

ZooKeeper——开源项目阅读与分析

寇伊 2015K8009915002

一、项目具体功能

从命名的角度讲，ZooKeeper，顾名思义，动物园的管理员，结合以往了解到的 IT 届各种项目的命名规则和以动物作为吉祥物的癖好，大数据处理器 Hadoop 是一只黄色的大象；负责数据仓库的 Hive 是虚拟的蜂巢；管理 web 容器的 tomcat 是一直雄猫；数据分析项目 Apache Pig 则是一头猪；那么 ZooKeeper 就非常形象的成为了所有系统的管理者，因此它是一个负责分布式协调的项目。

官方给出的解释是 ZooKeeper 分布式服务框架是 Apache Hadoop 的一个子项目，它主要用于解决分布式应用中的常见数据管理问题，比如：统一命名服务、状态同步服务、集群管理、分布式应用配置项的管理。ZooKeeper 通过封装服务以达到将接口简单化，使性能高效化以及使功能稳定化从而简化分布式协调及管理的难度，提供高性能的分布式服务。

下面列举一些 ZooKeeper 中举足轻重的部分：

首先是 Zab 协议，Zab 协议是原子广播协议的简称，它是一种在分布式协调系统中支持崩溃恢复的广播协议，基于该协议，ZooKeeper 才能实现主备模式系统架构中集群副本的数据一致性。

其次是 Znode 结点，作为一种独特的数据结构，ZooKeeper 系统通过基于该数据结构的原语定义了一系列相关的操作。

最后值得一提的是 Watcher 通知机制，这种机制服务于 ZooKeeper 在分布式环境下通过消息以网络的形式发送给分布式应用程序。

二、主要分析的功能以及针对它的具体分析

在之后的阅读与分析中，我将着重关注 ZooKeeper 中的 Watcher 通知机制。

Watcher 通知机制的整体流程如下图所示，客户端先向 ZooKeeper 服务端成功注册想要监听的节点状态，同时客户端本地会存储该监听器相关的信息在 WatchManager 中，当 ZooKeeper 服务端监听的数据状态发生变化时，ZooKeeper 就会主动通知发送相应事件信息给相关会话客户端，客户端就会在本地响应式的回调相关 Watcher 的 Handler。

其中这项通知机制遵循如下特征：

■ Watch 是一次性的，每次都需要重新注册，并且客户端在会话异常结束时不会收到任何通知，而快速重连接时仍不影响接收通知。

📌 Watch 的回调执行都是顺序执行的，并且客户端在没有收到关注数据的变化事件通知之前是不会看到最新的数据，另外需要注意不要在 Watch 回调逻辑中阻塞整个客户端的 Watch 回调

📌 Watch 是轻量级的，WatchEvent 是最小的通信单元，结构上只包含通知状态、事件类型和节点路径。ZooKeeper 服务端只会通知客户端发生了什么，并不会告诉具体内容。

