

并行与分布式计算导论 作业 3

PDC 2025s Homework 3

截止期限 2025 年 6 月 4 日 23:59

DDL: 2025 Jun. 4 23:59 (GMT+8)

1. 假设现在有一份全国人口普查的数据库，目的是统计北京市 80 岁以上（即出生年份为 1945 年及以前）的人数。输入为一系列键值对“(身份证号, 人口记录)”，每条人口记录为字符串，格式为：“姓名,籍贯,性别,户籍所在地”，例如“(110101199001011234, “张三,北京,男,北京”)”，其中出生年月可以从身份证号 7~10 位提取，输出格式为：“(“北京”, 人数)”。你的任务是使用 MapReduce 的编程接口，统计当前户籍在北京且年龄在 80 岁以上的人数。定义编程接口如下：

```
def mapper(key, value):
```

```
    # 填写伪代码完成功能
```

```
    # 用 emit(k, v)接口生成 map 所得的键值对
```

```
def reducer(key, values):
```

```
    # 填写伪代码完成功能
```

```
    # 用 emit(k, v)接口生成 reduce 所得的键值对
```

2. 为了评估系统在长时间运行下的稳定性，我们使用平均故障间隔时间（Mean Time Between Failures, MTBF）这一指标，该指标表示系统或组件在两次故障之间的平均运行时间。某系统由 500 个独立节点组成，每个节点的 MTBF 为 1500 天。由于未考虑容错，该系统只要任意一个节点发生故障，整个系统就会停止工作。假设各节点的故障事件相互独立，求该系统整体的 MTBF。