

异地多活-携程MySQL 实时双向(多向)复制

携程框架架构部 / 李明冬

QCon+ 案例研习社



扫码学习大厂案例

学习前沿案例，向行业领先迈进

40⁺

热门专题

—
行业专家把关内容筹备，
助你快速掌握最新技术发展趋势

200⁺

实战案例

—
了解大厂前沿实战案例，
为 200 个真问题找到最优解

40⁺

直播答疑

—
40 位技术大咖，每周分享最新
技术认知，互动答疑

365⁺

持续学习

—
视频结合配套 PPT
畅学 365 天

大纲

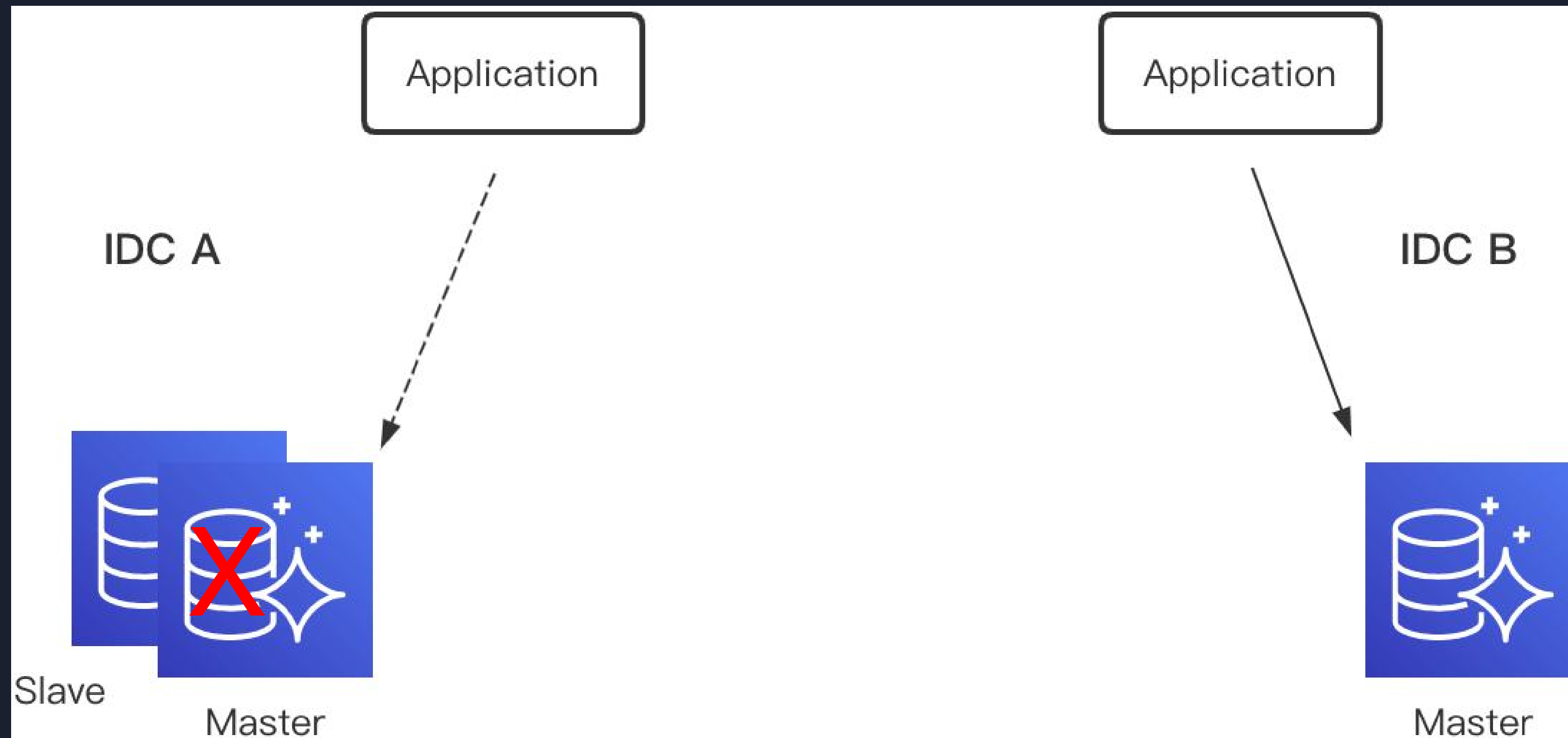
- 异地多活整体介绍
- DRC 详细设计
- 公有云复制
- 业务落地遇到的问题

1

异地多活整体介绍

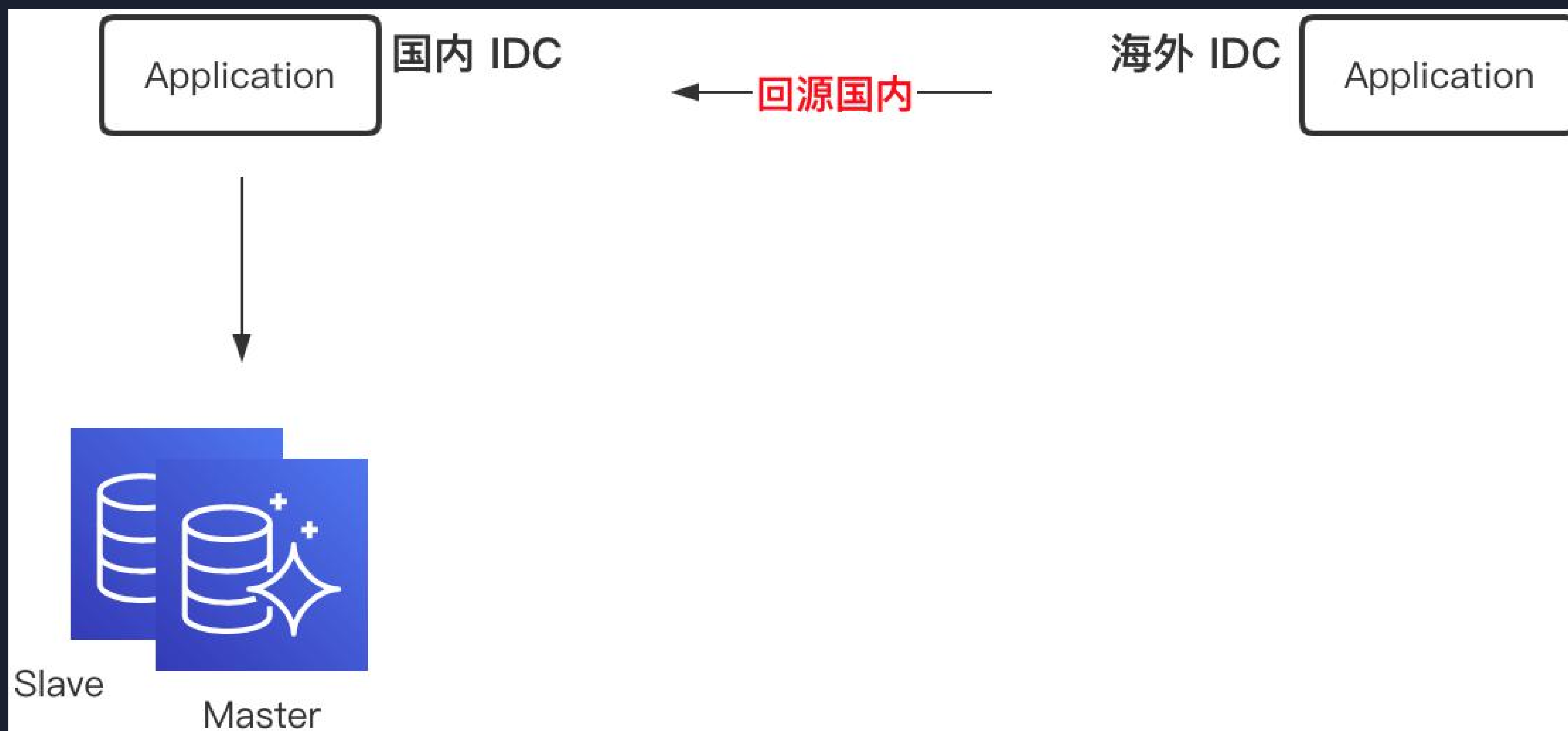
机房故障

- 机房故障，快速恢复？



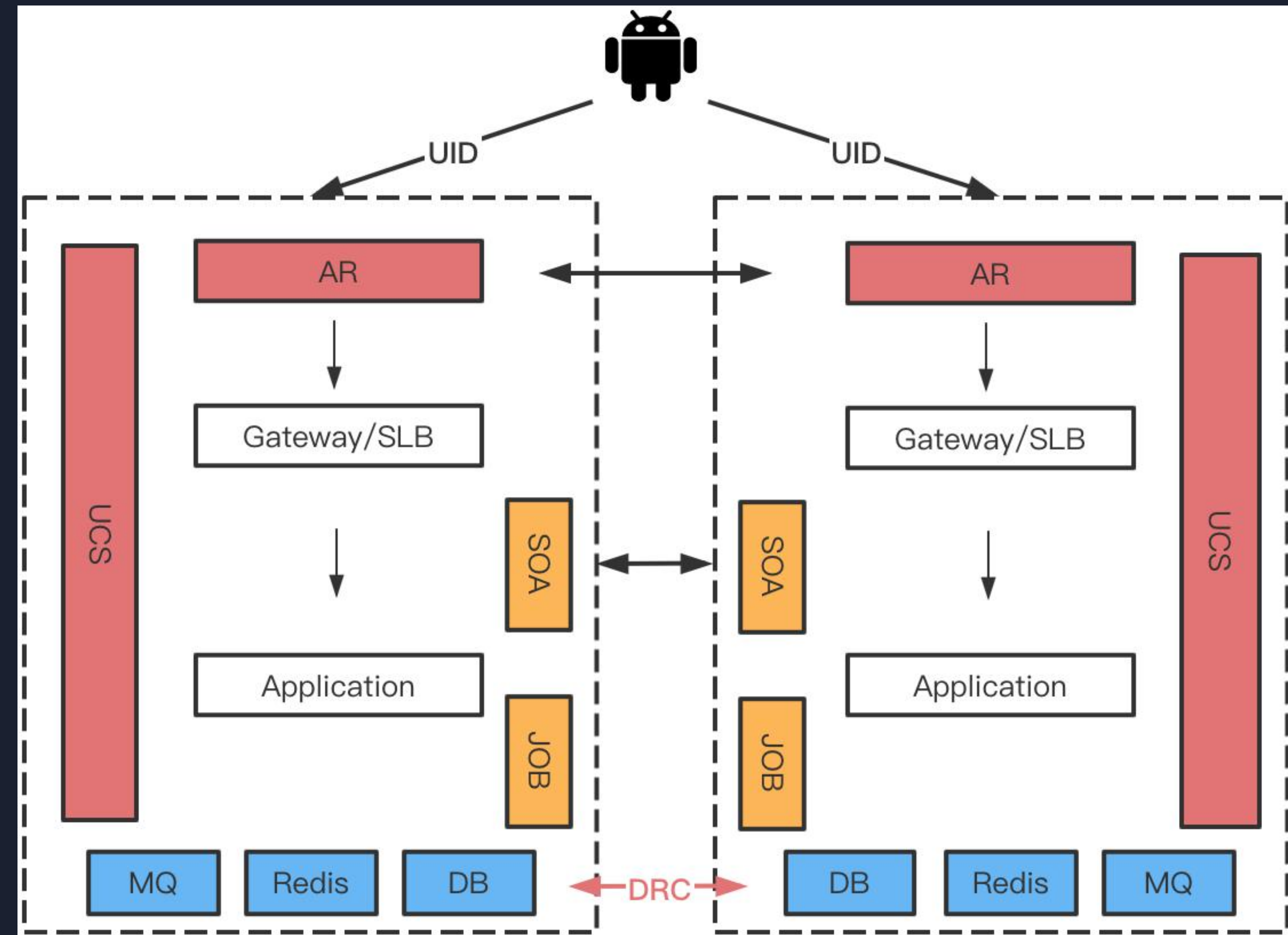
业务国际化

- 业务出海，数据回源国内？



异地多活

- **AP**: 最终一致性, 数据复制
- **CP**: 强一致性, 分布式数据库
- **API Router**: UID维度流量路由
- **Unit Control Service**: 管理多活路由
- **Data Replication Center**: 数据复制



数据出海

- 数据双向复制，就近读写
- 海外故障，流量切回国内

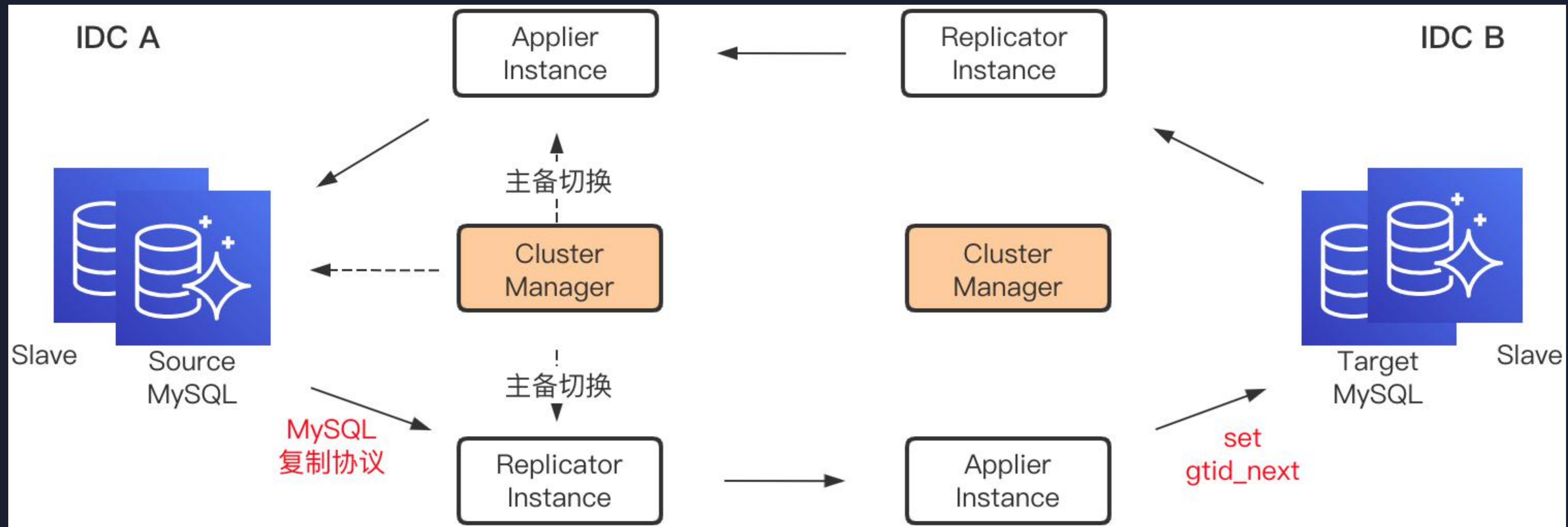


2

DRC 详细设计

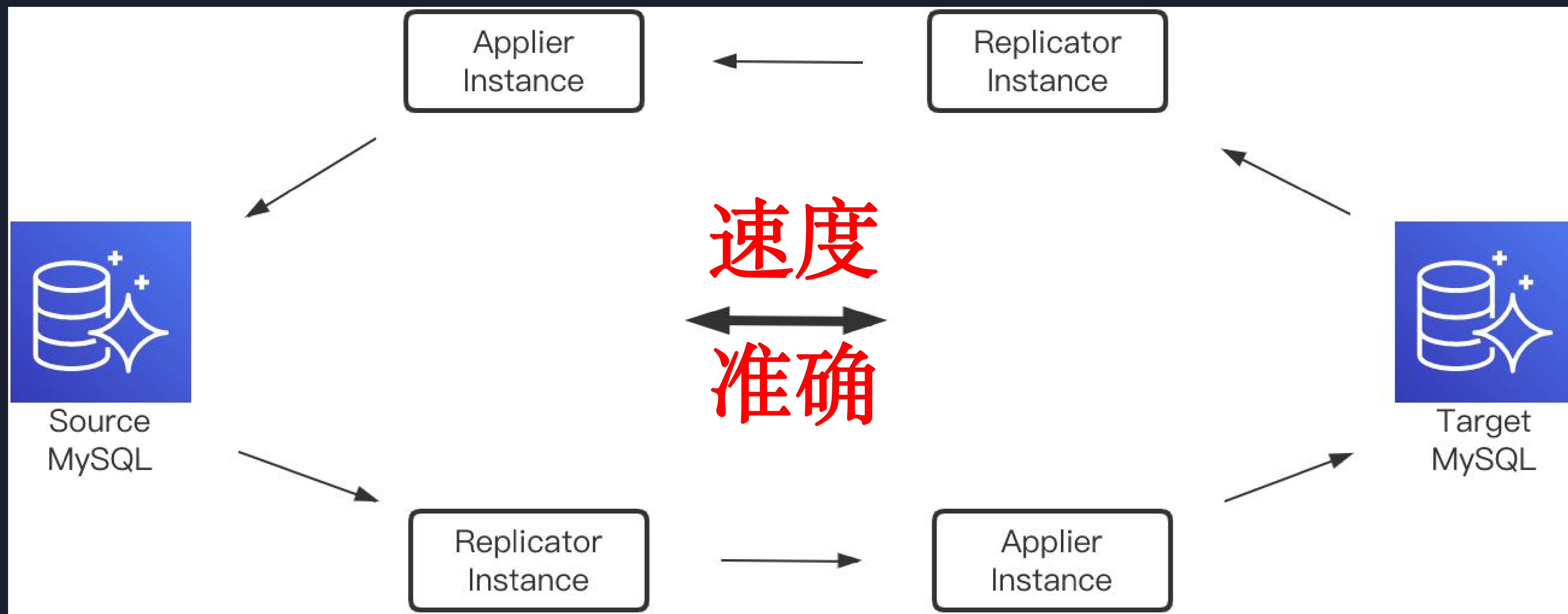
DRC架构

- Replicator: 拉取Binlog
- Applier: 解析Binlog, 复制到目标端MySQL
- Cluster Manager: 负责系统高可用



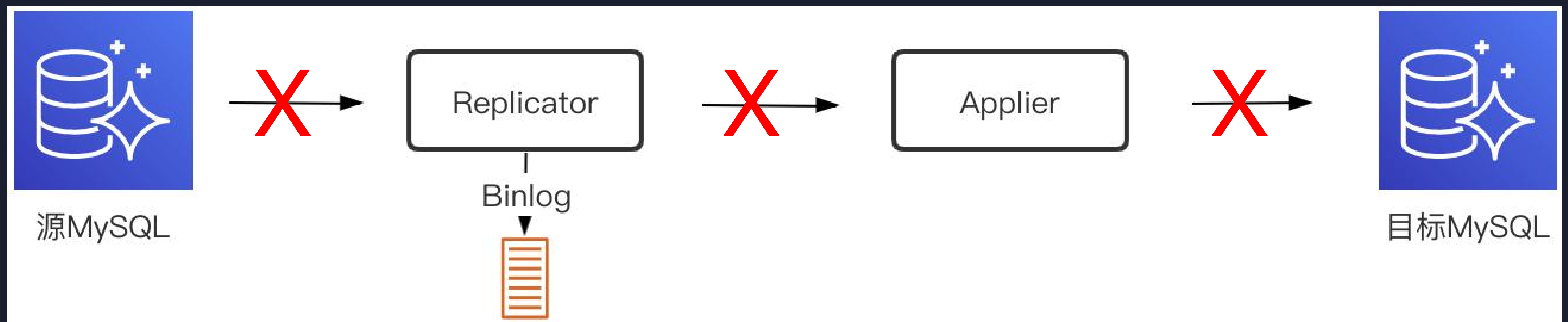
核心指标

- 复制延迟
- 数据一致性



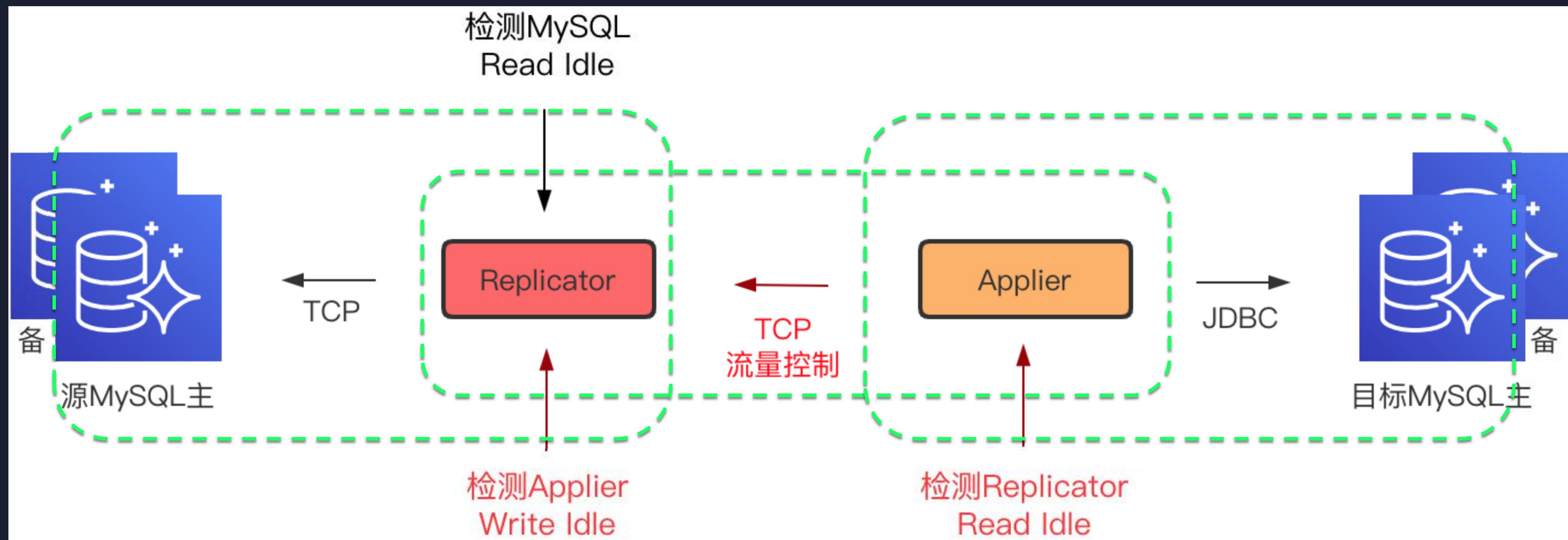
复制延迟

- 网络故障：快速检测恢复
- Replicator：高效处理I/O
- Applier：大流量、大事务场景处理



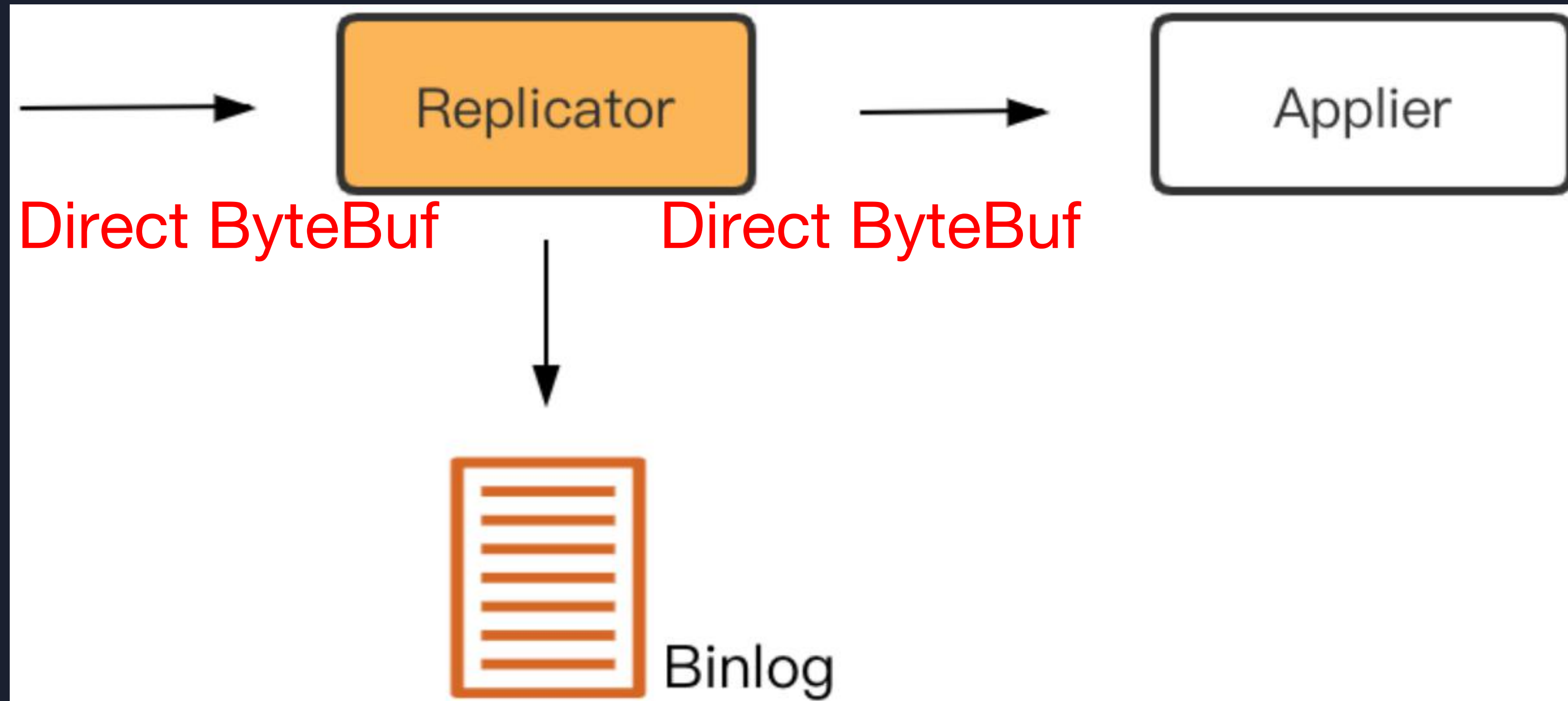
复制延迟-空闲检测

- 接收Binlog：检测读空闲，断开连接
- 发送Binlog：检测写空闲，发送心跳



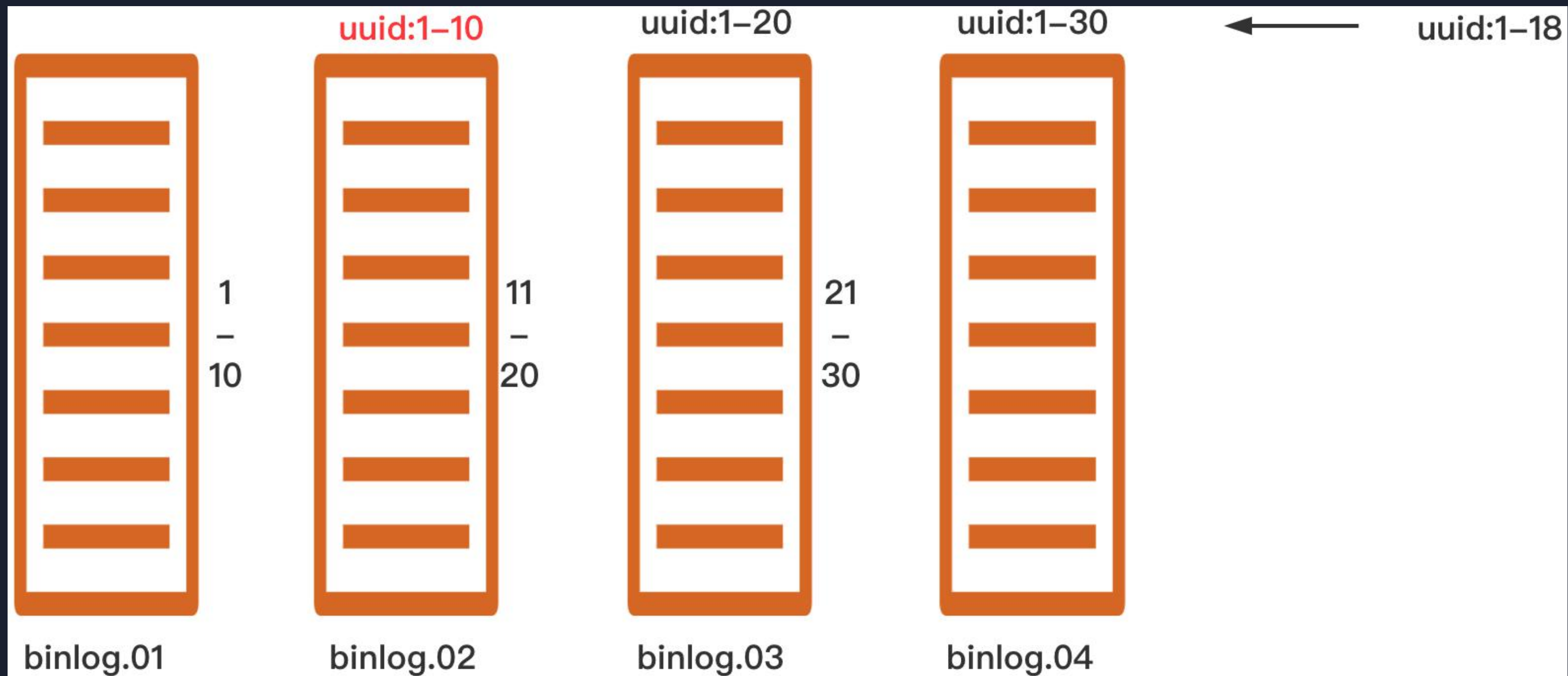
复制延迟-I/O

- 接收发送Binlog: Direct ByteBuf
- Binlog落盘: Page Cache, 异步刷盘



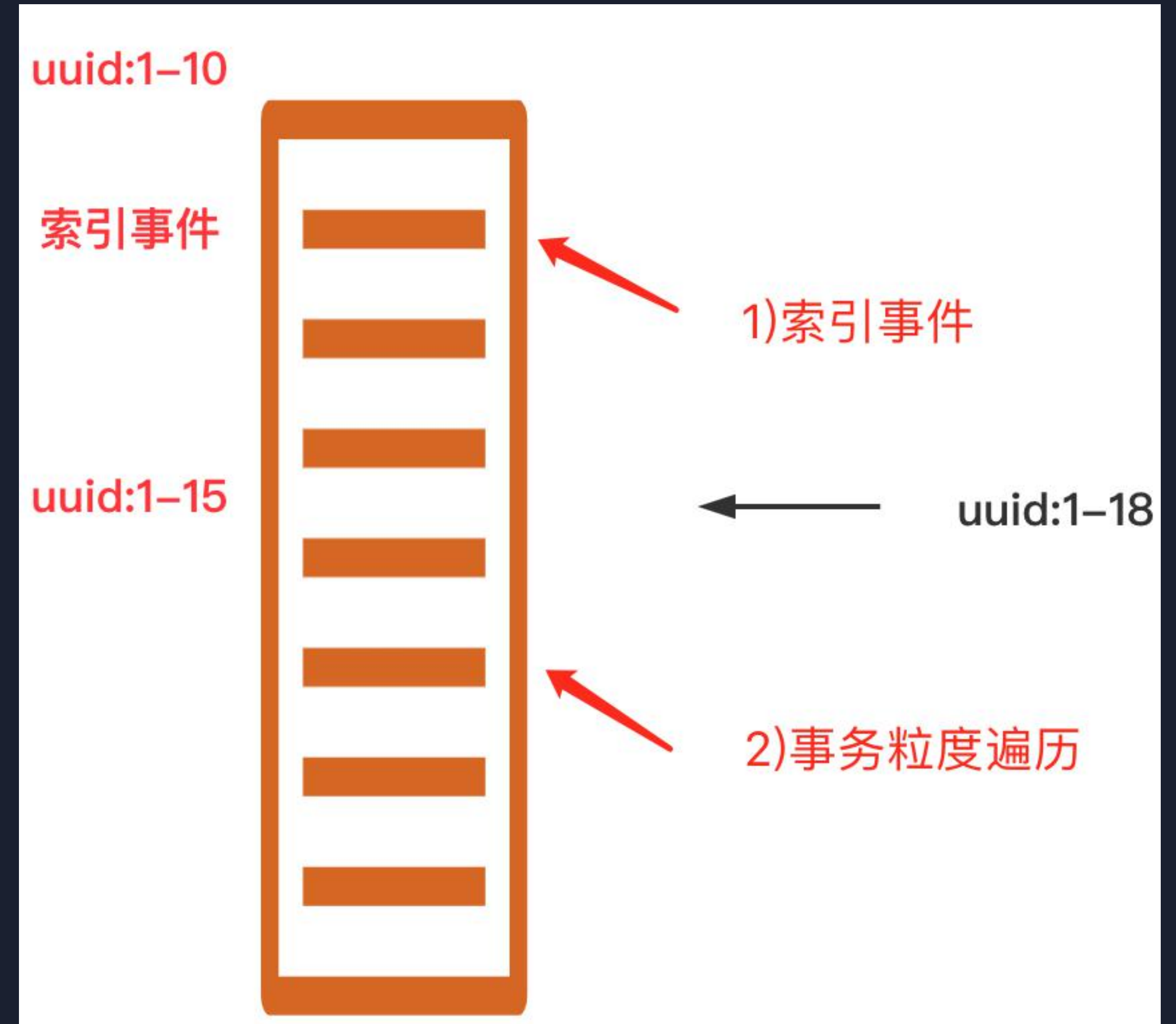
复制延迟-位点定位

- 文件定位：倒序、gtidset子集
- 事件定位：事件粒度顺序遍历



复制延迟-位点定位优化

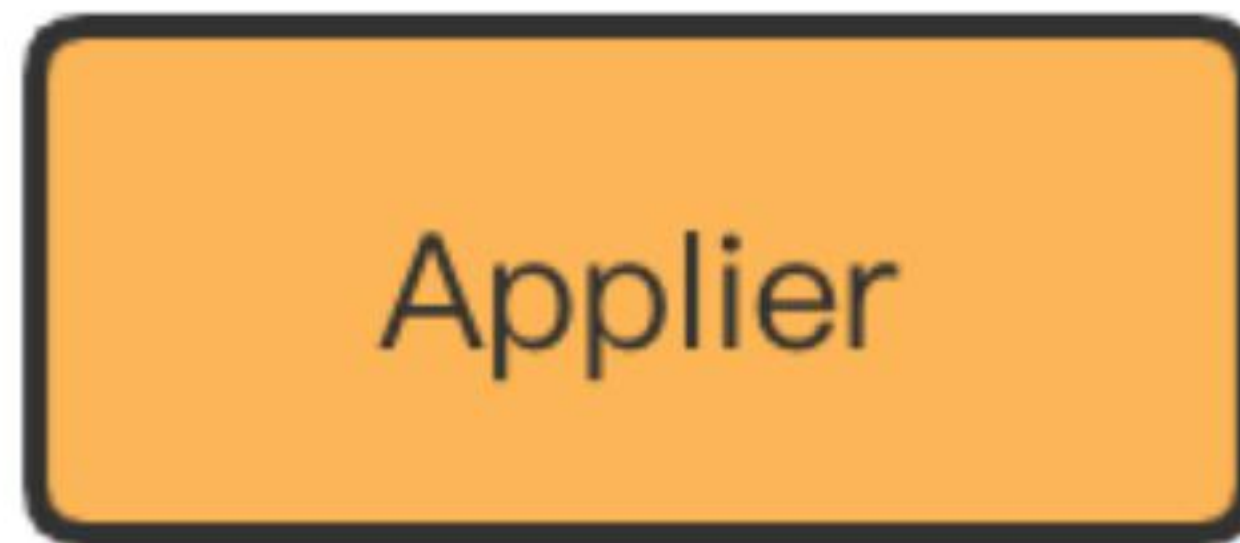
- 块粒度：50M间隔插入
previous_gtid_event
- 事务粒度：gtid事件增加事务偏移量



复制延迟-大流量

- 度量指标：事务数和直接内存使用量
- 限流响应：关闭autoRead

→
autoRead

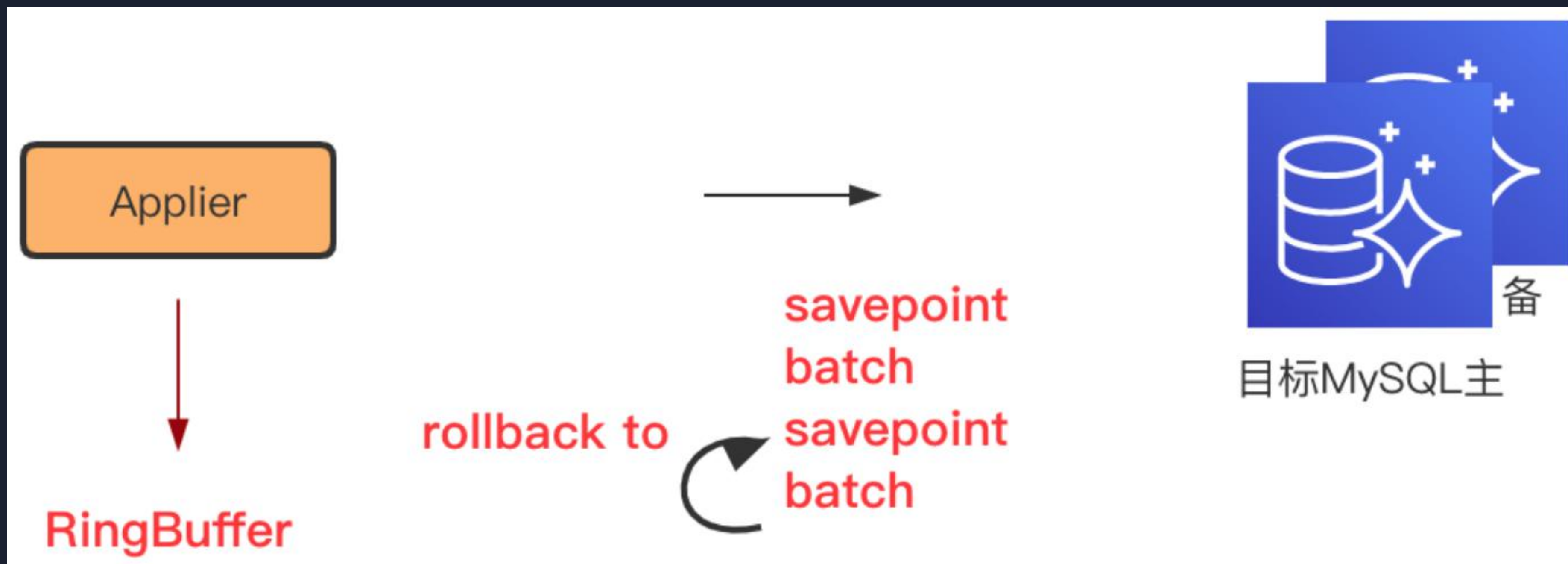


目标MySQL主

1. Writeset并行复制
2. 事务数
3. 直接内存使用量

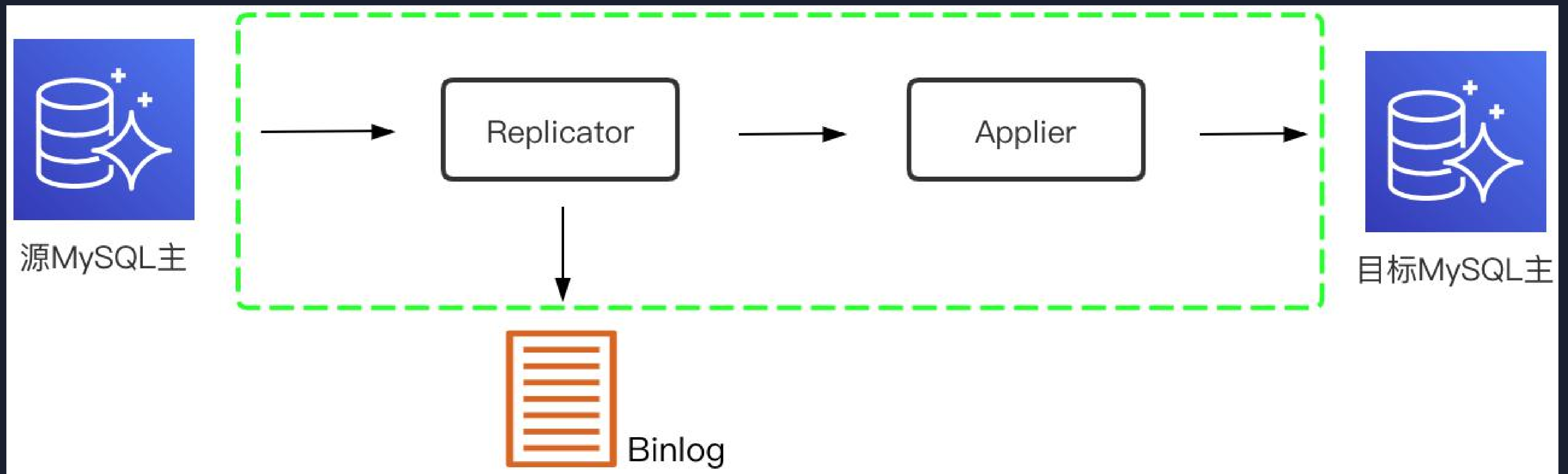
复制延迟-大事务

- 存储: RingBuffer
- 执行: 分割多个子事务



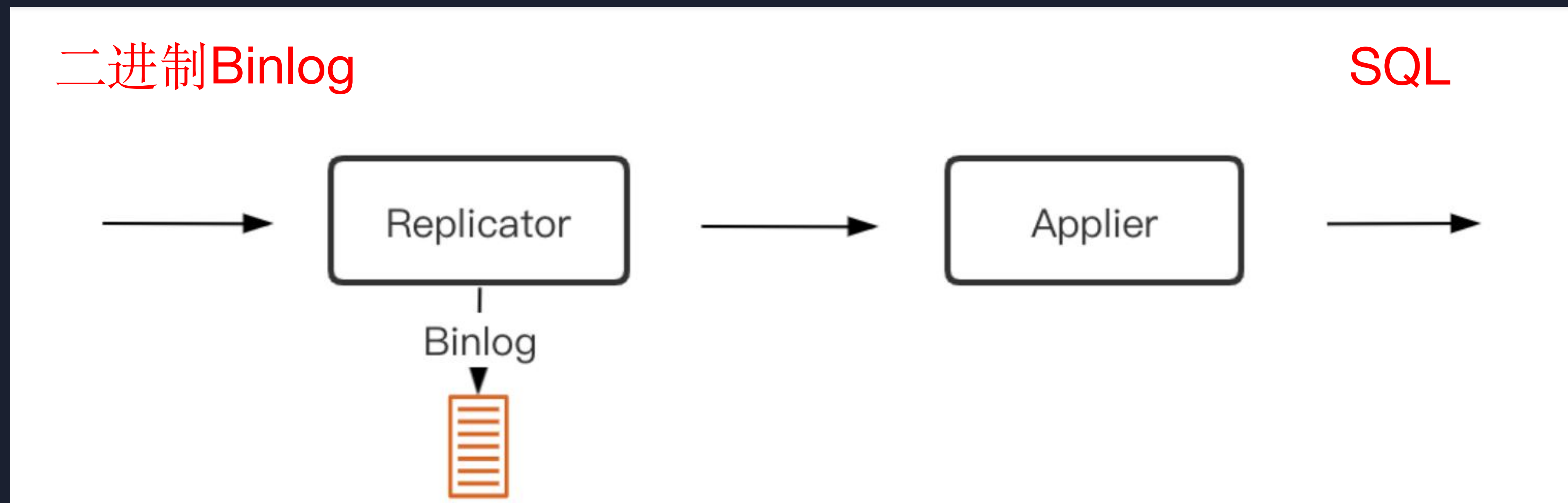
数据一致性

- 自增ID
- 多机房MySQL同一条数据被修改



数据一致性

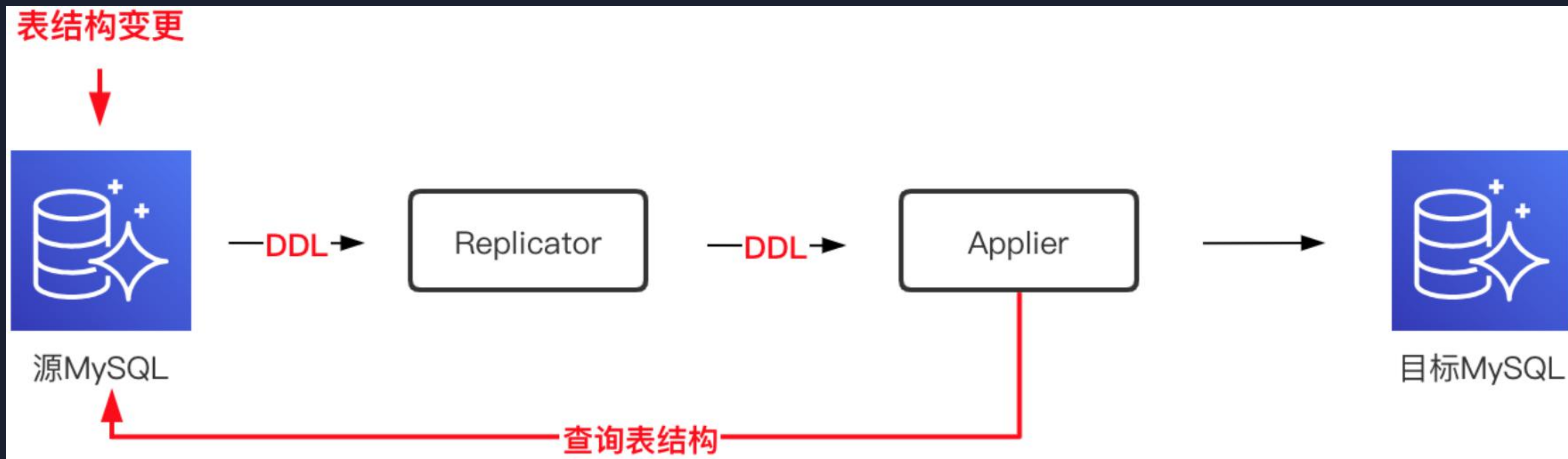
- Binlog正确解析为SQL语句



数据一致性-数据冲突

- 自增ID不冲突：分布式唯一ID生成算法或MySQL auto_increment
 - $\text{auto_increment_increment} = 2^N$ ($N=1,2,3,\dots$)
 - $\text{auto_increment_offset} = M$ ($M=[1, 2^N]$)
- 多个机房同时修改同一条数据
 - 以ON UPDATE CURRENT_TIMESTAMP时间戳新的数据为准
 - 用户可手动处理

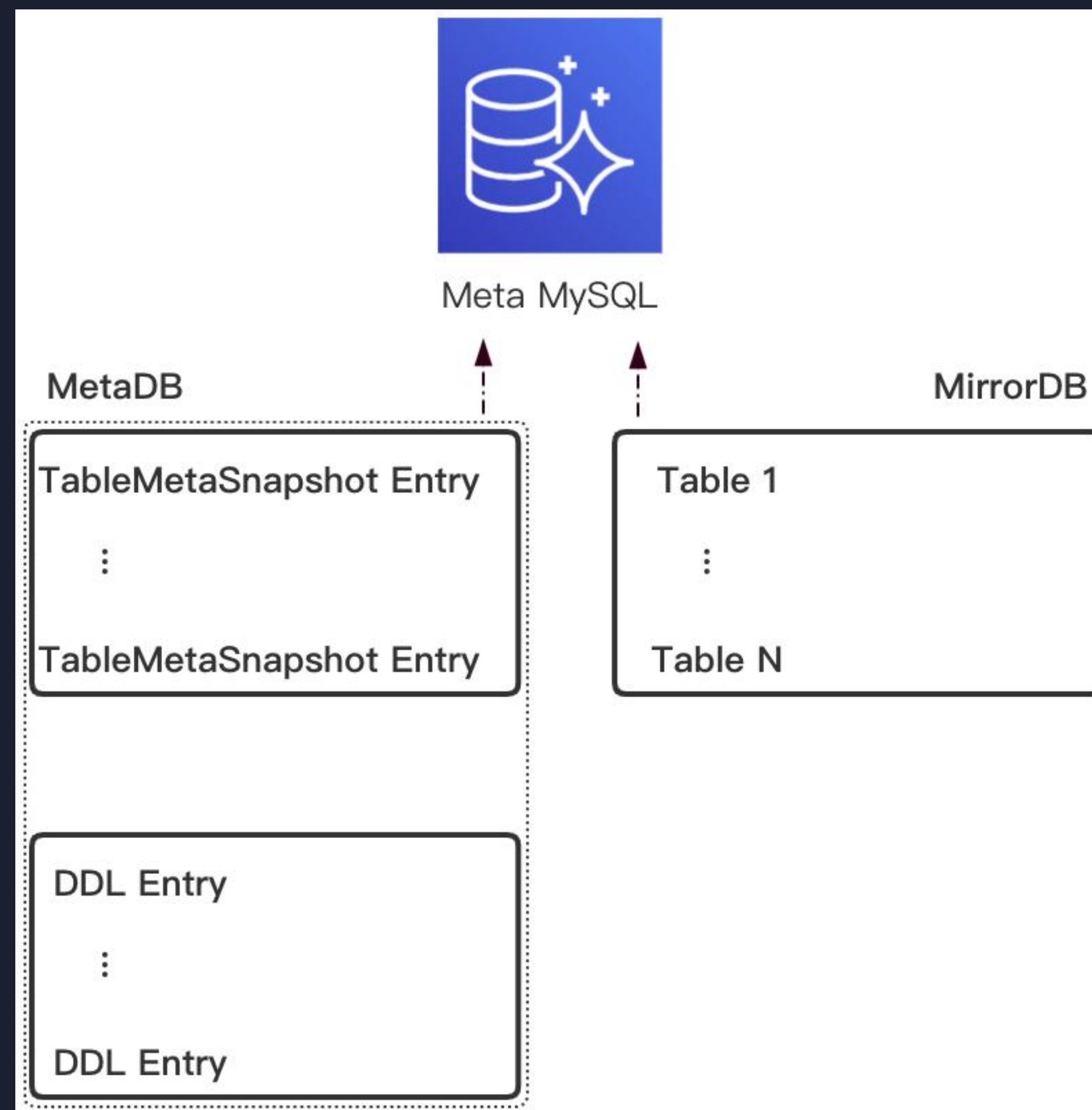
数据一致性-表结构变更



- Binlog异步发送，从何处拉取表结构信息？
- Applier消费速率不匹配时，历史表结构如何获取？

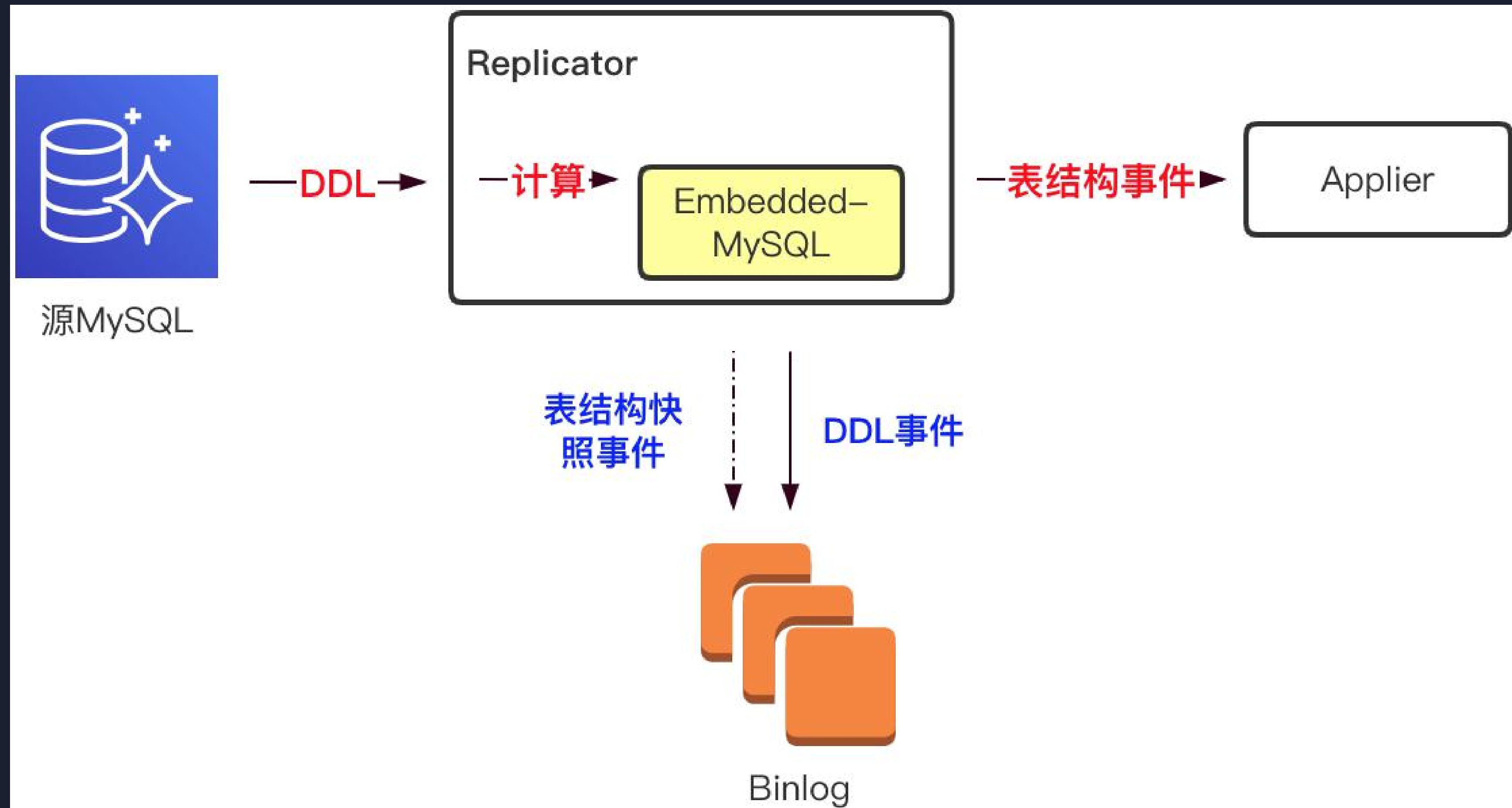
数据一致性-表结构变更业界方案

- 镜像库实时**计算**表结构信息
- 元数据库**存储**表结构信息



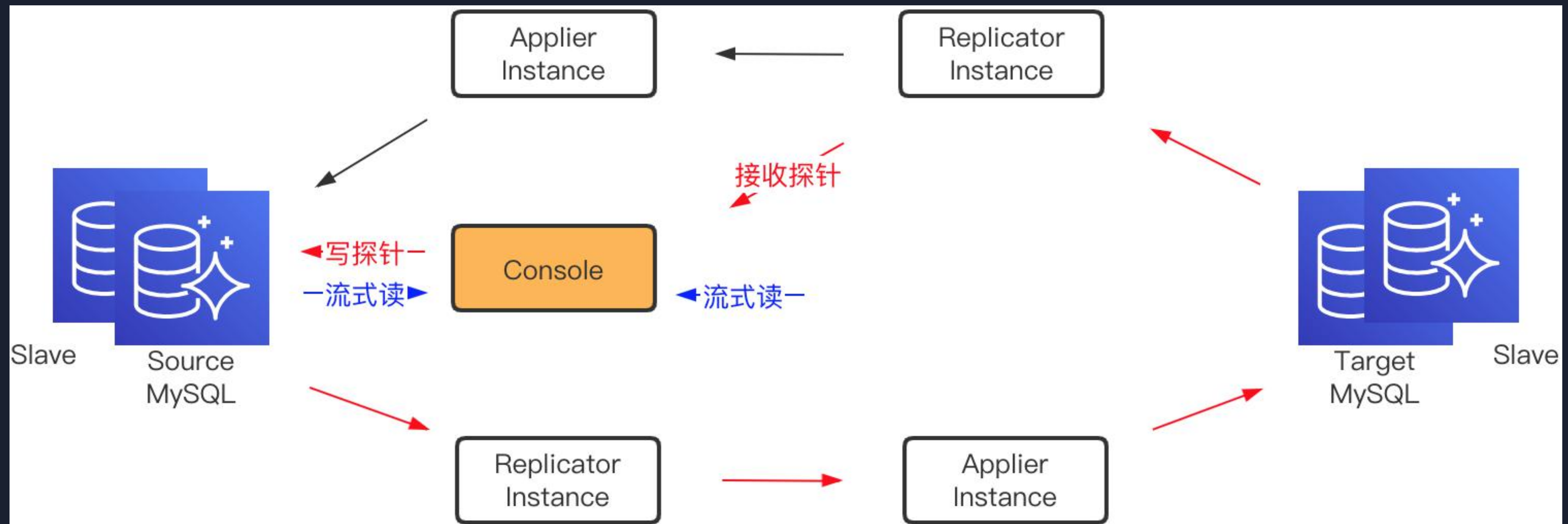
数据一致性-表结构变更Binlog方案

- 计算：内存数据库
- 存储：Binlog文件，
时序恢复



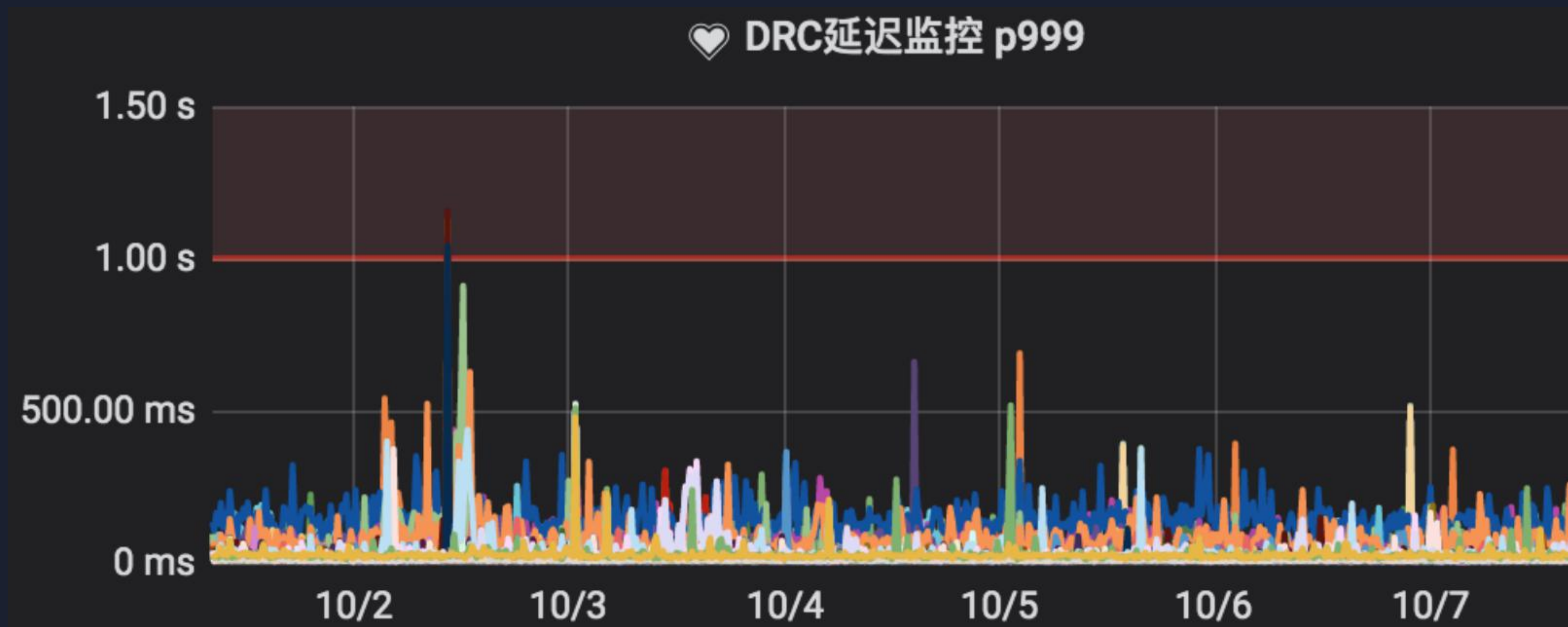
核心指标监控告警

- **延迟**: 同机房写探针, 跨机房读探针
- **一致性**: 新增数据实时比较, 全量数据按需比较



复制延迟监控

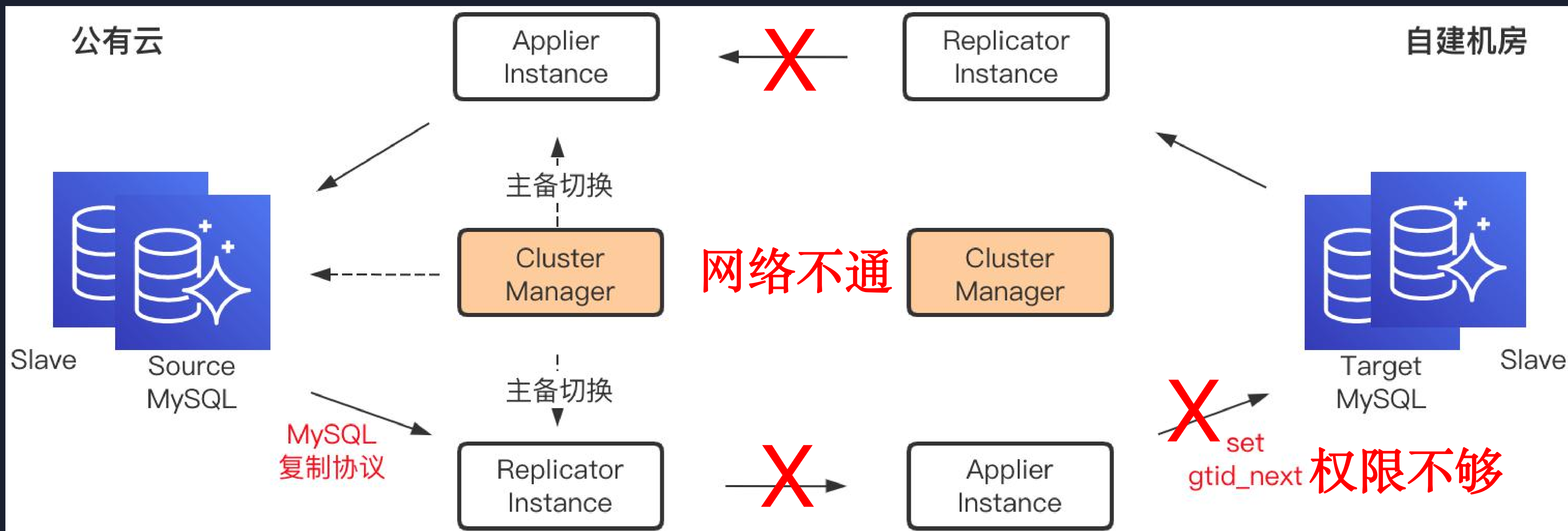
- 999线: $< 1s$



3

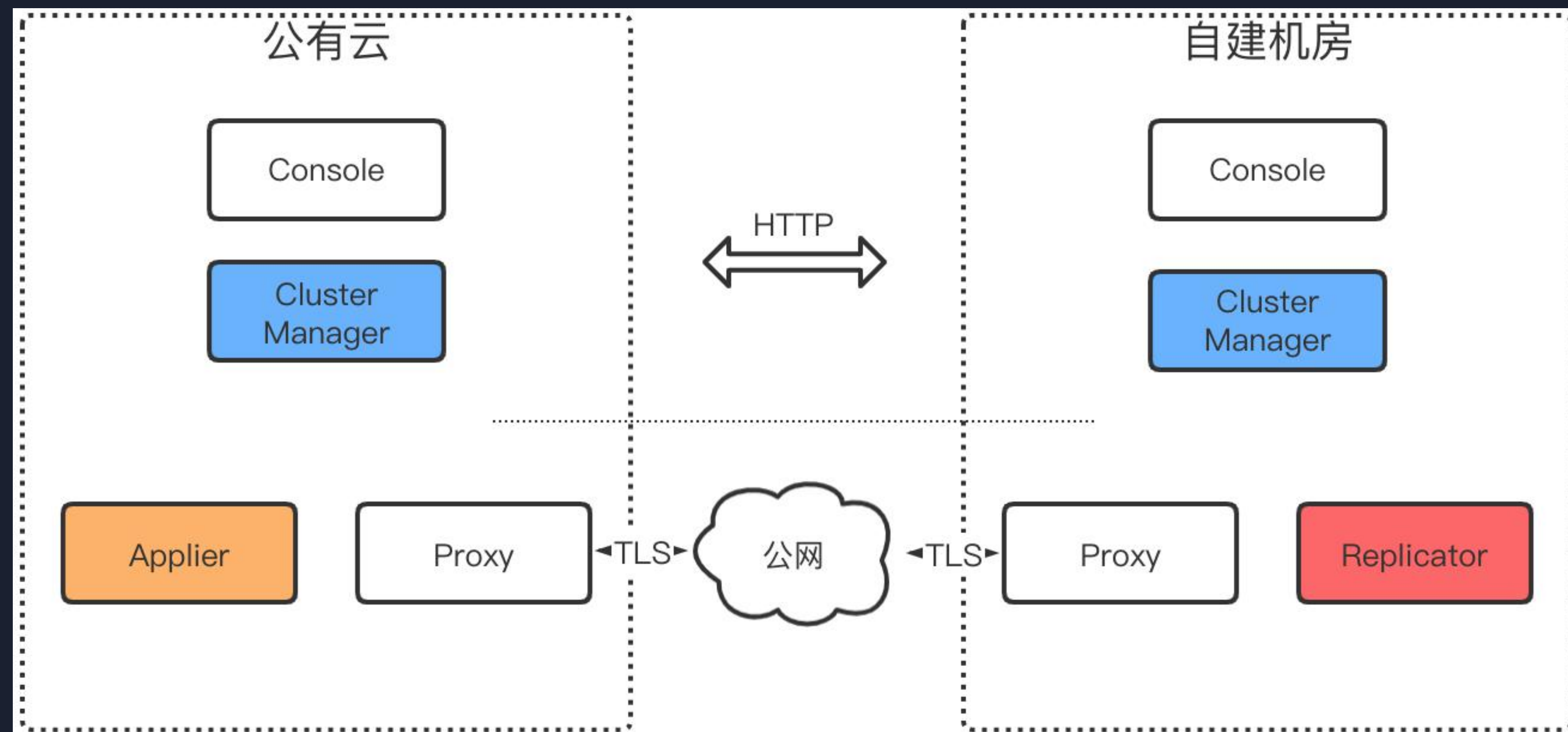
公有云复制

公有云复制



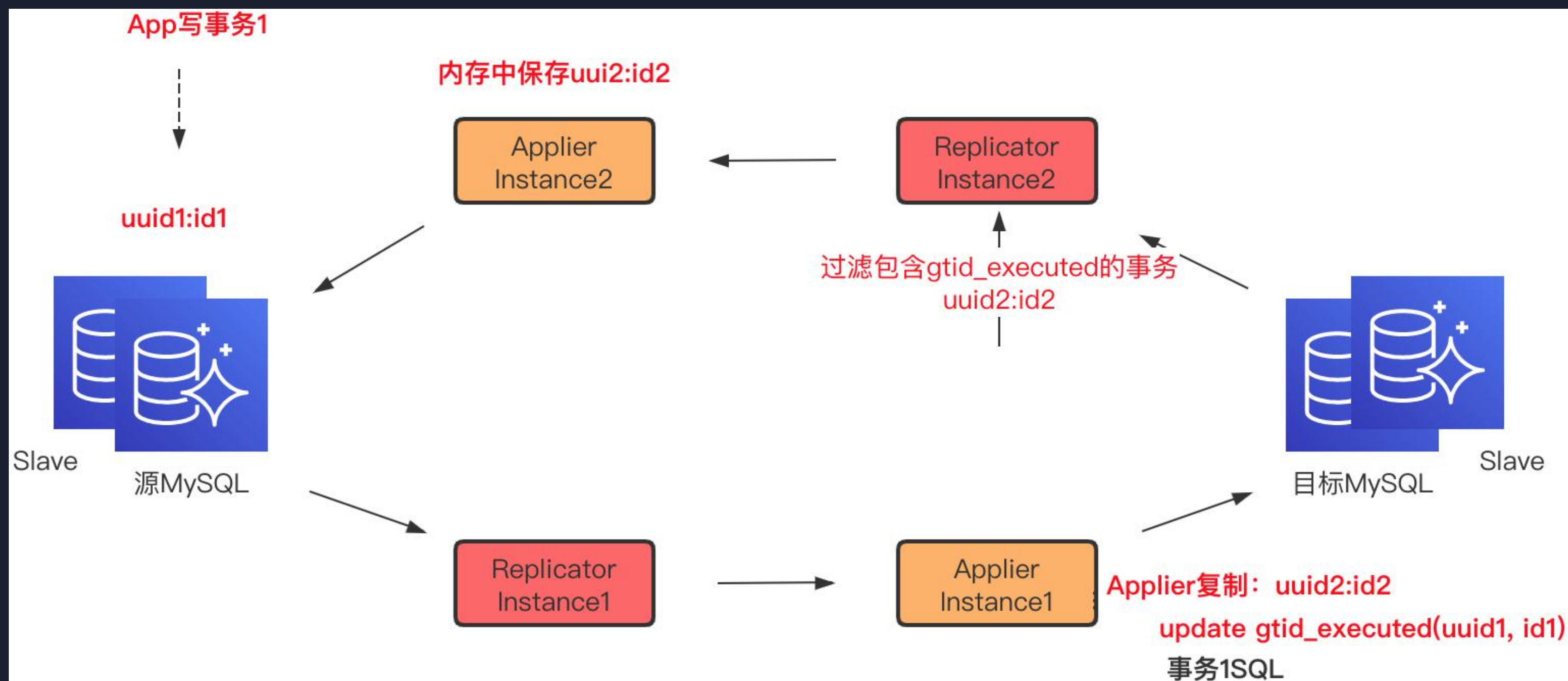
公网传输

- 链路：TLS
- 内网/公网：Proxy转发
- Proxy 客户端



基于事务表复制

- 位点记录：事务表持久化
 - 循环复制：以事务表开头的事务
- 的事务

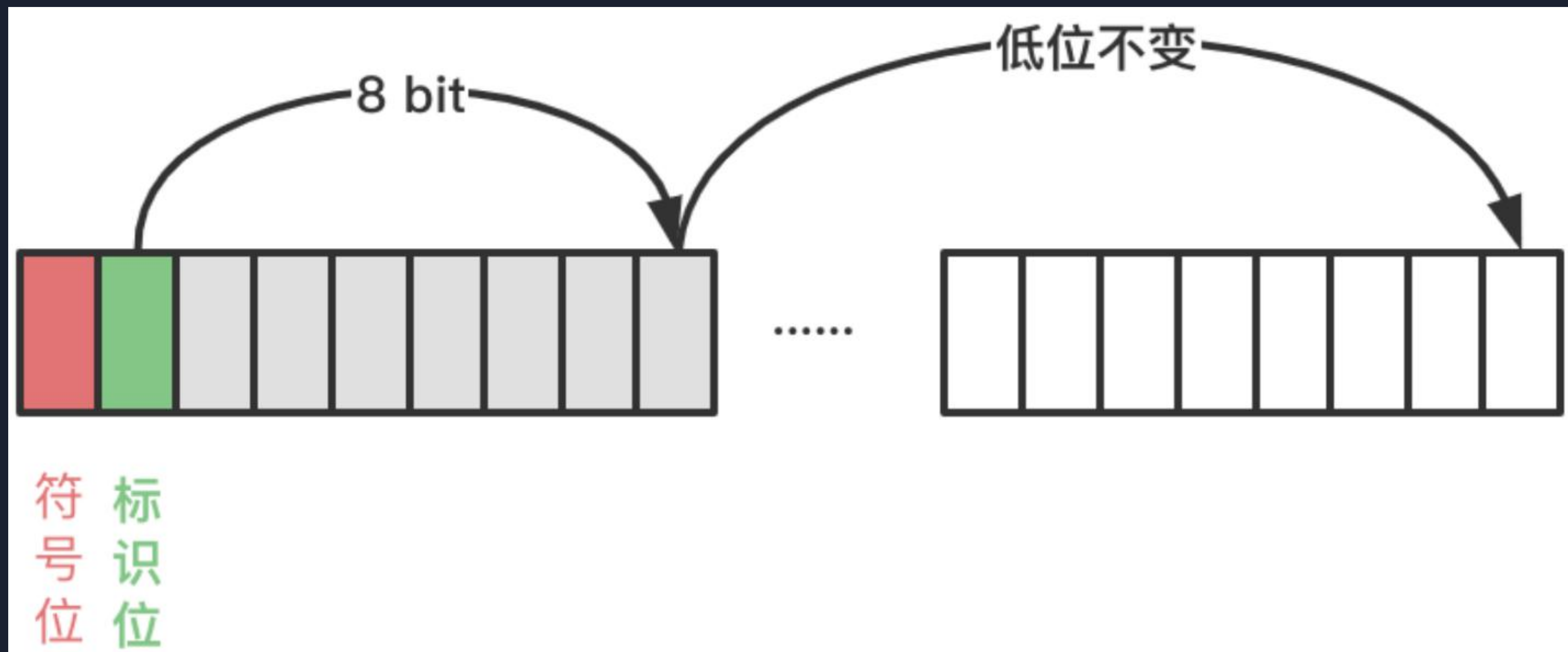


4

业务落地遇到的问题

非UID维度-OrderId改造

- OrderId 高位种入UID模值



总结

- 借助数据复制DRC构建分布式AP系统
- 优化I/O、处理极端场景保证系统可用性
- 引入事务表支持公有云复制， Proxy支持公网
- 非UID维度转化为UID维度减少业务改造成本

为一线互联网公司核心技术 人员提供优质内容

☑ TGO专访

☑ 技术干货

☑ 每周精要

☑ 行业趋势



关注 InfoQ 公众号



THANKS