



下载APP



结束语 | 搞懂etcd，掌握通往分布式存储系统之门的钥匙

2021-03-22 唐聪

etcd实战课

[进入课程 >](#)



唐聪

腾讯云资深工程师、etcd 活跃贡献者

你好，我是唐聪。

期待你能通过问题及任务式驱动，搞懂 etcd 背后分布式系统核心技术点。学以致用，打开分布式存储系统之门。

 极客时间



讲述：王超凡

时长 07:44 大小 7.09M



你好，我是唐聪。

时间过得真快，这就到了我们的定期更新的最后一节课了。从筹备、上线到今天专栏完结，过去了将近 7 个多月的时间。

说句实在话，刚开始筹备专栏的时候，我没想过战线会拉得如此之长。当时就是简单地觉得，我的经验也比较丰富了，输出应该很简单。但是其实做专栏耗费的心力远超我的预期：每一节课的构思写作都会花费我大量的时间，而且写完后还得考虑文章逻辑是否有优化的空间，怎样加配图、加一个什么样的配图可以更加形象，甚至部分文章写完自己不注意我还会重写一遍。

细心的你应该能发现，其实这个专栏每一节课的内容都是比较多的。一开始的筹划是每篇文章 3500 字左右，但最后为了讲清楚、讲明白，每一节课大部分都是到了 6000 字到 7000 字的内容（有的文章字数是破万了）。在此特别感谢我的“好基友”王超凡非常用心地和我一块深度 review 每一篇文章，因为平时工作也很忙，还经常得封闭式开发，所以录音只能放在凌晨。

在这里和你分享一件有意思的小事，专栏上线的前一天凌晨，我们和编辑都激动得睡不着，在群里预览文章，聊上线后会是怎么样的一番景象。我们甚至想，会不会上线后被各位疯狂吐槽，以至于不得不录一个“负荆请罪”视频。

1月20日 凌晨03:28

[开篇词的预览链接](https://time.geekbang.org/preview/article/b593d554809bff2c8f3b2bdfd34a8731)



[01 的预览链接](https://time.geekbang.org/preview/article/82a941dc93b80ca9eec8234014bcc096)



1月20日 凌晨04:49

辛苦 这预览效果感觉可以



1月20日 凌晨04:53



当然，你在学习和使用 etcd、Kubernetes 过程中遇到的问题肯定远远不止这些，下面我用思维导图给你总结了更多类似问题，你可以对照自身的经历去看一下。

```
graph TD; A[etcd 常见问题] --> B[etcd 和 Kubernetes 的区别]; A --> C[etcd 和 ZooKeeper 的区别]; A --> D[etcd 和 Raft 的区别]; A --> E[etcd 和 Consul 的区别]; A --> F[etcd 和 Etcd3 的区别]; A --> G[etcd 和 Etcd4 的区别]; A --> H[etcd 和 Etcd5 的区别]; A --> I[etcd 和 Etcd6 的区别]; A --> J[etcd 和 Etcd7 的区别]; A --> K[etcd 和 Etcd8 的区别]; A --> L[etcd 和 Etcd9 的区别]; A --> M[etcd 和 Etcd10 的区别]; A --> N[etcd 和 Etcd11 的区别]; A --> O[etcd 和 Etcd12 的区别]; A --> P[etcd 和 Etcd13 的区别]; A --> Q[etcd 和 Etcd14 的区别]; A --> R[etcd 和 Etcd15 的区别]; A --> S[etcd 和 Etcd16 的区别]; A --> T[etcd 和 Etcd17 的区别]; A --> U[etcd 和 Etcd18 的区别]; A --> V[etcd 和 Etcd19 的区别]; A --> W[etcd 和 Etcd20 的区别]; A --> X[etcd 和 Etcd21 的区别]; A --> Y[etcd 和 Etcd22 的区别]; A --> Z[etcd 和 Etcd23 的区别];
```

这门课就是为了帮助你解决这些问题而生。不过你可能会想，你能把这些东西都讲明白么？我先和你聊聊我的个人 etcd 经历，你就知道我为什么有自信能带你学好 etcd 了。

我和 etcd 的那些事

现在回想起来，真的是做好了被大家吐槽的准备。但你们给我的是超出预期的热情。不少同学从上线到结束，都在时刻关注、学习每一节课，并留下优质的提问以及鼓励、认可。

有的同学是比较资深的 etcd 使用者，会独立分析源码，撰写高质量的技术博客，并给出精彩的回答；

有的同学是刨根问底的 etcd 兴趣用户，会细致思考每一个异常场景，给出精彩的提问；

有的同学刚刚入门 etcd 用户，正因为你们的提问，让我意识到需要在基础篇中多去增加一些特性初体验的案例；

还有的同学着急说面试要用，所以春节期间我们没有筹划春节特别活动，而是正常更新课程正文；

.....

当然在这过程中，我也收获满满。为了解答你们的疑问，我必须得更加深入地阅读 etcd 源码，也是倒逼着我去进一步成长。

编辑半开玩笑地和我说，我们是以百米冲刺的速度去跑马拉松。这段经历真的很难忘，你们的评论和收藏证明了我们的付出是值得的。

在这最后一节课里，我想最后和你再分享下我个人的 etcd 学习经验，以及这整个专栏设计和写作思路。

如果要用一个核心词来总结这个专栏，那我希望是问题及任务式驱动。

从我的个人经验上来看，我每次进一步学习 etcd 的动力，其实都是源于某个棘手的问题。数据不一致、死锁等一系列棘手问题，它们会倒逼我走出舒适区，实现更进一步成长。

从专栏目录中你也可以看到，每讲都是围绕着一个**问题**深入展开。在具体写作思路上，我会先从整体上给你介绍整体架构、解决方案，让你有个全局的认识。随后围绕每个点，按照由浅入深的思路给你分析各种解决方案。

另外，**任务式驱动**也是激励你不断学习的一个非常好的手段，通过任务实践你可以获得满满的成就感，建立正向反馈。你在学习 etcd 专栏的过程中，可结合自己的实际情况，为自

己设立几个进阶任务，下面我给你列举了部分：

从 0 到 1 搭建一个 etcd 集群（可以先从单节点再到多节点，并进行节点的增删操作）；

业务应用中使用 etcd 的核心 API；

自己动手实现一个分布式锁的包；

阅读 etcd 的源码，写篇源码分析博客（可从早期的 etcd v2 开始）；

基于 raftexample 实现一个支持多存储引擎的 KV 服务；

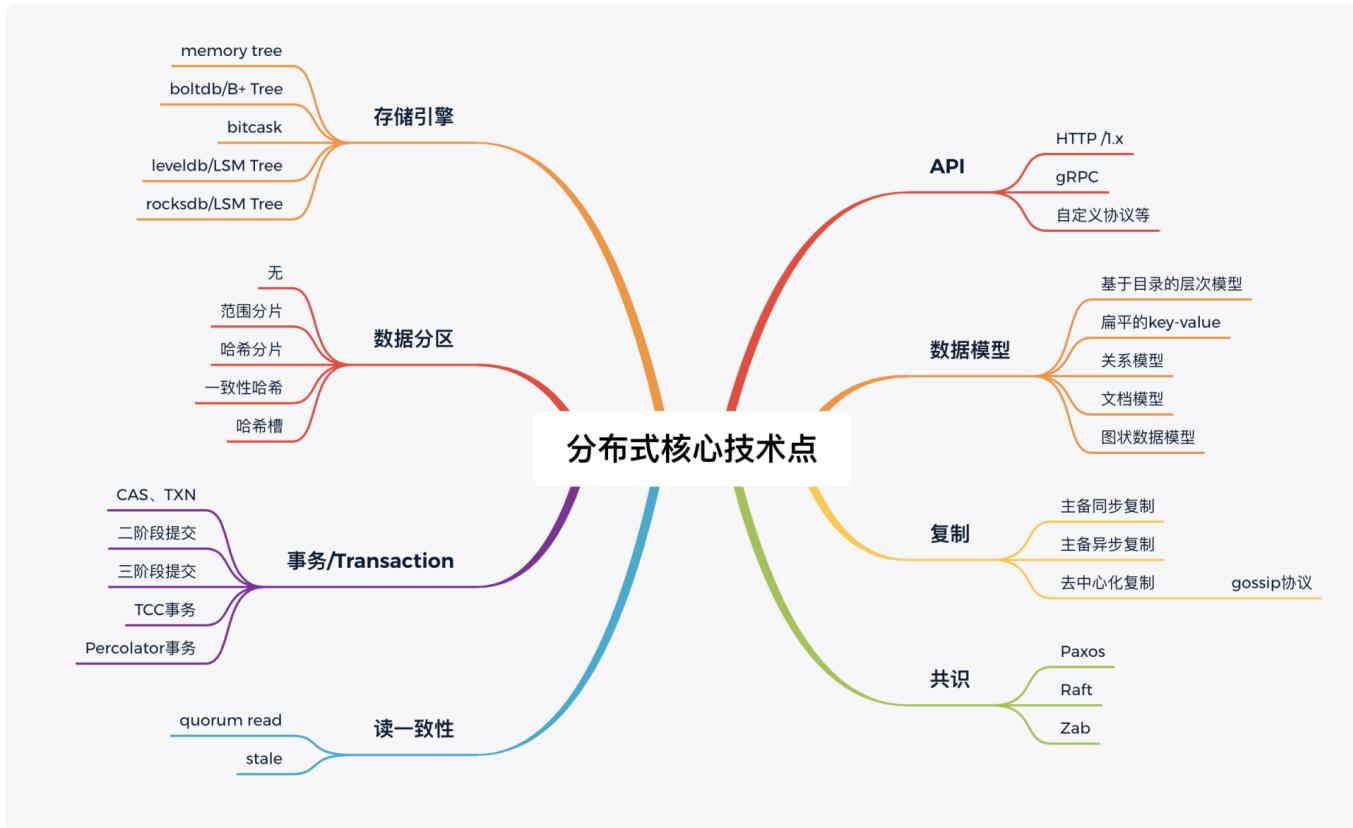
基于 Kubernetes 的 Operator 机制，实现一个 etcd operator，创建一个 CRD 资源就可新建一个集群；

.....

我希望带给你的不仅仅是 etcd 原理与实践案例，更希望你收获的是一系列分布式核心问题解决方案，它们不仅能帮助你搞懂 etcd 背后的设计思想与实现，更像是一把通往分布式存储系统之门的钥匙，让你更轻松地学习、理解其他存储系统。

那你可能会问了，为什么搞懂 etcd 就能更深入理解分布式存储系统呢？

因为 etcd 相比其他分布式系统如 Hbase 等，它足够简洁、轻量级，又涵盖了分布式系统常见的问题和核心概念，如 API、数据模型、共识算法、存储引擎、事务、快照、WAL 等，非常适合新人去学习。



上图我为你总结了 etcd 以及其他分布式系统的核心技术点，下面我再和你简要分析一下几个分布式核心问题及解决方案，并以 Redis Cluster 集群模式作为对比案例，希望能够帮助你触类旁通。

首先是服务可用性问题。分布式存储系统的解决方案是共识算法、复制模型。etcd 使用的是 Raft 共识算法，一个写请求至少要一半以上节点确认才能成功，可容忍少数节点故障，具备高可用、强一致的目标。Redis Cluster 则使用的是主备异步复制和 Gossip 协议，基于主备异步复制协议，可将数据同步到多个节点，实现高可用。同时，通过 Gossip 协议发现集群中的其他节点、传递集群分片等元数据信息等操作，不依赖于元数据存储组件，实现了去中心化，降低了集群运维的复杂度。

然后是数据如何存取的问题。分布式存储系统的解决方案是存储引擎。除了 etcd 使用的 boltdb，常见的存储引擎还有我们实践篇 18 中所介绍 bitcask、leveldb、rocksdb (leveldb 优化版) 等。不同的分布式存储系统在面对不同业务痛点时（读写频率、是否支持事务等），所选择的解决方案不一样。etcd v2 使用的是内存 tree，etcd v3 则使用的是 boltdb，而 Redis Cluster 则使用的是基于内存实现的各类数据结构。

最后是如何存储大量数据的问题。分布式存储系统的解决方案是一系列的分片算法。etcd 定位是个小型的分布式协调服务、元数据存储服务，因此 etcd v2 和 etcd v3 都不支持分

片，每个节点含有全量的 key-value 数据。而 Redis Cluster 定位是个分布式、大容量的分布式缓存解决方案，因此它必须要使用分片机制，将数据打散在各个节点上。目前 Redis Cluster 使用的分片算法是哈希槽，它根据你请求的 key，基于 crc16 哈希算法计算 slot 值，每个 slot 分配给对应的 node 节点处理。

 复制代码

```
1 HASH_SLOT = CRC16(key) mod 16384
```

etcd 作为最热门的云原生存储之一，在腾讯、阿里、Google、AWS、美团、字节跳动、拼多多、Shopee、明源云等公司都有大量的应用，覆盖的业务不仅仅是 Kubernetes 相关的各类容器产品，更有视频、推荐、安全、游戏、存储、集群调度等核心业务。

更快、更稳是 etcd 未来继续追求的方向，etcd 社区将紧密围绕 Kubernetes 社区做一系列的优化工作，提供集群降级、自动将 Non-Voting 的 Learner 节点提升为 Voting Member 等特性，彻底解决饱受开发者诟病的版本管理等问题。

希望这个专栏一方面能帮助你遵循最佳实践，高效解决核心业务中各类痛点问题，另一方面能轻松帮你搞定面试过程中常见 etcd 问题，拿到满意的 offer。

当然，我发现很多同学只是默默地收藏，一直在“潜水”。我希望在这最后一课里，大家一块来“灌灌水”，分享一下你自己的 etcd 学习方法以及你对这门课的感受。我为你准备了一份 [问卷](#)，希望你花两分钟填一下，说不定你就是我们这门课的“小锦鲤”~

最后，再次感谢，我们留言区和加餐见！

**唐聪**

腾讯云资深工程师、etcd 活跃贡献者

感谢一起走过的这段时间，非常想听听你对我和这门课程的反馈与建议。在 2021 年 4 月 6 日前提交，将有机会获得



极客原创 | 2021 赋能周历本
价值 ￥99

或



极客时间课程阅码
价值 ￥99

填写问卷 

[提建议](#)

更多课程推荐

Redis 核心技术与实战

从原理到实战, 彻底吃透 Redis

蒋德钧

中科院计算所副研究员



涨价倒计时

现仅半价 **¥89** 4月17日涨价至**¥199**

© 版权归极客邦科技所有, 未经许可不得传播售卖。 页面已增加防盗追踪, 如有侵权极客邦将依法追究其法律责任。

上一篇 [结课测试题 | 这些相关etcd知识你都掌握了吗?](#)

精选留言 (7)

[写留言](#)

不瘦二十斤
不改头像
jeffery
2021-03-22

真不舍! 太清爽了! 希望加餐! 😊 要求是不是太过分了 😊 多谢唐老师三个月份的付出分享! 确实牛

作者回复: 感谢jeffery同学认可, 短暂休息下, 近期先推出答疑文章, 随后再来加餐, 查漏补缺, 兼顾各个层次的同学, 帮助大家更好学习专栏内容



7



Geek_daf51a

2021-03-22

辛苦老师, 从专栏中学到了很多, 有etcd历史、有原理、有重点问题案例、有定位延时、

内存等问题方法以及最佳实践、还有有kubernetes、配置服务发现、分布式锁等剖析，过程中大量配图总结，覆盖了etcd方方面面，能感受到老师的用心与认真，值了，期待后面的加餐

展开 ▼

作者回复: 感谢支持!



mmm

2021-03-22

感谢老师 期待加餐

展开 ▼



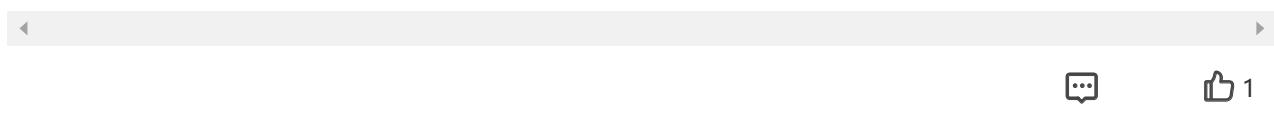
liudu_ec

2021-03-24

终于看完了，意犹未尽

展开 ▼

作者回复: 赞，能坚持这么快看完，了不起



Coder

2021-03-22

辛苦唐老师，收获满满！

展开 ▼



Geek_83edde

2021-03-31

潜水完，冒个泡，反反复复看了两三遍

展开 ▼



云原生工程师

2021-03-23

感谢老师，期待加餐

展开 

