# 实验四: SecondSort

组号:2020st32

小组成员:

171860607 白晋斌

171860033 黄诗涵

171180541 祖东珏

联系方式:810594956@qq.com

分工:因本次实验较为简单,小组成员决定分别独立完成代码编写. 此外,三位同学的源代码、JAR 包、JAR 包执行方式说明详见附件.

## 目录

实验任务	3
Map 和 Reduce 的设计思路	3
0.数据结构	3
1. Map	3
2.Reduce	3
3.自定义 compareTo 函数	3
4.自定义排序	4
5.其他思路	4
集群测试	4
1 运行 Jar 包	4
2 查看结果	
— WebUI 执行报告	
参考资料	5

### 实验任务

- 1. 使用 MapReduce 完成/data/exercise\_4 目录下 SecondSort.txt 的二次排序。
- 2. 实验数据分为两列,第一列为 1 到 10 之间的随机数字,第二列为 1 到 100 的随机数字。本次实验我们先对第一列数字按照升序排列,即所谓的分组,再在每组中按照第二列 数组进行降序排序完成二次排序。

## Map 和 Reduce 的设计思路

#### 0.数据结构

封装一个 Bean 存储每行的数据,每个 Bean 有两个成员,分别表示第一列和第二列.

```
1. public static class OrderBean implements WritableComparable<OrderBean> {
2.     private Integer first;
3.     private Integer second;
4.     //func
5. }
```

#### 1. Map

将读取到的数据存储为 Bean,然后将 Bean 作为 key, value 作为空发送.

#### 2.Reduce

Reduce 函数负责解包,将 Bean 解开,第一个成员作为 key,第二个成员作为 value 发送.

#### 3.自定义 compareTo 函数

按第一列升序、第二列降序排列.

```
1. public int compareTo(OrderBean o) {
2.    if (!this.first.equals(o.first)) {
3.        return this.first.compareTo(o.first);
4.    } else if (!this.second.equals(o.second)) {
5.        return -this.second.compareTo(o.second);
6.    } else {
7.        return 0;
8.    }
9. }
```

## 4.自定义排序

重载 Bean 的排序.

```
1. public static class SortComparator extends WritableComparator {
2.    public SortComparator() {
3.         super(OrderBean.class,true);
4.    }
5.    public int compare(OrderBean a,OrderBean b)
6.    {
7.         return a.compareTo(b);
8.    }
9. }
```

#### 5.其他思路

可直接在 map 函数中,将第一列作为 key,第二列作为 value 发送.利用系统自带的排序函数完成第一列的升序.

在 reduce 函数中,key 相同的为一组,编写排序函数对该组的数据降序排序,之后降序输出即可.

## 集群测试

#### 1 运行 Jar 包

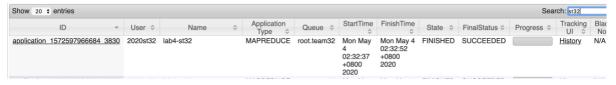
hadoop jar /home/2020st32/BigDataLab4.jar /data/exercise\_4/SecondSort.txt /user/2020st32/lab4

#### 2 查看结果

在大数据分布文件存储->HDFS 分布式存储中找到输出的文件



### WebUI 执行报告

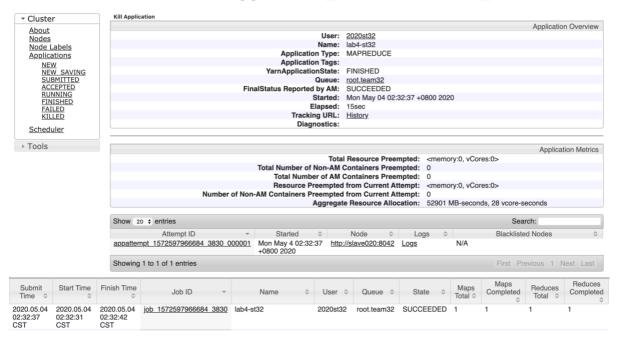




# Application application\_1572597966684\_3830

Logged in as: dr.who

Logged in as: dr.who



## ووماها

▼ Job

Overview
Counters
Configuration
Map tasks
Reduce tasks

▶ Tools

#### MapReduce Job job 1572597966684 3830

## 参考资料

- [1] https://blog.csdn.net/sinat 34022298/article/details/72854240
- [2] https://www.cnblogs.com/hrhguanli/p/4648740.html
- [3] https://blog.csdn.net/zhemeban/article/details/56843017
- [4] https://blog.csdn.net/u011463542/article/details/90693520
- [5] https://www.cnblogs.com/michaellfx/p/4001148.html