GOPS 2016 Shanghai

GOPS

全球运维大会

2016 重新定义运维

上海站

会议时间： 9月23日-9月24日

会议地点： 上海·雅悦新天地大酒店

主办单位： 开放运维联盟 OOPSA Open OPS Alliance　高效运维社区 GreatOPS Community

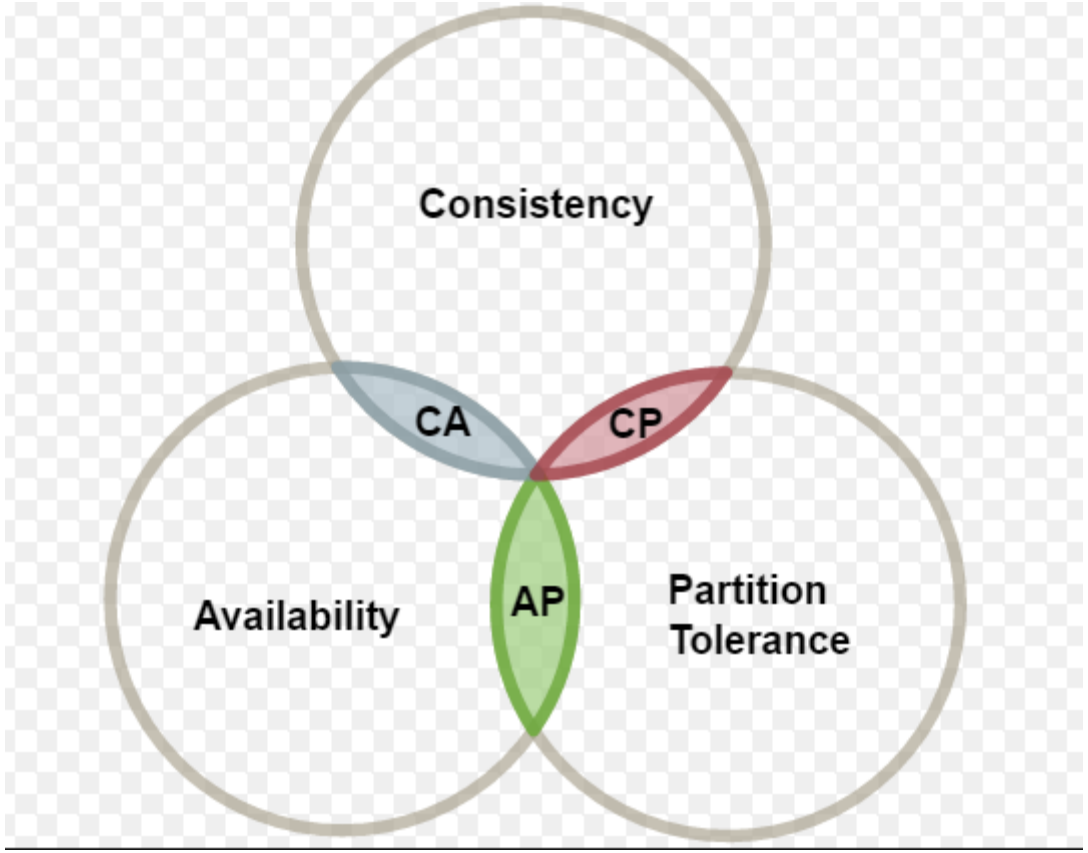
指导单位： 数据中心联盟 Data Center Alliance

分布式共识系统

在高可用架构中的应用

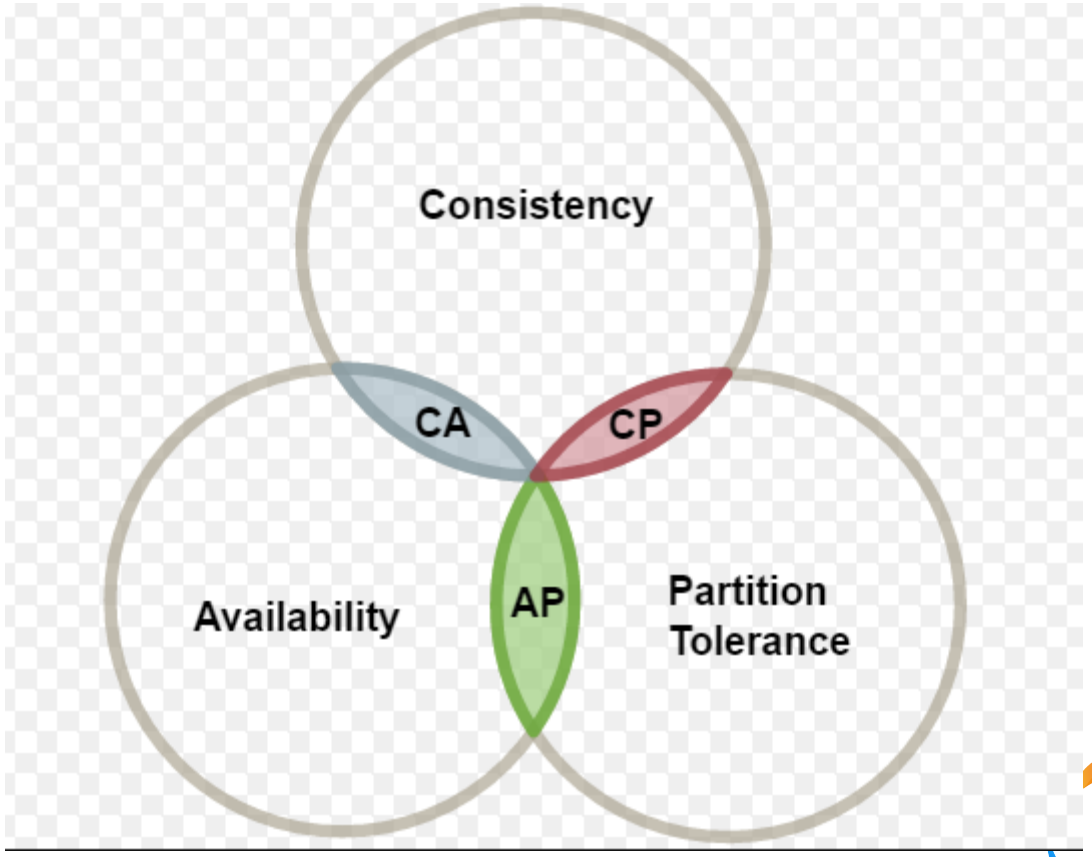
孙宇聪

无人值守的一致的高可用系统是不存在的



CA 系统的实际问题

- 分区问题是一定会发生的
- 分区就会发生脑裂
  - 简单超时心跳
  - 主从灾备切换
  - 小组领头人选举
- 脑裂是无法自动化解决的，人工解决则繁重而困难

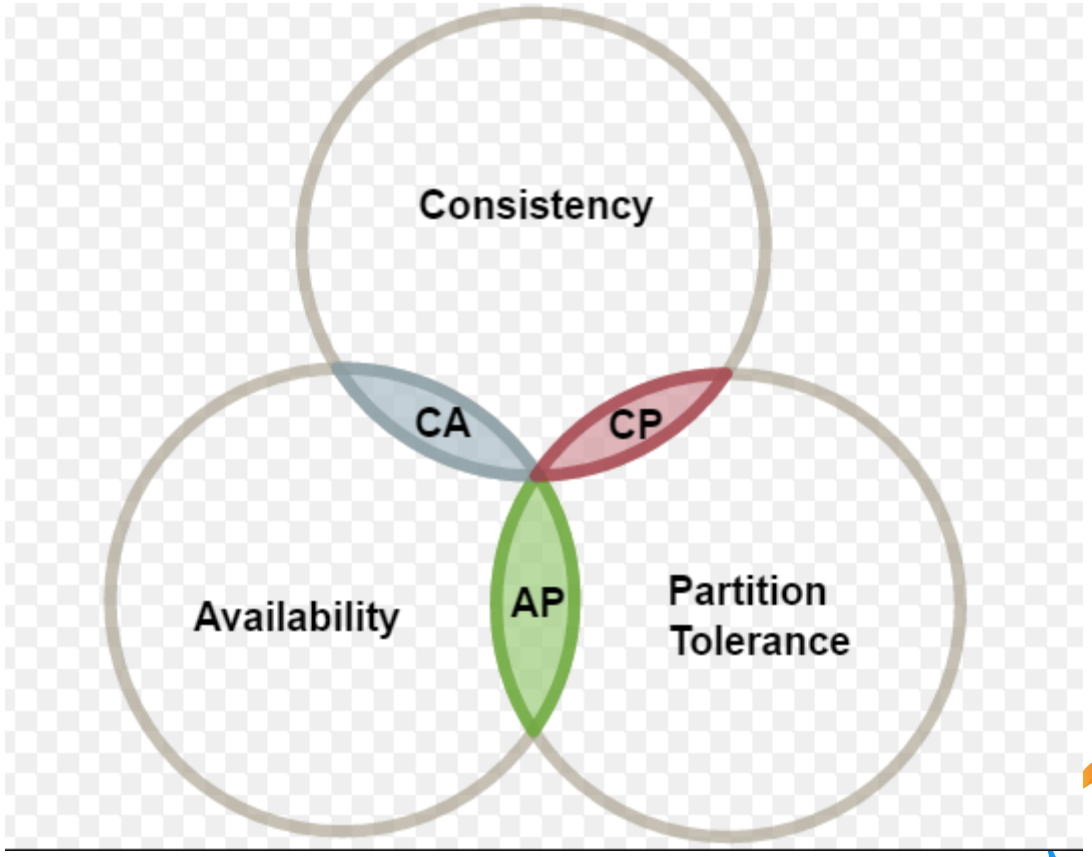


CP 系统 + A 是另外一种选项

- 理论上来说，有限时间内解决分布式共识问题是不可能的

- 优势环境
  - 必要时增加副本
  - 良好的网络链接
  - 避免决斗问题

- 人工干预要求低，风险小



分布式共识系统

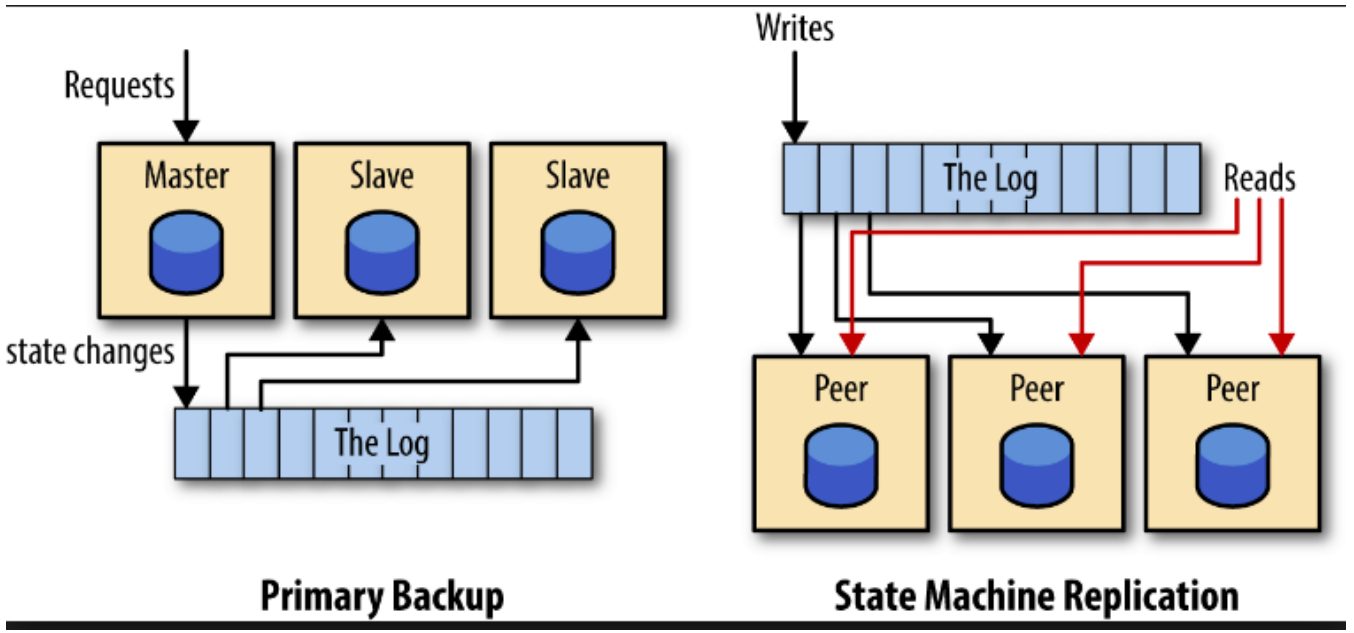
- Paxos, Zab, Raft…
  - Google Chubby, ZooKeeper, etcd, Consul …
- 拜占庭将军问题
  - 不稳定的通信环境下一组进程之间对某项事务（执行/不执行）达成一致的问题
- 稳定状态需要 3N + 1（拜占庭式失败）或 2N + 1（非拜占庭式失败）个实例
  - Quorum Voting
  - Round 1: Prepare / Promised
  - Round 2: Accept / Accepted

为什么要使用？

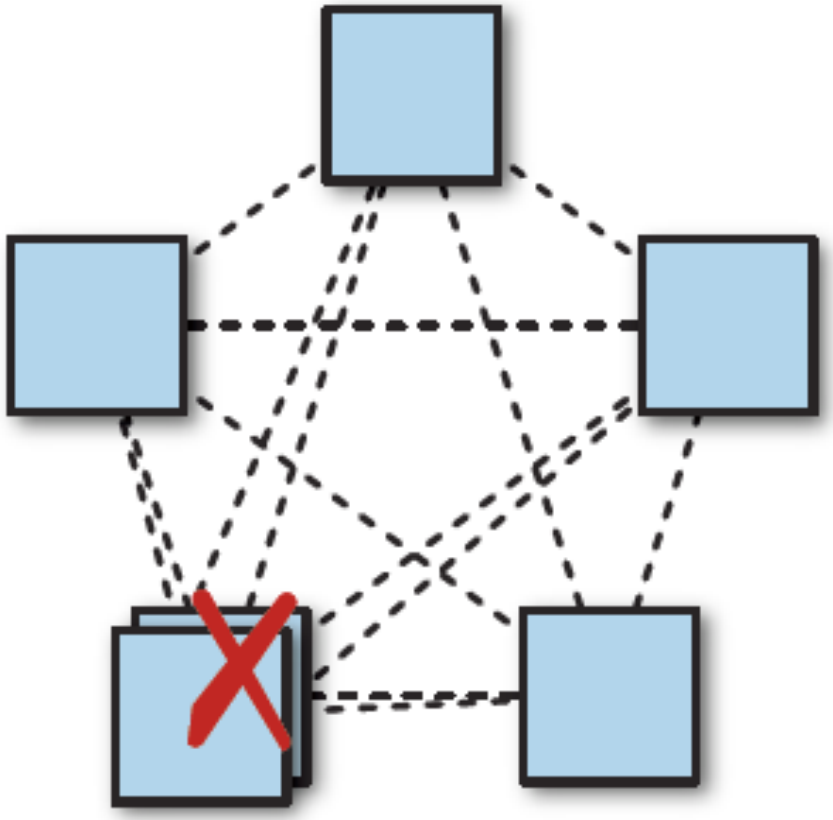
- 可用性传递定律
  - 欲使系统可用度达到 X，其所依赖的系统必须先达到 10 X
  - 链条中最弱的环节决定了系统可用性
- Availability = f(MTBF, MTTR)
  - 分布式共识系统的 MTBF 高, MTTR 低
  - 可以自动处理节点物理故障，容忍一定程度的节点间网络故障
- 微服务架构
  - 无状态的微服务（高可用的）需要存储状态（也需要高可用）
  - 分布式共识系统是实现高可用的共享状态最佳方法

复制状态机（RSM）

- 任何一个具有确定性的程序都可以采用RSM模式成为高可用系统



自动处理集群内节点故障

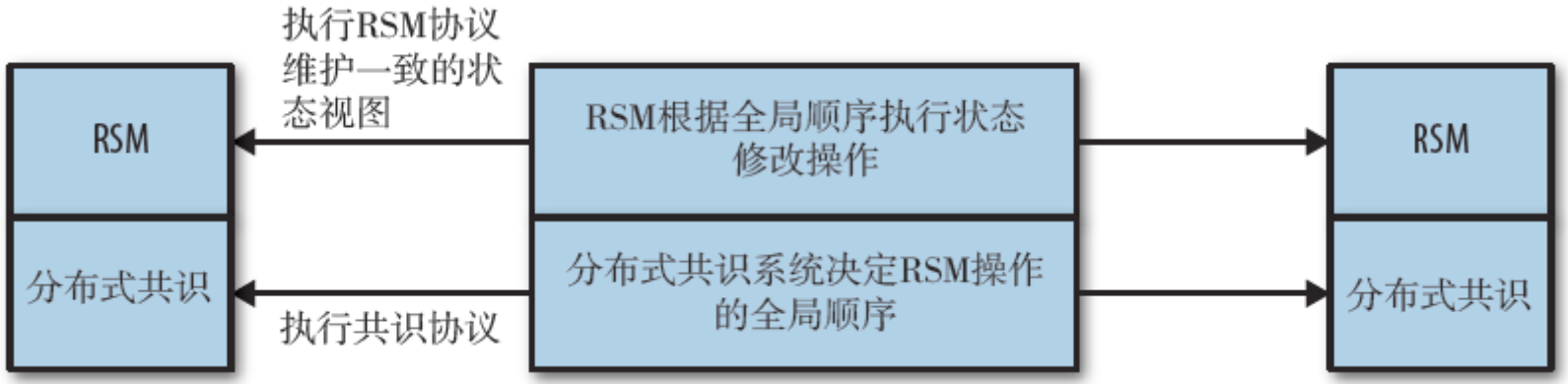


异地共识

应用场景

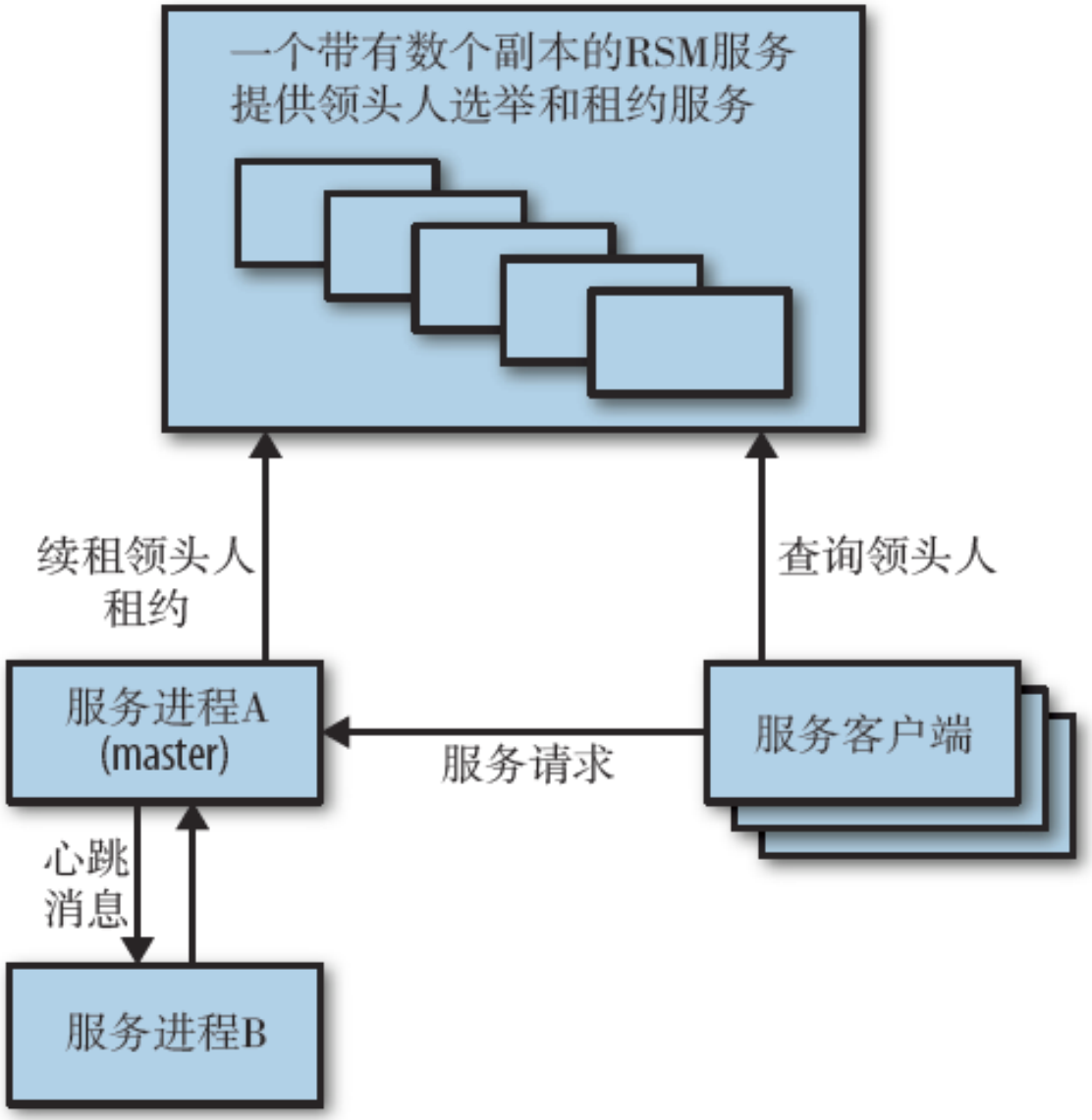
- Consistent View of system state
  - Config / Data Store
  - Queue
  - Lock Service
  - Leader Election

应用：一致性读写数据存储

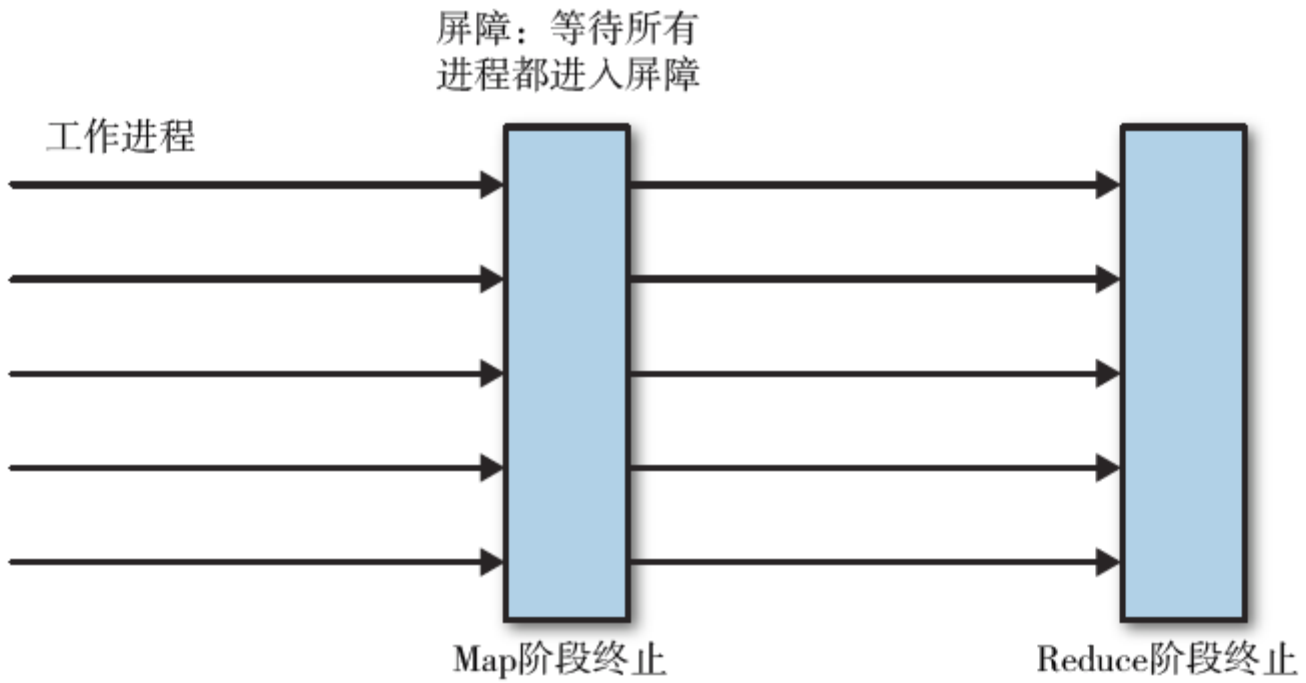


应用： L/F Election

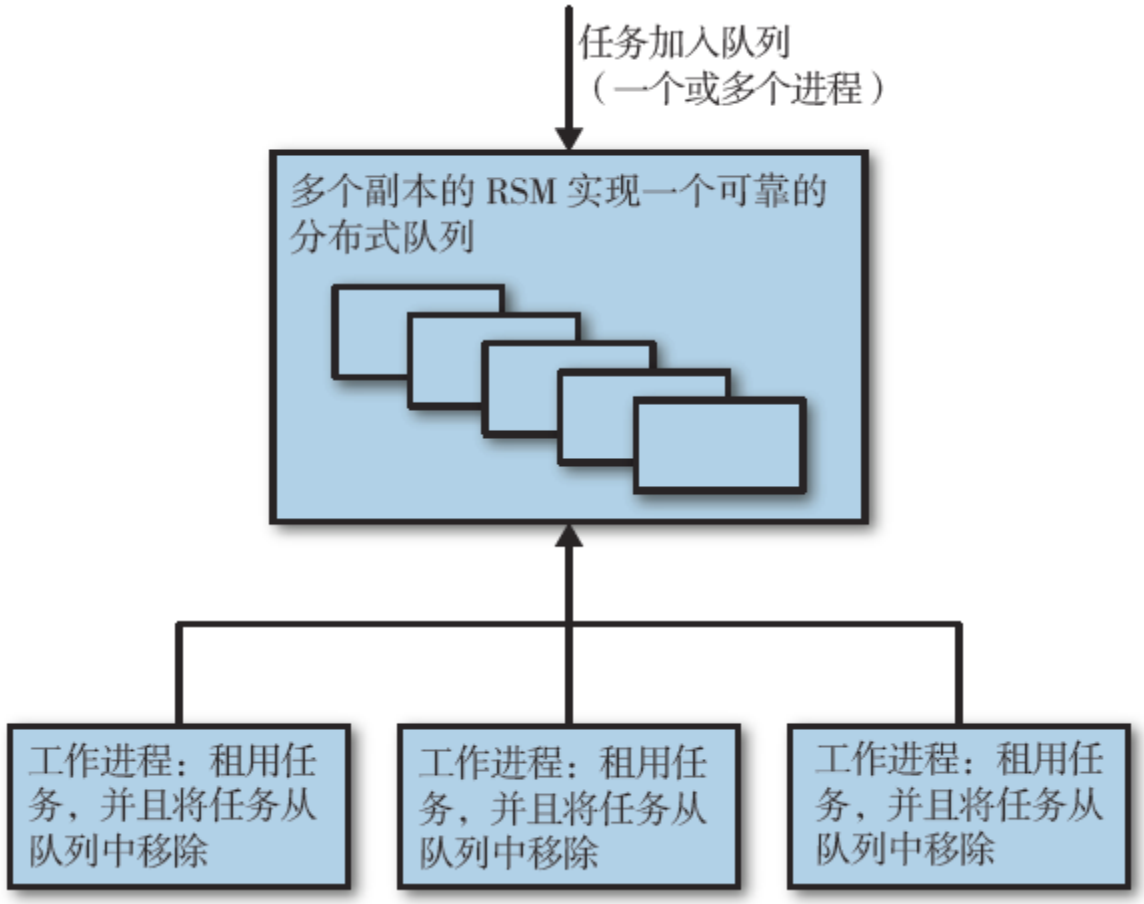
- 简单单机程序 + 复制
- 共识算法不在关键路径中



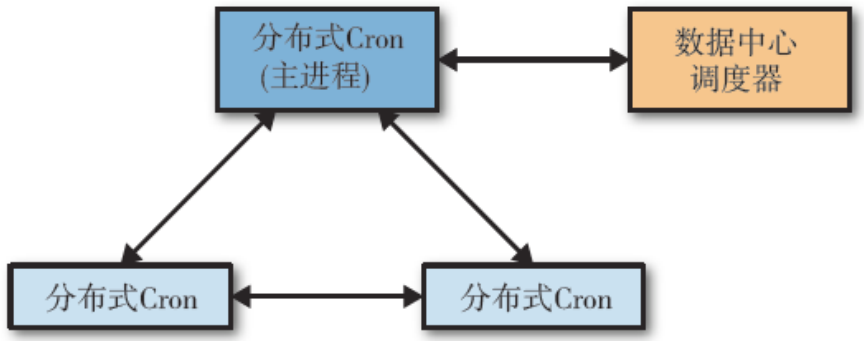
应用：锁 / 屏障

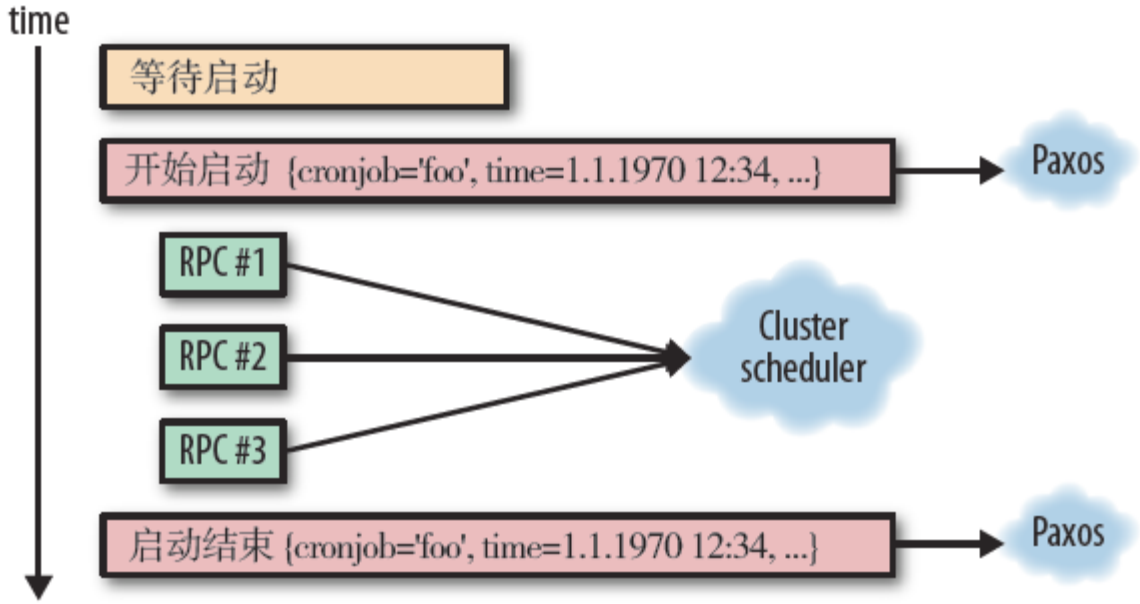


应用： 分布式队列


任务加入队列
（一个或多个进程）
多个副本的 RSM 实现一个可靠的分布式队列
工作进程：租用任务，并且将任务从队列中移除
工作进程：租用任务，并且将任务从队列中移除
工作进程：租用任务，并且将任务从队列中移除

应用：分布式 Cron 系统





回顾

- 关键状态的维护
  - 分布式共识系统是唯一的选择
- 无状态微服务架构
  - 需要分布式共识系统作为支撑