

华东师范大学数据科学与工程学院实验报告

课程名称： 分布式模型与编程	年级： 2016 级	上机实践成绩：
指导教师： 徐辰	姓名： 张宏伟	
上机实践名称： Spark Streaming 使用	学号： 10165101180	上机实践日期：
上机实践编号： 11	组号：	上机实践时间：

一、实验目的

使用 Spark Streaming

二、实验任务

阅读官方样例 NetworkWordCount 的代码，了解 Spark Streaming 的编程接口并运行。

三、使用环境

Spark2.3.2

四、实验过程

用 nc 命令向端口 9999 写数据

```
hadoop@PC-honwee:~$ nc -lk 9999
hello world
yes you are          Then, any lines
```

在另一个终端运行 NetworkWordCount，设定间隔为 1s，所以不太好找输出。

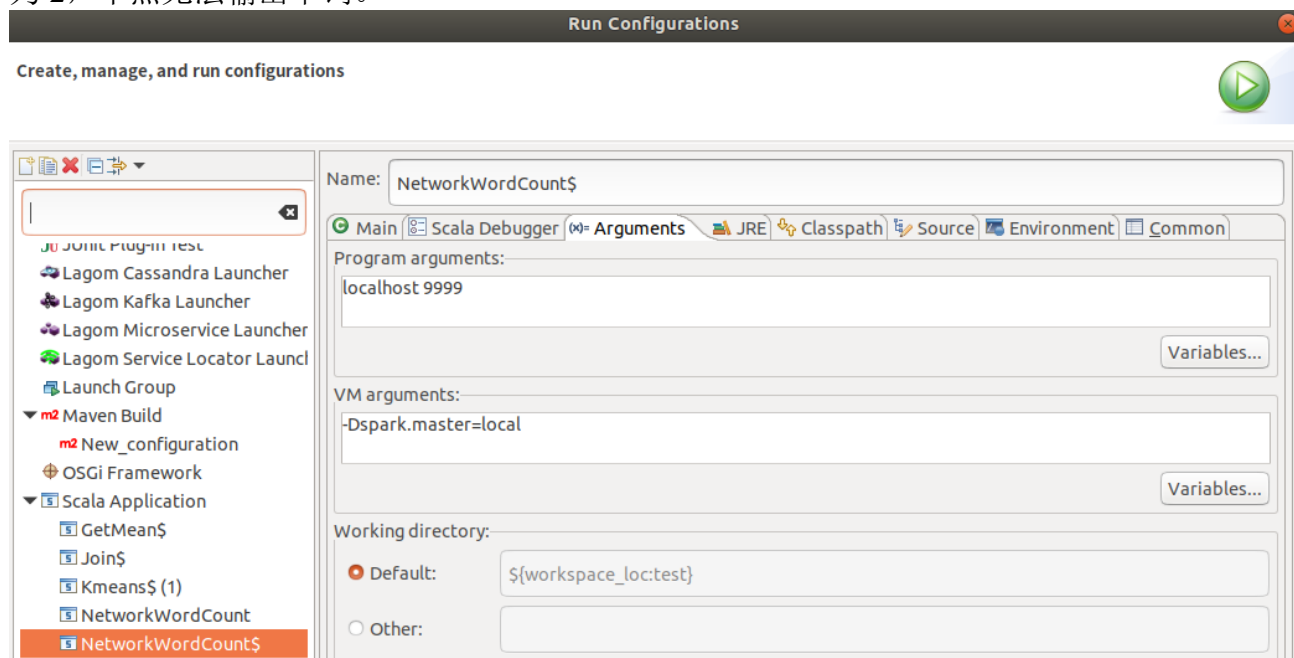
输入第一行

```
-----
Time: 1543280508000 ms
-----
(hello,1)
(world,1)          ost 9999
```

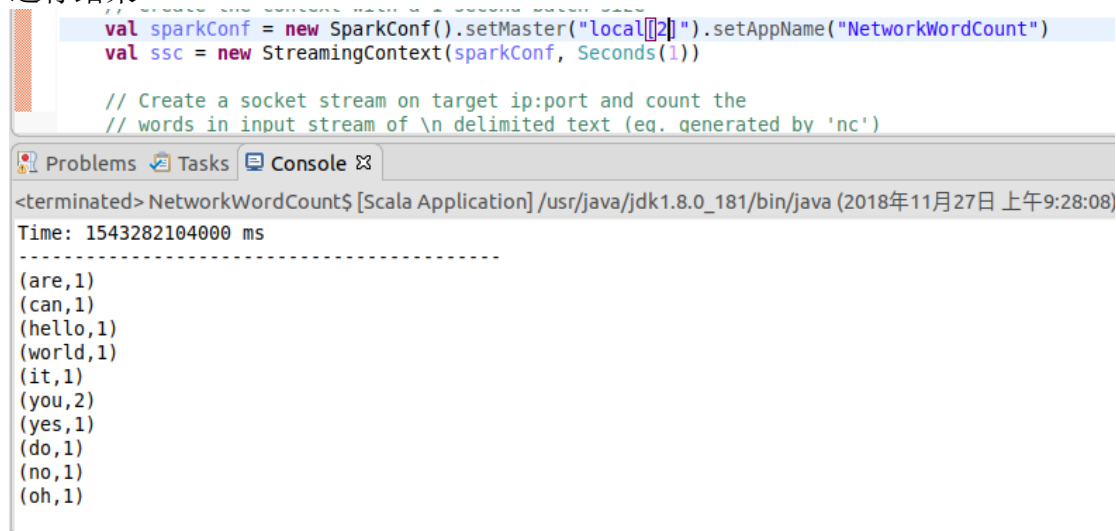
输入第二行

```
-----
Time: 1543280517000 ms
-----
(are,1)             counted and printed
(you,1)
(yes,1)
```

尝试在 scala 版的 eclipse 中配置 spark streaming 环境，先在 maven 的 pom.xml 中添加 spark streaming 的依赖，再添加 run configuration 的参数，注意需要更改代码中运行线程为 2，不然无法输出单词。



运行结果



五、总结

nc -lk port 命令可以用来对涉及网络的程序进行测试。

Eclipse 运行程序时可以添加参数模拟命令行启动。