**作业4**

**Spark环境搭建及程序编译运行**

**一、在Linux系统中安装Spark**

假设之前在Linux系统中已经按照伪分布式模式在机器上安装了Hadoop，请在此基础上，按照Local模式安装Spark，并做相关配置，使得Spark可以顺利访问HDFS。

**二、在spark-shell中运行代码**

请使用Local模式启动进入spark-shell环境，并书写几条Scala语句进行测试，然后退出spark-shell环境。

**三、编写Spark独立应用程序**

请编写一个简单的Scala代码文件（代码中需要包含println语句），然后，分别使用Maven和sbt工具对程序进行编译打包，最后，使用spark-submit命令运行程序。

**四、实验报告**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 实验报告 | | | | |
| 题目 | Spark环境搭建及程序编译运行 | 姓名 | 吴雨娟 | 日期：  2021.10.7 |
| 实验环境：VMware workstation 16 pro，Ubuntu-20.04操作系统，Hadoop-3.1.3，Spark-3.1.2,sbt-1.5.5,maven-3.8.3 | | | | |
| 实验内容与完成情况：  **一、在Linux系统中安装Spark**  假设之前在Linux系统中已经按照伪分布式模式在机器上安装了Hadoop，请在此基础上，按照Local模式安装Spark，并做相关配置，使得Spark可以顺利访问HDFS。  1. 按照Local模式安装Spark，并做相关配置。          2. Spark可以顺利访问HDFS。    **二、在spark-shell中运行代码**  请使用Local模式启动进入spark-shell环境，并书写几条Scala语句进行测试，然后退出spark-shell环境。          **三、编写Spark独立应用程序**  请编写一个简单的Scala代码文件（代码中需要包含println语句），然后，分别使用Maven和sbt工具对程序进行编译打包，最后，使用spark-submit命令运行程序。  1. 编写一个简单的Scala代码文件（代码中需要包含println语句）      2. 使用Maven工具对程序进行编译打包，并使用spark-submit命令运行程序。                3. 使用sbt工具对程序进行编译打包，并使用spark-submit命令运行程序。 | | | | |
| 出现的问题：  1.问题一：在安装sbt时，要对Shell脚本文件增加可执行权限，因为输错了字母的大小写，导致权限没有赋上。 | | | | |
| 解决方案（列出遇到的问题和解决办法，列出没有解决的问题）：  1.问题一解决方案：修改字母大小写后，可以正确赋上可执行权限。使用版本查询命令可以查看sbt版本信息。 | | | | |

（备注：相关实验过程可以截图放入实验报告）