

银河麒麟服务器操作系统 V4

zookeeper 软件适配手册



KYLIN
银河麒麟

天津麒麟信息技术有限公司

2019年6月

目 录

目 录.....	I
1 概述.....	2
1.1 系统概述.....	2
1.2 环境概述.....	2
1.3 ZOOKEEPER 软件简介.....	2
1.4 ZOOKEEPER 原理介绍.....	2
1.5 ZOOKEEPER 特性.....	3
1.6 ZOOKEEPER 选举流程.....	3
2 ZOOKEEPER 软件适配.....	4
3 ZOOKEEPER 命令操作.....	4

1 概述

1.1 系统概述

银河麒麟服务器操作系统主要面向军队综合电子信息系统、金融系统以及电力系统等国家关键行业的服务器应用领域，突出高安全性、高可用性、高效数据处理、虚拟化等关键技术优势，针对关键业务构建的丰富高效、安全可靠的功能特性，兼容适配长城、联想、浪潮、华为、曙光等国内主流厂商的服务器整机产品，以及达梦、金仓、神通、南大通用等主要国产数据库和中创、金蝶、东方通等国产中间件，满足虚拟化、云计算和大数据时代，服务器业务对操作系统在性能、安全性及可扩展性等方面的需求，是一款具有高安全、高可用、高可靠、高性能的自主可控服务器操作系统。

1.2 环境概述

服务器型号	长城信安擎天 DF720 服务器
CPU 类型	飞腾 2000+处理器
操作系统版本	Kylin-4.0.2-server-sp2-2000-19050910.Z1
内核版本	4.4.131
zookeeper 版本	3.5.5

1.3 zookeeper 软件简介

ZooKeeper 是一个分布式的，开放源码的分布式应用程序协调服务，是 GoogleChubby 一个开源的实现，是 Hadoop 和 Hbase 的重要组成部分。它是一个为分布式应用提供一致性服务的软件，提供的功能包括：配置维护、域名服务、分布式同步、组服务等。

ZooKeeper 的目标就是封装好复杂易出错的关键服务，将简单易用的接口和性能高效、功能稳定的系统提供给用户。

ZooKeeper 包含一个简单的原语集，提供 Java 和 C 的接口。

1.4 Zookeeper 原理介绍

ZooKeeper 是以 Fast Paxos 算法为基础的，Paxos 算法存在活锁的问题，即当有多个 proposer 交错提交时，有可能互相排斥导致没有一个 proposer 能提交成功，而 Fast Paxos 作了一些优化，通过选举产生一个 leader (领导者)，只有 leader 才能提交 proposer，具体算法可见 Fast Paxos。ZooKeeper 的基本运转流程：

- (1) 选举 Leader;
- (2) 同步数据;

- (3) 选举 Leader 过程中算法有很多，但要达到的选举标准是一致的；
- (4) Leader 要具有最高的执行 ID，类似 root 权限；
- (5) 集群中大多数的机器得到响应并接受选出的 Leader。

1.5 Zookeeper 特性

- (1) 最终一致性：client 不论连接到哪个 Server，展示给它都是同一个视图，这是 ZooKeeper 最重要的性能。
- (2) 可靠性：具有简单、健壮、良好的性能，如果消息 m 被到一台服务器接受，那么它将被所有的服务器接受。
- (3) 实时性：ZooKeeper 保证客户端将在一个时间间隔范围内获得服务器的更新信息，或者服务器失效的信息。但由于网络延时等原因，ZooKeeper 不能保证两个客户端能同时得到刚更新的数据，如果需要最新数据，应该在读数据之前调用 sync()接口。
- (4) 等待无关 (wait-free)：慢的或者失效的 client 不得干预快速的 client 的请求，使得每个 client 都能有效的等待。
- (5) 原子性：更新只能成功或者失败，没有中间状态。
- (6) 顺序性：包括全局有序和偏序两种：全局有序是指如果在一台服务器上消息 a 在消息 b 前发布，则在所有 Server 上消息 a 都将在消息 b 前被发布；偏序是指如果一个消息 b 在消息 a 后被同一个发送者发布，a 必将排在 b 前面。

1.6 Zookeeper 选举流程

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。Zk 的选举算法有两种：一种是基于 basic paxos 实现的，另外一种是基于 fast paxos 算法实现的。系统默认的选举算法为 fast paxos。

basic paxos 流程：

- (1) 选举线程由当前 Server 发起选举的线程担任，其主要功能是对投票结果进行统计，并选出推荐的 Server；
- (2) 选举线程首先向所有 Server 发起一次询问(包括自己)；
- (3) 选举线程收到回复后，验证是否是自己发起的询问(验证 zxid 是否一致)，然后获取对方的 id(myid)，并存储到当前询问对象列表中，最后获取对方提议的 leader 相关信息(id,zxid)，并将这些信息存储到当次选举的投票记录表中；
- (4) 收到所有 Server 回复以后，就计算出 zxid 最大的那个 Server，并将这个

Server 相关信息设置成下一次要投票的 Server;

- (5) 线程将当前 zxid 最大的 Server 设置为当前 Server 要推荐的 Leader, 如果此时获胜的 Server 获得 $n/2 + 1$ 的 Server 票数, 设置当前推荐的 leader 为获胜的 Server, 将根据获胜的 Server 相关信息设置自己的状态, 否则, 继续这个过程, 直到 leader 被选举出来。

通过流程分析我们可以得出: 要使 Leader 获得多数 Server 的支持, 则 Server 总数必须是奇数 $2n+1$, 且存活的 Server 的数目不得少于 $n+1$. 每个 Server 启动后都会重复以上流程。在恢复模式下, 如果是刚从崩溃状态恢复的或者刚启动的 server 还会从磁盘快照中恢复数据和会话信息, zk 会记录事务日志并定期进行快照, 方便在恢复时进行状态恢复。

2 zookeeper 软件适配

- 1) 下载并解压 zookeeper

```
$ wget https://www-eu.apache.org/dist/zookeeper/stable/apache-zookeeper-3.5.5-bin.tar.gz
$ tar xvf apache-zookeeper-3.5.5-bin.tar.gz -C /usr/local/
$ cd /usr/local/apache-zookeeper-3.5.5-bin/
```

- 2) 修改 zookeeper 配置文件

```
$ cd conf
$ mv zoo_sample.cfg zoo.cfg
```

- 3) 启动 zookeeper 服务

```
$ /usr/local/apache-zookeeper-3.5.5-bin/bin/zkServer.sh start
```

```
root@sp2-arm64-hargrove-01:/usr/local/apache-zookeeper-3.5.5-bin# /usr/local/apache-zookeeper-3.5.5-bin/bin/zkServer.sh start
/usr/bin/java
ZooKeeper JMX enabled by default
Using config: /usr/local/apache-zookeeper-3.5.5-bin/bin/../conf/zoo.cfg
Starting zookeeper ... STARTED
root@sp2-arm64-hargrove-01:/usr/local/apache-zookeeper-3.5.5-bin# /usr/local/apache-zookeeper-3.5.5-bin/bin/zkServer.sh status
/usr/bin/java
ZooKeeper JMX enabled by default
Using config: /usr/local/apache-zookeeper-3.5.5-bin/bin/../conf/zoo.cfg
Client port found: 2181. Client address: localhost.
Mode: standalone
```

3 Zookeeper 命令操作

- 1) 进入 shell 命令行

```
$ /usr/local/apache-zookeeper-3.5.5-bin/bin/zkCli.sh
```

```

root@ip2-8rnq4-hargrove-01:/usr/local/apache-zookeeper-3.5.5-bin# /usr/local/apache-zookeeper-3.5.5-bin/bin/zkCli.sh
/usr/bin/java
Connecting to localhost:2181
2019-06-12 11:06:15,996 [myId:] - INFO [main:Environment@109] - Client environment:zookeeper.version=3.5.5-398fe37ea45dee01bf87dc1c042b5d3dce88653, built on 05/03/2019 12:07 GMT
2019-06-12 11:06:15,994 [myId:] - INFO [main:Environment@109] - Client environment:host.name=ip2-8rnq4-hargrove-01
2019-06-12 11:06:15,994 [myId:] - INFO [main:Environment@109] - Client environment:java.version=1.8.0_191
2019-06-12 11:06:15,999 [myId:] - INFO [main:Environment@109] - Client environment:java.vendor=Oracle Corporation
2019-06-12 11:06:15,999 [myId:] - INFO [main:Environment@109] - Client environment:java.home=/usr/lib/jvm/java-8-openjdk-arm64/jre
2019-06-12 11:06:15,999 [myId:] - INFO [main:Environment@109] - Client environment:java.class.path=/usr/local/apache-zookeeper-3.5.5-bin/bin/./zookeeper-server/target/classes:/usr/local/apache-zookeeper-3.5.5-bin/bin/./build/classes:/usr/local/apache-zookeeper-3.5.5-bin/bin/./lib/zookeeper-jute-3.5.5.jar:/usr/local/apache-zookeeper-3.5.5-bin/bin/./lib/zookeeper-3.5.5.jar:/usr/local/apache-zookeeper-3.5.5-bin/bin/./lib/slf4j-log4j12-1.7.25.jar:/usr/local/apache-zookeeper-3.5.5-bin/bin/./lib/slf4j-api-1.7.25.jar:/usr/local/apache-zookeeper-3.5.5-bin/bin/./lib/netty-all-4.1.29.Final.jar:/usr/local/apache-zookeeper-3.5.5-bin/bin/./lib/log4j-1.2.17.jar:/usr/local/apache-zookeeper-3.5.5-bin/bin/./lib/json-simple-1.1.1.jar:/usr/local/apache-zookeeper-3.5.5-bin/bin/./lib/jline-2.11.jar:/usr/local/apache-zookeeper-3.5.5-bin/bin/./lib/jetty-util-9.4.17.v20190418.jar:/usr/local/apache-zookeeper-3.5.5-bin/bin/./lib/jetty-server-9.4.17.v20190418.jar:/usr/local/apache-zookeeper-3.5.5-bin/bin/./lib/jetty-http-9.4.17.v20190418.jar:/usr/local/apache-zookeeper-3.5.5-bin/bin/./lib/javax-servlet-api-3.1.0.jar:/usr/local/apache-zookeeper-3.5.5-bin/bin/./lib/jackson-databind-2.9.8.jar:/usr/local/apache-zookeeper-3.5.5-bin/bin/./lib/jackson-core-2.9.8.jar:/usr/local/apache-zookeeper-3.5.5-bin/bin/./lib/jackson-annotations-2.9.8.jar:/usr/local/apache-zookeeper-3.5.5-bin/bin/./lib/commons-cli-1.2.jar:/usr/local/apache-zookeeper-3.5.5-bin/bin/./lib/audience-annotations-0.5.0.jar:/usr/local/apache-zookeeper-3.5.5-bin/bin/./zookeeper-*.jar:/usr/local/apache-zookeeper-3.5.5-bin/bin/./zookeeper-server/src/main/resources/lib/*.jar:/usr/local/apache-zookeeper-3.5.5-bin/bin/./conf:
2019-06-12 11:06:15,999 [myId:] - INFO [main:Environment@109] - Client environment:java.library.path=/usr/java/packages/lib/aarch64:/usr/lib/aarch64-linux-gnu:/usr/lib/aarch64-linux-gnu:/usr/lib/aarch64-linux-gnu:/usr/lib/jni:/usr/lib:/usr/lib
2019-06-12 11:06:15,999 [myId:] - INFO [main:Environment@109] - Client environment:java.io.tmpdir=/tmp
2019-06-12 11:06:15,999 [myId:] - INFO [main:Environment@109] - Client environment:java.compiler=NA
2019-06-12 11:06:15,999 [myId:] - INFO [main:Environment@109] - Client environment:os.name=Linux
2019-06-12 11:06:15,999 [myId:] - INFO [main:Environment@109] - Client environment:os.arch=aarch64
2019-06-12 11:06:15,999 [myId:] - INFO [main:Environment@109] - Client environment:os.version=4.50-20180615.kylin.server.YUN-generice
2019-06-12 11:06:15,999 [myId:] - INFO [main:Environment@109] - Client environment:user.name=root
2019-06-12 11:06:15,999 [myId:] - INFO [main:Environment@109] - Client environment:user.home=/root
2019-06-12 11:06:15,999 [myId:] - INFO [main:Environment@109] - Client environment:user.dir=/usr/local/apache-zookeeper-3.5.5-bin
2019-06-12 11:06:15,999 [myId:] - INFO [main:Environment@109] - Client environment:os.memory.free=235MB
2019-06-12 11:06:15,999 [myId:] - INFO [main:Environment@109] - Client environment:os.memory.max=241MB
2019-06-12 11:06:15,999 [myId:] - INFO [main:Environment@109] - Client environment:os.memory.total=241MB
2019-06-12 11:06:15,999 [myId:] - INFO [main:ZooKeeper@868] - Initiating client connection, connectingLocalHost:2181 sessionTimeout=30000 watcher=org.apache.zookeeper.ZooKeeperMain$MyWatcher@4c4d35a9
2019-06-12 11:06:15,999 [myId:] - INFO [main:X509Util@79] - Setting -Djdk.tls.rejectClientInitiatedRenegotiation=true to disable client-initiated TLS renegotiation
2019-06-12 11:06:15,999 [myId:] - INFO [main:ClientCnxnSocket@237] - jute.maxbuffer value is 4194304 bytes
2019-06-12 11:06:15,999 [myId:] - INFO [main:ClientCnxn@1653] - zookeeper.request.timeout value is 0. Feature enabled=
Welcome to ZooKeeper!
2019-06-12 11:06:15,998 [myId:localhost:2181] - INFO [main:SendThread(localhost:2181):ClientCnxnSendThread@1112] - Opening socket connection to server localhost/127.0.0.1:2181. Will not attempt to authenticate using SASL (unknown error)
JLine support is enabled
2019-06-12 11:06:15,799 [myId:localhost:2181] - INFO [main:SendThread(localhost:2181):ClientCnxnSendThread@959] - Socket connection established, initiating session, client: /127.0.0.1:51680, server: localhost/127.0.0.1:2181
[zk: localhost:2181(CONNECTING) 0] 2019-06-12 11:06:15,814 [myId:localhost:2181] - INFO [main:SendThread(localhost:2181):ClientCnxnSendThread@1394] - Session establishment complete on server localhost/127.0.0.1:2181, sessionId = 0x10023540110000, negotiated timeout = 30000

WATCHER::
MatchedEvent state:SyncConnected type:None path:null
[zk: localhost:2181(CONNECTED) 0]
    
```

2) 查看当前 zookeeper 包含的内容

\$ ls /

```

[zk: localhost:2181(CONNECTED) 0] ls /
[zookeeper]
[zk: localhost:2181(CONNECTED) 1]
    
```

3) 创建新的节点及相关字符

\$ create /zk mydata

```

[zk: localhost:2181(CONNECTED) 2] create /zk mydata
Created /zk
[zk: localhost:2181(CONNECTED) 3] ls /
[zk, zookeeper]
[zk: localhost:2181(CONNECTED) 4]
    
```

4) 查看节点内容

\$ get /zk

```

[zk: localhost:2181(CONNECTED) 4] get /zk
mydata
[zk: localhost:2181(CONNECTED) 5]
    
```

5) 修改节点内容

\$ set /zk kylin2019

```

[zk: localhost:2181(CONNECTED) 5] set /zk kylin2019
[zk: localhost:2181(CONNECTED) 6] get /zk
kylin2019
[zk: localhost:2181(CONNECTED) 7]
    
```

6) 删除节点

\$ delete /zk

```

[zk: localhost:2181(CONNECTED) 7] delete /zk
[zk: localhost:2181(CONNECTED) 8] ls /
[zookeeper]
    
```