

## 大数据技术-第六章：Zookeeper组件安装配置 ZooKeeper相关知识



# CONTENTS

---

01. 什么是ZooKeeper

02. ZooKeeper重要概念

03. ZooKeeper特点



# 01

## 什么是ZooKeeper



## >> 什么是ZooKeeper

- Zookeeper 是Google 的Chubby一个开源的实现，是Hadoop 的分布式协调服务。
- 分布式应用程序可以基于它实现同步服务，配置维护和命名服务等。
- 大部分分布式应用需要一个主控、协调器或控制器来管理物理分布的子进程（如资源、任务分配等）。
- 目前，大部分应用需要开发私有的协调程序，缺乏一个通用的机制。
- 协调程序的反复编写浪费，且难以形成通用、伸缩性好的协调器。
- ZooKeeper：提供通用的分布式锁服务，用以协调分布式应用。

# 02

## ZooKeeper重要概念



会话 (Session) : Session指的是ZooKeeper服务器与客户端会话。在ZooKeeper中, 一个客户端连接是指客户端和服务端之间的一个TCP长连接。客户端启动的时候, 首先会与服务器建立一个TCP连接, 从第一次连接建立开始, 客户端会话的生命周期也开始了。通过这个连接, 客户端能够通过心跳检测与服务器保持有效的会话, 也能够向ZooKeeper服务器发送请求并接受响应, 同时还能够通过该连接接收来自服务器的Watch事件通知。Session的sessionTimeout值用来设置一个客户端会话的超时时间。当由于服务器压力太大、网络故障或是客户端主动断开连接等各种原因导致客户端连接断开时, 只要在sessionTimeout规定的时间内能够重新连接上集群中任意一台服务器, 那么之前创建的会话仍然有效。在为客户端创建会话之前, 服务端首先会为每个客户端都分配一个sessionID。由sessionID是ZooKeeper会话的一个重要标识, 许多与会话相关的运行机制都是基于这个sessionID的, 因此, 无论是哪台服务器为客户端分配的sessionID, 都务必保证全局唯一。



Znode：在谈到分布式的时候，我们通常说的“节点”是指组成集群的每一台机器。然而，在ZooKeeper中，“节点”分为两类，第一类同样是指构成集群的机器，我们称之为机器节点；第二类则是指数据模型中的数据单元，我们称之为数据节点Znode。ZooKeeper将所有数据存储在内存中，数据模型是一棵树（Znode Tree），由斜杠（/）的进行分割的路径，就是一个Znode，例如/foo/path1。每个上都会保存自己的数据内容，同时还会保存一系列属性信息。在ZooKeeper中，node可以分为持久节点和临时节点两类。所谓持久节点是指一旦这个Znode被创建了，除非主动进行Znode的移除操作，否则这个Znode将一直保存在ZooKeeper上。而临时节点就不一样了，它的生命周期和客户端会话绑定，一旦客户端会话失效，那么这个客户端创建的所有临时节点都会被移除。另外，ZooKeeper还允许用户为每个节点添加一个特殊的属性：SEQUENTIAL。一旦节点被标记上这个属性，那么在这个节点被创建的时候，ZooKeeper会自动在其节点名后面追加上一个整型数字，这个整型数字是一个由父节点维护的自增数字。

- 版本：在前面我们已经提到，ZooKeeper的每个ZNode上都会存储数据，对应于每个ZNode，ZooKeeper都会为其维护一个叫作Stat的数据结构，Stat中记录了这个ZNode的三个数据版本，分别是version（当前ZNode的版本）、cversion（当前ZNode子节点的版本）和zxversion（当前ZNode的ACL版本）。
- Watcher（事件监听器），是ZooKeeper中的一个很重要的特性。ZooKeeper允许用户在指定节点上注册一些Watcher，并且在一些特定事件触发的时候，ZooKeeper服务端会将事件通知到感兴趣的客户端上去，该机制是ZooKeeper实现分布式协调服务的重要特性。
- ACL：ZooKeeper采用ACL（AccessControlLists）策略来进行权限控制，类似于UNIX文件系统的权限控制。



03

## ZooKeeper特点



- 顺序一致性：从同一客户端发起的事务请求，最终将会严格地按照顺序被应用到ZooKeeper中去。
- 原子性：所有事务请求的处理结果在整个集群中所有机器上的应用情况是一致的，要么整个集群中所有的机器都成功应用了某一个事务，要么都没有应用。
- 单一系统映像：无论客户端连到哪一个ZooKeeper服务器上，其看到的服务端数据模型都是一致的。
- 可靠性：一旦一次更改请求被应用，更改的结果就会被持久化，直到被下一次更改覆盖。

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

