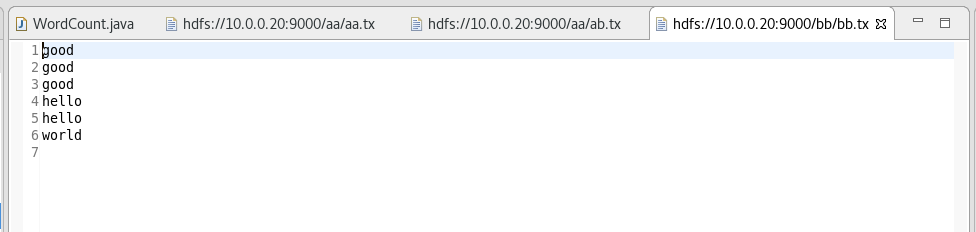
大数据第8周

1. wordcount程序的运行
   1. 上传文本文件（英文）到hdfs文件系统

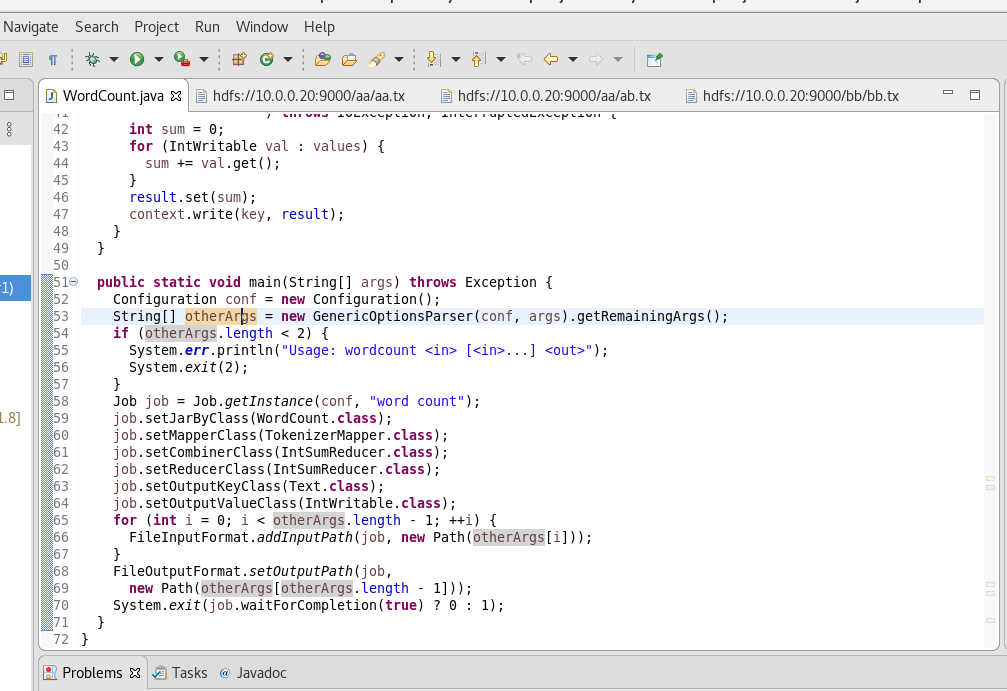
上传文件后，可以在eclipse环境中看到路径。



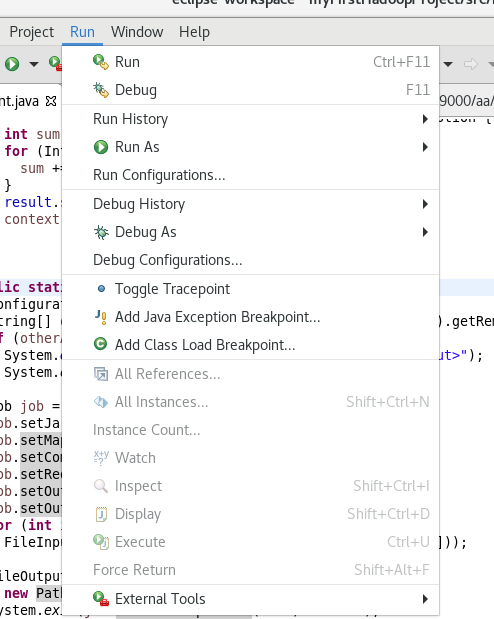
这些路径作为main的运行参数，需要填入configuration里。

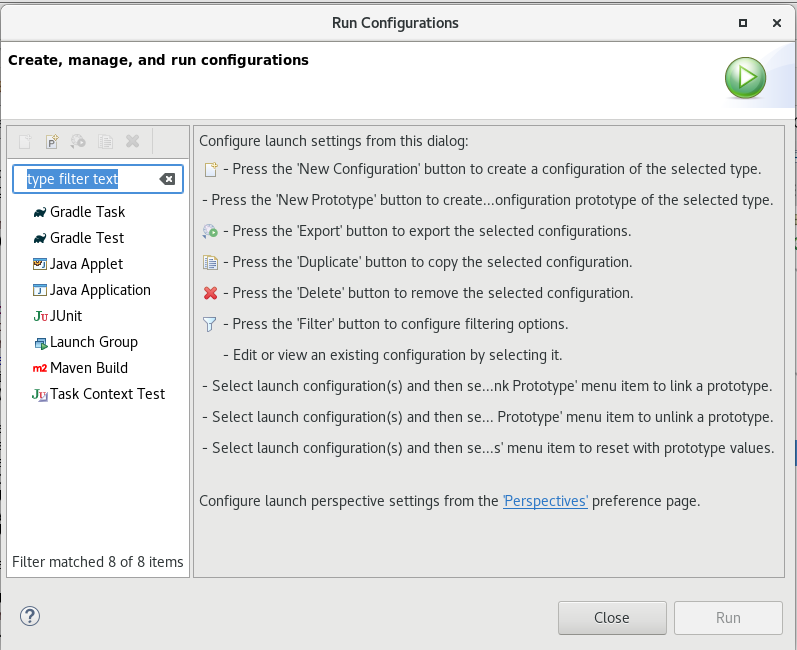
1.2设置configuration

首先，eclipse的激活页面要是程序界面：

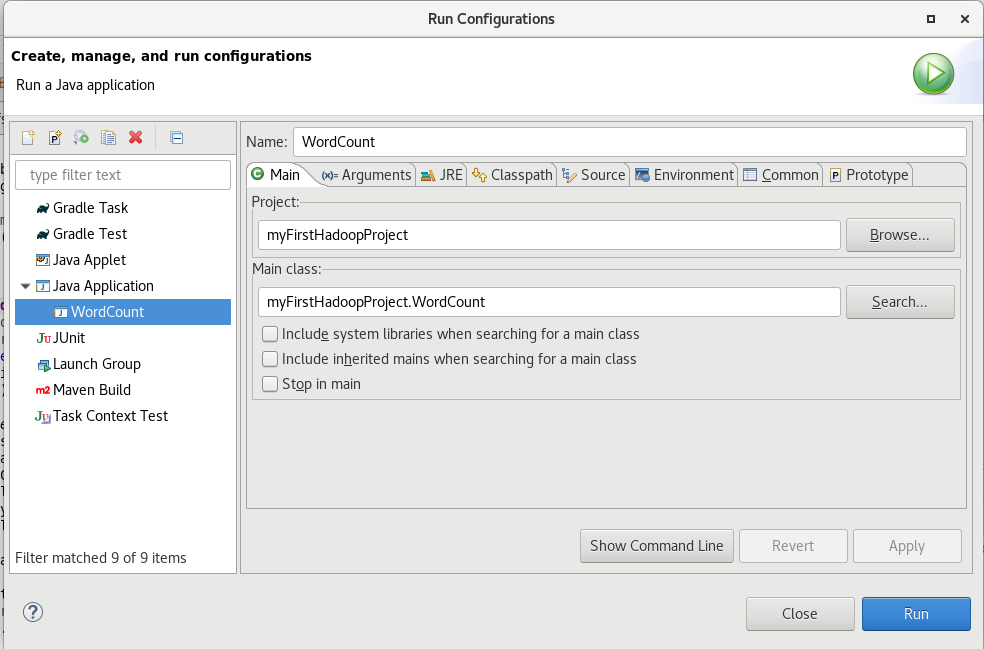


在主菜单选择“Run”，然后“Run Configurations”。

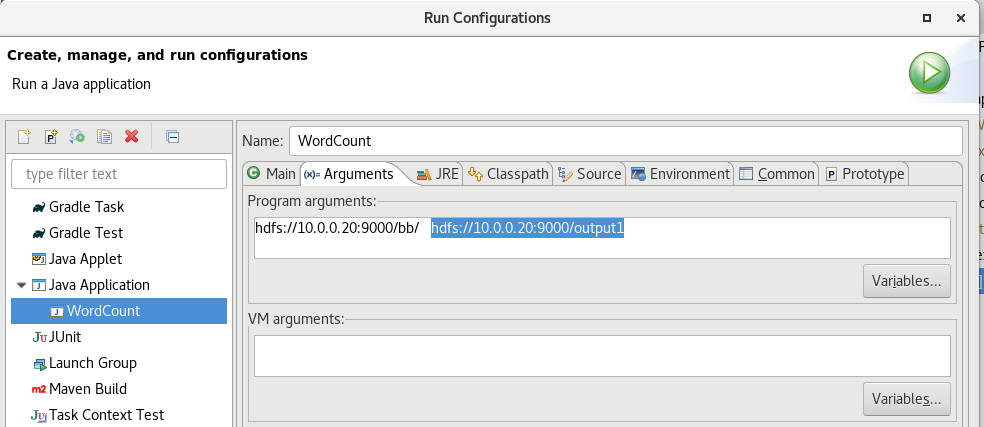




双击“Java Application”，检查自动填入的内容是否正确。

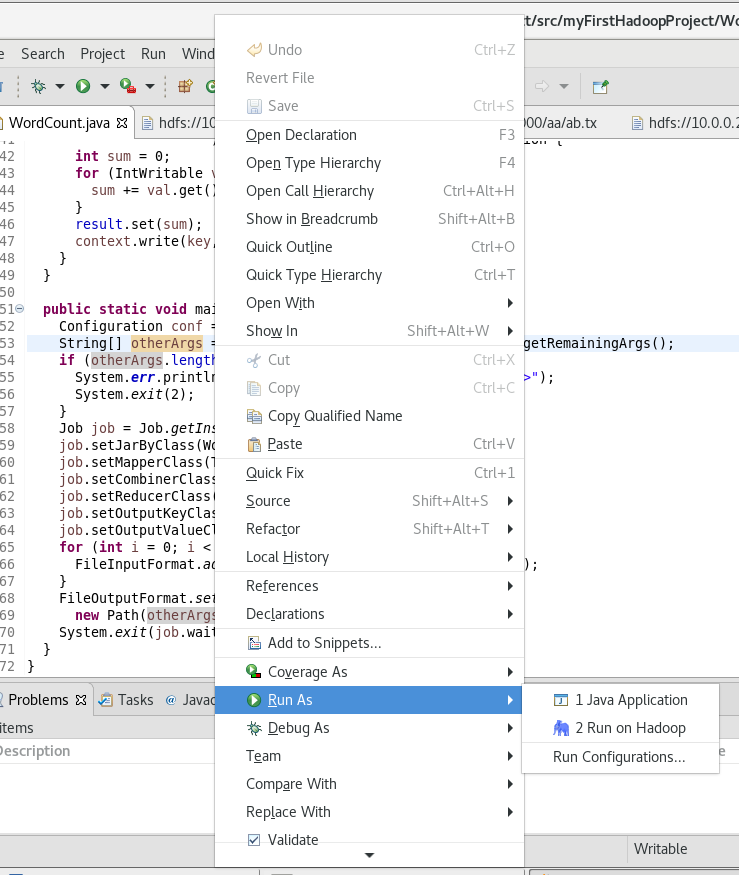


打开“Arguments”标签页，填入路径。



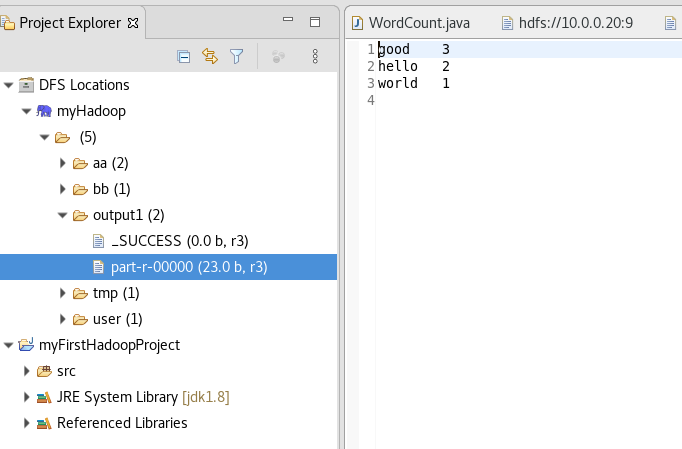
填好后点击“Apply”，然后 “Close”。

在程序界面上右击：



选择“Run As”，“Run on Hadoop”。

运行之后，可以在右侧看到hdfs文件的变化：



1. spark相关信息

下载解压spark2.4.5。

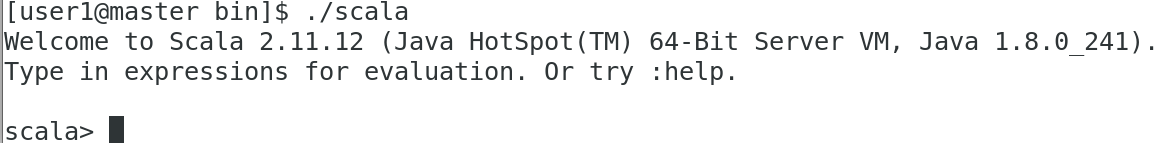
在路径：/home/user1/spark-2.4.5/bin下，运行./spark-shell。



1. scala下载解压与配置

3.1下载scala2.11.12，解压：tar -zxvf scala-2.11.12.tgz。

进入目录：/home/user1/scala-2.11.12/bin，运行./scala：



说明scala运行正常。

3.2配置路径

在.bash\_profile里添加如下内容：

export SCALA\_HOME=/home/user1/scala-2.11.12

export PATH=$SCALA\_HOME/bin/:$PATH

然后运行.bash\_profile：source .bash\_profile。

3.3 scala环境退出

“:q”。

1. scala语言简介

4.1值和变量

值：val；

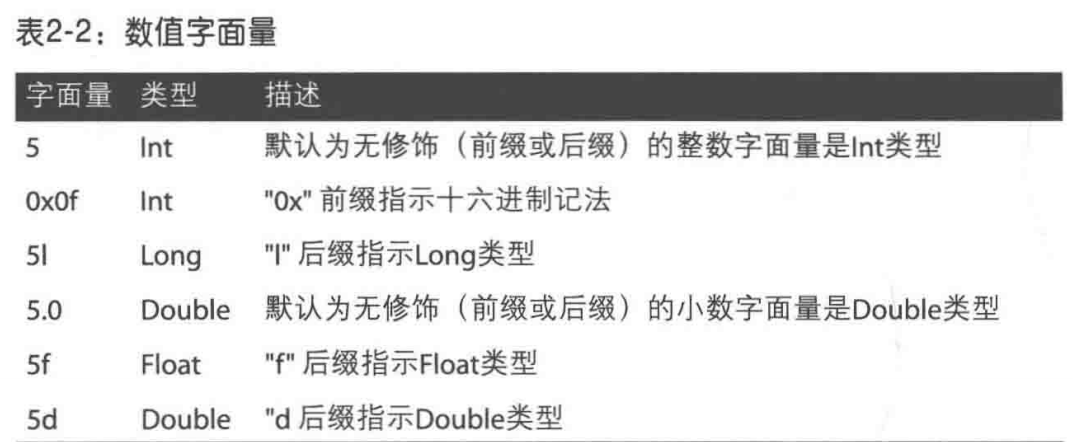
变量：var。

值：声明后不能修改；变量可以。



在计算过程，低级数值类型会自动向高级数值类型转换，但不会自动从高级数据类型向低级数据类型转换。

类型转换可以强制执行，使用toType方法。



字面量中的字母不区分大小写。

4.2 字符串String

用双引号。

String类型比数值类型级别高，所以计算过程中数值类型会转换为String类型，进行String类型的计算（拼接）。

4.3 scala类型概述

