MySQL Webinar Series

Using Oracle Cloud Infrastructure Data Integration with MySQL Database Service and HeatWave

Nicolas de Rico
MySQL Master Principal Solution Engineer
Oracle Solution Engineering

Julien Testut
Senior Principal Product Manager
OCI Data & AI Services
Agenda

• MySQL Database Service & HeatWave
• Oracle Cloud Infrastructure Data Integration
  • Overview
  • Demo
  • Use Cases
  • User Experience
• Q&A
MySQL is the most popular database for developers

<table>
<thead>
<tr>
<th>Most popular databases</th>
<th>Which databases have you used in the last 12 months?</th>
</tr>
</thead>
<tbody>
<tr>
<td>MySQL</td>
<td>MySQL</td>
</tr>
<tr>
<td>PostgreSQL</td>
<td>PostgreSQL</td>
</tr>
<tr>
<td>Microsoft SQL Server</td>
<td>MongoDB</td>
</tr>
<tr>
<td>SQLite</td>
<td>SQLite</td>
</tr>
<tr>
<td>MongoDB</td>
<td>Redis</td>
</tr>
</tbody>
</table>

Stackoverflow survey 2020

Jetbrains survey
Innovative enterprises across many industries run MySQL

Social
- facebook
- twitter
- LinkedIn
- WeChat
- Pinterest

E-Commerce
- Booking.com
- Netflix
- Uber
- Airbnb
- Taobao.com

Tech
- AppDynamics
- GitHub
- HubSpot
- Zendesk
- Mint
- New Relic

Finance
- Bank of America
- JPMorgan
- Citigroup
- Fidelity
- Visa

Manufacturing
- Tesla
- Volkswagen
- Toyota
- Caterpillar
Problem: MySQL users need separate systems for OLTP and OLAP
MySQL Database Service & HeatWave

Easily run high performance analytics against your MySQL database, no ETL required

Single MySQL database for OLTP & analytics applications

All existing applications work without any changes

Extreme performance: 400x faster than MySQL, scales to thousands of cores
All MySQL-compatible analytics applications run without any changes.
HeatWave dramatically speeds up analytic queries

400G, 64 cores

*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.
HeatWave vs. Amazon Aurora

4TB

- 30x Sooner
- 1100x Faster
- 3.5x Cheaper

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

Copyright © 2021, Oracle and/or its affiliates. All rights reserved.
HeatWave vs. Amazon Redshift’s low-cost shape

4TB

*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.
HeatWave vs. Amazon Redshift’s fastest shape
4TB, 10 nodes

- **2.7x Faster**
- **3x Cheaper**

**Geo Mean of 19 derived TPC-H Queries**
- Redshift (4 * dc2.8xlarge): 19.7 seconds
- HeatWave (10 E3 nodes): 7.3 seconds

**Annual Cost**
- Redshift: $110,560
- HeatWave: $37,022

*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.
MySQL database service costs less

MySQL Database Service: Standard E3 AMD 16GB/Core, all regions have the same price.

Amazon RDS: Intel R5 16GB/Core, AWS US East.

Microsoft Azure: Memory Optimized Intel 20GB/Core, MS Azure US-East.

Google: High Memory N1 Standard Intel 13GB/Core, GCP Northern Virginia.

Configuration: 100 OCPUs, 1TB Storage.

Annual cost for a 100 OCPUs, 1TB Storage configuration

MySQL Database Service: $65,833
Amazon RDS: $215,652
Microsoft Azure: $212,974
Google Cloud SQL: $170,244

Copyright © 2021, Oracle and/or its affiliates. All rights reserved.
MySQL Database Service is integrated with other Oracle Services
End-to-end Integration from data ingestion to data visualization

Data Sources
- Enterprise
- Applications
- Devices
- Sensors
- Events
- Social voice
- Any digital asset

Data Consumers
- People
- Applications

Oracle Cloud Infrastructure

MySQL Database Service

Data Ingestion

MySQL Database Service

Data Integration Service

OLTP Engine

Analytics Engine

Oracle Analytics Cloud

Data Visualization
Data Integration Challenges

- Inaccessible data, lost opportunities
- Machine learning based on data you can’t trust
- Incomplete, stale data leads to poor decisions
Cloud, Big Data Analytics and AI/ML

Convergence is driving new opportunities

- **Cloud**: Instant, elastic compute on infinite storage
- **Big Data**: Data lakes enables predictive analytics
- **Data Science AI & ML**: Digital transformation to gain competitive advantage
Introducing OCI Data Integration
Oracle Cloud Infrastructure (OCI) Data Integration with MySQL Database Service and HeatWave

A cloud native, serverless ETL service for integrating, transforming, and moving data within the OCI ecosystem

- Graphical, code-free designer
- Native integration with MySQL Database Service and HeatWave
- Interactive Data Preparation and Profiling
- Schema evolution protection
- Powered by Spark ETL or E-LT Push-Down

Copyright © 2021 Oracle and/or its affiliates.
OCI Data Integration
Features
Maximize Developer Productivity

No Code Data Flow Design

- Powerful graphical editor for building Data Flows
- Visually preview data with Data Xplorer
- Parameterize Data Flows for maximum flexibility

Benefits

- Empowers developers to innovate faster
- Simplified ETL design and maintenance
- Powerful and flexible data integration transformations
Visually Preview and Prepare Data

- Visually discover insights
- Access statistical profiling
- Prepare and transform data iteratively

Benefits
- Boost productivity & load data in a few minutes
- Promote best practices and reuse data flow rules
More Time to Innovate and Improve Your Business

Schema Drift Protection

- Schema evolution (drift) often invalidates processes, requiring costly manual maintenance
- Rule-based design protects from schema drift by handling schema changes dynamically

Benefits
- Minimize development costs with simpler maintenance
- Free up developer time to get more value from data
Broad Connectivity
Optimized for Oracle Cloud

- Secure public or private access even on-premises
- Optimized connectivity for Oracle Cloud including
  - MySQL
  - Oracle Autonomous Database (ADW/ATP)
  - Oracle Database & Exadata DB Systems
  - Oracle Object Storage: CSV, JSON, Parquet, Avro
  - Microsoft SQL Server / PostgreSQL / Apache Hive

Benefits
- Best in class connectivity for Oracle Cloud
- Expanding set of easy-to-use native adapters
Connecting OCI Data Integration to MySQL Database Service

- Native Data Asset Type for MySQL Databases
- Using MySQL Connector / JDBC
- Faster push-down E-LT processing using HeatWave
- Allows MySQL users to take full advantage of all capabilities with OCI Data Integration
Pipeline & Scheduling

- Design Pipelines to run your Data and AI processes end-to-end with Data Integration tasks along with Data Flow applications and Data Science models invocation.
- Create sequential or parallel executions, add conditional links and retry logic to handle errors.
- Schedule executions to run hourly, daily, monthly and more.
Industry Leading Performance
Powered by Spark ETL or SQL Push-Down

• Supports both ETL and E-LT processing
• Generates executable code for the best engine
  • Spark-based ETL for distributed data processing
  • Push-down optimization for E-LT processing on database engines including MySQL

Benefits
• Highly efficient execution to process data at scale
• Fully managed serverless runtime
Demo
Key Use Cases

**Data Integration for Data Marts and Analytics**

Efficiently load and transform data at scale into DWs (e.g. MySQL, Autonomous Data Warehouse) used for analytics purposes

**Data Integration for Data Lakes & Data Science**

Efficiently load and transform data at scale into Data Lakes used for data science and analytics purposes
Data Integration for MySQL Database Service and HeatWave

Business Drivers
• Operational or Advanced Analytics in the Cloud requires reliable data feeds
• Work with data coming from on-premise data sources or other cloud sources

Key Features
• Prepare and ingest data to Data Warehouses and Data Marts running on MySQL
• Transform and cleanse data using powerful hybrid execution layer (Spark/SQL Push-Down)
OCI Data Integration
User Experience
Home

• Landing page for Data Integration Workspaces
• Quickly create new Data Assets, new Tasks, Pipelines or Data Integration Flows using Quick Actions tile
• Navigate to your Projects, Data Assets and Applications
Data Flows

- Powerful Data Integration Flow Editor
- Quickly create simple to complex ETL processes with Joins, Filters, Union or Aggregate Functions
- Native schema drift handling
- Immersive Data Xplorer to preview transformation results
Tasks

• Prepare, Integrate or Load Data easily using Tasks
• Configure processes for MySQL, Autonomous Databases or Object Storage Data Lakes in a few clicks using powerful data preparation transformations
• Augment Data Integration Flows with advanced parameters
Applications

- Publish Tasks into Applications to execute them
- Iterate on your data integration processes without impacting runtime operations
- Configure your Tasks with runtime configuration
- Schedule Tasks
- Monitor your Task Runs
Follow us on Social Media

mysql.com
oracle.com/mysql/analytics
twitter.com/mysql
facebook.com/mysql
linkedin.com/company/mysql
Q&A

Thank You!