**电 子 科 技 大 学**

**实 验 报** **告**

**学生姓名：** 赵则橦 **学号：** 2017221305013 **指导教师：黄俊**

**实验地点：** 信软楼 **实验时间：** 2019/11

**一、 实验名称：**Eclipse编译MapReduce程序

**二、** **实验学时：**4学时

**三、 实验目的：**

1. 掌握用 Eclipse 和 Hadoop 开发应用基础
2. 熟悉 Eclipse 和 hadoop 编程基本操作

**四、 实验原理**

在这个实验中，我们需要在 Eclipse 上运行程序，在 Eclipse 上对我们的 hdfs 进行相关配置，使得我们可以在 Eclipse 上观察到我们 hdfs 上的文件路径。然后我们就可以进行 MapReduce 执行，MapReduce 是一种编程模型，用于大规

模数据集（大于 1TB）的并行运算。使用 mapReduce 能够使得程序具有更高的并发性，从而获得更好的运算效率。

**五、 实验内容**

首先这个实验的主要内容为在 Eclipse 上运行一个 MapReduce 程序，所以第一部要在设备上安装 Eclipse 所以要进行解压，然后在把它拷贝到相应目录，因为大多数 ubuntu 上许多软件都是所谓的绿色软件，只需要自己配置相应运行环境就可以运行，然后把 hadoop-eclipse-plugin 复制到 Eclipse 上，这个是后面运行 MapReduce 所需要的包，接着打开 Eclipse 新建属于我们的 MapReduce，然后在 Eclipse 配置 Hadoop，如果没有之前的包，进行相关配置的。完成配置后就可以在项目目录上看到我们的 hdfs 上的文件。

最后编写我的 wordCount 代码，在运行之前需要对运行环境进行配置。确

定输入和输出路径，当然这两个路径都是我们 hdfs 上的文件路径。接着我们就

可以利用 cat 命令查看我们的输出结果。

**六、 实验设备及环境**

Ubuntu jdk 1.8 hadoop 2.7.x Eclipse

1. **实验步骤**

1.安装eclipse

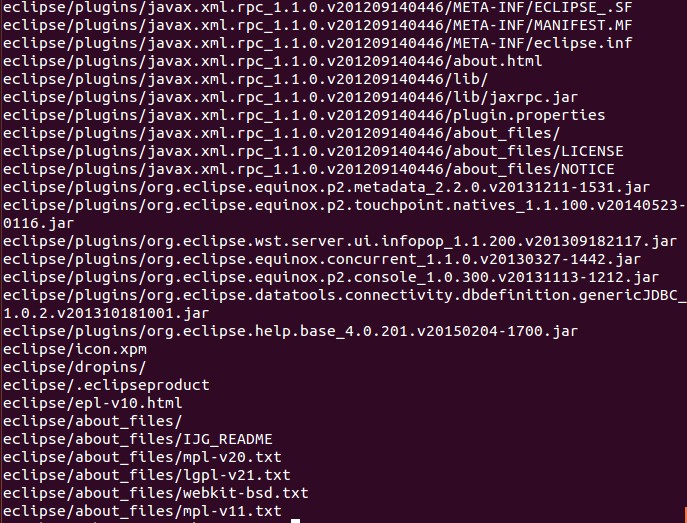
2.安装 hadoop-eclipse-plugin

3.查看 HDFS 中的文件内容

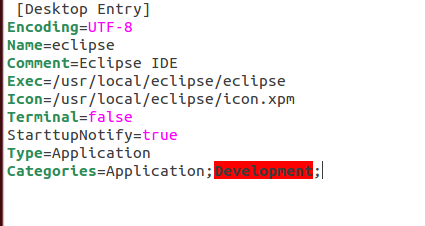
4.在 Eclipse 中编写WordCount程序

5.在 Eclipse 中运行 MapReduce程序

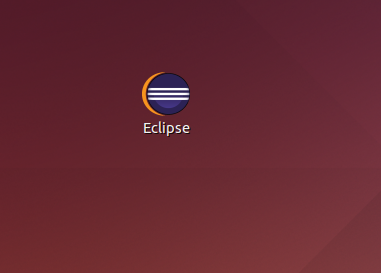
1. **实验结果与分析（含重要数据结果分析或核心代码流程分析）**
2. **解压eclipse**



创建eclipse桌面快捷图标，配置eclipse，将eclipse变为可执行文件，最后将其图标拷到桌面





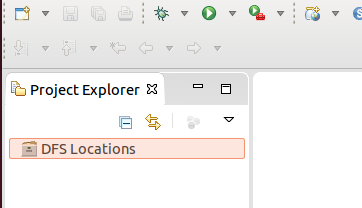


1. 安装配置hadoop-eclipse-plugin

拷贝hadoop-eclipse-plugin-2.7.3.jar到Eclipse/plugins目录下

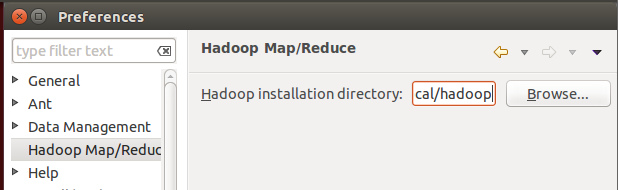


启动 Eclipse 后就可以在左侧的Project Explorer中看到 DFS Locations。

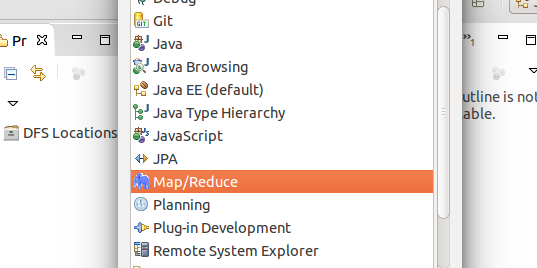


插件需要进一步的配置

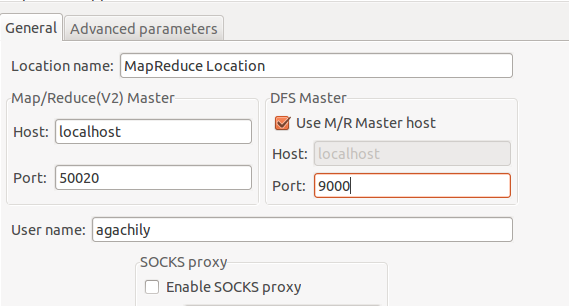
此时会弹出一个窗体，窗体的左侧会多出 Hadoop Map/Reduce 选项，点击此选项，选择 Hadoop 的安装目录

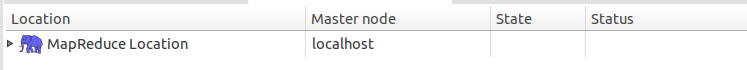


切换 Map/Reduce 开发视图



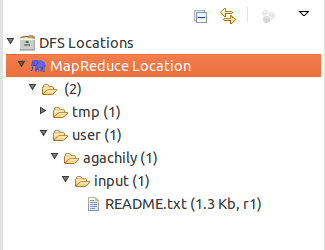
建立与 Hadoop 集群的连接





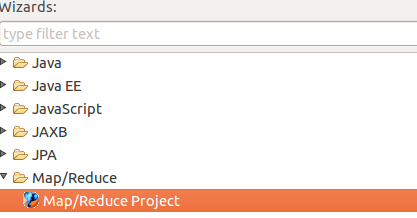
1. 查看 HDFS 中的文件内容

配置成功后，点击左侧 Project Explorer中的 MapReduce Location 就能直接查看 HDFS 中的文件内容了（如下图是 WordCount 的输出结果

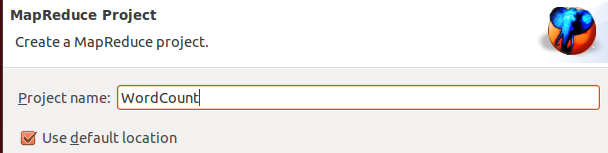


1. **在 Eclipse 中创建 MapReduce 项目**

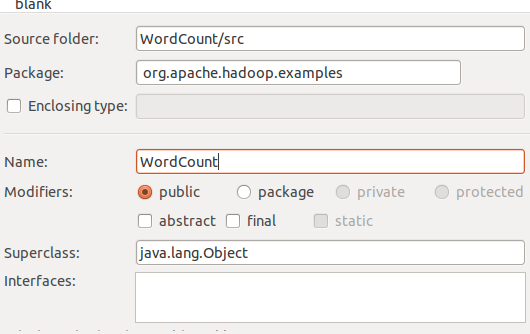
点击 File 菜单，选择 New -> Project…:  
选择 Map/Reduce Project，点击 Next。

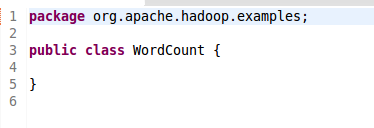


填写 Project name 为 WordCount 即可，点击 Finish 就创建好了项目

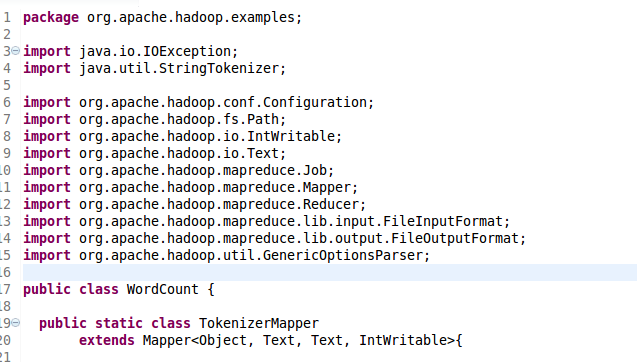


接着右键点击刚创建的 WordCount 项目，选择 New -> Class  
需要填写两个地方：在 Package 处填写 org.apache.hadoop.examples；在 Name 处填写 WordCount。



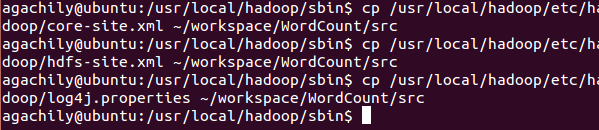


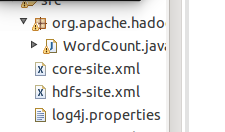
创建 Class 完成后，在 Project 的 src 中就能看到 WordCount.java 这个文件。将如下 WordCount 的代码复制到该文件



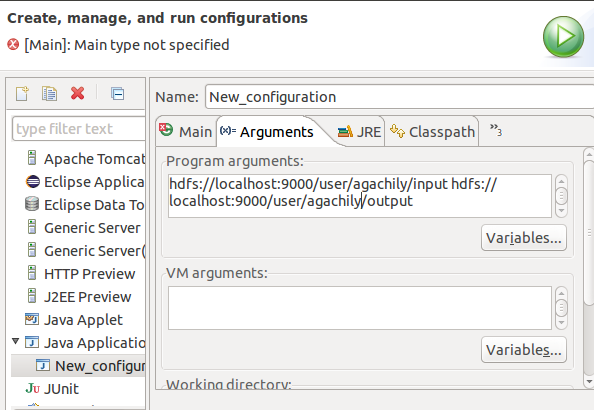
1. 通过 Eclipse 运行 MapReduce

对修改过后的配置文件进行操作





Run As -> Run Configurations，在此处可以设置运行时的相关参数

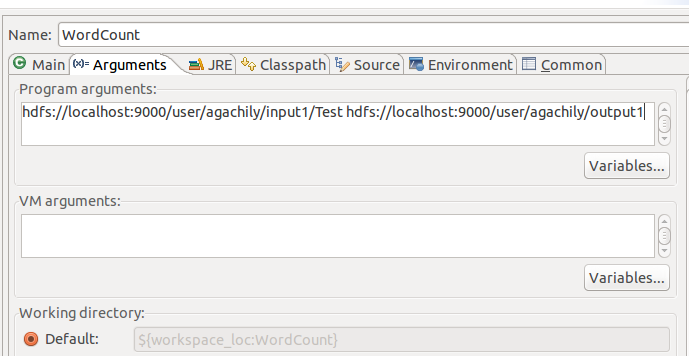


以readme文件进行实例

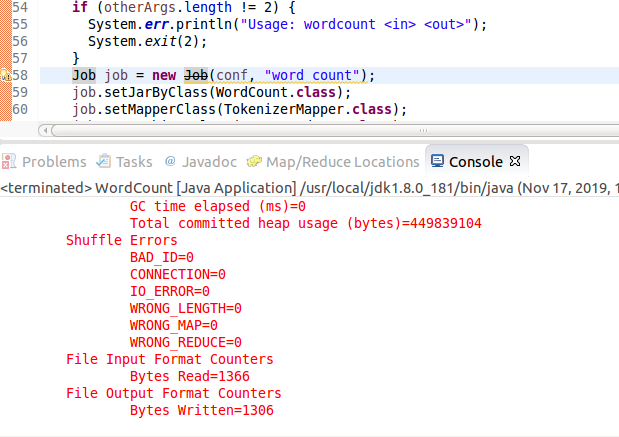
上传readme

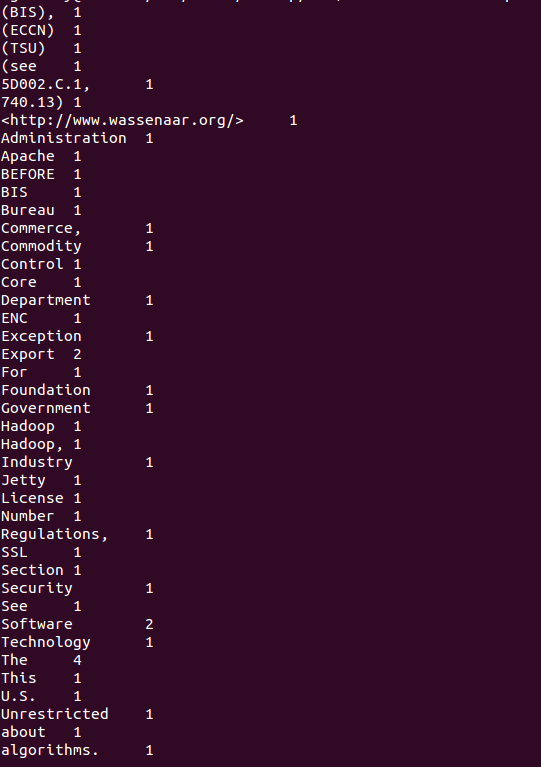


对wordcount配置进行更改



查看运行结果





1. **总结及心得体会：**

通过这次试验我在Ubuntu环境下安装了eclipse，并且在eclipse上对hdfs进行了相关的配置，然后在eclipse上运行了一个MapReduce的程序，在安装eclipse的时候我发现了一些坑，首先要对eclipse中的进行java路径的配置，否则哪怕已经安装了java依旧是无法成功运行eclipse。其次在配置文件中的Exec和Icon行中一定要与eclipse实际所在的环境相适应。在对eclipse进行与hadoop有关的操作时，我了解到了eclipse是如何与hadoop进行交互运作的。同时这也是实验指导书的一个问题，在进行实验3和4的时候是需要运行hadoop的，否则无法进行操作。另外在运行每一个程序时，都需要对程序的run

Configation进行涉资，而且要提前将文件上传到hdfs集群中。并在run configation中指明集群和输出路径。 并且要注意端口号的一致。

这次实验帮助我了解了如何用Eclipse去编译运行mapreduce文件，熟悉了Eclipse和hadoop的基本操作，也锻炼了的动手能力，要时刻注意着各种先前步骤的配置完善，这对我是个很大的提高，最后我用hadoop中的readme对程序进行了检测。

**十、对本实验过程及方法、手段的改进建议：**

无

**报告评分：**

**指导教师签字：**