第八次实验报告

学号：518030910308
姓名：刘文轩

一、实验准备

1、实验环境介绍

操作系统：ubuntu 14.04

架构：Hadoop

2、实验目的

2.1 安装 Hadoop

2.2 实现 wordcount

2.3 运用 Hadoop 粗略地计算 π，选择适当的参数精确到更多的位数

3、实验思路

3.1 根据教程新建 hduser

3.2 在 hduser 中安装 Hadoop

3.3 运用 Hadoop，统计 txt 文件中的单词个数

3.4 运用 Hadoop，估算 π 的值

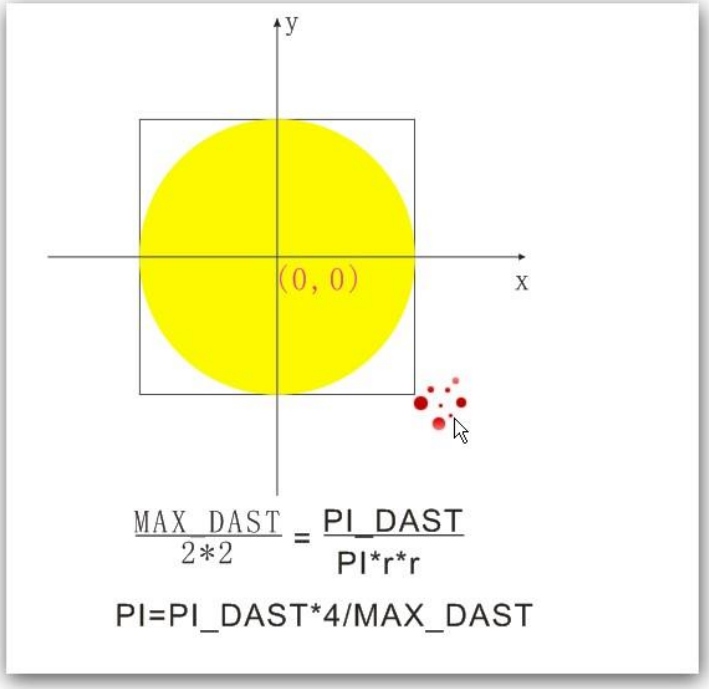
二、实验过程

1、根据不同的 map 和 reduce，估算π

1.1 计算原理

我们运用 Hadoop 自带的代码估算 π 的值。

Hadoop 采用 Quasi-Monte Carlo 算法来估算 PI 的值。原理图如下：



这是蒙特卡洛算法，我们取一个单位的正方形（1×1）里面做一个内切圆（单位圆），则单位正方形面积：内切单位圆面积=单位正方形内的飞镖数：内切单位圆内的飞镖数，通过计算飞镖个数就可以把单位圆面积算出来，通过面积，把圆周率计算出来。

因此：精度和投掷的飞镖次数成正比。

1.2 实验结果

我们使用如下的代码：

```
1. $ hadoop jar /usr/local/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-
   examples-2.2.0.jar pi  <nMaps>  <nSamples>
```

第 1 个 nMap 指的是要运行 map 任务的次数，第 2 个 nSamples 指的是每个 map 任务，要投掷多少次。2 个参数的乘积就是总的投掷次数。

第一次实验，我们分别选取两个参数为 2 和 10，运行的部分结果如下：

```
hduser@ubuntu:~$ hadoop jar /usr/local/hadoop/share/hadoop/mapreduce/hadoop-mapr
educe-examples-2.2.0.jar pi  2 10
Number of Maps  = 2
Samples per Map = 10
```

```
Job Finished in 21.628 seconds
Estimated value of Pi is 3.80000000000000000000
```

此时的精度还不是很高，我们选取不同的参数，得到的结果如下表：

Number of Maps	Number of Samples	Time	π
2	10	21.628	3.80000000
5	10	32.346	3.28000000
10	10	45.869	3.20000000
2	100	20.618	3.12000000
10	100	46.686	3.14800000

可见，整体上，随着投掷次数的增加，实验的时长在增加，估算的精度在提高。同时，降低 Maps 次数，提高样本数量，能降低实验的时长。

2、 获得更加精确的π

2.1 实验结果

要获得更加精确的π，我们只需要提高“投掷次数”即可。

在此，我选择 Maps 任务次数为 100 次，每次样本数为 10000000。

执行结果如下：

```
hduser@ubuntu:~$ hadoop jar /usr/local/hadoop/share/hadoop/mapreduce/hadoop-mapr
educe-examples-2.2.0.jar pi  100 10000000
Number of Maps  = 100
Samples per Map = 10000000
```

```
Job Finished in 345.423 seconds
Estimated value of Pi is 3.14159273600000000000
```

我们计算得到的 π 为 3.141592736，精度已经达到了小数点后 5 位，完成了实验要求。

三、实验总结

1、 实验概述

本次实验的主要任务，可以总结为安装 Hadoop，并运用它完成简单的任务，是崭新的篇章。

2、感想总结

在这次的实验中，学会了的东西有很多，其中最重要的就是提高了自己处理问题、收集相关资料、解决问题的能力，这在我们将来的学习和工作生活中都是很重要的。而具体细化开来，在本次实验中：

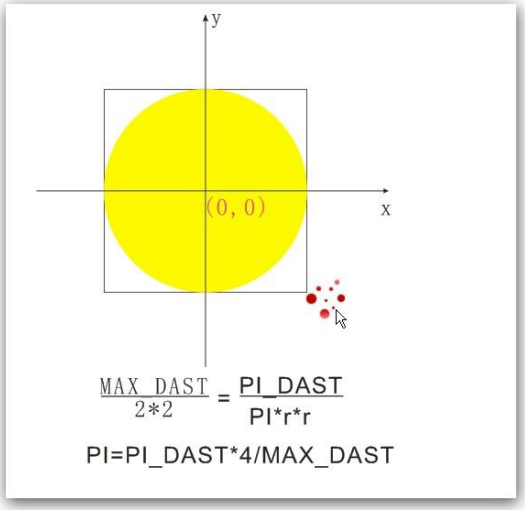
2.1 学会了如何安装 Hadoop

2.2 学会了如何运用 Hadoop 进行词数统计

2.3 学会了如何运用 Hadoop 进行π的计算

3、创新点

本次实验中我主要的创新点是阐明了 Hadoop 估算π的算法，这是运用了蒙特卡洛算法。



4、遇到的问题

本次实验中遇到的主要问题是安装 Hadoop 的过程中可能会遇到各种各样的异常情况，然而我并没有遇到，在了解了 vim 的使用方法后，本次实验还是很简单的。