实验2：多机开发环境部署

1. 实验目的

掌握多机Hadoop分布式运行环境

1. 实验步骤
2. 操作系统和安装软件

方案一：曙光Ubuntu虚拟机环境，两人或者以上组合

方案二：多PC机上Linux虚拟机组合，两人或者以上组合

方案三：单PC机，多虚拟机组合

1. 配置hosts 以及 ssh 免密登录
2. 修改 /etc/hosts 增加其他主机的hostname
3. 使用 ssh-copy-id 复制公钥到所有节点
4. 用ssh 登录IP地址以及hostname，确保都可以免密登录
5. 在.bashrc中增加环境变量 HADOOP\_HOME = /home/xxxx/opt/hadoop目录
6. 配置HDFS以及Yarn
7. 配置文件
8. etc/hadoop/hadoop-env.sh：

修改 JAVA\_HOME 为 openjdk 的路径

1. etc/hadoop/hdfs-site.xml：

修改数据块复制属性，修改为2，默认是3

修改备份NameNode地址，修改为主 NameNode 地址

1. etc/hadoop/core-site.xml

修改HDFS NameNode的地址

修改DataNode存储数据地址

1. etc/hadoop/mapred-site.xml：

修改Mapreduce使用Yarn作为调度

1. etc/hadoop/yarn-site.xml

修改ResouceManager地址，可以用IP也可以用hostname

修改NodeManager，增加shuffle辅助服务

1. 启动分布式的HDFS
2. 启动分布式的Yarn
3. 测试与验证



