Open Source Data Warehousing & Business Intelligence

Zack Urlocker
VP Products, Database Group
ZUrlocker@mysql.com
Why Did Sun Acquire MySQL?

The acquisition...
- Drives market tipping change in the $15B database marketplace
- Positions Sun as the leading platform for the Web economy
- Expands Sun's reach into MySQL's LAMP user base
- Expands MySQL's capacity to deliver global service and support
- Affirms Sun's position as the largest commercial open source contributor

Alignment in Culture and Vision
Industry-Leading Customers

Open-source is powering the Web
Leadership, innovation, market acceptance

12 years old
400+ employees
750 partners
70K downloads / day
Customers across every major operating system, hardware vendor, geography, industry, and application type
Typical Pattern of Disruption

**Innovator’s Dilemma**
- Entrant considered inferior
- Entrant gains toehold
- Incumbent waits & watches
- Entrant becomes “good enough”
- Niche market booms
- Incumbent “crams” disruption
- New model wins

**Implications**
- Aim for new applications, not legacy
- Co-exist with legacy database systems
- Many customers are also disruptive
MySQL’s Disruptive Sneak Attack

**Disruptive Innovation**
- Defy perceived rules
- Change the game
- Radical business model
- Missing key features
- Targeted niche approach

*New entrants can win big*

**Implications**
- Focus on specific use cases where partners extend our reach
- Don’t fight “feature wars” on incumbent’s turf
- Look for repeatable patterns in target markets
Who’s Using Open Source BI / DWH?

• Wide range of use cases
  > Reporting, Analysis
  > Dashboards
  > Data Integration
  > Data Mining

• Large & small deployments
  > Dozen to Thousands of users
  > Megabytes - Terabytes of data

• Multiple deployment models
  > New deployments
  > Co-existence
  > Replacement
Range of Open Source Usage

- **Frontier Airlines – Yield Management Analysis**
  - 4 years of historical ticket data to the individual ticket
  - 800 million records; Half a terabyte of live OLAP data
  - 4 server Greenplum cluster
  - Most queries under 8 seconds

- **Orbitz – Agent web portal**
  - Self-service portal travel agents with integrated reporting
  - 2,500 users with contract renewal, ordering, reporting
  - Using Red Hat, MySQL, Pentaho
  - $1m in TCO savings

- **Enterasys - Embedded BI Reporting**
  - Dragon Intrusion Detection / Prevention appliance
  - Log millions of “events” per day for 3,000 customers
  - Widely used in government, financial services
  - Requires highly scalable custom reporting
  - Using Debian, MySQL, JasperReports
Sun BI / Data Warehouse Partners

Apps
- Operational Data Store
- Source Systems
- ETL Process
- Data Warehouse
- Metadata Repository
- Data Mart

Tools
- Reporting & Analytics
- ETL
- Replication
- Development & Admin

Platforms
- Database
- Storage (SAN – NAS)
- Operating System
- Hardware

Embracing a suite of comprehensive solutions and tools, Sun BI / Data Warehouse Partners offers a robust ecosystem for businesses aiming to enhance their business intelligence and data warehouse capabilities. The integration of leading technologies from companies like SAS, Business Objects, Jaspersoft, Actuate, Birt, Pentalo, GoldenGate, Quest Software, Informatica, IBM, Talend, Jaspersoft, Pentaho, MySQL, KickFire, PostgreSQL, Greenplum, EMC, Sun Microsystems, IBM, HP, NetApp, Hitachi, Dell, HP, Sun Microsystems, AMD, Intel, and SCO, ensures a comprehensive approach to data management, analysis, and operational efficiency.
# Data Warehouse: DBMS Technologies

<table>
<thead>
<tr>
<th>Small Data Warehouses</th>
<th>Medium-Large Data Warehouses</th>
<th>Emerging Technologies</th>
</tr>
</thead>
<tbody>
<tr>
<td>MySQL</td>
<td>Oracle</td>
<td>Aster Data Systems</td>
</tr>
<tr>
<td>PostgreSQL</td>
<td>DB2</td>
<td>HyperRoll</td>
</tr>
<tr>
<td>SQL Server</td>
<td>Teradata</td>
<td>Dataupia</td>
</tr>
<tr>
<td></td>
<td>Neoview</td>
<td>QlikTech</td>
</tr>
<tr>
<td></td>
<td>Sybase IQ</td>
<td>TeraManager</td>
</tr>
<tr>
<td></td>
<td>Netezza</td>
<td>SAND DNA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Calpont</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Truvisio</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EAXSOL</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Kickfire</td>
</tr>
</tbody>
</table>
Typical Data Warehouse Size

• According to IDC, data warehouses are growing, but not as large as the press reports.
• 60% of data warehouses are under a terabyte
• Only 4% are larger than 25TB
MySQL – A Popular Data Warehouse Choice

- MySQL DW deployed at 28% of MySQL customers
- Strong DW ecosystem support

What types of applications are you deploying on MySQL?

- Custom Applications
- Web (non-transacational)
- Web (transactional)
- Data Warehouse
- IT Infrastructure
- Messaging & Collaboration
- CRM
- Financials
- BI/Analytics
- Retail
- Supply Chain
- Telecom
- Other
- ERP
- Security (SFA)
- HRMS

Legend:
- Enterprise
- Community
MySQL Data Warehousing Strategy

• Strongly support common data warehouse use cases
• Partner with major BI/ETL vendors
• Offer highly attractive total cost of ownership
Common Use Cases

1. Small, semi real-time data marts
2. Continuous, real-time/query data warehousing
3. Traditional, standard reporting warehouse
4. Massive historical, with ad-hoc queries warehouse
5. BI, analytic in OLTP applications (emerging…)

Data Mart | Real-Time | Traditional | Historical | Analytic-OLTP
MySQL Technical Strategy

- Open source architecture to maximize innovation
- Core data warehousing feature set
- Specialized data warehouse engines for key use cases
- Mixed workload with multiple engines
Core Features for Data Warehousing

- No practical storage limits (1 tablespace=110TB)
- Automatic storage management
- ANSI-SQL support for all datatypes (including BLOB and XML)
- Data/Index partitioning (range, hash, key, list, composite)
- Built-in Replication
- Main memory tables (for dimension tables)
- Variety of indexes (b-tree, fulltext, clustered, hash, GIS)
- Multiple-configurable data/index caches
- Pre-loading of index data into index caches
- Unique query cache (caches result set + query; not just data)
- Multi-insert DML
- Data compression
- Read-only tables
- Fast connection pooling
- Cost-based optimizer
- Wide platform support
The MySQL Storage Engine Advantage

Modern, Flexible Architecture
  +
Strong Internal Engineering Staff
  +
Growing ISV Engine Program
  +
Thriving Community Development

✓ Unique position in the market
✓ Custom DB for perfect application fit
✓ High performance
✓ Greater efficiency
✓ Rapidly expanding reach of MySQL
Data Warehouse Engines Overview

**MyISAM**
- High-speed query/insert engine
- Non-transactional, table locking
- Perfect for data marts, small warehouses
- Compresses data by up to 80%
- Fast table scans for large tables
- Only allows inserts/selects
- Great for seldom accessed data
- Main memory tables
- Perfect for small dimension tables
- B-tree and hash indexes
- Comma separated values
- Allows both flat file access and editing as well as SQL query/DML
- Allows instantaneous data loads

**Archive**
- Compresses data by up to 80%
- Fast table scans for large tables
- Only allows inserts/selects
- Great for seldom accessed data

**Memory**
- Main memory tables
- Perfect for small dimension tables
- B-tree and hash indexes

**CSV**
- Comma separated values
- Allows both flat file access and editing as well as SQL query/DML
- Allows instantaneous data loads

Also:
- Federated for remote data access
Data Warehouse Engines Overview

**InfoBright**
- Column-oriented structure
- Powerful high-speed loader
- Strong data compression (10:1)
- Non-transactional engine
- Auto-partitioning; no indexing

**Kickfire**
- Transactional data warehouse appliance
- SQL-chip for performance acceleration
- Column-oriented, data compression

**Nitrosecurity**
- Specializes in aggregation style queries via N-tree indexes
- Non-transactional database engine
- Supports heavy data loading, DML, and simultaneous queries
### Warehouse use cases/engine mapping

<table>
<thead>
<tr>
<th>Data Mart</th>
<th>Real-Time</th>
<th>Traditional</th>
<th>Historical</th>
<th>Analytic-OLTP</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="#" alt="Diagram" /></td>
<td><img src="#" alt="Diagram" /></td>
<td><img src="#" alt="Diagram" /></td>
<td><img src="#" alt="Diagram" /></td>
<td><img src="#" alt="Diagram" /></td>
</tr>
</tbody>
</table>
| - MyISAM  
  - Kickfire  
  - Memory | - Nitro  
  - Kickfire | - MyISAM  
  - Archive  
  - Merge  
  - Infobright  
  - Kickfire | - Archive  
  - Infobright  
  - Kickfire | - InnoDB  
  - PBXT |

- Data Mart: Real-time analytics, historical analytics, and traditional analytics.
- Real-Time: Stream processing, event sourcing, and real-time warehousing.
- Traditional: Decision support, data warehousing, and data mining.
- Historical: Data archiving, query optimization, and data migration.
- Analytic-OLTP: Transaction processing, data manipulation, and analytics support.
Brighthous from Infobright

MySQL Provides:

- Mature connectors, tools, resources
- Interconnectivity and certification with BI Tools
- Management Services and Utilities

Infobright Provides

- Load function that compresses data
- Column-oriented data storage engine
- Knowledge Grid metadata layer that contains information about the compressed data
- Optimizer/executor that uses the Knowledge Grid.
Infobright Customer Example

- Online marketing research firm
- Analyze 20 billion online transactions a month
- Beat out Oracle solution
- Achieved 20-30:1 data compression ratio

“Brighthouse performed extremely well,” noted Mats Johansson, senior consultant at Lincube Group AB, the organization that conducted the trials/testing. “It handled huge volumes of data and resolved queries faster than I have ever witnessed after years in this industry—and did so using standard hardware.”
Kickfire

World’s first high-performance appliance for MySQL

- Makes MySQL faster for reporting and queries
- Affordable, low-power, load-and-go appliance
- Scalable from GBs to TBs
- Achieved TPC-H record-breaking benchmarks (100, 300GB)
Kickfire Example

Customer: Public company managing online forums

- Test platform
  - One 45GB customer DB
  - 50 million rows in fact table
  - Sample query below

```sql
SELECT
  sql_date,
  forum_name,
  COUNT(DISTINCT visitor_id) AS Visitors,
  COUNT(DISTINCT user_key) AS Users
FROM
  customer_mart.clickstream_fact cf,
  customer_mart.forum_dim fd,
  forum_dim.user_agent_dim_robots ua,
  forum_dim.day_dim dd
WHERE
  cf.forum_key = fd.forum_key AND
  cf.user_agent_key = ua.user_agent_key AND
  robot_flag = '1' AND
  cf.day_key = dd.day_key
GROUP BY
  sql_date,
  forum_name
ORDER BY
  sql_date,
  Visitors DESC;
```

- Results
  - Average 35X improvement

<table>
<thead>
<tr>
<th>Customer System Performance</th>
<th>Kickfire Appliance Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>01 232s</td>
<td>22s</td>
</tr>
<tr>
<td>02 3646s</td>
<td>24s</td>
</tr>
<tr>
<td>03 1189s</td>
<td>39s</td>
</tr>
<tr>
<td>04 568s</td>
<td>12s</td>
</tr>
<tr>
<td>05 18s</td>
<td>6s</td>
</tr>
<tr>
<td>06 1952s</td>
<td>96s</td>
</tr>
</tbody>
</table>
Supporting the Warehouse - MySQL Enterprise

**Server**
- MySQL Enterprise Server
- Monthly Rapid Updates
- Quarterly Service Packs
- Hot Fix Program
- Extended End-of-Life

**Monitor**
- Global Monitoring of All Servers
- Web-Based Central Console
- Built-in Advisors
- Expert Advice
- Specialized Scale-Out Help

**Support**
- 24 x 7 x 365 Production Support
- Web-Based Knowledge Base
- Consultative Help
- Bug Escalation Program
Added Value of MySQL Enterprise

- Comprehensive offering of production support, monitoring tools, and MySQL database software
- Optimal performance, reliability, security, and uptime
MySQL Unlimited

- Fixed Annual Subscription:
  - Unlimited Servers
  - Unlimited CPUs
  - Unlimited Cores

- Simplify
  - No Counting
  - No Compliance Issues

- Pricing
  - No proprietary DBMS license fees
  - Price starts at $40K/year
MySQL Services

Consulting
• Architecture and Design
• Database/Application Migrations
• Performance Tuning and Optimization
• Data Warehousing
• High Availability

Training
• In-house and Distance Learning
• DBA, Developer, and Sysadmin Courses
• MySQL Certification Training and Testing
• Hands On Experience

Support
• 24 x 7 x 365 Production Support
• Design Reviews
• Consultative Help
Data Warehouse Success Strategies

• Select the right hardware for the job
• Select the right engine(s) for each scenario
• Use core MySQL data warehouse features
• Tune key MySQL configuration parameters
• Leverage open source ETL, BI and Reporting
“Open source software represents the most significant all-encompassing and long-term trend that the software industry has seen since the early 1980s."

“Open-source BI software is probably in your future; the real issue is not whether, but when.”

“Open Source BI is here to stay.”

“Interest in open-source BI technology is high as companies seek alternatives to higher-priced, commercially available products.”
Resources

- MySQL Partner program (MECA)
  > http://solutions.mysql.com/
- MySQL Data Warehouse Central
  > http://www.mysql.com/datawarehouse
- MySQL Case Studies
  > http://www.mysql.com/casestudies
- MySQL Whitepapers
  > http://www.mysql.com/why-mysql/white-papers/
- MySQL Enterprise
  > http://www.mysql.com/products/enterprise
- MySQL Services
  > http://www.mysql.com/services/
Open Source Data Warehousing & Business Intelligence

Zack Urlocker
VP Products, Database Group
ZUrlocker@mysql.com