去哪网数据库架构发展历程

Qunar数据库架构师

黄勇

自我介绍

2007~2011, Oracle DBA：智联、淘宝

去IOE大潮下的改变

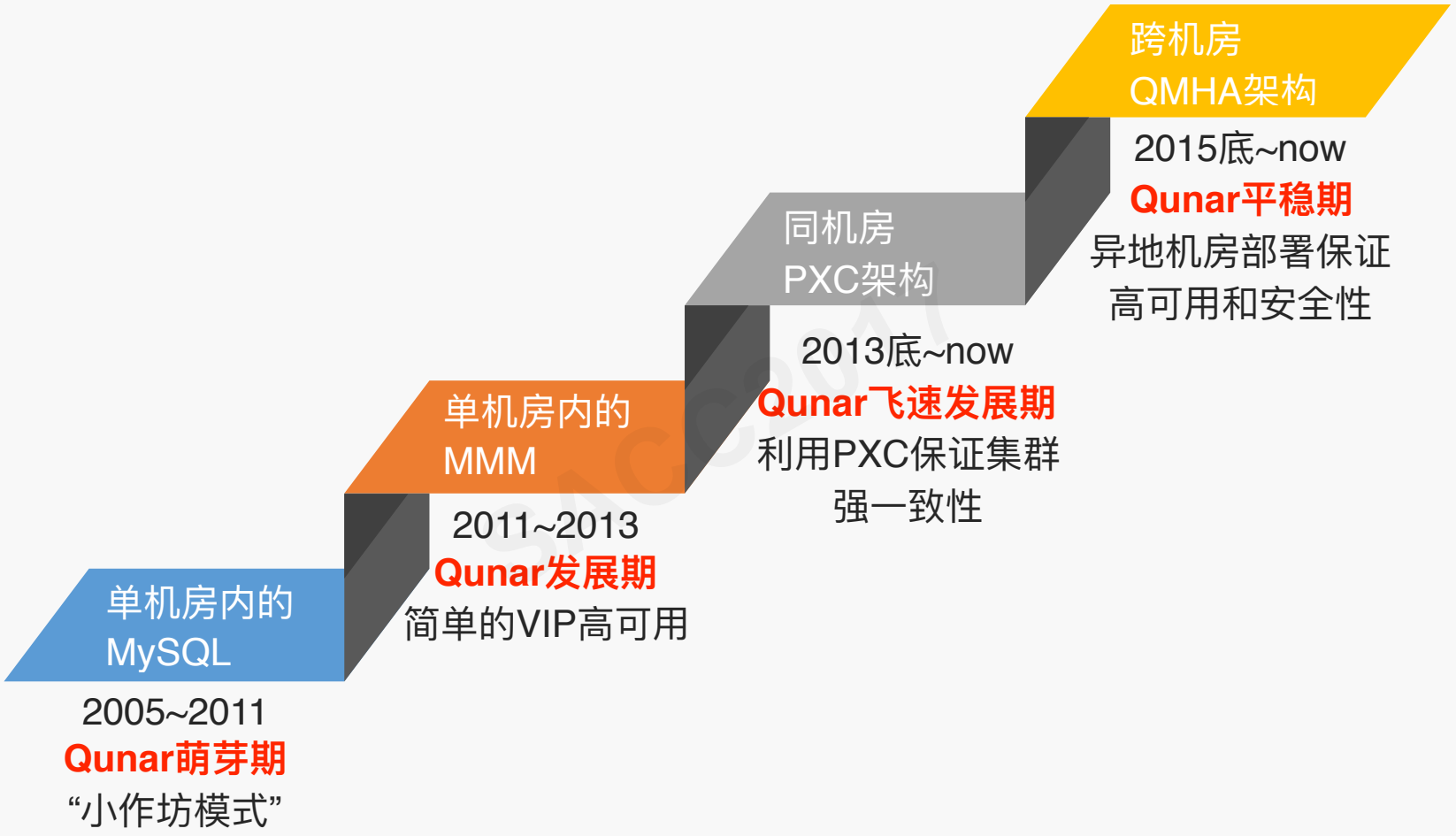
2011~now, MySQL DBA：百度、去哪

thunderbird.huang@gmail.com

wx: elnino_1114

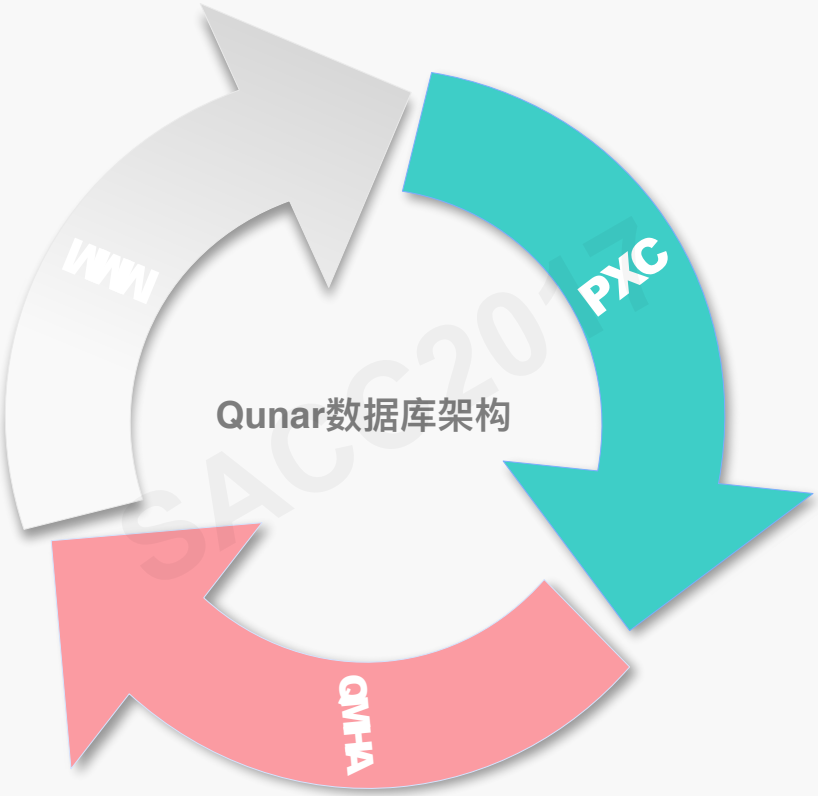
Contents

Qunar数据库的四个时代

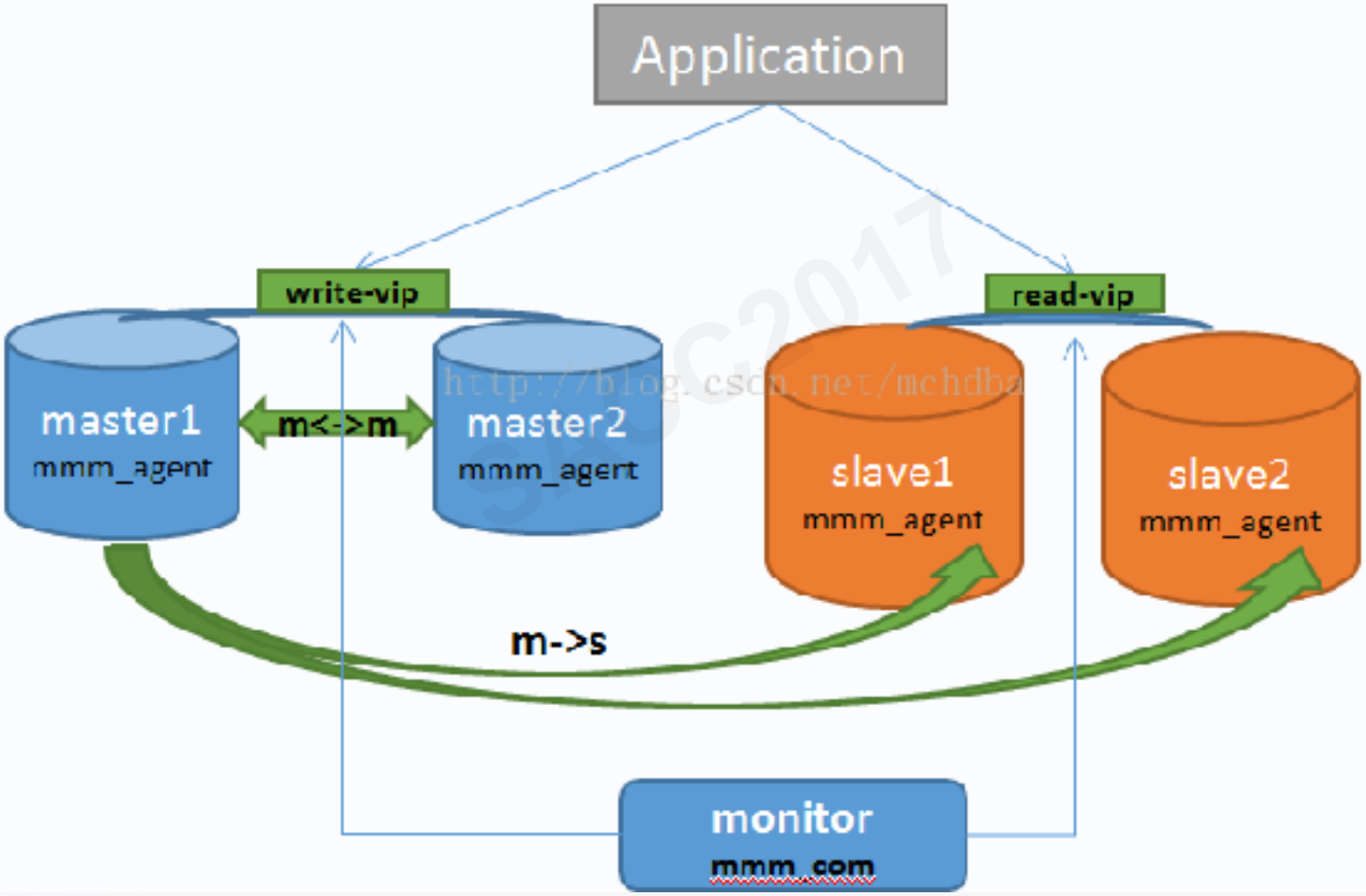


Qunar数据存储时间历程大事记

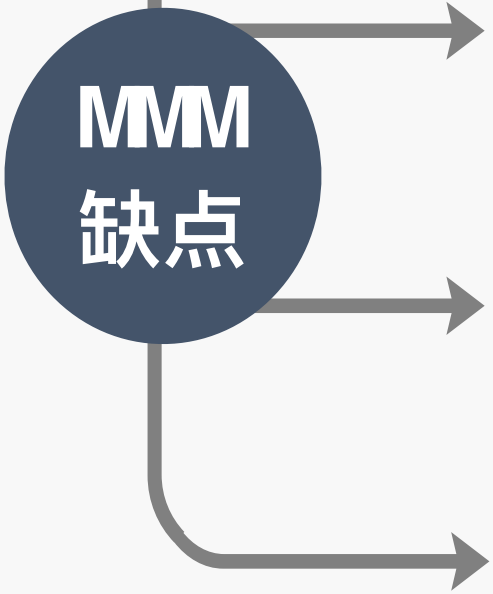
Qunar数据库架构的组成



MMM的基本架构



MMM架构的缺点



运维复杂

绑定VIP
部署和修改配置文件
周边监控工具匮乏

网络分区

Master“假死”导致误切换
数据库双写，导致数据错乱
VIP没有漂移或者漂移失败

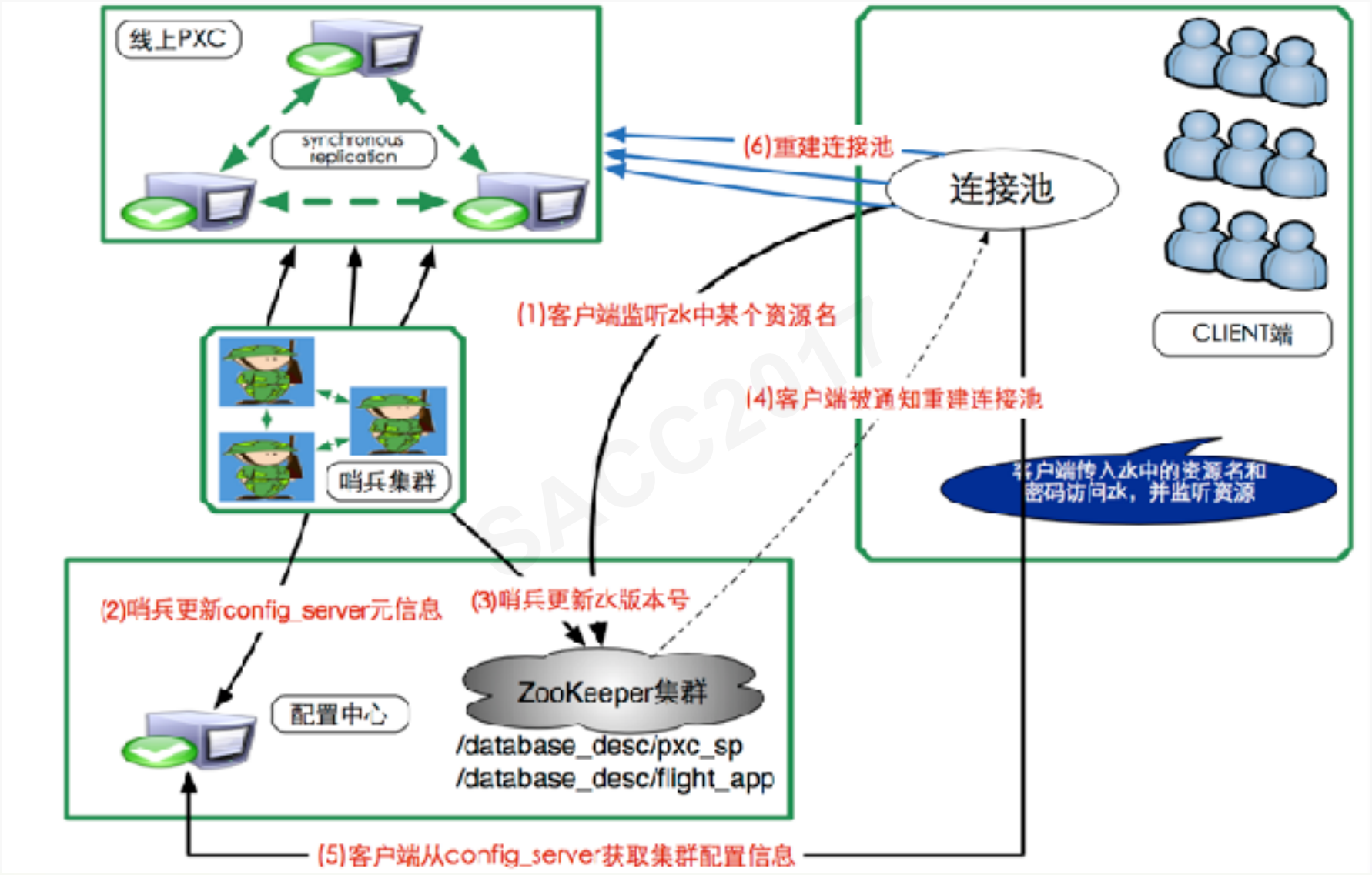
无法跨网段

VIP不能跨网段
VIP不能跨机房
机房容灾根本无从谈起

新特性的不支持

2012年已经停止版本更新
MySQL5.6以上版本新特性的不支持
落后的高可用无法匹配新技术的发展

PXC的基本架构



PXC架构特点

自动切换

自动failover

手动switchover

读写分离

读写分离

负载均衡

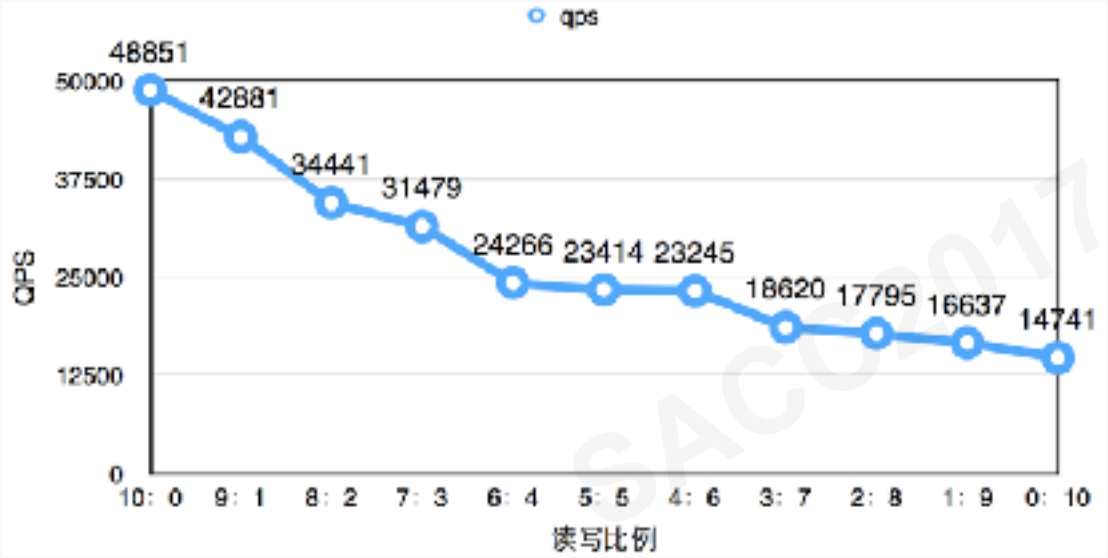
全局服务

namespace服务

全局唯一、透明

扩容、迁移和升级

PXC的性能

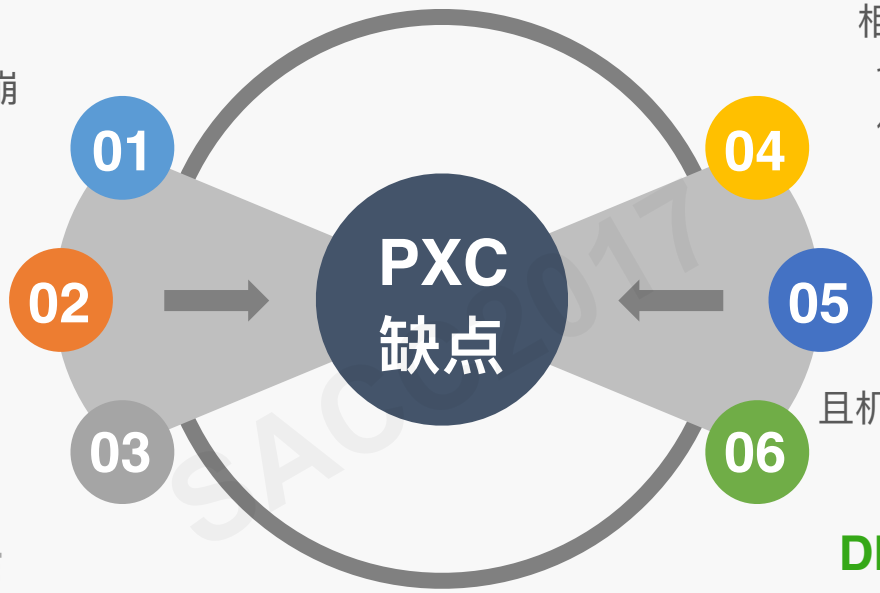


PXC单节点读取可达5W qps

PXC单节点写入可达15K qps

以7:3的读写比，单节点可达3W qps

PXC的缺点



Flow Control

节点间机器木桶短板
流量控制，客户端容易雪崩

多节点写入

相互校验，写入性能下降
切换时不影响前端写入
但尽量不要长时间多写

大事务

大事务和密集事务
PXC节点压力高，fc产生

集群节点跨机房

机房间网络延迟高
影响客户端QPS
且机器节点越多QPS影响越大

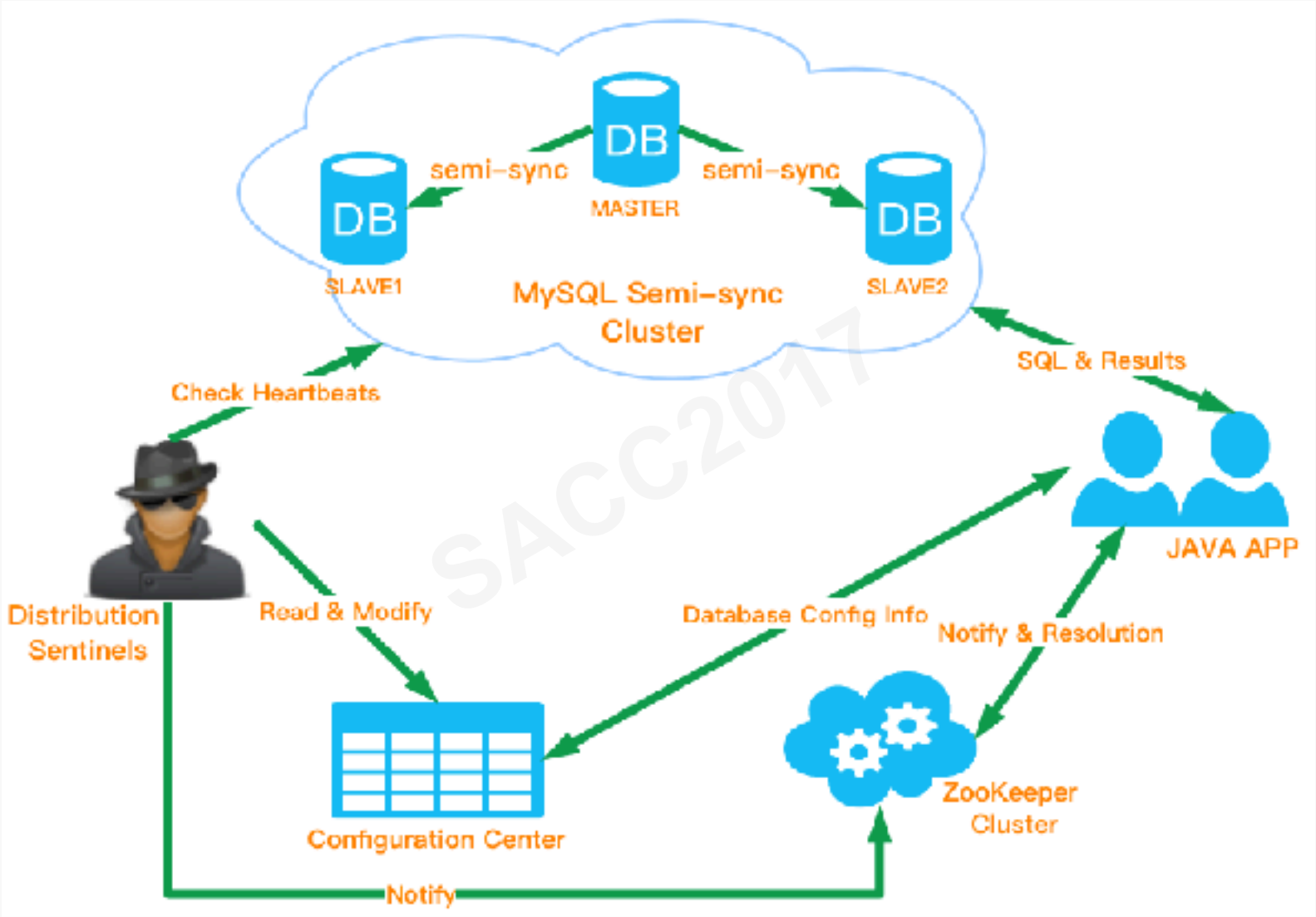
DDL操作

DDL杀死其他事务
DDL不能取消

DBA运维起点高

PXC和MGR等新兴结构
DBA学习成本高
长期的学习和经验

QMHA的基本架构



QMHA的技术特点

GTID

GTID易于维护和切换
主从节点间可知数据差异

Semi-Sync

提高数据节点一致性的同时
提高集群安全性和可用性
多线程复制
且可以跨机房和网段部署

SentineId

分布式哨兵
减少误切换和网络分区
raft算法，自动切换

ZooKeeper

全局namespace
通知客户端更新配置

QMHA/PXC解决的问题

无网络分区

多机房的分布式哨兵判断
MySQL实例的健康情况

0事务丢失

在failover和switchover时，
没有事务丢失
且PXC的集群数据强一致性，
QMHA的弱之但性能较好，而且
机器越多同步越快

集中配置管理

后台数据库配置中心存储
和维护线上所有PXC和

跨机房网段

QMHA的特点，多机房部
署提高节点间的同步效率
提高机房容灾的安全性

快速切换

failover切换只需要8-16s且
没有误切换
switchover只需要2秒内

切换逻辑可控

切换逻辑可以由情况和参数而定，
大事务或者主从延迟时不发生
switchover和不提供线上服务等

QMHA后续改进的问题

自动补全binlog

MHA可以自动补全binlog，PXC可以IST

QMHA需要能在failover后自动补全binlog给原master节点

延迟处理

某个从库因为某种原因出现延迟时，需要特殊处理

所有从库都出现延迟又该如何?

权重控制

PXC和QMHA都需要做到：

只读数据源可以根据权重配比进行流控，有助于对特殊机器的特殊处理

MMM、PXC和QMHA的对比

各个架构对比	MMM/MHA	PXC	QMHA
一致性	一般	强一致	较好
可用性	一般，受网络影响	一般，受网络影响	很好，网络影响小，可
数据丢失	主从切换可能会数据丢	0数据丢失	semi-sync时0数据丢失
运维成本	至少2台，运维要求低	至少3台，PXC运维门	至少2台，运维要求低

两手抓，两手都要硬

我们的平台 - 补天

Qunar

DBA操作平台

我们的平台 - 补天

数据库基础平台

PXC管理 | QMHA管理 | MMM管理 | REDIS管理 | HBASE管理

机器资源 | 账号管理 | 备份归档 | DBA值班 | DBA工具

工单申请

- SQL审核
- 数据库申请
- 账号权限
- 数据库迁移

DBA运维

- 数据库巡检
- 慢查询分析
- 监控报警
- 自助信息查询

内部服务

- 内部服务监控
- DBA小工具
- 需求反馈吐槽
- 一键初始化

The Future


THANKS