操作系统版本为：Ubuntu 18.04.1 LTS

使用的IDE: IntelliJ IDEA 2018.2.6 linux版

1.java环境安装

在命令行中输入命令：sudo apt get install openjdk-8-jdk，安装openjdk-8

1. 安装ssh,rsync

linux环境下可直接使用命令行安装：

$ sudo apt-get install ssh

$ sudo apt-get install rsync

1. 安装hadoop-3.1.1
2. 官网下载hadoop-3.1.1.tar.gz(官网下载地址为：https://hadoop.apache.org/releases.html)
3. 将文件解压到指定目录：tar -zxvf hadoop-3.1.1.tar.gz 安装目录
4. 修改hadoop目录中的hadoop-env.sh文件，添加Java环境变量：
5. export JAVA\_HOME=/usr/lib/jvm/java-8-openjdk-amd64（在这里由于直接安装的openjdk-8,因此目录为/usr/lib/jvm/java-8-openjdk-amd64）
6. hadoop添加到环境变量，然后更新一下环境变量。修改/etc/profile文件，添加以下内容：
7. export HADOOP\_HOME=hadoop安装目录(这里我的安装目录为：/hadoop/hadoop-3.1.1)
8. export PATH=$PATH:$HADOOP\_HOME/bin
9. 可以利用source /etc/profile更新环境变量
10. 安装spark
11. 官网下载spark,选择安装包种类为：Pre-built for Apache Hadoop 2.7 and later(已经安装hadoop后选择的版本)，下载压缩包：spark-2.4.0-bin-hadoop2.7.tgz（官网地址为:https://spark.apache.org/downloads.html）
12. 将压缩文件解压到指定目录
13. 将文件夹下conf目录下的spark-env.sh.template文件修改为spark-env.sh：

cp conf/spark-env.sh.template conf/spark-env.sh

1. 修改spark-env.sh文件内容，添加java，hadoop等环境变量：

export JAVA\_HOME=/usr/lib/jvm/java-8-openjdk-amd64(java安装目录)

export HADOOP\_HOME=/hadoop/hadoop-3.1.1(hadoop安装目录)

export HADOOP\_CONF\_DIR=/hadoop/hadoop-3.1.1/etc/hadoop(hadoop安装目录/etc/hadoop)

1. 配置环境变量：修改/etc/profile文件，添加：

export SPARK\_HOME=/spark/spark-2.4.0

export PATH=$PATH:$HADOOP\_HOME/bin:$SPARK\_HOME/bin

1. 安装maven
2. 下载maven安装包：apache-maven-3.6.0-bin.tar.gz

（官网下载地址为：http://maven.apache.org/download.cgi）

1. 解压到指定目录
2. 添加环境变量：

export M2\_HOME=/apache-maven-3.6.0

export PATH=$PATH:$HADOOP\_HOME/bin:$SPARK\_HOME/bin:$SCALA\_HOME/bin:$M2\_HOME/bin