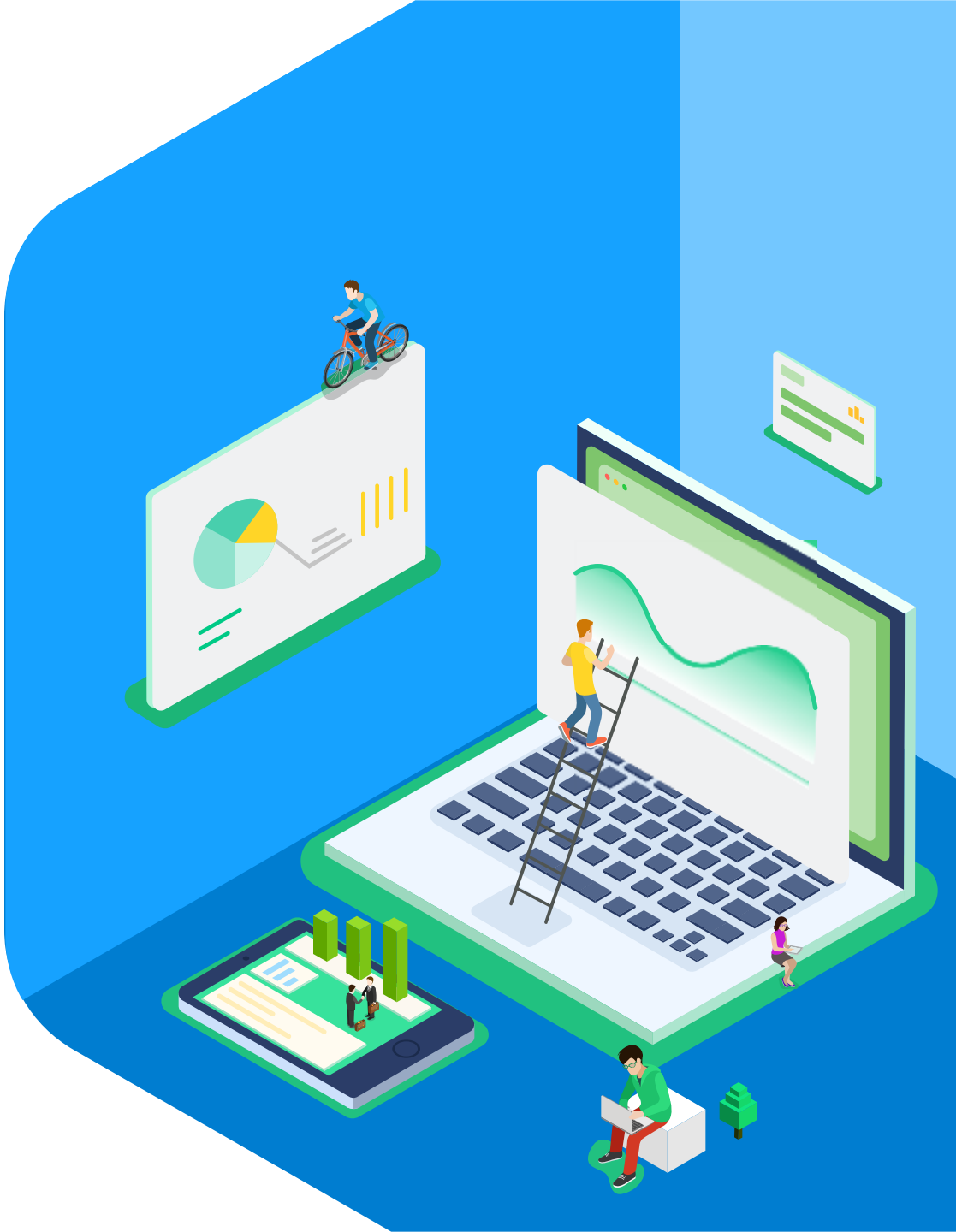
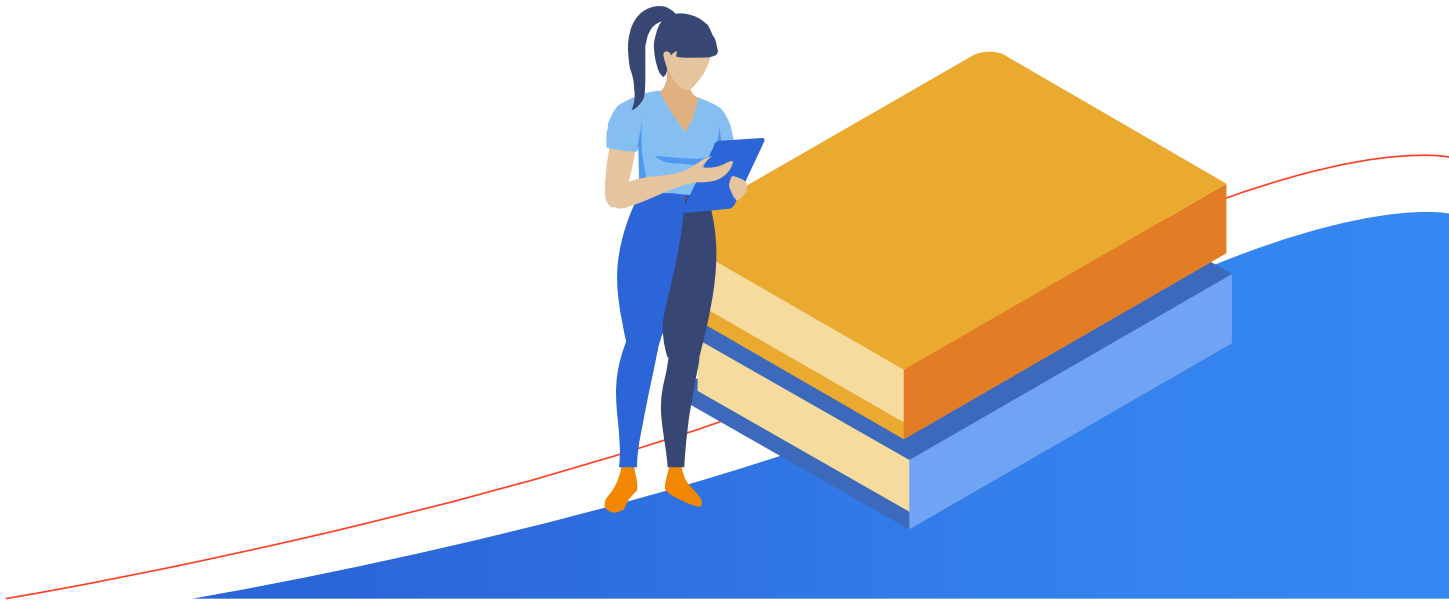
Turing AI 万维图灵 | 大数据系列课程



大数据技术-第六章：Zookeeper组件安装配置

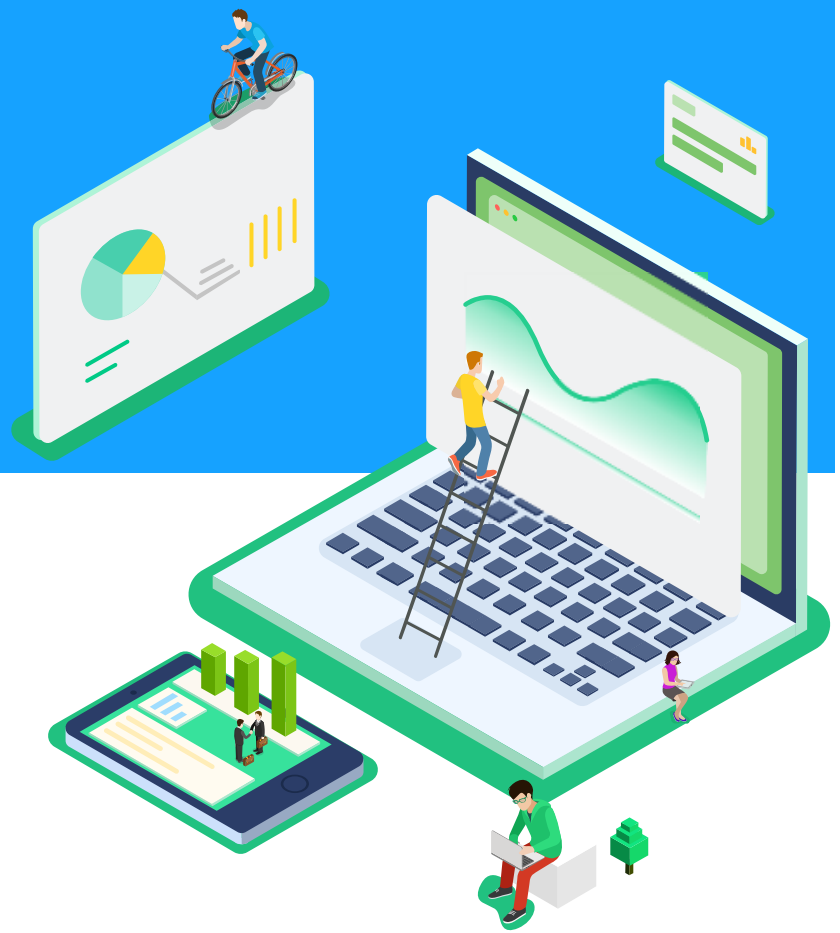
ZooKeeper角色选举

CONTENTS



01

ZooKeeper角色关系

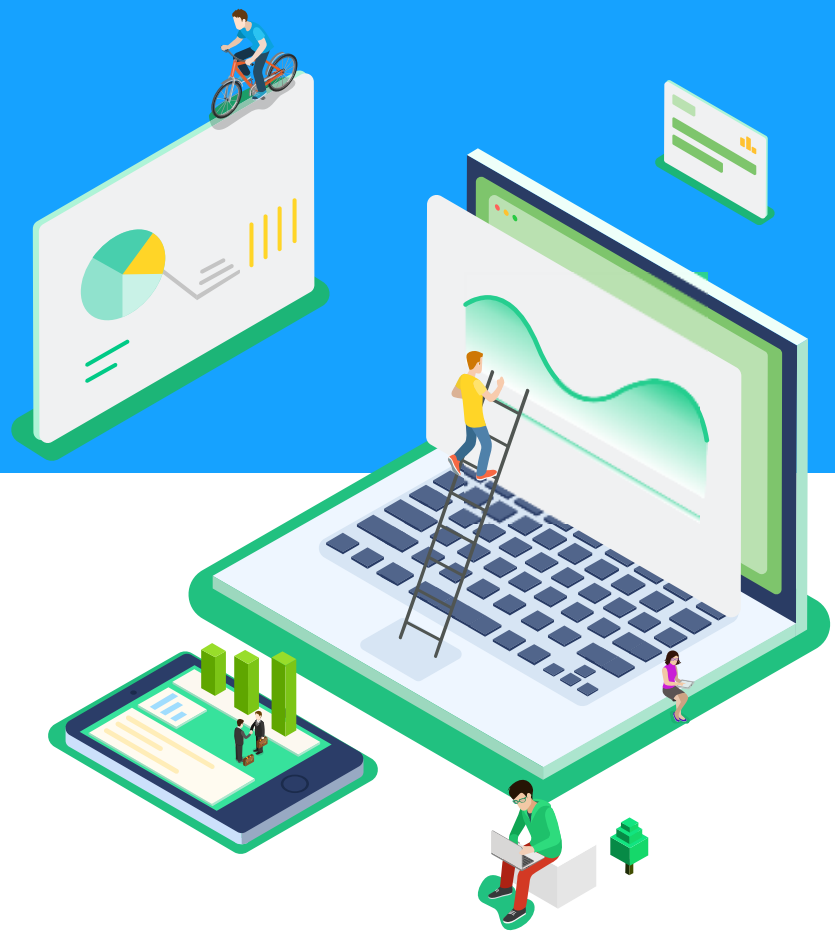


ZooKeeper角色关系

- 最典型集群模式：Master/Slave模式（主备模式）。在这种模式中，通常Master服务器作为主服务器提供写服务，其他Slave从服务器通过异步复制的方式获取Master服务器最新的数据提供读服务。
- 在ZooKeeper中没有选择传统的Master/Slave概念，而是引入了leader、follower和observer三种角色。ZooKeeper集群中的所有机器通过一个leader选举过程来选定一台称为“leader”的机器，leader既可以为客户端提供写服务又能提供读服务。
- 除了leader外，follower和observer都只能提供读服务。follower和observer唯一的区别在于observer机器不参与leader的选举过程，也不参与写操作的“过半写成功”策略，因此observer机器可以在不影响写性能的情况下提升集群的读性能。

02

Zookeeper工作原理



ZooKeeper角色关系

角色		描述
领导者（leader）		负责进行投票的发起和决议，更新系统状态。
学习者（learner）	跟随者（follower）	用于接收客户请求并向客户端返回结果，在选主过程中参与投票。
	观察者（observer）	接收客户端连接，将写请求转发给leader节点。observer只同步leader的状态，不参加投票过程。observer的目的是为了扩展系统，提高读取速度。
客户端（client）		请求发起方。

Zookeeper工作原理

- Zookeeper的核心是原子广播，这个机制保证了各个server之间的同步。实现这个机制的协议叫做Zab协议。Zab协议有两种模式，它们分别是恢复模式和广播模式。当服务启动或者在领导者崩溃后，Zab就进入了恢复模式，当领导者被选举出来，且大多数server的完成了和leader的状态同步以后，恢复模式就结束了。状态同步保证了leader和server具有相同的系统状态。
- 一旦leader已经和多数的follower进行了状态同步后，他就可以开始广播消息了，即进入广播状态。这时候当一个server加入zookeeper服务中，它会在恢复模式下启动，发现leader，并和leader进行状态同步。待到同步结束，它也参与消息广播。Zookeeper服务一直维持在Broadcast状态，直到leader崩溃了或者leader失去了大部分的followers支持。

Zookeeper工作原理

- 广播模式需要保证proposal被按顺序处理，因此zk采用了递增的事务id号(zxid)来保证。所有的提议(proposal)都在被提出的时候加上了zxid。实现中zxid是一个64为的数字，它高32位是epoch用来标识leader关系是否改变，每次一个leader被选出来，它都会有一个新的epoch。低32位是个递增计数。
- 当leader崩溃或者leader失去大多数的follower，这时候zk进入恢复模式，恢复模式需要重新选举出一个新的leader，让所有的server都恢复到一个正确的状态。

Turing AI 万维图灵 | 大数据系列课程

大数据

BIG DATA

智 / 能 / 科 / 技　　放 / 眼 / 未 / 来