

阿里云 × CLOUD NATIVE
COMPUTING FOUNDATION
云原生技术公开课

第 17 讲

深入理解 etcd: etcd性能优化实践

陈星宇 阿里巴巴技术专家



关注“阿里巴巴云原生”公众号
获取第一手技术资料



本节大纲

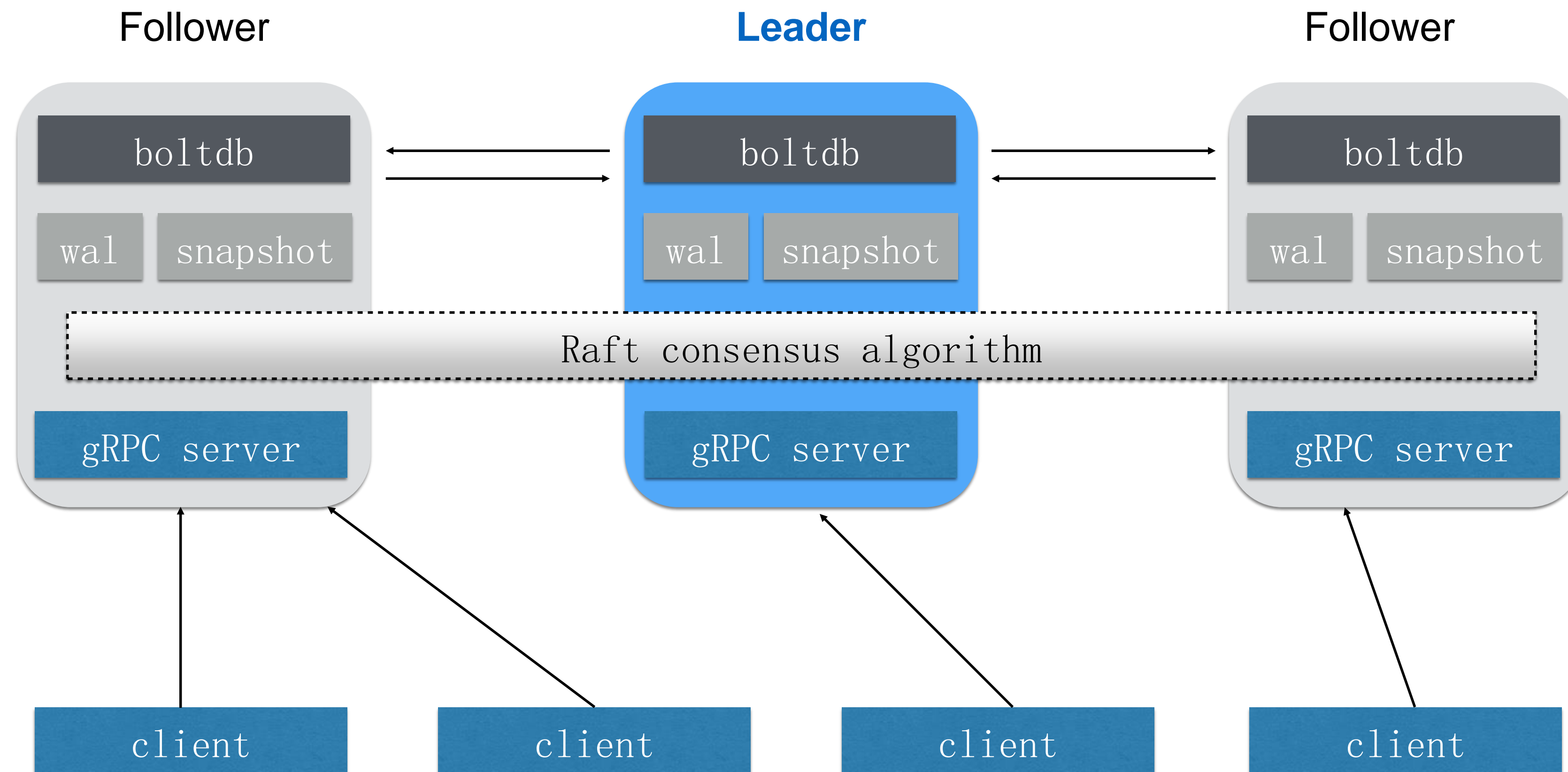
- etcd 前节课程回顾复习
- 理解etcd性能
- etcd 性能优化-server端
- etcd 性能优化-client端
- 本节总结

1 etcd 前节课程回顾复习

etcd 前节回顾复习



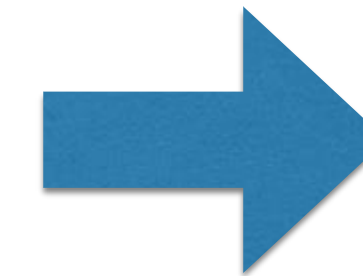
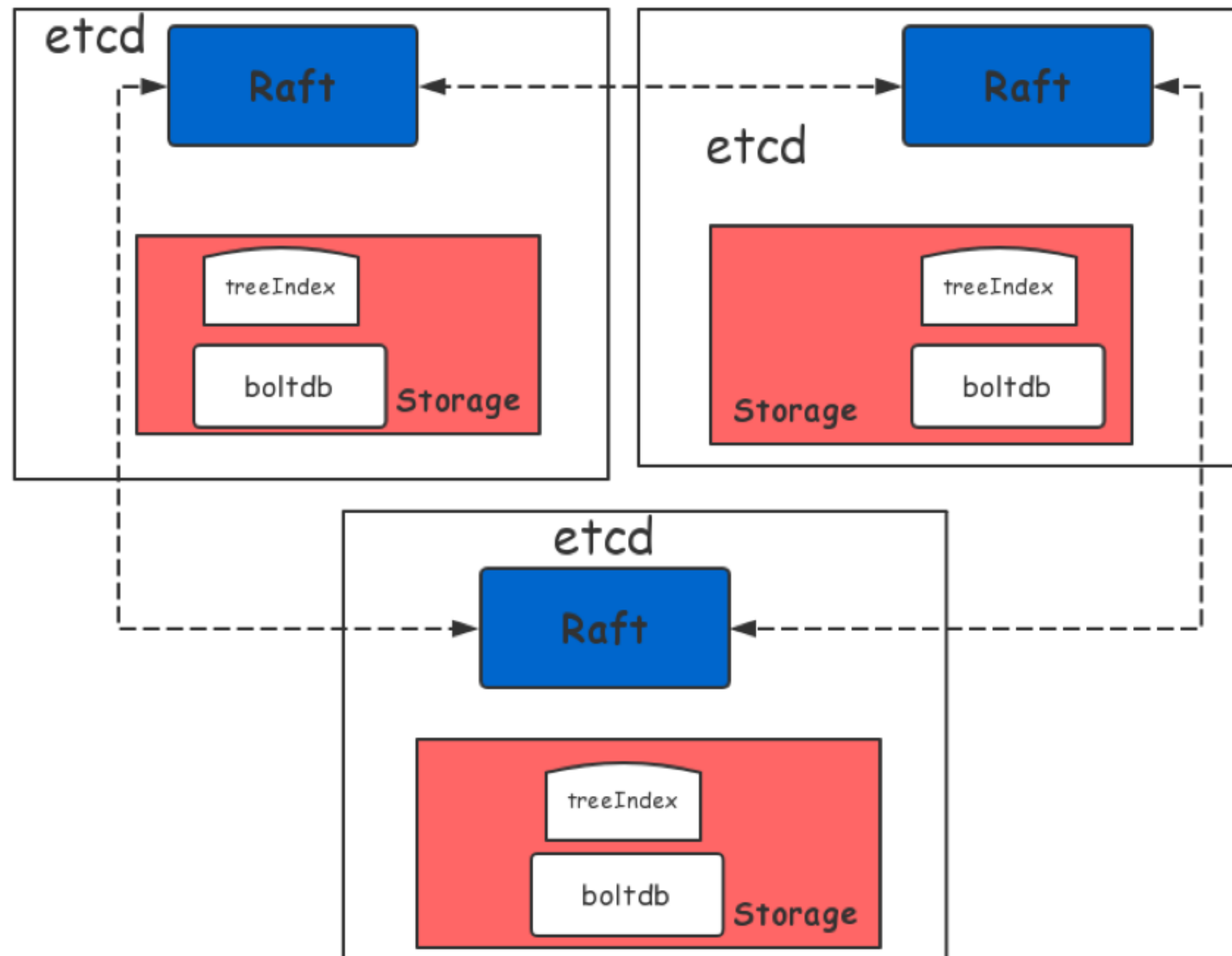
etcd 诞生于 CoreOS 公司，最初用于解决集群管理系统中 OS 升级时的分布式并发控制、配置文件的存储与分发等问题。基于此，etcd 被设计为提供高可用、强一致的小型 KeyValue 数据存储服务。项目当前隶属于 CNCF 基金会，被包括 AWS、Google、Microsoft、Alibaba 等大型互联网公司广泛的使用。



基本架构和内部机制

2 理解etcd性能

理解etcd性能



Raft

网络IO 节点之间的RTT/带宽
WAL 受到磁盘IO 写入延迟

Storage

磁盘IO fdatasync延迟
索引层锁的block
boltdb Tx 的锁
boltdb本身的性能

其他

内核参数
grpc api层延迟

3 etcd性能优化-server端

etcd server性能优化 - 硬件部署

升级CPU
Memory

选取性能优秀的ssd

网络带宽优先级

独占部署，减少其他程序的运行时干扰

详见<https://coreos.com/etcd/docs/latest/op-guide/hardware.html>

etcd server性能优化 - 软件

内存索引层

提升etcd内存索引性能，优化内部锁的使用减少等待时间 <https://github.com/coreos/etcd/pull/9511>

lease 规模使用

优化lease revoke和过期失效的算法，解决了lease规模性的问题 <https://github.com/etcd-io/etcd/pull/9418>

后端boltdb 使用优化

后端batch size limit/interval，可根据不同的硬件和工作负载配置(以前是固定保守值)
<https://github.com/etcd-io/etcd/commit/3faed211e535729a9dc36198a8aab8799099d0f3>

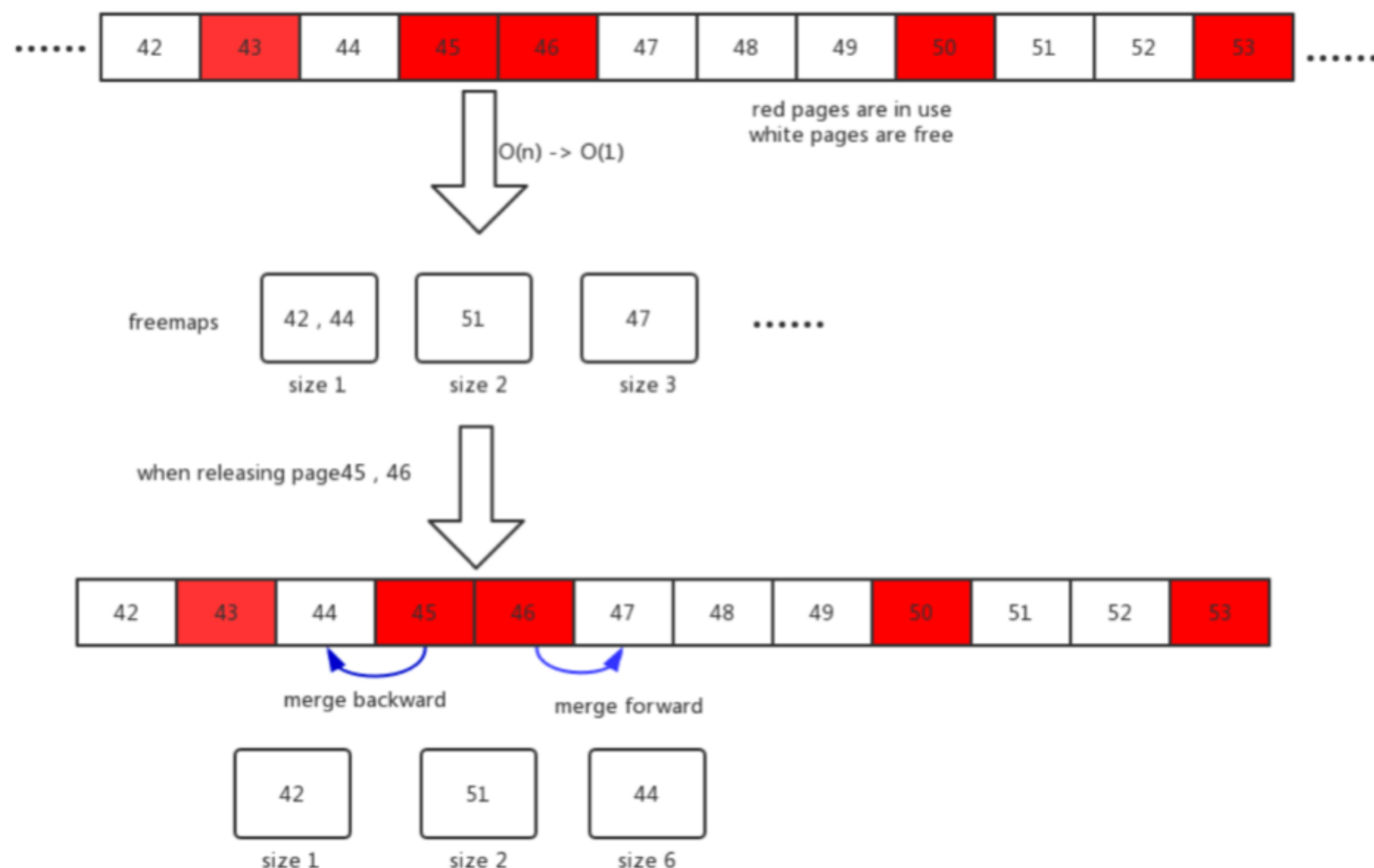
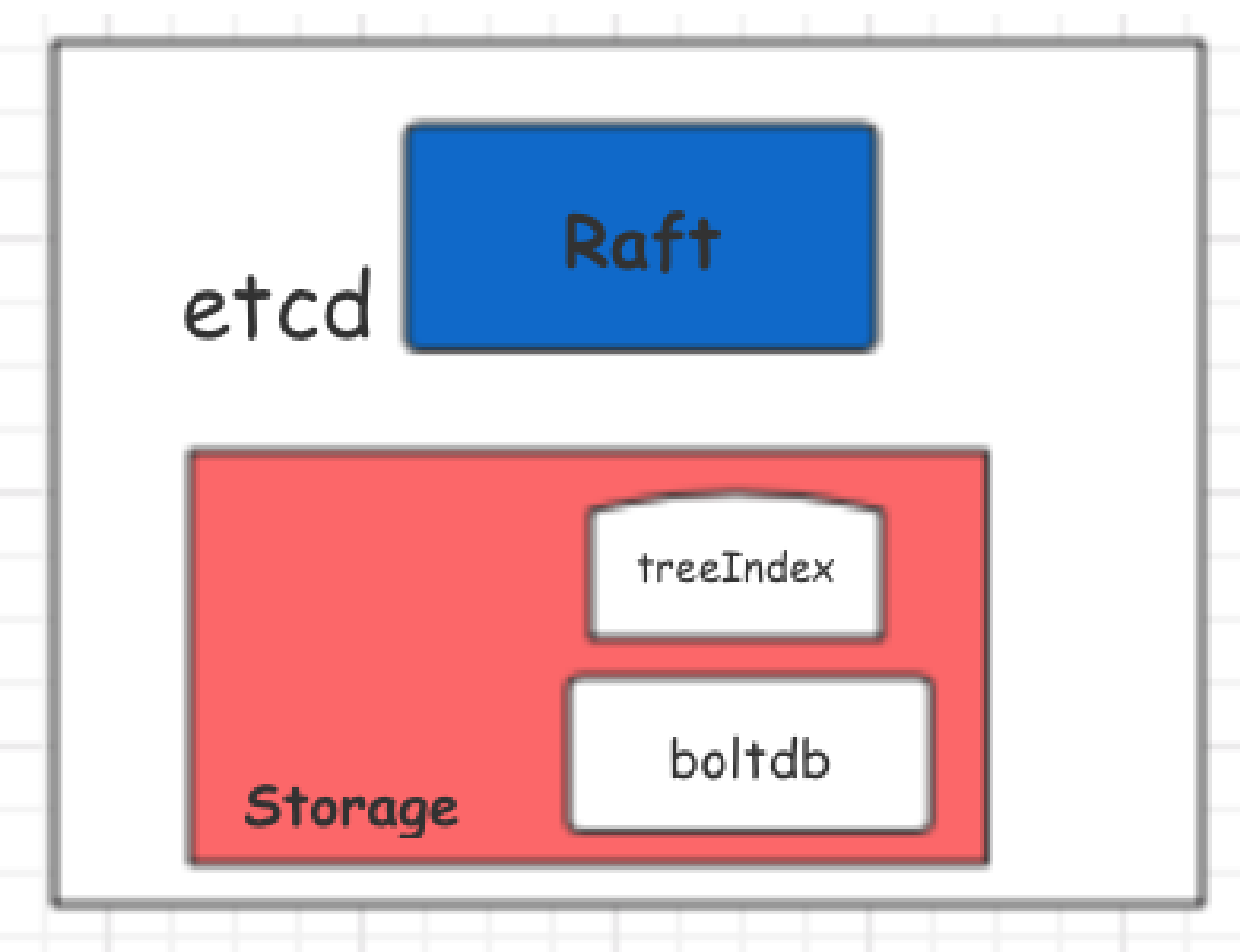
完全并发读 优化调用boltdb tx 读写锁使用，提升读性能 <https://github.com/etcd-io/etcd/pull/10523>

.....

etcd server性能优化 - 软件

基于segregated hashmap的etcd内部存储freelist分配回收算法

cncf文章<https://www.cncf.io/blog/2019/05/09/performance-optimization-of-etcd-in-web-scale-data-scenario/>



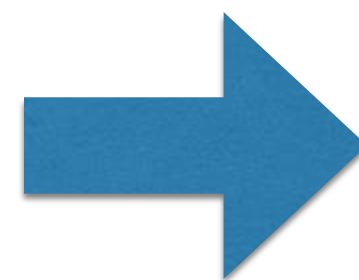
将etcd内部存储空间分配算法时间复杂度从 $O(n)$ 优化到 $O(1)$ ，回收从 $O(n \lg n)$ 优化到 $O(1)$

极大地提高了单个etcd集群的能力 真实场景 存储容量2GB->100GB / 读写性能24x

4 etcd 性能优化 client端

etcd client性能优化实践

- Put(key, value) / Delete(key)
- Get(key) / Get(keyFrom, keyEnd)
- Watch(key / keyPrefix)
- Transactions(if / then / else ops).Commit()
- Leases: Grant / Revoke / KeepAlive



put时避免大value, **精简再精简!**
例如k8s下crd使用

避免创建频繁变化的key/value
例如k8s下 node数据上传

避免创建大量lease, 尽量选择复用
例如k8s下event数据管理

保持客户端使用最佳实践, 将保证你的etcd集群稳定高效运行

本节总结

- 理解etcd性能背景，了解潜在性能瓶颈点
- 解析etcd server端性能优化：硬件/部署/内部核心软件算法等优化
- 了解etcd client 使用最佳实践



关注“阿里巴巴云原生”公众号
获取第一手技术资料