

**ZooKeeper 常见问题维护手册 V1.0**

# **ZooKeeper 常见问题维护手册 V1.0**

文档版本      01  
发布日期      2016-03-31

华为技术有限公司



**版权所有 © 华为技术有限公司 2016。 保留一切权利。**

非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。

## 商标声明



和其他华为商标均为华为技术有限公司的商标。

本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

## 注意

您购买的产品、服务或特性等应受华为公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为公司对本文档内容不做任何明示或默示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

# 华为技术有限公司

地址：                深圳市龙岗区坂田华为总部办公楼                邮编：518129

网址：                <http://www.huawei.com>

客户服务邮箱：      [support@huawei.com](mailto:support@huawei.com)

客户服务电话：      4008302118

# 目 录

---

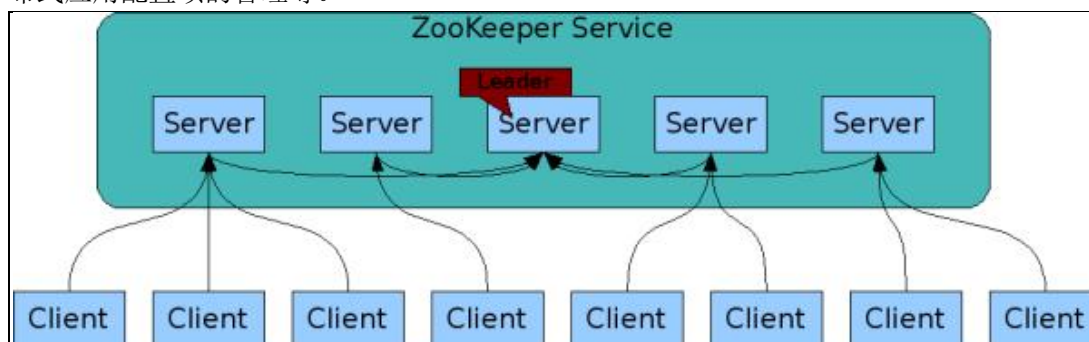
Zookeeper .....	3
1、基本概念.....	3
【概述】 .....	3
【日志概述】 .....	3
【客户端工具】 .....	3
【常用参数】 .....	5
2、常见问题.....	5
【启动监控告警】 .....	5
[ZooKeeper-10001] ZooKeeper 安装启动失败。 .....	5
[ZooKeeper-10002] ZooKeeper 实例状态异常。 .....	7
[ZooKeeper-003] ZooKeeper 服务不可用告警。 .....	8
【Shell 客户端操作问题】 .....	9
[ZooKeeper-20001] ZooKeeper Shell 客户端登录失败。 .....	9
【二次开发问题】 .....	11
[ZooKeeper-30001] 使用 IBM JDK 的时候，连接 ZooKeeper 失败。 .....	11
【咨询】 .....	12
[ZooKeeper-40001]怎样查看哪个 ZooKeeper 实例是 leader? .....	12
[ZooKeeper-40002]怎样查看 znode 的临时节点还是永久节点? .....	12
[ZooKeeper-40003]能否为临时节点创建子节点? .....	12

# Zookeeper

## 1、基本概念

### 【概述】

Zookeeper 分布式服务框架是 Apache Hadoop 的一个子项目，它主要是用来解决分布式应用中经常遇到的一些数据管理问题，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等。



### 【日志概述】

ZooKeeper 日志的默认存储路径为 “/var/log/Bigdata/zookeeper/quorumpeer”。

说明：zookeeper 日志详细说明。

运行日志：/var/log/Bigdata/zookeeper/quorumpeer/zookeeper-omm-server-主机名.log

健康状态检查日志：/var/log/Bigdata/zookeeper/quorumpeer/check-serviceDetail.log

审计日志：/var/log/Bigdata/audit/zookeeper/quorumpeer/zk-audit-quorumpeer.log

### 【客户端工具】

#### Shell 客户端使用

在 FI 页面，可以下载安装客户端。安装客户端后，可以使用如下命令进入 ZooKeeper 命令行。

```
sh zkCli.sh -server 192.168.0.100:24002
```

```
sh zkCli.sh -server 192.168.0.100:24002,192.168.0.101:24002
```

说明：

- 1) 端口：默认为 24002，可查看 zoo.cfg 中的 clientPort 值获取该端口
- 2) IP：需要是 Zookeeper 服务端业务平面 IP。

## Shell 客户端命令

在客户端中，输入 help 可以看到 zookeeper 的客户端命令。

命令概述	实例	参数详解
查看帮助	help	查看各参数的使用说明
查子节点	ls / 或者 ls /hadoop-ha	说明: ls 后的节点必须以/开头, 且绝对路径
创建节点	create /test testdata create /test/t1 t1data	创建节点
查看节点内容	get /test	查看节点/test 中存储的数据
查看节点权限	getAcl /test	查看节点的 acl 信息, 默认节点的 acl 信息为'world,' anyone: cdrwa
节点统计信息	stat /test	查看节点的统计信息, 如创建时间、修改时间、子节点个数, 节点大小
删除节点	rmr /test deleteall /test	递归删除/test 节点及子节点。 建议使用 deleteall 命令。 删除性能: 1 小时 200 万 znode
修改节点	set /test "this is update msg"	修改/test 的内容
退出	quit	退出 shell 客户端

## 四字命令

ZooKeeper 支持某些特定的四字命令字母与其的交互。它们大多是查询命令, 用来获取 ZooKeeper 服务的当前状态及相关信息。用户在客户端可以通过 nc 或 netcat 向 ZooKeeper 提交相应的命令

概述	命令实例	命令详情
统计	echo stat   netcat 192.168.0.100 24002	来查看哪个节点被选择作为 follower 或者 leader
是否启动	echo ruok   netcat 192.168.0.100 24002	测试是否启动了该 Server, 若回复 imok 表示已经启动
未处理	echo dump  netcat 192.168.0.100 24002	列出未经处理的会话和临时节点
配置	echo conf  netcat 192.168.0.100 24002	输出相关服务配置的详细信息
连接数	echo cons   netcat 192.168.0.100 24002	列出所有连接到服务器的客户端的完全的连接 / 会话的详细信息。
环境	echo envi   netcat 192.168.0.100 24002	输出关于服务环境的详细信息 (区别于 conf 命令)

未处理请求	echo reqs   netcat 192.168.0.100 24002	列出未经处理的请求
watch	echo wchs   netcat 192.168.0.100 24002	列出服务器 watch 的详细信息
Session	echo whc   netcat 192.168.0.100 24002	通过 session 列出服务器 watch 的详细信息,它的输出是一个与 watch 相关的会话的列表。

## 【常用参数】

ZooKeeper 的常用参数不多，以下列举出一些重要的参数。

参数名	默认值	参数介绍
skipACL	no	是否跳过 ACL 检查。在配置互信切域的时候,需要改为yes。当切换为yes后,所有kerberos认证后的用户,操作 zookeeper 节点会跳过内部权限检查。
clientPort	24002	Zookeeper 对外服务端口。
dataDir	/srv/BigData/zookeeper	zookeeper 数据目录。修改 ZooKeeper 数据目录的时候, 需要将旧的目录数据拷贝到新目录。
maxClientCnxns	2000	在集群中 ZooKeeper 连接数满的情况下,可以调大。
GC_OPTS	-Xms512M -Xmx1G	ZooKeeper 进程内存设置, 在 ZooKeeper 中数据量大的情况下需要加大

## 2、常见问题

### 【启动监控告警】

[ZooKeeper-10001] ZooKeeper 安装启动失败。

#### 【问题背景与现象】

使用 FusionInsight 安装集群，启动 ZooKeeper 失败。

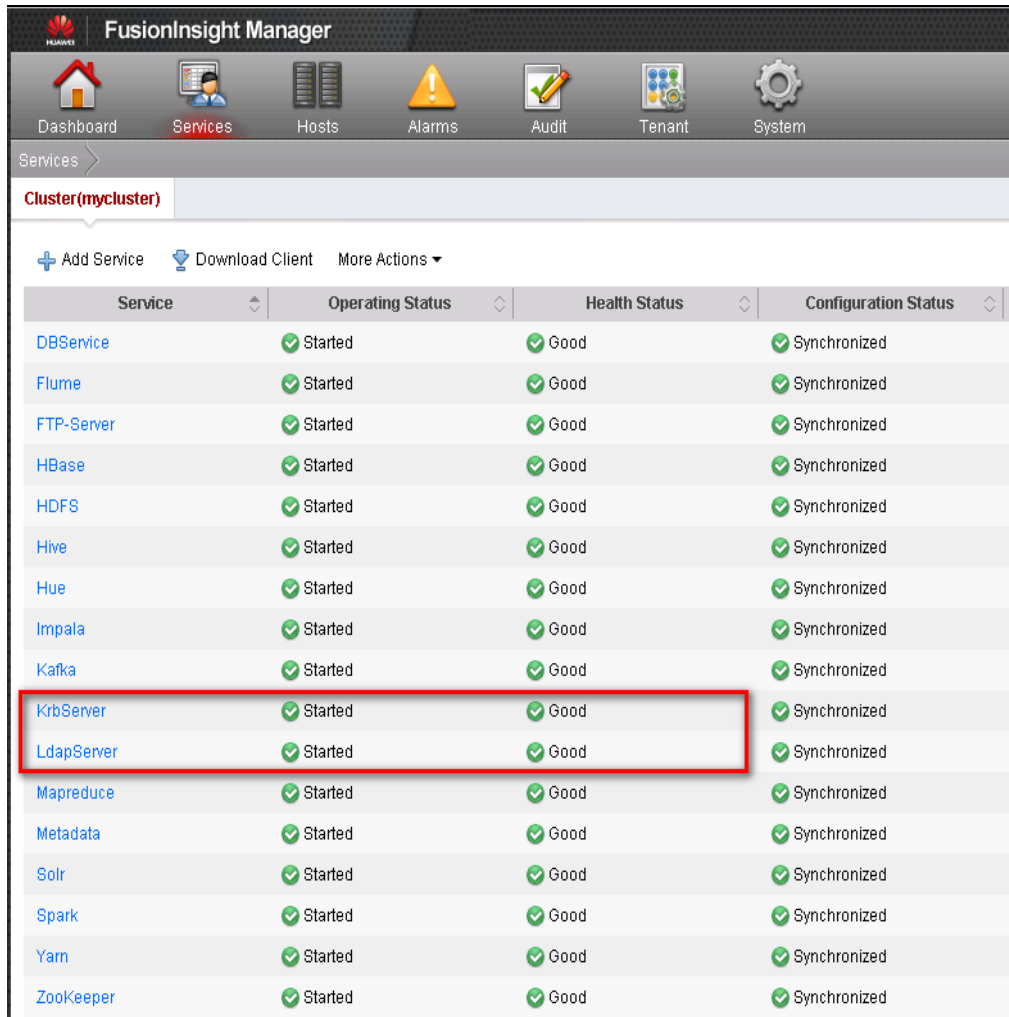
#### 【可能原因】

1. ZooKeeper 启动依赖的 LdapServer 或者 KrbServer 底层服务异常
2. ZooKeeper 数据目录权限不对

## 【原因分析】

在 ZooKeeper 启动依赖 LdapServer 或者 KrbServer，可能 LdapServer 或者 KrbServer 组件异常导致 ZooKeeper 安装启动失败。

1. 通过 FusionInsight Manager 页面，点击 Service，查看当前 LdapServer 和 KrbServer 组件当前状态，发现操作状态 **Started** 且健康状态为 **Good**。如果状态不正常，即先解决底层组件异常问题，否则继续排查。



Service	Operating Status	Health Status	Configuration Status
DBService	Started	Good	Synchronized
Flume	Started	Good	Synchronized
FTP-Server	Started	Good	Synchronized
HBase	Started	Good	Synchronized
HDFS	Started	Good	Synchronized
Hive	Started	Good	Synchronized
Hue	Started	Good	Synchronized
Impala	Started	Good	Synchronized
Kafka	Started	Good	Synchronized
KrbServer	Started	Good	Synchronized
LdapServer	Started	Good	Synchronized
Mapreduce	Started	Good	Synchronized
Metadata	Started	Good	Synchronized
Solr	Started	Good	Synchronized
Spark	Started	Good	Synchronized
Yarn	Started	Good	Synchronized
ZooKeeper	Started	Good	Synchronized

2. ZooKeeper 数据目录没权限  
登录 ZooKeeper 实例所在节点，查看 ZooKeeper 数据目录权限是否正确（默认 /srv/BigData/zookeeper）。如果 zookeeper 目录的权限不是 700 且属组为 omm:wheel，即权限设置有问题。需要手动调整。

```
189-39-172-210: /srv/BigData # ll /srv/BigData/
total 36
d----- 2 omm wheel 4096 Feb 25 19:09 Manager
drwx----- 3 omm wheel 4096 Mar 7 19:27 dbdata_om
drwx----- 3 omm wheel 4096 Mar 7 19:27 dbdata_service
drwxr-x--- 5 omm wheel 4096 Feb 25 19:09 hadoop
drwx----- 3 omm wheel 4096 Mar 7 21:38 journalnode
drwx----- 3 omm wheel 4096 Mar 7 19:27 namenode
drwx----- 3 omm wheel 4096 Mar 8 10:45 solr
drwxr-x--- 3 omm wheel 4096 Mar 2 11:29 varn
drwx----- 4 omm wheel 4096 Mar 7 21:25 zookeeper
```

## 【解决办法】

1. 如果是 LdapServer 或者 KrbServer 状态异常，请先解决对应组件问题，然后尝试重启 ZooKeeper
2. 如果是磁盘权限不对，修改目录权限后重启 ZooKeeper。

```
chown -R omm:wheel /srv/BigData/zookeeper
chmod 700 /srv/BigData/zookeeper
```

3. 如果是磁盘空间不足，请确认 ZooKeeper 是否独立挂盘，如果是共盘，需要清理不需要的数据后尝试重启 ZooKeeper。

## [ZooKeeper-10002] ZooKeeper 实例状态异常。

## 【问题背景与现象】

使用 FusionInsight R2C30 或者 C50 版本安装集群后，ZooKeeper 实例状态为启动失败状态或者 ConCern 状态。

## 【可能原因】

1. ZooKeeper 所在节点配置了 DNS 服务。
2. ZooKeeper 所在节点的 “/etc/hosts” 配置错误。
3. ZooKeeper 元数据磁盘空间使用率超过 95%

## 【原因分析】

“/etc/hosts” 不对，或者配置 DNS 会导致 ZooKeeper 实例状态时好时坏。

1. 使用 PuTTY 工具，依次以 root 用户连接到故障 ZooKeeper 所在节点上。
2. 查看 “/etc/hosts” 文件是否正确。
3. 如果节点安装并启用了 DNS 服务，需要停用该节点的 DNS。

### **service named stop**

如果出现如下结果，即说明 DNS 服务停止成功。

Shutting down name server BIND waiting for named to shut down (28s)

4. 执行以下命令，查看节点是否安装 DNS 客户端。

```
cat /etc/resolv.conf
```



如果出现了类似如下内容，说明节点安装并启用了 DNS 客户端服务。

nameserver 8.8.8.8

nameserver 8.8.4.4

5. 如果节点安装并启用了 DNS 客户端服务，需要停用该节点的 DNS 客户端服务。

**vi /etc/resolv.conf**

将内容全部用“#”注释并保存文件。

# nameserver 8.8.8.8

# nameserver 8.8.4.4

6. 在 FI C50SPC200 及后续版本， ZooKeeper 会监控磁盘剩余空间，如果磁盘空间使用率超过 95%，即无法正常运行。

```
189-39-172-210: /srv/BigData # df -h /srv/BigData/zookeeper/  
Filesystem      Size  Used Avail Use% Mounted on  
/dev/xvdel      20G   183M   20G    1% /srv/BigData/zookeeper
```

### 【解决办法】

1. 如果/etc/hosts 文件不对，请删除错误的项目。
2. 如果是配置了 DNS，需要停用 DNS。如果必须使用，请联系华为的技术支持。
3. 如果是磁盘空间不足，请确认 ZooKeeper 是否独立挂盘，如果是共享磁盘，需要清理不需要的数据后尝试重启 ZooKeeper。

## [ZooKeeper-003] ZooKeeper 服务不可用告警。

### 【问题背景与现象】

使用 FusionInsight R2C30 或者 C50 版本安装集群，出现 ZooKeeper 服务不可用告警。

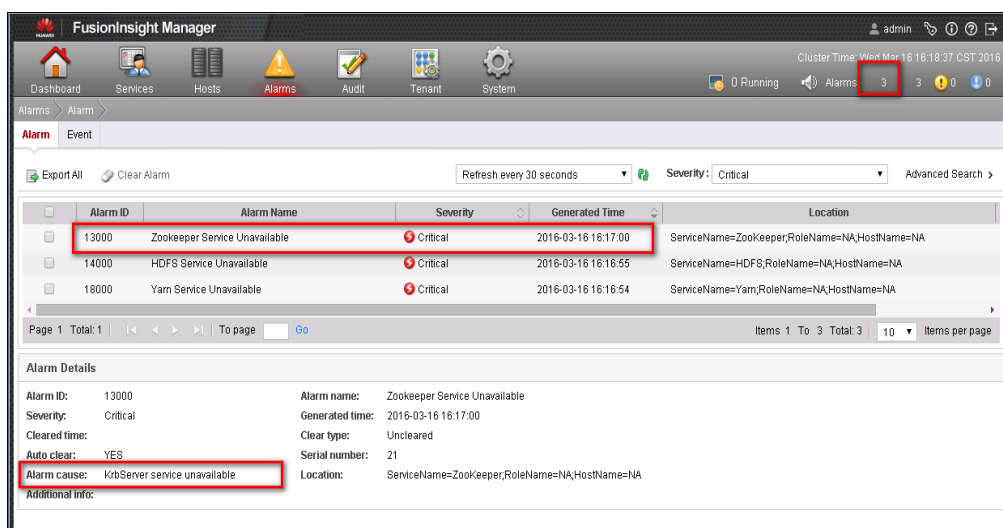
### 【可能原因】

1. ZooKeeper 启动依赖的 LdapServer 或者 KrbServer 底层服务异常
2. 如果数据节点和 ZooKeeper 合布
3. ZooKeeper 未单独挂载磁盘

### 【原因分析】

在 ZooKeeper 启动依赖 LdapServer 或者 KrbServer，可能 LdapServer 或者 KrbServer 组件报服务不可用，导致 ZooKeeper 服务不可用告警。

1. 通过 FusionInsight Manager 页面，点击有上交告警按钮，即可进告警列表页面
2. 点击告警编号，查看告警详情，如果“Alarm Cause”是 KrbServer 或者 LdapServer 服务不可用，需要查看对应底层组件服务状态，确保启动且运行状态正常。



3. 如果 ZooKeeper 异常的原因不是底层组件异常，请排查 ZooKeeper 实例是否有异常状态。请参考《[ZooKeeper-10002] ZooKeeper 实例状态异常》章节处理。

### 【解决办法】

1. 确保底层组件 LdapServer 和 KrbServer 运行正常。
2. 如果有实例异常，解决 ZooKeeper 实例异常的问题。
3. 如果偶现服务不可用，需要检查 ZooKeeper 所在节点资源情况。
4. 另外需要查看此告警是否上报后马上清除，在 C30SPC600 和 C30SPC601 中，ZooKeeper 健康检查比较敏感。当 ZooKeeper 数据目录没有挂载独立磁盘或者节点 CPU 使用率高的情况下，会出 ZooKeeper 报服务不可用，且马上清除的情况。

### 【Shell 客户端操作问题】

[ZooKeeper-20001] ZooKeeper Shell 客户端登录失败。

### 【问题背景与现象】

ZooKeeper 的 Shell 客户端失败。

说明：ZooKeeper 客户端命令为：sh zkCli.sh -server 192.168.1.100:24002

其中：192.168.1.100 是任意一个 ZooKeeper 实例的业务 IP（Business IP）

### 【可能原因】

1. ZooKeeper 命令使用不正确，如果 ZooKeeper 命令没有传递节点信息，会出现如下错误 “Empty nameStrings not allowed”

```
2016-03-16 20:15:09,658 [myid:] - ERROR [main-SendThread(localhost:2181):ZooKeeperSaslClient@295] - Exception while trying to create SASL client
java.lang.IllegalArgumentException: Empty nameStrings not allowed
```

2. 执行命令前，没有“source 环境变量”或者“使用 kinit 登录账号”

如果没有登录账号，会出现如下的日志。

```
[zk: 192.168.0.100:24002(CONNECTING) 0] Debug is true storeKey false
useTicketCache true useKeyTab false doNotPrompt false ticketCache is null
isInitiator true KeyTab is null refreshKrb5Config is false principal is null
tryFirstPass is false useFirstPass is false storePass is false clearPass is
false
Acquire TGT from Cache
Principal is null
null credentials from Ticket Cache
2016-03-16 20:19:51,643 [myid:] - WARN [main-
SendThread(192.168.0.100:24002):ZooKeeperSaslClient$ClientCallbackHandler@492]
- Could not login: the client is being asked for a password, but the Zookeeper
client code does not currently support obtaining a password from the user.
Make sure that the client is configured to use a ticket cache (using the JAAS
configuration setting 'useTicketCache=true') and restart the client. If you
still get this message after that, the TGT in the ticket cache has expired and
must be manually refreshed. To do so, first determine if you are using a
password or a keytab. If the former, run kinit in a Unix shell in the
environment of the user who is running this Zookeeper client using the command
'kinit <princ>' (where <princ> is the name of the client's Kerberos
principal). If the latter, do 'kinit -k -t <keytab> <princ>' (where <princ> is
the name of the Kerberos principal, and <keytab> is the location of the keytab
file). After manually refreshing your cache, restart this client. If you
continue to see this message after manually refreshing your cache, ensure that
your KDC host's clock is in sync with this host's clock.
[Krb5LoginModule] authentication failed
No password provided
2016-03-16 20:19:51,670 [myid:] - WARN [main-
SendThread(192.168.0.100:24002):ClientCnxn$SendThread@1029] - SASL
configuration failed: javax.security.auth.login.LoginException: No password
provided Will continue connection to Zookeeper server without SASL
authentication, if Zookeeper server allows it.
```

### 3. 网络不通，或者节点开启了防火墙。

如果客户端机器和 ZooKeeper 所在主机网络不通，或者使用了管理平面的 IP，即会出现如下的错误：“Can not get the principle name from server”。客户端或者服务端防火墙打开的情况下，也会出现此错误。

```
2016-03-16 20:23:26,919 [myid:] - WARN [main-
SendThread(192.168.0.100:24002):ClientCnxn$SendThread@958] - Can not get the
principle name from server 192.168.0.100
JLine support is enabled
[zk: 192.168.0.100:24002(CONNECTING) 0] Debug is true storeKey false
useTicketCache true useKeyTab false doNotPrompt false ticketCache is null
isInitiator true KeyTab is null refreshKrb5Config is false principal is null
tryFirstPass is false useFirstPass is false storePass is false clearPass is
false
Acquire TGT from Cache
Principal is admin@HADOOP.COM
Commit Succeeded
[zk: 192.168.0.100:24002(CONNECTING) 0] 2016-03-16 20:23:27,520 [myid:] -
ERROR [main-SendThread(10.64.35.11:24002):ZooKeeperSaslClient@295] - Exception
while trying to create SASL client
java.security.PrivilegedActionException: javax.security.sasl.SaslException:
Failure to initialize security context [Caused by GSSException: Invalid name
provided (Mechanism level: Illegal character in realm name; one of: '/', ':',
'' (600))]
```

## 【解决办法】

1. 使用正确的客户端命令，正确的ZooKeeper服务端IP。  
说明：ZooKeeper 客户端命令为：sh zkCli.sh -server 192.168.1.100:24002  
其中：192.168.1.100 是任意一个 ZooKeeper 实例的业务 IP（Business IP）  
使用客户端前，需要 source /opt/client/bigdata\_env 和 kinit 账号名。登录
2. 检查客户端节点和服务端节点的防火墙状态，确保防火墙关闭。

## 【二次开发问题】

[ZooKeeper-30001] 使用 IBM JDK 的时候，连接 ZooKeeper 失败。

### 【问题背景与现象】

使用 IBM 的 JDK 的情况下客户端连接 ZooKeeper 失败。

### 【可能原因】

1. jaas.conf文件内容不对

### 【原因分析】

1. IBM的JDK和普通JDK的jaas.conf文件格式不一样。

## 【解决办法】

1. 若使用的JDK版本为IBM JDK，请使用如下jaas.conf文件模板，其中“useKeytab”中文件路径必须以“file:”开头，后面跟绝对路径。

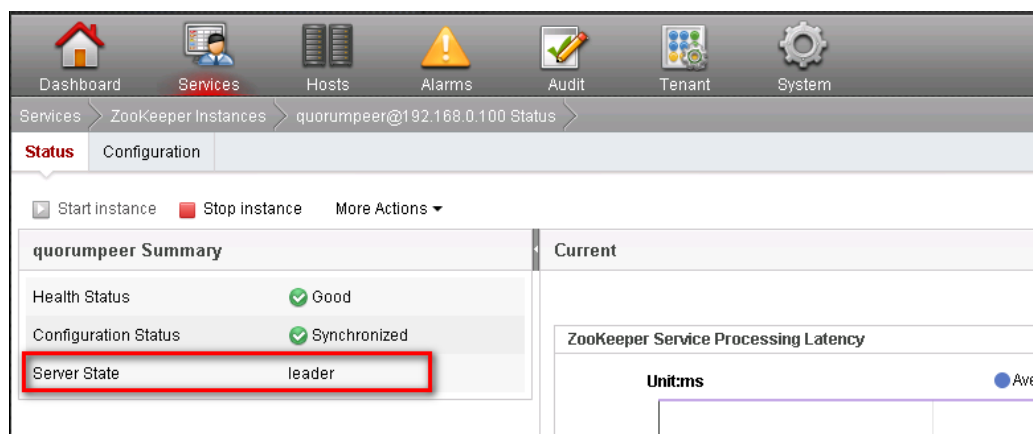
```
Client {  
    com.ibm.security.auth.module.Krb5LoginModule required  
    useKeytab="file:///D:/install/HbaseClientSample/conf/user.keytab"  
    principal="hbaseuser1"  
    credsType="both";  
};
```

## 【咨询】

[ZooKeeper-40001]怎样查看哪个 ZooKeeper 实例是 leader?

## 【解决办法】

1. 通过FI页面，进入任意ZooKeeper实例的详情页面，可以看到该实例的状态



[ZooKeeper-40002]怎样查看 znode 的临时节点还是永久节点?

## 【解决办法】

1. 登录ZooKeeper Shell客户端，使用get /znode名称，即可查看节点是否临时节点。

如果ephemeralOwner对应的值是“0x0”即节点为永久节点。当ephemeralOwner的值非零，即临时节点，该节点会在对应session（如示例的：0x70071eaba61065a）结束的时候消失。

```
[zk: 192-168-0-100:24002 (CONNECTED) 4] get /t6
...
ephemeralOwner = 0x70071eaba61065a
...
[zk: 192-168-0-100:24002 (CONNECTED) 5] get /t5
...
ephemeralOwner = 0x0
```

[ZooKeeper-40003]能否为临时节点创建子节点?

## 【解决办法】

1. 不能。

例如给临时节点/t6创建子节点会报错：“Ephemerals cannot have children”。

```
[zk: 192-168-0-100:24002(CONNECTED) 10] create /t6/test testdata  
Ephemerals cannot have children: /t6/test
```