

基于Zookeeper+MHA 的mysql高可用架构设计

DTCC

2016中国数据库技术大会

DATABASE TECHNOLOGY CONFERENCE CHINA 2016

数据定义未来

SequeMedia
盛拓传媒

IT168.com

ChinaUnix

ITPUB

Lianjia刘世勇
2016年05月14日

关于我

- 曾就职于华为、网易
 - ✓ Oracle运维
 - ✓ Mysql运维
- 2015年初加入链家网
 - ✓ 任职链家网DBA
 - ✓ 负责链家网oracle和mysql数据库的运维、数据库架构设计、DB性能调优和SQL优化、DB自动化运维平台的构建等工作



DTCC

2016年中国数据库技术大会
DATABASE TECHNOLOGY CONFERENCE CHINA 2015

SequeMedia
世纪传媒

IT168

ChinaUnix

ITPUB

分享什么

- 基于MHA的常用mysql HA架构
- 为什么要改造常用方案
- Lianjia当前的架构
- 核心组件实现
- 流程分析
- 优化



DTCC

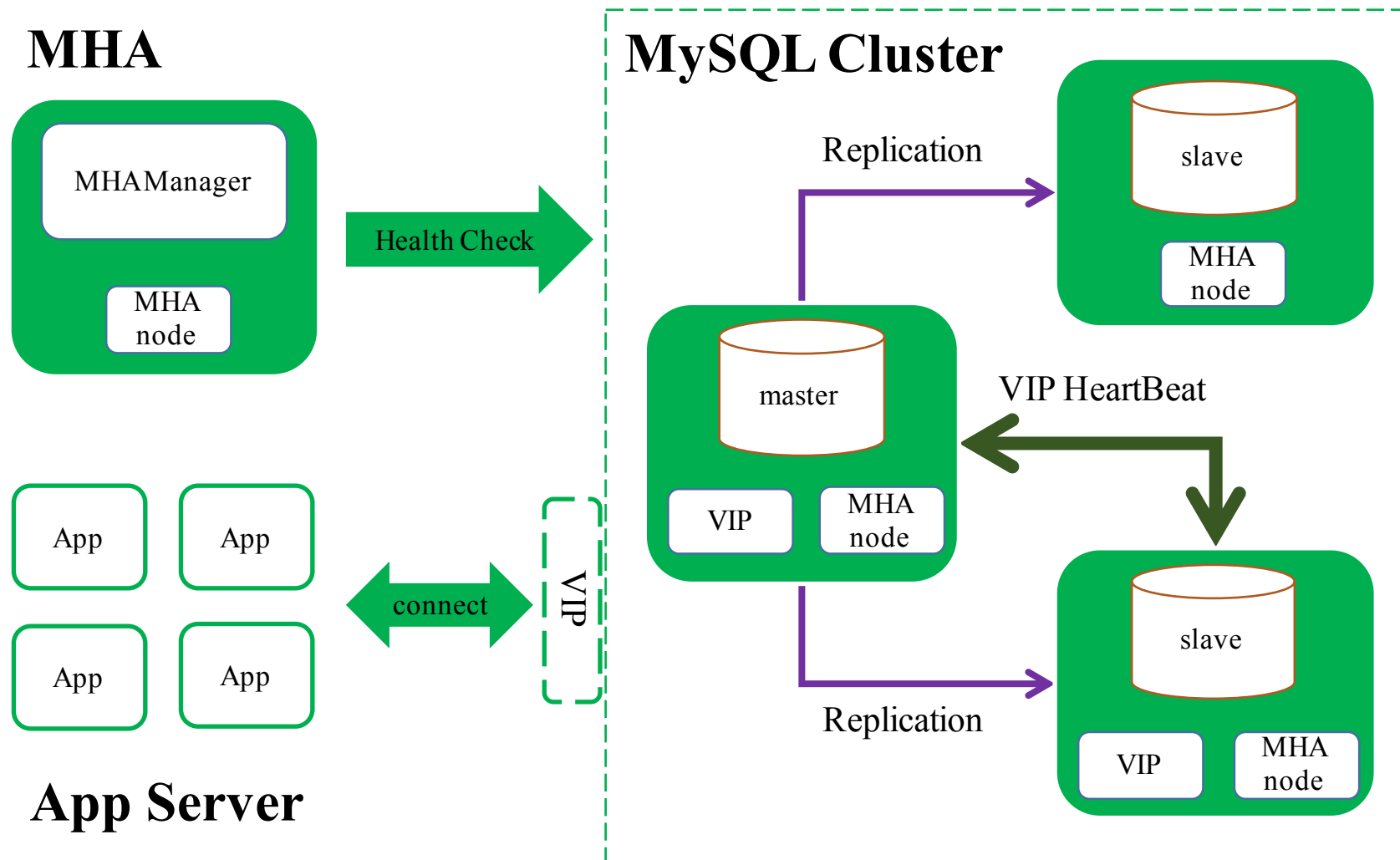
2016年中国数据库技术大会
DATABASE TECHNOLOGY CONFERENCE CHINA 2015

SequeMedia
世纪佳缘

IT168

ChinaUnix

ITPUB



基于MHA的经典mysql HA架构



DTCC

2016年中国数据库技术大会
DATABASE TECHNOLOGY CONFERENCE CHINA 2015

SequeMedia
世纪传媒

IT168

ChinaUnix

mpub

有哪些问题？

- VIP变成了单点
- keepalived本身的脑裂问题
- 单机多实例混部时，VIP如何应对



DTCC

2016年中国数据库技术大会
DATABASE TECHNOLOGY CONFERENCE CHINA 2015

SequeMedia
世纪传媒

IT168

ChinaUnix

ITPUB

改造目的&思路

- 解决VIP存在的问题
- 使用命名服务，对上层应用屏蔽mysql集群的拓扑信息，达到底层mysql集群的变更对上层透明的目的



DTCC

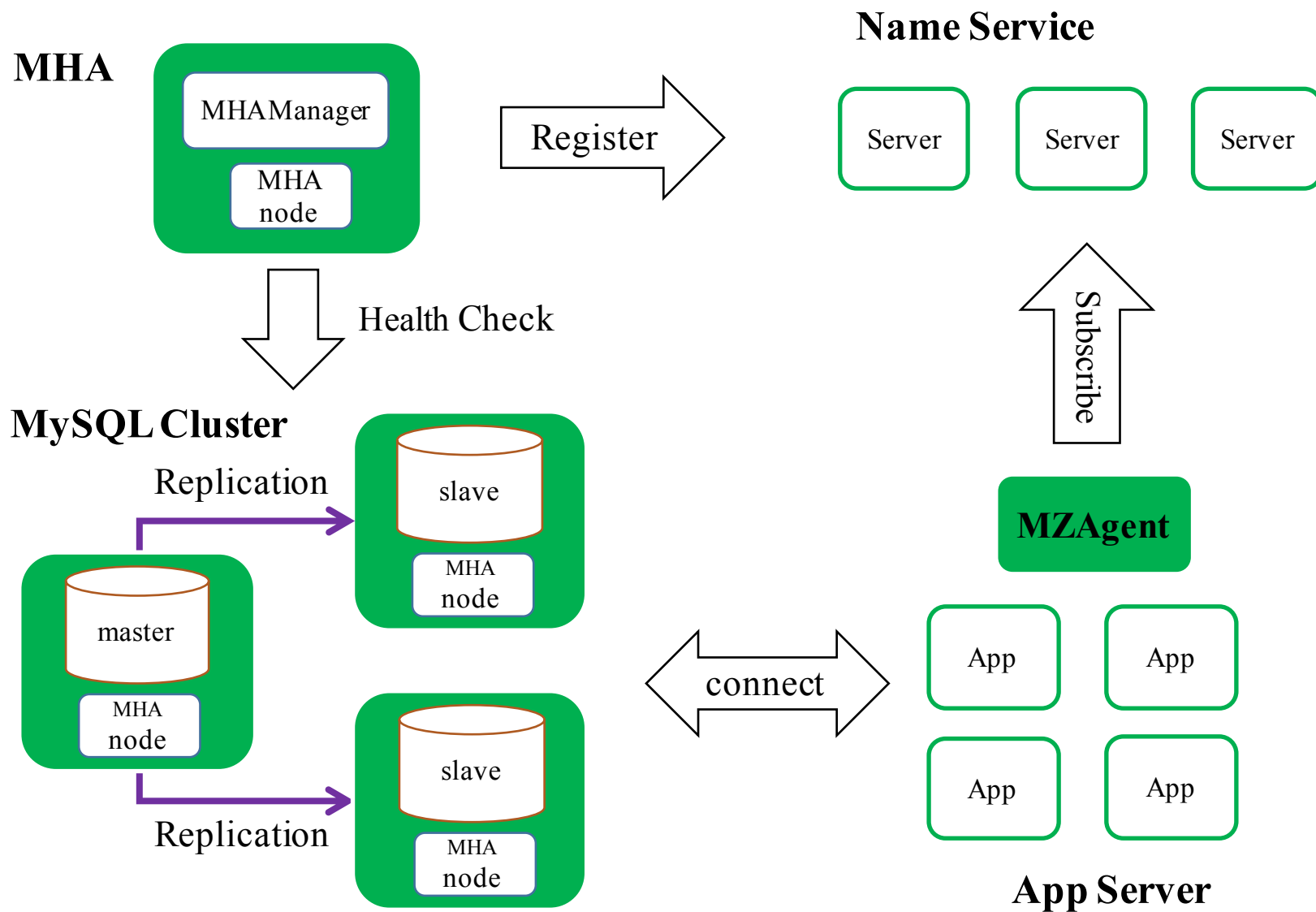
2016年中国数据库技术大会
DATABASE TECHNOLOGY CONFERENCE CHINA 2015

SequeMedia
世纪传媒

IT168

ChinaUnix

ITPUB



Lianjia 基于MHA的mysql HA架构



DTCC

2016年中国数据库技术大会
DATABASE TECHNOLOGY CONFERENCE CHINA 2015

SequeMedia

17173

ChinaUnix

mpub

MHA

- 集中管理mysql集群
- 负责mysql切换
- 向name service注册mysql服务信息
- 切换时发布mysql服务信息变更



DTCC

2016年中国数据库技术大会
DATABASE TECHNOLOGY CONFERENCE CHINA 2015

SequeMedia
世纪传媒

IT168

ChinaUnix

mpub

Name Service

- 提供命名服务
- 存储mysql服务信息，包括Port， IP， 主从拓扑
- 基于Zookeeper实现



DTCC

2016年中国数据库技术大会
DATABASE TECHNOLOGY CONFERENCE CHINA 2015

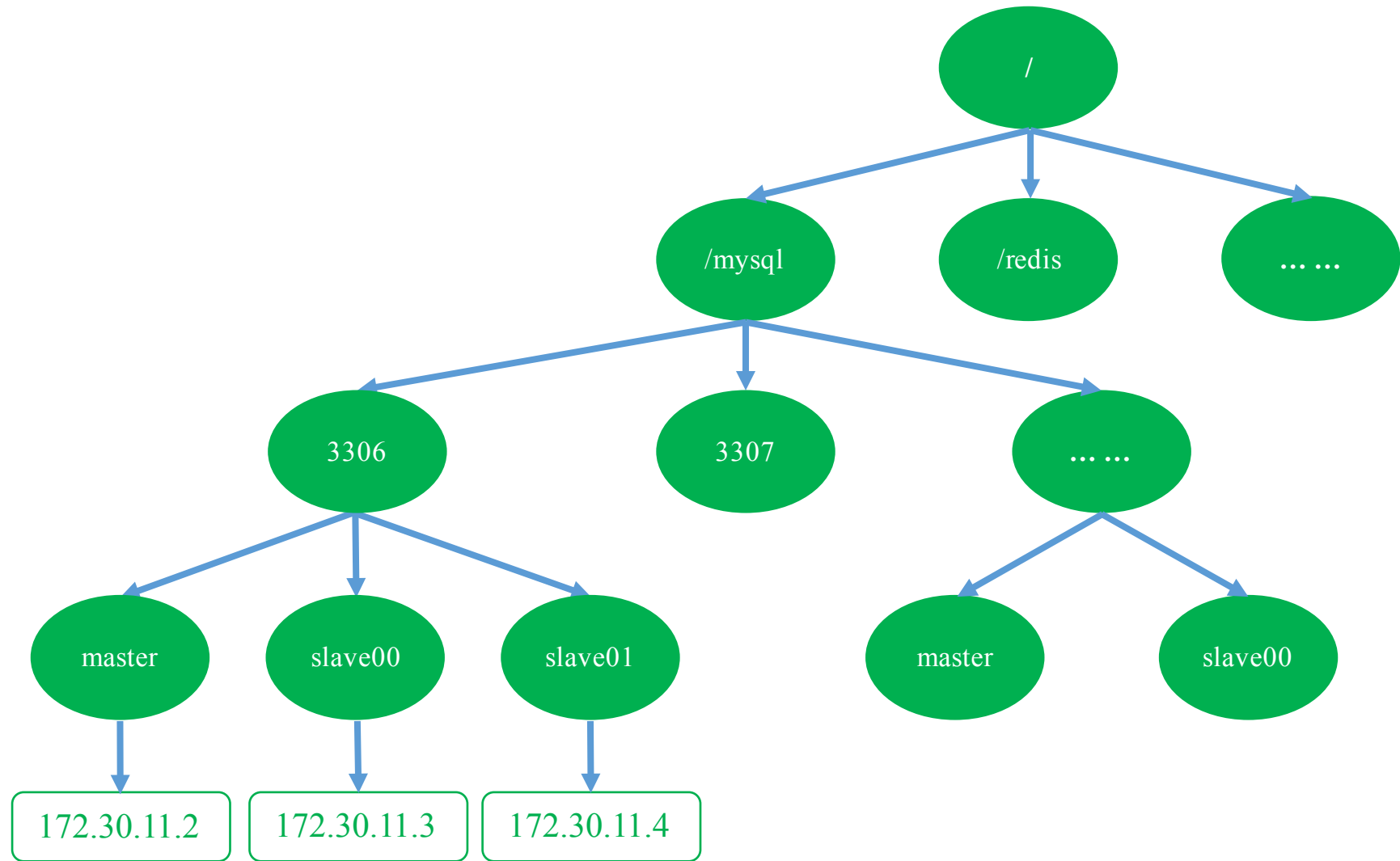
SequeMedia
世纪传媒

IT168

ChinaUnix

ITPUB

Mysql服务信息在zookeeper中存储结构



MZAgent

- 部署在app server
- 订阅在name service注册的mysql服务信息，并持久化到本地/etc/hosts
- 订阅变更，实时修改本地/etc/hosts
- 基于zkclient实现



DTCC

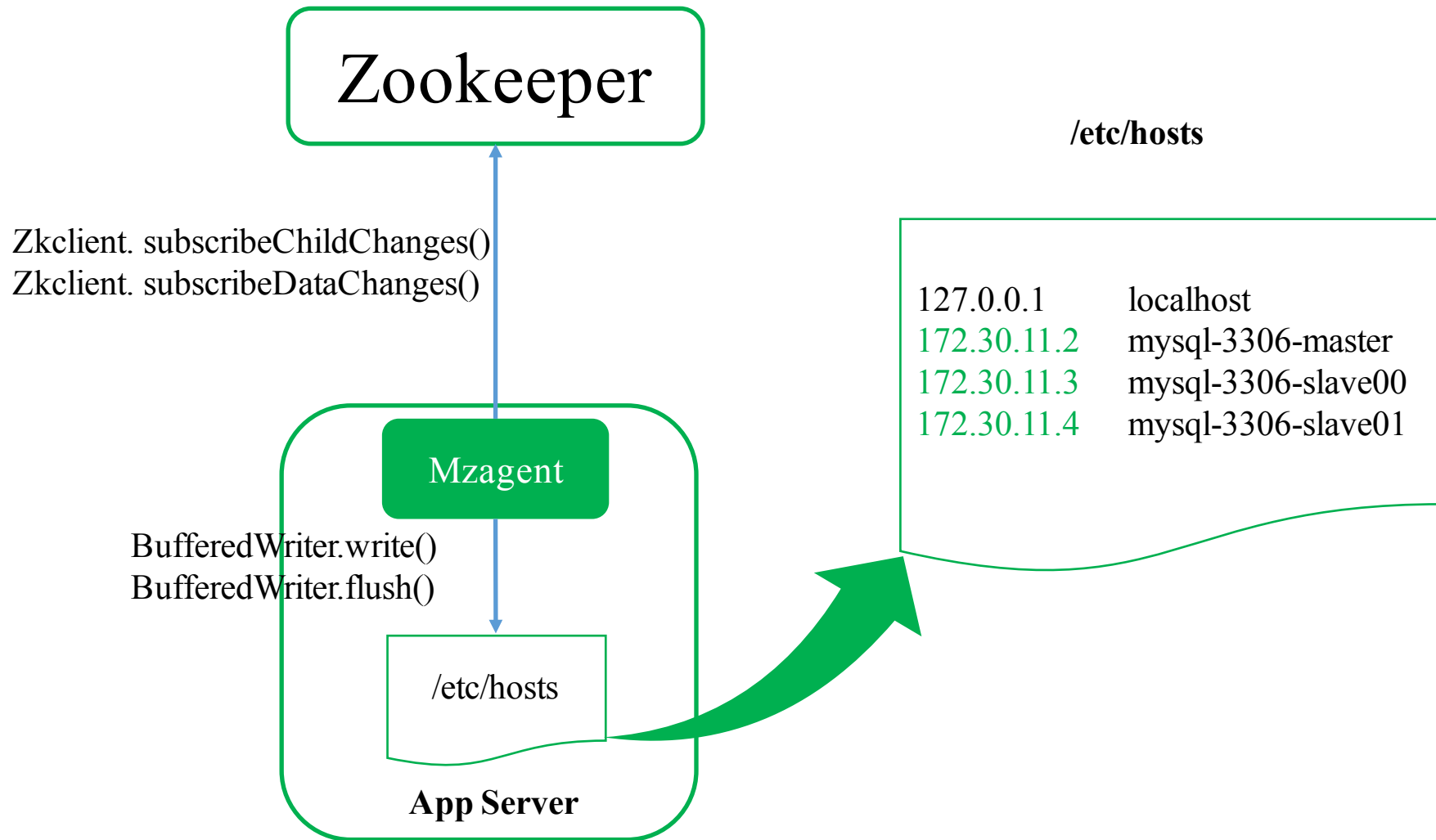
2016年中国数据库技术大会
DATABASE TECHNOLOGY CONFERENCE CHINA 2015

SequeMedia
世纪传媒

IT168

ChinaUnix

ITPUB



DTCC

2016年中国数据库技术大会
DATABASE TECHNOLOGY CONFERENCE CHINA 2015

SequeMedia
世纪传媒

IT168

ChinaUnix

ITPUB

Mysql服务注册流程

1. MHA监控进程启动
2. MHA向ZK注册mysql服务信息
3. MZAgent启动，订阅mysql服务信息
4. 持久化mysql服务信息到/etc/hosts
5. 应用使用hostname连接mysql



DTCC

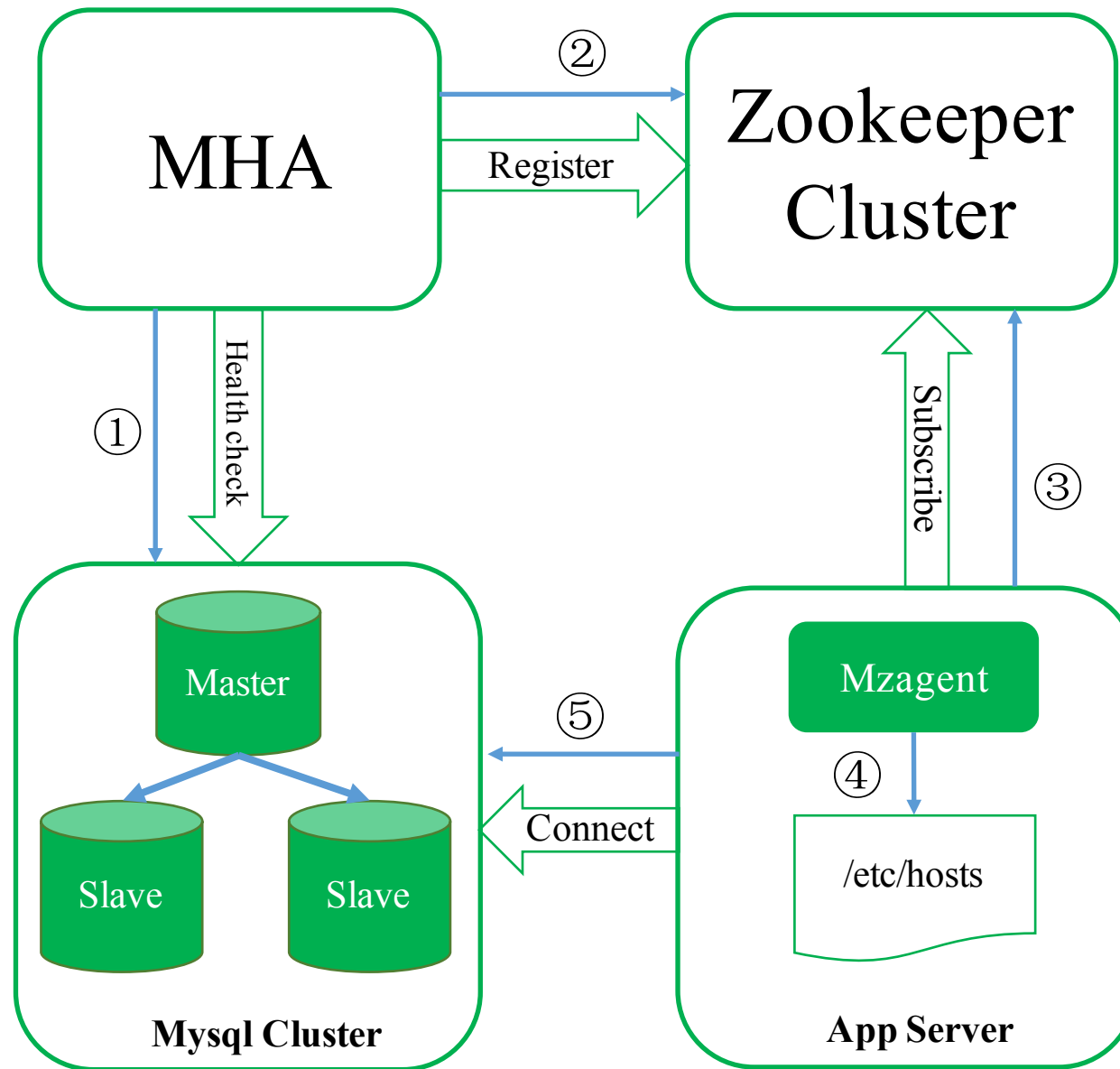
2016年中国数据库技术大会
DATABASE TECHNOLOGY CONFERENCE CHINA 2015

SequeMedia
世纪传媒

IT168

ChinaUnix

ITPUB



DTCC

2016年中国数据库技术大会
DATABASE TECHNOLOGY CONFERENCE CHINA 2015

SequeMedia
世纪传媒

IT168

ChinaUnix

mpub

Mysql切换流程

1. MHA做mysql切换
2. MHA向ZK发布mysql服务信息变更
3. MZAgent订阅到变更，并修改/etc/hosts中的hostname
4. 应用使用新的hostname连接mysql



DTCC

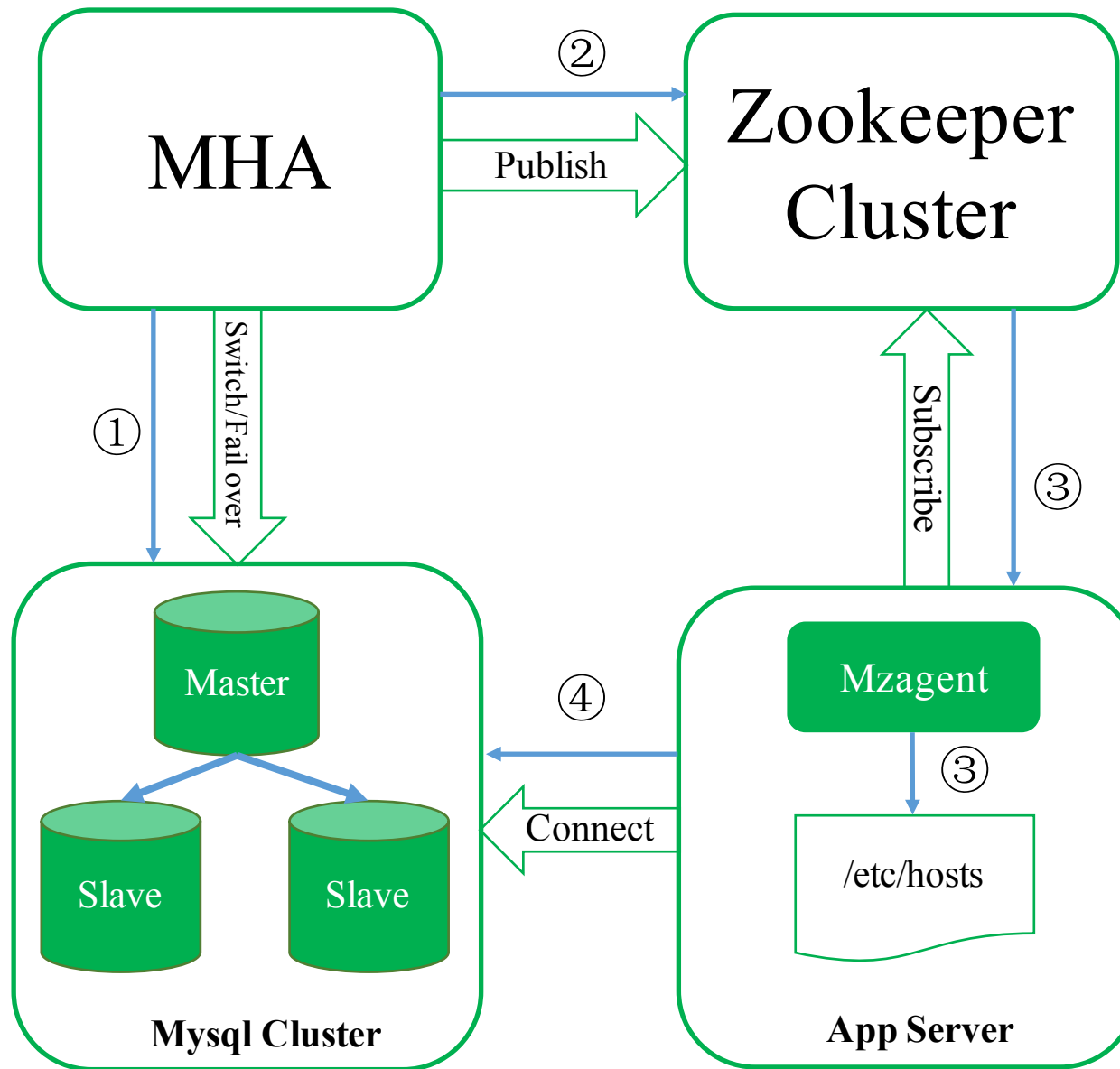
2016年中国数据库技术大会
DATABASE TECHNOLOGY CONFERENCE CHINA 2015

SequeMedia
媒体传媒

IT168

ChinaUnix

ITPUB



解决了哪些问题

- 命名服务提供者无单点问题
 - ✓ Mzagent单点，但是故障不影响访问数据库
- 规避VIP脑裂对上层应用的影响
- 单机多实例部署，管理方便，切换时集群间互不影响



DTCC

2016年中国数据库技术大会
DATABASE TECHNOLOGY CONFERENCE CHINA 2015

SequeMedia
世纪传媒

IT168

ChinaUnix

ITPUB

持续优化

- Agent的问题
 - ✓mysql集群扩/缩容时，应用需要做相应地配置更新
 - ✓/etc/hosts容易误操作，可能导致应用访问DB异常
 - ✓App server订阅mysql服务信息不同，带来额外的管理成本，不利于自动化
 - ✓额外的开发和维护成本



DTCC

2016年中国数据库技术大会
DATABASE TECHNOLOGY CONFERENCE CHINA 2015

SequeMedia
世纪传媒

IT168

ChinaUnix

ITPUB

持续优化

- 使用DNS接口
 - ✓ 使用DNS实现Name Service
 - ✓ 为mysql服务分配内部域名
 - ✓ 注册mysql服务到内网DNS server
 - ✓ App server使用dnsmasq，做DNS请求路由



DTCC

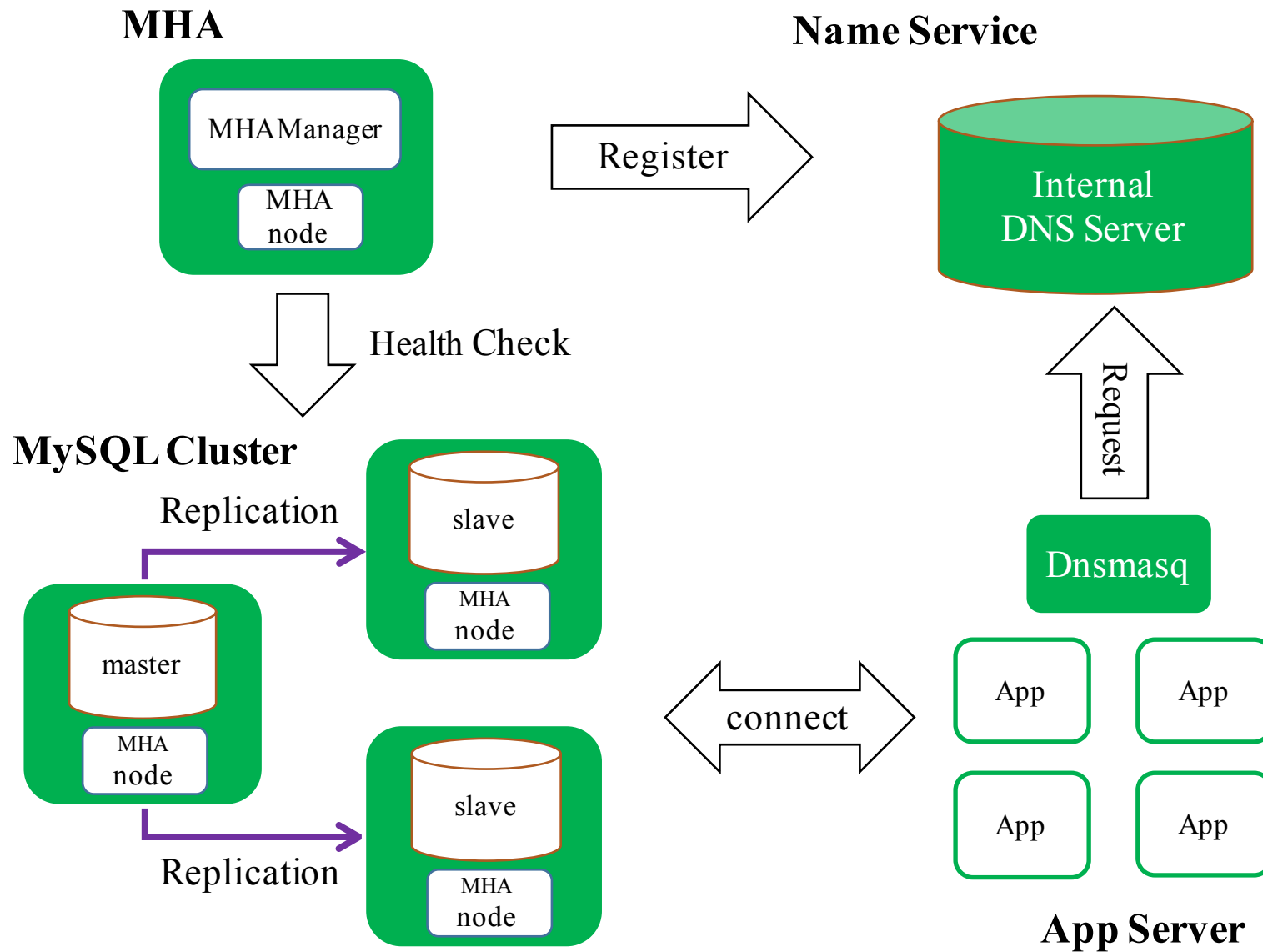
2016年中国数据库技术大会
DATABASE TECHNOLOGY CONFERENCE CHINA 2015

SequeMedia
世纪传媒

IT168

ChinaUnix

ITPUB



DTCC

2016年中国数据库技术大会
DATABASE TECHNOLOGY CONFERENCE CHINA 2015

SequeMedia
世纪传媒

IT168

ChinaUnix

mpub

持续优化

- DNS Cache带来的问题
 - ✓切换时mysql变更对上层不能及时生效
- 如何解决？
 - ✓设置合理的TTL
 - ✓切换时，主动purge cache记录



DTCC

2016年中国数据库技术大会
DATABASE TECHNOLOGY CONFERENCE CHINA 2015

SequeMedia
世纪传媒

IT168

ChinaUnix

ITPUB

持续优化

- 使用DNS接口解决了哪些问题
 - ✓ 多个slave共用同一个域名，读请求负载均衡，mysql集群扩/缩容对应用透明
 - ✓ 规避了人为误操作影响上层业务的风险
 - ✓ 消除app server和mysql对应关系的管理成本
 - ✓ 更好地支持自动化
 - ✓ 无需再维护额外的agent



DTCC

2016年中国数据库技术大会
DATABASE TECHNOLOGY CONFERENCE CHINA 2015

SequeMedia
世纪传媒

IT168

ChinaUnix

ITPUB

广告时间

Lianjia诚聘资深DBA，欢迎推荐和自荐
请赐简历到liushiyong@lianjia.com

万亿级房产O2O平台，等你一起来打造！



Q&A

THANKS

SequeMedia
盛拓传媒

IT168.com

ChinaUnix

ITPUB