Clustering for the Masses
A Gentle Introduction to Tungsten for MySQL

Robert Hodges
CTO, Continuent, Inc.
Topics

/ What is the Problem?
/ What is Tungsten and how does it work?
/ What can you do with it?
/ Prove it!
/ Summary and Questions
About Continuent

Our Business: Full service database management

Value:
- Ensure data are available when and where you need them
- Make open source DBMS scale economically

Technical Expertise
- Database replication
- Cluster management
- Application connectivity
- SaaS applications
The Problem: Scaling Businesses on MySQL
MySQL Is Not as Cheap as You Think

We have a system in place based on shared mysql + memcache but it’s quickly becoming prohibitively costly (in terms of manpower) to operate. We need a system that can grow in a more automated fashion and be highly available.

Ryan King - Twitter
MySQL Scaling Problems

/ Failures and maintenance on complex topologies
/ SLAs on shared resources
/ Multi-tenant operations
/ Growing transaction and data volumes
/ Administrative labor costs
Convert to NoSQL?? NOT!

- Integrate clusters of off-the-shelf MySQL databases
- Hide management complexity behind simple interfaces
- Use hardware efficiently
- Manage multi-tenant data and resource use
- Provide seamless upgrade
- Operate across multiple sites
- Protect data with decent backup/restore and consistency checking
What is Tungsten and How Does It Work?
What Is Tungsten?

Tungsten implements data services consisting of off-the-shelf databases linked by replication to:

- Protect data
- Maintain high availability
- Improve resource utilization
- Raise performance

Install and set up in a few minutes

Integrated backup/restore and data integrity checks

Rule-driven management with automated failover

Simple procedures for updates and maintenance

SQL scaling for multi-tenant systems
Tungsten Data Service Architecture

Apache/Mod_PHP
libmysqlclient.a
Manager
Connector
DBMS
Replicator
Manager
Slave

Apache/Mod_PHP
libmysqlclient.a
Connector
Manager
DBMS
Replicator
Manager
Master

DBCMS
Replicator
Manager
Slave

MySQL Conference 2010
Flexible Replication Pipelines

Tungsten Replicator Process

Pipeline

Stage

Extractor → Filters → Applier → Extractor → Filters → Applier

Stage

binlogs

THL

Transaction History Log

Slave DBMS

MySQL Conference 2010
Distributed Rule-Based Management

- Business Rules
- Manager #1 (Coordinator)
- Group Communications
- Broadcast commands and monitoring data
- Manager #2
- Local DBMS Resources
- Cctrl Client #1
- Cctrl Client #2
- Manager #3
- Cctrl Client #3
- Local DBMS Resources

MySQL Conference 2010
Managed Routes to Databases

Java App Server
- Tungsten SQL Router
- MySQL JDBC Driver

PHP Application
- libmysqlclient.a

Clustered Databases

Admin & Monitoring

Virtual IP Address

Tungsten Connector
How Do You Set Up A Cluster?

- **Fulfill host pre-reqs**
  - MySQL, Ruby, Java, network hygiene

- **Download Tungsten software to host and un-tar**
  - Or build it yourself from source!

- **Run ‘configure’ script**

- **Services start automatically and join cluster as soon as replicator goes online for the first time**

- Rinse and repeat on each host
How Do You Administer Clusters?

/ ‘cctrl’ is the Tungsten management client
/ Connect to any manager to administer cluster
/ List data sources in the cluster
  [LOGICAL] /cluster/demo/> ls
/ Check cluster liveness
  [LOGICAL] /cluster/demo/> cluster heartbeat
/ Switch master and slave
  [LOGICAL] /cluster/demo/> switch to prod01
/ Start and stop operating system services
  [LOGICAL] /cluster/demo/> service */mysql restart
/ Backup and restore databases
  [LOGICAL] /cluster/demo/> datasource prod02 backup
  [LOGICAL] /cluster/demo/> datasource prod02 restore
How Do You Scale Reads?

/ Tungsten “session-consistency” distributes tenant reads to slaves if caught up with last tenant write

/ Works with tenants sharded by database or with independent database connections
How Do You Handle Failures?

/ **Automatic policy mode** performs master failover automatically whenever:
  • A database crashes
  • A replicator fails or crashes
  • A server reboots or drops off the network

/ **Manual policy mode** lets you failover manually
  [LOGICAL] /cluster/demo/> failover

/ **Failover procedure** promotes most advanced slave
  • And points other slaves to the slave

/ **Recover command** brings back failed master
  [LOGICAL] /cluster/demo/> datasource centos5a recover

/ **Slaves recover without admin intervention in automatic mode**
How Do You Check Data Consistency?

/ Tungsten incremental checks verify data without stopping replication
  • Implemented as replication events inside Tungsten Replicator
  • Compute MD5 on all or part of tables
  • Can do single tables or all tables in a database

/ Run checks from manager in combination with flush command
  [LOGICAL] /cluster/demo/> cluster check tenant1.*
  [LOGICAL] /cluster/demo/> cluster flush

/ Slaves can either fail or warn of failures
How Do You Perform Maintenance?

// Tungsten supports “daylight maintenance”

// To perform maintenance on a slave, just take it offline
[LOGICAL] /cluster/demo/> set policy maintenance
[LOGICAL] /cluster/demo/> datasource centos5b offline
[LOGICAL] /cluster/demo/> replicator centos5b offline

// To perform maintenance on a master, first do a switch to promote a new master
[LOGICAL] /cluster/demo/> switch

// Then perform maintenance using the slave procedure

// Switch the master back when you are done
What Can I Do with Tungsten?
Lots of Things…

/ Raise application availability without restarts
/ Scale applications using reads on slaves
/ Perform daylight maintenance on hosts
/ Archive selected data off-site
/ Ensure applications are online
/ Replicate from databases into files, queues, etc.
  • Some assembly required

How about online schema upgrade?
Online Upgrade

Application Server

SQL Router/Connector

DB1
Master

Writes to master

DB2
Slave

Reads to slave

Step 0: Master and slave both online
Online Upgrade

Step 1: Shun Slave and Upgrade

- Writes and reads to master
- Run upgrade script

Application Server
- SQL Router/Connector

DB1
- Master

DB2
- Slave
Online Upgrade

Step 2: Welcome Slave Back

 Writes to master

DB1 Master

SQL

(Catching up…)

DB2 Slave

Application Server

SQL Router/Connector

Reads to slave; max latency in router ensures timely app data
Online Upgrade

Application Server

SQL Router/Connector

DB1 Slave

DB2 Master

Reads to slave

Writes to master

Step 3: Switch master
Online Upgrade

Step 4: Shun Slave and Upgrade

Application Server
SQL Router/Connector

DB1 Slave

DB2 Master

 Writes and reads to master

Run upgrade script
Online Upgrade

Reads to slave; max latency in router ensures timely app data

 Writes to old slave

DB1 Slave

(Catching up…)

DB2 Master

Step 5: Welcome slave back
Online Upgrade

Step 5: Switch back to original master and check consistency
Does This Really Work?

Time for a demo!
Conclusion
Tungsten Development History

2007  Initial idea

2008  Replicator alpha but mostly vaporware

2009  Integrated management & first deployments

2010  Full parsing of MySQL binlog & advanced SaaS usage
2010 Development Roadmap

/ Tungsten is adding features rapidly for clusters based on MySQL 5.0/5.1

<table>
<thead>
<tr>
<th>Feature</th>
<th>Beta</th>
<th>Prod</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flexible replication pipelines</td>
<td>Q1</td>
<td>NOW</td>
</tr>
<tr>
<td>100% of binlog replicated</td>
<td>Q1</td>
<td>NOW</td>
</tr>
<tr>
<td>Shard-based parallel replication</td>
<td>Q2</td>
<td>Q2-3</td>
</tr>
<tr>
<td>Fast event logging</td>
<td>Q2</td>
<td>Q2-3</td>
</tr>
<tr>
<td>Low-latency WAN replication</td>
<td>Q2</td>
<td>Q2-3</td>
</tr>
</tbody>
</table>

/ PostgreSQL 8 warm standby support and adding features to manage PostgreSQL 9

/ Official Drizzle support as soon as we get customers
Conclusion

/ Scaling MySQL economically is harder than it looks

/ Tungsten solves the problem through easy-to-deploy data services that integrate:
  • Flexible replication
  • Unitary management
  • Transparent connectivity

/ Check out Tungsten at http://www.continuent.com
Contact Information

HQ and Americas
560 S. Winchester Blvd., Suite 500
San Jose, CA 95128
Tel (866) 998-3642
Fax (408) 668-1009

e-mail: sales@continuent.com

robert.hodges at continuent.com
http://scale-out-blog.blogspot.com

EMEA and APAC
Lars Sonckin kaari 16
02600 Espoo, Finland
Tel +358 50 517 9059
Fax +358 9 863 0060

Continuent Web Site:
http://www.continuent.com

MySQL Conference 2010