# zookeeper

## 一.定义

Zookeeper是一个分布式的，开放源代码的分布式应用程序协调服务，是Google的Chubby一个开源实现

## 二.Paxos 算法

### 1.Paxos 算法

解决的问题是一个[分布式系统](https://baike.baidu.com/item/%E5%88%86%E5%B8%83%E5%BC%8F%E7%B3%BB%E7%BB%9F)如何就某个值（决议）达成一致

### 2.Paxos 算法

Paxos 算法就是一种基于消息传递模型的一致性算法。

### 3.节点通信存在两种模型：

[共享内存](https://baike.baidu.com/item/%E5%85%B1%E4%BA%AB%E5%86%85%E5%AD%98)（Shared memory）和消息传递（Messages passing）Paxos 算法用消息传递实现

### 4.Google Chubby :

Google公司开发的一个分布式锁服务

### 5.GFS(google文件系统)

 GFS是一个可扩展的[分布式文件系统](https://baike.baidu.com/item/%E5%88%86%E5%B8%83%E5%BC%8F%E6%96%87%E4%BB%B6%E7%B3%BB%E7%BB%9F)，用于大型的、分布式的、对大量数据进行访问的应用。

### 6.QPS

QPS 规定时间内所处理数据流量多少

### 7.Zookeeper

Zookeeper 开放源码的[分布式应用程序](https://baike.baidu.com/item/%E5%88%86%E5%B8%83%E5%BC%8F%E5%BA%94%E7%94%A8%E7%A8%8B%E5%BA%8F)协调服务

### 8.ACL

[访问控制列表](https://baike.baidu.com/item/%E8%AE%BF%E9%97%AE%E6%8E%A7%E5%88%B6%E5%88%97%E8%A1%A8)（Access Control List，ACL） 是[路由器](https://baike.baidu.com/item/%E8%B7%AF%E7%94%B1%E5%99%A8)和[交换机](https://baike.baidu.com/item/%E4%BA%A4%E6%8D%A2%E6%9C%BA)接口的指令列表，用来[控制端口](https://baike.baidu.com/item/%E6%8E%A7%E5%88%B6%E7%AB%AF%E5%8F%A3)进出的数据包。ACL适用于所有的[被路由协议](https://baike.baidu.com/item/%E8%A2%AB%E8%B7%AF%E7%94%B1%E5%8D%8F%E8%AE%AE)，如IP、IPX、AppleTalk等。

## 三.zookeeper的特性

### 1.顺序一致性：

从一个客户端发起一个请求，最终会严格按照其发起的顺序被应用到zookeeper

### 2.原子性：

所有事务请求的处理结果，在整个集群上的所有机器的应用情况是一致的。

### 3.单一视图：

无论客户端连接到zookeeper集群中的那个服务器，看到的服务端数据都是一致的

### 4.可靠性：

一旦服务端成功的应用了一个事务，并完成了对客户端的相应，那么这个事务所引起的服务端状态的变更，会一直的保留下来，除非有另外一个事务又对他进行了修改

### 5.实时性：

zookeeper保证在一段时间内，客户端最终一定能从服务端读取到最新的服务其状态

## 四.zookeeper的典型应用场景

1.数据发布、订阅（推、拉）

2.负载均衡

3.命名服务：顺序增长的 可以在集群环境下使用的，命名易于理解的ID

4.分布式协调、通知（心跳检测）

### 用到过zookeeper的

hadoop HBase Storm  solr

## 五.zookeeper中基本概念

### 1.zookeeper中的角色

Leader服务器是整个Zookeeper工作机制中的核心

Follower服务器是zookeeper集群状态的跟随者

Observer服务器充当一个观察者

### 2.会话：

会话是指客户端和Zookeeper服务器的链接，zookeeper中的会话叫做Session，客户端靠与服务器建立一个TCP的长连接来维持一个Session，客户端在启动的时候首先会与服务器建立一个TCP链接，通过这个链接,客户端能通过心跳检测与服务器保持有效的会话，也能像Zookeeper服务器发送请求获得相应。

### 3.节点：

Zookeeper中的节点有两类

1.集群中的一台机器称为一个节点

2.数据模型中的数据单元Znode，分为持久节点和临时节点

3.Zookeeper的数据模型就是一棵树，树的节点就是Znode，Znode中可以保存信息。

### 4.版本：

Version   当前数据节点数据内容的版本号

cversion  当前数据节点子节点的版本号

aversion  当前数据节点ACL变更的版本号

         ACL（控制列表）（权限）

### 5.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要有更高的zxid；
5. 集群中大多数的机器得到响应并follow选出的leader。

## 六.悲观锁和乐观锁

### 1.悲观锁

悲观锁又叫悲观并发锁，是数据库中一种非常严格的锁策略，具有强烈的排他性，能够避免不同事物对同一数据并发更新造成的数据不一致，在上一个事物没有完成之前，下一个事物不能访问相同的资源，适合数据更新非常激烈的场景。

### 2.乐观锁

相比悲观锁，乐观锁使用场景会更多，悲观锁认为访问相同数据的时候一定会出现相互干扰，所以简单粗暴的使用排他访问方式，而乐观锁认为不同事物访问相同资源是很少出现相互干扰的情况，因此在事物处理期间不需要进行并发控制，当然，客观锁也是锁，他还是会有并发的控制！对于数据库我们通常的做法是在每个表中增加一个version版本字段，事物修改数据之前先读出数据，当然版本号页顺势取出来，然后把这个读取出来的版本号加入到更新语句的条件中，比如读取出来的版本号是1，我们修改数据的语句可以这样写，update某某表set字段一=某某值，where id=1 and version=1，那如果更新失败了说明，其他事务已经修改过数据了，那系统需要抛出异常给客户端，让客户端自行处理，客户端看可以选择充实。

## 七.watcher （事件监听器）

Zookeeper允许永久在指定节点上注册一些Watcher，数据节点发生变化的时候，Zookeeper服务器会把这个变化通知发生给感兴趣的客户端

ACL是Access Control Lists的简写，Zookeeper采用ACL策略来进行权限控制，有以下权限。

Create：创建子节点的权限

Read:获取节点数据和子节点列表的权限

Write：更新节点数据的权限

Delete：删除子节点的权限

Admin：设置节点ACL的权限

## 八．zookeeper的集群配置

1.安装JDK

2.解压zookeeper安装包

3.新建文件夹 mkdir /data/zookeeper/zookeeper3.5

新建文件夹 mkdir /data/zookeeper/zk\_data

新建文件夹 mkdir /data/zookeeper/zk\_data\_logs

4.移动解压出来的包到/data/zookeeper/zookeeper3.5

5.进入/data/zookeeper/zookeeper-3.5/conf

6.拷贝cp zoo\_sample.cfg ./zoo.cfg

7.修改配置

dataDir=/data/zookeeper/zk\_data #事务日志

dataLogDir=/data/zookeeper/zk\_data\_logs #快照日志

8.配置集群

# 唯一标示和myid文件相同 ip 集群通信端口 集群leader挂掉后的选举端口

server.1=192.168.253.128:2888:3888

server.2=192.168.253.129:2888:3888

server.3=192.168.253.130:2888:3888

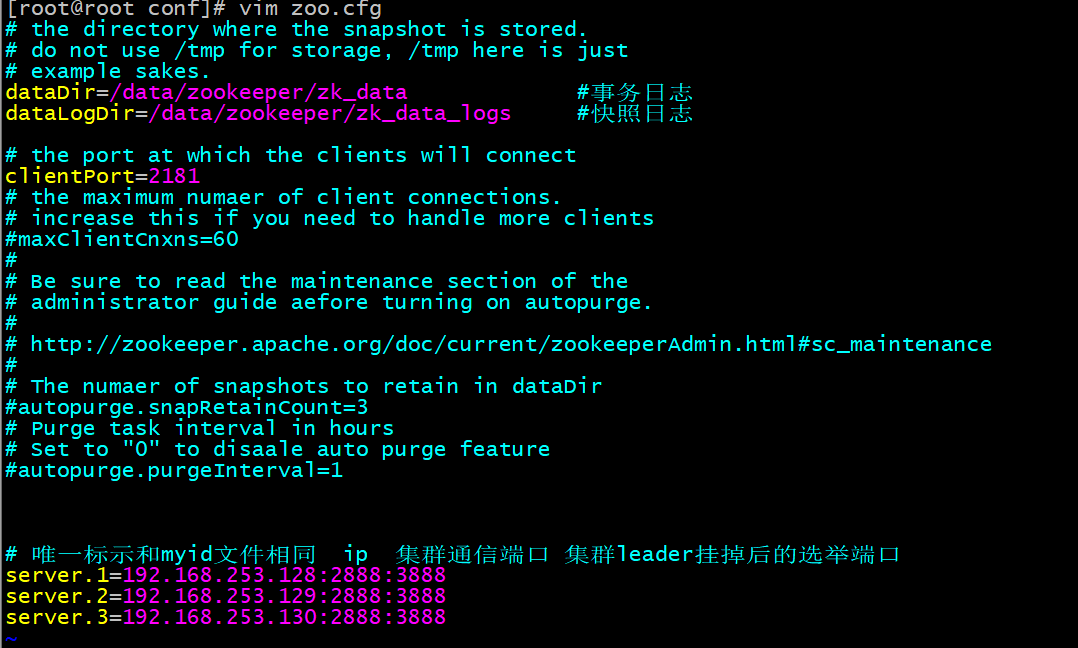
9.创建myid文件

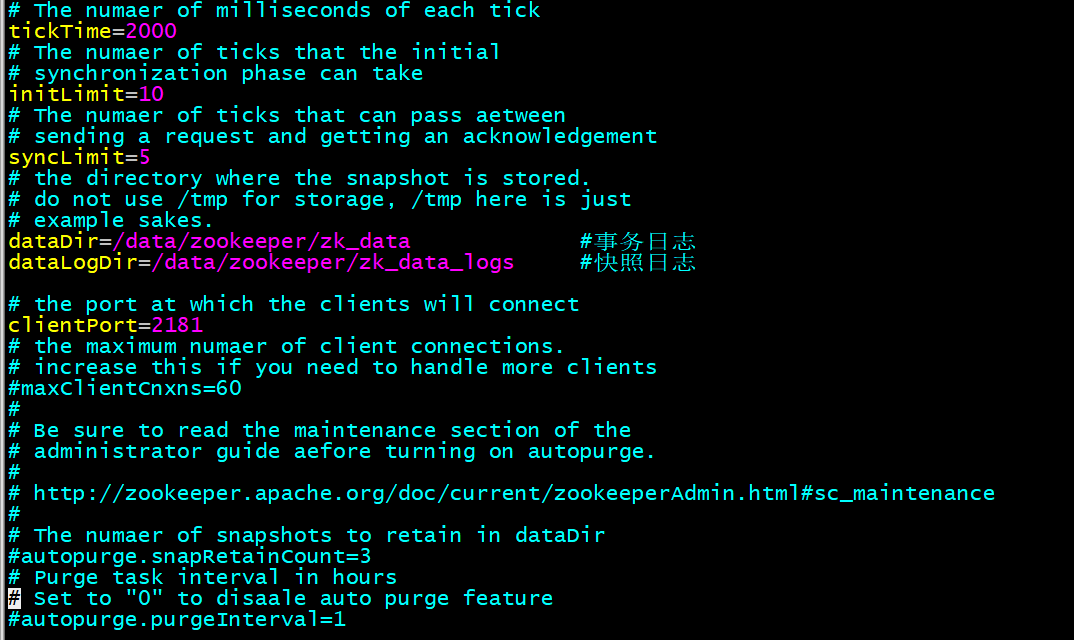
在/data/zookeeper/zk\_data中创建myid文件

在myid文件中写入与配置文件中本IP对于的数字即可

10.配置完成 ./zkServer.sh start 启动

./zkServer.sh status 查看状态





**虚拟化:**是指通过虚拟化技术将一台计算机虚拟为多台逻辑计算机。在一台计算机上同时运行多个逻辑计算机，每个逻辑计算机可运行不同的操作系统，并且[应用程序](http://baike.baidu.com/item/%E5%BA%94%E7%94%A8%E7%A8%8B%E5%BA%8F)都可以在相互独立的空间内运行而互不影响，从而显著提高计算机的工作效率。

**把有限的固定的资源根据不同需求进行重新规划以达到最大利用率的思路，在**[**IT**](http://baike.baidu.com/item/IT)**领域就叫做**[**虚拟化技术**](http://baike.baidu.com/item/%E8%99%9A%E6%8B%9F%E5%8C%96%E6%8A%80%E6%9C%AF)**。**

Hypervisor(虚拟机监视器):是一种**运行在物理服务器和操作系统之间的中间软件层**,可允许多个操作系统和应用共享一套基础物理硬件.Hypervisor是所有虚拟化技术的核心

**云计算:云计算是通过网络提供可伸缩的廉价的分布式计算能力**

云计算是一种商业计算模型，它**将计算任务分布在大量计算机构成的资源池**上，使用户能够按需获取计算力、存储空间和信息服务。

**大数据:**它的特色在于对**海量数据的挖掘**,但它必须依托云计算的分布式处理、分布式数据库、[云存储](http://www.chinacloud.cn/show.aspx?id=757&cid=30)和[虚拟化](http://www.chinacloud.cn/show.aspx?id=2107&cid=12)技术。

**大数据:**指无法在**一定时间范围内用常规软件工具进行捕捉、管理和处理的数据集合**，是需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力的海量、高增长率和多样化的信息资产。

**分布式处理:**就是多台相连的计算机**各自承担同一工作任务的不同部分**，在人的控制下，同时运行，共同**完成同一件工作任务**。

**分布式:**一个业务分拆多个子业务，部署在不同的服务器上

**集群:**同一个业务，部署在多个服务器上

**高并发:**大量请求同时发起

**@author** CGB

@createtime 2018年2月27日22:00:44

@Public email cgb2499404424@163.com

@Location Beijing China

@github https://github.com/chenguangbo