

## 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： 基于 Dataflow 系统的图算法编程	学号： 10165101180	上机实践日期：
上机实践编号： 14	组号：	上机实践时间：

### 一、实验目的

体会用非面向图运算的分布式系统

### 二、实验任务

分别使用 MapReduce、Spark、Flink 实现 pageRank

### 三、使用环境

### 四、实验过程

Flink:

输入

```
1 2
1 15
2 3
2 4
2 5
2 6
2 7
3 13
4 2
5 11
5 12
6 1
6 7
6 8
7 1
7 8
8 1
8 9
8 10
```

输出

```
Printing result to stdout. Use --out
(1,0.2507740474831439)
(2,0.14680796376156166)
(3,0.03476167589543225)
(4,0.03476167589543225)
(5,0.03476167589543225)
(6,0.03476167589543225)
(7,0.04464948608422618)
(8,0.03891448491322475)
(9,0.021024659120945137)
```

## MapReduce:

### 输入

```
1 8 2,4,5,6,7,8,9
2 10 3,4,7,8,9
3 9 1,4,6,8,10
4 8 2,3,5,8,9,10
5 9 1,2,3,6,8
6 7 2,3,4,7,9
7 5 1,2,3,5,6,8,9
8 6 1,3,5,6,9,10
9 5 2,3,4,5,7
10 8 1,3,4,5,6,8,9
```

### 输出

```
1 6.457143
2 13.847619
3 24.238094
4 32.72381
5 39.05714
6 46.65714
7 52.199977
8 62.133327
9 70.86667
10 75.00001
```

## Spark:

### 输入

```
A B,C
B A,C
C A,B,D
D C
```

### 输出

```
A 0.985024
B 0.980243
C 1.462103
D 0.567848
```

## 五、总结

实现过程的代码较为复杂。