

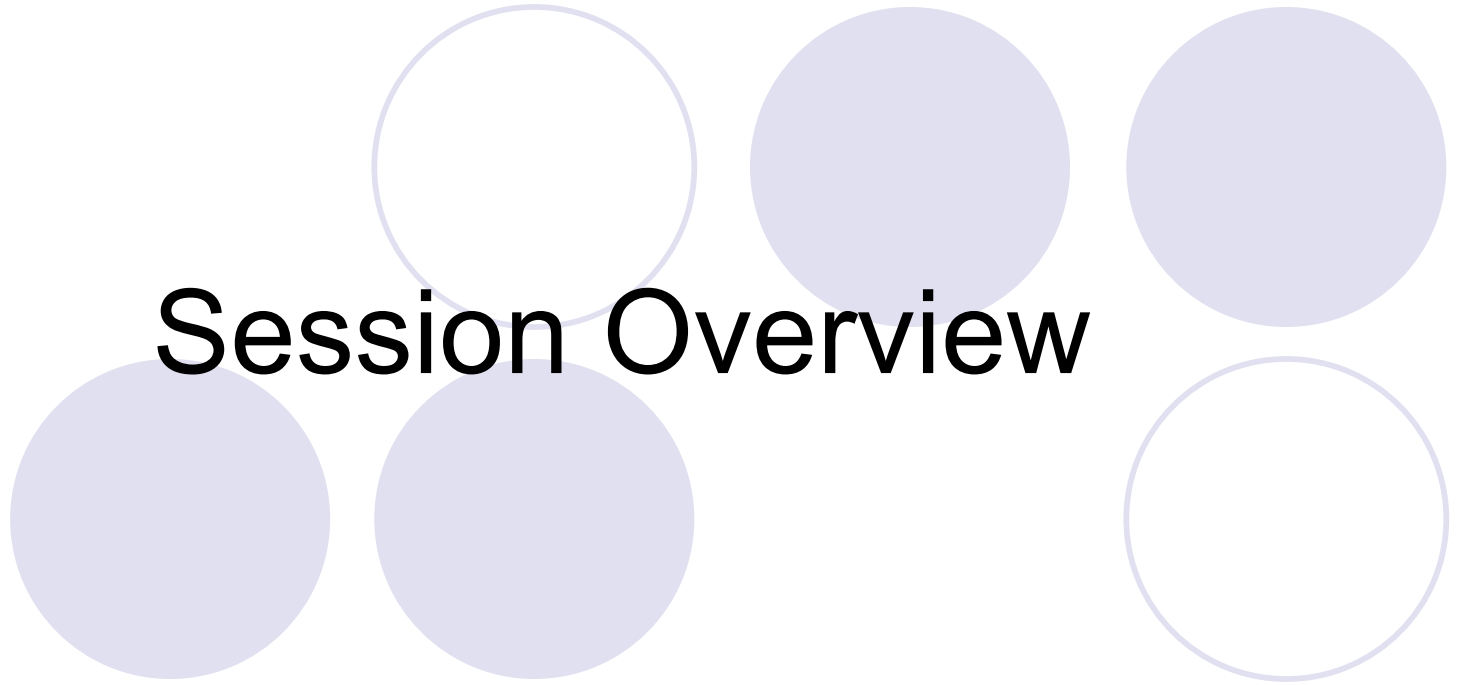


An Introduction to MySQL Cluster

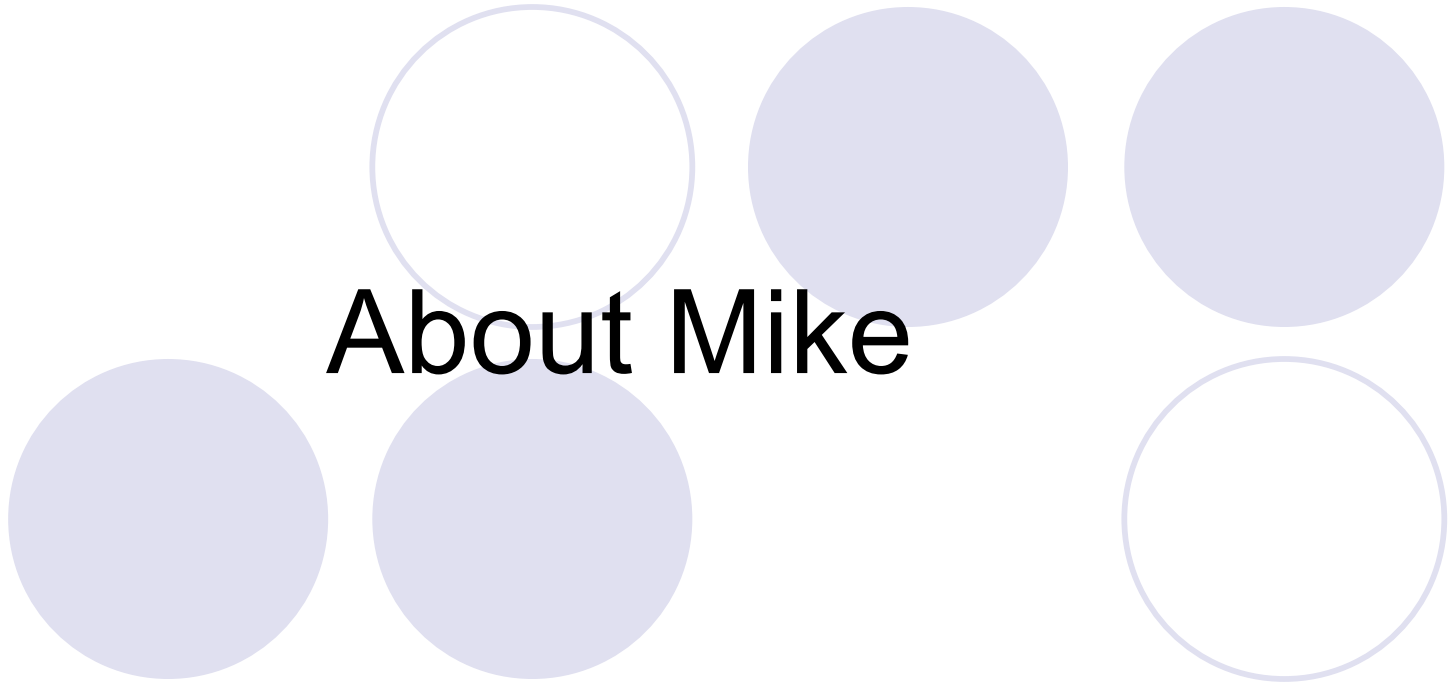
PHP Québec
March 31, 2006

Mike Kruckenberg
mike@kruckenberg.com
<http://mike.kruckenberg.com>

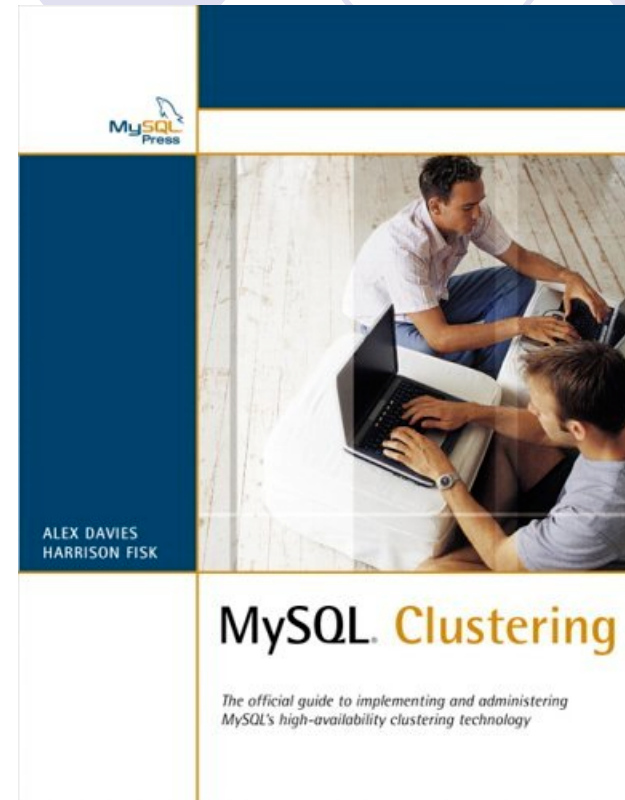
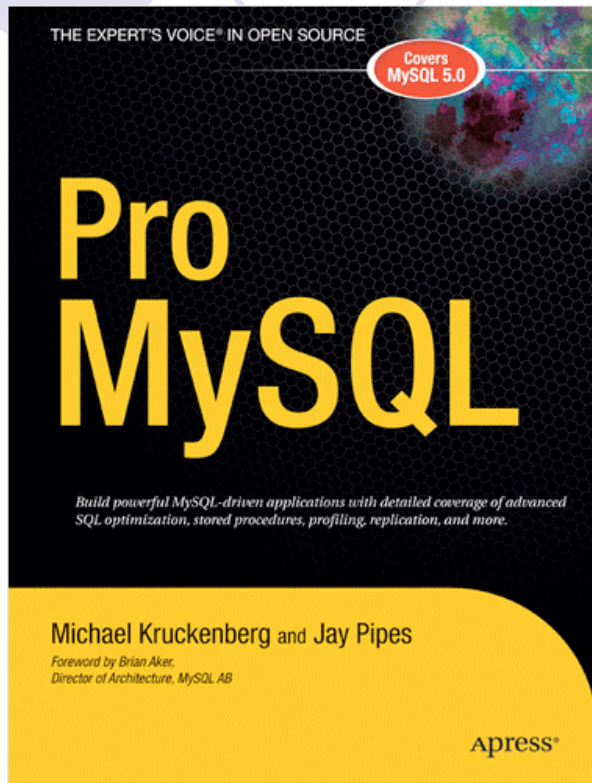
Session Overview



About Mike



Two Books





When to Think About Cluster

- Standalone server(s) overloaded
- Replication lag time unacceptable
- Failover process is inadequate
- Uptime is critical (really)
- Others...

Executive Overview



- MySQL Cluster is a fault-tolerant, clustered, distributed database
- Acquired from the telecom industry in 2003
- Integrated, and released in 2004
- Provides synchronous data replication

Executive Overview (cont)



- No single point of failure
- Share nothing architecture
- Automatic failover & recovery
- Runs on commodity hardware

Technical Overview



- Is just another MySQL storage engine

```
CREATE TABLE t1 (id INTEGER) ENGINE=NDB;
```

- Requires multiple machines for true redundancy
- Data written in two-phase commit

Technical Overview (cont)



- Transactional
- In-memory data storage (pre-5.1)
- Data written to disk at checkpoints and shutdown
- Communicates via TCP/IP or other
- Rolling upgrades

Anatomy of a MySQL Cluster

- Three Types of Processes

- Management Node (ndb_mgmd)

- Storage Node (ndbd)

- SQL Node (mysqld)

- Spread across multiple machines

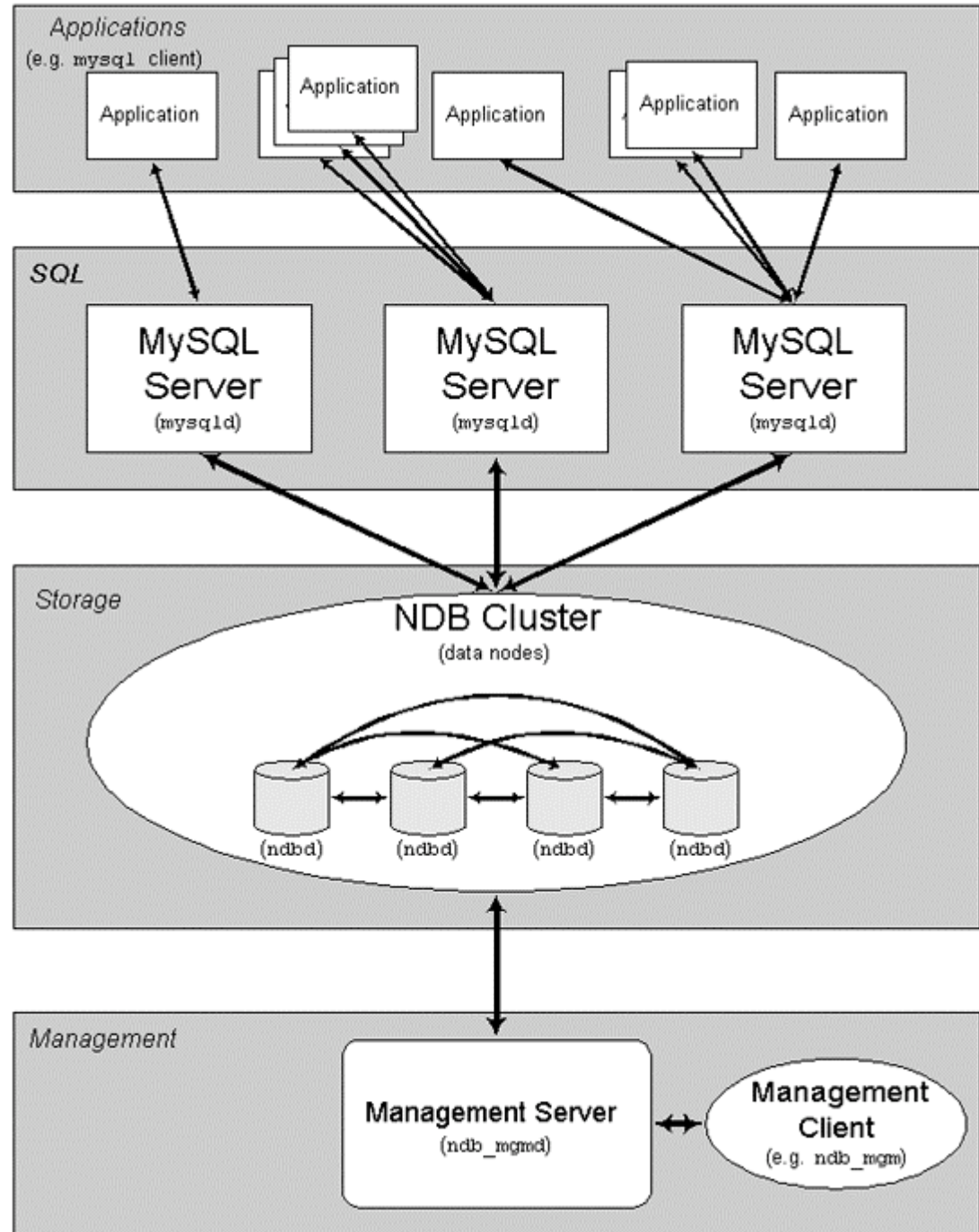
- Storage node count based on amount of data and number of replicas

`(data MB * no. of replicas * 1.1) / nodes = MB/node`

ie. `(5000MB * 2 replicas * 1.1) / 4 nodes = 2750MB`

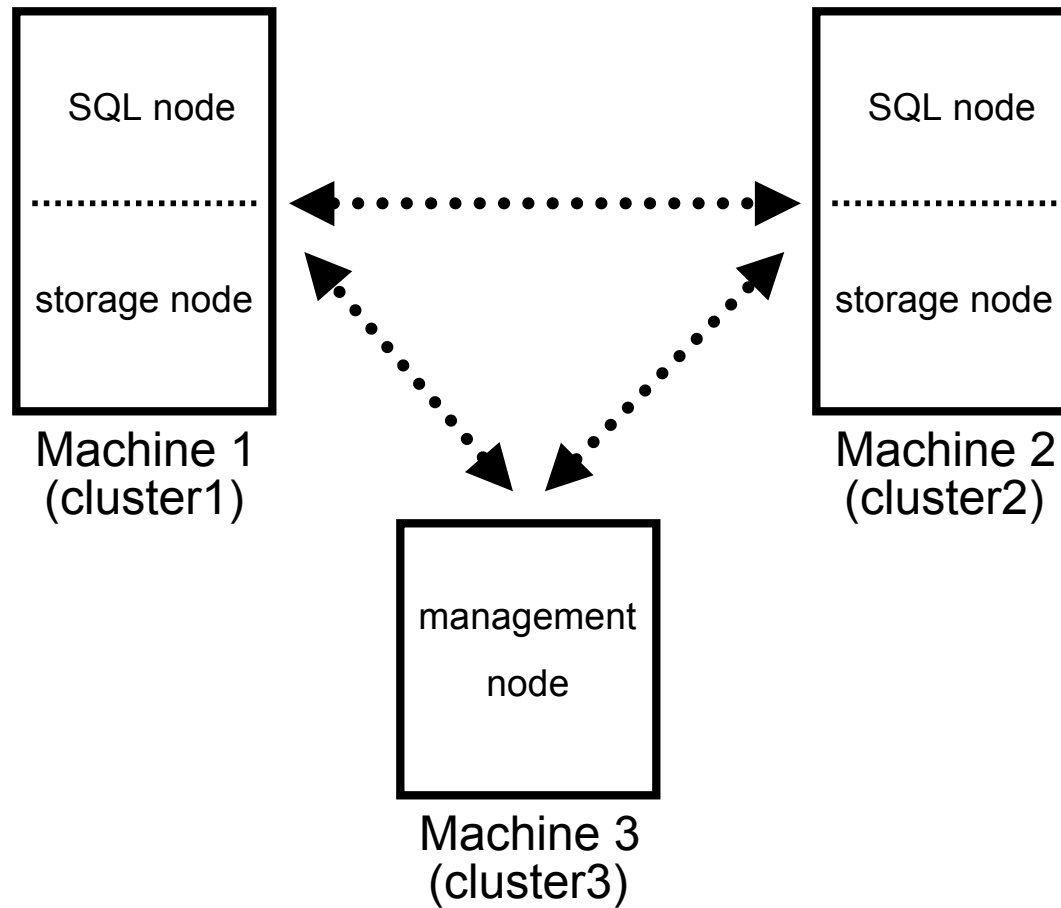
Cluster Diagram

(Courtesy MySQL AB)



Live Demo

Demonstration of setting up three-machine cluster





Notable Configuration Options

- All Nodes

- ndb-connectstring

- Management Node

- noofreplicas - NDBD

- datamemory - NDBD

- indexmemory - NDBD

- datadir – NDBD

- hostname – MYSQLD

- arbitrationrank – MYSQLD or NDB_MGMD



Management Commands

- SHOW – shows status
- SHUTDOWN – stops all nodes
- <id> STATUS – show status of node
- <id> START – start individual node
- <id> STOP – stop individual node
- START BACKUP – all nodes backup their data
- ENTER SINGLE USER MODE <id>
- CLUSTERLOG ON|OFF <LEVEL> – start/stop logging at different detail levels
- <id> CLUSTERLOG ... - change logging for node
- QUIT – exit the management client

Detecting Node Failure



- Loss of communication or heartbeat failure
 - Heartbeat performed by each node
 - All nodes notified on failure
- Remaining Nodes Take Over
 - Arbiter used to decide if data is complete
 - Failed nodes attempt restore
- 100 milliseconds for TCP/IP failover
- System failure requires restart using logs



Performance

- Depends on hardware, network
- 10,000 transactions/second on 2 node cluster
- 100,000 transactions/second on 4 node cluster
- Memory is faster than disk



Be Aware

- SQL nodes do not failover
- NDB does not support foreign keys
- Limitations on data types
- Transactions will abort on storage node failure
- Plan memory allocation carefully
- Can't add new nodes without restart

Be Aware (cont)



- Servers must be same architecture
- Network traffic not encrypted
- Non-NDB tables not clustered
- Limited OS availability
- Can consume lots of CPU and network

Future



- 5.0 (now)
 - Condition push-down (w/EXPLAIN support)
 - Query cache enabled
 - Metadata objects increased from 1600 to +20K
- 5.1
 - Disk-based storage (keys in memory)
 - Replication support
 - Partitioning
- Beyond
 - Dynamic provisioning
 - MS Windows (are working on it)
 - MySQL Administrator GUI support

Thank You



- Slides
 - <http://mike.kruckenberg.com>
- MySQL Docs
 - <http://dev.mysql.com/doc/refman/5.0/en/ndbcluster.html>
- Quick Setup Docs
 - <http://dev.mysql.com/doc/refman/5.0/en/mysql-cluster-quick.html>
- Architecture Overview
 - <http://www.mysql.com/why-mysql/white-papers/cluster-technical.php>
- Bredbandsbolaget Case Study
 - <http://www.mysql.com/why-mysql/case-studies/mysql-b2-casestudy.pdf>