

大数据实验总结

原创

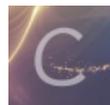
Volcano3511 于 2017-11-30 10:54:44 发布 19113 收藏 5

分类专栏: [课程学习 大数据实验](#) 文章标签: [大数据 实验报告 总述](#)

版权声明: 本文为博主原创文章, 遵循 [CC 4.0 BY-SA](#) 版权协议, 转载请附上原文出处链接和本声明。

本文链接: <https://blog.csdn.net/Volcano3511/article/details/78673586>

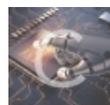
版权



[课程学习](#) 同时被 2 个专栏收录

18 篇文章 0 订阅

订阅专栏



[大数据实验](#)

7 篇文章 0 订阅

订阅专栏

大数据实验总结

[大数据实验总结](#)

[实验总要求](#)

[基础环境](#)

[心得体会](#)

实验总要求

给五台主机, 先后配置torque, mpich, mongodb,hadoop,zookeeper,hbased

torque

有主有从

master: lk

slave: slave2~5

mpich

各台机器是平等的

mongodb

需要分片, 副本集, 还有各种角色的配置

lk	slave2	slave3	slave4	slave5
mongos	mongos	mongos		
config server	config server	config server	shard1	shard1
shard1	shard2	shard2	shard2	shard3

lk	slave2	slave3	slave4	slave5
shard3	shard3	shard4	shard4	shard4

一台机器可以有不同的角色，但是要给不同角色配置不同的端口

副本集	端口
config server	21000
mongos	20000
shard1	27001
shard2	27002
shard3	27003
shard4	27004

-hadoop, zookeeper

lk	slave2	slave3	slave4	slave5
namenode	y	y	n	n
datanode	n	n	y	y
resourcemanager	y	y	n	n
journalnode	y	y	y	y
zookeeper	y	y	y	n

zookeeper只在前面三台机器上安装，安装结果应该是一个leader和两个follower

- hbase
 master: lk
 backup-master: slave2
 regionservers: slave3~5

基础环境

基础环境就是所有软件安装中都需要

免密码登陆
方便交互

关闭防护墙
防止各台机器在交互时候通信受到影响

`/etc/hosts` 和 `/etc/hostname`

给各个机器取别名，要一致

心得体会

整个环境其实搭了很久，属于边摔跤边学会走路的那种，参考了很多网上的资料，可以说是既是帮了很大的忙，又是挖了很大的坑。

因为以前没有接触过这些，所以是网上的教程把我带上路，但是因为每个人的机器什么的都是不同的，要求也不一样，所以单纯的粘贴复制肯定是不行的，要小心的把每个点都看清楚然后再进行配置，比如，我在配置MongoDB的时候，有一个地方IP忘了改成自己机器的IP，导致浪费了很多时间。

出现问题不可怕，多去网上看看别人的解决方法就可以了

整个集群的环境，有一些关键词吧：交互通信，主从控制等，我感觉还是在配置过程中感受到了这方面的思想。

torque和MongoDB是比较难搭的。torque是一个搭的，一是刚刚开始做，所以比较不熟悉，二是torque主从的搭建不一样，需要理清哪些在master上面搭建，哪些在slave上面搭建；MongoDB的难点在于组件多和配置多，config server，shard分片和mongos的配置，比较繁琐。

hadoop中等难度吧，其他的mpich，hbase都不是很难

我这次实验的最大感受是要学会看日志，日志真的很有用，只有知道哪里错了才能对症下药

还有，要学会积累工具。在hadoop搭建的时候，看到网上基本都用到了 `deploy.conf`，`deploy.sh` 和 `runRemoteCmd.sh` 这三个文件，通过这三个文件，可以很方便的在一台机器上实现对多台机器的操作（前提是免密登陆和权限足够），就不需要在集群搭建的时候经常进行切换，对每个机器做重复的操作。