# 如何安装配置一个 ZooKeeper 生产环境

极客时间

扫码试看/订阅

《ZooKeeper实战与源码剖析》视频课程

# 配置项

ZooKeeper 的配置项在 zoo.cfg 配置文件中配置，另外有些配置项可以通过 Java 系统属性来进行配置。

- clientPort ：ZooKeeper 对客户端提供服务的端口。
- dataDir ：来保存快照文件的目录。如果没有设置 dataLogDir ，事务日志文件也会保存到这个目录。
- dataLogDir ：用来保存事务日志文件的目录。因为 ZooKeeper 在提交一个事务之前，需要保证事务日志记录的落盘，所以需要为 dataLogDir 分配一个独占的存储设备。

# ZooKeeper 节点硬件要求

给 ZooKeeper 分配独占的服务器，要给 ZooKeeper 的事务日志分配独立的存储设备。

1. 内存：ZooKeeper 需要在内存中保存 data tree 。对于一般的 ZooKeeper 应用场景，8G 的内存足够了。
2. CPU：ZooKeeper 对 CPU 的消耗不高，只要保证 ZooKeeper 能够有一个独占的 CPU 核即可，所以使用一个双核的 CPU 。
3. 存储：因为存储设备的写延迟会直接影响事务提交的效率，建议为 dataLogDir 分配一个独占的 SSD 盘。

# 日志配置文件

# 配置一个 3 节点的 ZooKeeper 的集群

# 安装配置步骤

1. 申请 ZooKeeper 节点服务器。每个 ZooKeeper 节点有两个挂载盘。
2. 每个节点安装 JDK 8 。
3. 在每个节点为 dataLogDir 初始化一个独立的文件系统 /data ，编辑 myid 。
4. 各个节点之间配置基于 public key 的 SSH 登录。
5. 在一个节点上下载解压 apache-zookeeper-3.5.5-bin.tar.gz ，配置 zoo.cfg 。使用 rsync 把解压的目录同步到其他节点。

# 演示

# 如何进行 ZooKeeper 的监控

# The Four Letter Words

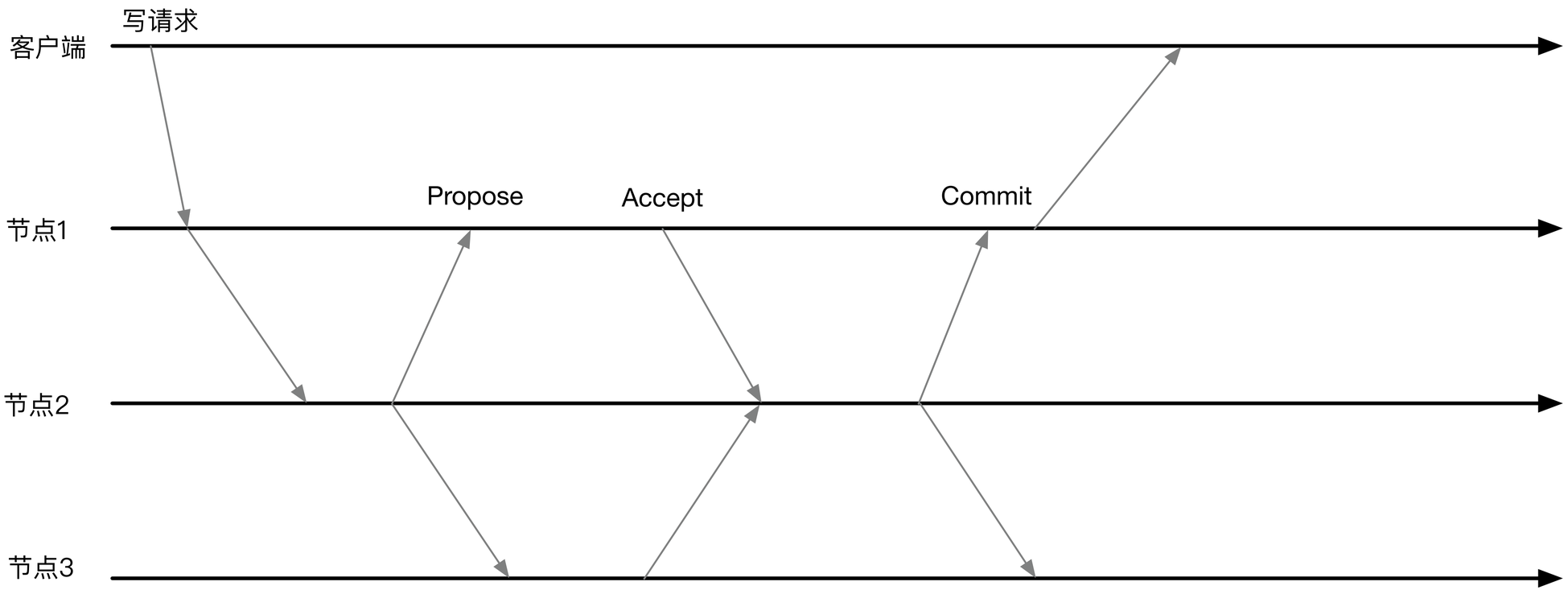
一组检查 ZooKeeper 节点状态的命令。每个命令由四个字母组成，可以通过 telnet 或 ncat 使用客户端端口向 ZooKeeper 发出命令。

# JMX

ZooKeeper 很好的支持了 JMX ，大量的监控和管理工作多可以通过 JMX 来做。

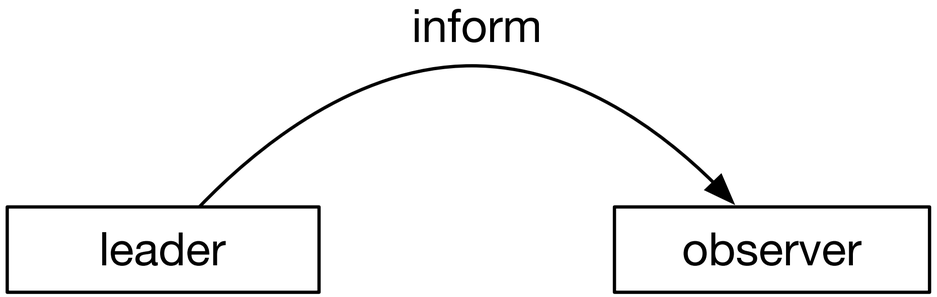
# 通过 ZooKeeper Observer 实现跨区域部署

# ZooKeeper 处理写请求时序



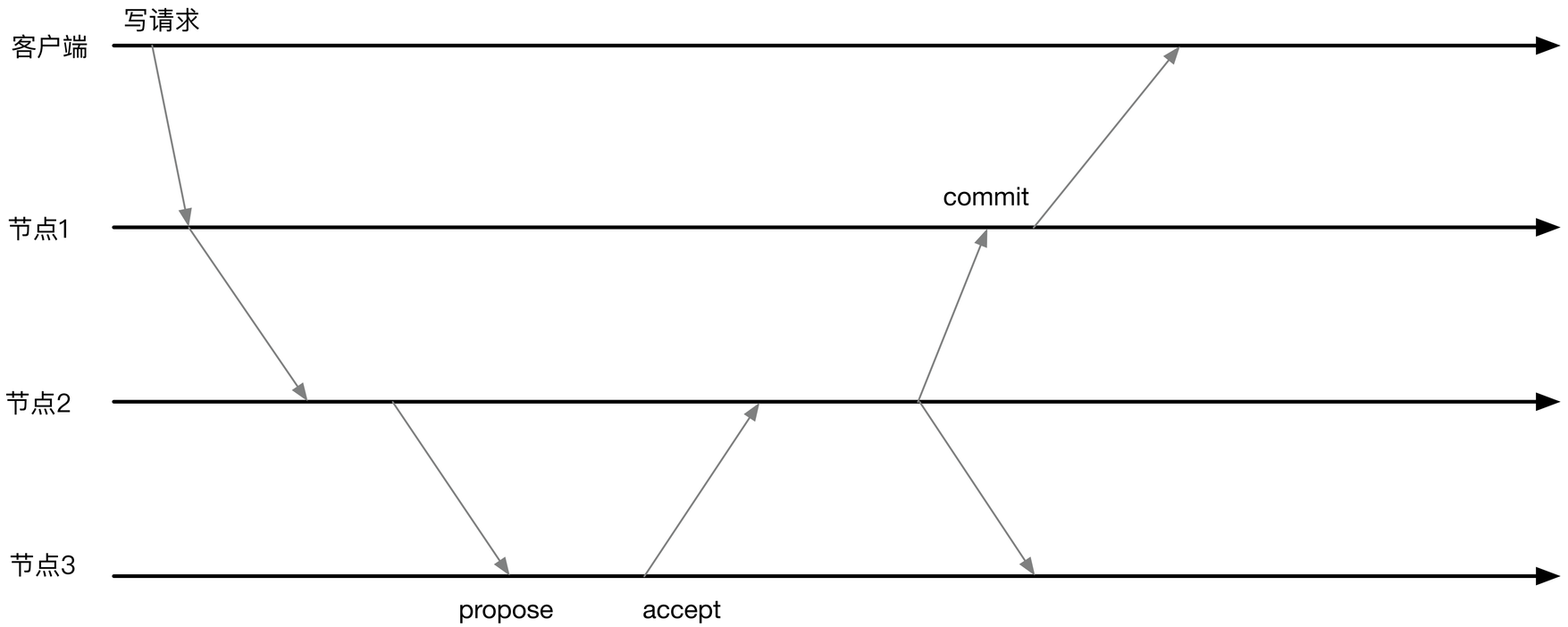
# 什么是 Observer?

Observer 和 ZooKeeper 机器其他节点唯一的交互是接收来自 leader 的 inform 消息，更新自己的本地存储，不参与提交和选举的投票过程。



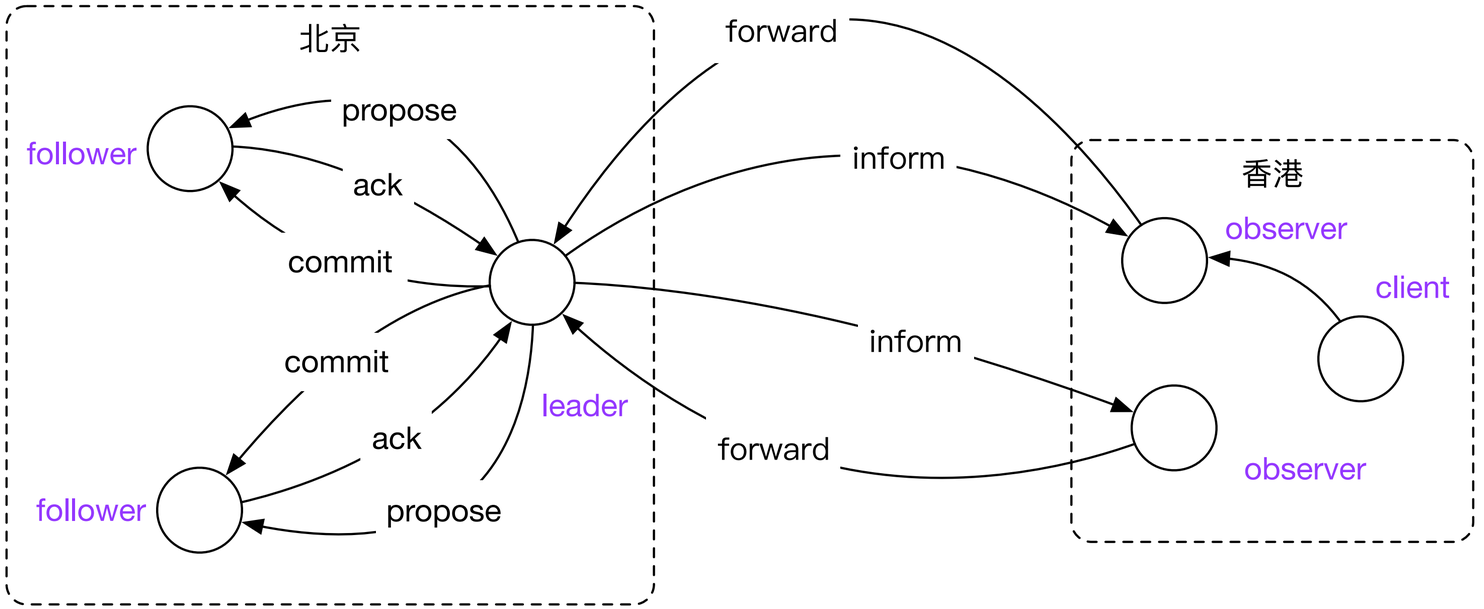
# Observer 应用场景 - 读性能提升

Observer 和 ZooKeeper 机器其他节点唯一的交互是接收来自 leader 的 inform 消息，更新自己的本地存储，不参与提交和选举的投票过程。因此可以通过往集群里面添加 Observer 节点来提高整个集群的读性能。



# Observer 应用场景 - 跨数据中心部署

我们需要部署一个北京和香港两地都可以使用的 ZooKeeper 服务。我们要求北京和香港的客户端的读请求的延迟都低。因此，我们需要在北京和香港都部署 ZooKeeper 节点。我们假设 leader 节点在北京。那么每个写请求要涉及 leader 和每个香港 follower 节点之间的 propose 、ack 和 commit 三个跨区域消息。解决的方案是把香港的节点都设置成 observer 。上面提的 propose 、ack 和 commit 消息三个消息就变成了 inform 一个跨区域消息消息。



# 通过动态配置实现不中断服务的集群成员变更

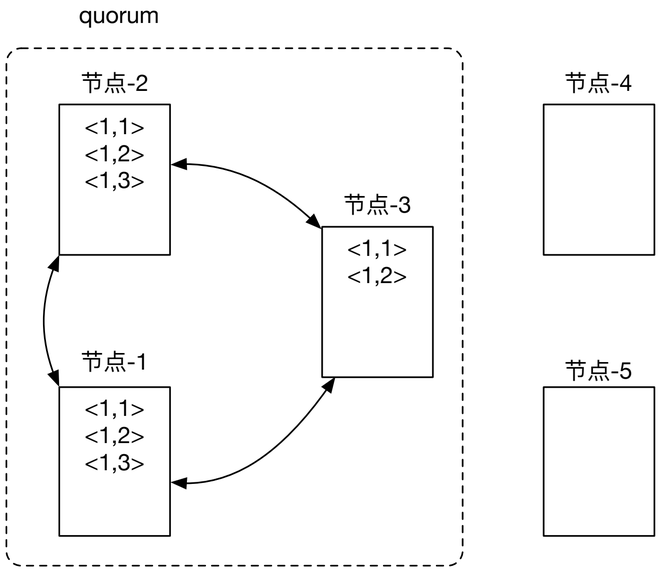
# 手动集群成员调整

1. 停止整个 ZooKeeper 现有集群。
2. 更改配置文件 zoo.cfg 的 server.n 项。
3. 启动新集群的 ZooKeeper 节点。

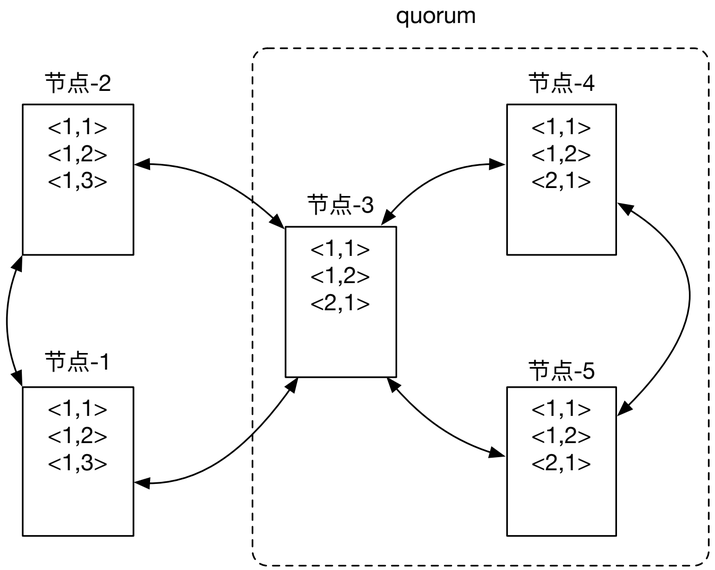
问题1：需要停止 ZooKeeper 服务。

# 问题 2

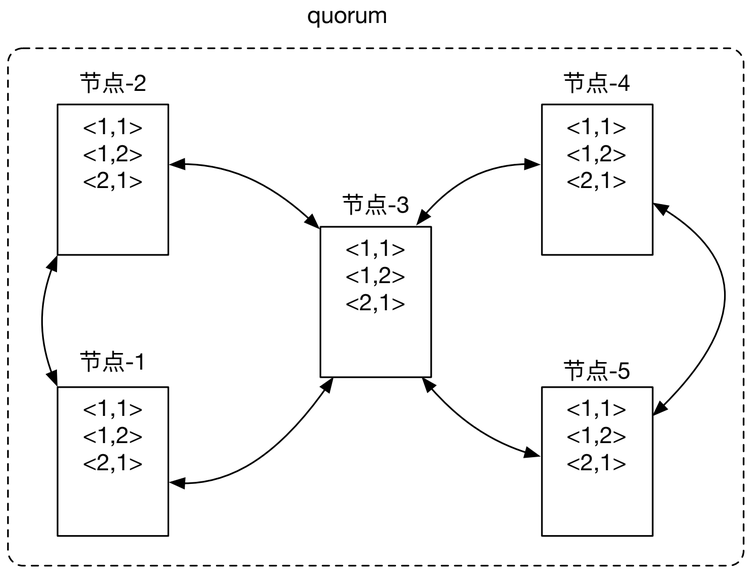
可能会导致已经提交的数据写入被覆盖。




节点-3的数据旧一些




启动节点–4和节点–5，重启节点–1、节点–2和节点–3。
节点–3、节点–4和节点–5先形成一个quorum。



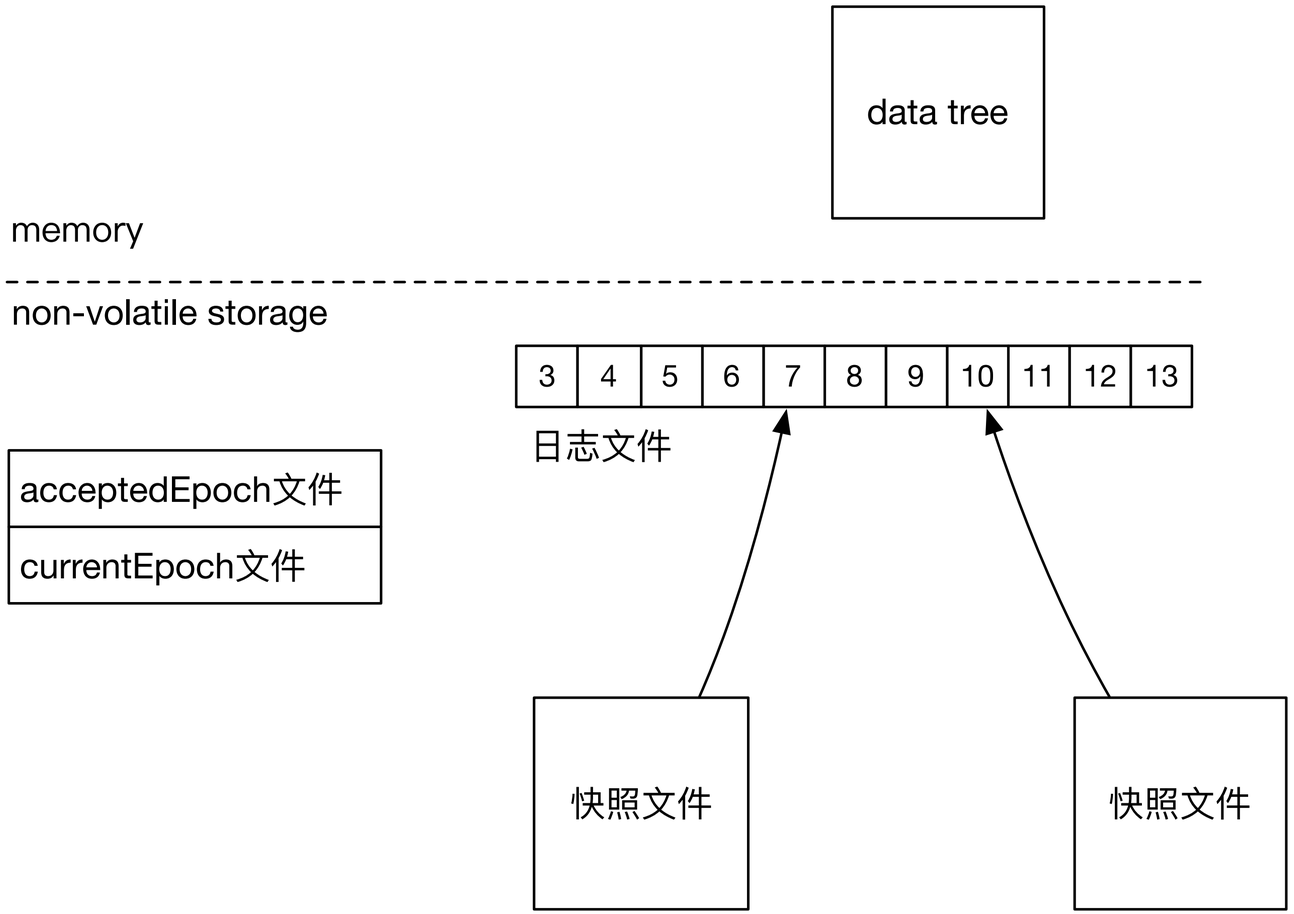

节点–1和节点–2加入quorum之后，<1,3>的事务日志会被覆盖

# 3.5.0 新特性 - dynamic reconfiguration

可以在不停止 ZooKeeper 服务的前提下，调整集群成员。

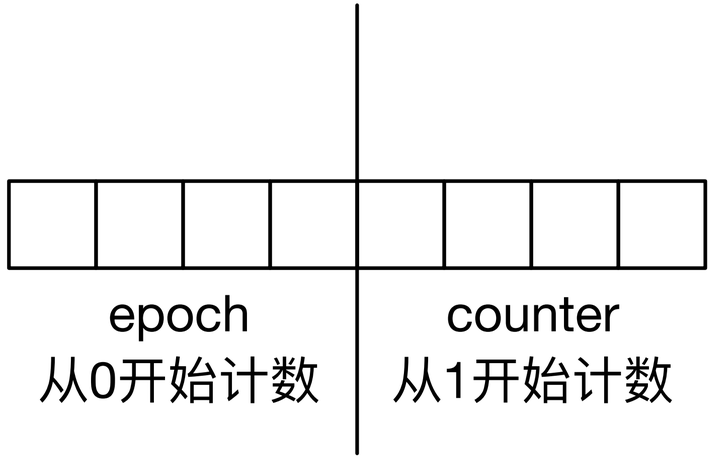
# ZooKeeper 内部数据文件介绍

# ZooKeeper 节点本地存储架构



# zxid

每一个对 ZooKeeper data tree 都会作为一个事务执行。每一个事务都有一个 zxid。zxid 是一个 64 位的整数（Java long 类型）。zxid 有两个组成部分，高 4 个字节保存的是 epoch，低 4 个字节保存的是 counter 。



# 事务日志（Transaction Logs）

# 快照（Snapshots）

# Epoch 文件

极客时间

扫码试看/订阅

《ZooKeeper实战与源码剖析》视频课程