ZooKeeper是如何服务HDFS的?

厉梦如 2019-05-19 发表

问题描述

Q：ZooKeeper是如何服务HDFS的?

解决方法

A：ZooKeeper提供了一个简单的机制来保证只有一个NameNode是活动的。如果当前的活动NameNode失效了，那么另一个NameNode将获取ZooKeeper的独占锁，表明自己是活动的节点。
ZKFailoverController process（ZKFC）是用于监控和管理NameNode状态的ZooKeeper客户端。每一个运行NameNode的机器还会运行ZKFC， ZKFC主要负责以下几点：
l 健康检查：ZKFC定期对本地NameNode发启健康检查命令，只要NameNode会及时响应，ZKFC认为该节点是健康。如果节点失效，冻结或以其他方式进入不健康状态，ZKFC将其标记为不健康即失效。
l ZooKeeper的会话管理：当本地NameNode处于健康状态，ZKFC将在ZooKeeper中持有一个会话。如果本地NameNode为Active，那么ZKFC还有持有一个"ephemeral"的节点作为锁，一旦本地NameNode失效了，那么这个节点将会被自动删除。
l 以ZooKeeper为基础选择：如果本地NameNode是健康的，并且ZKFC发现没有其他的NameNode持有那个独占锁。那么他将试图去获取该锁，一旦成功，那么它就需要执行Failover，然后成为Active的NameNode节点。Failover的过程是：第一步，对之前的NameNode执行fence。第二步，将本地NameNode转换到Active状态。