### **安装 Java 8**

安装 Java 8  
sudo apt-get install oracle-java8-installer

验证 java 是否安装成功  
java -version

### **配置 SSH** ssh-keygen -t rsa -P '' -f ~/.ssh/id\_rsa cat ~/.ssh/id\_rsa.pub>>~/.ssh/authorized\_keys chmod 0600 ~/.ssh/authorized\_keys ssh localhost

### **安装 Hadoop**

wget <https://mirrors.tuna.tsinghua.edu.cn/apache/hadoop/common/hadoop-3.3.0/hadoop-3.3.0-src.tar.gz>

**解压tar包**  
tar -xzf hadoop-3.3.0-src.tar.gz -C /opt/

### **Hadoop 配置**

**配置系统环境变量**  
打开.bashrc文件  
nano ~/.bashrc

在该文件末尾假如下面环境变量：

export HADOOP\_PREFIX="/opt/hadoop-3.3.0"

export PATH=$PATH:$HADOOP\_PREFIX/bin

export PATH=$PATH:$HADOOP\_PREFIX/sbin

export HADOOP\_MAPRED\_HOME=${HADOOP\_PREFIX}

export HADOOP\_COMMON\_HOME=${HADOOP\_PREFIX}

export HADOOP\_HDFS\_HOME=${HADOOP\_PREFIX}

export YARN\_HOME=${HADOOP\_PREFIX}

使环境变量生效：  
Source ~/.bashrc

**配置 hadoop-env.sh**  
在 $HADOOP\_HOME/etc/hadoop 目录下，编辑 hadoop-env.sh，并配置JAVA\_HOME路径。  
export JAVA\_HOME=/usr/lib/jvm/java-8-oracle/

**配置 core-site.xml**  
在 $HADOOP\_HOME/etc/hadoop 目录下，打开 core-site.xml 文件，新增下面配置：

<property>

<name>fs.defaultFS</name>

<value>hdfs://localhost:9000</value>

</property>

<property>

<name>hadoop.tmp.dir</name>

<value>/home/dataflair/hdata</value>

</property>

**配置 hdfs-site.xml**  
在 $HADOOP\_HOME/etc/hadoop 目录下，打开 hdfs-site.xml，新增下面配置：

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

**配置 mapred-site.xml**  
在 $HADOOP\_HOME/etc/hadoop 目录下，如果找不到 mapred-site.xml，可以从 mapred-site.xml.template 文件复制一份，并把文件名修改为 mapred-site.xml  
cp mapred-site.xml.template mapred-site.xml

在 mapred-site.xml 增加下面配置：

<property>

<name>mapreduce.framework.name</name>

<value>yarn</value>

</property>

**配置 yarn-site.xml**  
在$HADOOP\_HOME/etc/hadoop目录下，打开yarn-site.xml，新增下面配置：

<property>

<name>yarn.nodemanager.aux-services</name>

<value>mapreduce\_shuffle</value>

</property>

<property>

<name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>

<value>org.apache.hadoop.mapred.ShuffleHandler</value>

</property>

### **启动 Hadoop**

**格式化 HDFS**  
启动 hadoop 之前先对 hdfs 做格式化操作，把 hdfs 里面的文件全部清空。  
hdfs namenode -format

**启动HDFS**  
start-dfs.sh

**启动 Yarn**  
start-yarn.sh

**jps命令查看启动的进程**

$jps

2961 ResourceManager

2482 DataNode

3077 NodeManager

2366 NameNode

2686 SecondaryNameNode

3199 Jps