

## 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：吴双	
上机实践名称：Spark Streaming 使用	学号：10164102141	上机实践日期：
上机实践编号：#11	组号：23	上机实践时间：

### 一、实验目的

Spark Streaming 的使用与相关样例的实现

### 二、实验任务

阅读 Spark Streaming 的 A Quick Example 部分；运行 wordcount 例子

### 三、使用环境

Ubuntu LTS 18.04  
Spark 2.3.2

### 四、实验过程

#### 1. 用 nc 命令向端口 9999 写数据

```
hadoop@Master ~$ nc -lk 9999
hello world
yes you are
>
```

#### 2. 在另一个终端运行 NetworkWordCount

##### i. The first line

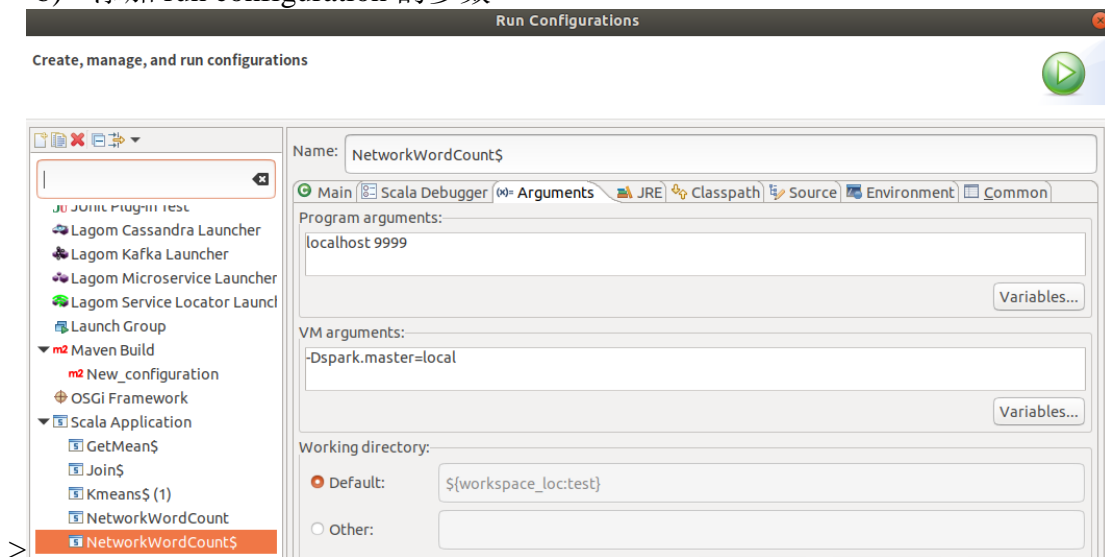
```
Time: 1543280508000 ms
-----
(hello,1)
(world,1)
>
```

##### ii. The second line

```
Time: 1543280517000 ms
-----
(are,1)
(you,1)
(yes,1)
>
```

## 3. 尝试在 scala 的 eclipse 中配置 spark streaming 环境:

- 在 maven 的 pom.xml 中添加 spark streaming 的依赖;
- 添加 run configuration 的参数



## 运行结果

```
<terminated> NetworkWordCount$ [Scala Application] /usr/java/jdk1.8.0_181/bin/java (2018年11月27日 上午9:28:08)
Time: 1543282104000 ms
-----
(are,1)
(can,1)
(hello,1)
(world,1)
(it,1)
(you,2)
(yes,1)
(do,1)
(no,1)
(oh,1)
```

## 五、总结

命令 `nc -lk <port>` 相当好用，可以用来对涉及网络的程序进行测试。