

南京信息工程大学 实验（实习）报告

实验（实习）名称 实验三：Hadoop 单机模式环境搭建 （实习）日期 2023.3.23 得分
指导教师 孙乐 专业 软件工程（中外合办） 年级 2020 级 班次 2 班
姓名 颜晓雨 学号 202083020070

实验目的：通过部署 hadoop 本地运行模式，初步认识 hadoop 的使用。

实验内容：

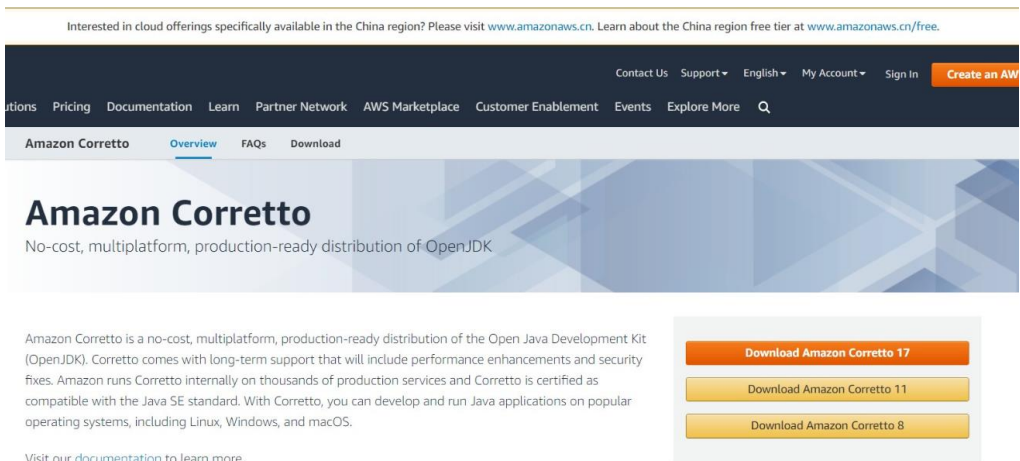
- 1. jdk 和 hadoop 的安装以及环境配置
- 2. 测试 MapReduce 案例

实验环境：Linux 系统、VMware

实验步骤：

1.jdk 安装以及环境配置

1) 下载



今天 (1)

amazon-corretto-8.322.06.1-window... 2022/3/1 10:27 WinRAR ZIP 压缩... 119,379 KB

2) 解压

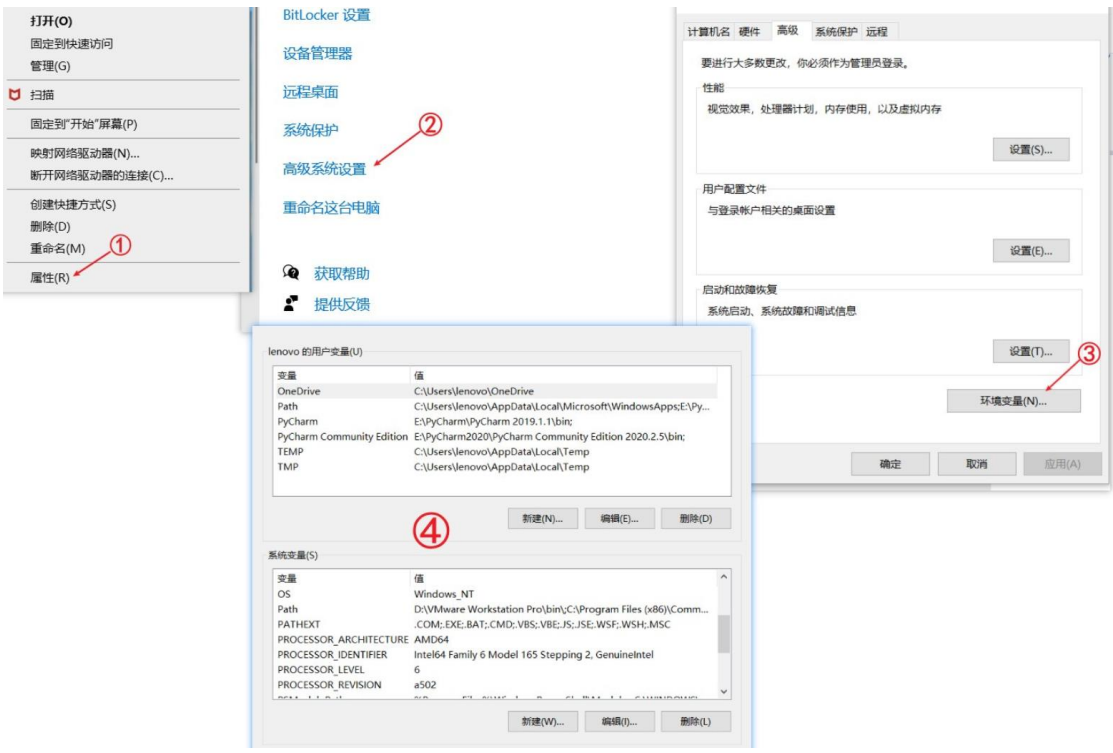
名称	修改日期	类型	大小
bin	2022/1/14 0:07	文件夹	
demo	2022/1/14 0:07	文件夹	
include	2022/1/14 0:07	文件夹	
jre	2022/1/14 0:07	文件夹	
lib	2022/1/14 0:07	文件夹	
sample	2022/1/14 0:07	文件夹	
ASSEMBLY_EXCEPTION	2022/1/14 0:05	文件	2 KB
javafx-src.zip	2022/1/14 0:05	WinRAR ZIP 压缩...	6,384 KB
LICENSE	2022/1/14 0:05	文件	19 KB
release	2022/1/14 0:05	文件	1 KB
src.zip	2022/1/14 0:05	WinRAR ZIP 压缩...	50,715 KB
THIRD_PARTY_README	2022/1/14 0:05	文件	154 KB
version.txt	2022/1/14 0:05	文本文档	1 KB

3) JDK 目录介绍

目录	描述
bin	包含一些用于开发Java程序的工具，例如：编译工具(javac.exe)、运行工具 (java.exe)、打包工具 (jar.exe)等。
include	C语言的头文件，用于支持Java程序设计。
jre	Java运行时环境的根目录。
jre\bin	包含Java平台所使用工具和类库的可执行文件和DLL文件。
jre\lib	Java运行时环境所使用的核心类库。
lib	包含开发Java程序所使用的类库文件。
demo	包含一些用于演示的示例程序。
sample	包含一些示例程序。

4) JDK 环境配置

打开环境变量：



配置 JAVA_HOME

配置 JAVA_HOME 的表单，包含变量名和变量值的输入框，以及确定和取消按钮。

变量名(N):	JAVA_HOME		
变量值(V):	D:\Java\jdk1.8.0_271		
浏览目录(D)...	浏览文件(F)...	确定	取消

配置 path

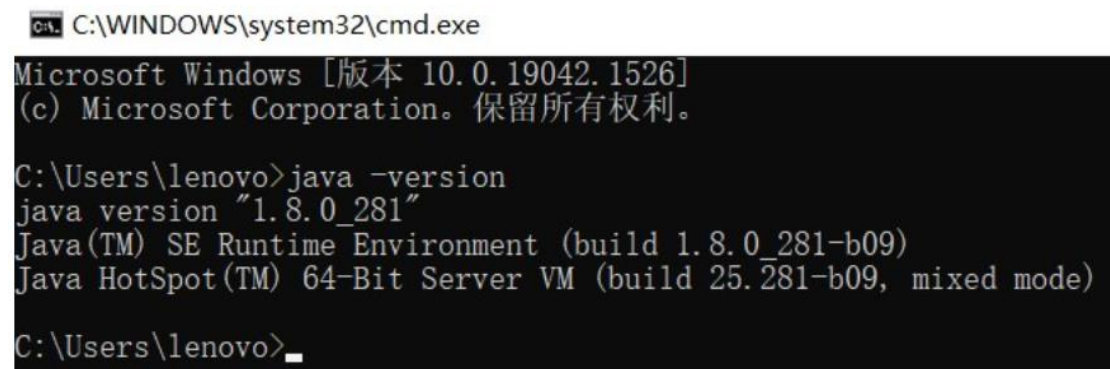
测试

用 dos 窗口检测 (win+R):



输入命令: java-version 回车

配置成功:



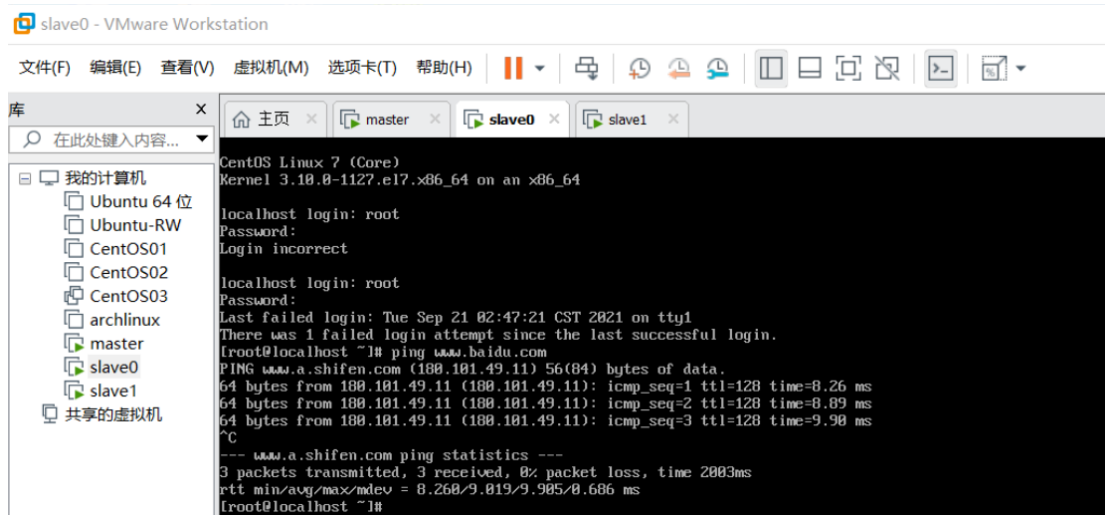
2.hadoop 安装以及环境配置

- 1) 新建虚拟机
- 2) 配置静态网络, 关闭防火墙

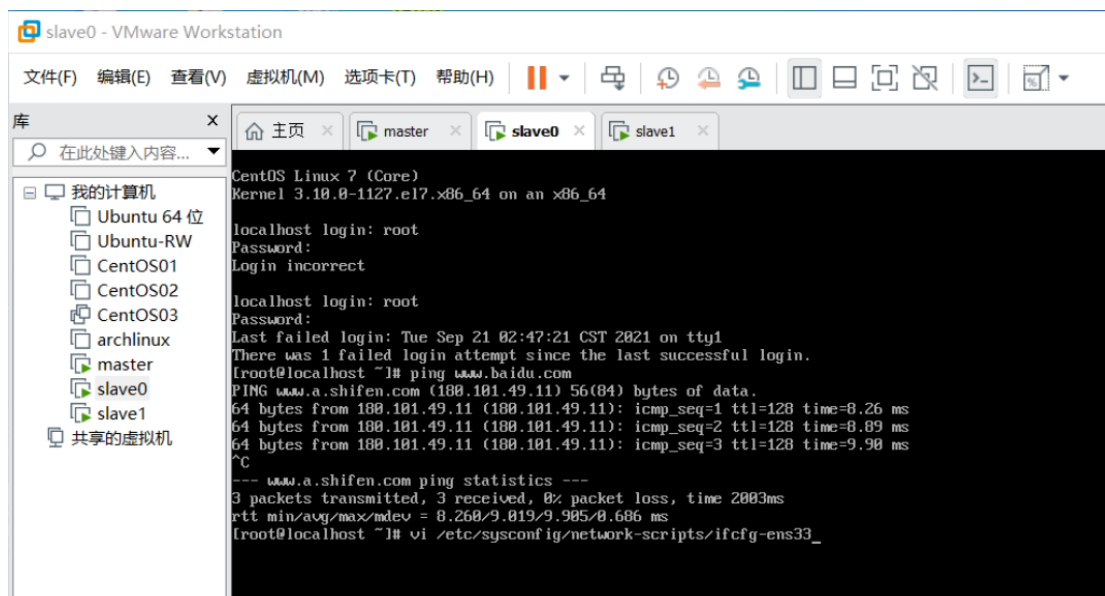
进入到 hadoop

在其中输入 root 回车, 密码回车, 登陆成功后输入命令 `ping www.baidu.com`

查看网络是否能 ping 通, Ctrl+C 停止



输入命令 vi/etc/sysconfig/network-scripts/ifcfg-ens33



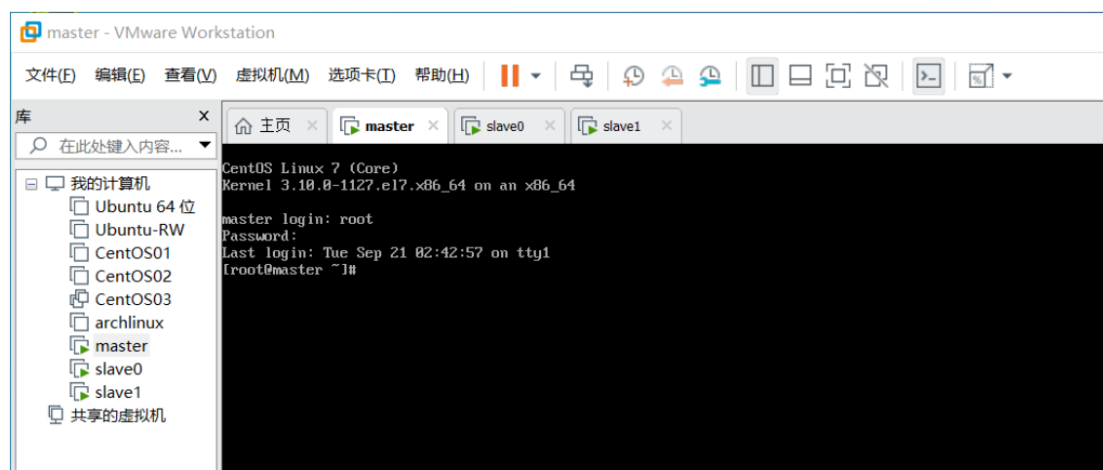
3) 配置主机名，配置 hosts 文件

在 master 输入命令 hostnamectl set-hostname master

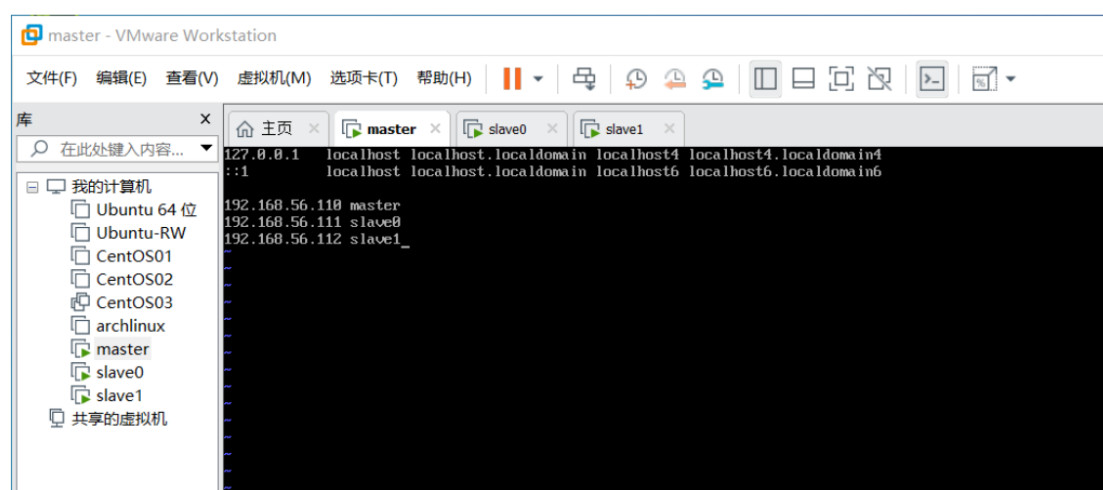
在 slave0 中输入命令 hostnamectl set-hostname slave0

在 slave1 中输入命令 hostnamectl set-hostname slave1

最后都输入 reboot 重启，可以发现登录名已经更改过来了



如图



4) 重要目录

- (1) bin目录：存放对Hadoop相关服务（HDFS,YARN）进行操作的脚本
- (2) etc目录：Hadoop的配置文件目录，存放Hadoop的配置文件
- (3) lib目录：存放Hadoop的本地库（对数据进行压缩解压缩功能）
- (4) sbin目录：存放启动或停止Hadoop相关服务的脚本
- (5) share目录：存放Hadoop的依赖jar包、文档、和官方案例

3. 测试 MapReduce 案例

切换路径到 `/opt/module/hadoop-3.1.3/` 目录下，在该目录下创建一个文件 `data.txt`:

```
good better best
never let it rest
till good is better
and better best
```

```
good good study
day day up
today is a good day
```

```
[root@localhost hadoop-3.1.3]# vim data.txt
```

```
[root@localhost hadoop-3.1.3]# cat data.txt
```

```
good better best
never let it rest
till good is better
and better best
```

```
good good study
day day up
today is a good day
```

使用 `hadoop jar` 命令来运行一个 jar 包，在 jar 包中可能存在多个可执行的类，目前需要运行的是其中的 `wordcount`；运行时需要两个参数：

- 输入路径：数据源文件或目录
- 输出路径：一个还不存在的目录，会自动创建
查看生产的 output 目录：

cd 进入 output 文件：

```
public class NumberSortMapper extends Mapper<LongWritable, Text, ComparableKey> {

    @Override
    protected void map(LongWritable key, Text value, Context context) throws IOException, InterruptedException {
        String[] strings = value.toString().split(",");
        long firstNum = Long.parseLong(strings[0]);
        long secondNum = Long.parseLong(strings[1]);
        ComparableKey comparableKey = new ComparableKey(firstNum, secondNum);
        context.write(comparableKey, NullWritable.get());
    }
}
```

实验时间：2 学时