HA Solutions for MySQL

Joffrey Michaie
Consultant

SkySQL

Agenda

- Introduction
- MySQL Replication
  - Semi-Synchronous Replication
  - MySQL 5.6
  - MariaDB 10.0
- MHA
- Galera
- Shared Disk
- DRBD
- MySQL Cluster

SkySQL

Introduction to HA

“High availability is a system design protocol and associated implementation that ensures a certain degree of operational continuity during a given measurement period”

SkySQL

Uptime, Downtime, 9s

Availability = uptime / (uptime + downtime)

90%	1 nine	36.5 days / year
99%	2 nines	3.65 days / year
99.9%	3 nines	8.76 hours / year
99.99%	4 nines	52 minutes / year
99.999%	5 nines	5 minutes / year
99.9999%	6 nines	31 seconds / year

Availability = MTBF / (MTBF + MTTR)

SkySQL

Terminology

- Synchronous vs. Asynchronous
- Shared-Disk vs. Shared-Nothing vs. Shared-Memory
- Single Point Of Failure - SPOF
- Failover vs. Switchover
- Split Brain
- Node Fencing, STONITH, Quorum

Designing for HA

- Which level of availability do I need?
  - How many nines?
- Do I require no loss of data?
  - Could I loose some transactions?
  - Will my users notice or care?
- Do I need automatic failover or is manual switchover ok?
  - How do I test this?
- Can I provide a reasonable service when X is down?
  - Replace X with each component of the service

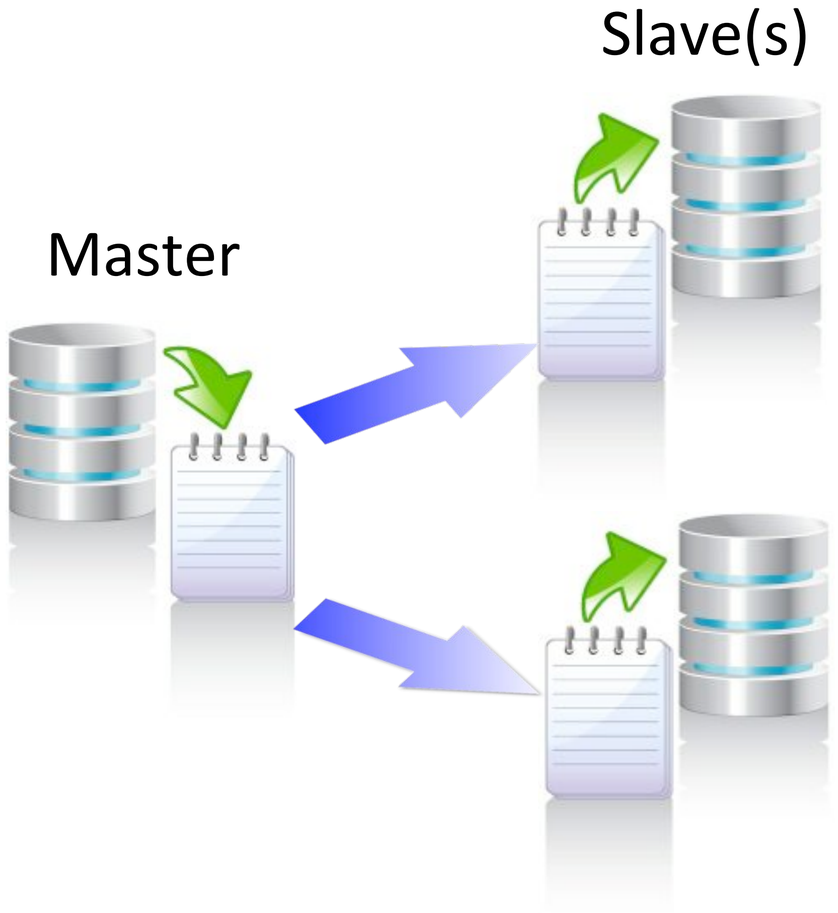
SkySQL

Before we talk about solutions

- A high availability setup does not replace backups
- Check parameters
  - flush_at_trx_commit
  - sync_binlog
  - expire_logs_day
  - binlog_format
  - sync_master_info (on slaves)
  - …
- Requirements for parameters may change if using battery-backed disk cache

MySQL Replication - Asynchronous

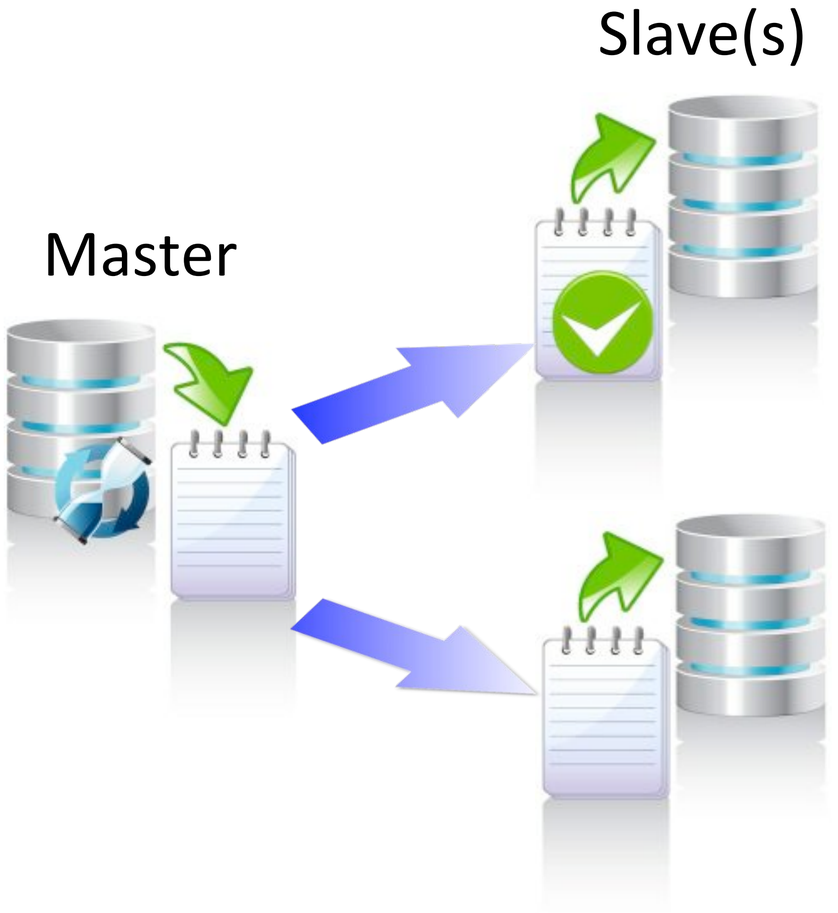
- Asynchronous: 3 Phases
  1. Commit and write to binlog on Master
  2. Ship changes to relay log on slave
  3. Apply changes on slave
- Master -> slave relationship
- Mono-threaded on slaves until MySQL 5.5
  - MySQL 5.6 allows multi-threaded
- No conflict resolution
  - Master-master replication or circular replication need application logic



SkySQL

MySQL Replication – Semi-Synchronous

- Added as a plugin in MySQL® 5.5
- Ensures that changes have been shipped to at least one slave (or timeouts)
- A COMMIT on the master waits for ONE Slave to acknowledge the transaction
  - Important: The Master does not wait for the Slave to execute the transaction, only to write it to the relay log
  - So the Slave SQL Thread may still lag behind the Master and queries to the Slave may still return old data
- Potentially adds latency to queries



SkySQL

MySQL Replication for HA

- Master <-> standby master - manual failover
- Minimal downtime for changes and upgrades
- Semi-synchronous should be used
- Used in combination with other HA solutions for geographical replication

SkySQL

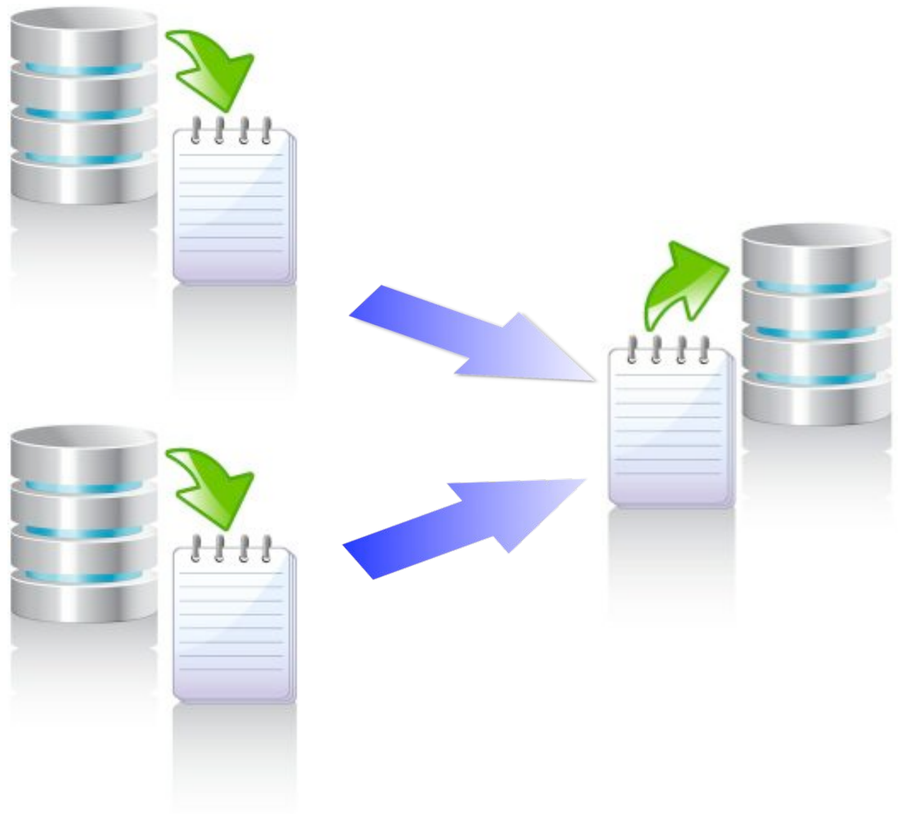
MySQL® 5.6 – New Replication Features

- Global Transaction ID (GTID)
  - Makes it easy to automate failover and slave promotion
- Replication failover and admin utilities
- Multi-threaded slaves
- Replication event checksums
- Time-delayed replication

SkySQL

MariaDB 10.0 – New Replication Features

- Multi-source replication
- Global Transaction ID (GTID)
  - Different implementation from MySQL 5.6
  - GTID per domain instead of server



MHA
Master High Availability Manager

- Automates master failover and slave promotion
  - Monitors the master or can integrate with Pacemaker/Heartbeat
  - Failover is an online operation
  - Also allows manual switchover
- Short downtime: often a few seconds
- MySQL-Replication consistency
- No performance penalty
- Drop in solution on existing deployment

SkySQL

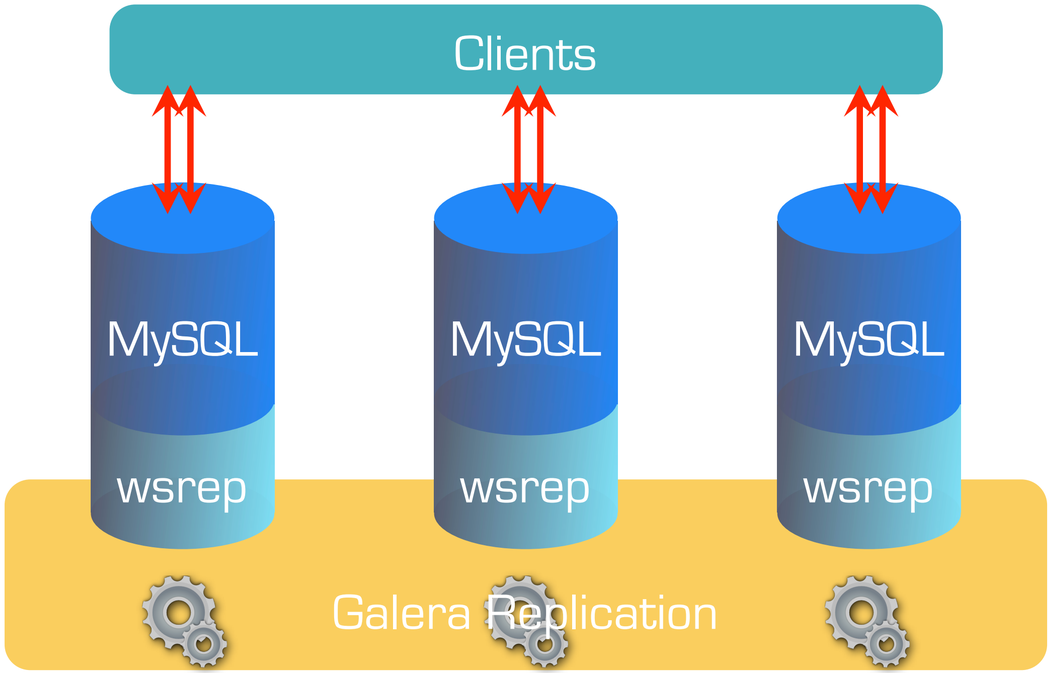
MHA
Failover Process

1. Attempt to contact MySQL master server by SSH
2. If master server is alive access the binary log and recover events
3. Find the slave with the most advanced relay log
4. Sync all slaves to the latest available binlog event
5. STONITH master if necessary
6. Promote slave to master

SkySQL

Galera / MariaDB Galera Cluster

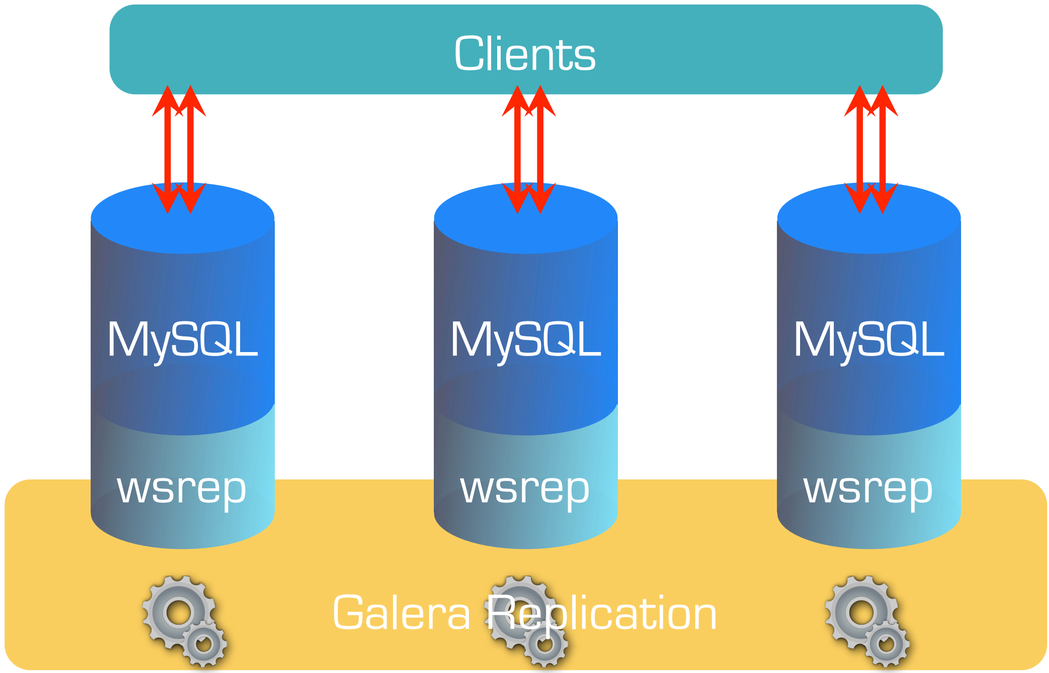
- Provides “virtually” synchronous replication
- Works with InnoDB
- No slave lag
- Transactions are validated by slaves upon commit
  - Certification and quorum
  - Transactions may be rolled back at this stage
- Multi-master or master-slave possible



SkySQL

Galera / MariaDB Galera Cluster

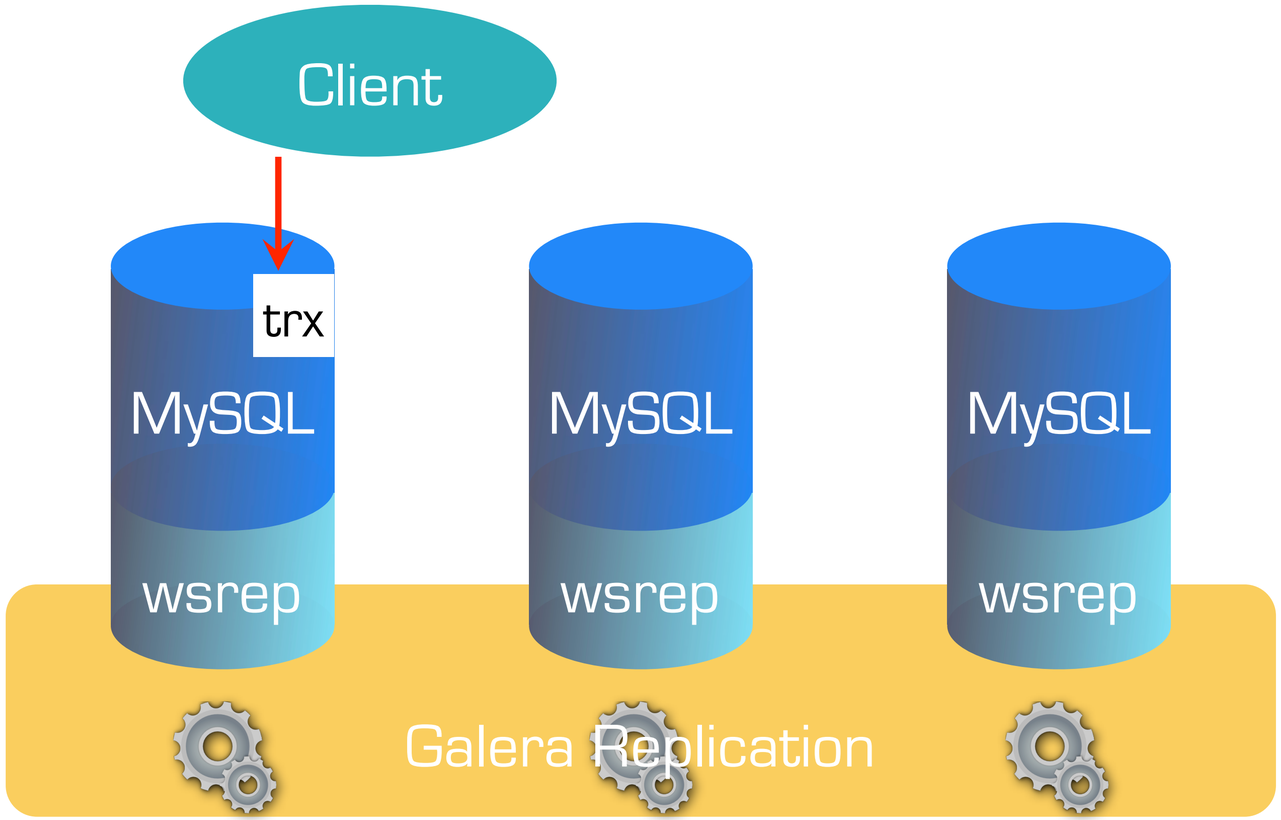
- Provides “virtually” synchronous replication
- Works with InnoDB
- No slave lag
- Transactions are validated by slaves upon commit
  - Certification and quorum
  - Transactions may be rolled back at this stage
- Multi-master or master-slave possible



SkySQL

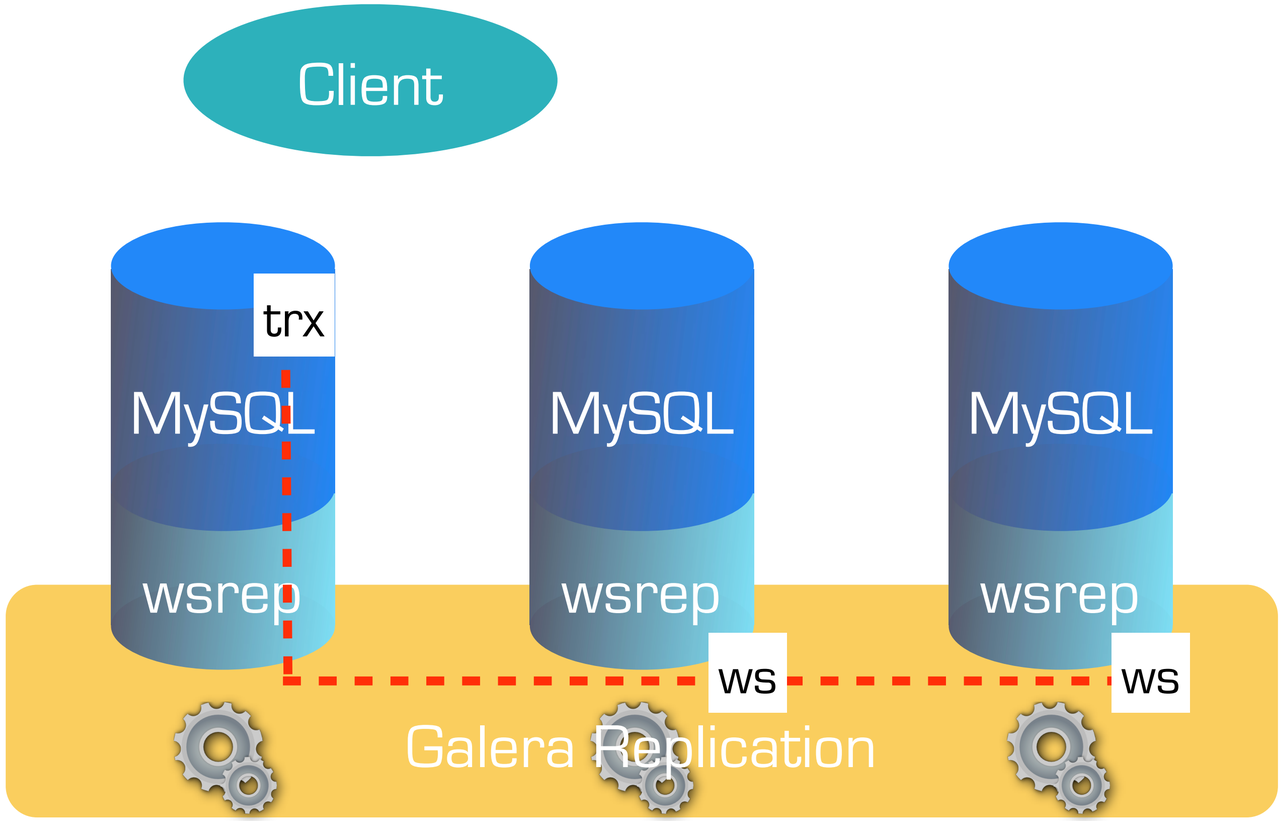
Galera – Transaction Process

- Transaction is processed locally up to commit time



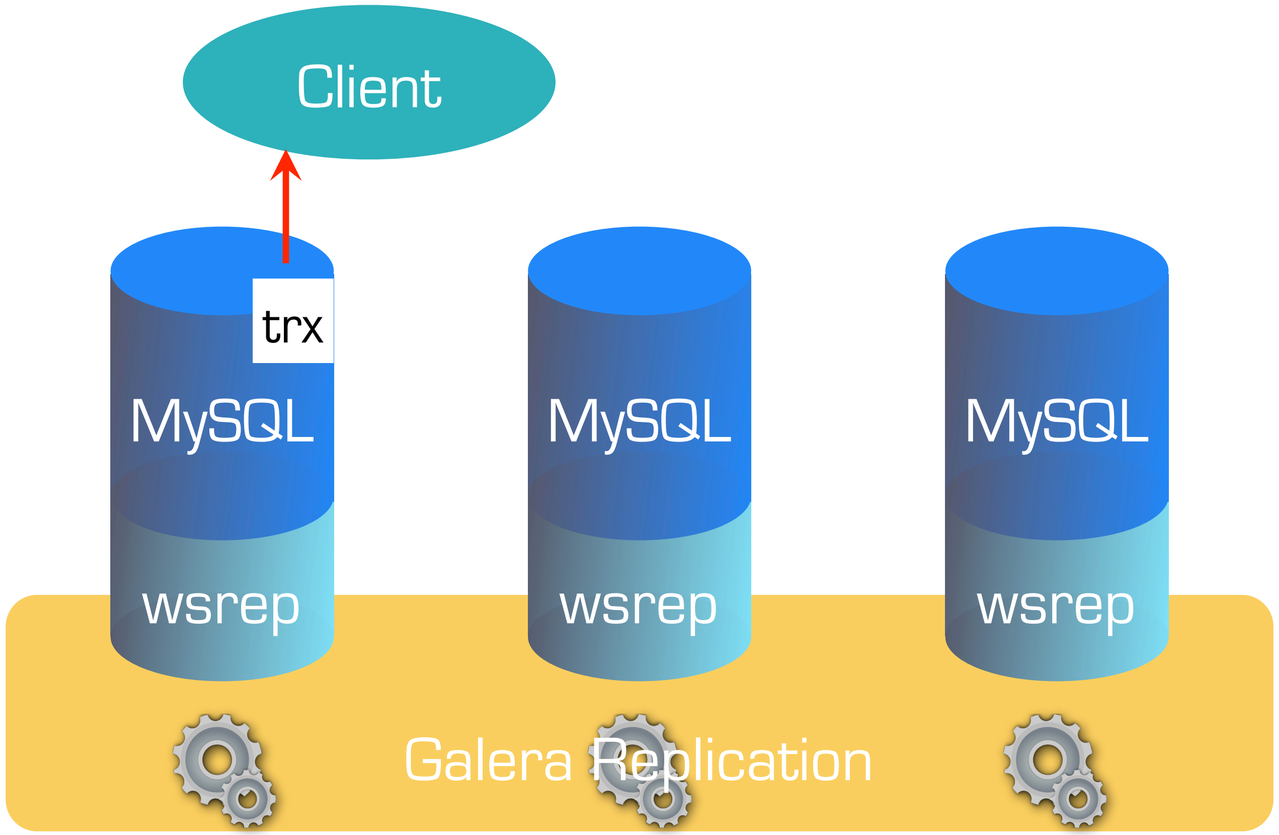
Galera – Transaction Process

- Transaction is processed locally up to commit time
- Transaction is replicated to whole cluster



Galera – Transaction Process

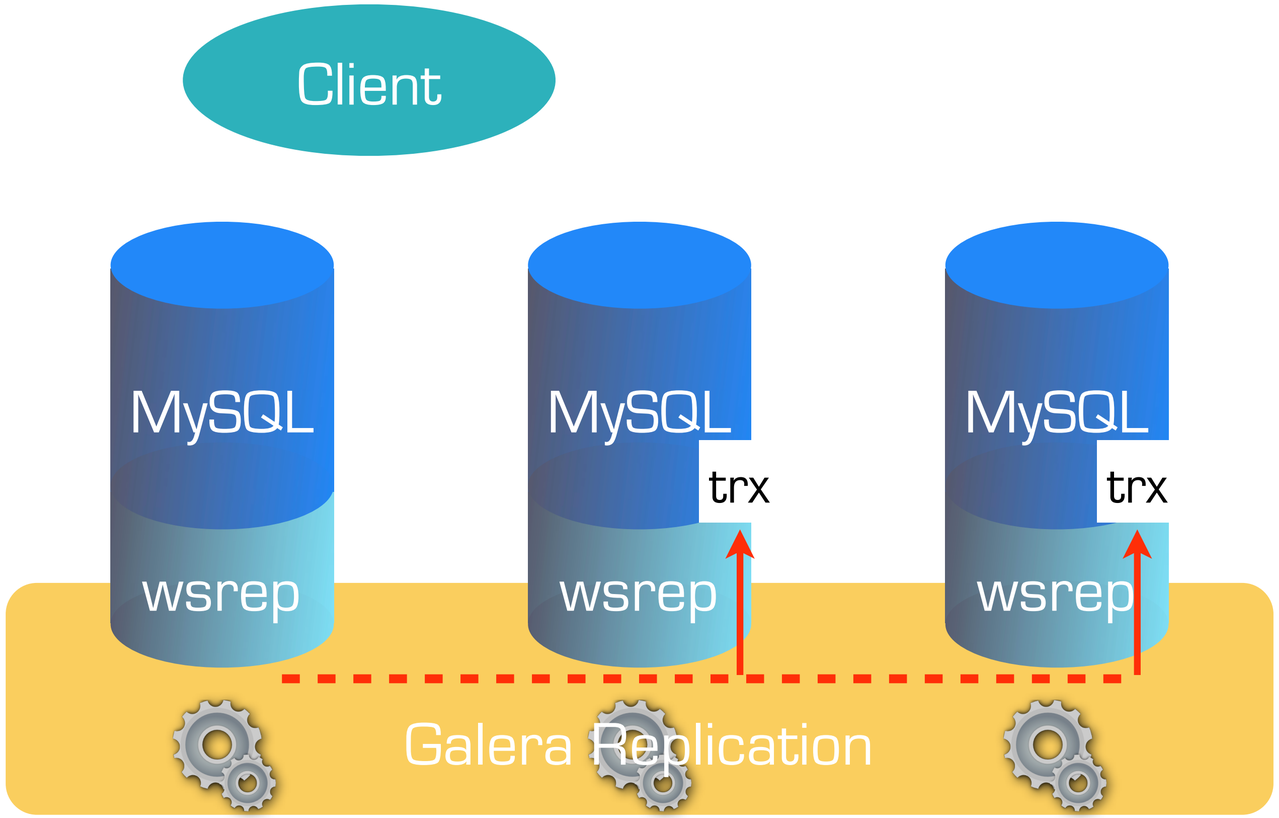
- Transaction is processed locally up to commit time
- Transaction is replicated to whole cluster
- Client gets OK status



SkySQL

Galera – Transaction Process

- Transaction is processed locally up to commit time
- Transaction is replicated to whole cluster
- Client gets OK status
- Transaction is applied in slaves



SkySQL

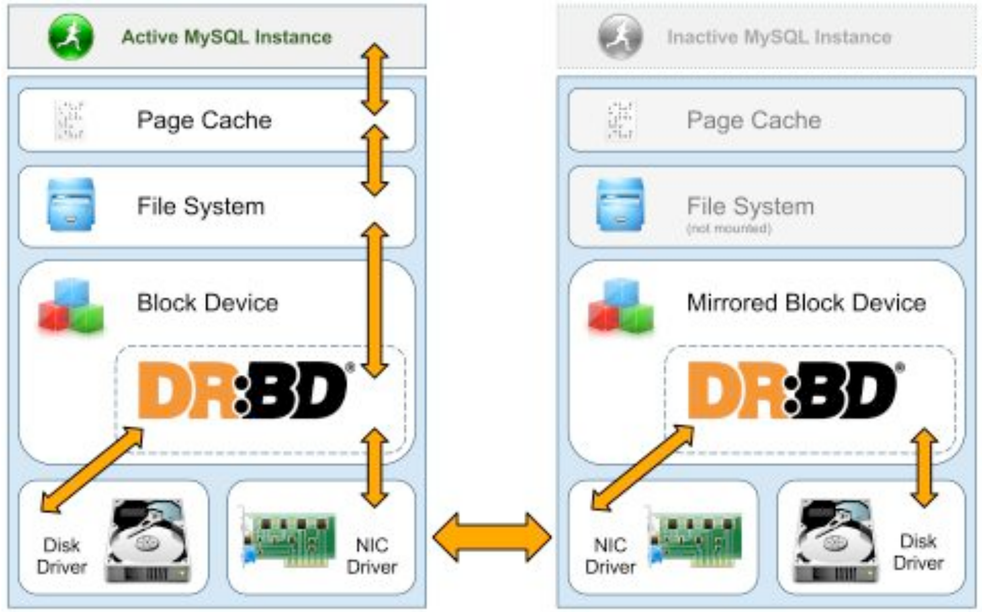
Shared-disk Solution

- Active – Passive replication
  - Failover requires MySQL crash recovery (and often file system crash recovery)
- Combined with Pacemaker/Heartbeat for automatic failover
  - Virtual IP most often used to fail over
- In theory the SAN is a SPOF

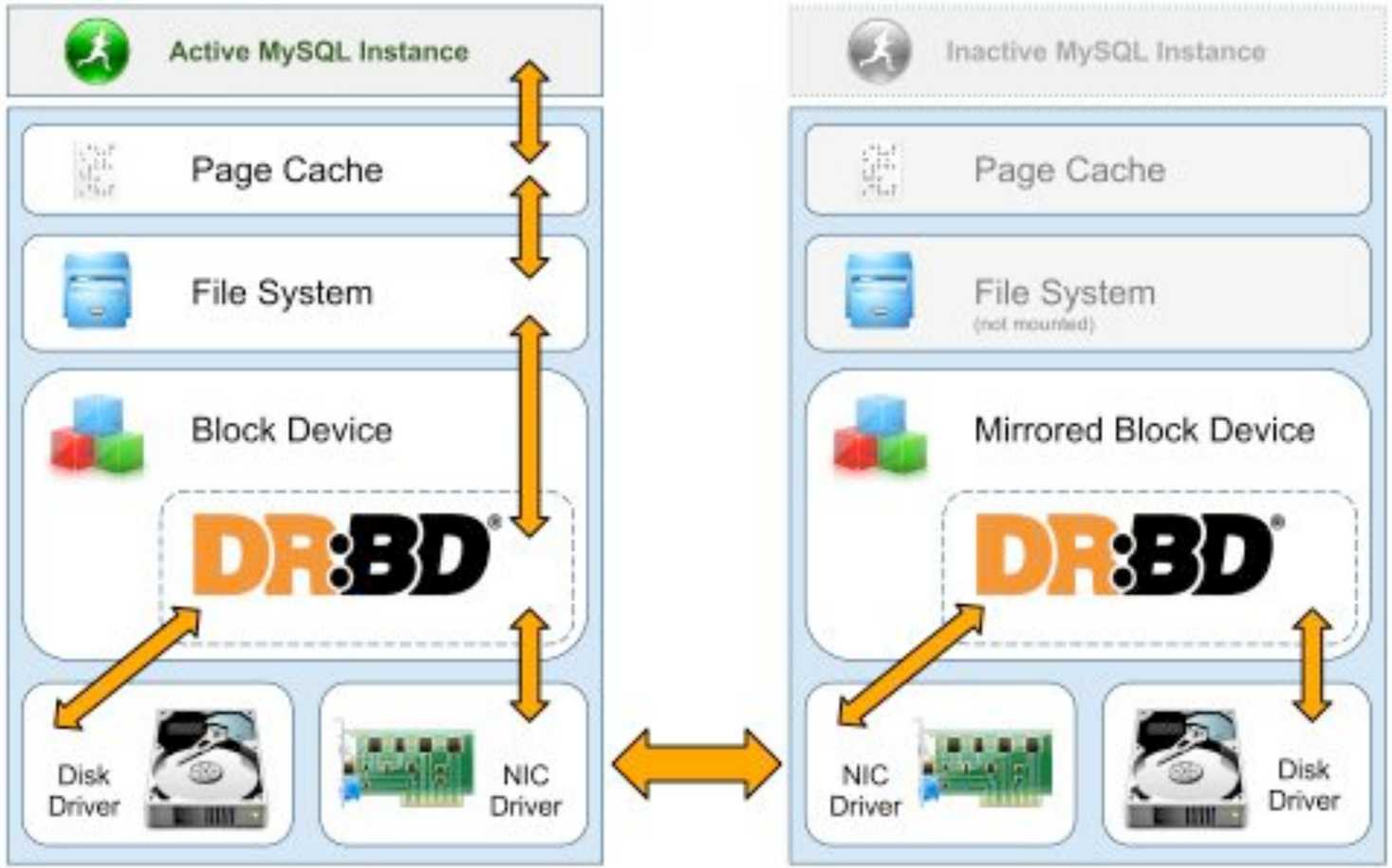
SkySQL

DRBD®

- Synchronous replication (three modes)
- Active – Passive replication
  - Failover requires MySQL crash recovery (and often file system crash recovery)
- Combined with Pacemaker/Heartbeat for automatic failover
  - Virtual IP most often used to fail over
- STONITH or other fencing mechanism needed to avoid split-brain scenarios
- Available on Linux



DRBD® - Architecture

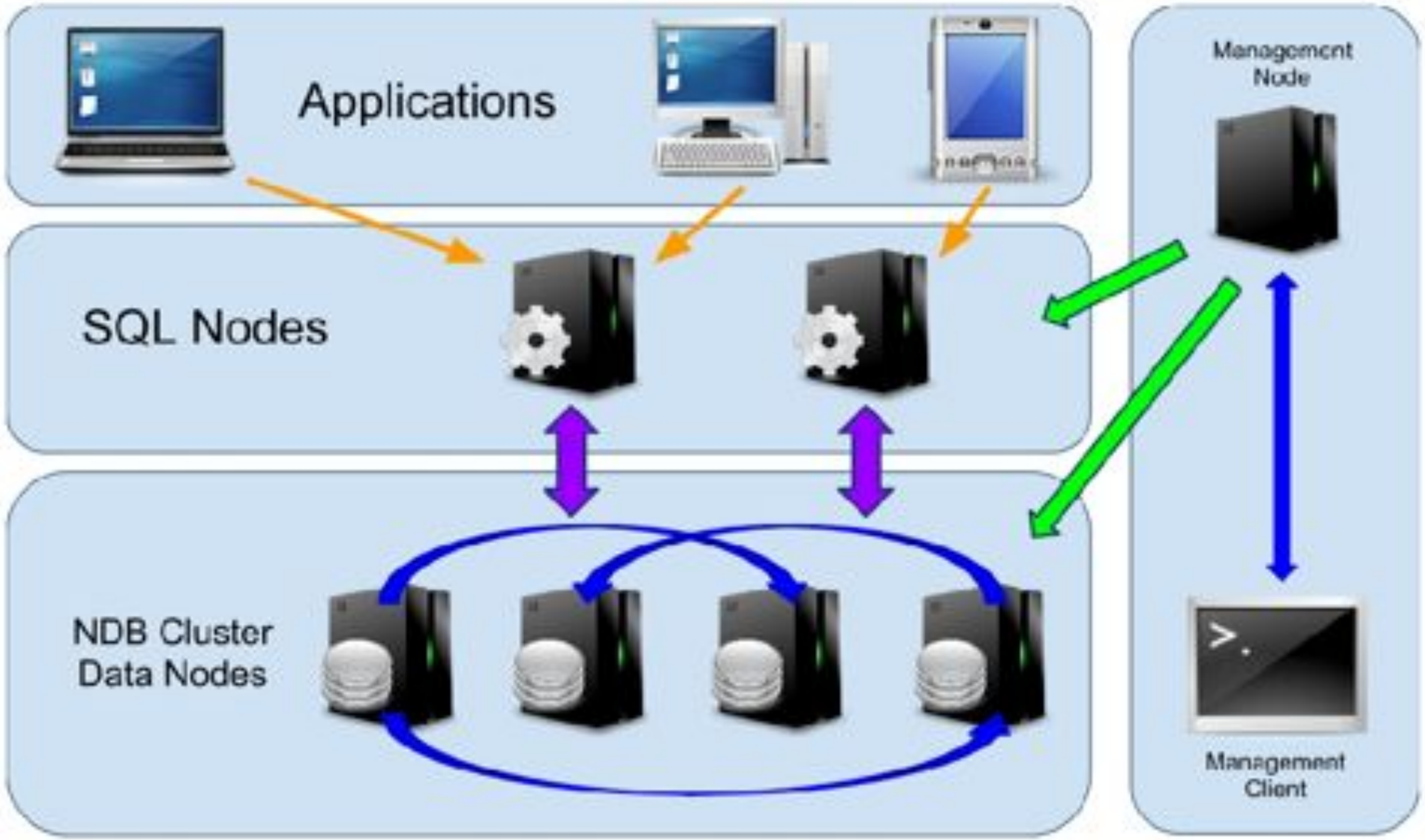


MySQL Cluster - Features

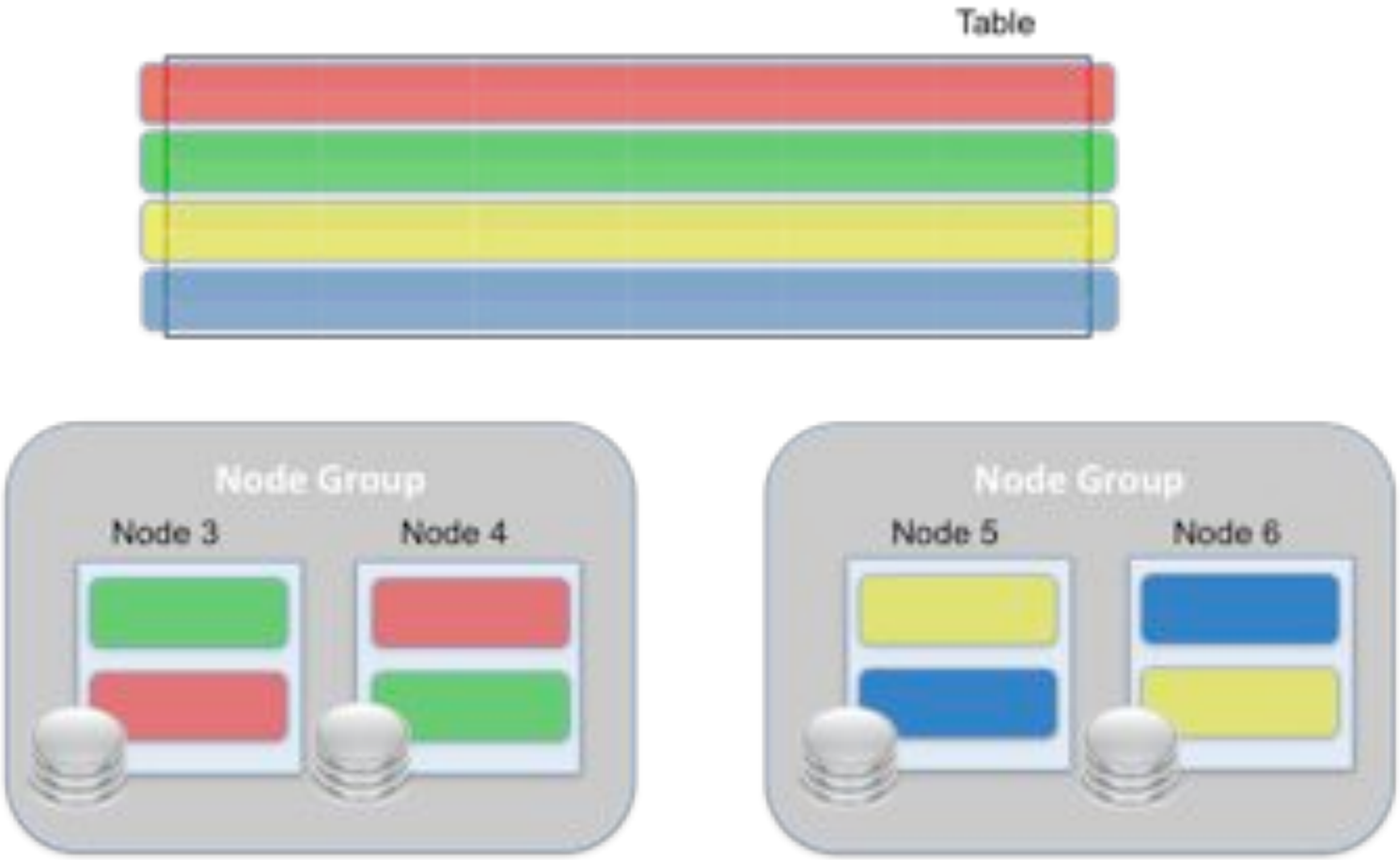
- Synchronous replication between nodes
  - Through Two-Phase Commit Protocol
- ACID transactions
- Row level locking
- Shared nothing architecture
  - No single point of failure
- Automatic failover
- In-memory storage
  - Some data can be stored on disk
  - Checkpointing to disk for durability
- Two types of indexes
  - Ordered T-trees
  - Unique hash indexes
- Online operations
  - Add node groups
  - Software upgrade
  - Some table alterations

SkySQL

MySQL Cluster - Architecture



MySQL Cluster - Partitioning



MySQL Cluster
Network partitioning protocol

Designed to avoid split-brain

1. Is at least one node from each node group present?
   - If not then the cluster cannot continue - shutdown
2. Are all nodes present from any node group?
   - If so then this is the only viable cluster - continue
3. Ask the arbitrator
   - The arbitrator decides which ”cluster” continues
   - If the arbitrator is not available the cluster will shutdown

SkySQL

Where does MySQL Cluster fit?

- High demands on availability (5 nines)
- You need write scalability
- You have 3 or more ”machines” available
- Where the queries and data model are simple
- When the data fits in memory
- When you have skilled people

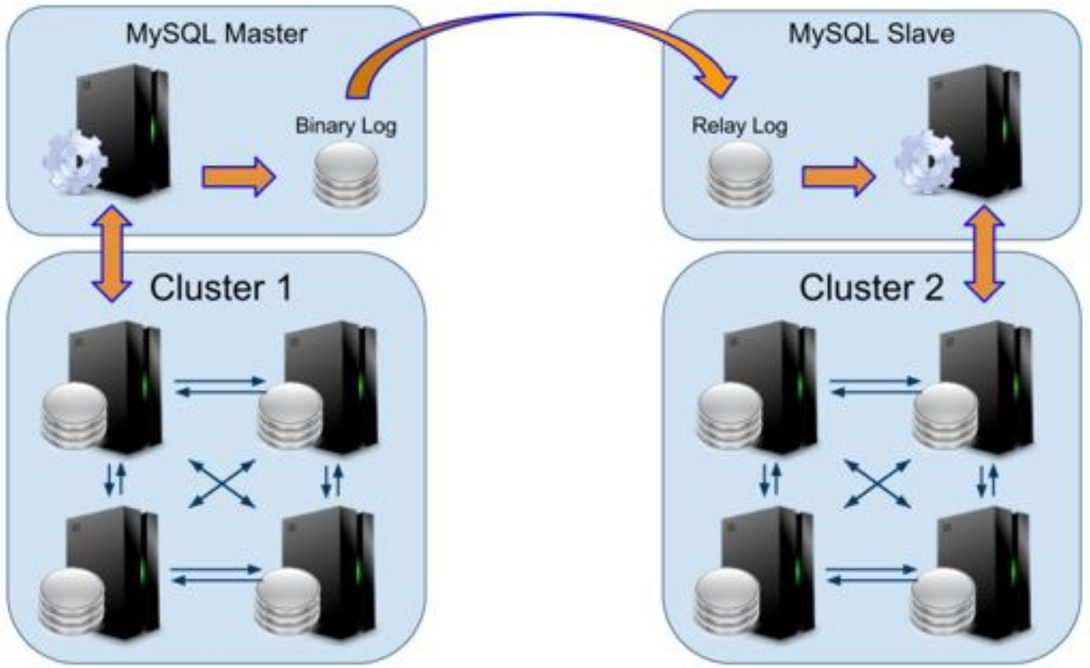
SkySQL

Geographical Replication

- For geographical (multi-site) redundancy synchronous solutions are often not desirable
- Standard Replication can be combined with any synchronous solution locally
  - MySQL Cluster
  - DRBD / shared disk
  - Galera

Geographical Replication
MySQL Cluster

- MySQL Cluster has additional features related to Geographical replication
- Conflict detection and automatic resolution
  - Several possible resolution methods exist
- Multiple replication channels possible
- Multi-source replication possible
- Binlog injection ensures the consitency of binlogs



SkySQL

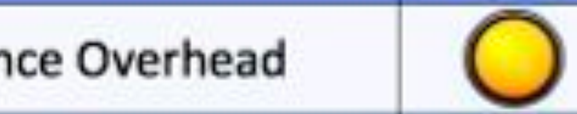
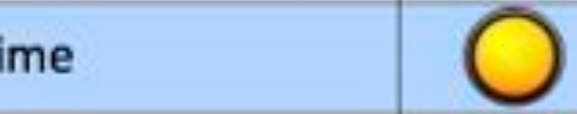
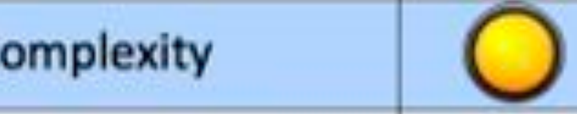
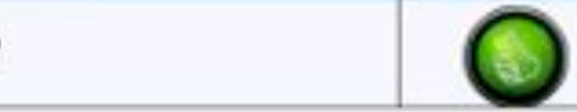
Quick Comparison Chart

	MySQL Replication	Shared Storage	MHA	DRBD	Galera
Asynchronous Replication	✓	-	✓	✗	✗
Synchronous Replication	✗	-	✗	✓	✓
No-Data Loss	✗	✓	✗	✓	✓
Active-Active	✓	✗	✓	✗	✓
Integrated Failover	✗	✓	✓	✓	✓
Management Interface	✗	✗	✓ (beta version)	✗	✓ (beta version)
All MySQL Storage Engines	✓	✓	✓	✓	✗
Monitor Extensions	✗	✗	✓	✗	✓
Multi-Master	✗	✗	✗	✗	✓

*

SkySQL

Quick Comparison Chart

	MySQL Replication	Shared Storage	MHA	DRBD	Galera
Performance Overhead					
Failover Time					
Failover Complexity					
Failback Complexity					
Configuration					
Administration					
Scalability					

SkySQL

Questions?

SkySQL

Thank you!

Joffrey Michaie

Joffrey.michaie@skysql.com

SkySQL Ab

www.skysql.com

www.facebook.com/skysql

www.linkedin.com/company/skysql

SkySQL