100x Acceleration on Analytics and Machine Learning with HeatWave on Transactional Data Outside of MySQL databases

Mandy Pang
Senior Principal Product Manager
MySQL HeatWave
Agenda

• MySQL HeatWave overview
• Use Cases
MySQL HeatWave overview

Transactions, real-time analytics, machine learning and GenAI across data warehouse and data lake in one service.


MySQL HeatWave

Queries → Results

MySQL storage

Object Store

For both non-MySQL and MySQL workloads

OLTP
Analytics
In-database ML
MySQL Autopilot
Vector store*
GenAI*

* Coming soon

Data Sources

Enterprise Apps
Web/Social
IoT
Log files
Streaming data

Database exports

CSV
Parquet
AVRO
AVRO
PDF
CSV
PPT
TXT

Database exports

amazon REDSHIFT
Aurora

MySQL Autopilot

Queries

Results

* Coming soon
Best performance in industry for data warehouse

FASTER TIME TO INSIGHTS = FASTER BUSINESS RESPONSE TO MARKET TRENDS

* 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.

* Results from March 2023
Lowest cost in industry for data warehouse

**PRICE PERFORMANCE COMPARISON FOR 10TB TPC-H**

- **HeatWave (10 nodes)**: 23X
- **Redshift (10 * ra3.4xlarge)**: 27X
- **Snowflake (X-Large)**: 27X
- **Google BigQuery (800 slots)**: 61X
- **Databricks (Large)**: 61X

*Only compute costs are considered above*

- Pricing for Redshift is based on 1-year reserved instance, paid upfront. Snowflake is based on **standard edition**
- Pricing for Google Big Query is based on monthly flat rate commitment. Databricks 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
* Results from Sept 2022
HeatWave AutoML: In-database machine learning

- Eliminates tedious and laborious steps
- Simple to use interface for beginner or advanced ML users
- Automatically selects algorithm and tunes it
- Explainable model behavior and predictions
- Fast training allows to quickly iterate to achieve desired outcome

In-database ML

- Model training
- Model inference
- Model explanations

Preprocessing -> Algorithm Selection -> Adaptive Sampling -> Hyperparameter Optimization -> Model Explainer -> Prediction Explainer -> Tuned Model

InnoDB

MySQL

AWS Aurora export
AWS Redshift export
MySQL export
Parquet
CSV
HeatWave AutoML use-cases

**Classification**
- Player churn prediction
- Classify warranty claims

**Anomaly Detection**
- Defective part identification
- Identify game hackers
- Predict when failure will occur
- IoT digital twin failure prediction
- Predict air pollution
- Return on advertising spend prediction
- Utilization demand forecasting

**Regression**
- Loan default prediction
- Demand forecasting
- Rain fall amount prediction

**Recommender System**
- Identify similar users
- Recommend movies to viewers
- Suggest substitute products
- Recommend new products
- Timeseries Forecasting
Industries and use cases with HeatWave AutoML

**Digital Marketing**
- Cost per acquisition
- Targeted campaigns
- Customer classification

**E-Commerce**
- Videos for users
- Lottery suggestions
- Product upsell

**Education**
- Predict student success
- Monitor student behavior
- HIPPA Compliance

**Services**
- Erroneous ledge entries
- Predict future losses
- Predict price elasticity

**FinTech**
- Loan default prediction
- Identify loan extensions
- Loan approval

**Gaming**
- Player churn detection
- Adjust game difficulty
- Identify game hackers

**Internet Of Things**
- Airport ticketing
- Rain water level
- Air pollution

**Manufacturing**
- Reduce warranty claims
- Defective part identification
- Detect anomalies in supplies
Machine learning with HeatWave is fast, cost effective, accurate and scalable.

- 25x faster than Redshift
- 1% of the cost of Redshift
Customers

Success Stories on MySQL workload
Easily run analytics/ML against on-premises MySQL databases

No other cloud vendor provides this capability
Estuda.com achieves real-time insights

“MySQL HeatWave improved our complex query performance 300X for responses in seconds and at 85% of the cost compared to Google BigQuery with no code changes. Now we can better deliver real-time analytics at a scale of 3 million users and continually improve our application to enhance student performance.”

Vitor Freitas
CTO, Estuda.com

Business Challenge:
Brasil’s leading ed-tech serves over 8 million students from more than 500 K-12 schools to enhance student performance. It needed a data platform to deliver real-time insights by reducing ETL complexity and costs in moving data from AWS RDS to Google BigQuery to scale for 3 million users per month.

Results:
- 300X faster performance from migrating from BigQuery to MySQL HeatWave with no code changes and low-latency
- 85% cost reduction by eliminating ETL processes and pay-for-use consumption model
- Real-time analytics enable faster development to improve app usability and adoption
- Scales queries to any data size for more flexibility growth to impact more students

Products Used:
MySQL HeatWave

Read story
Fintech company (MySQL mixed workload from AWS)

Replicate from Percona MySQL to MySQL HeatWave for analytics

**Company**: leading NBFC, process 30K loan a day, loan tickets size is 5K to 500k, serving 28000 pin code in India.

**Use Case**: The application and databases are hosted on AWS Environment. They use multiple Percona MySQL instances running from AWS EC2 Instances with Read Replica for reporting and data sharing for different business use cases.

**Challenges**: Consolidation of data from multiple MySQL deployments for high query performance for reporting (Total 30TB of data)

**Solution:**
1. Hybrid solution – data consolidation by replicating multiple Percona MySQL deployments into single MySQL HeatWave

**Customer chose MySQL HeatWave:**
1. **845 X** better query performance
2. No need of ETL tools to move data from the MySQL database
3. Real-time insights to better analyze and understand customer behavior to continuously improve its application with rapid development
4. Reduced TCO; compared to AWS costs
5. Uses the native analytics capabilities of MySQL
6. Enhanced data security and ensured regulatory compliance (MySQL EE and OCI security)
Ebook company (MySQL workload –> ETL –> Teradata)

Migrated data warehouse from Teradata, now expanded to use HeatWave Lakehouse

**Company:** Provides e-book service and game service in Japan, has 35 million (30% of Japan population) unique active subscribers.

**Use Case:** For their e-book business,
1. MySQL Enterprise Edition for OTLP
2. Teradata as data warehouse (10TB of data)

**Challenges:** They were starting to migrate to Google BigQuery

**Solution:**
1. Hybrid solution – replicate data from on-premise MySQL to MySQL HeatWave for data warehousing
2. Looker as the BI tool

**Customer chose MySQL HeatWave:**
1. They are already familiar with MySQL
2. No change in applications and existing tools that support MySQL
3. Provides real-time data analytics via MySQL replication, without using ETL tool
4. High query performance. “Never-ending query" in MySQL ow runs in a few seconds
5. Predictable pricing model – compared to Google BigQuery (they were using on-demand pricing)
Easily run analytics/ML for non-MySQL workload
Logistics company (non MySQL workload)

*MySQL HeatWave for fast dashboard/reporting along with ATP for OLTP*

**Company:** global supply-chain services to help enable sustainable trade and commerce in key markets

**Use Case:** Track and Trace about the Cargo shipments and Business decision reports in their Visibility and Reporting application running on on-premise Oracle database by both Internal/External Stake holders.

**Challenges:** application is a heavy data processing and integration-oriented platform. SQL queries are taking longer time and users are experiencing slower performance with the data growth

**Customer Looking to:** Modernize the VNR application by on multiple data stores (approx. 5-6 transaction database sources) for scalability and *near real-time data with interactive dashboard* for visibility and reporting

**Data Size:**
1. Total ~3TB data, 1TB (1-2 year) of data for dashboard
2. ~2GB/day data growth
3. Concurrent users ~200

**Solution:**
1. Replicate data from on-premise Oracle Database using GoldenGate
2. Use ATP for OLTP
3. MySQL for reporting and interactive dashboard

**Results:**
1. Up to 3,000X faster than on-premise deployment
HeatWave AutoML – Customer momentum
Company: Provides Automation of ITSM & GRC in an Integrated Platform

Use Cases:
1. Pre-configured processes and workflows eliminating spreadsheets and manual work
2. Maximum visibility and data insights allows users to correlate, analyze, and remediate issues
3. Flexible Platform that can scale and simplify existing stack

Solution: Leveraging HeatWave to Automate IT & Security Management

50% Reduction in ML Activity
Reduction in Data Cleaning, Model Selection, Model Tuning and Training Time

15-25% Performance Improvement
Using Auto-Indexing, Auto-ML and Javascript
Move processing closer to Data to improve latency (Javascript)

Tech. Consolidation
Seamlessly extending relational model to support OLAP/ML/AI workload
Reduction in maintaining and deploying multiple Technology stack and training
India’s leading private sector bank, offers Online NetBanking Services & Personal Banking Services like Accounts & Deposits, Cards, Loans etc.

**Use Case**
- Identify upsell opportunities
- Detect fraudulent accounts

**Model Type**
- Anomaly Detection, Generative AI

**Results**
- One patented algorithm addressed various types of anomalies leading to effective multi layer fraud detection
- Database developers were able to build the models without ML expertise
- Was able to create thousands of predictions per second to meet high throughput requirement from the bank
Dubai-based online aggregator connecting thousands of its users to their favorite restaurants, making online ordering easier, reliable and convenient.

Use Case
- Predict food deliver time.
- Suggest food / restaurant based on past actions.
- Summarize menu for the selected restaurant.

Model Type
- Regression, Recommendation, Generative AI

Results
- Developed ML models in days that would have otherwise taken months
- Database developers were able to build the models without ML expertise
- Simplified infrastructure with no complex ETL to manage and one platform providing OLTP, Analytics, ML and GenAI
- Consistent interface across various ML model types simplified learning for Eat Easy development team