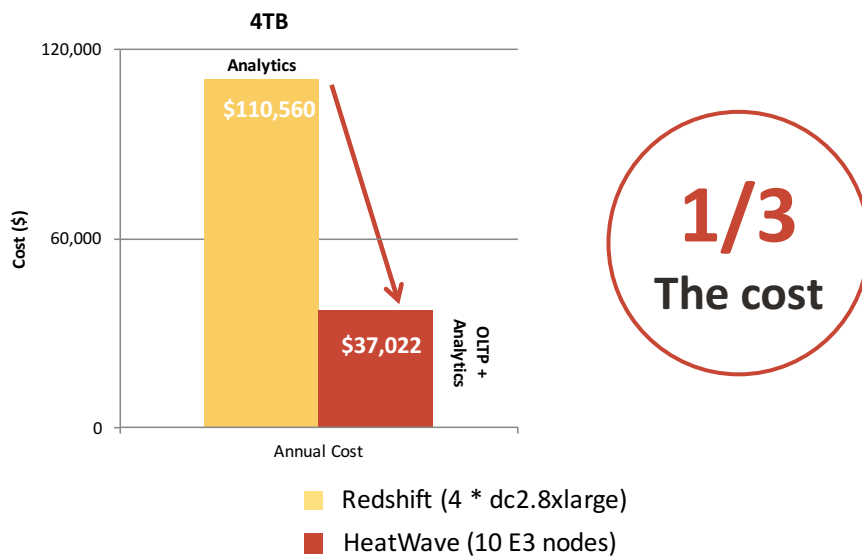

Figure 15. HeatWave is 1/3 the cost of Amazon Aurora

4TB

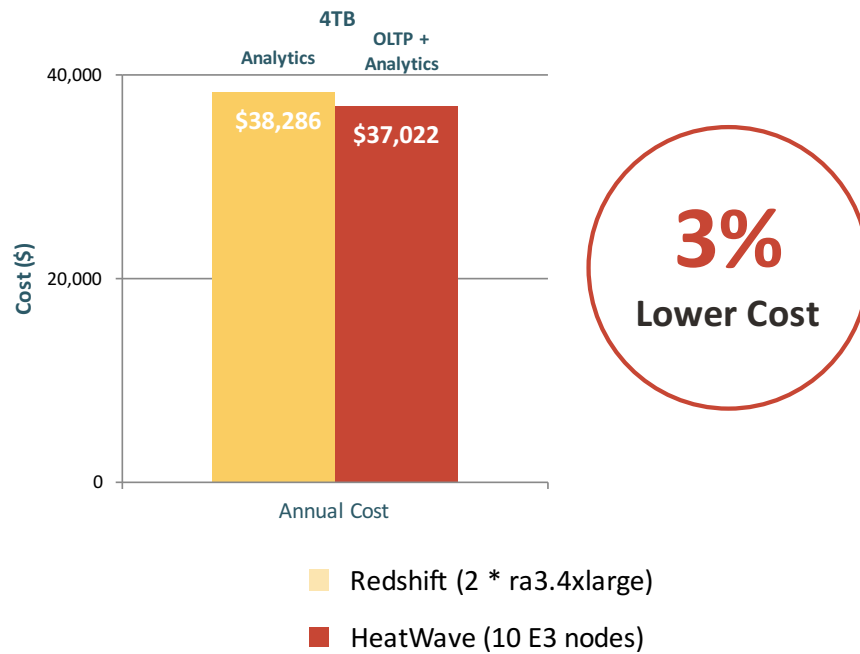Figure 16. HeatWave is 1/3 the cost of Amazon Redshift's fastest shape

ORACLE

4TB

	Analytics	OLTP + Analytics
$38,286	$37,022	

3%
Lower Cost

■ Redshift (2 * ra3.4xlarge)

■ HeatWave (10 E3 nodes)

Figure 17. HeatWave is 3% lower cost than Amazon Redshift's cheaper shape and 17.7x faster.

HeatWave offers pay as you go and Universal Credit Annual Flex pricing. Pay as you go allows customers to pay for only what they use, lowering the cost when they don't need to use the service 24/7. In both pricing models, organizations only need to pay for what they use.

Conclusion

HeatWave is a cloud native service which is exclusively available in Oracle cloud Infrastructure and provides compelling performance and cost for analytic workloads. Organizations using MySQL database for managing their enterprise data can now run analytic queries with HeatWave with significantly better performance, lower cost, not requiring ETL and support for real- time analytics. The service can be deployed in a cloud only or in a hybrid environment, and it simplifies management for both transactional and analytic applications.

ORACLE

[1] IDC's Global DataSphere Forecast Shows Continued Steady Growth in the Creation and Consumption of Data. https://www.idc.com/getdoc.jsp?containerId=prUS46286020. May 2020

[2] How 6 Companies are Using Technology and Data to Transform Themselves. https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/how-six-companies-are-using-technology-and-data-to-transform-themselves. August 2020

Connect with us

Call +**1.800.ORACLE1** or visit **oracle.com**. Outside North America, find your local office at: **oracle.com/contact**.

blogs.oracle.com   facebook.com/oracle   twitter.com/oracle