# Hadoop部署实践1

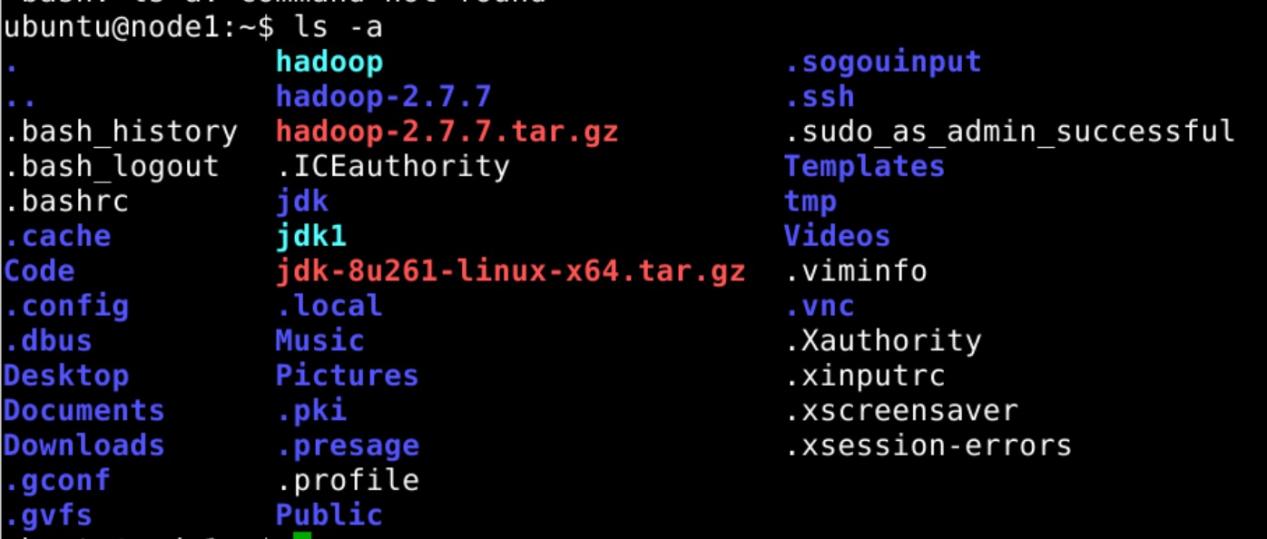
f21010726 苏一桐

## 搭建Hadoop伪分布式模式环境

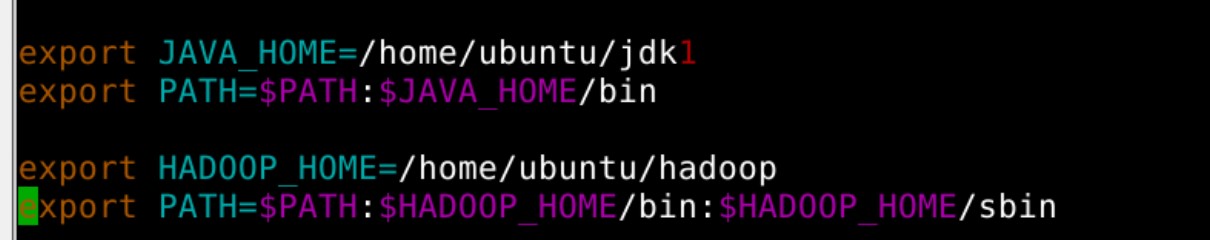
1. 安装对应JDK8、Hadoop-2.7.7以及ssh\_server

|  |
| --- |
| wget “http://10.90.3.2/HUP/Hadoop/jdk-8u261-linux-x64.tsr.gz”  解压：tar -xvf jdk-8u261-linux-x64.tsr.gz |
| wget “http://10.90.3.2/HUP/Hadoop/hadoop-2.7.7.tar.gz  解压：tar -xvf hadoop-2.7.7.tar.gz -C ~ |
| sudo apt install openssh-client openssh-server |

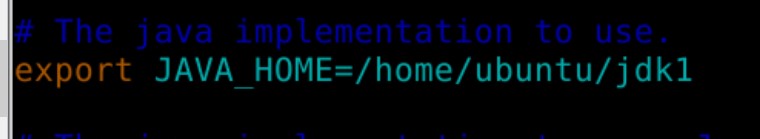
设置软连接：



1. 配置环境变量，source .bashrc激活变量



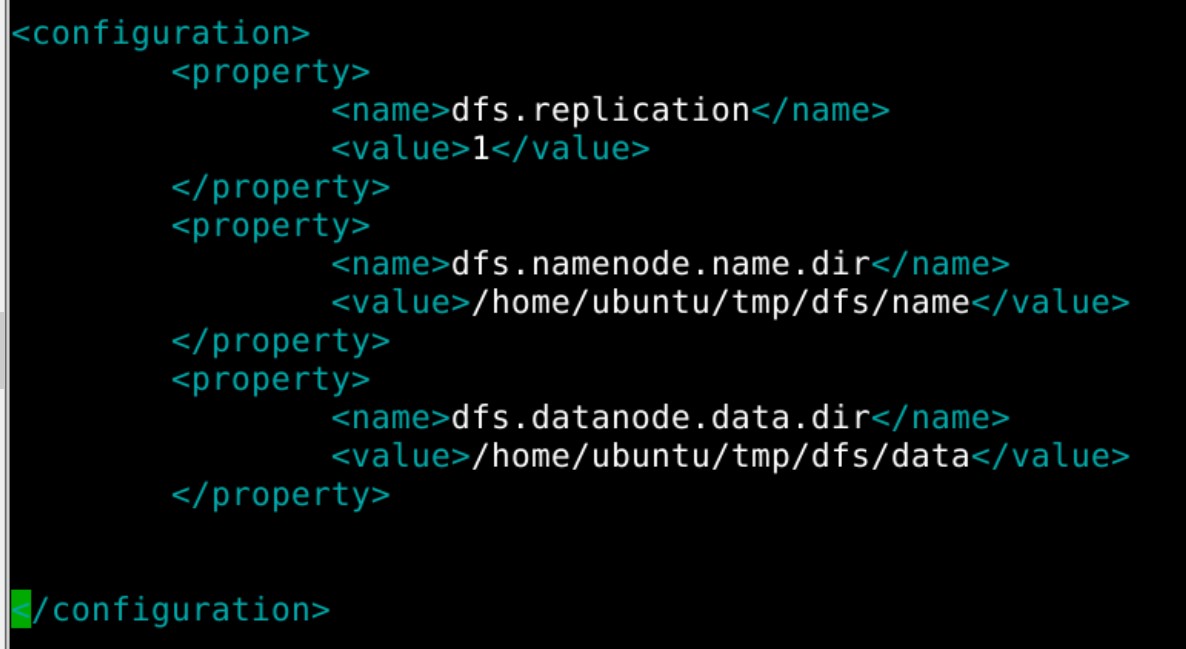
1. 配置hadoop-env.sh



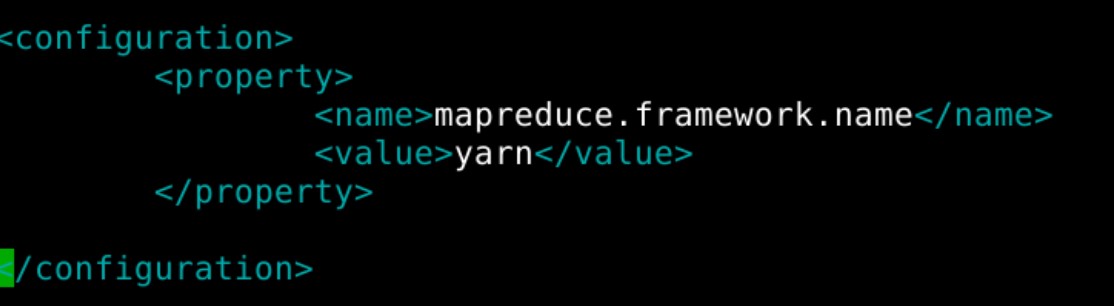
1. 配置core-site.xml



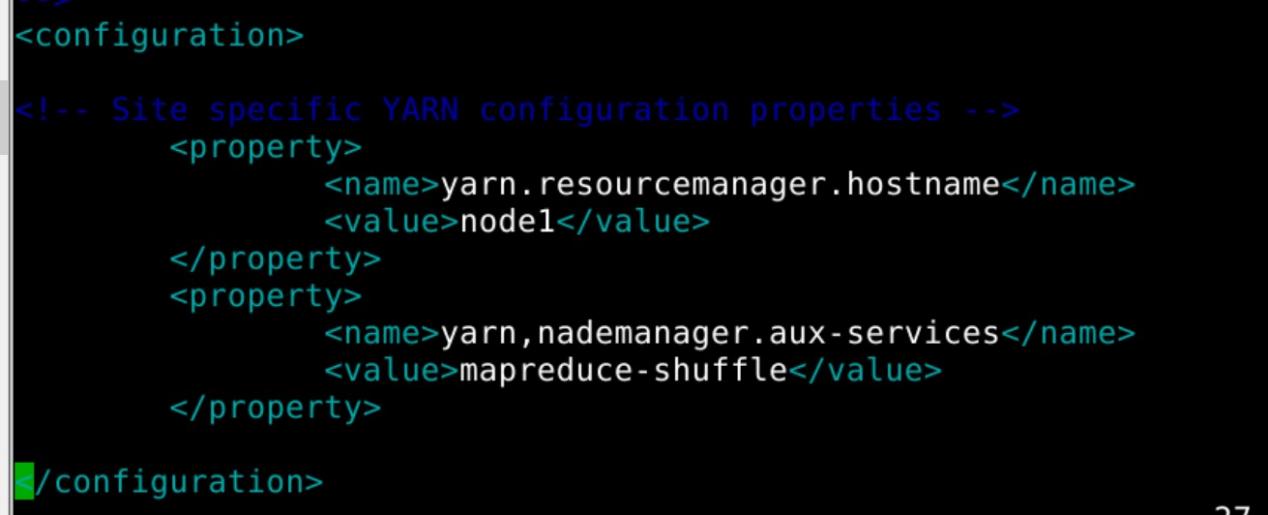
1. hdfs配置



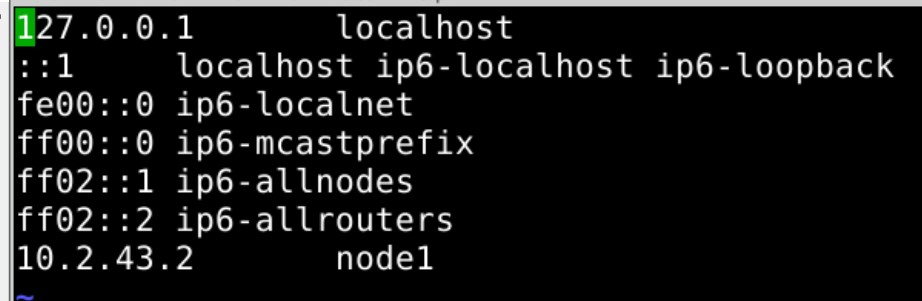
1. mapred配置



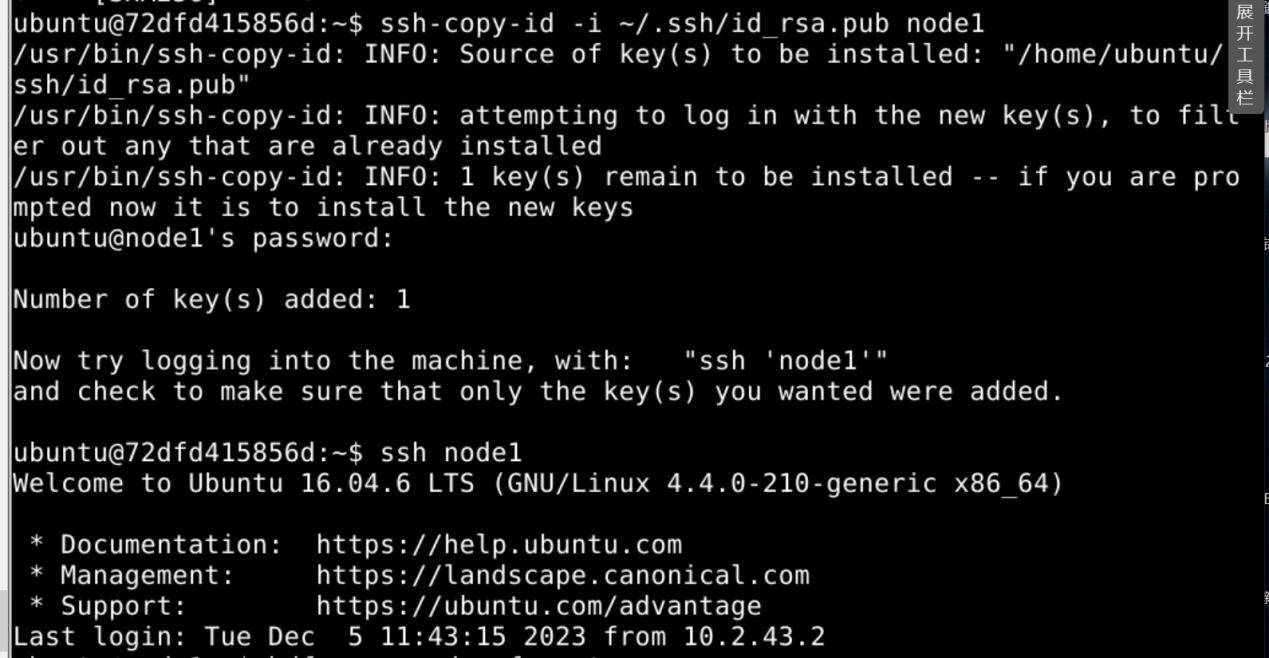
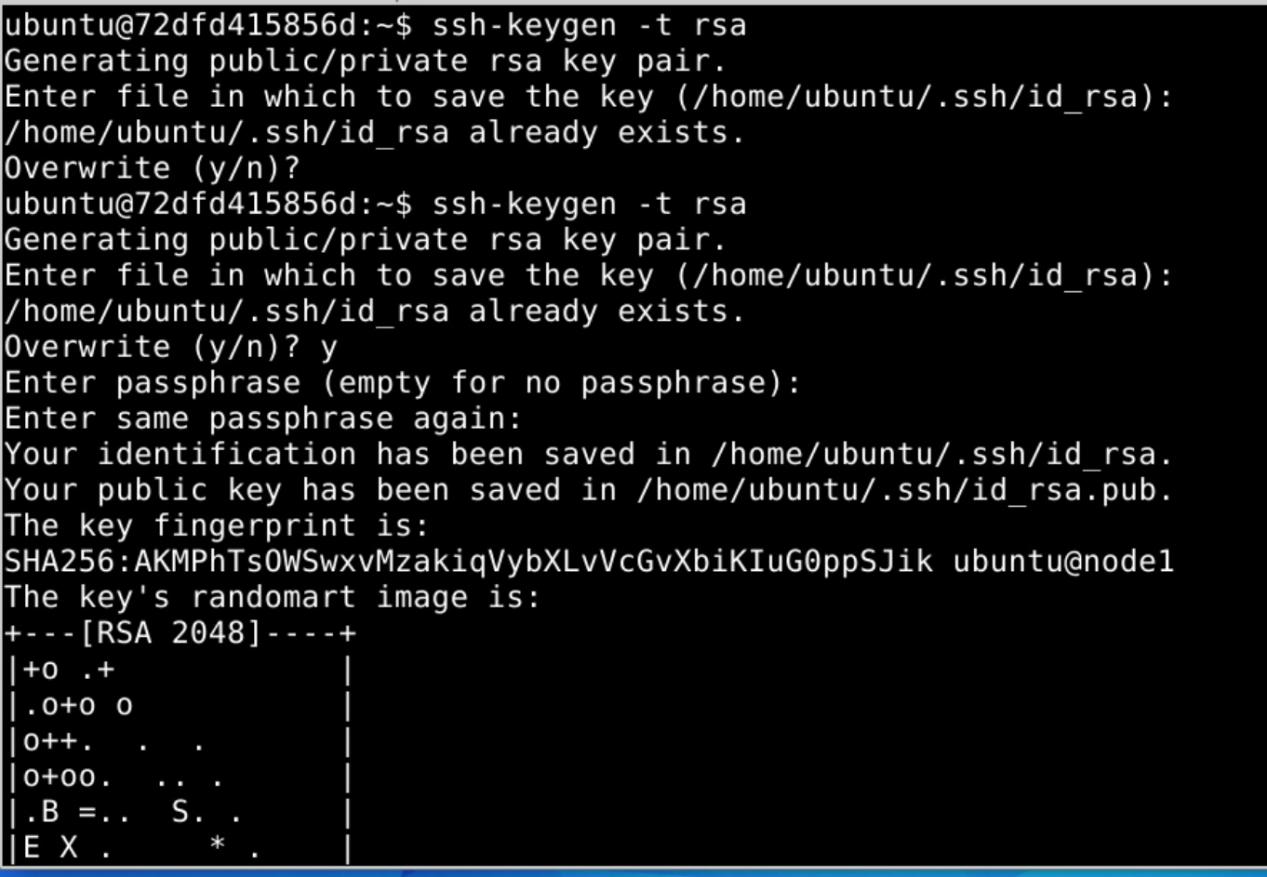
1. yarn配置



1. 修改主机名



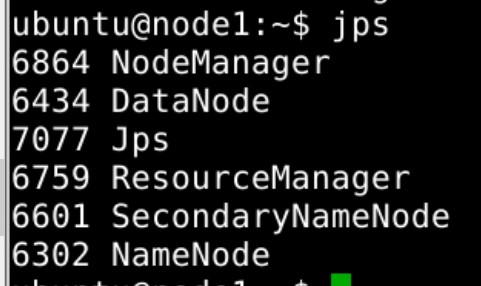
1. 设置免密登录并拷贝公钥到主机



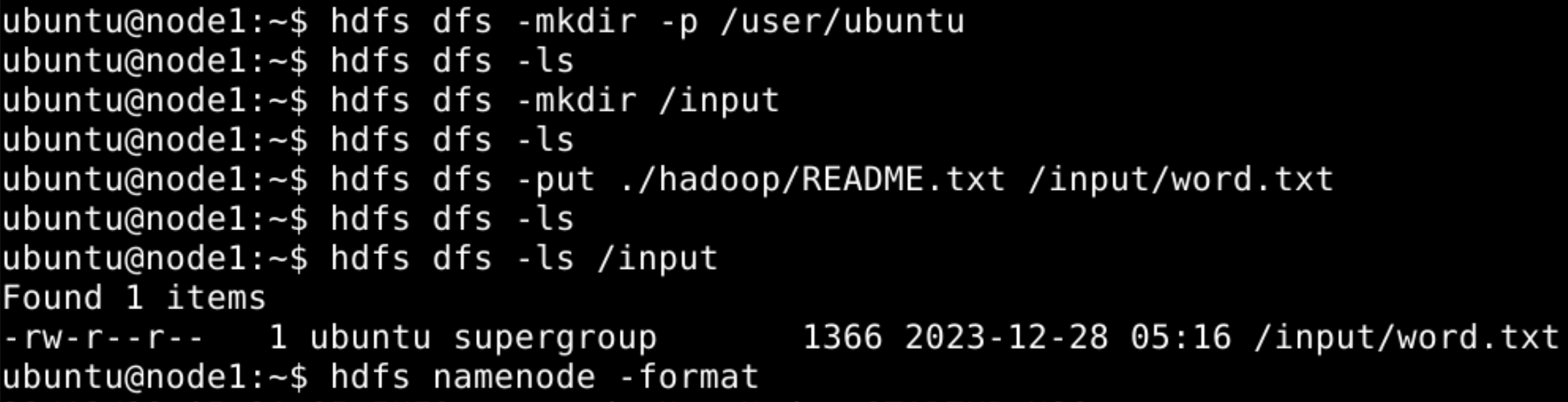
1. 格式化hdfs并启动，jps查看

hdfs namenode -format

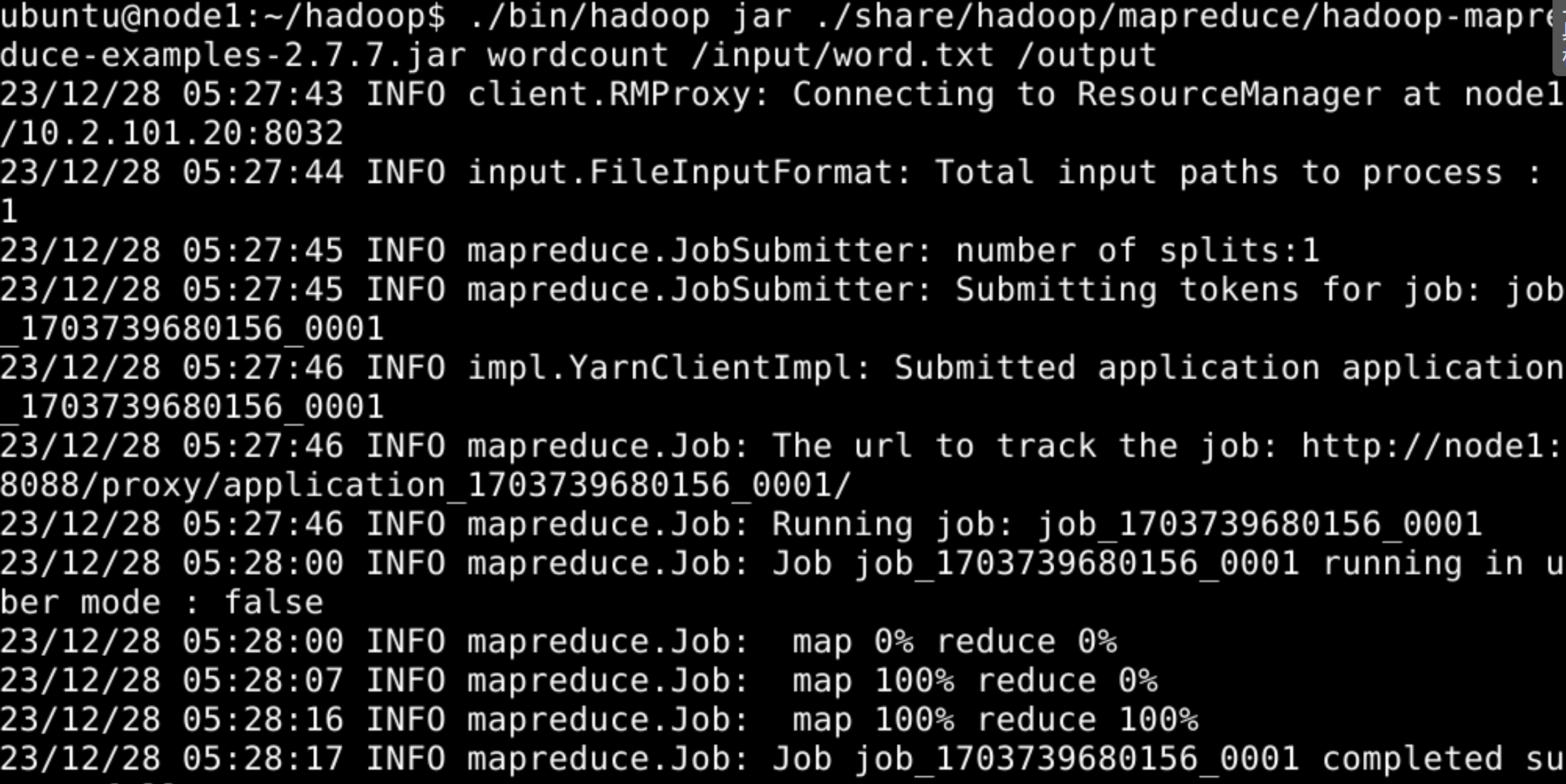
start-all.sh



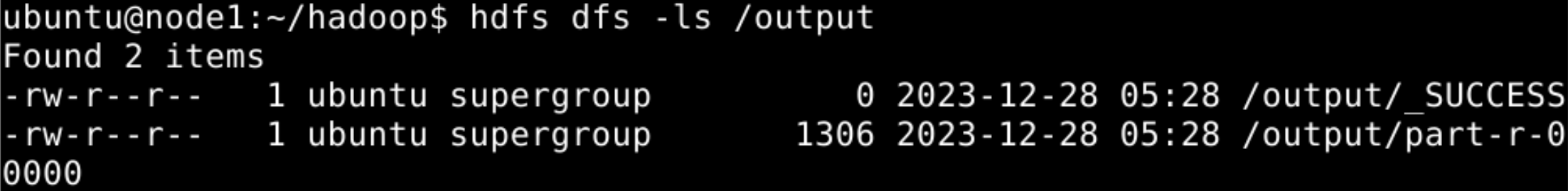
1. 在HDFS上创建文件夹并上传文件



1. 执行wordcount命令



1. 查看输出结果





## 实验总结

本次实验实操了Hadoop环境的部署，了解其分布式存储和计算框架。完成了HDFS和MapReduce的变量配置，明白了它们是如何协同工作以处理大规模数据的。

过程中也不少出现报错情况。如配置环境变量时由于变量符号使用不规范、路径位置错误等问题导致Hadoop无法启动。其次是设置免密登录，学会了获取公钥实现节点间的免密通信，但遇到设置后需要重启生效配置，否则无法直接登陆（sudo service ssh restart使用此命令重启ssh，配置生效）。此外，若出现重复启动时报错，可将tmp文件夹中的缓存文件全部删除，再次启动。

接下来也将逐渐接触多机配置，有关Hadoop生态中其他组件的功能使用，进一步加深对大数据技术应用技术的理解、操作。