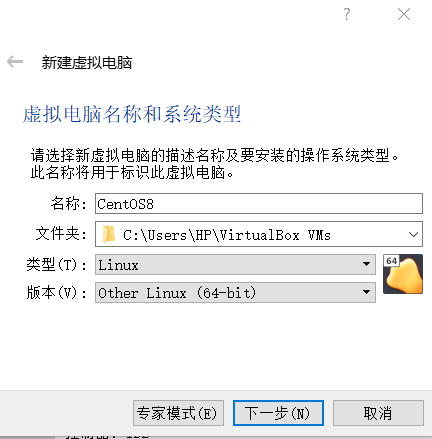
搭建Hadoop伪分布式大数据环境

1、安装VirtualBox虚拟机

环境：系统：win10内存：16 GB

2、在虚拟机环境中安装Linux操作系统

1.左上角“新建”，设置名称为“CentOS8“，选择类型为Linux，版本为Other Linux(64-bit)



2.设置内存大小8024MB，选择现在创建虚拟硬盘，硬盘类型为vdi，存放方式为动态分配，硬盘大小选40G。创建完成。

3、在Linux系统中安装JDK并完成环境变量配置

1.下载jdk-8u231-linux-x64.tar.gz，Linux系统使用tar zxvf命令解压。

2.使用“vi .bash\_profile “配置环境变量，配置如下：

JAVA\_HOME=/usr/jdk1.8.0\_231

PATH=$JAVA\_HOME/bin:$PATH

export JAVA\_HOME

export PATH

3.使用“source .bash\_profile “命令使配置生效。

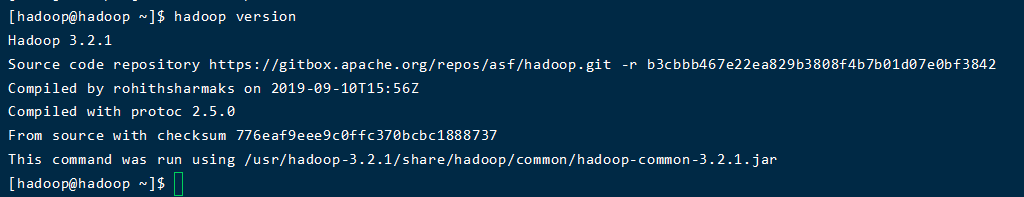
1. 安装Hadoop，实现伪分布式配置

1.环境：系统：centos8，JDK：jdk8，hadoop：3.2.1

2.使用vi /etc/hostname，vi /etc/hosts命令设置主机名与IP地址。同时关闭防火墙，创建用户，命名为Hadoop。

3.Hadoop的tar包安装，使用tar zxvf hadoop-3.2.1.tar.gz命令解压。

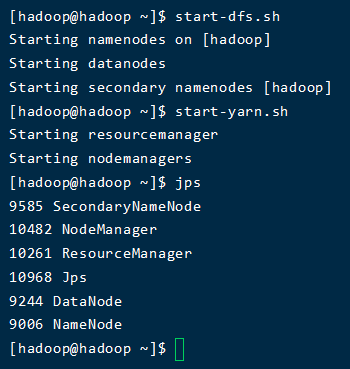
4.配置环境变量：/.bash\_profile，hadoop-env.sh。测试安装是否完成，如下图所示即完成.



5.Hadoop核心配置：配置core-site.xml，hdfs-site.xml，mapred-site.xml，yarn-site.xml，workers。

6.SSH免密配置，生成公钥与私钥

7.启动集群（hadoop），首次启动hadoop时，必须namenode格式化。之后的命令及结果如图所示：



1. 执行wordcount程序，统计指定目录下的词频

1.建立data文件夹，创建数据文件data.txt,内容为：Hello world!

2.上传数据文件至dfs文件系统，执行WordCount程序，命令及结果如图所示：