MySQL Cluster Introduction

Harrison Fisk
MySQL User Conference 2005
April 18-21, 2005
Overview

- Shared-nothing clustering
- Automatic Partitioning
- Automatic Synchronous Replication
- Main-memory engine currently
  - Disk based in progress
- Integrated as a new storage engine
- Part of Max release in MySQL 4.1
  - Only some platforms supported at this time
    - Linux, Solaris, AIX, HP-UX
    - More in future (Windows)
Shared-Nothing

• No single point of failure
  – Any single (or multiple depending on setup) server can fail avoiding a total cluster shutdown
• No extra special hardware required (SAN, etc.)
• Everything is independent to each node
Data Partitioning

- Data is horizontally partitioned automatically over the data nodes
- Each node in charge of only a piece of the data in a table
- Data reading can be done in parallel across nodes
- Hashing based on the **PRIMARY KEY**
- Example: 4 data nodes would have 4 data fragments with a quarter of the data each
Synchronous Replication

• Data is automatically replicated to 2 or more (NoOfReplicas) servers
• High availability
• Guaranteed at COMMIT time to be present in multiple nodes
• Automatic takeover in the event of node failure
Main Memory System

• Data and indexes are stored totally in memory
• High performance
• Data only written to disk asynchronously
  – Only required in event of a complete cluster shutdown
• Amount of memory across all nodes restricts database size

Example:

4 data nodes with 16GB of RAM with \texttt{NoOfReplicas} set to 2 would give:

\[ 4 \times 16\text{GB} = 64\text{GB} / 2 = 32\text{GB} \text{ of RAM available} \]
Other HA Features

• Online backup
  – Possible to take a backup without setting any locks or otherwise interfering with a running server

• Rolling software upgrades
  – Possible to upgrade software across cluster without shutting down entire server in many cases

• Online Configuration Changes
  – Able to change many configuration settings by doing a rolling upgrade as well
  – Some (NoOfReplicas, number of data nodes, etc..) can not be done without a full cluster restart
    • May also require re-initialization in certain cases
MySQL Cluster Nodes

- Three node types
  - MySQL Server (*mysqld*)
  - Data nodes (*ndbd*)
  - Management nodes (*ndb_mgmd*)
- Can run on same or different servers
MySQL Node

- Normal MySQL Server with cluster enabled
- Create tables with \texttt{ENGINE=NDBCluster}
- Possible to use other storage engines at same time
- Regular SQL that is used with other storage engines
- Common to run on the same system with Application Server
Data Nodes

- In charge of the actual data storage
- Handle replication/partitioning/failover transparent to applications
- Has to be a multiple of `NoOfReplicas`
- Maximum of 48 per cluster by default in binaries
Management Node

• Node in charge of storing cluster configuration
  – Central configuration location (config.ini)
  – Required to be running for node startup
• Also handles other centralized cluster ideas
  – Starts and controls system backups
  – Monitoring node status
  – Central Logging
  – Arbitration
• Not involved at all in actual data storage and query execution
Cluster Components

- Application
- MySQL Server
- MySQL Server
- MySQL Server
- MGM client
- MGM Server
- NDB Cluster (Data nodes)
- DB
- DB
- DB
- Storage
- SQL
Physical Layout

NDB Cluster
(Data nodes)

App

mysqld

App

mysqld

App

mysqld

App

mysqld

App

mysqld

App

mysqld

ndbd

ndbd

ndbd

ndbd

ndb_mgmd

App

mysqld

App

mysqld
Cluster Setup

- Install software
  - `max` release in MySQL 4.1+
- Create `config.ini`
  - Defines nodes/settings/etc…
- Start `ndb_mgmd`
- Start `ndbd`
  - Use `--initial` for first startup
  - Specify `--ndb-connectstring` to point to `ndb_mgmd`
- Start `mysqlld`
  - Requires `--ndbcluster` and `--ndb-connectstring`
- Create tables with `ENGINE=ndbcluster`
config.ini

shell> cat config.ini

[ndb_mgmd]
hostname=192.168.0.5

[ndbd default]
NoOfReplicas = 2
Datadir = /var/lib/mysqlcluster/

[ndbd]
hostname = 192.168.0.6

[ndbd]
hostname = 192.168.0.7

[mysqld]
[mysqld]
[mysqld]
[mysqld]