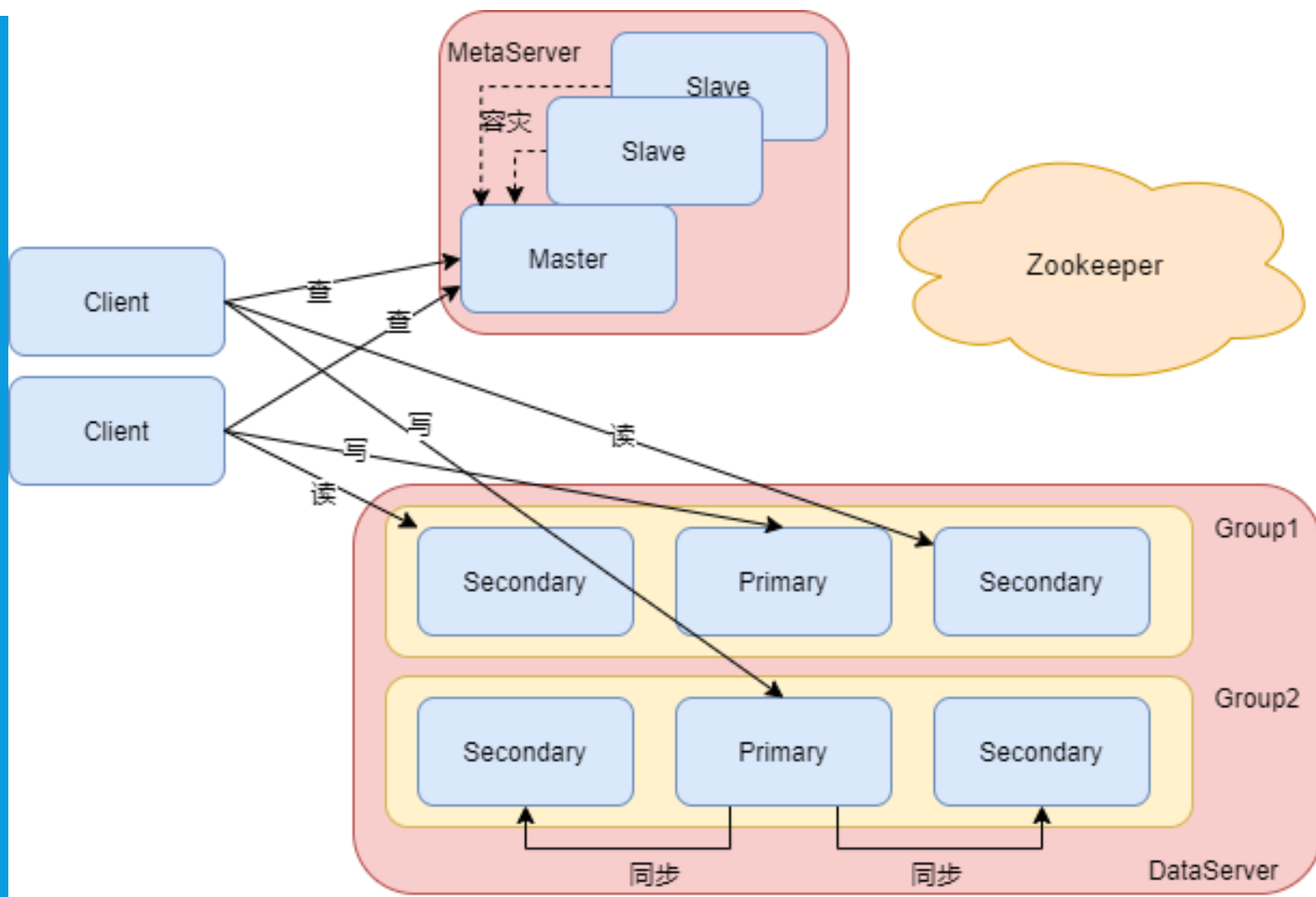
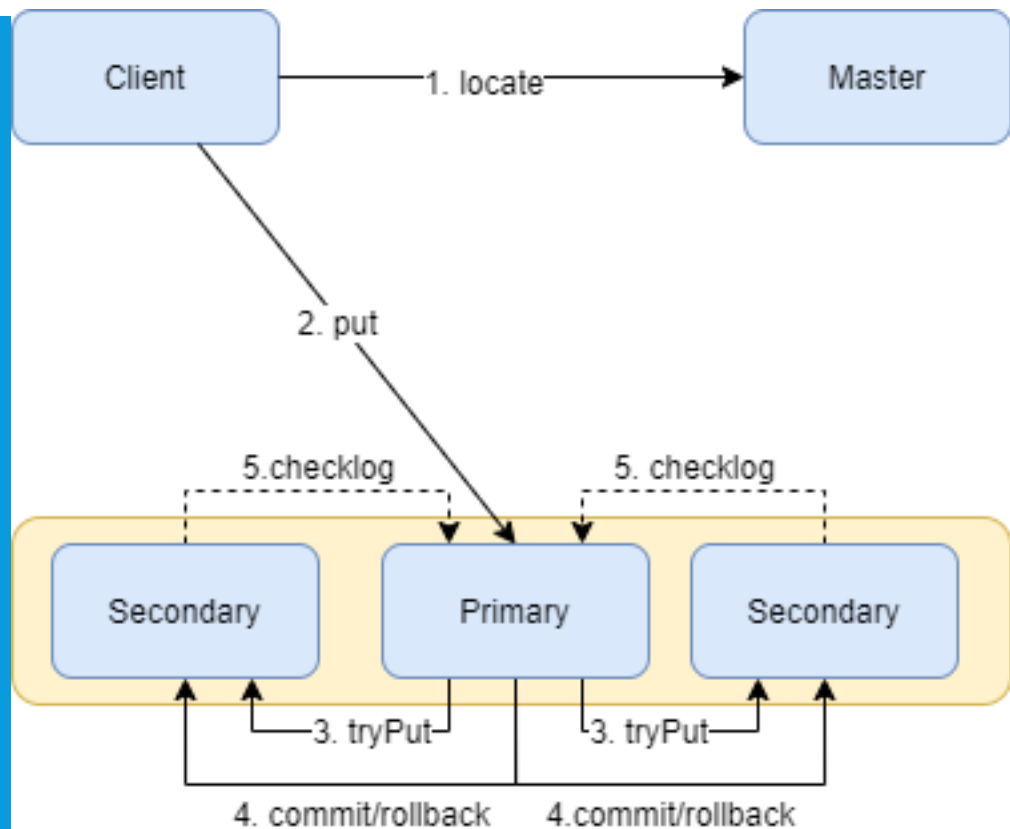


整体架构

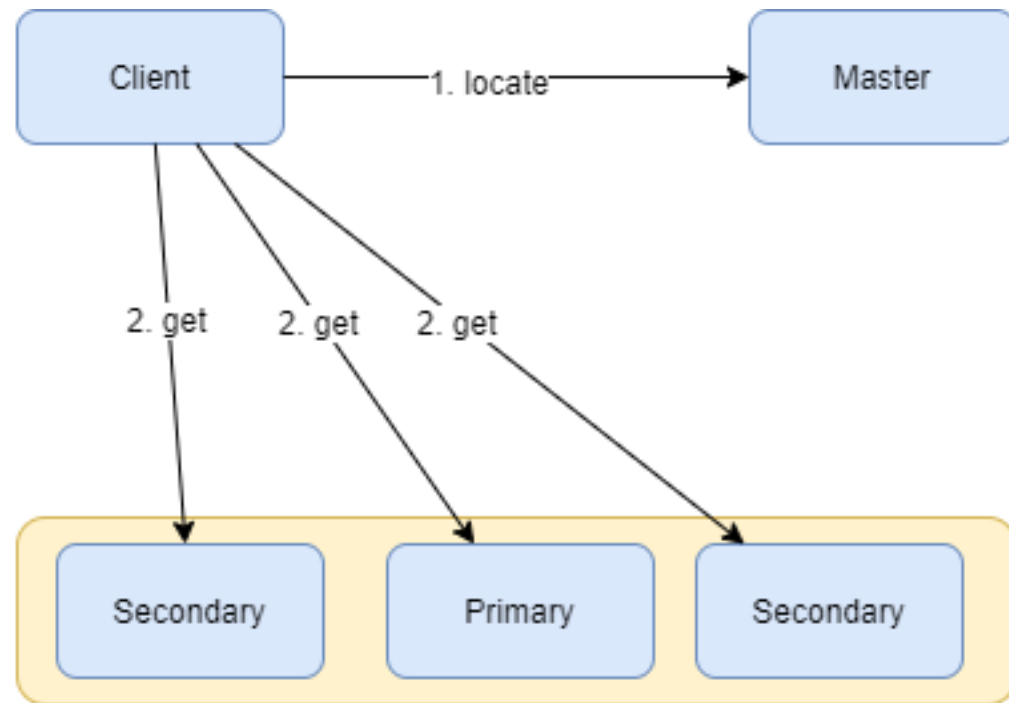
- 多Client并发
- Master/Slave集群
- Primary/Secondary集群
- Zookeeper集群
- 无单点故障！



业务流程

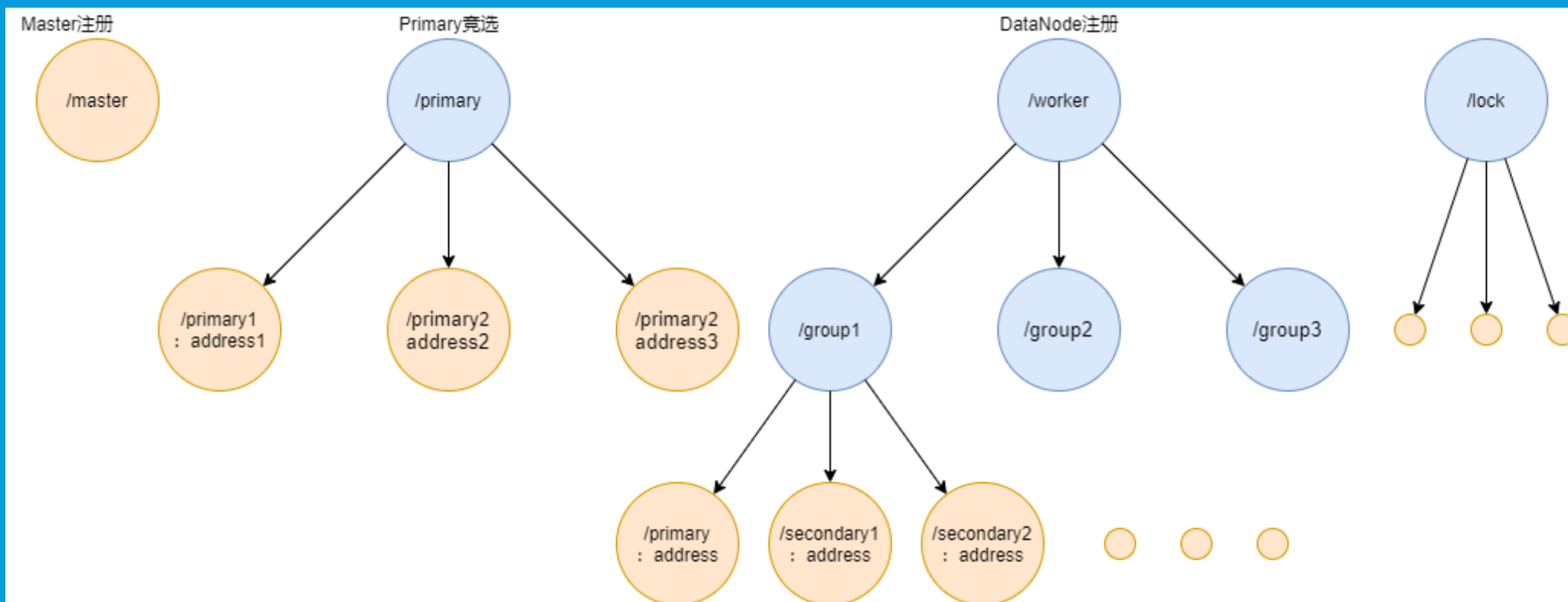


写请求处理流程



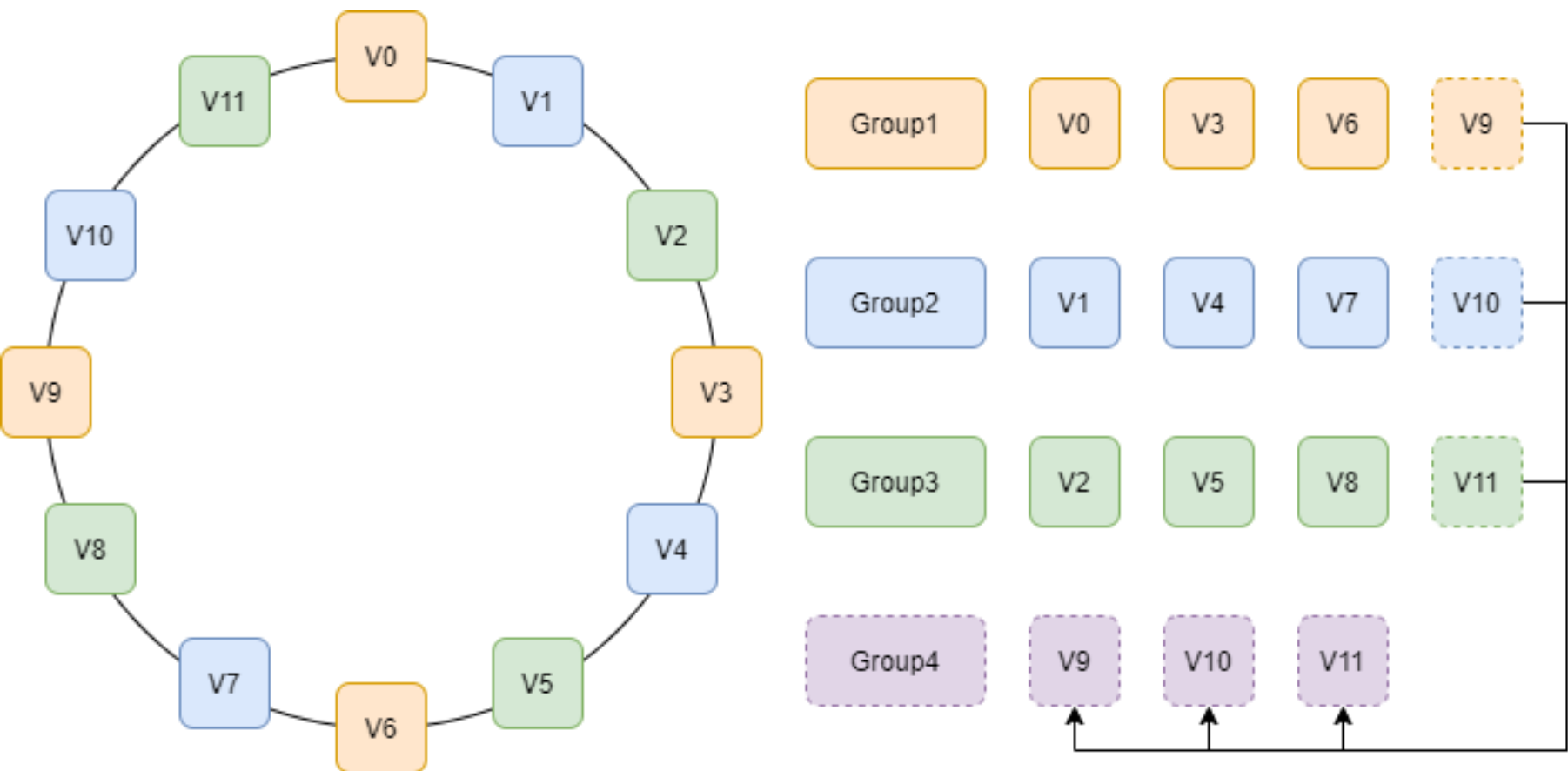
读请求处理流程

ZOOKEEPER结构



PARTITION & SCALABILITY

——改进型一致性哈希

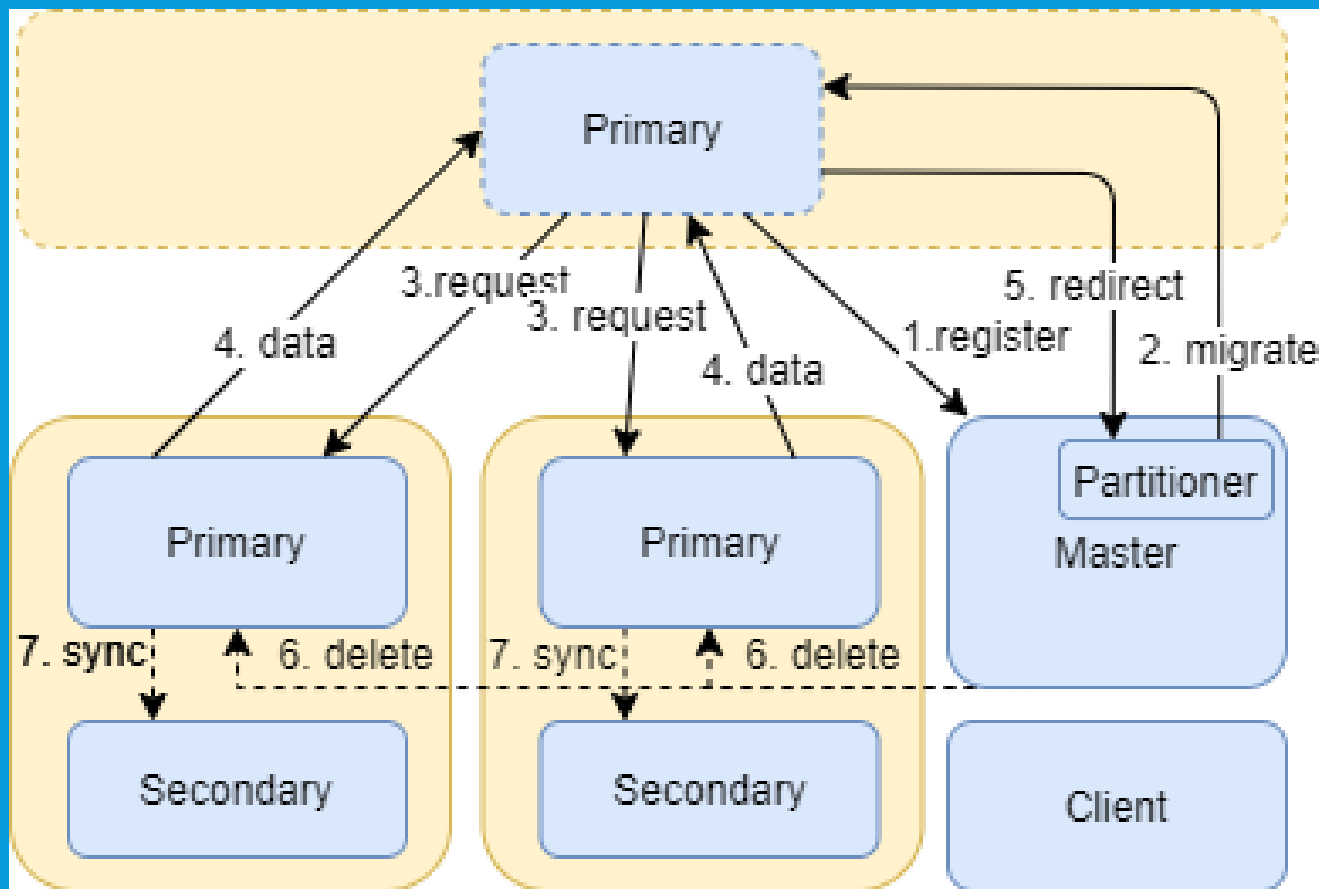


- 数据平衡分布

- 迁移压力分散

PARTITION & SCALABILITY

——启发自原子更新技术的原子迁移



· 原子迁移:

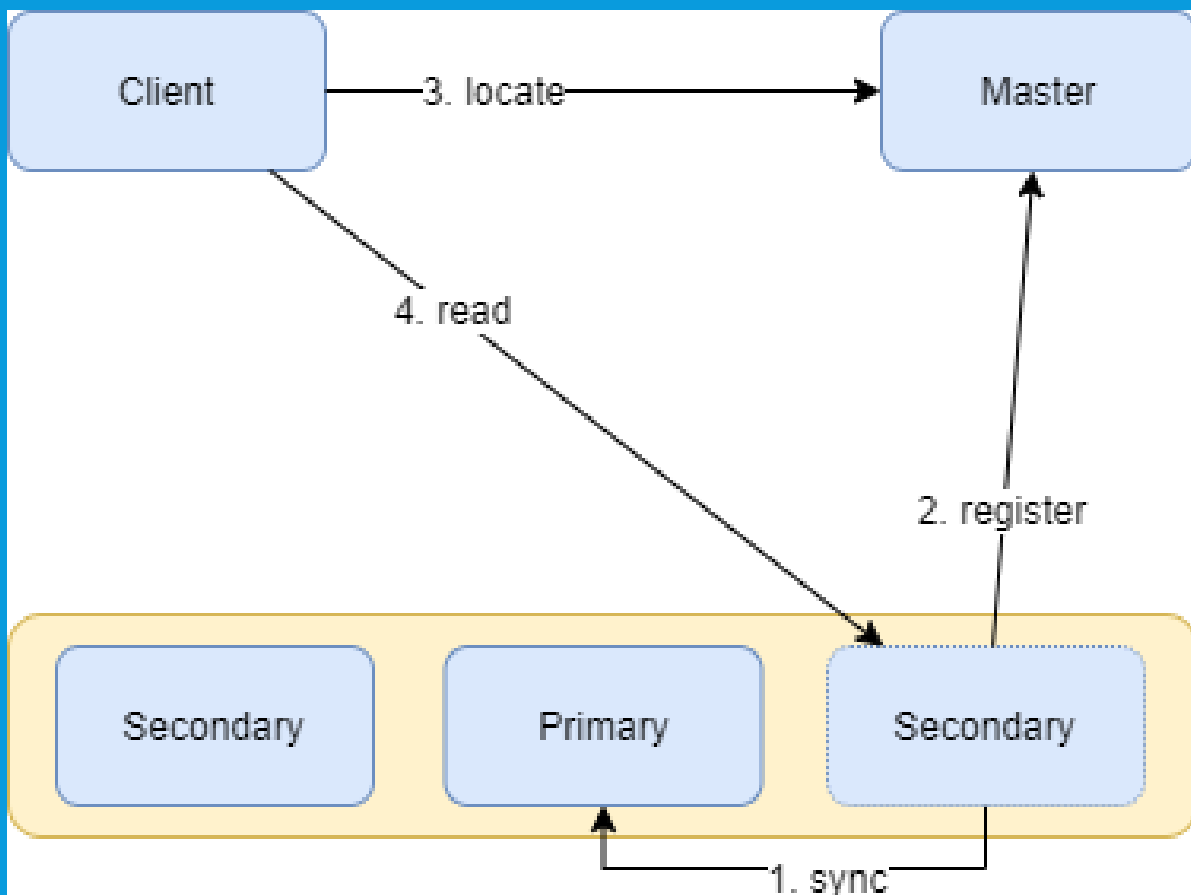
1. 数据迁移
2. 重定向
3. 旧数据清理

· 双写技术:

1. 避免全局分布式锁的引入
2. 不拒绝写请求
3. 提高系统可用性

AVAILABILITY

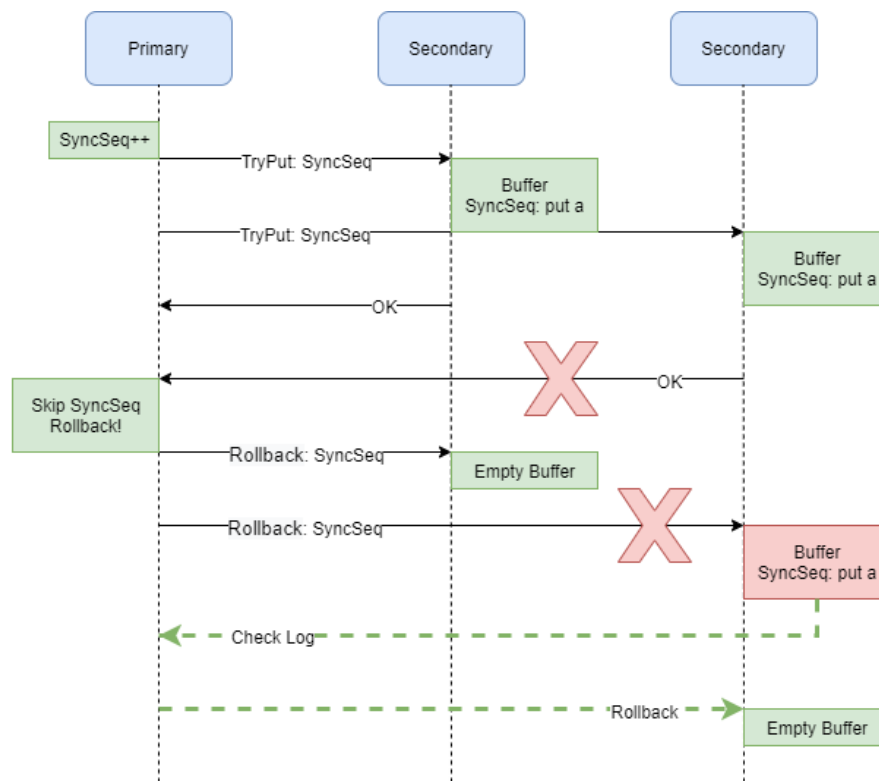
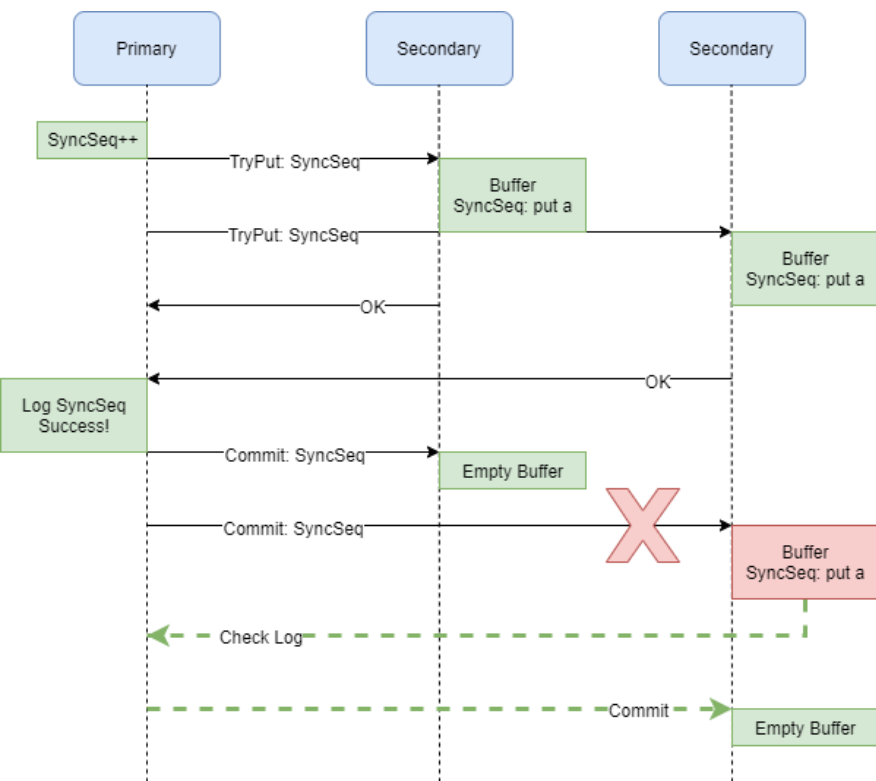
——全量同步



- 备份节点新上线时同步
- Group粒度的分布式锁
- 禁止写，允许读
- Client缓存写操作（可用性）

AVAILABILITY

—增量同步 (2PC+LOG)



- 写时增量同步
- 解决2PC的缺陷:
 1. 同步阻塞
 2. 无限等待/不一致
- 异步（可用性）
- 强一致性由Client保证

Commit异常: check log & Commit

Rollback异常: check log & Rollback

总结

核心功能

1. 数据同步
 - 全量同步（分布式锁，Client缓存）
 - 增量同步（2PC+log）
2. 改进型一致性哈希
3. 原子数据迁移（双写技术）

其他功能

1. Master容灾
 - 无状态，易恢复，Slave闲置等待
2. 并发控制
 - 数据同步：分布式排他锁
 - 日志层面：本地读写锁
 - 请求层面：本地分段锁
3. 负载均衡
 - 数据分布
 - 请求分发
4. 定期持久化备份