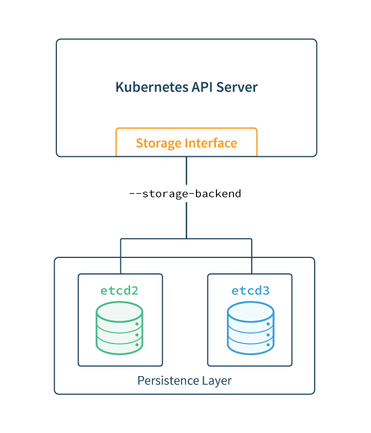
# 概述

Zookeeper 和 Etcd 都是非常优秀的分布式协调系统，zookeeper 起源于 Hadoop 生态系统，起步比较早，而 etcd 算是后起之秀，它的流行是因为它是 kubernetes 的后台支撑。

Etcd是用go开发的，出现的时间并不长，不像zookeeper那么悠久和有名，但是前景非常好。

Etcd是因为kubernetes而被人熟知的，kubernetes的kube master使用etcd作为分布式存储获取分布式锁，这为etcd的强大做了背书。



**Etcd使用RAFT算法实现的一致性，比zookeeper的ZAB算法更简单。**

Etcd没有使用zookeeper的树形结构，而是提供了一个分布式的key-value存储。

# 特点

原子性

一致性

顺序一致性

可串行化级别

高可用

可线性化

## 优点

支持增量快照，避免了zookeeper的快照暂停问题

堆外存储，没有垃圾回收暂停问题

无需像zookeeper那样为每个watch都做个socket连接，可以复用

zookeeper每个watch只能收到一次事件通知，etcd可以持续监控，在一次watch触发之后无需再次设置一次watch

zookeeper会丢弃事件，etcd3持有一个事件窗口，在client断开连接后不会丢失所有事件

## 缺点

如果超时，或者client与etcd网络中断，client不会明确的知道当前操作的状态。

在leader选举时，etcd会放弃操作，并且不会给client发送放弃响应。

在网络分区时，当leader处于小分区时，读请求会继续被处理。

# 原理

# API

etcd3 提供了如下操作接口：

put - 添加一个新的 key-value 到存储中

get - 获取一个 key 的 value

range - 获取一个范围的 key 的 value，例如：key1 - key10

transaction - 读、对比、修改、写的组合

watch - 监控一个或一个范围的 key，发生变化后就会得到通知

# 对比

zookeeper是用java开发的，被Apache很多项目采用。

etcd是用go开发的，主要是被Kubernetes采用。

zookeeper非常稳定，是一个著名的分布式协调系统，etcd是后起之秀，前景广阔。

因为etcd是用go写的，现在还没有很好的java客户端库，需要通过http方式调用。

而zookeeper在这方面就成熟很多，对于java之外的其他开发语言都有很好的客户端库。

具体选择zookeeper还是etcd，需要根据您的需求结合它们各自的特性进行判断，还有您所使用的开发语言。

# 应用

在腾讯TDSQL的元数据管理MC模块中，底层的存储使用的就是Etcd。