

# 基于PXC的 MySQL高可用架构探索

周彦伟  
2014.10

# 个人简介

- ▶ 周彦伟去哪儿
  - ▶ 去哪儿网数据库总监
    - MySQL、Redis、Hbase、SQL Server、Oracle，中间层和源码开发
    - 招人是必须的
  - ▶ 中国MySQL用户组（CMUG）
    - [acmug.com](http://acmug.com)
    - 最纯粹的MySQL社区组织

# 从HA谈起

HA (可用水平)	T (每年可中断时间)
99.9999%	< 1 分钟
99.999%	< 5.3 分钟
99.99%	< 53 分钟
99.9%	< 8 小时 46 分
99%	< 87 小时 36 分

- ▶ QUNR Q2营收 4亿 51/s? 5000/s

# HA最容易忽视的问题

- ▶ 备份
  - 冷备
  - 热备
  - 逻辑备份
  - 物理备份
  - 容灾备份
- ▶ 没有备份 HA = 0%

# MySQL HA

- ▶ MySQL replication
  - M-S
  - MMM
  - MHA
- ▶ 异步复制和数据修复是软肋

# MySQL HA

- ▶ DRBD (Distributed Replicated Block Device)
  - 资源浪费
  - MySQL recovery, 故障迁移时间成本高

# MySQL HA

- ▶ shared storage
  - Redhat Cluster Suite
- ▶ 数据集唯一
- ▶ 对DBA来说维护复杂

# MySQL HA

- ▶ MySQL Proxy
  - 官方版本性能低，多年不更新
  - Proxy本身是瓶颈
- ▶ 二次开发版本可堪一用，维护成本高

# MySQL HA

## ▶ MySQL Cluster

- share nothing
- 基于内存
- 维护成本高
- 复杂查询性能受限

# HA去哪儿？

- ▶ MMM
- ▶ <http://mysql-mmm.org/>

## Multi-Master Replication Manager for MySQL

**NOTE:** By now there are some good alternatives to MySQL-MMM. Maybe you want to check out  Galera Cluster which is part of  MariaDB Galera Cluster and  Percona XtraDB Cluster.

# Galera

[HOME](#)[PRODUCTS](#)[DOWNLOADS](#)[SUPPORT](#)[COMMUNITY](#)

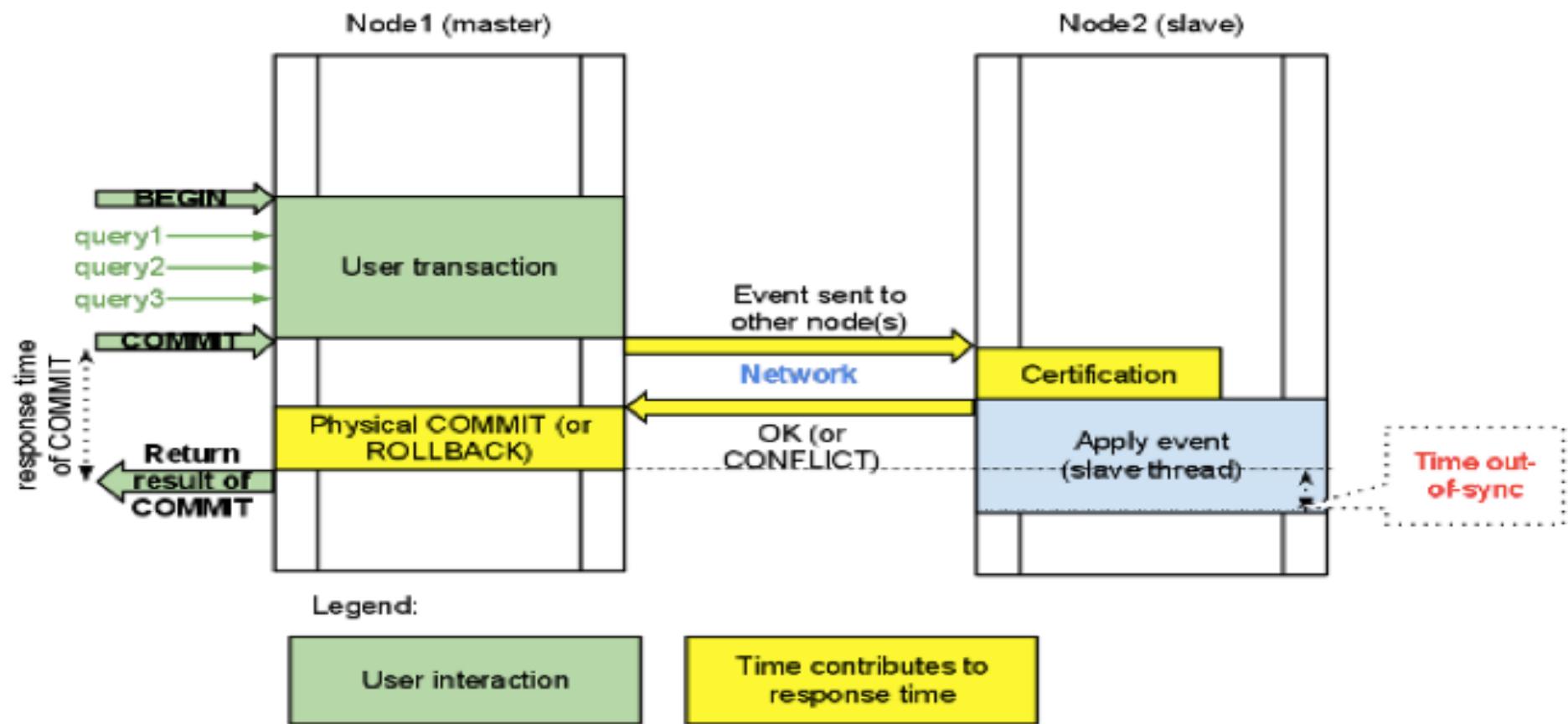
## THE WORLD'S MOST ADVANCED OPEN SOURCE DATABASE CLUSTER

Galera Cluster for MySQL is an easy-to-use high-availability solution with high system up-time, no data loss, and scalability for future growth

# Galera



# Galera



# Galera

- ▶ multi-master
- ▶ 准同步复制
- ▶ 行级并行复制
- ▶ 节点数据维护 SST & IST
- ▶ 自动节点管理
- ▶ innodb & row statement

# Percona Xtradb Cluster (PXC)

- ▶ Percona出品
- ▶ 基于Percona Server
- ▶ 封装了Galera，更易用
- ▶ 社区活跃

# 怎么访问DB?

- ▶ VIP (LVS, haproxy)
- ▶ MySQL-proxy
- ▶ API

# q-db-pool

- ▶ 可配置连接池
- ▶ 可配置连接池参数
- ▶ 智能扩展连接数
- ▶ 自动平滑切换链接
- ▶ 读写分离
- ▶ 负载均衡
- ▶ 自动重连

# 配置和通知

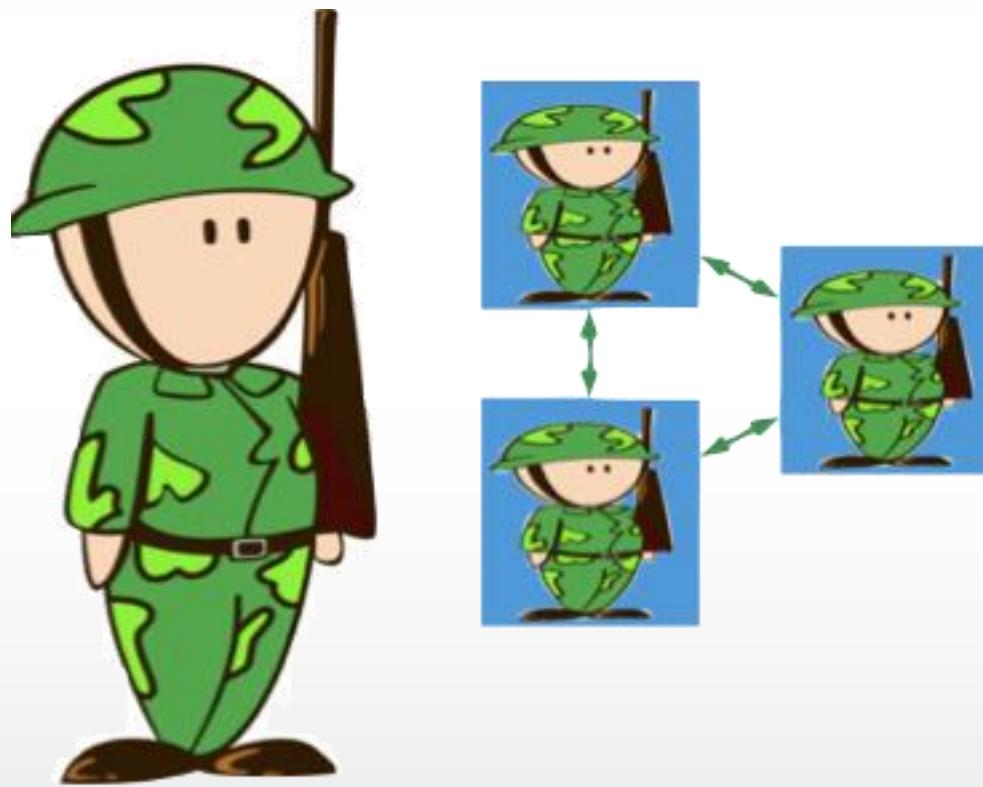
- ▶ XML+ICE
- ▶ Zookeeper
- ▶ PXC+Zookeeper



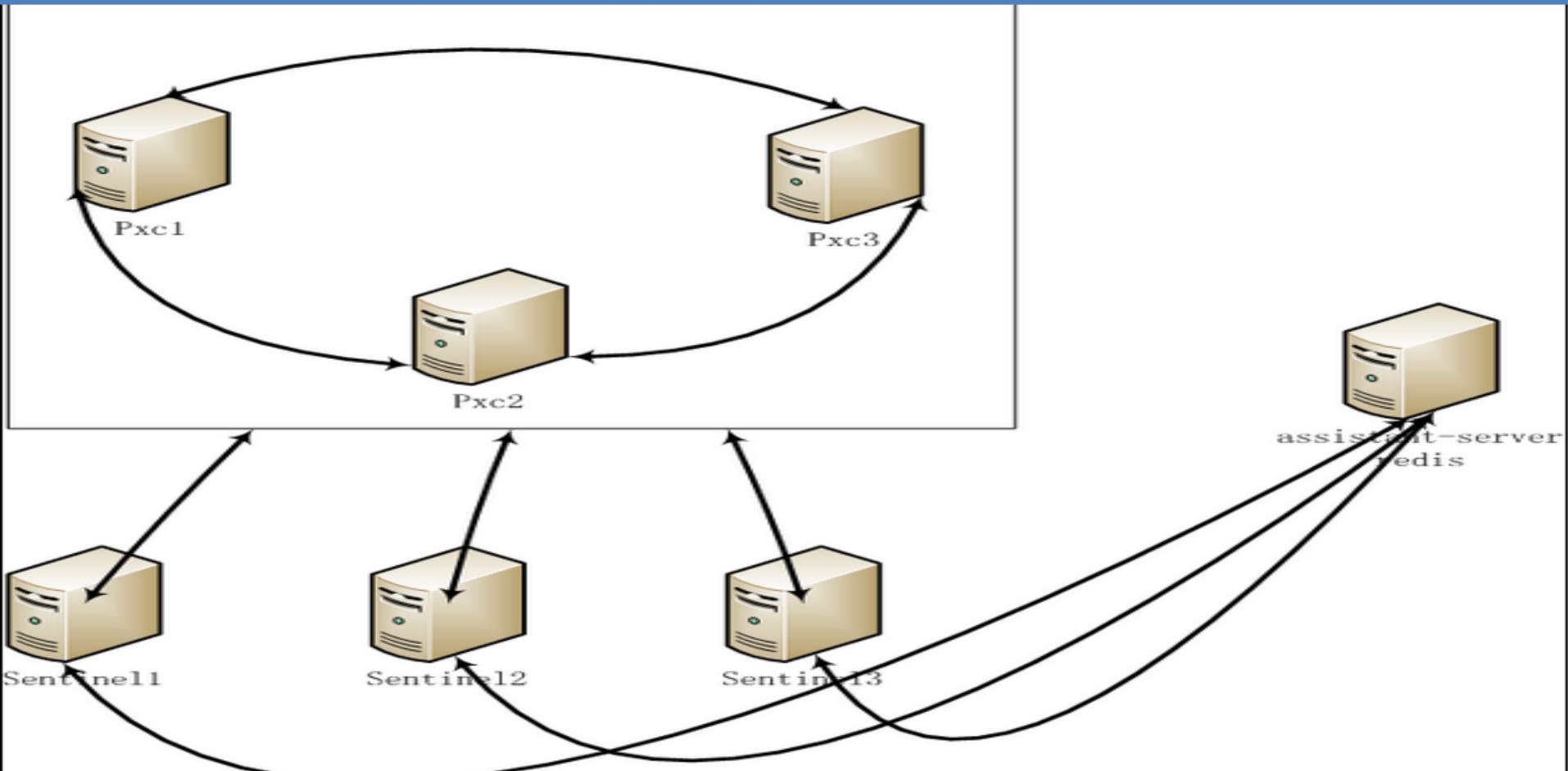
# 故障检测

- ▶ ps aux|grep mysql
- ▶ nagios
- ▶ mmm-monitor
- ▶ mysql-sentinel?

# mysql-sentinel



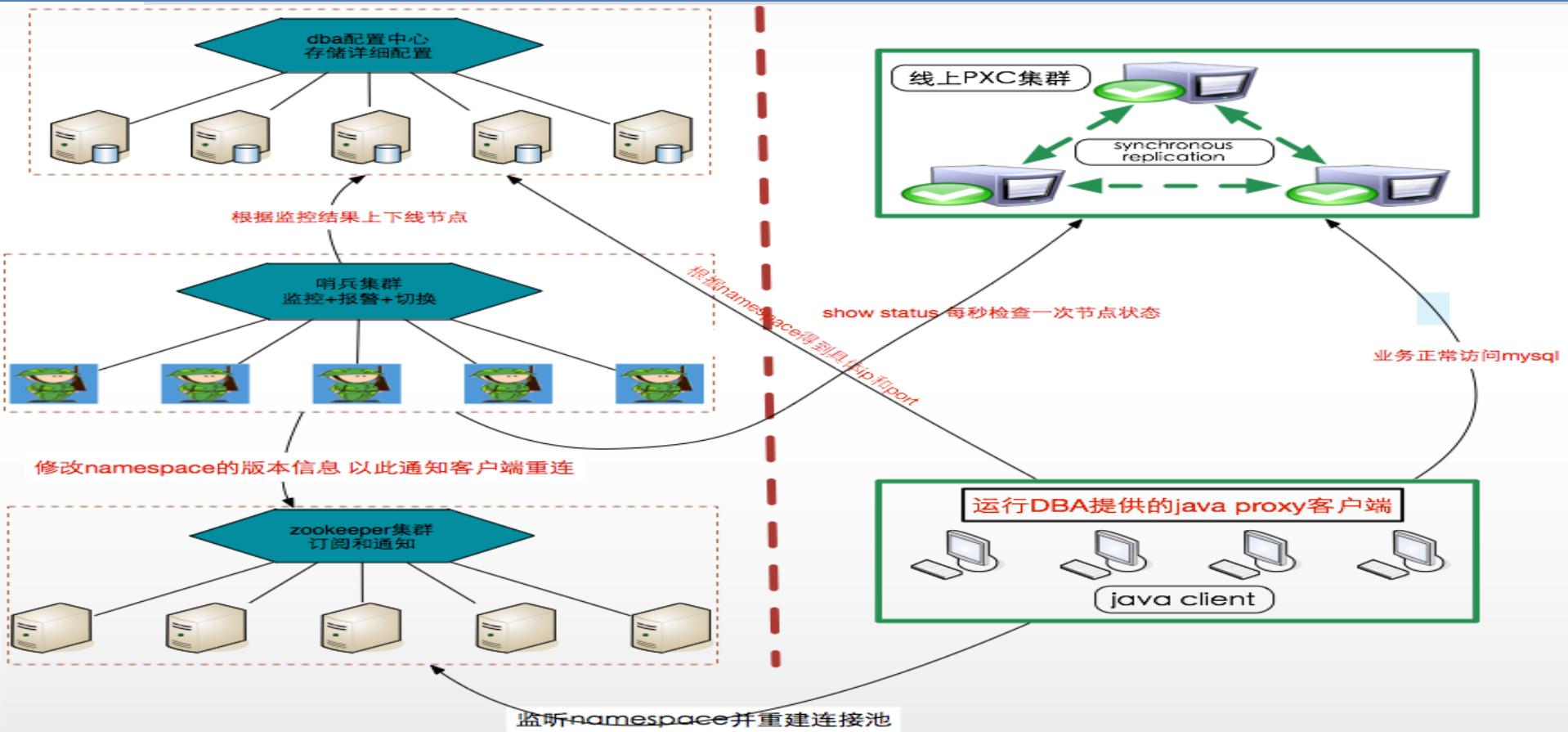
# mysql-sentinel



# mysql-sentinel

- ▶ 监控mysqld和galera信息
- ▶ 分布式选举
- ▶ 选举leader，并由leader通知zk
- ▶ 报警通知和配置切换API
- ▶ 交互式操作工具

# 最终的样子



# 注意事项

- ▶ Galera性能损失
- ▶ 避免多点写入
- ▶ 目前仅支持java
- ▶ 多个组件需要修改源码



招人不会结束（MySQL、Redis、Hbase、Java）  
[zhouyanwei@gmail.com](mailto:zhouyanwei@gmail.com)