

数据之大，云动未来

——弹性敏捷的RDS架构
与互联网最佳实践

爱可生公司



——中国领先的企业数据处理技术整体方案提供商
开源数据库商用一体化产品领军企业

SACC 2016 第八届中国系统架构师大会

SequeMedia
盛拓传媒

IT168.com
中国IT网 16 年

ChinaUnix

ITPUB
www.itpub.net

爱可生用户情况

十年磨砺，薄积厚发

1000+ 需求分析；300+ 项目合作；
100+ 大型用户；30+ 行业标杆

金融行业



通信行业



广电行业



互联网金融行业



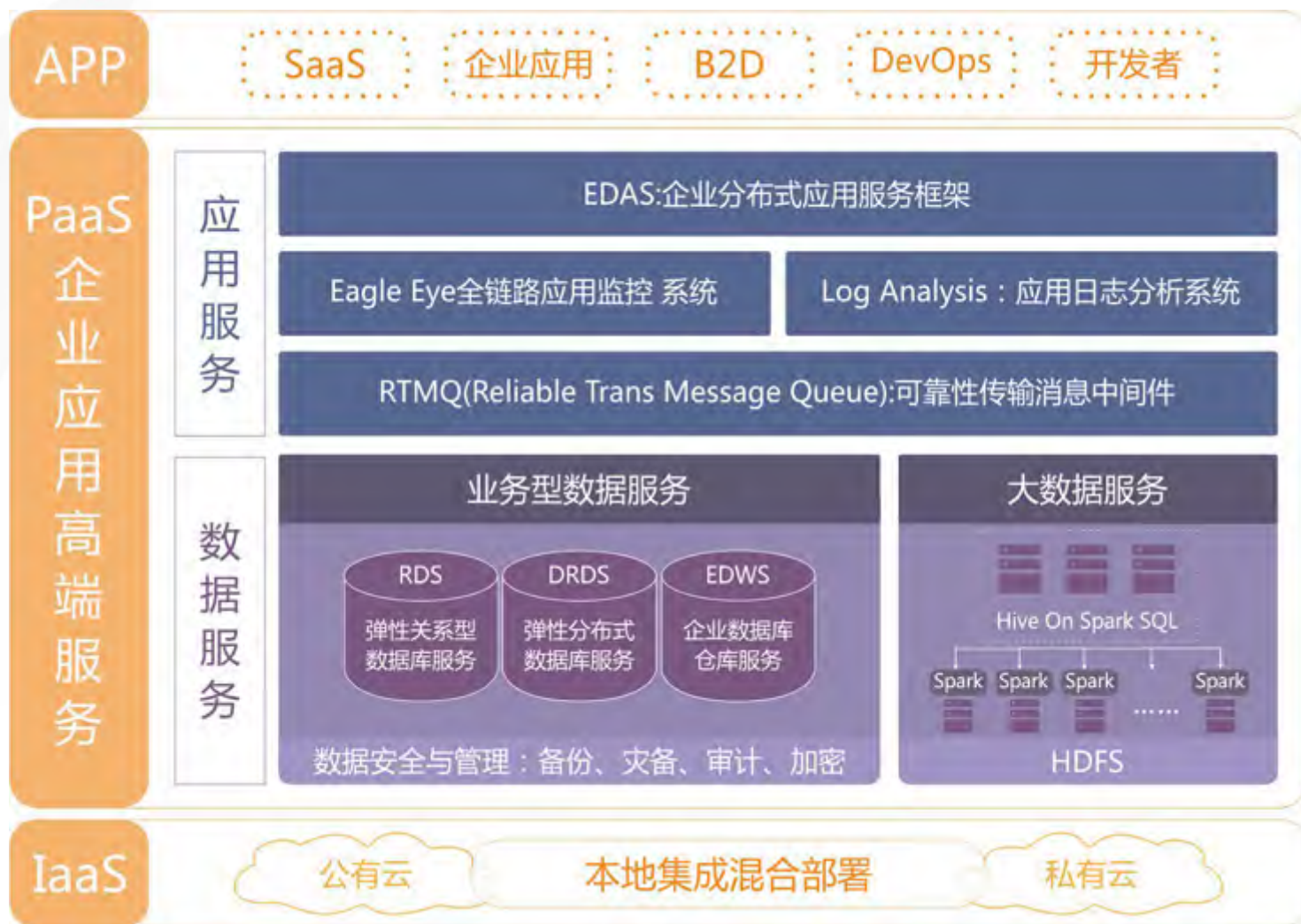
其他行业



传统企业互联网转型的应用架构

——以数据处理技术为核心，持续为互联网转型企业打造高端服务PaaS平台

爱可生服务范围



Enterprise Ready Open Source

开源软件的特点

和在企业应用中的问题
——以MySQL为例



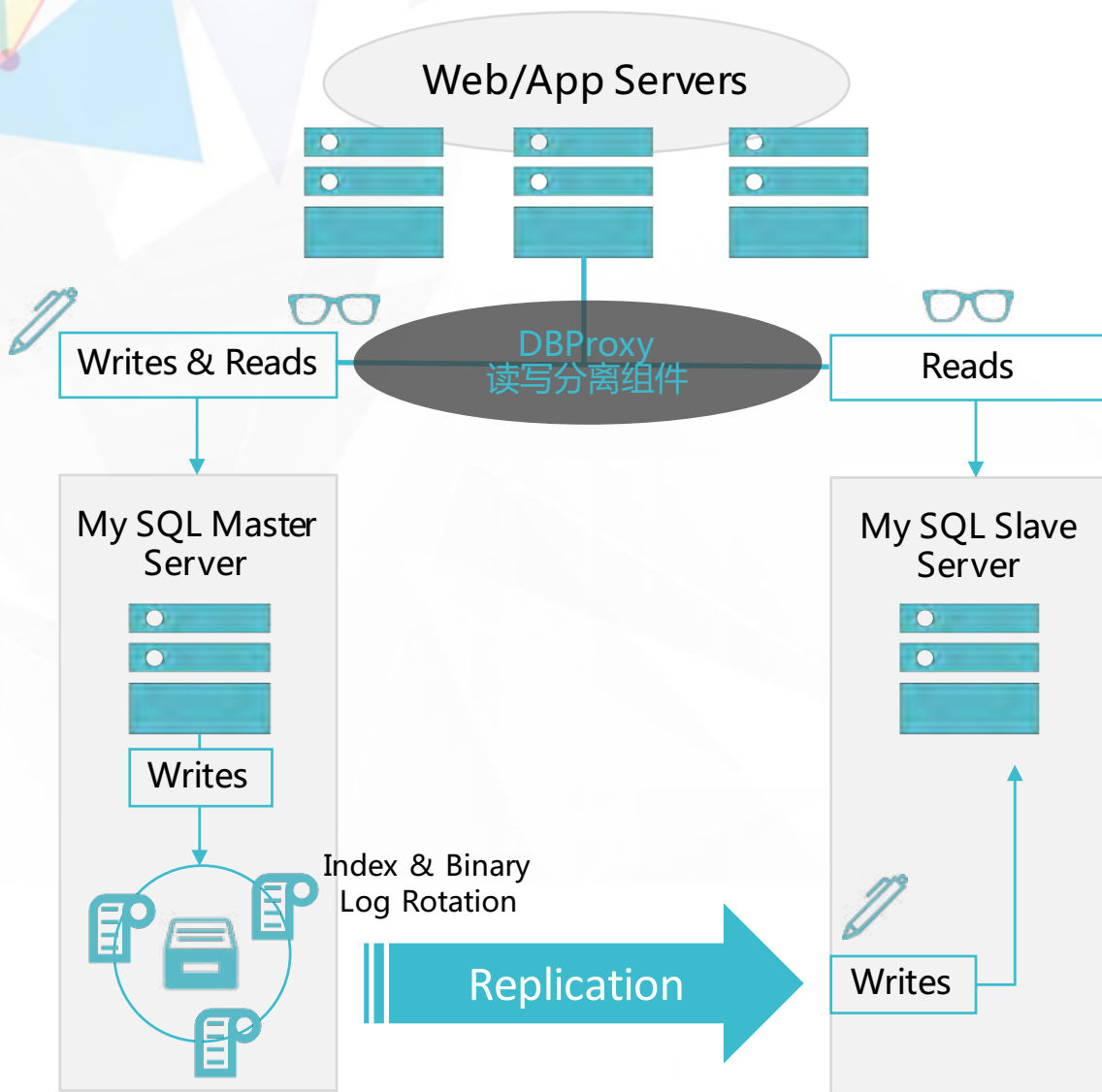
简单，易用，十分钟数据库



生产应用中

普遍面临**扩展性**
和**高可用**的难题

扩展性：读写分离是互联网行业提升数据库吞吐量的常规方案



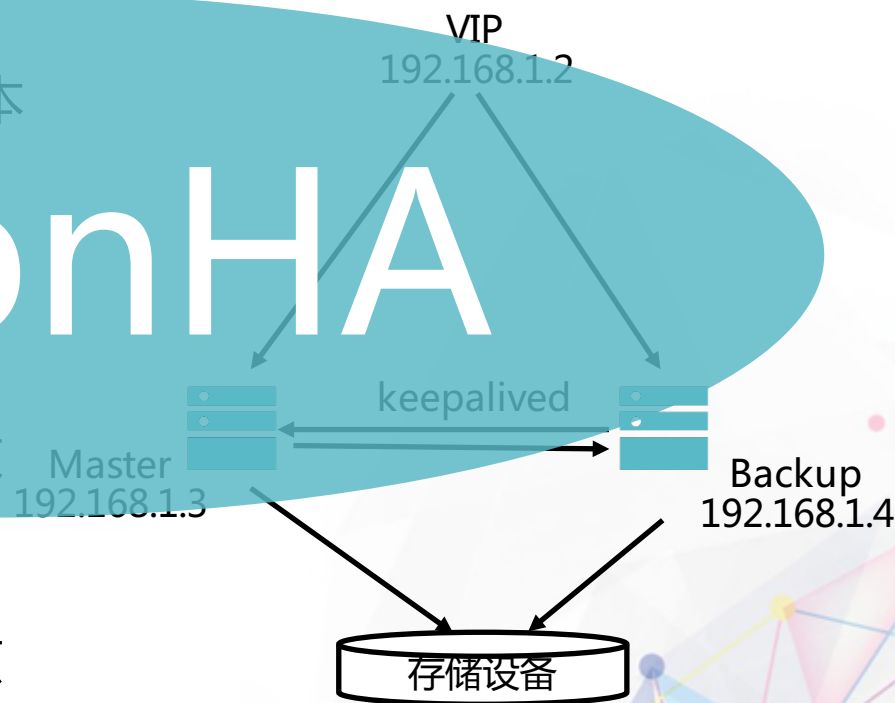
存在的问题：

- 在应用程序中包含
 - 读JDBC/ODBC
 - 写JDBC/ODBC
- 应用程序逻辑决定哪些SQL调用读数据源，哪些调用写数据源
- 已有应用需做大量修改才可使用
- 数据复制的效率得不到有效保障，复制延迟影响业务的结果

各类高可用方案不能保证数据完整性

- Pacemaker/Keepalived/MHA/MySQL Cluster.....
- 配置和运维复杂；
- 针对MySQL特点需要人工开发额外脚本
- 不可避免脑裂的问题
- 在业务高峰期出现问题不敢切换；
- 数据文件存在于存储设备，是个单独故障点。
- 冷备，备机不提供服务，浪费一半资源
- 不支持故障节点自动修复

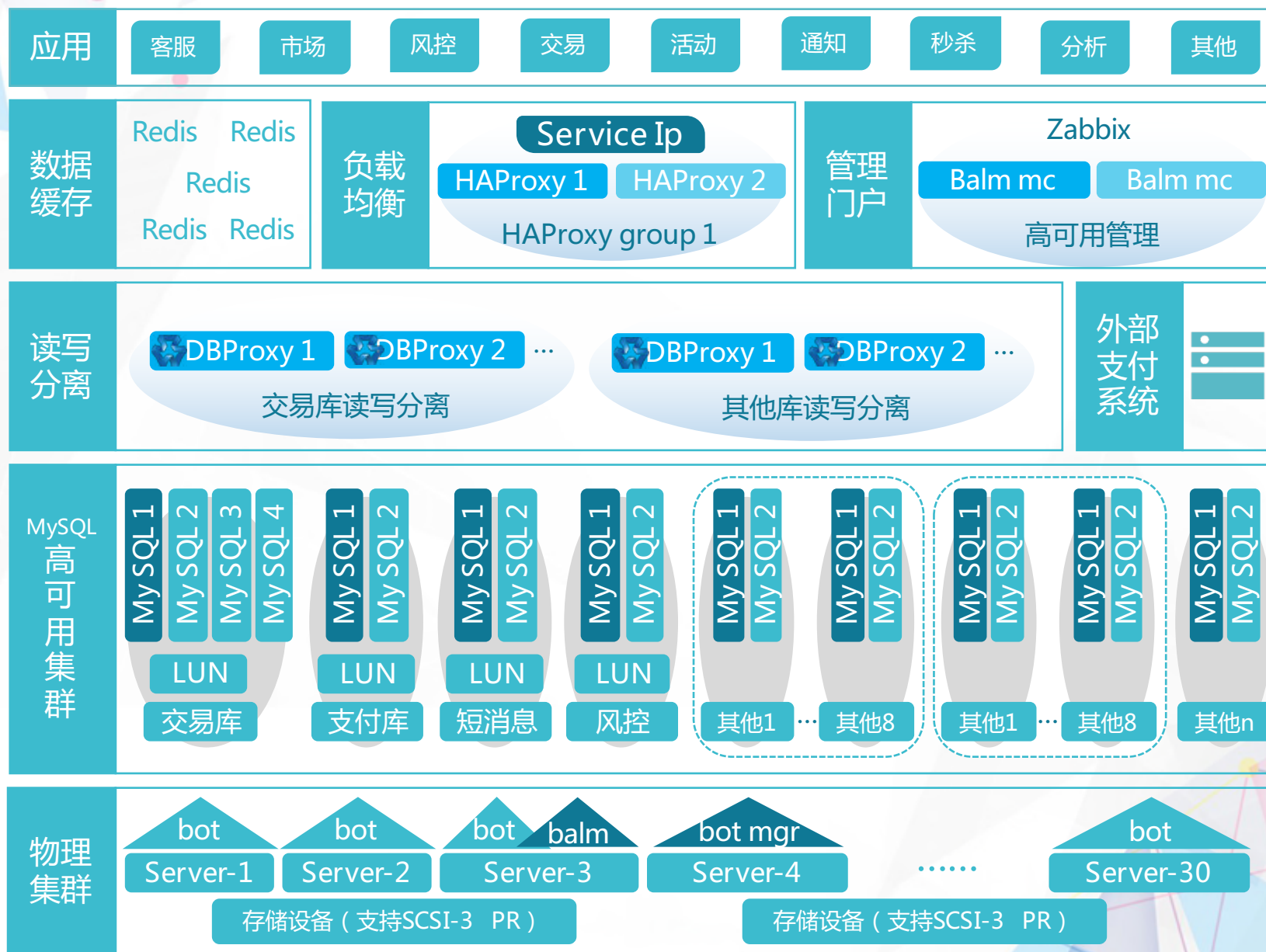
ActionHA



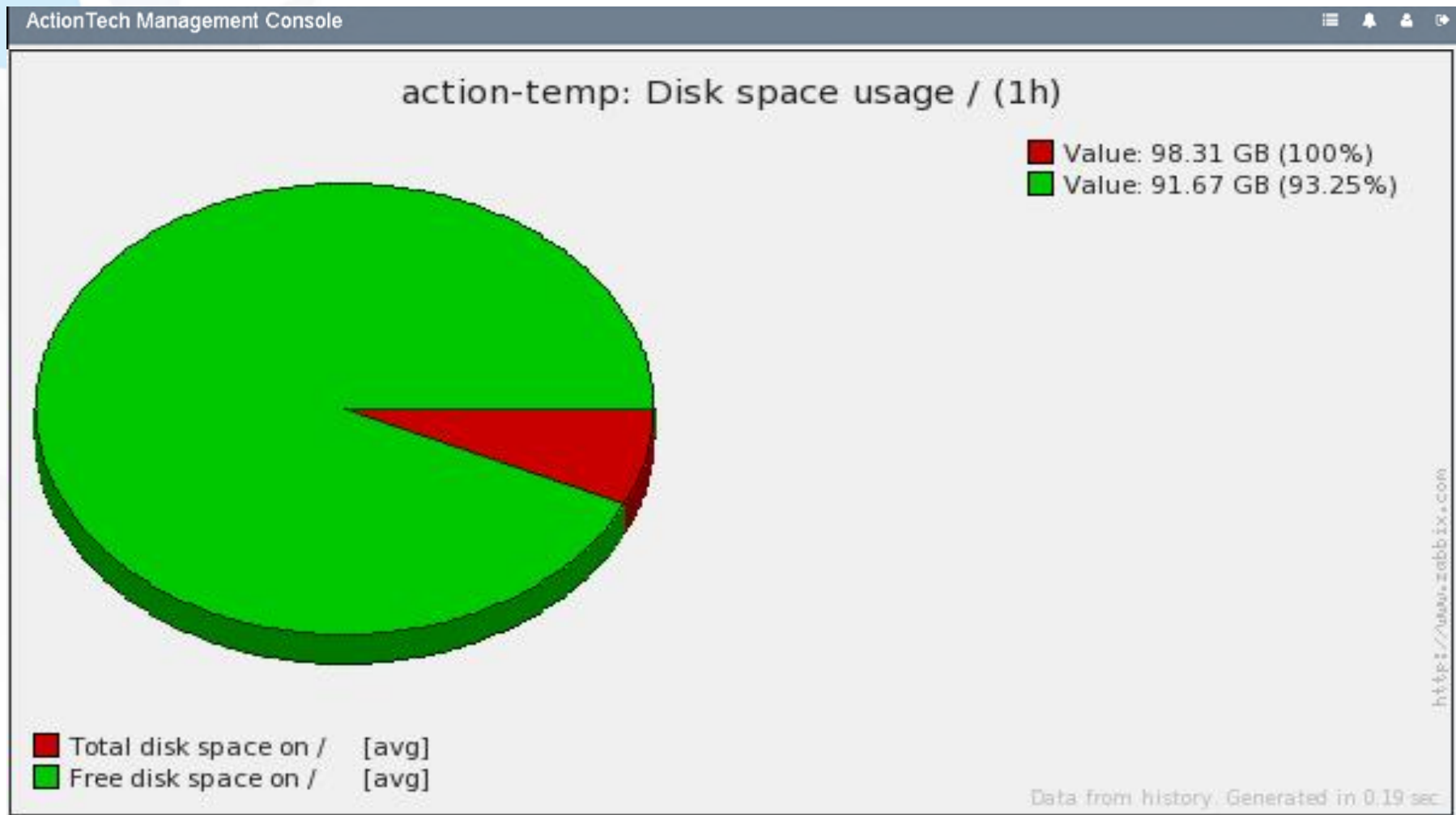
“互联网+” 需要这样的数据库云

**集高可用技术之大成
合横向扩展能力于一身**

互联网金融高可用数据云平台架构

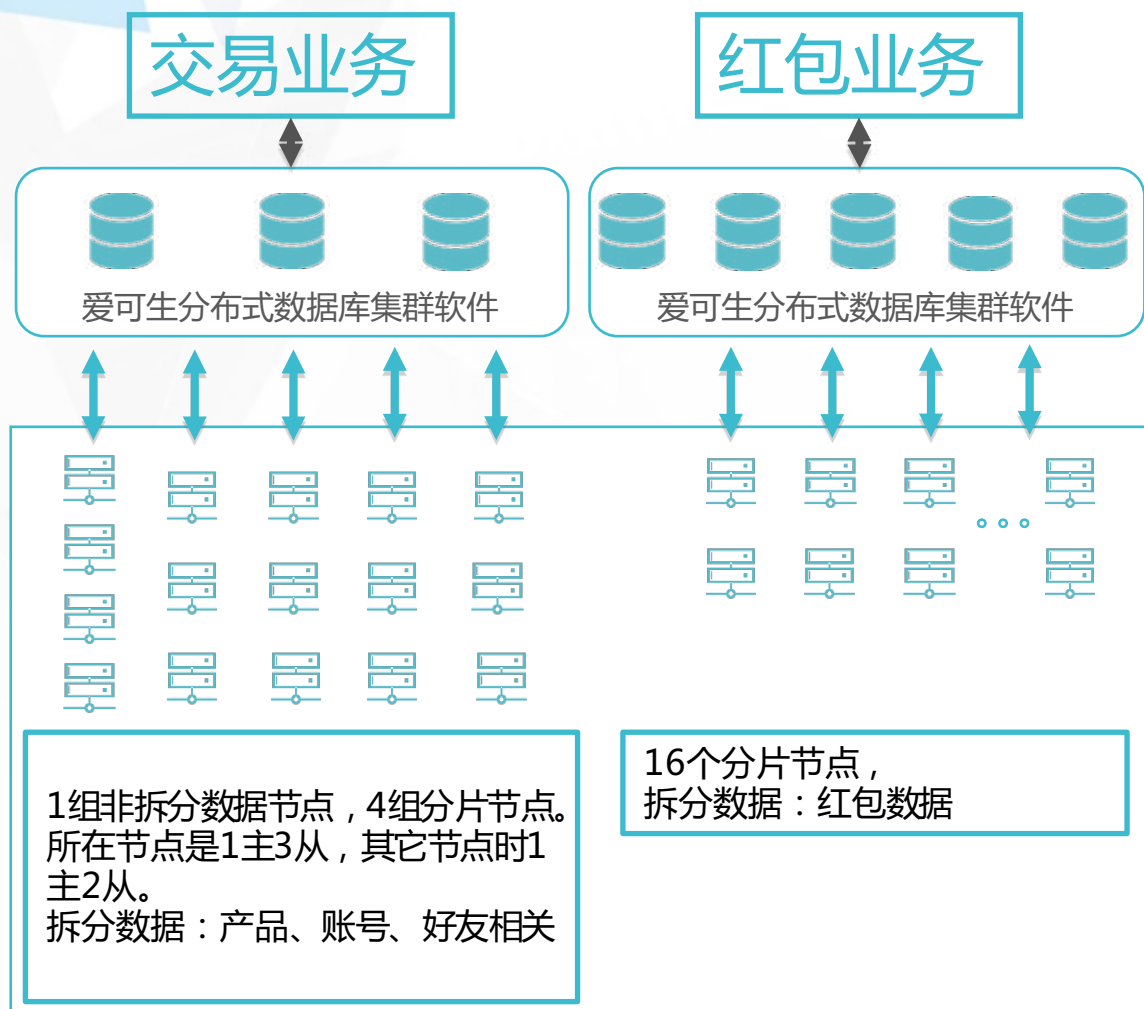


管理平台展示



技术演进：分布式数据库集群的无缝迁移

——支撑100万/秒红包设计容量



业务系统

- 线上借贷产品交易，红包库

数据规模

- 总量1TB
- 注册用户6千万以上

性能指标

- 红包数量按照**100万/秒**容量设计

系统特性

- 全局自增ID
- 拆分表多维度复制
- 中间层1%影响以内的异步审计日志
- 在水平拆分基础上支持单节点自定义分表

运行级别

- 7x24x365

数据库云PaaS平台的优势

多年数据库服务积累

- 十年数据服务经验
- 完善的数据生态系统
- 自主知识产权产品

快速投产无需更改应用

- 中间层透明MySQL接口
- 中间层自动读写SQL路由
- 事务可强制读主库
- 业务维度的延迟检测和自动处理



集高可用之大成

- 部署简单，运维方便
- 极端情况不丢任何数据
- 切换时间短，5秒以内
- 多套灵活高可用分组
- 同城灾备，异地灾备一体化
- 贴近MySQL设计，自动处理上百种MySQL故障

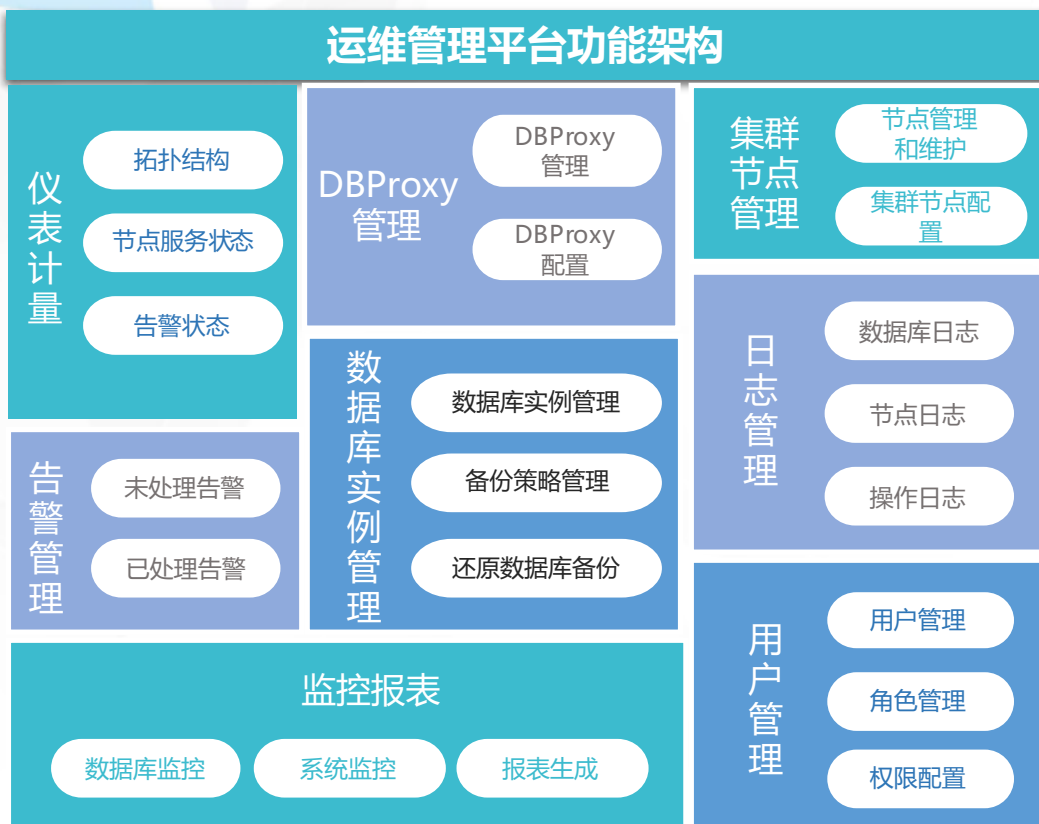
主从读写分离水平扩展

- 降低主节点的负载压力
- 提高资源利用效率
- 在线添加MySQL从机，不影响业务
- 根据业务SQL统计调整主从比例
- 高性能，中间层的性能损耗低于10%

MySQL数据库自动化运维管理

自动化运维管理平台方案

运维管理平台功能架构



数据库智能管理平台软件功能：

- **Dashboard**：显示平台中各类节点的逻辑拓扑图和健康状态，并提供标准接口，供第三方软件对接
- **集群节点管理**：实现集群节点主机的配置、管理及维护，如添加/删除主机、停止/启动HA等
- **数据库管理**：MySQL数据库的日常管理和维护，如高可用组的动态创建、扩展等，数据库实例的创建、激活、移除/加入高可用组、移出/加入路由、停止/启动实例等，备份策略管理，还原数据库备份等
- **监控报表**：实现对集群内的节点主机、MySQL数据库的运行状态进行监控，并生成性能诊断报表。
- **告警管理**：实现预设报警规则的配置，提供邮件、短信的异常告警提醒
- **DBProxy管理**：实现DBProxy（读写分离中间件）的全生命周期的管理，如创建、配置、删除、添加/删除路由等
- **用户及日志管理**：实现用户及权限、各节点及组件日志的管理

物理主机管理

← → ↻ 10.1.5.11:5710/server ☆

ActionTech Management Console 172

Dashboard

主机

MC组

数据库

Proxy

安装部署

路由设置

负载均衡

安装部署

路由设置

告警

告警配置

告警日志

日志

操作日志

组件日志

主机

刷新 添加主机 删除主机 启动Agent 启动BA 筛选

	主机名	IP地址	Agent状态	Ha状态	实例数
>	caict1	10.1.5.11	Online	Manager	1 x 1 x
>	caict2	10.1.5.12	Online	Online	1 x
>	caict3	10.1.5.13	Online	Online	1 x

管理控制台组件实例管理

← → ↻ 10.1.5.11:5710/mc_group

ActionTech Management Console

Dashboard

主机

MC组

数据库

Proxy

- 安装部署
- 路由设置

负载均衡

- 安装部署

MC组

刷新 配置SIP 扩展MC实例 删除MC实例 踢出HA 触发全备

筛选

组名	实例名	主机名	HA 管理状态	My SQL 角色	My SQL 端口	My SQL 状态	My SQL 复制状态	MC 状态	My SQL 可用备份
> mc	caict1	caict1	接受管理(SIP)	Master	5711	Running	Running	Running	有

数据库实例管理

← → ↻ 10.1.5.11:5710/database ☆

ActionTech Management Console 172

Dashboard

主机

MC组

数据库

Proxy

- » 安装部署
- » 路由设置

负载均衡

- » 安装部署
- » 路由设置

告警

数据库

刷新 创建实例 激活实例 删除实例 配置SIP 踢出HA 踢出路由 停止实例 触发全备 筛选

	组名	实例名	主机名	HA 管理状态	路由 管理状态	My SQL 角色	My SQL 端口	My SQL 状态	My SQL 复制状态	My SQL 可用备份
>	caicldb	caicldb1	caict2	接受管理	已加入	Slave	3306	Running	Running	无
>	caicldb	caicldb2	caict3	接受管理(SIP)	已加入	Master	3306	Running	Running	有

读写分离组件管理

← → ↻ 10.1.5.11:5710/proxy ☆

ActionTech Management Console 172

Dashboard 主机 MC组 数据库 Proxy

Proxy 安装部署

刷新 创建实例 删除实例 配置SIP 踢出HA 停止实例 筛选

	组名	实例名	主机名	HA 管理状态	Proxy 服务端口	Proxy 角色	当前动作	Proxy 状态
>	proxy	proxy1	caict1	接管管理(SIP)	13306	Master		Running

» 安装部署

» 路由设置

负载均衡组件管理

← → ↻ 10.1.5.11:5710/lb ☆

ActionTech Management Console 172

Dashboard 主机 MC组 数据库 Proxy 安装部署

负载均衡 安装部署

刷新 配置SIP 创建实例 删除实例 踢出HA 筛选

主机名	HA 管理状态	负载均衡 角色	当前动作	负载均衡 状态
No data available in table				

监控告警管理

10.1.5.11:5710/alert_log

ActionTech Management Console

172

Dashboard

主机

MC组

数据库

Proxy

- 安装部署
- 路由设置

负载均衡

- 安装部署
- 路由设置

告警

- 告警配置
- 告警日志

日志

- 操作日志

告警日志

刷新 全部标记为已读 筛选

时间戳	告警内容	状态
2016-08-11 19:58:31.788	INFO instance(instance-proxy-proxy1) service restart succeed	未读
2016-08-11 19:58:31.739	WARN instance(instance-proxy-proxy1) service is restarting because extra service is unavailable	未读
2016-08-11 19:46:04.102	ERROR server-caict2 is offline.	未读
2016-08-11 17:48:41.743	WARN restart instance-mysql-caictdb1 mysql replication, reason is replication fail(Slave_SQL_Running=No,Last_SQL_Ermo=0)	未读
2016-08-11 17:48:12.287	WARN restart instance-mysql-caictdb1 mysql replication, reason is replication fail(Slave_SQL_Running=No,Last_SQL_Ermo=0)	未读
2016-08-11 17:37:59.070	WARN restart instance-mysql-caictdb1 mysql replication, reason is replication fail(Slave_SQL_Running=No,Last_SQL_Ermo=0)	未读
2016-08-11 17:37:56.898	INFO instance(instance-mysql-caictdb1) service restart succeed	未读
2016-08-11 17:37:41.789	WARN instance(instance-mysql-caictdb1) service is restarting because mysql is not started	未读
2016-08-11 17:27:08.406	ERROR exclude instance-mysql-caictdb1 from ha and routing, reason is "mysql data is invalidated"	未读
2016-08-11 17:27:05.944	WARN instance-mysql-caictdb1 mysql is unavailable, reason is "mysql is not started and restart exceed limit"	未读
2016-08-11 17:26:56.297	WARN restart instance-mysql-caictdb1 mysql replication, reason is replication fail(Slave_SQL_Running=No,Last_SQL_Ermo=0)	未读
2016-08-11 17:26:54.135	INFO instance(instance-mysql-caictdb1) service restart succeed	未读
2016-08-11 17:26:39.021	WARN instance(instance-mysql-caictdb1) service is restarting because mysql is not started	未读

用户审计

10.1.5.11:5710/action_log

ActionTech Management Console

172

Dashboard

主机

MC组

数据库

Proxy

安装部署

路由设置

负载均衡

告警

告警配置

告警日志

日志

操作日志

组件日志

操作日志

刷新

筛选

	用户名	时间戳	日志
>	admin	2016-08-11 17:39:58.240	加入路由实例 instance-mysql-caictdb1 完成
>	admin	2016-08-11 17:39:54.693	加入路由实例 instance-mysql-caictdb1...
>	admin	2016-08-11 17:37:39.340	加入HA实例 instance-mysql-caictdb1 完成
>	admin	2016-08-11 17:37:38.658	加入HA实例 instance-mysql-caictdb1...
>	admin	2016-08-11 16:46:27.620	触发数据库 caictdb2 的全量备份 完成
>	admin	2016-08-11 16:46:24.162	触发数据库 caictdb2 的全量备份...
>	admin	2016-08-11 16:15:27.347	触发数据库 caictdb2 的全量备份 完成
>	admin	2016-08-11 16:15:24.079	触发数据库 caictdb2 的全量备份...
>	admin	2016-08-11 14:32:46.188	加入路由实例 instance-mysql-caictdb1 完成
>	admin	2016-08-11 14:32:42.317	加入路由实例 instance-mysql-caictdb1...
>	admin	2016-08-11 14:31:51.030	加入HA实例 instance-mysql-caictdb1 完成
>	admin	2016-08-11 14:31:50.101	加入HA实例 instance-mysql-caictdb1...
>	admin	2016-08-11 12:26:33.381	加入路由实例 instance-mysql-caictdb2 完成
>	admin	2016-08-11 12:26:33.381	加入路由实例 instance-mysql-caictdb2...

用户管理

← → ↻ 10.1.5.11:5710/user ☆

ActionTech Management Console 172

Dashboard

主机

MC组

数据库

Proxy ▾

- » 安装部署

负载均衡 ▾

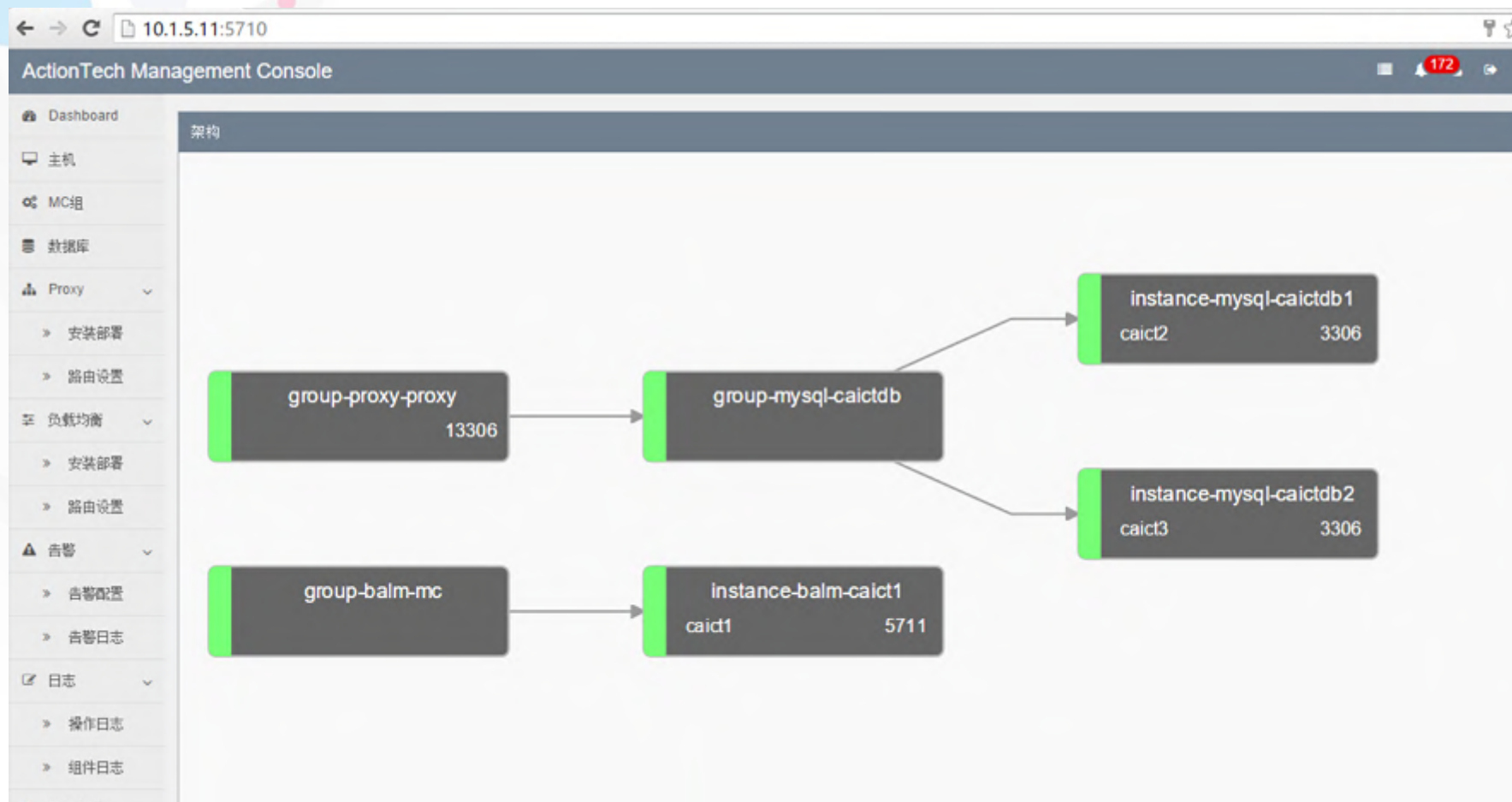
- » 安装部署

用户管理

刷新 创建用户 修改用户 删除用户 筛选

用户名	用户组	状态	创建时间
admin	administrator	启用	2016-08-10 11:14:17.752

平台组件拓扑结构图

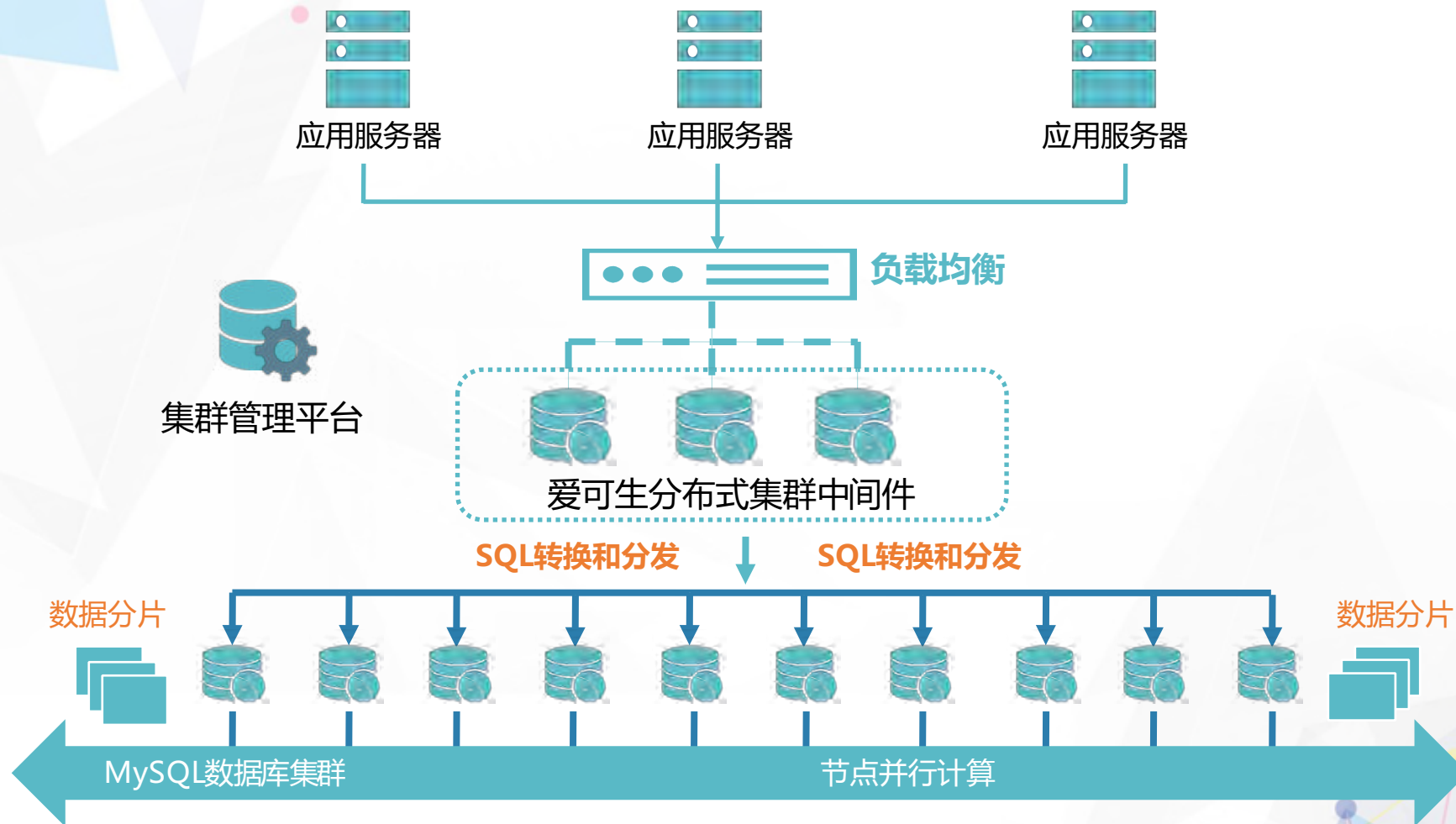


“互联网+” 对于大规模云数据集的处理技术需求

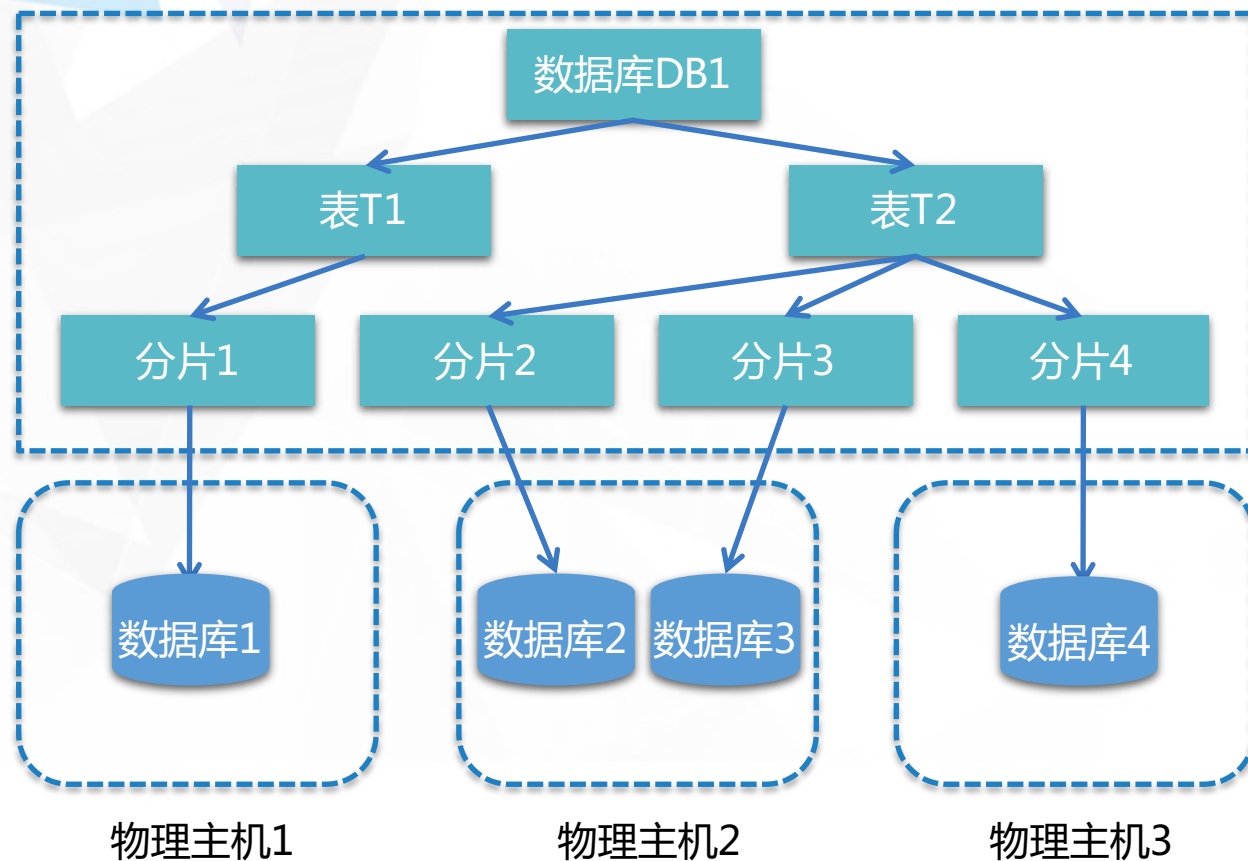
分布式集群

分库分表

分布式MySQL集群原理



分布式数据库分库分表模型



- 逻辑库(schema)- DB1
- 逻辑表(table)-T1
- 分片节点(datanode)-分片1
- 实例(database)-数据库1
- 节点主机(datahost)-物理主机1

“互联网+” 对开发测试云的要求

DevOps开发运维一体化

DB as a Service自服务

某银行面向私有云的云数据库平台，打造DevOps模式应用

● 项目背景：

该银行的数据中心拥有规模庞大的IT基础设施和种类繁多的线上和开发测试项目。需构建完整统一的私有云中的数据库即服务平台，并与IaaS整合。

● 解决方案：

- 我们提供了基于MySQL的云数据库即服务平台。
- 实现了面向多租户开发测试环境和生产环境的数据库自服务、按需创建扩缩、便捷自动管理和运维的统一的数据库云平台。

● 方案价值：

目前DBaaS平台已投产系统63套，开发环境200个数据库实例，测试环境100个实例。物理服务器共计30台左右，虚拟机200个左右。

- 提升了快速的MySQL服务交付能力
- 形成了标准的MySQL运维管理规范
- 实现了全面的MySQL运维自动化能力

已经支持了**手机银行、财富体验、个贷系统、互联网在线融资、信用卡企业服务**等**30余个项目**，加快了开发、单元测试、功能测试、集成测试和生产部署的处理流程，从基础设施层面满足业务敏捷性需求。

界面展示---实例管理

物理机资源管理 ▾ 服务配置管理 ▾ 数据库实例管理 ▾ 日志管理 ▾ 告警管理 ▾ 系统管理 ▾

Dashboard 数据库实例管理 ×

刷新 申请实例 性能扩容 存储扩容 HA扩容 备份 注销 删除 恢复 修改密码 修改proxy连接数

	实例名称	虚IP	端口	服务配置类型	角色	HA状态	实例状态	启停操作	详细信息
= 实例名称: test5									
1	☐	test5	10.186.24.36	5100	超级mini型	主	Running	运行中	停用 查看详情
2	☐	test5	10.186.24.37	5100	超级mini型	备	Running	运行中	停用 查看详情
= 实例名称: HA1									
3	☐	HA1	10.186.24.36	5101	超级mini型	主	Running	运行中	停用 查看详情
4	☐	HA1	10.186.24.37	5101	超级mini型	备	Recovering	运行中	停用 查看详情

界面展示---备份策略

Dashboard

数据库实例管理 ×

备份策略管理 ×

刷新

添加备份策略

修改备份策略

		数据库名称
1	+	test5
2	+	HA1

添加备份策略

全备自动备份配置

备份周期:

月备份

周备份

☒ 日备份

00:00

备份保存套数(一个全备带多个增长为一套):

10

套

增备自动备份配置

备份周期:

月备份

周备份

日备份

☒ 小时备份

分钟备份

每小时

添加

取消

20

第 1

界面展示---实例性能扩容

The screenshot displays a database management interface with a top navigation bar containing 'Dashboard', '数据库实例管理' (Database Instance Management), and '备份策略管理' (Backup Strategy Management). Below the navigation bar is a toolbar with various actions: '刷新' (Refresh), '申请实例' (Apply Instance), '性能扩容' (Performance Expansion), '存储扩容' (Storage Expansion), 'HA扩容' (HA Expansion), '备份' (Backup), '注销' (Deregister), '删除' (Delete), '恢复' (Restore), '修改密码' (Change Password), and '修改proxy连接数' (Change proxy connection count). The '性能扩容' button is highlighted with a red box and a red arrow pointing to a modal dialog box titled '性能扩容'.

The modal dialog box '性能扩容' contains a dropdown menu for '服务配置' (Service Configuration) set to '百分之50虚拟机 (50%-2G)'. At the bottom of the dialog are '提交' (Submit) and '取消' (Cancel) buttons.

The background interface shows a table of database instances. The table has columns for '实例名称' (Instance Name), '实例状态' (Instance Status), '启停操作' (Start/Stop Operation), and '详细信息' (Detailed Information). The instances are grouped by name: 'test5' and 'HA1'.

	实例名称	实例状态	启停操作	详细信息
实例名称: test5				
1	test5	运行中	停用	查看详情
2	test5	运行中	停用	查看详情
实例名称: HA1				
3	HA1	运行中	停用	查看详情
4	HA1	运行中	停用	查看详情

界面展示---实例存储扩容

The screenshot displays a database management console with a 'Database Instance Management' tab. A red arrow points to the 'Storage Expansion' button in the top toolbar. A modal dialog titled 'Storage Expansion' is open, showing the current data directory size as 10GB and the required expansion size as 20GB. The background table lists instances, with instance HA1 highlighted.

Dashboard | 数据库实例管理 | 备份策略管理

刷新 | 申请实例 | 性能扩容 | **存储扩容** | HA扩容 | 备份 | 注销 | 删除 | 恢复 | 修改密码 | 修改proxy连接数

	实例名称	端口	服务器类型	角色	HA状态	实例状态	启停操作	详细信息
实例名称: test5								
1	test5	10.1				运行中	停用	查看详情
2	test5	10.1				运行中	停用	查看详情
实例名称: HA1								
3	HA1	10.1				运行中	停用	查看详情
4	HA1	10.1				运行中	停用	查看详情

存储扩容

当前数据目录大小为10GB
需要扩容的大小: 20

提交 取消

界面展示---实例HA扩容

Dashboard 数据库实例管理 备份策略管理

刷新 申请实例 性能扩容 存储扩容 HA扩容 备份 注销 删除 恢复 修改密码 修改proxy连接数

	实例名称	虚IP	端口	服务配置类型	角色	HA状态	实例状态	启停操作	详细信息
= 实例名称: test5									
1	test5	10.186.24.36						停用 查看详情	
2	test5	10.186.24.37						停用 查看详情	
= 实例名称: HA1									
3	HA1	10.186.24.36						停用 查看详情	
4	HA1	10.186.24.37						停用 查看详情	
= 实例名称: HA2									
5	HA2	10.186.24.213(eth0)					启用	查看详情	

HA扩容

是否进行HA扩容? 点击确定继续

注: 扩容时需要停止数据库服务

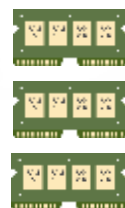
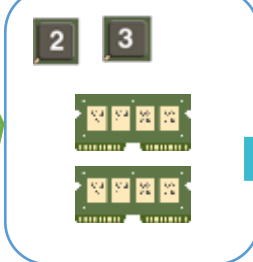
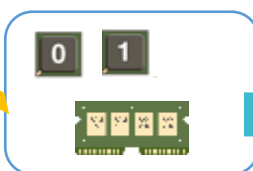
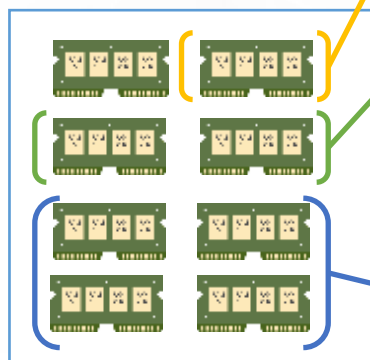
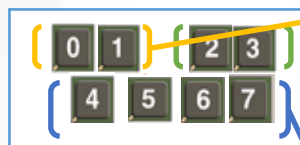
确定 取消

资源隔离：Container与Docker技术

物理机资源

Control Groups

资源用户



平台系统
资源调度

- 单机多实例，充分利用资源

- DBaaS提供了CPU计算资源、内存资源的隔离手段

MySQL A型实例

- DBaaS中的数据库服务实例运行时在各自的容器中运行，互相不可见，不影响

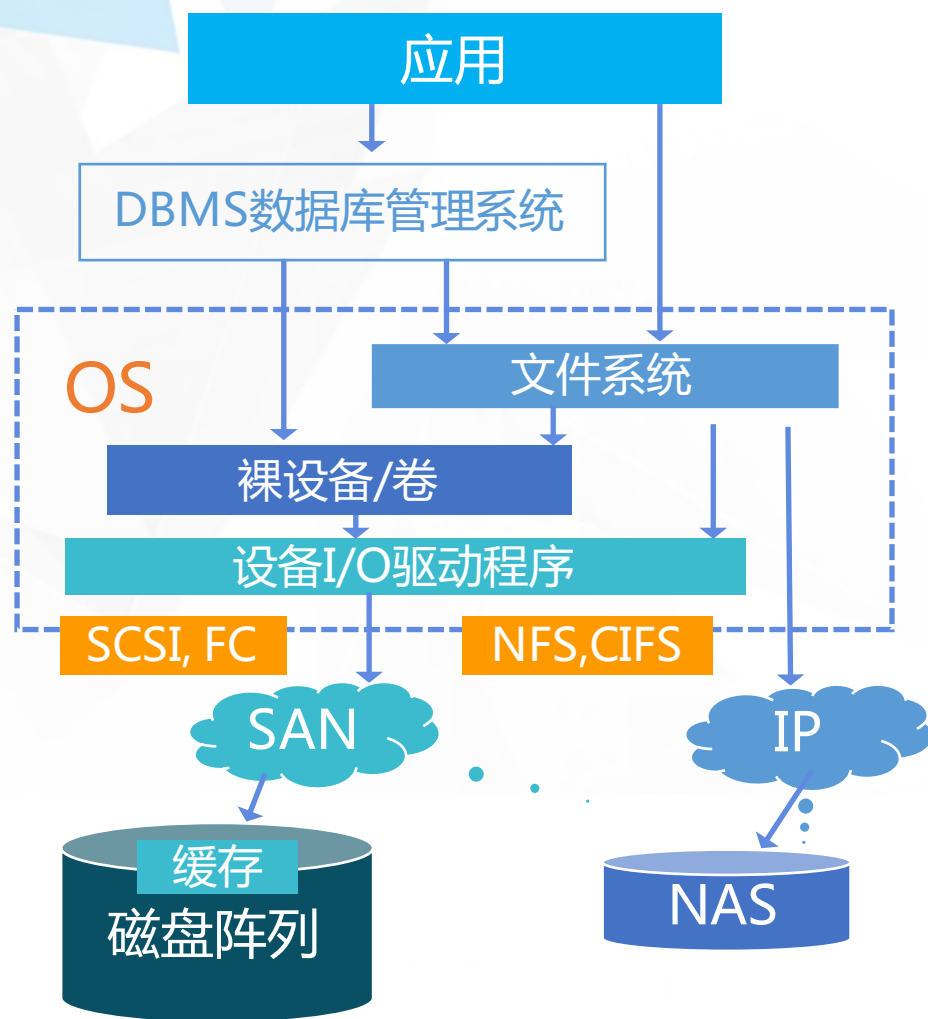
- CGroups是Linux内核提供的一种资源隔离手段，资源调度损耗小，性能高

MySQL B型实例

- Docker容器采用了linux内核中的cgroup技术来实现container的资源隔离和控制。

MySQL数据库灾备数据复制

容灾技术层次及分类



5. 基于应用复制

4. 基于数据库复制

3. 基于OS复制

2. 基于SAN复制

1. 基于阵列复制

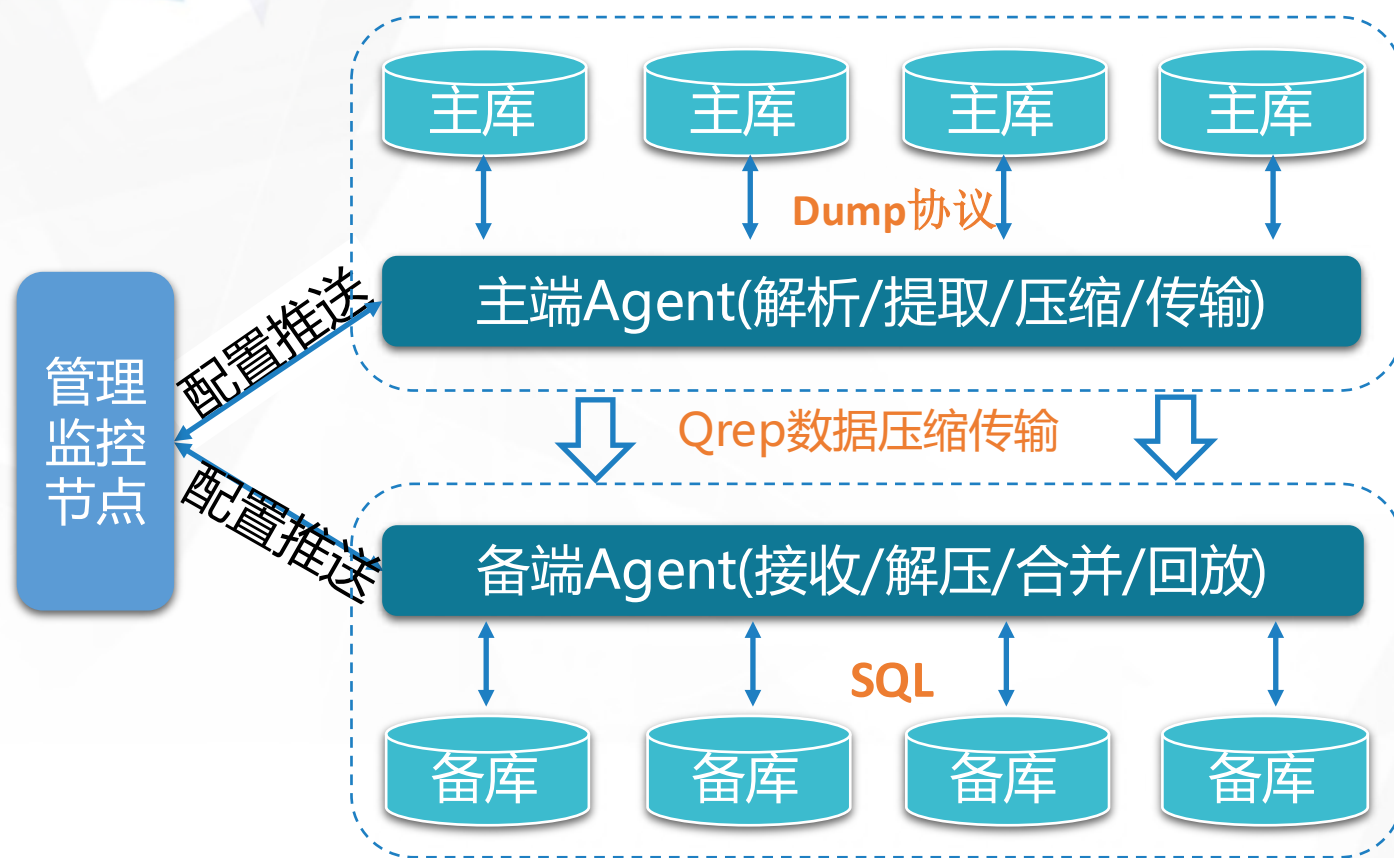
扩展性

复杂度

兼容性

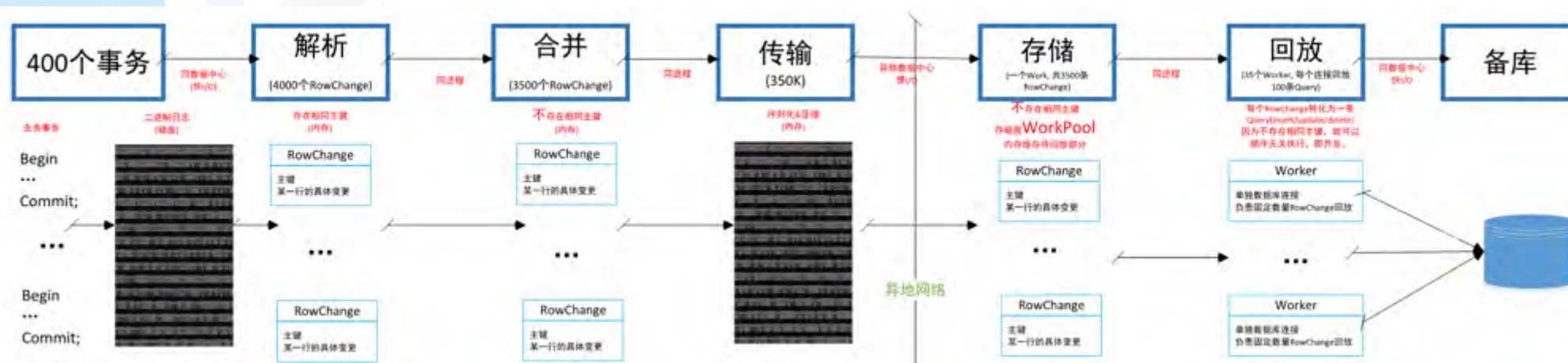
带宽

基于QRep的数据库容灾架构原理



- 数据复制基于binlog日志进行解析、过滤、传输、回放
- 优化了异地网络传输，采用SQL合并及压缩提升处理性能
- 基于行格式binlog的高效处理及传输，有效提升RPO
- 备库在复制过程中始终处于运行等待状态，基于主键归并，加速并行回放效率，有效提升RTO

QRep数据复制组件架构



Gargou : 主端组件

Emmet : 备端组件

1. 主端并行收集各个业务库的事务binlog，基于主键合并RowChange，序列化压缩成一个work进行传输
2. 备端接收work，解压，反序列化成RowChange，置于workpool
3. 备端启动多个worker并行回放。基于主键合并的RowChange不存在相同主键，与顺序无关。

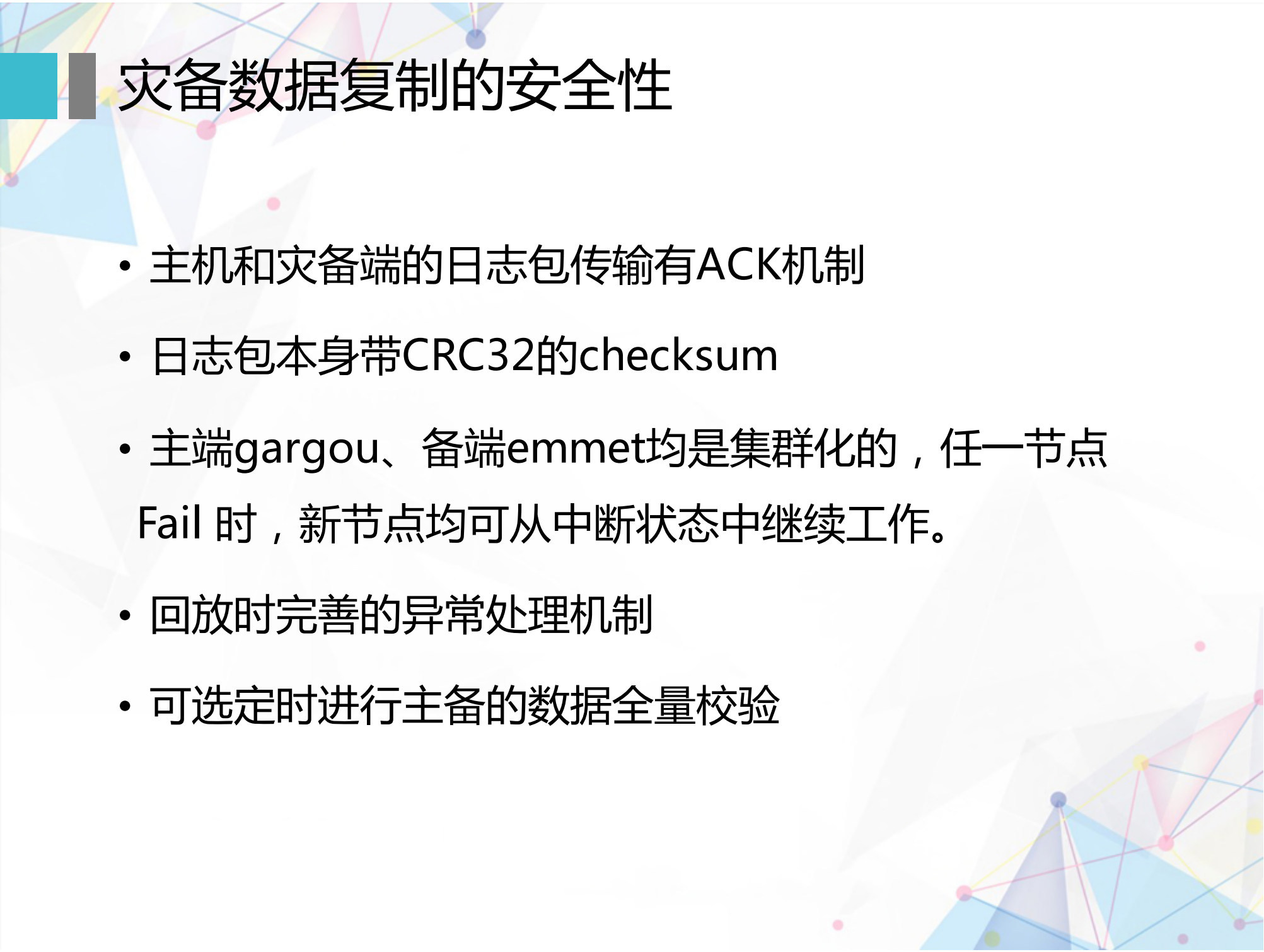
爱可生灾备数据复制组件设计原理

- 数据传输

- 原生replication是流式，基于event，粒度小，压缩比小，压缩和解压消耗时间和传输收益相比较较大
- 灾备数据复制工具可配置最大粒度（缺省3M），较原生的粒度大，压缩比高，所以压缩后传输效率更高（gzip压缩算法）

- 并行回放

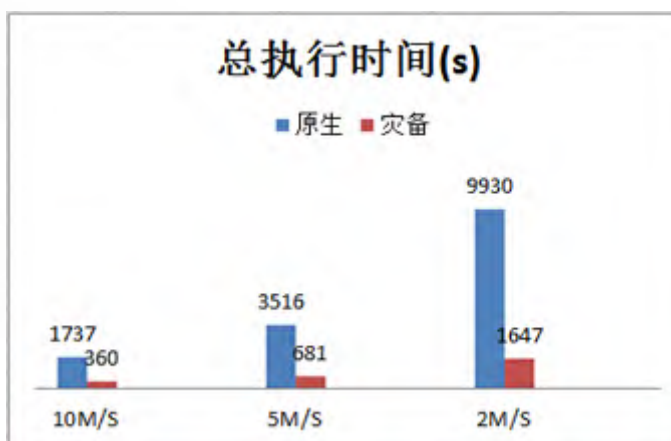
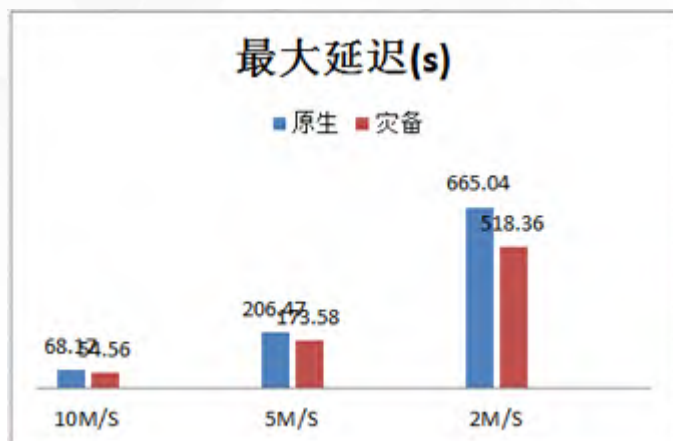
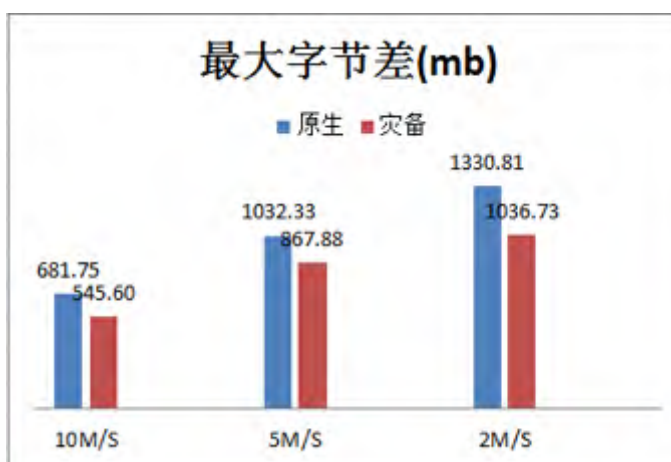
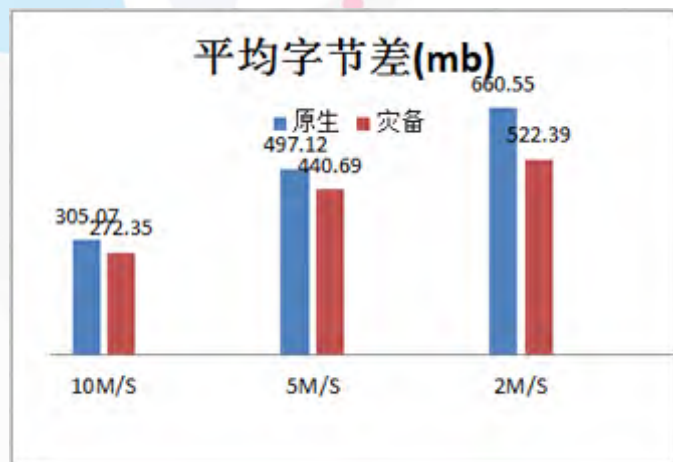
- 基于主键合并的并行回放，提升回放效率
 - insert + update = insert
 - update + update = update
 - insert/update + delete = delete
 - delete + insert = insert
 - delete + delete = delete



■ 灾备数据复制的安全性

- 主机和灾备端的日志包传输有ACK机制
- 日志包本身带CRC32的checksum
- 主端gargou、备端emmet均是集群化的，任一节点Fail时，新节点均可从中断状态中继续工作。
- 回放时完善的异常处理机制
- 可选定时进行主备的数据全量校验

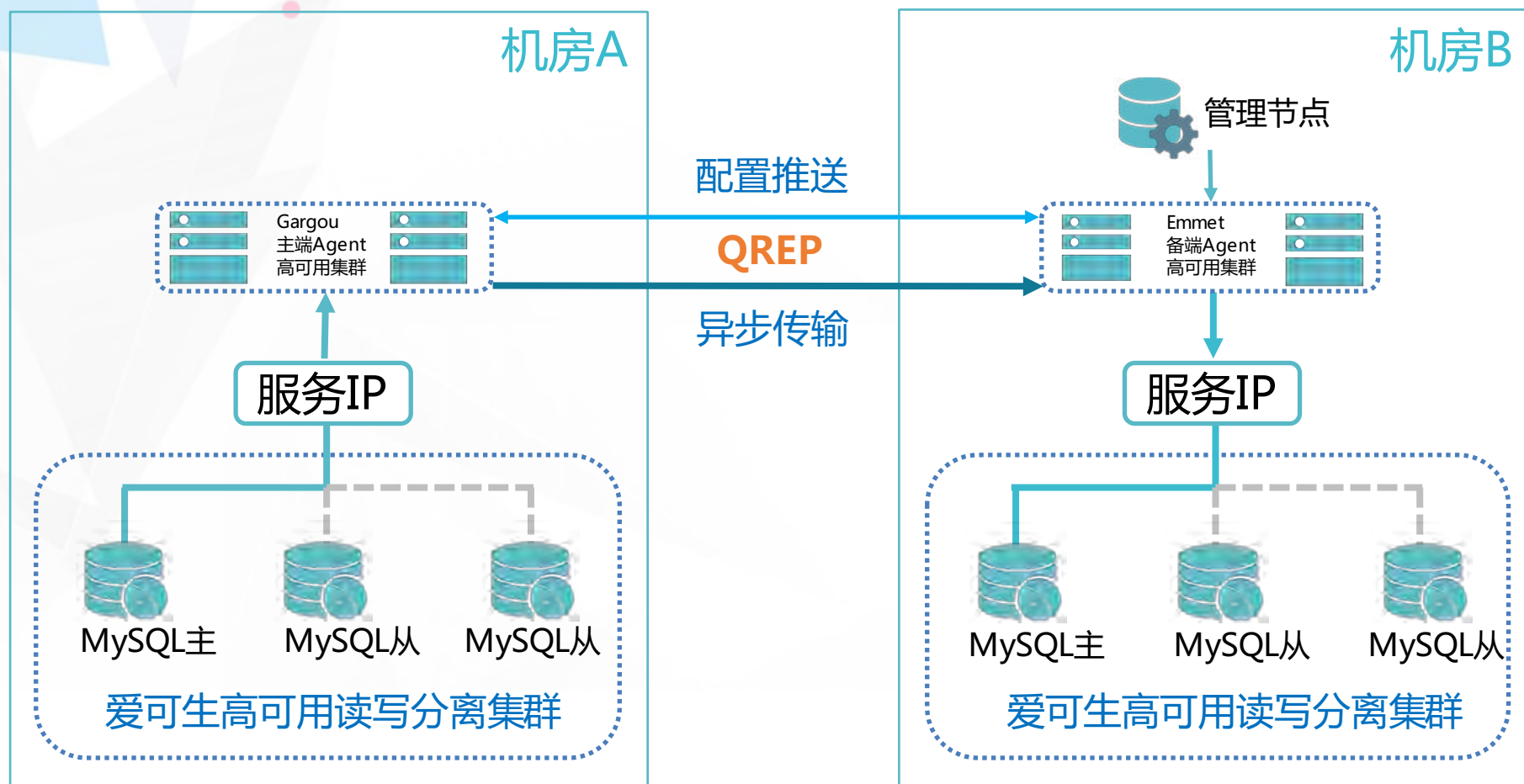
爱可生灾备数据复制组件评测结果



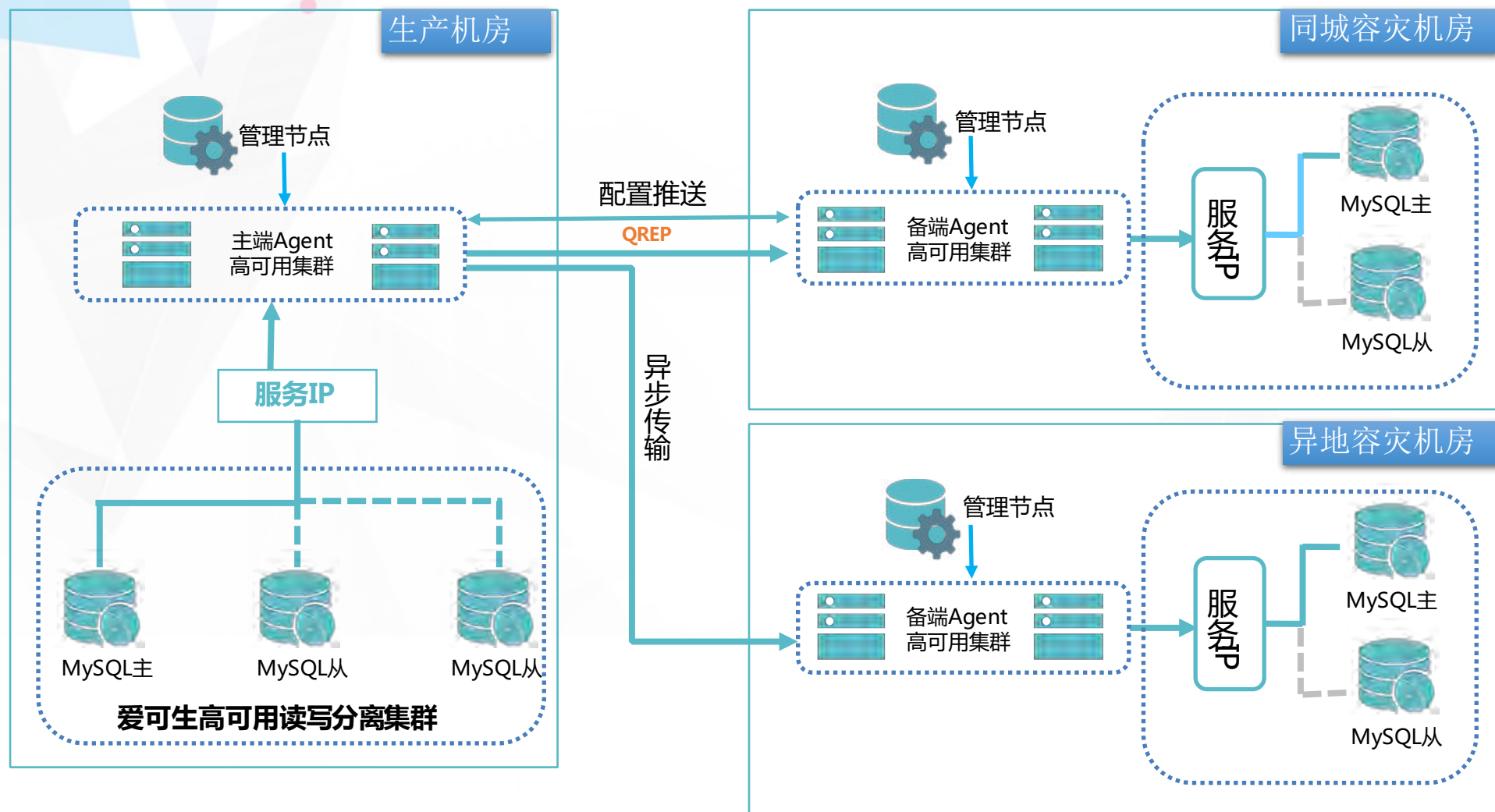
- 分别在2M、5M、10M三种网络环境进行压测
- 总执行时间爱可生复制组件比原生组件快**4倍以上**
- **在2万左右QPS，灾备复制工具可以基本无延迟**

基于 QRep 的灾备场景解析

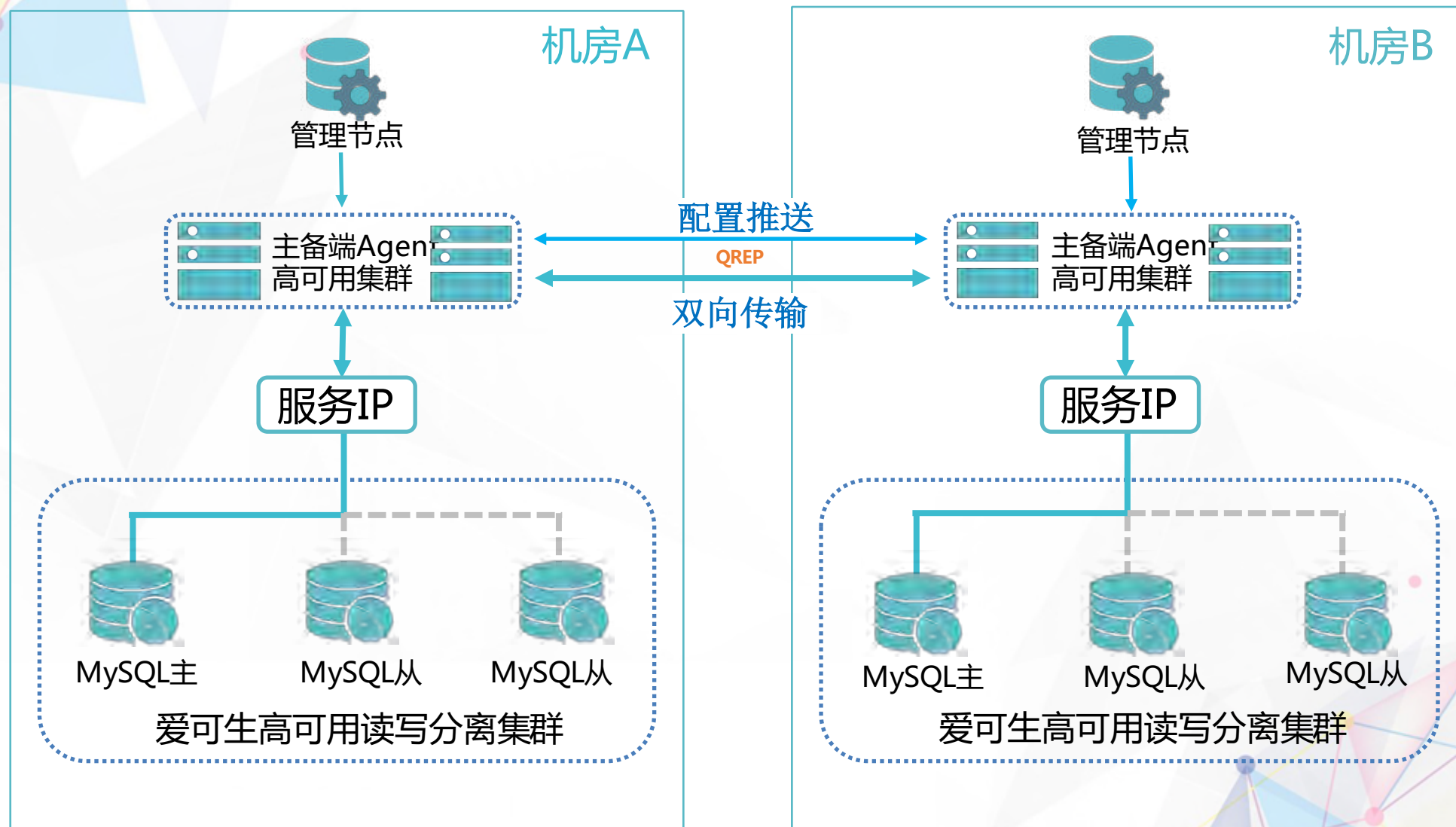
1. 基于Qrep的异地容灾方案



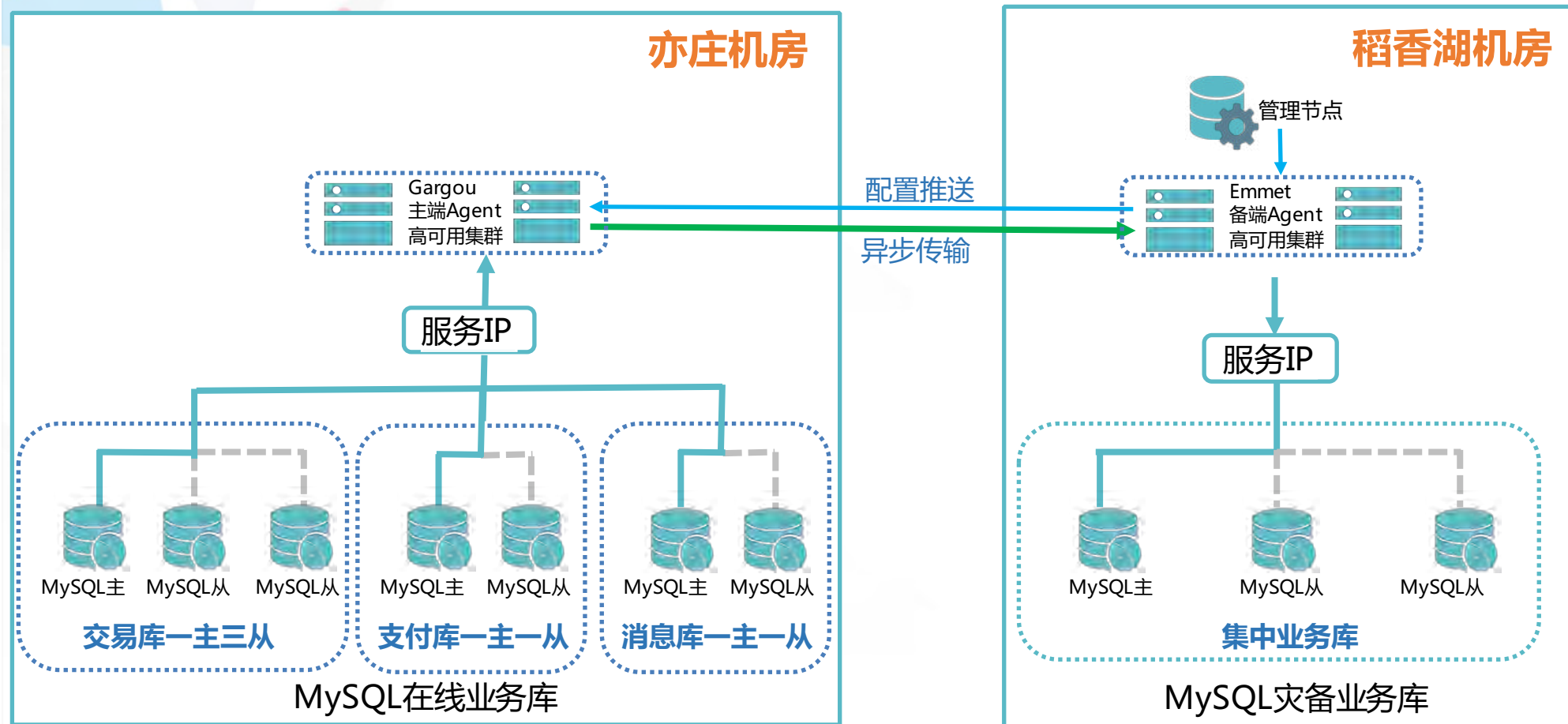
2. 基于QRep的两地三中心场景



3. 基于Qrep的双活互备场景



案例：某互联网金融MySQL灾备



THANKS

SequeMedia
盛拓传媒

IT168.com
专注 移动 10 年

ChinaUnix

ITPUB
www.itpub.net