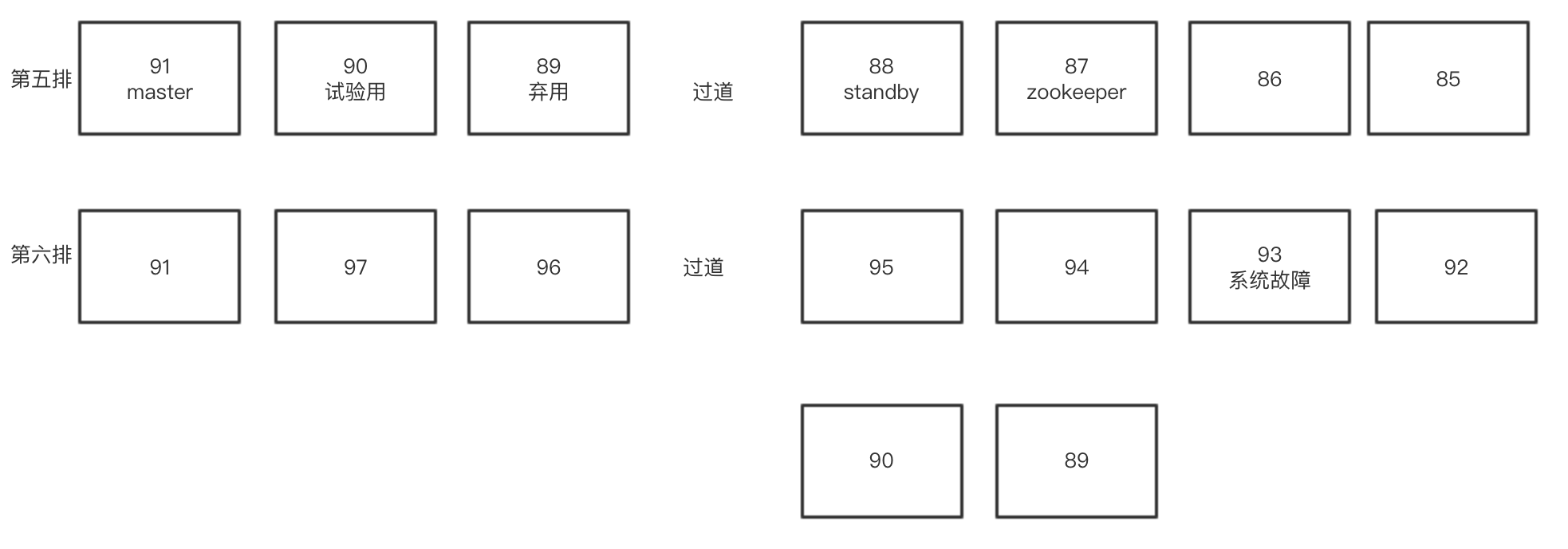
Hadoop配置文档

# 1、现有环境

现有总共14台台式电脑作为实验机，它们的物理位置、IP设置以及集群角色如下：



\*其中所有电脑的IP前缀皆为116.56.136.90

其中：IP尾缀为89的电脑由于不明错误弃用（后查清为IP冲突），IP尾缀为93的电脑系统故障无法正常开机。

# 2、配置版本：

hadoop使用从 <http://archive.cloudera.com/cdh5/cdh/5/hadoop-2.6.0-cdh5.13.3.tar.gz> 下载的5.13.3版本

zookeeper使用的版本为目前最新版

# 3、配置过程：

## 1）设置可远程root登录：

此步骤作用是方便远程批量操作，不是配置过程的必须项。

把每一台电脑的/etc/ssh/sshd\_config的PermitRootLogin设置改为yes，并重启，即可外部使用root账户登录到主机。

## 2）设置无密码登录：

此步骤为必须项。

在master 和standby上面使用“ssh\_keygen -t rsa”命令生成公钥，储存在~/.ssh/ id\_rsa.pub内，使用cat把两个公钥放在同一个authorized\_keys文件内，然后把这个authorized\_keys覆盖到各个电脑的~/.ssh内的同名文件。

我们在配置过程中先把standby的公钥放在master的.ssh内，然后使用脚本rsa\_for\_master.sh即可分发到各个slave。此过程需要多次人工键入yes和密码，考虑数量不多，没有采用自动化输入。

设置完成后，使用check\_ssh.sh检查，如果都可以正常访问，则输出公钥的内容，如有无法无密访问的，则需要手动输入密码，则可以查出哪台无密访问不正常。

## 3）批量改host文件：

此步骤的作用是为了方便集群管理以及实现配置文件可读性，此非必选项但是不应该跳过此步骤。此步骤我们使用脚本批量远程登录修改，需要步骤一作为基础。我们使用的脚本文件为send\_host.sh 。具体的步骤为先使用scp把本机的hosts文件发送到目标电脑，然后使用cat覆盖原电脑的/etc/hosts文件，然后重启网络服务。

## 4）其它准备：

此步骤包括apt-get update 和apt-get upgrade、安装java、检查系统等其它操作。我们使用在此过程中发现实验室内有其它电脑与实验电脑IP冲突。

## 5）发送Hadoop安装包以及相关环境修改

我们把hadoop的安装包以及相关的环境变量配置文件如.bash\_profile打包，使用脚本批量SCP到目标电脑并解压