ORACLE

# HeatWave on AWS

Integrated, automated, and secure generative AI and machine learning in one fully managed cloud service for transactions and lakehouse scale analytics—on Amazon Web Services (AWS).

## 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.

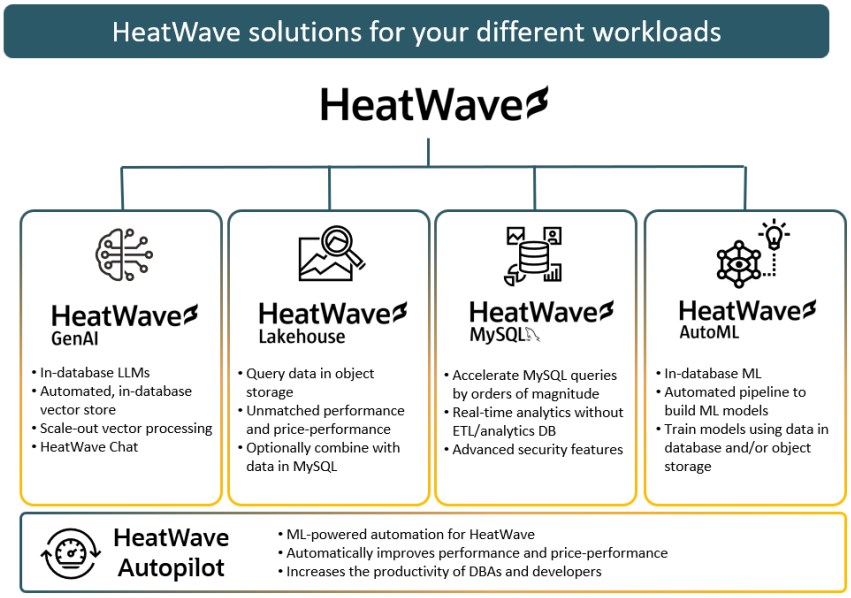
This document is for informational purposes only and is intended solely to assist you in planning for the implementation and upgrade of the product features described. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. 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 and TPC-DS benchmark, but results are not comparable to published TPC-H and TPC-DS benchmark results since they do not comply with the TPC-H TPC-DS specification.

Table of contents

## Introduction

HeatWave is a fully managed cloud service that enables organizations to efficiently run transactions, real-time analytics across data warehouses and data lakes, plus automated generative AI and machine learning. The scale out design of HeatWave enables organizations to achieve very high performance and price performance for processing structured, semi-structured and unstructured data. With HeatWave GenAI, organizations can create a vector store and run large language models (LLMs) inside the database, providing them a secure, simple, efficient, and low cost choice to build a new class of generative AI applications.




HeatWave eliminates the need for complex, time-consuming ETL operations and unnecessary data duplication between separate databases and tools for generative AI, machine learning, analytics, and transaction processing use cases. Customers avoid the latency and security risks of data movement between data stores while reducing costs. HeatWave also includes HeatWave Autopilot, providing workload-aware, machine learning-powered automation of various aspects of the database system lifecycle, including provisioning, data management, query execution, and failure handling for both OLTP and OLAP workloads.

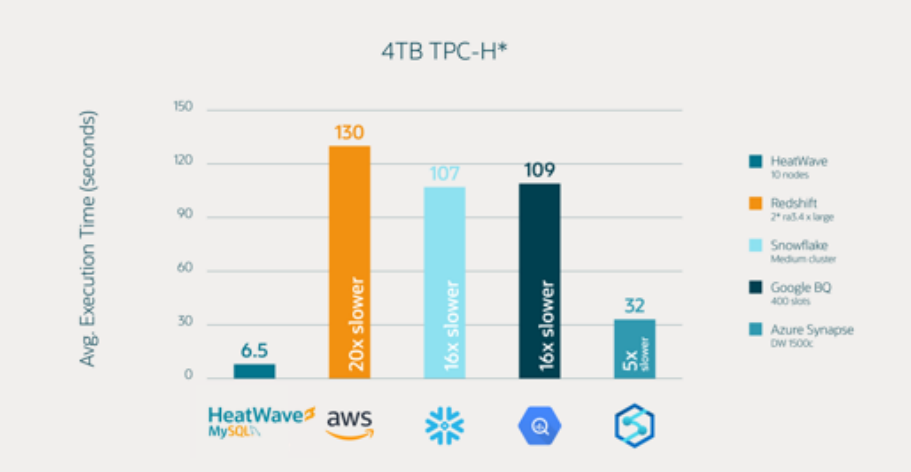
Oracle makes all these HeatWave capabilities, which are built, managed, and supported by the HeatWave development team, natively available on AWS. All components of the HeatWave service on AWS, namely the service console, control plane, and data plane, are built and optimized for AWS. HeatWave's native integration with AWS enables customers with applications already deployed on AWS to benefit from HeatWave without incurring the latency of accessing a database service running outside of AWS. This eliminates the data egress fees charged by AWS to move data to a service running outside of AWS. Lastly, the tight integration of HeatWave with AWS services such as Amazon S3,

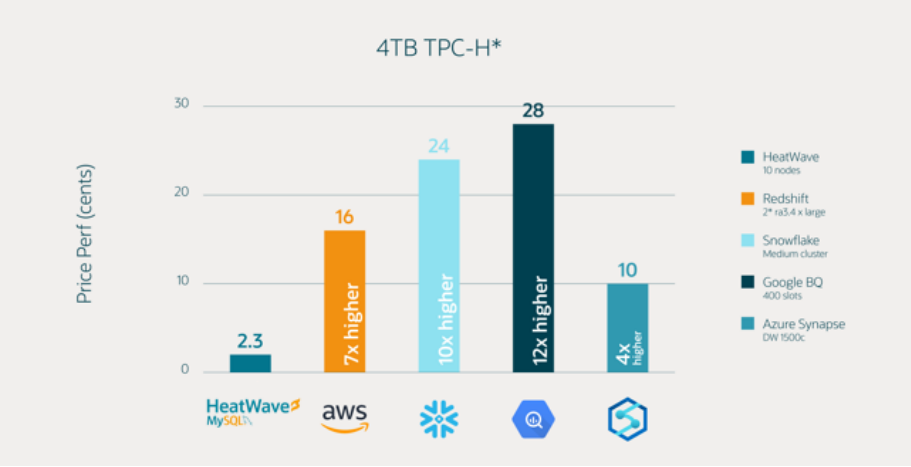
> "HeatWave GenAI makes it extremely simple to take advantage of generative AI. The support for in-database LLMs and in-database vector creation leads to significant reduction in application complexity, predictable inference latency, and, most of all, no additional cost to us to use the LLMs or create the embeddings."
>
> **Vijay Sundhar**
>
> Chief Executive Officer, SmarterD

CloudWatch, and PrivateLink, makes it easy for developers to rely on HeatWave for new applications.

## Performance and Price Performance Advantages

With its highly performant data-processing architecture and optimization for AWS infrastructure, HeatWave on AWS delivers unmatched performance and price-performance. As demonstrated by a 4TB TPC-H benchmark, HeatWave MySQL on AWS delivers 7X better price performance than Amazon Redshift, 10X better than Snowflake, 12X better than Google BigQuery, and 4X better than Azure Synapse. For machine learning, HeatWave MySQL on AWS is 25X faster than Redshift ML. On a 10GB TPC-C workload, HeatWave MySQL offers up to 10X higher and sustained throughput compared to Amazon Aurora at high concurrency. The benchmark scripts are available on GitHub for customers to replicate.




*Only compute costs are considered in the above graphs. Pricing for Redshift is based on 1-year reserved instance, paid upfront. Pricing for Snowflake is based on standard edition. Pricing for Google Big Query is based on monthly flat rate commitment. Pricing for Azure Synapse is based on 1-year reserved pricing. *Benchmark queries are derived from the TPC-H benchmarks, but results are not comparable to published TPC-H benchmark results since these do not comply with the TPC-H specifications.*

## HeatWave MySQL

HeatWave MySQL provides a single database system for high-performance and secure transaction processing, real-time analytics, and machine learning where the data is stored inside MySQL database. Each HeatWave instance consists of a MySQL Database node, mainly targeting transaction processing, and a configurable number of HeatWave nodes for analytical processing, machine learning, and generative AI tasks. HeatWave on AWS strives to provide the best price-performance by optimizing all components of the underlying AWS infrastructure—while keeping costs as low as possible.

Oracle provides the latest MySQL Enterprise Edition version in HeatWave MySQL on AWS, which contains additional functionalities and security features developed, maintained, and supported by the MySQL team at Oracle. With each new release of the MySQL server, customers of HeatWave MySQL on AWS will have immediate access to the new features that are developed and tailored for the AWS infrastructure to maximize performance and security. Some of those features of HeatWave MySQL on AWS are presented in the following sections.

> "HeatWave on AWS has 139X faster complex queries in comparison to AWS RDS and Aurora. This provides a significant opportunity to simplify the existing data infrastructure for both OLTP and OLAP, along with sub-second response time, to deliver an optimal experience."
>
> **Anish Kumar**
> Associate Vice President
> 6D Technologies

### MySQL Database

#### Tuned for Peak Performance

HeatWave MySQL on AWS is available on several shapes (i.e., AWS EC2 instance types) with different vCPU counts and memory sizes, on which the MySQL server can run for transaction processing. Customers can select a shape based on their workload needs. Each MySQL shape comes with a default MySQL configuration tailored for that shape to maximize performance on the target AWS infrastructure. Depending on the selected shape and storage size specified by the customer, the underlying storage system built on AWS EBS is tuned for peak performance without magnifying the associated costs.

#### High Availability

HeatWave MySQL on AWS supports high availability that enables applications to meet higher uptime requirements and zero data loss tolerance. A high availability DB system consists of three MySQL instances provisioned across different availability zones. The data is replicated among the instances using a Paxos-based consensus protocol implemented by the MySQL Group Replication technology. Applications connect to a single endpoint to read and write data to the database. In case of a failure, the service will automatically failover within minutes to a secondary instance without data loss and without requiring reconfiguring the application. High availability is currently available in limited availability (LA).

#### Advanced Security and Compliance Features

HeatWave MySQL on AWS includes several comprehensive security features natively implemented in the MySQL server, as opposed to other services such as Amazon Aurora, which provide security methods as an additional layer on top of the database.

1. **Data masking and de-identification**: Helps organizations protect sensitive data from unauthorized users by hiding and replacing real values with substitutes.
2. **Asymmetric encryption**: Helps developers and DBAs increase the protection of confidential data and comply with regulatory requirements including HIPAA, Sarbanes-Oxley, and the PCI Data Security Standard, through encryption, key generation, digital signatures, and other cryptographic features.
3. **Database firewall**: Helps provide real-time protection against database-specific attacks such as SQL injections by monitoring, alerting, and blocking unauthorized database activity without any changes in the application.
4. **MySQL Enterprise Audit**: Provides a robust and powerful auditing mechanism to help meet data governance, compliance, and security requirements such as FedRAMP, HIPAA, GDPR, and PCI-DSS. It helps database administrators define filters that specify which events and activities are collected. These database events provide details like who, what, when, where, how, and more. Additionally, optional query execution metrics can be included, which are useful for pinpointing issues, such as slow queries.

"HeatWave with HeatWave Autopilot on AWS is a gift from the database gods."

**Matt Kimball**
Senior Analyst
Moor Insights & Strategy

### JavaScript Support (GraalVM)

This feature supports JavaScript stored programs in HeatWave MySQL. It helps developers build rich procedural logic inside the database and access their MySQL datasets seamlessly. The JavaScript stored programs are executed in an environment running Oracle GraalVM Enterprise edition. GraalVM is an Oracle compiler ecosystem that includes JDK, language implementation such as JavaScript, R, Python, Ruby, and Java. It includes just-in-time (JIT) and head-of-time (AOT) compilation technology and provides a fully managed virtual machine with sandboxing capability and tooling support.

### Data Import from Amazon S3

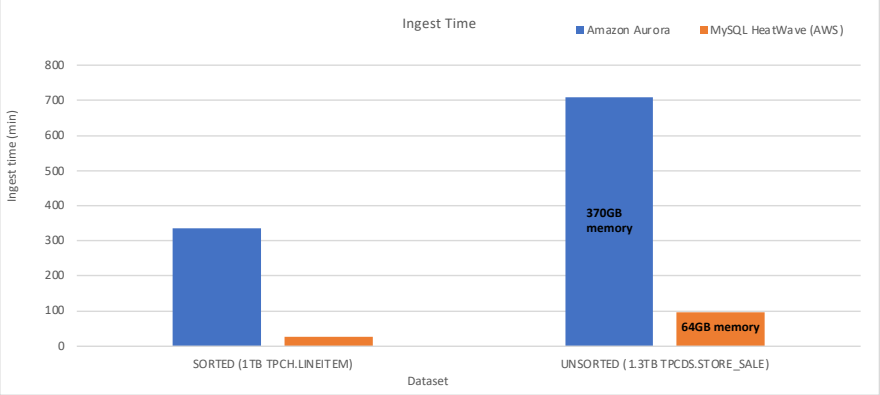
Data Import for HeatWave MySQL on AWS allows users to import data directly to HeatWave MySQL from Amazon S3. It provides a simple and intuitive user interface on the HeatWave AWS console. It supports data formats such as MySQL shell dump and delimited text files.

The data import capability provides enhanced security and better performance for importing data on Amazon S3 into HeatWave MySQL. It supports both AWS User Access Keys and AWS IAM Roles for authentication to Amazon S3. AWS IAM Roles grant exclusive access to user data on Amazon S3 to specific HeatWave instance(s). This helps provide tight security, which is better suited for enterprise production workloads.

Users can more easily migrate data to HeatWave MySQL on AWS from MySQL-compatible databases such as Amazon Aurora, RDS for MySQL, MySQL on AWS EC2 instance, or MySQL on-premises. Users also have the flexibility to export data from Amazon Redshift, Snowflake, or Google BigQuery in delimited text files, and then import to HeatWave MySQL.

### Native Bulk Ingest of Data Files in Amazon S3 into HeatWave on AWS

To provide a fast mechanism for importing large data sets from Amazon S3 to HeatWave MySQL, the MySQL native data import functionality has been enhanced to support the direct import of data files stored on Amazon S3 into HeatWave, and a new Bulk Ingest algorithm provides better data import performance with low memory usage.




Compared to Amazon Aurora, importing a 1TB dataset with data already sorted from Amazon S3, HeatWave on AWS is 12.4X faster. For data that is not sorted, HeatWave on AWS is 7.5X faster and uses 5.8X less memory than Aurora.

> "For cost conscious IT teams and developers, HeatWave on AWS represents a whole new TCO calculation with zero cost for what are add-on services on AWS and no data egress fees."
>
> **Marc Staimer**
> Senior Analyst
> Wikibon

### Inbound Replication

For organizations seeking to keep their existing transactional workloads either in the cloud or on-premises but wanting to use HeatWave for their dev/test, analytics, and machine learning use cases, they can easily set up inbound replication to replicate their existing databases to HeatWave on AWS.

### AWS PrivateLink Support

With the support of AWS PrivateLink, organizations can connect their applications to HeatWave through a private connection. All traffic to and from the HeatWave instance remains private, secured, and always stays within the AWS network. In addition, HeatWave on AWS uses AWS Identity and Access Management, which enables customers to control the creation of AWS PrivateLink VPC endpoints that are required to connect to HeatWave. This allows customers to have better control of their network configurations to HeatWave, as well as increased security.

## Analytical Processing

### Optimized and Tuned for Peak Performance and Best Price Performance

The HeatWave analytical processing engine is tuned to achieve the best performance on the AWS infrastructure. To minimize costs on AWS, HeatWave compresses the data in-memory, which helps reduce the memory footprint and the number of required processing nodes, without sacrificing performance.

### Support for Small Workloads

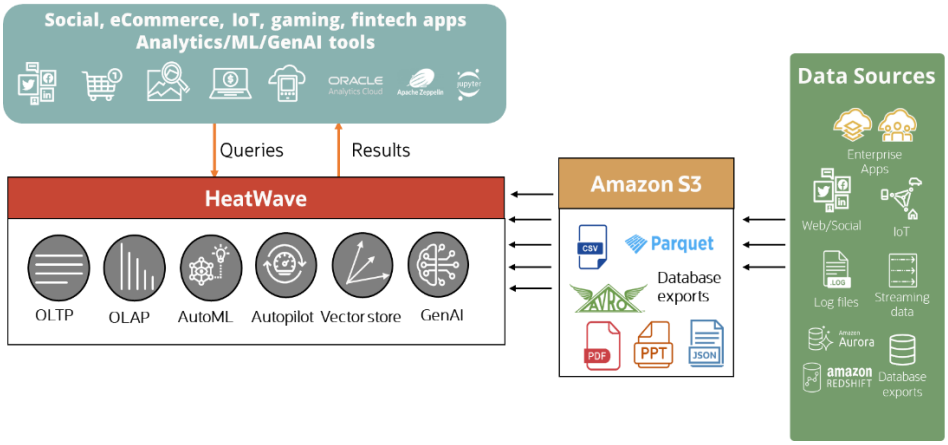
HeatWave has been designed to meet the needs of very large datasets by utilizing many processing nodes with large memory capacities. To allow OLTP databases with smaller datasets to benefit from the capabilities of the HeatWave

engine, HeatWave on AWS provides customers the option of selecting a cluster with a smaller HeatWave shape, enabling them to benefit from the high performance of HeatWave at a lower cost.

## HeatWave Lakehouse

HeatWave Lakehouse enables querying data in Amazon S3 in a variety of file formats, such as CSV, Parquet, Avro, JSON, as well as exports from databases (e.g., Aurora, Redshift, MySQL). Customers can query hundreds of terabytes of data in Amazon S3 and optionally combine it with transactional data in MySQL databases, without copying the data from Amazon S3 into the MySQL database. Users can perform machine learning (ML) tasks, like training, predictions, and explanations on this data. They can also create a vector store from the files in object storage and perform similarity search, enabling enterprises to bring the power generative AI to their enterprise content—without moving data to other services. Querying data in Amazon S3 is as fast as querying the data in the database. HeatWave Lakehouse scales out to 512 nodes.

> "For any developers working with MySQL on AWS, Oracle has just dropped a big productivity boost on your doorstep without the big price tag."
>
> **Carl Olofson**
> Research Vice President
> IDC

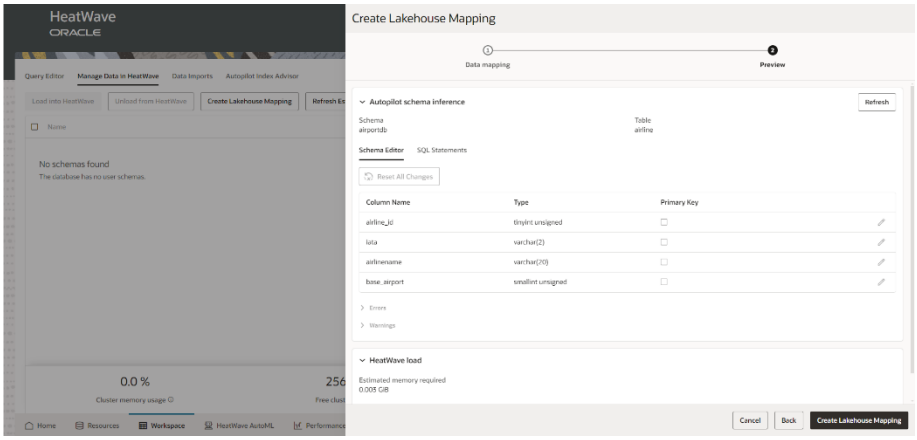



### HeatWave Lakehouse Advantages

1. Highly partitioned architecture to quickly process data from Amazon S3.
2. HeatWave Autopilot support for automatically inferring schema from raw semi-structured data in CSV and Parquet formats.
3. Full suite of in-database GenAI and AutoML capabilities
4. Secure access to customer data in Amazon S3 buckets using AWS IAM roles, where the customer has complete control over what data is shared with the HeatWave on AWS service.
5. Ability to provision large cluster sizes in minutes.

### HeatWave Lakehouse Console

HeatWave on AWS offers an interactive console to map data from Amazon S3 to a Lakehouse table in HeatWave (as data is not stored in the MySQL DB system). Users don't need to manually specify the schema, column name, data type and precision and other details of their data in Amazon S3. HeatWave Autopilot Auto schema inference automatically performs the mapping of data in S3 to data types in the database, saving significant time and effort.




Learn more.

## HeatWave AutoML

### Current Challenges of Machine Learning in Databases

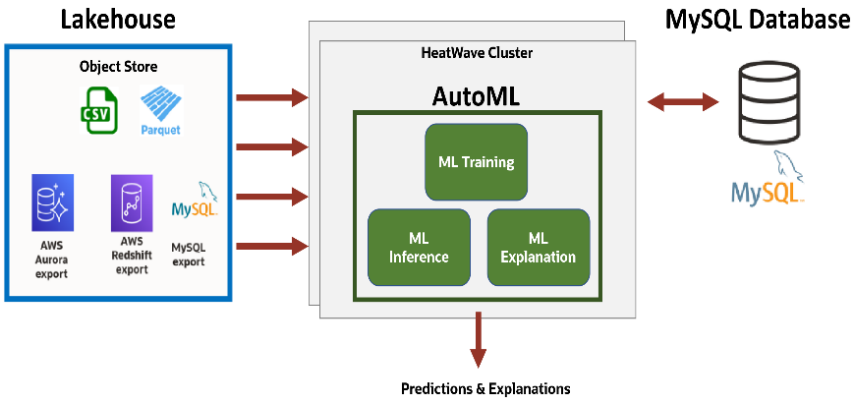
Developing and using machine-learning models requires skill sets such as:

- Candidate algorithms/models to select from
- Hyperparameters that need to be tuned per algorithm
- Features to engineer and select from
- Data preprocessing approach per data type
- Drift detection and retraining
- Knowledge of Python, as most ML algorithm frameworks are available only in Python

Even with the above expertise, users still need to extract data out of the database to train and test the model, which can lead to trust and security issues.

### HeatWave AutoML Approach

HeatWave AutoML enables users to train ML models and generate inferences and explanations across data stored in data lakes and MySQL databases. Machine learning training, inference, and explanation activities are performed inside the database, eliminating the need to extract data out of the database.




HeatWave AutoML is a native, in-database solution for data stored in MySQL or Amazon S3

### HeatWave AutoML Advantages

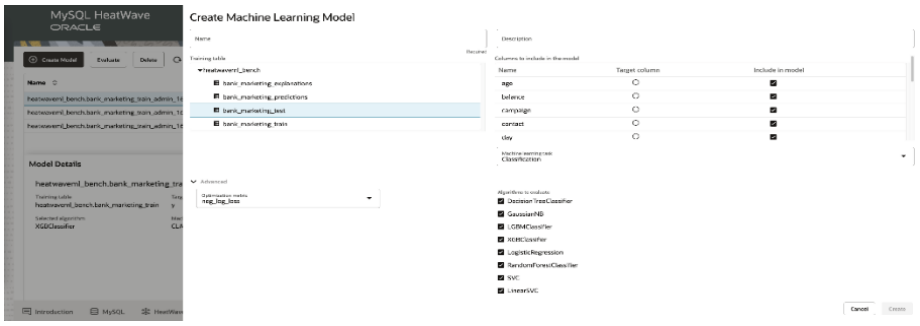
HeatWave AutoML enables users to train a model, generate inferences, and generate explanations, without extracting data out of the database. It provides several advantages:

1. **Fully Automated:** HeatWave AutoML fully automates the creation of tuned models, generating inferences and explanations, thus eliminating the need for the user to be an expert ML developer.
2. **SQL interface**: Provides the familiar MySQL interface for invoking machine learning capabilities.
3. **Security and Efficiency:** Data and models never leave the Database. Clients or any other services never see the data or models stored in the DB service.
4. **Explanations:** All models created by HeatWave AutoML can be explained. Enterprises have a growing need to explain the predictions of machine learning models to build trust, demonstrate fairness, and comply with regulatory requirements.
5. **Performance and Scalability:** The performance of HeatWave AutoML is much better, at a lower cost, than competing services such as Redshift ML. Furthermore, HeatWave AutoML scales with the size of the cluster.
6. **Easy Upgrades:** HeatWave AutoML leverages state-of-the-art open-source Python ML packages that enable continual and swift uptake of newer (and improved) versions.
7. **Supported Models**: HeatWave AutoML supports multiple model types such as Classification, Regression, Time Series Forecasting, Anomaly Detection, and Recommender System. This enables organizations to use HeatWave AutoML for different types of business use cases.

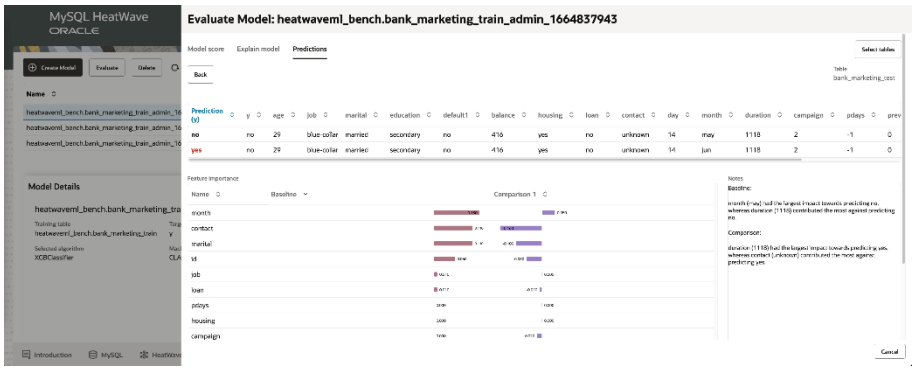
All these capabilities are available to HeatWave customers at no additional cost. The HeatWave console on AWS gives access to these HeatWave AutoML features through a user-friendly interface, further increasing ease of use.

### HeatWave AutoML Console

The HeatWave console on AWS allows customers to train machine learning models, to use the trained models for predictions, and to inspect the explanations for the models and predictions. Users can tune the models by adjusting the training parameters (such as algorithms and features to include), specify different scoring metrics for model or prediction evaluation, and evaluate what-if scenarios by manipulating different feature values to understand how they affect the model's predictions. The console enables business analysts to build, train, and run ML models without using SQL commands or any coding, simply using the visual interface.




HeatWave AutoML – create machine learning model

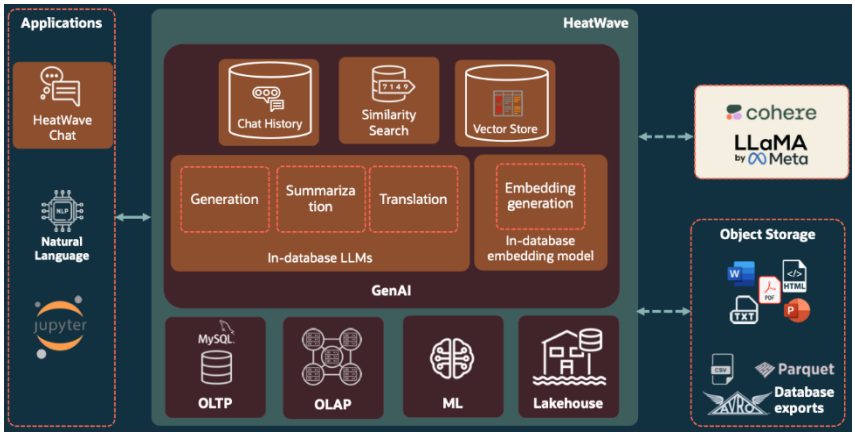



HeatWave AutoML – what if analysis

Learn more.

## HeatWave GenAI

HeatWave GenAI provides integrated, automated, and secure generative AI with the industry's first in-database large language models (LLMs); an automated, in-database vector store; the industry's best performance and price-performance vector processing (similarity search); and the ability for customers to interact with unstructured contents in HeatWave using natural language. HeatWave GenAI enables enterprises to take advantage of generative AI in a secure environment, without requiring AI expertise or the need to move data to other services.



### In-database LLMs

In-database LLMs simplify the development of generative AI applications—at a lower cost. Customers can benefit from generative AI without the complexity of external LLM selection and integration, and without worrying about the availability of LLMs in various cloud providers' data centers. The in-database LLMs enable customers to search data, generate or summarize content, and perform retrieval-augmented generation (RAG) with HeatWave Vector Store. In addition, they can combine generative AI with other built-in HeatWave capabilities such as AutoML to build richer applications. HeatWave GenAI is also integrated with Amazon Bedrock foundation models such as Cohere and Llama.

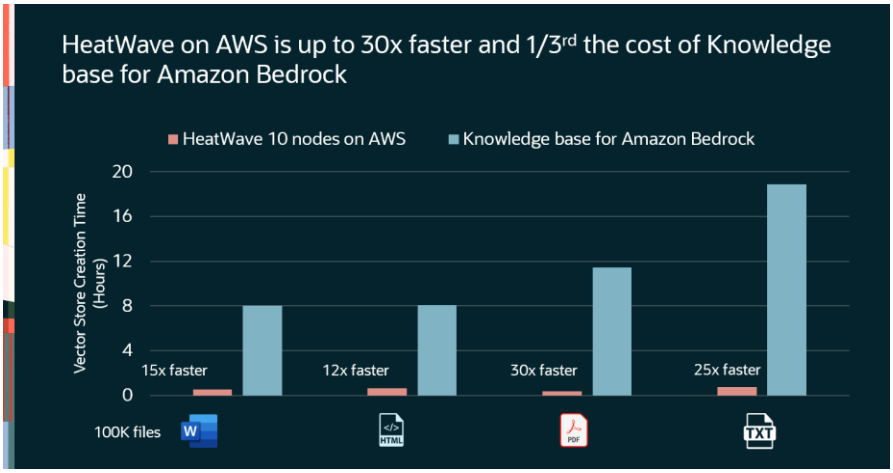
### Automated, In-database Vector Store

The automated, In-database HeatWave Vector Store enables customers to use generative AI with their business documents without moving data to a separate vector database and without AI expertise. All the steps to create a vector store and vector embeddings are automated and executed inside the database, including discovering the documents in object storage, parsing them, generating embeddings in a highly parallel and optimized way, and inserting them into the vector store, making HeatWave Vector Store efficient and easy to use. Using a vector store for RAG helps solve the hallucination challenge of LLMs as the models can search proprietary data with appropriate context to provide more accurate and relevant answers.

### Scale-out Vector Processing

Scale-out vector processing delivers very fast semantic search results without any loss of accuracy. HeatWave supports a new, native VECTOR data type and an optimized implementation of the distance function**,** enabling customers to perform semantic queries with standard SQL. In-memory hybrid columnar representation and the scale-out architecture of HeatWave enable vector processing to execute at near-memory bandwidth and parallelize across up to 512 HeatWave nodes. As a result, customers get their questions answered rapidly. Users can also combine semantic search with other SQL operators to, for example, join several tables with different documents and perform similarity searches across all documents.

For similarly search, HeatWave GenAI provides 39X better price-performance than snowflake, 96X better than Databricks and 48X better than Google Big Query.

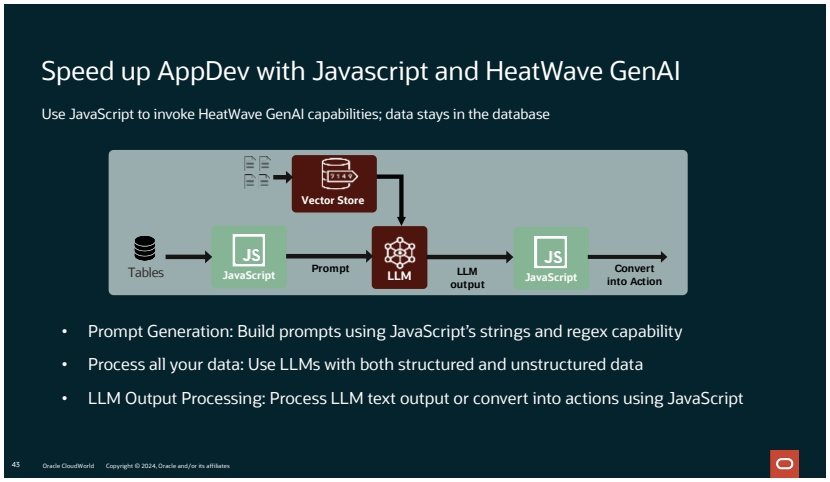



### HeatWave Chat

HeatWave Chat is a Visual Studio code plug-in for MySQL shell which provides a graphical interface for HeatWave GenAI. Developers can ask questions in natural language or use SQL. The integrated Lakehouse Navigator enables users to select files from object storage and create a vector store. Users can search across the entire database or restrict the search to a folder. HeatWave maintains context with the history of questions asked, citations of the source documents, and the prompt to the LLM. This facilitates a contextual conversation and allows users to verify the source of answers generated by the LLM. This context is maintained in HeatWave and is available to any application using HeatWave.

### In-database JavaScript Support

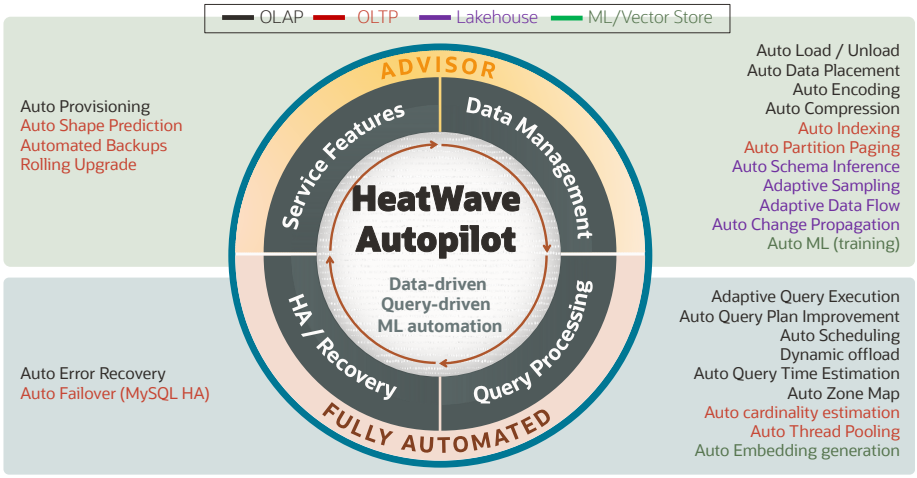
In-database JavaScript support for GenAI adds native support for the vector data type in JavaScript and the ability to invoke GenAI capabilities from a JavaScript program. As generative AI and LLMs primarily handle textual and JSON data, this allows developers to more easily pre-process prompts based on SQL data, invoke LLMs directly within the database, and post-process the responses in JavaScript.




Learn more.

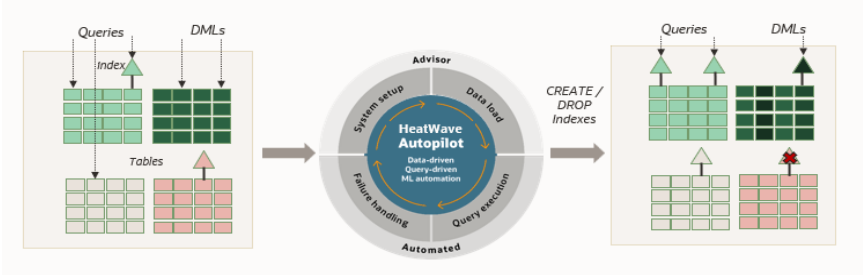
## HeatWave Autopilot

HeatWave Autopilot automates many of the most important and often challenging aspects of achieving high query performance at scale - including system setup, data loading, query execution and failure handling. It uses advanced techniques to sample data, collect statistics on data and queries, and build machine learning models to model memory usage, network load and execution time. HeatWave Autopilot uses these ML models to intelligently learn from queries executed in the system, resulting in continually improving system performance over time. HeatWave Autopilot also provides capabilities to improve the performance and price-performance of OLTP workloads.



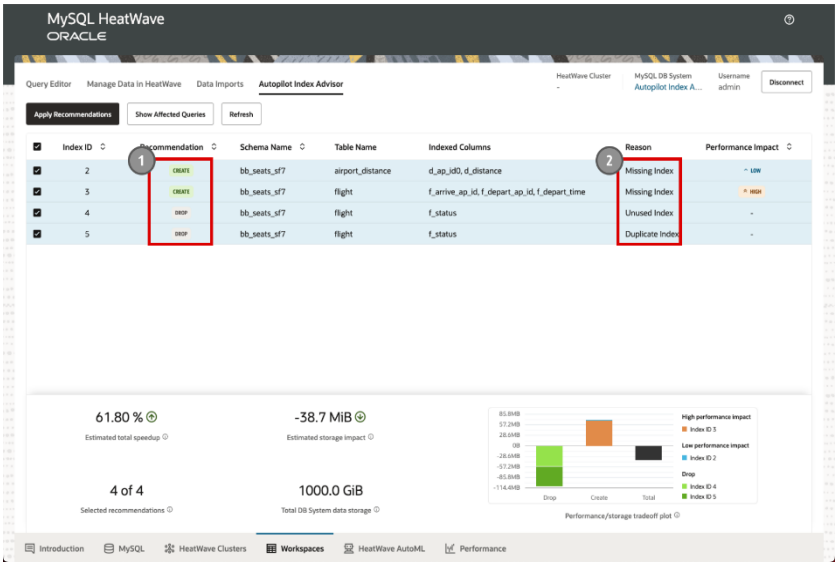

### HeatWave Autopilot Features for OLTP

1. **Autopilot Indexing**: HeatWave Autopilot Indexing recommends the right set of indexes for columns to improve OLTP query performance. It balances cost, storage space, and performance by adding or removing indexes.

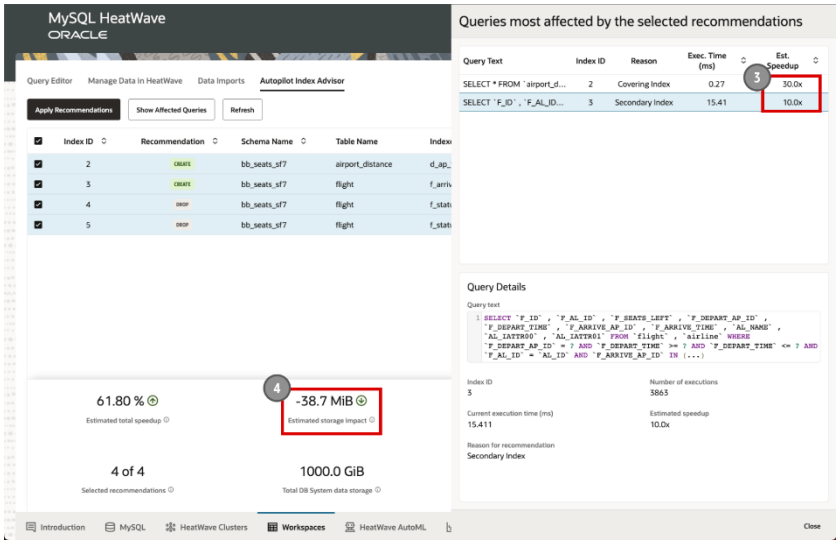



With Autopilot Indexing, database administrators no longer need to manually identify which indexes are most beneficial for their workload. Autopilot Indexing automatically generates secondary index recommendations to create or drop indexes based on the current workload. Autopilot Indexing considers both the query performance and the cost of maintaining the indexes when generating recommendations. It provides performance and storage estimations, as well as explanations for the recommendations it generates. The Autopilot Indexing interface consists of a simple and intuitive console that customers can use to view and analyze the projected performance and storage impact of recommended index

suggestions. This makes it easy to foresee the impact of changes to the database systems before applying the suggestions.




Example of Autopilot Indexing showing (i) CREATE and DROP index suggestions, (ii) explanations for the recommendations.

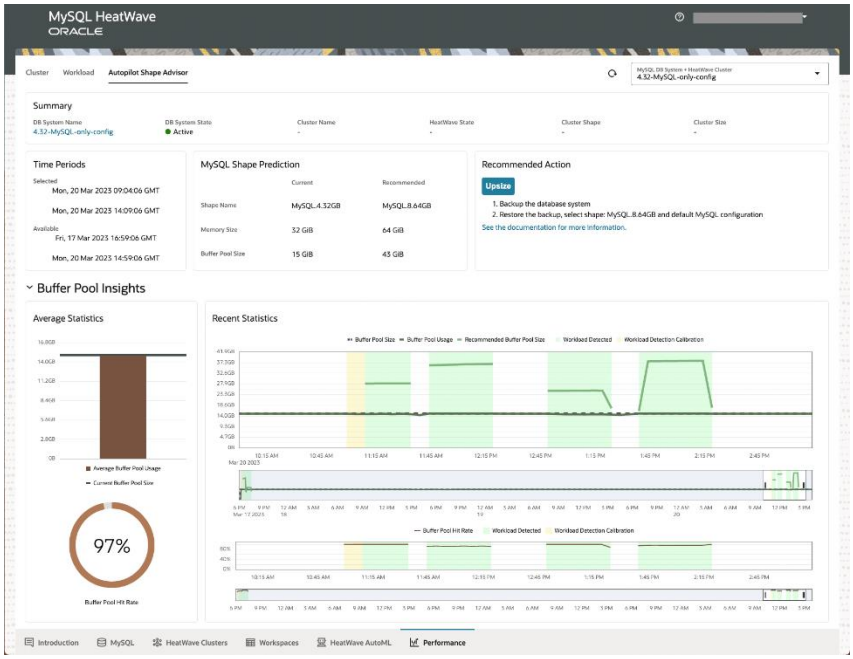



Example of Autopilot Indexing showing (iii) estimated performance impact, and (iv) estimated storage impact from the recommendations.

2. **Auto Shape Prediction**: To alleviate the burden of experimenting with different MySQL shapes to determine the most performant one for a given workload, Auto Shape Prediction recommends the optimal MySQL server shape, based on highly accurate predictions from machine-learning models inside the MySQL server, along with the most recent query execution metrics and traces.

   HeatWave Autopilot Auto shape prediction continuously monitors OLTP workloads to provide suggestions that adapt to evolving workload patterns, allowing customers to maintain the best OLTP price-performance over time. It also shows trends on the workload with metrics such as buffer pool hit ratio and buffer pool utilization factor, providing insights for customers to better

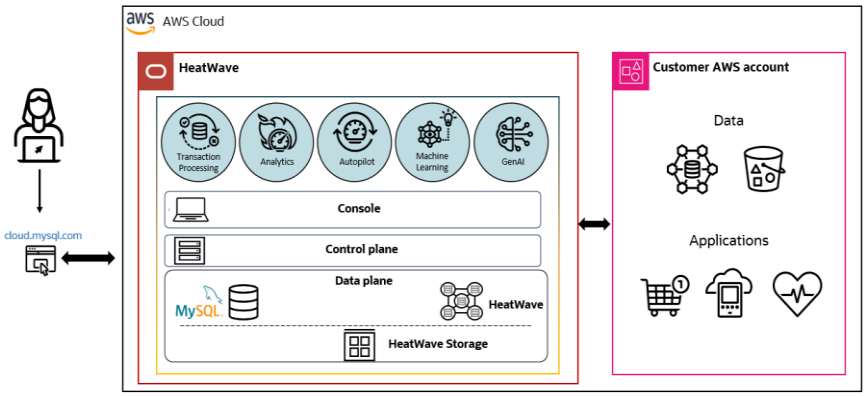
understand their workloads. The visual representation within the console makes it easy for database users to upsize or downsize their database shape to optimize price-performance.




3. **Auto Thread Pooling**: With Auto Thread Pooling, HeatWave MySQL prioritizes not only peak single-thread performance, but also high throughput in the presence of concurrent clients running concurrent queries on a MySQL server. With this feature, the MySQL server can perform workload-aware admission control of the incoming transactions. It eliminates the resource contention created by too many awaiting transactions, automatically queuing them to maximize performance while sustaining the throughput in the face of high concurrency.

## Architecture of HeatWave on AWS

HeatWave on AWS delivers a native experience for AWS customers. The console, control plane, and data plane reside in AWS and are responsible for managing the HeatWave database resources in AWS. The control plane communicates with Oracle Cloud Infrastructure (OCI) Identity for account management, and with OCI metering & billing for monitoring and managing the usage and expenses associated with the customer's account.



Once the user signs up for an OCI cloud account and registers their OCI account with HeatWave on AWS, the main interactions with the HeatWave service take place in AWS, through the service console hosted at cloud.mysql.com.

### HeatWave Control Plane

The HeatWave control plane, which enables the management of HeatWave and maintains the necessary metadata, is built on publicly available AWS services. The control plane components are hosted in the Oracle AWS account dedicated to the HeatWave service and are tightly controlled with AWS Identity and Access Management permissions and policies.

The HeatWave control plane is responsible for the management of the database system lifecycle, including for provisioning/de-provisioning/pausing, configuring the database system, orchestrating backups, security patching, upgrades, monitoring, as well as helping to isolate the different database systems from each other.

### HeatWave Data Plane

HeatWave on AWS hosts all the customer databases' components in a dedicated AWS account and strictly isolates them from the service control plane components and other database systems managed by the control plane.

HeatWave databases are hosted on AWS services, which are publicly available to AWS customers, such as Elastic Compute Cloud (EC2) for data processing, Elastic Block Storage (EBS) for storage, and VPC for resource isolation. Customers do not have access to the infrastructure on which the database runs, such as the database host machines or storage; they are provided with a MySQL endpoint, which can be accessed by a standard MySQL client. Customers can restrict the client addresses that can access their database systems through the MySQL endpoint and the connection between the client and database system is secured with TLSv1.2. Using the MySQL endpoint, customers can run any application against their HeatWave database systems.

### HeatWave Console

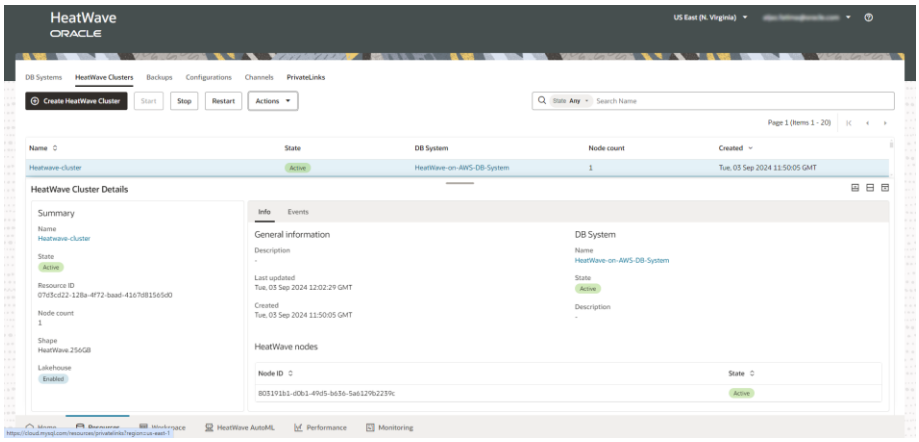
The HeatWave console is designed to facilitate:

1. Lifecycle management of the HeatWave resources, such as DB systems, HeatWave clusters, backups, configurations, channels, and PrivateLinks.
2. Sample data, schema and queries, and one-click starter database system for quick start with HeatWave on AWS.
3. HeatWave data management including querying data in the DB system and Amazon S3, importing data from other databases such as Amazon Aurora and RDS, and the relevant HeatWave Autopilot capabilities such as Autopilot index advisor.
4. Management and visualization of training, inference, and explanation of machine learning algorithms offered by HeatWave AutoML.

5. Monitoring HeatWave, MySQL, and workload performance through visualization of the schema metadata and statistics that are stored on the MySQL server.

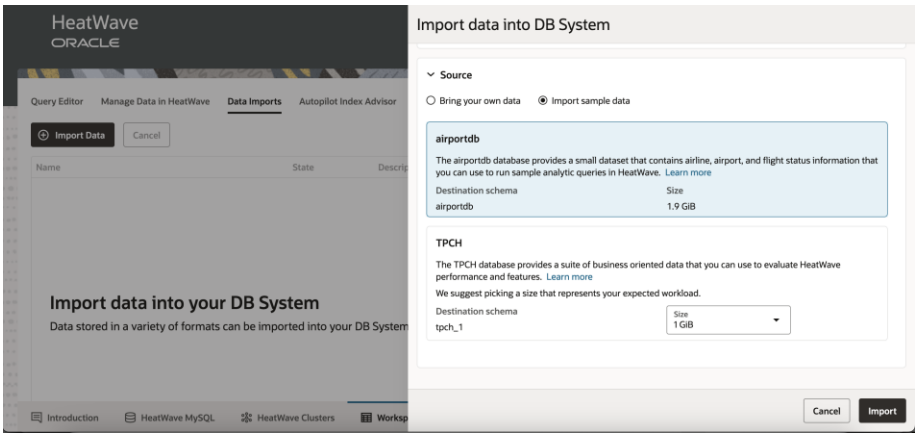
### Lifecycle Management

Customers can manage the HeatWave resources associated with their accounts through the HeatWave console. The HeatWave console provides a unified interface to manage database systems, HeatWave clusters attached to the database systems, configurations, backups, and networking setup (such as inbound replication channels and PrivateLink) of the database systems. The service console allows users to, for example, provision/deprovision/configure resources, pause a database system, and pause the HeatWave cluster attached to it if no user activity is expected to save costs.

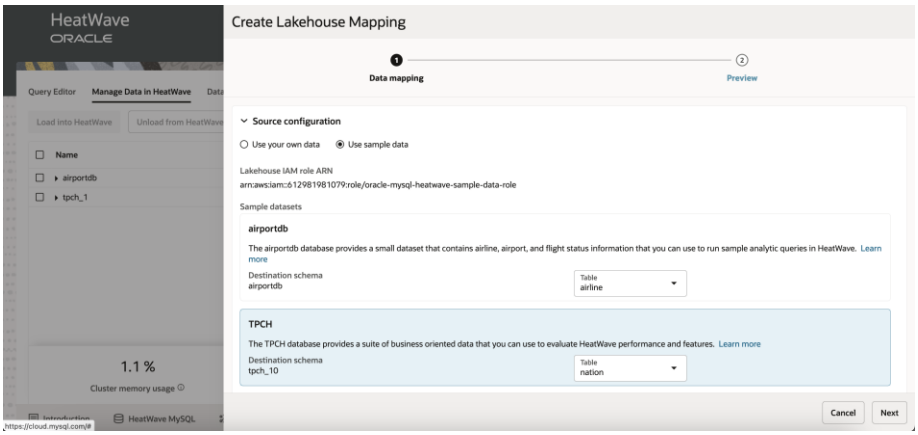



### Sample data, Schema and Queries

HeatWave provides sample data sets, such as the industry-standard TPC-H benchmarks and open-sourced dataset airportdb, for evaluating and testing a broad range of HeatWave's SQL support. Users can ingest the sample data through HeatWave Lakehouse and/or HeatWave MySQL.

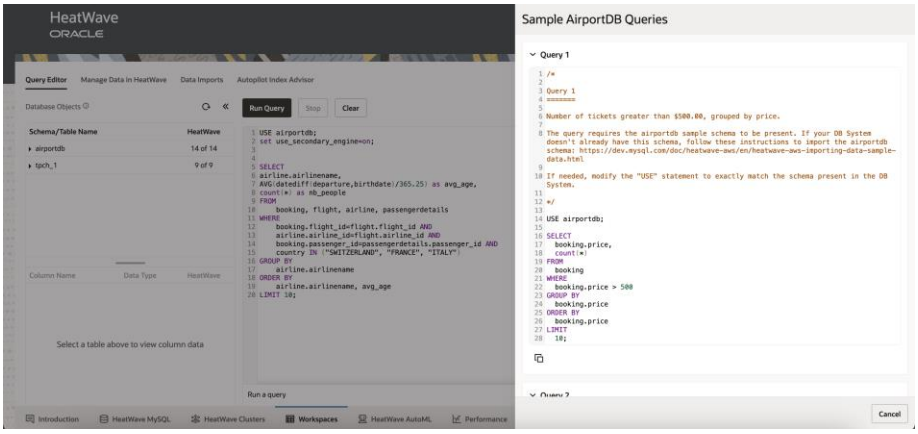



Sample data set for HeatWave MySQL




Sample data set for HeatWave Lakehouse

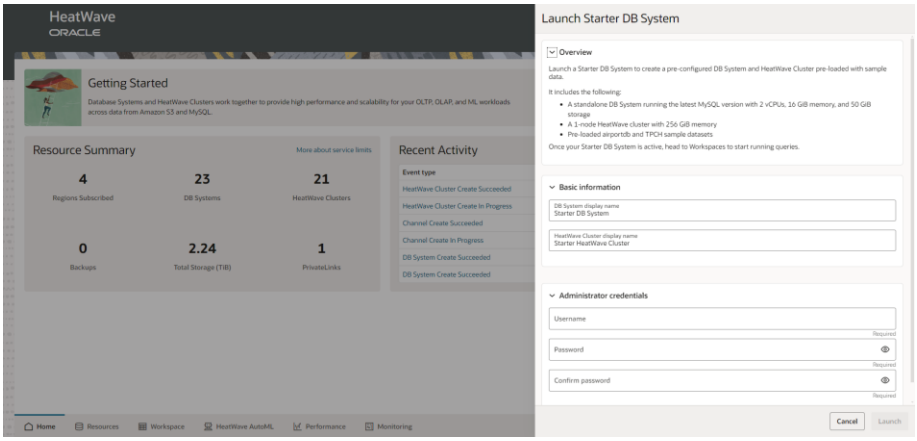
Once the data is loaded into HeatWave, the Query Editor on the console provides samples queries for the sample datasets, allowing users to test the performance of HeatWave.




Sample queries in Query Editor

## One-click Starter Database System

A one-click starter db system is available. This is a pre-configured DB System with preloaded sample data to explore HeatWave on AWS.

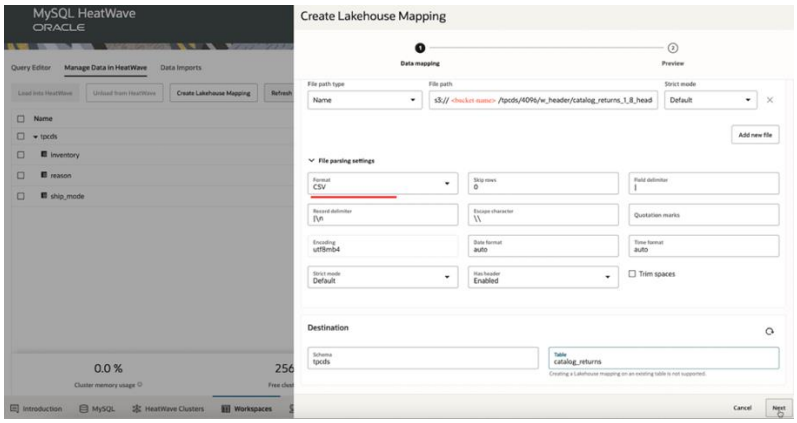


### Interactive Schema and Data Management

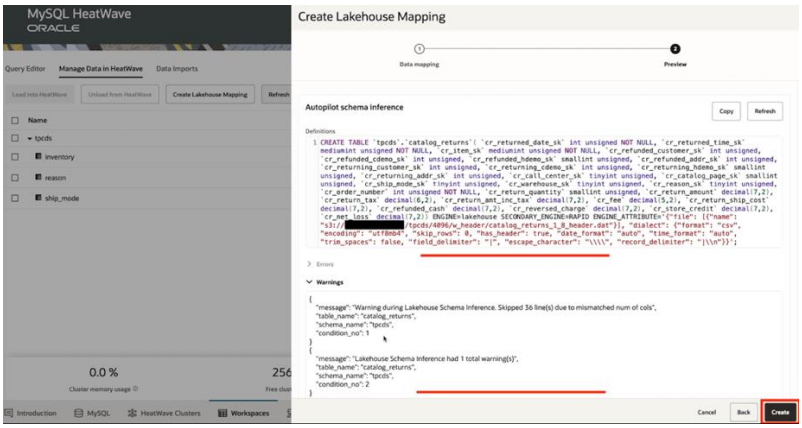
HeatWave on AWS can process data in MySQL or Amazon S3. The console provides intuitive user interfaces for users to load and manage data in HeatWave.

1. Manage data in Amazon S3 using HeatWave Lakehouse

   To process data in Amazon S3, HeatWave Lakehouse needs to first map the data to a table, as data is not stored in the MySQL database system. HeatWave Lakehouse uses HeatWave Autopilot auto schema inference to automatically infer the mapping of file data to data types in the database. As a result, customers do not need to manually specify the mapping for each new file to be queried by HeatWave Lakehouse, saving time and effort.



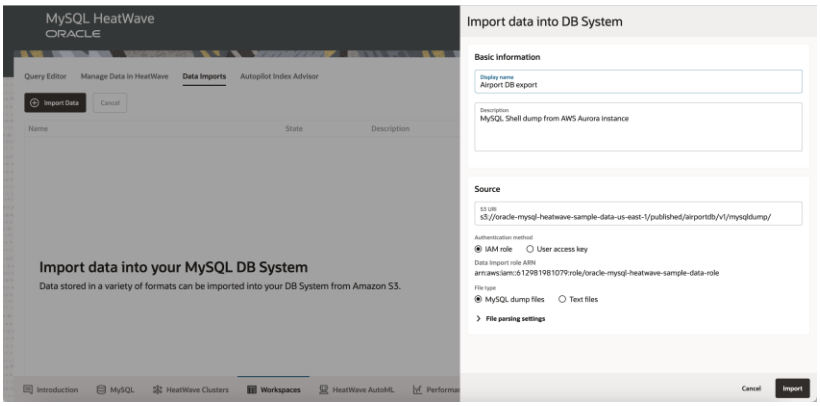

Create Lakehouse mapping using the console




HeatWave Autopilot – Auto Schema Inference suggests the schema with column name mapping, data type and precision.

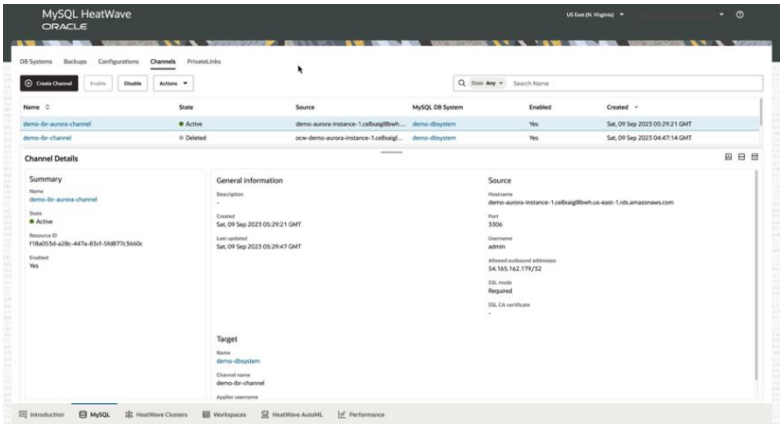
2. Import data on Amazon S3 to HeatWave

   Data import provides a simple way to import data to HeatWave database systems such as MySQL dump or CSV files on Amazon S3 using the HeatWave on AWS console.

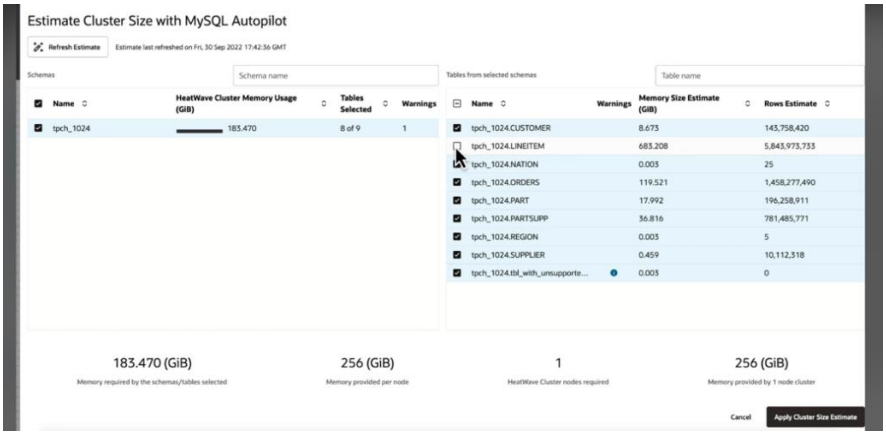



3. Replicate data from a source MySQL database using inbound replication

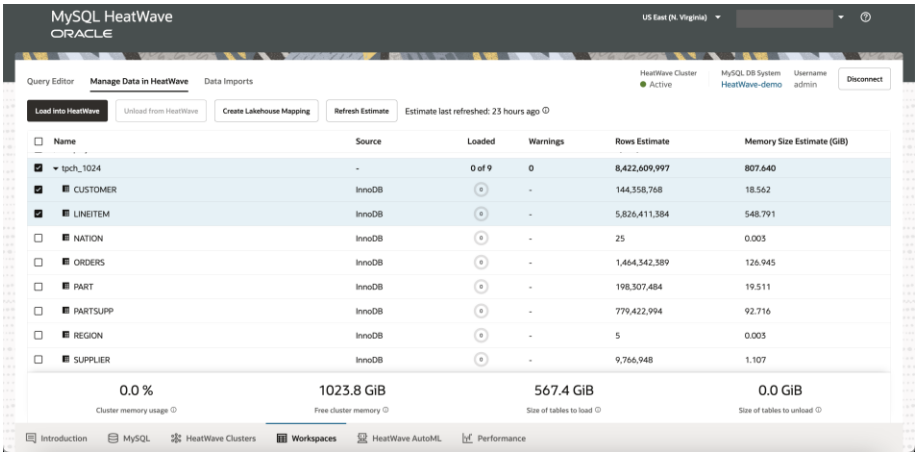
   Customers can set up an inbound replication channel to replicate data from their existing MySQL deployment, including Amazon RDS/Aurora, to HeatWave MySQL. The console provides a simple interface to set up the replication channel, which internally uses MySQL native replication.



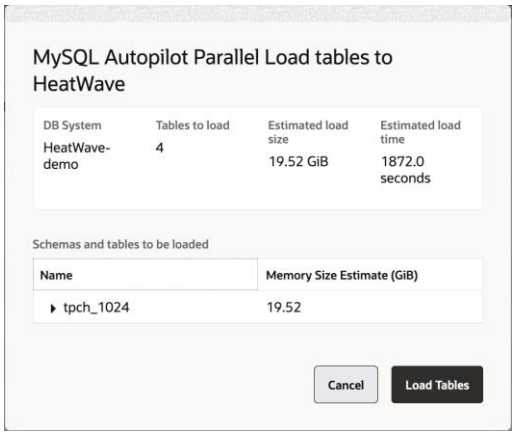

Once the data mapping is set up on HeatWave Lakehouse or data is imported or replicated into the MySQL server, customers can leverage the HeatWave Autopilot Auto Provisioning capability to estimate the HeatWave cluster size needed for their workload.



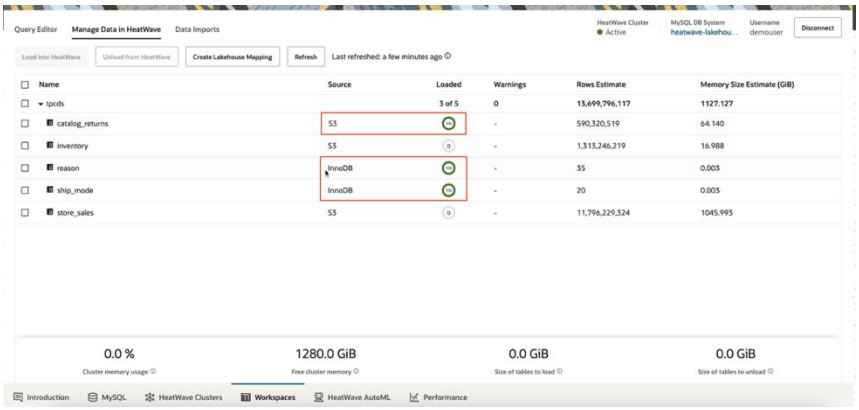
The HeatWave console provides users with a list of schemas and tables in the MySQL server. Depending on which tables the user wants to run analytical queries on, Auto Provisioning estimates the total memory usage in the HeatWave cluster memory. Based on the HeatWave shape selected by the user, it estimates the number of HeatWave cluster nodes needed to accommodate the target dataset. Users can then provision a HeatWave cluster based on the number of nodes suggested by Auto Provisioning and load the desired tables to the HeatWave memory for analytical processing.




Customers can continuously monitor the schemas and tables loaded in the MySQL server as well as the HeatWave memory. This monitoring information helps loading the desired tables into the HeatWave cluster memory in the most efficient way. The user first needs to select which tables to load into the HeatWave memory and then use Auto Parallel Load, which optimizes the load time and memory usage of the data load operation into HeatWave by predicting the optimal degree of parallelism for the set of tables selected by the user. To provide better visibility into storage and memory usage, the HeatWave console also provides detailed information about the estimated memory footprint of each table in the HeatWave cluster memory, encoding type, load status, as well as the predicted load time provided by HeatWave Autopilot.

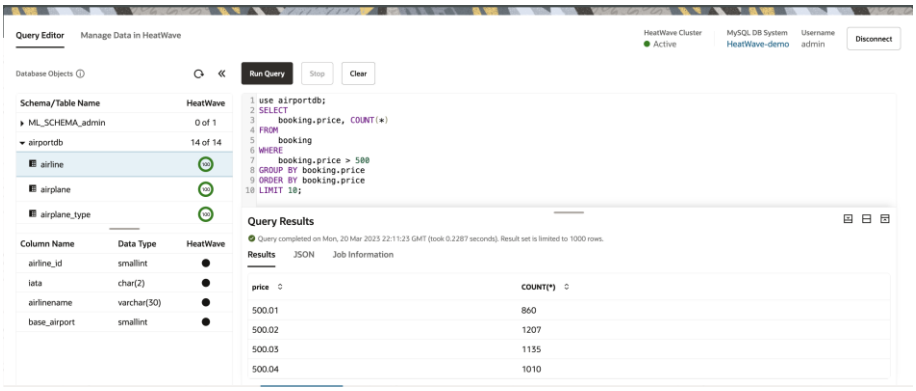


Once the desired tables are loaded into the HeatWave cluster memory, customers can monitor the actual breakdown of the memory consumption with the detailed breakdown of each table's in-memory footprint.

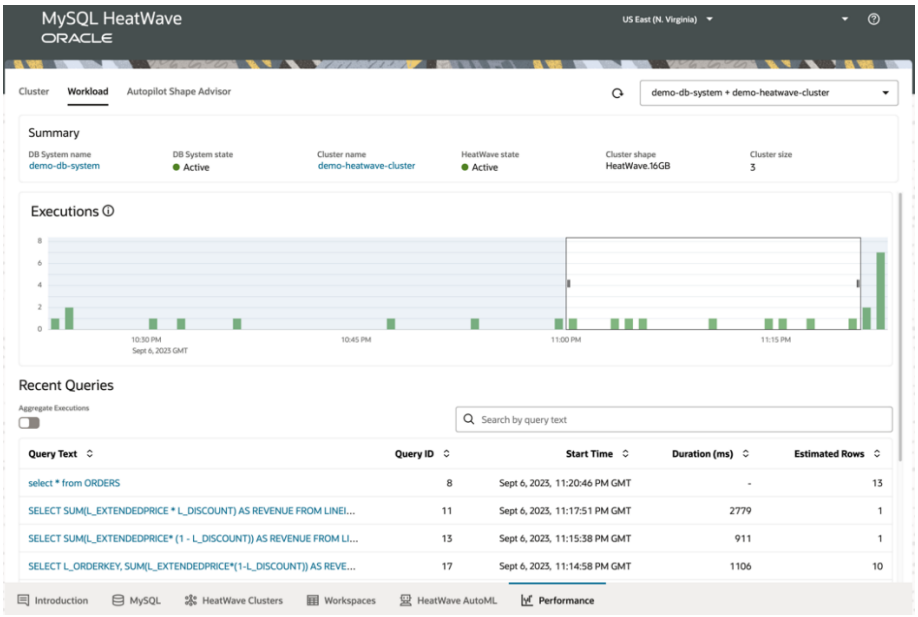



### Interactive Query Interface & Workload Monitoring

The HeatWave console provides a query editor to ease the customer's interaction with their database system by eliminating the need to go back and forth between the console and an external MySQL client for resource and data management. Customers can write and execute queries on HeatWave and view the query results through the query editor, while monitoring the state of the database, all in the same console.

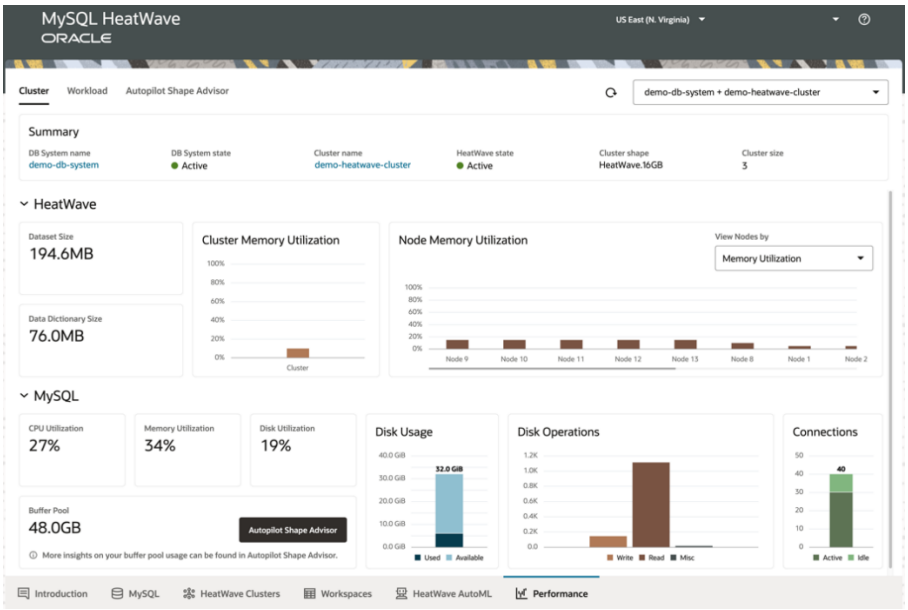



In addition, HeatWave workload monitoring summarizes the results of the most recent queries executed along with their execution time and the number of rows in the result set, allowing the customers to have access to the recent history of activities in the HeatWave database from the console.



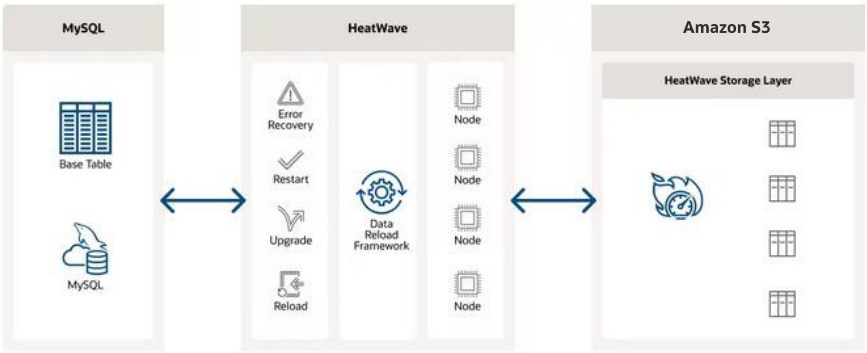

## Performance Monitoring

The HeatWave console allows users to monitor the overall and "per-node" utilization of HeatWave hardware resources such as CPU, memory, and storage. It also provides a detailed breakdown of your resource consumption, such as data dictionary size, buffer pool size, and database connections.




## HeatWave Scale-Out Data Management

HeatWave on AWS provides an optimized storage layer built on Amazon S3 to store the HeatWave in-memory hybrid columnar representation of the data. This allows data to be reloaded to each HeatWave node independently and in parallel, significantly improving the service uptime and the performance of operations such as error recovery, maintenance, and system restart.




The fast and automatic data reload from the HeatWave storage layer also enables customers to easily and quickly pause and resume a HeatWave cluster to save cost when HeatWave is not needed.

## Integration with AWS Services

HeatWave on AWS integrates natively with various AWS services to provide seamless integration with AWS applications.

### AWS Identity and Access Management (IAM)

HeatWave on AWS integrates with AWS Identity and Access Management, which enables HeatWave instances to securely access data or resources in customers' AWS accounts. It uses the cross-account roles mechanism provided by the AWS IAM service. With this mechanism, customers can delegate S3 access permission to a HeatWave instance, providing tighter security for enterprise production workloads.

### Amazon S3

Integration with Amazon S3 enables HeatWave instances to read or write data to S3 directly. It supports multiple use cases:

1. HeatWave Lakehouse – The integration with Amazon S3 enables HeatWave Lakehouse to securely access customers' data files stored in Amazon S3, providing fast analytics query processing and machine learning capability directly on data stored in Amazon S3.
2. Data import from Amazon S3 – This enables users to directly import data stored in Amazon S3 to HeatWave. Data import provides a simple way to import data to HeatWave database systems such as MySQL dump or CSV files on Amazon S3 using the HeatWave on AWS console.

### Amazon Bedrock

With HeatWave GenAI, users have the flexibility to choose to use HeatWave in-database LLMs or Amazon Bedrock foundation models such as Cohere and Llama models. To use Amazon Bedrock, there will be a separate charge by AWS for accessing the Amazon Bedrock foundation models.

## Conclusion

HeatWave is the only fully managed cloud service that provides integrated, automated, and secure generative AI and machine learning (ML) in one cloud service for transactions and lakehouse scale analytics—without the complexity, latency, risks, and cost of ETL duplication.

Customers don't have to face data egress fees charged by AWS and higher latency when accessing a database service running in Oracle's cloud. HeatWave on AWS is optimized for AWS with a superior architecture that delivers unmatched performance and price-performance at a lower cost, as demonstrated by industry-standard benchmarks.

Try HeatWave on AWS for free now!


**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