

# 电类工程学导论 C 实验报告 8

518030910406 郑思榕

## 一、实验准备

### 1. 实验环境介绍

- a) 环境：在 windows 系统中使用 VirtualBox 5.2.18 安装 Ubuntu14.04 虚拟机，从而在 UNIX 系统环境下进行本次实验。
- b) 工具：本实验主要使用了分布式系统基础架构 Hadoop

### 2. 实验目的

- a) 练习使用基础的 hadoop 命令来填写下列表格：

Number of Maps	Number of samples	Time(s)	$\pi$
2	10		
5	10		
10	10		
2	100		
10	100		

- b) 找到一个解决方法使得计算得出的 $\pi$ 近似到小数点后五位

### 3. 实验原理

利用 Hadoop 自带的 $\pi$ 计算程序进行计算。该程序模拟投掷飞镖的过程，当在单位正方形内做内切圆，在正方形内随机投掷飞镖，则单位正方形面积：单位内切圆面积=单位正方形内的飞镖数：内切圆内的飞镖数，记该比值为 k。应用圆面积= $\pi r^2$ 可以得到  $k = (r^2) / (\pi r^2)$ ，所以 $\pi = 4/k$ 。

## 二、实验过程

### 1. ex1

当安装好 hadoop 并配置好相关的环境变量后，使用 start-dfs.sh 和 start-yarn.sh 打开运行 Hadoop。然后使用下面的命令运行 Hadoop 自带的计算 $\pi$ 的程序。其中 jar 为运行一个 jar 文件。<nMaps>

```
$ hadoop jar /usr/local/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.7.3.jar pi <nMaps> <nSamples>
```

为 mapper 任务的数量，<nSamples>为每个 mapper 任务里要投掷的飞镖数，这两个量分别对应表格里的 Number of Maps 和 Number of samples。所以，当输入<nMaps>和<nSamples>的值分别为 2 和 10，得到结果如下图：

```
Job Finished in 27.81 seconds
Estimated value of Pi is 3.80000000000000000000
```

当运行完五个例子，得到最终表格如下图。可以见到 Number of Maps 增大时，运行时间大幅增加，

Number of Maps	Number of samples	Time(s)	$\pi$
2	10	24.852	3.800000000 000000000 0
5	10	32.283	3.280000000 000000000 0
10	10	39.915	3.200000000 000000000 0
2	100	21.647	3.120000000 000000000
10	100	41.782	3.148000000 000000000 0

Number of samples 增大时，运行时间较大幅度增加。两者增大时， $\pi$ 的计算精度提高。

## 2. ex2

要想至少得到小数点后五位的 $\pi$ 估计值，需要同时提高<nMaps>和<nSamples>的值。所以输入二者分别为 100 和 1000000，输入命令并运行，得到结果如下图。

```
hduser@zslr-VirtualBox: ~
Reduce output records=0
Spilled Records=400
Shuffled Maps =100
Failed Shuffles=0
Merged Map outputs=100
GC time elapsed (ms)=35466
CPU time spent (ms)=64880
Physical memory (bytes) snapshot=26809266176
Virtual memory (bytes) snapshot=193226043392
Total committed heap usage (bytes)=20206583808

Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0

File Input Format Counters
Bytes Read=11800
File Output Format Counters
Bytes Written=97
Job Finished in 293.329 seconds
Estimated value of Pi is 3.14159256000000000000
hduser@zslr-VirtualBox:~$
```

## 三、实验总结

### 1. 实验概述

本实验通过 Hadoop 自带的 map/reduce 程序 pi 来计算 $\pi$ 。该程序通过 quasi-Monte Carlo 方法模拟随机投掷飞镖的过程来计算 pi。随着<nMaps>和<nSamples>的提高，运行时间提高，计算得到的 $\pi$ 的精度也提高。

### 2. 实验心得

- 认识了解了 Hadoop 的 map 和 reduce 的过程。
- 学会了 Hadoop 的基本操作。

- c) 学会了如何运用 Hadoop 自带的 pi 计算程序来计算 pi 值。

最后，衷心感谢实验中老师和各位助教的帮助！