#### 实验目的:

- 1) 理解 mapreduce 执行原理
- 2) 理解 map, reduce 阶段
- 3) 熟悉统计的原理
- 4) 熟悉去重原理
- 5) 熟悉平均函数的使用
- 6) 了解排序的原理
- 7) 二次排序的原理
- 8) 了解 mapreduce 处理日志的原理

# 实验环境:

本次环境是: centos7.5 + jdk1.8.0\_79 + hadoop2.4.1 + eclipse 日志文件 source.txt 存放在/root/simple/目录下

# 实验技术原理:

## 1. MR1-1 词频统计

首先对待处理的信息进行拆分,拆分之后在 map 阶段,把拆分的每个单词作为 map 方法的输出键,而 map 的方法输出的值设置为 1,最后在 reduce 阶段 对每个键的值集合进行遍历并把遍历的值进行相加,输出结果即可。

# 2.MR1-2 单词个数

首先对待处理的信息进行拆分,拆分之后在 map 阶段,拆分后计算出单词个数并作为 map 方法的输出值,而 map 的方法输出键作为 NullWritable 即可,最后在 reduce 阶段对每个键的值集合进行遍历并把遍历的值进行相加,输出结果即可。

## 3.MR2-2 成绩统计

## 【案例 3.1】 mapreduce 学生总成绩报表

## 【案例 3.2】mapreduce 学生各科平均成绩报表

Map 处理的是一个纯文本文件,文件中存放的数据时每一行表示一个学生的姓 名和他相应一科成绩。Mapper 处理的数据是由 InputFormat 分解过的数据集, 其中 InputFormat 的作用是将数据集切割成小数据集 InputSplit,每一个 InputSplit 将由一个 Mapper 负责处理。此外, InputFormat 中还提供了一个 RecordReader 的实现,并将一个 InputSplit 解析成<key, value>对提供给了 map 函数。InputFormat 的默认值是 TextInputFormat,它针对文本文件,按 行将文本切割成 InputSlit,并用 LineRecordReader 将 InputSplit 解析成 <key, value>对,key 是行在文本中的位置,value 是文件中的一行。Map 的结 果会通过 partion 分发到 Reducer, Reducer 做完 Reduce 操作后,将通过以 格式 OutputFormat 输出。Mapper 最终处理的结果对<key, value>,会送到 Reducer 中进行合并,合并的时候,有相同 key 的键/值对则送到同一个 Reducer 上。Reducer 是所有用户定制 Reducer 类地基础,它的输入是 key 和 这个 key 对应的所有 value 的一个迭代器,同时还有 Reducer 的上下文。 Reduce 的结果由 Reducer. Context 的 write 方法输出到文件中。

#### 4.MR3 MRShuffle

### 【案例 4.1】mapreduce 整数排序并指定顺序编号

本试验要求对一系列数据进行排序你,自然就会想到只要把这些数据作为 map 阶段的输出键即可,排序后的数据在输出时,在数值前面添加一个数字的自 然顺序编号,此时只要考虑在 reduce 阶段定义一个全局变量,每次执行 reduce 方法时,就是对 map 阶段的 key 进行处理,此时全局变量就可以作为 编号,每执行一次 reduce,全局变量就会增加 1。

### 【案例 4.2】mapreduce 两列数据的二次排序

在 MapReduce 操作时,我们知道传递的<key,value>会按照 key 的大小进行排序,最后输出的结果是按照 key 排过序的。有的时候我们在 key 排序的基础上,对 value 也进行排序。这种需求就是二次排序。二次排序是在框架在对 key2 排序后再对 reduce 输出结果的结果 value3 进行二次排序的需求。在 map 阶段,使用 job. setInputFormatClass 定义的 InputFormat 将输入的数据集分割成小数据块 splites,同时 InputFormat 提供一个 RecordReder 的实现。本例子中使用的是 TextInputFormat,他提供的 RecordReader 会将文本的字节偏移量作为 key,这一行的文本作为 value。

#### 核心总结:

1、map 最后阶段进行 partition 分区,一般使用 job. setPartitionerClass 设置的类,如果没有自定义 Key 的 hashCode()方法进行排序。

2、(第一次排序)每个分区内部调用 job. setSortComparatorClass 设置的 key 的比较函数类进行排序,如果没有则使用 Key 的实现的 compareTo 方法。 3、(第二次排序)当 reduce 接收到所有 map 传输过来的数据之后,调用

job.setSortComparatorClass 设置的 key 比较函数类对所有数据对排序,如果没有则使用 Key 的实现的 compareTo 方法。

4、紧接着使用 job. setGroupingComparatorClass 设置的分组函数类,进行分组,同一个 Key 的 value 放在一个迭代器里面。

## 【案例 4.3】mapreduce 多条数据去重处理

数据去重的最终目标是让原始数据中出现次数超过一次的数据在输出文件中 只出现一次。具体就是 reduce 的输入应该以数据作为 key,而对 value-list 则没有要求。当 reduce 接收到一个<key, value-list>时就直接将 key 复制 到输出的 key 中,并将 value 设置成空值。

#### 【案例 4.4】mapreduce 非结构化日志文件处理

需求:根据 tomcat 日志计算 url 访问了情况,具体的 url 如下,要求:区别统计 GET 和 POST URL 访问量 结果为:访问方式、URL、访问量

#### 5.MR4 MR 调优

【案例 5.1】mapreducetopN 排名

原理:可以在每一个 map 端, 就求出开始 TopN, 这样可以减少在 reduce 端的

压力,即在每一个 map 中先求出 Top10, 然后将这 top10, 传给 reduce, 让后 reduce 把所有全部的传来的数据中找出 top10

### 【案例 5.2】mapreduce 统计拨打公共服务号码的电话信息

原理: 首先对待处理的信息进行拆分,拆分之后在 map 阶段,把公共服务电 话作为 map 方法的输出键,用户电话作为 map 方法的输出值,最后在 reduce 阶段对每个键的值集合进行遍历并按指定规则拼接起来,输出结果即可。

### 【案例 5.3】mapreduce 学生信息按年龄分区

原理: 首先把学生的信息进行封装成新的类型, 作为 map'阶段的输出值, 因为不需要按键排序, 所以键只需要设置为 NullWritable 类型即可。在 reduce 阶段对所有的值进行遍历之后直接输出即可

#### 【案例 5.4】mapreduce 电话号码流量封装类作为键并排序

原理: 首先按电话号码记录信息的封装类作为键进行排序, 同时为电话记录的封装类型指定所属的数据类型(Writable), 重写排序方法和规则。

## 【案例 5.5】mapreduce 电话号码流量统计并分区

原理: 首先按电话号码作为键进行排序,相同键的内容形成一个集合,然后把相同键的所有内容值进行流量相加,最后按照指定分区条件进行分区输出。



#### 1. MR1-1 词频统计

一、项目准备阶段

1.1 在 linux 系统的命令终端上切换到<sup>~</sup>/simple 目录,执行命令:touch source.txt 创建一个文件。

```
(base) [root@ferry ~]# cd ~/simple
(base) [root@ferry simple]! touch source.txt
(base) [root@ferry simple]# ls
core-site.xml Hadoop-2.4.1 hw Hw.c mapper.sh soft test.txt
hadoop-2.4.1 HelloWorld.class Hw jdk reducer.sh source.txt word.txt
(base) [root@ferry simple]#
```

1.2 在 simple 目录下,执行命令: vi <sup>~</sup>/simple/source.txt 编辑该文件,并把数据的信息内容拷贝到该文件中,然后在 simple 目录可以查看到 source.txt 文件。



1.3 本案例需要用到 hadoop 的存储和计算,所以在编写程序之前需要先 启动 Hadoop 服务,在命令终端执行命令: start-all.sh 把 hdfs 和 yarn 服务启动。



通过 jps 命令可以看到服务已启动



| 0   |                               | lava - Eclipse        |              |                         |
|---|-------------------------------|-----------------------|--------------|-------------------------|
| File Edit Source Refa                       | actor Navigate Search Project | Run Window Help       |              |                         |
| <b></b> • <b></b> • <b></b> • <b>.</b>      | * ▼ O ▼ G ▼ # G ▼             | 😂 😂 🖋 🔻 🔪 😥 🔻 🖏 🔻     |              | Quick Access            |
| Pac New                                     |                               | ا العربي Java Project |              | Task List               |
| Open in New                                 | Window                        | Project               |              | T V 9                   |
| 🛛 🧀 En 🛛 Open Type Hi                       | ierarchy                      | F4 瞪 Package          |              | 4                       |
| 📇 show In                                   | Shift+A                       | It+W Class            |              | Find 🤞 🕨                |
| A Depy                                      | ed Name                       | Ctrl+C Ctrl+C Ctrl+C  |              | () Common               |
| Paste                                       |                               | Ctrl+V @ Annotation   |              | Connect                 |
| X Delete                                    |                               | Delete Source Folder  |              | ALM tool:<br>local task |
| . Remove from                               | Context Shift+Ctrl+Alt        | t+Down 🗳 Folder       |              | PE Outline Si           |
| Source                                      | Shift+/                       | Alt+S                 |              | oe Oddine 2             |
| Refactor                                    | Shift+A                       | Alt+T                 |              | An outline is           |
| 🔤 Import                                    |                               | 💣 Task                |              |                         |
| 🚵 Export                                    |                               | 📑 Example             |              |                         |
| A Refresh                                   |                               | F5                    |              |                         |
|   | New Jay                       | va Class              | ×            |                         |
| ava Class                                   |                               |                       | 0            |                         |
| Create a new Java                           | a class.                      |                       |              |                         |
|   |                               |                       | -            |                         |
| Source folder:                              | EnglishWordsCount/src         |                       | Browse       |                         |
| Package                                     | com wcount                    |                       | Browse       |                         |
| Fackage.                                    | com.wcounq                    |                       | Browse       |                         |
| Enclosing type                              | 6                             |                       | Browse       |                         |
|   |                               | _                     |              |                         |
| Name:                                       | WordCountMapper               |                       | 1            |                         |
| Modifiers:                                  | public O package              | ○ private ○ protected |              |                         |
|   | abstract final                | 🗆 static              |              |                         |
| Superclass:                                 | java.lang.Object              |                       | Browse       |                         |
| Interfaces:                                 |                               |                       | Add          |                         |
|   |                               |                       |              |                         |
|   |                               |                       | Remove       |                         |
| Which method stu                            | bs would you like to create   | ?                     | c            |                         |
|   | public static void main       | (String[] args)       |              |                         |
|   | Constructors from sup         | erclass               |              |                         |
|   | Inherited abstract met        | hods                  |              |                         |
|   | 0                             |                       |              |                         |
|   |                               |                       |              |                         |
| 2.3 在编写                                     | "WordCountMapp                | er" 类之前需要把            | hadoop 相关    | 的 jar 包导                |
| 入,首先右语                                      | 击EnglishWords(                | Count 项目选择"N          | [ew" — "Fol  | der"创建一                 |
| 个 lib 文件                                    | - 夹并将 hadoop 材                | 目关的 jar 包导入到          | 到此 lib 文件    | 夹中。需要的                  |
| ; 穴 []; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; |                               |                       |              |                         |
|   | - / 1 /1 1                    | /1 10 /1 1 1 10       |              |                         |
| nadoop-2.4.                                 | l/share/hadoop                | /hdfs/hadoop-hdf      | s-2.4.1. jar |                         |
| nadoop-2.4.                                 | 1/share/hadoop                | /hdfs/lib/ 所有.        | jar 包        |                         |
| nadoop-2.4.                                 | 1/share/hadoop                | /common/hadoop-c      | ommon-2.4.1. | . jar                   |
| $r^{1}$                                     | 1/share/hadoon                | /common/lih/ 師才       | 旨 iar 句      | 5                       |
| 1 0 4                                       | 1/3hare/hauoop                |                       | Jar 🖸        |                         |
| nadoop-2.4.                                 | 1/share/hadoop                | /mapreduce/ 际         |              |                         |
| nadoop-mapi                                 | reduce-examples               | -2.4.1. jar 之外的       | j jar 包      |                         |
| nadoop-2.4.                                 | 1/share/hadoop                | /mapreduce/lib/       | 所有 jar 包     |                         |
| adoon-2                                     | 1/share/hadoon                | /varn 昕友 iar 句        |              |                         |
| 1 auoop 2.4.                                | 1/ Share/ Hau00p              |                       | Þ            |                         |
| 1adoop-2.4.                                 | 1/share/hadoop                | /yarn/lib/所有 ja       | ar 包         |                         |
|   |                               |                       |              |                         |

|  | New   |                                    | Java Project           |        |                   |
|--|---|------------------------------------|------------------------|--------|-------------------|
|  | Go Into   |                                    | Project                | -      | *00*0*            |
|  | Open Type Hierarchy   |                                    | F4 G Class             | I      | Quick Access      |
| Package Explore  | Show In   | Shift+Alt+W                        | Interface              |        | 🗖 🗖 🗐 Task List 🛱 |
| 1  | Сору  | Ctrl-                              | +C C Enum              |        | 🍐 💣 🔻 🔚           |
| EnglishWordsC  | Copy Qualified Name   | Chal                               | Source Folder          |        | ~                 |
| T # com.vcoun X  | Paste<br>Delete   | Ctri-                              | te 😚 Java Working      | Set    | Find 🔏 🕨          |
| ► D WordCour S   | Bemove from Context   | Shift+Ctrl+Alt+Do                  | 📫 Folder               |        | ① Connec          |
| 🕨 🔜 JRE System L   | Build Path  | Shire Certractoo                   | File File              | ile    | Connect to        |
| 😂 hadoop 💧   | Source  | Shift+Alt+S                        | JUnit Test Case        |        | local task.       |
|  | Refactor  | Shift+Alt+T                        | Task                   | [      | 🗄 Outline छ       |
|  |   |                                    |                        |        | 0                 |
|  |   | New Folder                         |                        |        | ×                 |
| older  |   |                                    |                        |        | -                 |
| Create a new fo  | older resource.   |                                    |                        | /      |                   |
| Enter or select t  | he parent folder:   |                                    |                        |        |                   |
| EnglishWordsCo   | ount  |                                    |                        |        | _                 |
|  |   |                                    |                        |        |                   |
| 🕨 😂 EnglishWor   | dsCount   |                                    |                        |        |                   |
| Addoop   |   |                                    |                        |        |                   |
|  |   |                                    |                        |        |                   |
|  |   |                                    |                        |        |                   |
| Folder name: li  | ib  |                                    |                        |        |                   |
| Folder name: li  | ib  |                                    |                        |        |                   |
| Folder name: li<br>Advanced >>   | ib  |                                    |                        |        |                   |
| Folder name: Ii<br>Advanced >>   | ib  |                                    |                        |        |                   |
| Folder name: Ii<br>Advanced >>   | ы   |                                    |                        |        |                   |
| Folder name: I   | ib  |                                    |                        |        |                   |
| older name: Ii<br>Advanced >>  | ib  |                                    | Cancel                 | Finish |                   |
| Folder name: [li<br>Advanced >><br>②<br>記成导入,  | 可以查看如   | 下图所示:                              | Cancel                 | Finish |                   |
| <sup>■</sup> older name: [i<br>Advanced >><br>?<br>完成导入,   | 可以查看如   | 下图所示:                              | Cancel                 | Finish |                   |
| Folder name: [i<br>Advanced >><br>②<br>完成导入,<br>(base) [root@f<br>(base) [root@f   | 时<br>可以查看如<br>ferry ~]# cd ~/w<br>ferry lib]# ls  | 下图所示:<br>orkspace/EnglishWo        | Cancel<br>rdsCount/lib | Finish |                   |
| ■older name: [i<br>Advanced >><br>名成导入,<br>(base) [root@f<br>base) [root@f<br>ctivation-1]   | 时以查看如<br>ferry ~]# cd ~/w<br>ferry lib]# ls<br>l.jar  | 下图所示:<br>orkspace/EnglishWo        | Cancel<br>rdsCount/lib | Finish |                   |
| Folder name: [i<br>Advanced >><br>名成导入,<br>(base) [root@f<br>(base) [root@f<br>(ctivation-1.1<br>appalliance-1.1<br>appalliance-1.2<br>appalliance-1.2   | 时以查看如<br>ferry ~]# cd ~/w<br>ferry lib]# ls<br>l.jar<br>.0.jar  | 下图所示:<br>orkspace/EnglishWo        | Cancel<br>rdsCount/lib | Finish |                   |
| Folder name: [i<br>Advanced >><br>名成导入,<br>base) [root@f<br>(base) [root@f<br>ctivation-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.3<br>appalliance-1.1<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appalliance-1.3<br>appallia   | 时以查看如<br>ferry ~]# cd ~/w<br>ferry lib]# ls<br>l.jar<br>0.jar   | 下图所示:<br>orkspace/EnglishWo        | Cancel<br>rdsCount/lib | Finish |                   |
| Folder name: [i<br>Advanced >><br>名成导入,<br>(base) [root@f<br>(base) [root@f<br>activation-1.1<br>aopalliance-1.1<br>aopalliance-1.1<br>aopalliance-1.1<br>aopalliance-1.1<br>aopalliance-1.1<br>aopalliance-1.1<br>aopalliance-1.1<br>aopalliance-1.1<br>aopalliance-1.1   | 时以查看如<br>ferry ~]# cd ~/w<br>ferry lib]# ls<br>l.jar<br>0.jar   | 下图所示:<br>orkspace/EnglishWo        | Cancel                 | Finish |                   |
| Folder name: [i<br>Advanced >><br>充成导入,<br>(base) [root@f<br>(base) [root@f<br>activation-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appalliance-1.1<br>appall   | 时以查看如<br>ferry ~]# cd ~/w<br>ferry lib]# ls<br>l.jar<br>0.jar<br>r<br>tils-1.7.0.jar<br>tils-2.ore-1.8.0.<br>2 jar  | 下图所示:<br>orkspace/EnglishWo        | Cancel<br>rdsCount/lib | Finish |                   |
| Folder name: [i<br>Advanced >><br>名成导入,<br>(base) [root@f<br>(base) [root@f] [root@f] [root@f<br>(base) [root@f] [root@f   | 时以查看如<br>ferry ~]# cd ~/w<br>ferry lib]# ls<br>l.jar<br>0.jar<br>rtils-1.7.0.jar<br>tils-1.8.0.<br>.2.jar   | 下图所示:<br>orkspace/EnglishWo<br>jar | Cancel<br>rdsCount/lib | Finish |                   |
| Folder name: [i<br>Advanced >><br>名成导入,<br>(base) [root@f<br>(base) [root@f] [root@f] [root@f<br>(base)  | 时以查看如<br>ferry ~]# cd ~/w<br>ferry lib]# ls<br>l.jar<br>0.jar<br>frils-1.7.0.jar<br>tils-1.7.0.jar<br>tils-1.8.0.<br>.2.jar<br>-1.4.jar<br>:tions-3.2.1.jar   | 下图所示:<br>orkspace/EnglishWo<br>jar | Cancel<br>rdsCount/lib | Finish |                   |
| Folder name: [i<br>Advanced >><br>名dvanced >><br>名成导入,<br>(base) [root@f<br>(base) [root   | 时以查看如<br>ferry ~]# cd ~/w<br>ferry lib]# ls<br>l.jar<br>0.jar<br>rtils-1.7.0.jar<br>tils-1.7.0.jar<br>tils-core-1.8.0.<br>.2.jar<br>-1.4.jar<br>ttions-3.2.1.jar  | 下图所示:<br>orkspace/EnglishWo<br>jar | Cancel<br>rdsCount/lib | Finish |                   |
| Folder name: [i<br>Advanced >><br>⑦<br>完成导入,<br>(base) [root@f<br>activation-1.1<br>apalliance-1.<br>asm-3.2.jar<br>avro-1.7.4.jar<br>commons-beanut<br>commons-beanut<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection<br>commons-colection   | 时以查看如<br>ferry ~]# cd ~/w<br>ferry lib]# ls<br>l.jar<br>0.jar<br>fils-1.7.0.jar<br>lils-core-1.8.0.<br>2.jar<br>1.4.jar<br>titons-3.2.1.jar<br>ss-1.4.1.jar<br>puration-1.6.jar   | 下图所示:<br>orkspace/EnglishWo<br>jar | Cancel<br>rdsCount/lib | Finish |                   |
| Folder name: [i<br>Advanced >><br>名dvanced >><br>②<br>完成导入,<br>(base) [root@f<br>(base) [root@f<br>(base) [root@f<br>(base) [root@f<br>(base) [root@f<br>activation-1.1<br>apalliance-1.<br>asm-3.2.jar<br>avro-1.7.4.jar<br>commons-beanut<br>commons-beanut<br>commons-collect<br>commons-collect<br>commons-collect<br>commons-collect<br>commons-collect<br>commons-collect<br>commons-collect<br>commons-collect<br>commons-collect<br>commons-collect<br>commons-collect<br>commons-collect<br>commons-collect<br>commons-collect<br>commons-collect<br>commons-collect<br>commons-collect<br>commons-collect<br>commons-collect<br>commons-collect<br>commons-collect<br>commons-collect<br>commons-collect<br>commons-collect<br>commons-collect   | 时以查看如<br>ferry ~]# cd ~/w<br>ferry lib]# ls<br>l.jar<br>0.jar<br>fr<br>tils-1.7.0.jar<br>tils-core-1.8.0.<br>.2.jar<br>1.4.jar<br>tions-3.2.1.jar<br>esss-1.4.1.jar<br>juration-1.6.jar<br>1-1.0.13.jar<br>ter-1.8.jar  | 下图所示:<br>orkspace/EnglishWo        | Cancel<br>rdsCount/lib | Finish |                   |
| Folder name: [i<br>Advanced >><br>名成导入,<br>(base) [root@f<br>(base) [root@f<br>( | 时以查看如<br>ferry ~]# cd ~/w<br>ferry lib]# ls<br>l.jar<br>0.jar<br>tils-1.7.0.jar<br>tils-core-1.8.0.<br>.2.jar<br>1.4.jar<br>tilos-3.2.1.jar<br>ess-1.4.1.jar<br>juration-1.6.jar<br>h-1.0.13.jar<br>ter-1.8.jar   | 下图所示:<br>orkspace/EnglishWo        | Cancel<br>rdsCount/lib | Finish |                   |
| Folder name: [i<br>Advanced >><br>名dvanced >><br>名成导入,<br>base) [root@f<br>base) [root@f<br>base) [root@f<br>ctivation-1.1<br>opalliance-1.1<br>opalliance-1.1<br>commons-beanut<br>commons-beanut<br>commons-cli-1.2<br>commons-colled<br>commons-config<br>commons-config<br>commons-digest<br>commons-digest<br>commons-ling()  | 时以查看如<br>ferry ~]# cd ~/w<br>ferry lib]# ls<br>.jar<br>.0.jar<br>tils-1.7.0.jar<br>tils-core-1.8.0.<br>.2.jar<br>1.4.jar<br>:tions-3.2.1.jar<br>tess-1.4.1.jar<br>ter-1.8.jar<br>.jar<br>.jar<br>ter-1.8.jar  | 下图所示:<br>orkspace/EnglishWo        | Cancel<br>rdsCount/lib | Finish |                   |
| Folder name: [ii<br>Advanced >><br>名成导入,<br>base) [root@f<br>base) [root@f<br>base] [root@f   | 可以查看如<br>ferry ~]# cd ~/w<br>ferry lib]# ls<br>l.jar<br>.0.jar<br>tils-1.7.0.jar<br>tils-core-1.8.0.<br>2.jar<br>ss-1.4.jar<br>tiions-3.2.1.jar<br>ss-1.4.jar<br>uration-1.6.jar<br>h-1.0.13.jar<br>ter-1.8.jar<br>).jar<br>lient-3.1.jar<br>4.jar<br>2.6.jar                                       | 下图所示:<br>orkspace/EnglishWo<br>jar | Cancel<br>rdsCount/lib | Finish |                   |
| Folder name: [ii<br>Advanced >><br>名成导入,<br>志成导入,<br>base) [root@f<br>base) [root@f<br>base]  | 时以查看如<br>ferry ~]# cd ~/w<br>ferry lib]# ls<br>l.jar<br>entils-1.7.0.jar<br>tils-1.7.0.jar<br>tils-core-1.8.0.<br>2.jar<br>1.4.jar<br>tiions-3.2.1.jar<br>tiions-3.2.1.jar<br>tes-1.4.1.jar<br>puration-1.6.jar<br>h-1.0.13.jar<br>ter-1.8.jar<br>lient-3.1.jar<br>t.jar<br>2.6.jar<br>ng-1.1.3.jar | 下图所示:<br>orkspace/EnglishWo<br>jar | Cancel<br>rdsCount/lib | Finish |                   |

深圳大学学生实验报告用纸









| Pack       | Go Into                        |                                  | Project          |       |
|------------|--------------------------------|----------------------------------|------------------|-------|
|            | Open in New Window             | 100                              | Package          |       |
|            | Open Type Hierarchy            | FA C                             | Class            |       |
| 3 Enç      | Show In                        | Shift+Alt+W                      | Interfa          |       |
| 🦀 sr       | Copy                           | Ctrl+C 🕼                         | Enum             |       |
|            | Copy Qualified Name            | eurre                            | Annotation       | ongWr |
|            | Paste                          | Ctrl+V 🇳                         | Source Folder    | ue,   |
|            | Delete                         | Delete 😫                         | Java Working Set | >.Con |
| Blile      | Remove from Context            | Shift+Ctrl+Alt+Down              | Folder           |       |
|            | Build Path                     |                                  | File             |       |
|            | Source                         | Shift+Alt+S 🔹 📴                  | IUnit Test Case  |       |
|            | Refactor                       | Shift+Alt+T 🕨 🦷                  | Task             |       |
| <b>e</b> . |                                |                                  |                  |       |
|            | Net                            | w Java Class                     | ×                |       |
| va Class   | 5                              |                                  | a                |       |
| reate a n  | iew Java class.                |                                  |                  |       |
|            |                                |                                  |                  |       |
| ource fold | der: EnglishWordsCount/        | src                              | Browse           |       |
| ackage:    | com.wcount                     |                                  | Browse           |       |
| Enclosi    | ng type:                       |                                  | Browse           |       |
|            |                                | -                                |                  |       |
| ame:       | WordCountReducer               |                                  | J                |       |
| lodifiers: | public Opac                    | kage O private O protecte        | d                |       |
|            | 🗌 abstract 🗌 final             | static                           |                  |       |
| uperclass  | ; java.lang.Object             | Δ.                               | Browse           |       |
| nterfaces: |                                |                                  | Add              |       |
|            |                                |                                  | Remove           |       |
|            |                                |                                  |                  |       |
| /hich met  | hod stubs would you like to cr | eate?                            |                  |       |
|            | Constructors from              | main(String[] args)              |                  |       |
|            | Inherited abstract             | methods                          |                  |       |
| io you wa  | nt to add comments? (Configu   | re templates and default value h | ere)             |       |
| -          | Generate comme                 | nts                              |                  |       |
|            |                                |                                  |                  |       |
|            |                                |                                  |                  |       |
|            |                                |                                  |                  |       |
|            |                                |                                  |                  |       |
| 2          |                                | Cancel                           | Finish           |       |
| 2          |                                | Cancel                           | Finish           |       |









| <pre>File Edit View Search Terminal Help (base) [root@ferry ~]# c1 ~/simple (base) [root@ferry simple]# cd output (base) [root@ferry output]# ls part-r-00000 _SUCCESS (base) [root@ferry output]# cat ^C (base) [root@ferry output]# cat part-r-00000 44 (base) [root@ferry output]#</pre>   |
|---|
| 3.MR2-2 成绩统计<br>【案例 3.1】 mapreduce 学生总成绩报表  |
| 一、项目准备阶段<br>需求描述:   |
| □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□  |
| 和他相应的成绩,如果有多门学科,则每门学科为一个文件。   |
| 要求在输出中每行有两个间隔的数据,其中,第一个代表学生的姓名,第二个代表其   |
| 总成绩。  |
| 原始数据:   |
| 张三 88   |
| 李四 99   |
| 土山 00   |
| 此八 //   |
| <u>水</u> 78<br>李四 89  |
| 王五 96   |
| 赵六 67   |
| 张三 80   |
| 李四 82   |
| 王五 84   |
| 赵六 86   |
| <pre>(base) [root@ferry ~]# cd ~/simple<br/>(base) [root@ferry simple]# ls<br/>core-site.xml HelloWorld.class Hw.c output source.txt<br/>hadoop-2.4.1 hw jdk reducer.sh test.txt<br/>Hadoop-2.4.1 Hw mapper.sh soft word.txt<br/>(base) [root@ferry simple]# cat source.txt<br/>张三 88<br/>李四 99<br/>王五 66<br/>赵六 77<br/>张三 78<br/>李四 89<br/>王五 96<br/>赵六 67<br/>张三 80<br/>李四 82<br/>王五 84<br/>赵六 86</pre> |
|   |

| ( · • ·                  |  | New Java  | a Project   |   | ×             |
|--------------------------|--|---|---|---|---------------|
| cr                       | reate a Java Proje   | ct  |   | _   | 1             |
| C                        | reate a Java project   | in the workspace or in a  | an external location.                               | 1   | 7             |
|                          |  |   |   |   |               |
|                          | Project name Stud  | lentTotalScore  |   |   | ]             |
|                          | Vise default local   | tion  |   |   |               |
|                          | Location: /root/wor  | rkspace/StudentTotalSco   | re  | Browse  | 1             |
|                          | IRE  |   |   |   |               |
|                          | () 11-2  |   | laures a p  |   |               |
|                          | O Use an execution   | on environment JRE:   | Javase-1.8  | Ŷ.  |               |
|                          | Use a project s  | pecific JRE:  | jdk1.8.0_241  |   |               |
|                          | <ul> <li>Use default JRE</li> </ul>  | currently 'jdk1.8.0_24  | (1')  | Configure JREs  |               |
|                          | Project layout   |   |   |   |               |
|                          | O Use project fol  | der as root for sources a   | and class files                                     |   |               |
|                          | <ul> <li>Create separat</li> </ul>   | e folders for sources an  | d class files                                       | Configure default   |               |
|                          | Marking sets   |   |   |   |               |
|                          | Add project to r   | working sets  |   |   |               |
|                          | Working sets   | norking sets  |   | A Select  |               |
|                          | Working sets.  |   |   | T Sciectin  |               |
| 2 在项目 srd                | <b>⑦</b><br>c 目录下,   | Lange Back No.  | ext >   Cancel<br>译"新建",                            | <b>Finish</b><br>创建一个类  | 文件名称为         |
| 2 在项目 srd<br>CotalScoreM | ⑦<br>c 目录下,<br>Ⅰapper"并  | ▲ Back Net  | ext >   Cancel<br>译 "新建"<br>com. simple.            | Finish<br>创建一个类<br>. total. sco   | 文件名称为<br>re"。 |
| 2 在项目 srd<br>fotalScoreM | ②<br>c 目录下,<br>fapper"并  | Contract Co | ext > Cancel<br>译"新建"<br>com. simple.               | <mark>Finish</mark><br>创建一个类<br>. total. sco<br>∝                                     | 文件名称为<br>re"。 |
| 2 在项目 srd<br>TotalScoreM | ② c 目录下,<br>Mapper"并 Java Class  | A Back New Jave 右键点击,选<br>指定包名" New Jave  | ext > Cancel<br>译"新建",<br>com. simple.              | Finish<br>创建一个类<br>. total. sco   | 文件名称为<br>re"。 |
| 2 在项目 src<br>fotalScoreM | ②<br>c 目录下,<br>Mapper"并<br>Java Class<br>Create a new Java c   | ass.  | ext > Cancel<br>择"新建"<br>com. simple.               | Finish<br>创建一个类<br>. total. sco<br>〇  | 文件名称为<br>re"。 |
| 2 在项目 srd<br>TotalScoreM | ②<br>c 目录下,<br>Mapper"并<br>Java Class<br>Create a new Java cl<br>Source folder:  | A 键点击,选<br>指定包名" New Jav lass. StudentTotalScore/src  | ext > Cancel<br>译"新建",<br>com. simple.              | Finish<br>创建一个类<br>total.sco  | 文件名称头<br>re"。 |
| 2 在项目 src<br>fotalScoreM | ②<br>c 目录下,<br>fapper"并<br>Java Class<br>Create a new Java c<br>Source folder:<br>Package:   | < Back No<br>右键点击,选<br>指定包名"<br>New Jav<br>lass.<br>StudentTotalScore/src<br>com.simple.total.score/  | ext > Cancel<br>译"新建",<br>com. simple.              | Finish<br>创建一个类<br>total.sco<br>©<br>Browse<br>Browse                                 | 文件名称为<br>re"。 |
| 2 在项目 sre                | ②<br>c 目录下,<br>Mapper"并<br>Java Class<br>Create a new Java c<br>Source folder:<br>Package:   | A 翻 A 翻 A 翻 A 翻 A 翻 A 翻 A 翻 A 翻 A 翻 A 翻   | ext > Cancel<br>译"新建"<br>com. simple.               | Finish<br>创建一个类<br>total.sco<br>Browse<br>Browse<br>Browse                            | 文件名称为<br>re"。 |
| 2 在项目 srd<br>TotalScoreM | ⑦ c 目录下, fapper"并 Java Class Create a new Java class Source folder: Package: Enclosing type: Name:   | < Back Not<br>右键点击,选<br>指定包名"<br>New Jav<br>lass.<br>StudentTotalScore/src<br>com.simple.total.score  | ext > Cancel<br>译"新建"<br>com. simple.               | Finish<br>创建一个类<br>total.sco<br>Browse<br>Browse<br>Browse                            | 文件名称头<br>re"。 |
| 2 在项目 sre<br>fotalScoreM | ②<br>c 目录下,<br>fapper"并<br>Java Class<br>Create a new Java c<br>Source folder:<br>Package:<br>□ Enclosing type:<br>Name:<br>Modifiers:   | < Back Not A Control of the second seco     | ext > Cancel<br>译"新建",<br>com. simple.<br>ra Class  | Finish<br>创建一个类<br>total.sco<br>Browse<br>Browse<br>Browse                            | 文件名称头<br>re"。 |
| 2 在项目 srd<br>TotalScoreM | ②<br>c 目录下,<br>Mapper"并<br>Java Class<br>Create a new Java cl<br>Source folder:<br>Package:<br>□ Enclosing type:<br>Name:<br>Modifiers:  | < Back Not Control of Contro     | ext > Cancel<br>译作"新建",<br>com. simple.<br>ra Class | Finish<br>创建一个类<br>total.sco<br>Browse<br>Browse<br>Browse                            | 文件名称为<br>re"。 |
| 2 在项目 sro<br>SotalScoreM | <ul> <li>②</li> <li>c 目录下, fapper"并</li> <li>Java Class</li> <li>Java Class</li> <li>Create a new Java class</li> <li>Source folder:</li> <li>Package:</li> <li>Enclosing type:</li> <li>Name:</li> <li>Modifiers:</li> <li>Superclass:</li> </ul>   | A Back New Jav<br>右键点击,选<br>指定包名"          New Jav         lass.         StudentTotalScore/src         com.simple.total.score         TotalScoreMapper         public       package         abstract       final         java.lang.Object   | ext > Cancel<br>译 "新建"<br>com. simple.<br>a Class   | Finish<br>创建一个类<br>total.sco<br>Browse<br>Browse<br>Browse                            | 文件名称头<br>re"。 |
| 2 在项目 srd<br>TotalScoreM | <ul> <li>②</li> <li>c 目录下, fapper"并</li> <li>Java Class</li> <li>Create a new Java class</li> <li>Interfaces:</li> </ul>                                   | <pre>&lt; Back New A 右键点击,选 指定包名" New Jav lass. StudentTotalScore/src com.simple.total.score  TotalScoreMapper public</pre>   | ext > Cancel<br>译 "新建",<br>com. simple.<br>a Class  | Finish<br>创建一个类<br>total.sco<br>Browse<br>Browse<br>Browse<br>Browse<br>Add           | 文件名称为<br>re"。 |
| 2 在项目 sro<br>FotalScoreM | ②<br>c 目录下,<br>Apper"并<br>Java Class<br>Create a new Java cl<br>Source folder:<br>Package:<br>Enclosing type:<br>Name:<br>Modifiers:<br>Superclass:<br>Interfaces:   | < Back New Jav<br>右键点击,选<br>指定包名"<br>New Jav<br>lass.          StudentTotalScore/src         com.simple.total.score         TotalScoreMapper         public       package         abstract       final         java.lang.Object   | ext > Cancel<br>注择"新建"<br>com. simple.<br>ra Class  | Finish<br>创建一个类<br>total.sco<br>Browse<br>Browse<br>Browse<br>Browse<br>Browse        | 文件名称为<br>re"。 |
| 2 在项目 sro<br>FotalScoreM | <ul> <li>②</li> <li>c 目录下,<br/>fapper"并</li> <li>Java Class<br/>Create a new Java cl<br/>Source folder:<br/>Package:</li> <li>D Enclosing type:</li> <li>Name:<br/>Modifiers:</li> <li>Superclass:<br/>Interfaces:</li> <li>Which method stubs</li> </ul>                                      | A Back Net A Constraint of the second se    | ext > Cancel<br>译 "新建"<br>com. simple.<br>a Class   | Finish<br>创建一个类<br>total.sco<br>Browse<br>Browse<br>Browse<br>Browse<br>Remove        | 文件名称头<br>re"。 |
| 2 在项目 sro<br>FotalScoreM | <ul> <li>②</li> <li>c 目录下,<br/>lapper"并</li> <li>Java Class</li> <li>Create a new Java class</li> <li>Super class:</li> <li>Interfaces:</li> <li>Which method stubs</li> </ul>              | A 键点击,选指定包名"          右键点击,选         指定包名"         New Jav         lass.         StudentTotalScore/src         com.simple.total.score/         public ● package         abstract ● final         java.lang.Object         would you like to create?         ● public static void main()   | ext > Cancel<br>译作"新建",<br>com. simple.<br>a Class  | Finish<br>创建一个类<br>total.sco<br>Browse<br>Browse<br>Browse<br>Browse<br>Add<br>Remove | 文件名称头<br>re"。 |
| 2 在项目 sro<br>FotalScoreM | ②<br>c 目录下,<br>Apper"并<br>Java Class<br>Create a new Java cl<br>Source folder:<br>Package:<br>Enclosing type:<br>Name:<br>Modifiers:<br>Superclass:<br>Interfaces:<br>Which method stubs   | A 键点击,选指定包名" A 键点击,选指定包名" New Jav lass. StudentTotalScore/src com.simple.total.score TotalScoreMapper public ● package abstract ● final java.lang.Object would you like to create? Public static void main@ Constructors from supe Identified abstract area   | ext > Cancel<br>译择"新建"<br>com. simple.<br>ra Class  | Finish<br>创建一个类<br>total.sco<br>Browse<br>Browse<br>Browse<br>Browse<br>Add<br>Remove | 文件名称为<br>re"。 |
| 2 在项目 sro<br>FotalScoreM | <ul> <li>②</li> <li>c 目录下, f</li> <li>fapper"并</li> <li>Java Class</li> <li>Greate a new Java class</li> <li>Create a new Java class</li> <li>Create a new Java class</li> <li>Greate a new Java class</li> <li>Interfaces:</li> <li>Which method stubs</li> <li>Do you want to add</li> </ul> | A Back Net A Construction of the second     | ext > Cancel<br>译择"新建"<br>com. simple.<br>a Class   | Finish<br>创建一个类<br>total.sco<br>Browse<br>Browse<br>Browse<br>Browse<br>Add<br>Remove | 文件名称为<br>re"。 |
| 2 在项目 sro<br>FotalScoreM | ② C 目录下, fapper"并 Java Class Create a new Java class Super class: Interfaces: Which method stubs Do you want to add class  | A Back New Jav<br>右键点击,选<br>指定包名"<br>New Jav<br>lass.          StudentTotalScore/src         com.simple.total.score/         TotalScoreMapper         ● public         ● constructors from supe         ● Inherited abstract mett         ○ Generate comments  | ext > Cancel<br>译 "新建"<br>com. simple.<br>a Class   | Finish<br>创建一个类<br>total.sco<br>Browse<br>Browse<br>Browse<br>Browse<br>Add<br>Remove | 文件名称为<br>re"。 |
| 2 在项目 sro<br>FotalScoreM | ②<br>c 目录下,<br>lapper"并<br>Java Class<br>Create a new Java cl<br>Source folder:<br>Package:<br>Enclosing type:<br>Name:<br>Modifiers:<br>Superclass:<br>Interfaces:<br>Which method stubs<br>Do you want to add  | A Back New Jav<br>右键点击,选<br>指定包名"<br>New Jav<br>lass.<br>StudentTotalScore/src<br>com.simple.total.score<br>public package<br>abstract final<br>java.lang.Object<br>would you like to create?<br>public static void main(<br>Constructors from supe<br>Inherited abstract mett<br>comments? (Configure ter<br>Generate comments   | ext > Cancel<br>译 "新建",<br>com. simple.<br>a Class  | Finish<br>创建一个类<br>total.sco<br>Browse<br>Browse<br>Browse<br>Browse<br>Add<br>Remove | 文件名称为<br>re"。 |
| 2 在项目 sro<br>FotalScoreM | ②<br>c 目录下,<br>fapper"并<br>Java Class<br>Create a new Java cl<br>Source folder:<br>Package:<br>Enclosing type:<br>Name:<br>Modifiers:<br>Superclass:<br>Interfaces:<br>Which method stubss<br>Do you want to add   | < Back N4 右键点击,选指定包名" New Jav lass. StudentTotalScore/src com.simple.total.score TotalScoreMapper ● public ● package ■ abstract ● final java.lang.Object would you like to create? ● public static void main( Constructors from supe ✓ Inherited abstract meth comments? (Configure ter ● Generate comments   | ext > Cancel<br>译择"新建"<br>com. simple.<br>a Class   | Finish<br>创建一个类<br>total.sco<br>Browse<br>Browse<br>Browse<br>Browse<br>Add<br>Remove | 文件名称为<br>re"。 |





2.6 在项目 src 目录下指定的包名" com. simple. total. score"下右键点 击,新建一个类名为"TotalScoreReducer"并继承 Reducer 类,然后添加 该类中的代码内容如下所示。



2.7 在项目 src 目录下指定的包名" com. simple. total. score"下右键点 击,新建一个测试主类名为" TotalScoreJob"并指定 main 主方法。





| (base) [root@ferry ~]# cd ~/simple         (base) [root@ferry simple]# ls         core-site.xml HelloWorld.class Hw         hadoop-2.4.1 Hw         Hadoop-2.4.1 Hw         (base) [root@ferry simple]# cat sou         张三       88         李四       99         玉五       66         赵六       77         张三       78         李四       89         玉五       66         赵六       67         张三       88         李四       89         玉五       96         赵六       67         张三       86         二、程序编写       2.1 在 eclipse 中的项目列表中, 在         Project…" 新建一个项目 "Studen | fi键点击,选择"new"一>"Java   |
|--|--|
| • New  | Java Project 🛞   |
| Create a Java Project<br>Create a Java project in the workspace o  | r in an external location.   |
| Project name StudentAvgScore<br>© Use default location<br>Location: /root/workspace/StudentAvg<br>JRE<br>© Use an execution environment JRE<br>© Use a project specific JRE:<br>© Use default JRE (currently 'jdk1.8.4)<br>Project layout<br>© Use project folder as root for source<br>Working sets<br>© Add project to working sets<br>Working sets:   | Score Browse<br>: JavaSE-1.8 *<br>k jdk1.8.0_241 *<br>D_241') Configure JREs<br>ces and class files<br>s and class files Configure default |
| (?) < Back   | Next > Cancel Finish   |
| 2.2 在项目 src 目录下,右键点击,<br>"AvgScoreMapper"并指定包名"  | 选择"新建"创建一个类文件名称为<br>com.simple.avg.score"。   |





2.6 在项目 src 目录下指定的包名" com. simple. avg. score"下右键点击, 新建一个类名为"AvgScoreReducer "并继承 Reducer 类, 然后添加该类中 的代码内容如下所示。

| 😫 Package Explore  | er 12 🗖 🗖  | *AvgScoremapper.java  *AvgScoreReducer.java  |
|--|--|--|
| <ul> <li>EnglishWordsC</li> <li>hadoop</li> <li>StudentAvgScc</li> <li>StudentAvgScc</li> <li>StudentAvgSccret</li> <li>AvgScoret</li> <li>AvgScor</li></ul> | count<br>re<br>avg.score<br>mapper.java<br>re<br>avg.score<br>mapper.java<br>re<br>avg.score<br>mapper.java<br>re<br>(jdk1.8.0_2<br>ar - //rot/simple/jdk/jdk1.8<br>pot/simple/jdk/jdk1.8<br>pot/simple/jdk/jdk1.9<br>avg.score<br>avg.score<br>ar - //rot/simple/jdk/jdk1.9<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.score<br>avg.sc | <pre>package com.simple.avg.score;<br/>import java.util.Iterator;<br/>import java.util.Iterator;<br/>import org.apache.hadoop.io.Text;<br/>import org.apache.hadoop.mapreduce.Reducer;<br/>public class AvgScoreReducer extends<br/>ReducerText, IntWritable, Text, IntWritable&gt; {<br/>@@Override<br/>protected void reduce(Text key, Iterable<intwritable> values,Context context)<br/>throws IOException, InterruptedException {<br/>//KRDyHu值集合编历对象<br/>I Tterator<intwritable> iterator = values.iterator();<br/>int count = 0;<br/>int sum =0;<br/>//@FAREWENDHOMSMARK()){<br/>int v = iterator.next().get();<br/>sum += v;<br/>count++;<br/>}<br/>int avg = sum/count;<br/>context.write(key, new IntWritable(avg));<br/>}<br/>//WIMENDHOMSMARK</intwritable></intwritable></pre> |
| →<br>新建一个测   | 目 src 日求<br>则试主类名〉   | 卜指定的包名" com.simple.avg.score" 卜石键点击,<br>勺" AvgScoreJob"并指定 main 主方法。   |
| 2.1 在坝F<br>新建一个》   | l src 日求<br>则试主类名グ   | ト指定的包名" com.simple.avg.score"ト右键点击,<br>为" AvgScoreJob"并指定 main 主方法。  |
| 新建一个》  | 目 src 日求<br>则试主类名<br>J <b>ava Class</b><br>Create a new Jav  | ト指定的包名" com. simple. avg. score"ト右键点击,<br>为" AvgScoreJob"并指定 main 主方法。<br>New Java Class   |
| 新建一个》  | I STC 日求<br>则试主类名之<br>Java Class<br>Create a new Jav<br>Source folder:<br>Package:<br>Enclosing type   | ト指定的包名" com. simple. avg. score"ト右键点击,<br>为" AvgScoreJob"并指定 main 主方法。<br>New Java Class a class.  StudentAvgScore/src Browse Browse Browse Browse Browse Browse Browse  |
| 2.7 在项F<br>新建一个》   | I STC 日求<br>则试主类名之<br>Java Class<br>Create a new Jav<br>Source folder:<br>Package:<br>Enclosing type<br>Name:<br>Modifiers:  | ▶指定的包名" com. simple. avg. score" 卜石键点击, 为" AvgScoreJob"并指定 main 主方法。 New Java Class a class. StudentAvgScore/src Browse Com.simple.avg.score Browse   |

Remove

Finish

Cancel

2!

Which method stubs would you like to create?

public static void main(String[] args) Constructors from superclass Inherited abstract methods Do you want to add comments? (Configure templates and default value here)

Generate comments

?

2.8 测试代码如下:







发现在项目下多一个列表项"Referenced Libraries"。如图7所示





2.6 在项目 src 目录下指定的包名" com. simple. intsort"下右键点击,新 建一个类名为"IntSortReducer"并继承 Reducer类,然后添加该类中的代 码内容如下所示。





| <pre>e.Job (Job.java:monitorAndPrintJob(1355)) - Job job local419858282 8001 running in uber mode<br/>e.Job (Job.java:monitorAndPrintJob(1362)) - map 100% reduce 100%<br/>e.Job (Job.java:monitorAndPrintJob(1373)) - Job Job_tocat419858282_8001 completed successfull<br/>e.Job (Job.java:monitorAndPrintJob(1380)) - Counters: 33</pre> |
|---|
| 2.10 程序执行完毕之后,可以到输出信息目录/simple/output下,执行查看命令:cat part-r-00000 查看对数据处理后产生的结果。   |
| <pre>(base) [root@ferry ~]# cd ~/simple<br/>(base) [root@ferry simple]# cd output<br/>(base) [root@ferry output]# ls<br/>reat = 00000CUCCTCC</pre>  |
| (base) [root@ferry output]# cat part-r-00000  |
| 1 2   |
| 2 6   |
| 4 22  |
| 5 26  |
| 6 32<br>7 54  |
| 8 92  |
| 9 650   |
| 10 654  |
| 12 5956   |
| 13 65223  |
| (base) [root@terry output]#   |
| 【案例 4.2】mapreduce 两列数据的二次排序<br>一、项目准备阶段   |
| 箭 <b>水:</b><br>对西列教提进行排 <b>点,第一列相同的</b> 比较第二列   |
| 而此数据,   |
| $20 \ 21$   |
| 50 51   |
| 50 52   |
| 50 53   |
| 50 54   |
| 60 51   |
| 60 53   |
| 60 52<br>60 FC  |
| 60 50<br>60 57  |
| 70 58   |
| 60 61   |
| 70 54   |
| 70 55   |
| 70 56   |
| 70 57   |

| 70 58  |
|--|
| 1 2  |
| 3 4  |
| 5 6  |
| 7 82   |
| 203 21   |
| 50 512   |
| 50 522   |
| 50 53  |
| 530 54   |
| 40 511   |
| 20 53  |
| 20 522   |
| 60 56  |
| 60 57  |
| 740 58   |
| 63 61  |
| 730 54   |
| 71 55  |
| 71 56  |
| 73 57  |
| 74 58  |
| 12 211   |
| 31 42  |
| 50 62  |
| 7 8  |
| (base) [root@ferry output] + cd ~/simple               |
| <pre>(base) [root@ferry simple] + cat source.txt</pre> |
|  |
| 20 21  |
| 50 52  |
| 50 53  |
| 50 54  |
| 60 51<br>60 52   |
| 60 52  |
| 60 56  |
| 60 57  |
| 70 58  |
| 70 54  |
| 70 55  |
| 70 56  |
|  |
|  |
| 2.1 仕 eclipse 甲的坝目列表甲,石键点击,选择"new"一>" Java             |



深圳大学学生实验报告用纸

2.3 在项目 src 目录下,右键点击包名" com. simple. secondarysort"选择"New"—"Class"创建一个类文件名称为"FirstPartitioner",该类 是对数据处理后的结果进行分区设置。代码实现如下:

😫 Package Explorer 🕴 📮 🗊 👔 \*IntPair.java\_ 👔 \*FirstPartition 🕱 📝 \*GroupingCompar 33 - 0 age com.simple.secondarysort; P 5 import org.apache.hadoop.io.Text; EnglishWordsCount Þ import org.apache.hadoop.mapreduce.Partitioner; ÷ 分区函数类 IntSortCode · C SecondarySort public class FirstPartitioner extends Partitioner<IntPair,Text>{ V 🕮 STC @Override public int getPartition(IntPair key, Text value, int numPartitions) { com.simple.secondarysort return Math.abs(key.getFirst().hashCode() \* 127) % numPartitions; ▶ J FirstPartitioner.java GroupingComparator.java 🕨 🗊 IntPair.java SecondarySortMapper.jav ▶ ➡ JRE System Library [jdk1.8.0 StudentAvgScore StudentTotalScore 2.5 在项目 src 目录下,右键点击包名" com. simple. secondarysort"选 择"New"一"Class"创建一个类文件名称为"SecondarySortMapper"。 New Java Class Java Class Create a new Java class. Source folder: SecondarySort/src Browse. Package: com.simple.secondarysort Browse. D Enclosing type: SecondarySortMapper Name: Modifiers: private O protected 🗌 abstract 🗌 final Superclass: java.lang.Object Browse .... Interfaces: Add .. Which method stubs would you like to create? public static void main(String[] args) Constructors from superclass Inherited abstract methods Do you want to add comments? (Configure templates and default value here) Generate comments ? Cancel Finish 2.6 在编写 "SecondarySort Mapper" 类之前需要把 hadoop 相关的 jar 包 导入,首先右击项目选择"New"一"Folder"创建一个 lib 文件夹并把指定



2.9 在项目 src 目录下指定的包名" com. simple. secondarysort"下右键点击,新建一个类名为"SecondarySortReducer"并继承 Reducer类,然后添加该类中的代码内容如下所示。

| Fackag      | e Explorer 🕱 🗖 🗖         | *GroupingCompar  *SecondarySortM  | ] *SecondarySortR ☎ *2 = □  |
|-------------|--------------------------|---|---|
|             |                          | <pre>1 package com.simple.secondarysort;</pre>                                |   |
| 🕨 😂 Englis  | hWordsCount              | <pre>20 import java.io.IOException; 8 import org.apache.hadoop.io.Text;</pre> |   |
| ► 😂 hadoo   | p                        | import org.apache.hadoop.mapreduce.R  | educer;   |
| IntSor      | tCode                    | public class SecondarySortReducer ex<br>private static final Text SEPARATOR   | <pre>tends Reducer<intpair, text,="" text,<br="">= new Text(""</intpair,></pre> |
| 🔻 😂 Secon   | darySort                 | Popublic void reduce(IntPair key, Iter  | able <text> values, Context context</text>                                      |
| 🔻 🚈 src     |                          | <pre>context.write(SEPARATOR, null);</pre>                                    | Tion (  |
| V 🕀 CON     | m.simple.secondaryso     | <pre>p for (Text val : values) {     context write(null wall); </pre>         | <u>*</u>  |
| ► [j] Fi    | irstPartitioner.java     | 2 }   |   |
|             | atPair iav               | 18 }<br>4 1   |   |
| ► 🕅 S       | econ arySortMapper.ja    |   |   |
| ► 🖸 S       | econdarySortReducer.j    |   |   |
| V 🛋 JRE S   | System Library [jdk1.8   |   |   |
| res         | ources.jar - /root/simp  |   |   |
| ► 🔂 rt.j    | ar - /root/simple/jdk/jc |   |   |
| iss         | e.jar - /root/simple/jdk |   |   |
| ► 🔐 jce     | .jar - /root/simple/jdk/ |   |   |
| ► 🔤 cha     | iar - /root/simple/idk/i |   |   |
| 2 10 7      | 车项目 src 目-               | 录下指定的包名" com simpl  | e secondarysort"下右键   |
|             |                          |   |   |
| <b>点</b> 击, | 新建一个测证                   | 、主尖名为 SecondarySort   | Job <sup>®</sup> 升指定 main 主力法。  |
|             |                          | New Java Class  | (*)   |
|             | Java Class               |   | O   |
|             | Create a new Java        | class.  |   |
|             |                          |   |   |
|             | Source folder:           | SecondarySort/src   | Browse  |
|             | Package:                 | com simple secondarysort  | Browse  |
|             | C Fallerin I             | containpic.secondarysore  | Browsen.  |
|             | Enclosing type:          |   | Browse  |
|             |                          |   |   |
|             | Name:                    | SecondarySortJob  |   |
|             | Modifiers:               | public O package O private  | o protected   |
|             |                          | 🗆 abstract 🗌 final 📄 static   |   |
|             | Superclass:              | java.lang.Object  | Browse  |
|             | Interfaces:              |   | Add   |
|             |                          |   |   |
|             |                          |   | Remove  |
|             |                          |   |   |
|             | Which method stul        | os would you like to create?  |   |
|             |                          | <pre>wublic static void main(String[] args)</pre>                             |   |
|             |                          | Constructors from superclass  |   |
|             |                          | Inherited abstract methods  |   |
|             | Do you want to ad        | d comments? (Configure templates and def                                      | ault value <u>here</u> )  |
|             |                          | <ul> <li>Generate comments</li> </ul>   |   |
|             |                          |   |   |
|             |                          |   |   |
|             |                          |   |   |
|             |                          |   |   |
|             | ?                        |   | Cancel Finish   |
|             |                          |   |   |
| 2.11 测      | 则试代码如下所                  | 斤示。   |   |




|        | 0          |                                    | New Java P  | roject                           |                | × .      |
|--------|------------|------------------------------------|---|----------------------------------|----------------|----------|
|        | Create a   | lava Project                       |   |                                  |                |          |
|        | Create a J | ava project in t                   | the workspace or in an  | external location.               |                |          |
|        |            |                                    |   |                                  |                |          |
|        | Project r  | name: Dedupli                      | icationInfo   |                                  |                |          |
|        | 🗹 Use d    | default locatior                   | ı   |                                  |                |          |
|        | Location   | : /root/works                      | pace/DeduplicationInfo  |                                  | Brows          | e        |
|        | JRE        |                                    |   |                                  |                |          |
|        | 🔿 Use      | an execution                       | environment JRE:  | JavaSE-1.8                       |                | A        |
|        | 🖲 Use      | a project spec                     | ific JRE:   | jdk1.8.0_241                     |                | ÷        |
|        | O Use      | e default JRE (co                  | urrently 'jdk1.8.0_241'   | )                                | Configure J    | REs      |
|        | Project l  | ayout                              |   |                                  |                |          |
|        | 🔿 Use      | project folder                     | as root for sources and   | l class files                    |                |          |
|        | 🖲 Cre      | ate separate fo                    | olders for sources and c  | lass files                       | Configure defa | ault     |
|        | Working    | sets                               |   |                                  |                |          |
|        | Add        | project to wor                     | king sets   |                                  |                |          |
|        | Workin     | g sets:                            |   |                                  | * Select.      |          |
|        |            |                                    |   |                                  |                |          |
|        |            |                                    |   |                                  |                |          |
|        |            |                                    |   |                                  |                |          |
|        | ?          |                                    | < Back Next   | > Cance                          | el Fini        | ish      |
| 2.2 在1 | 页目 src     | 目录下,                               | 右键点击,选择   | 译"新建"创                           | 建一个类           | 文件名称为    |
| "Dedup | olicatio   | onMapper"                          | '并指定包名"   | com.simple                       | e.deduplic     | eation"。 |
|        |            | o<br>Java Class                    | New Java C  | lass                             | ×              |          |
|        |            | Create a new Java                  | a class.  |                                  | C              |          |
|        |            | Source folder:                     | DeduplicationInfo/src   |                                  | Browse         |          |
|        |            | Package:                           | com.simple.deduplication  |                                  | Browse         |          |
|        |            | <ul> <li>Enclosing type</li> </ul> |   |                                  |                |          |
|        |            | Name:<br>Modifiers:                | DeduplicationMapper   | private O protected              |                |          |
|        |            | Houners.                           | abstract final st   | atic                             |                |          |
|        |            | Superclass:<br>Interfaces:         | java.lang.Object  |                                  | Browse         |          |
|        |            |                                    |   |                                  | Remove         |          |
|        |            | Which method stu                   | ibs would you like to create?   |                                  |                |          |
|        |            |                                    | <ul> <li>public static void main(Stri</li> <li>Constructors from superclassion</li> </ul> | ing[] args)<br>ass               |                |          |
|        |            | Do you want to ar                  | Inherited abstract methods<br>id comments? (Configure templa)                             | ates and default value he        | re)            |          |
|        |            | So you want to at                  | Generate comments   | ines and deraute value <u>ne</u> |                |          |
|        |            |                                    |   |                                  | 4              |          |
|        |            | 0                                  |   | Cancel                           | Finish         |          |
|        |            |                                    |   |                                  |                |          |

深圳大学学生实验报告用纸





2.5 让类 "DeduplicationMapper"继承类 Mapper 同时指定需要的参数类型,根据业务逻辑修改 map 类的内容如下。







| New Java Project   |    |
|--|----|
| Create a Java Project  |    |
|  |    |
| Project name MrLogger  |    |
| Use default location   |    |
| JRE  |    |
| O Use an execution environment JRE: JavaSE-1.8   |    |
| Use a project specific JRE:     jdk1.8.0_241      jdk1.8.0_241   |    |
| Desired layout   |    |
| Use project folder as root for sources and class files   |    |
| Create separate folders for sources and class files <u>Configure default</u>   |    |
| Working sets   |    |
| Add project to working sets     Working sets:     Select   |    |
|  |    |
|  |    |
| ? < Back Next > Cancel Finish  |    |
|  | L. |
| 2.2 住坝日 Src 日求下,石健点击,远择 胡建 "创建一个突叉件名称/<br>"LoggerMapper"并指定包名" com.simple.log" 。   | 7  |
| New Java Class   |    |
| Java Class<br>Create a new Java class.   |    |
| Source folder: MrLogger/src Browse   |    |
| Package: Com.simple.lpg Browse   |    |
| Enclosing type: Browse   |    |
| Name: LoggerMapper   |    |
| Modimers:     Image: Description of the protected       Image: Image: Image: Image: Image: Description of the protected       Image: Image |    |
| Superclass: java.lang.Object Browse  |    |
| Add  |    |
| Remove   |    |
| Which method stubs would you like to create?   |    |
| Constructors from superclass   |    |
| Do you want to add comments? (Configure templates and default value <u>here</u> )  |    |
| Generate comments  |    |
|  |    |
|  |    |

深圳大学学生实验报告用纸

2.3 在编写"LoggerMapper"类之前需要把 hadoop 相关的 jar 包导入,首先 右击项目选择"New"—"Folder"创建一个 lib 文件夹并把指定位置中(桌 面 lib 文件夹)的包放入该文件中。



2.4 把 lib 下所有的 jar 包导入到环境变量,首先全选 lib 文件夹下的 jar 包文件,右键点击,选择 "build path" --> "add to build path",添加后,发现在项目下很多奶瓶图表的 jar 包。



2.5 让类"LoggerMapper"继承类 Mapper 同时指定需要的参数类型,根据业务逻辑修改 map 类的内容如下。





pool-3-thread-1] mapred.Task (Task.java:sendDone(1115)) - Task 'attempt local1312657434 0001 r 0000 pool-3-thread-1] mapred.LocalJobRunner (LocalJobRunner.java:run(325)) - Finishing task: attempt\_loc Thread-11] mapred.LocalJobRunner (LocalJobRunner.java:runTasks(456)) - reduce task executor complet lob iob local1312657434 0001 running in u main] mapreduce.Job (Job.java:monitorAndPrintJob(1355)) main] mapreduce.Job (Job.java:monitorAndPrintJob(1362)) map 100% reduce 100% main] mapreduce.Job (Job.java:monitorAndPrintJob(1373)) Job job tocati312037434 0001 completed su main] mapreduce.Job (Job.java:monitorAndPrintJob(1380)) - Counters: 33 bytes read=1792 bytes written=438128 read operations=0 large read operations=0 write operations=0 2.10 程序执行完毕之后,可以到输出信息目录/simple/output下,执行查看 命令:cat part-r-00000 查看对数据处理后产生的结果。 (base) [root@ferry simple] cd output (base) [root@ferry output]# ls part-r-00000 SUCCESS (base) [root@ferrv output]# cat part-r-00000 GET / /html/20140620/872.html 1 GET GET /html/20140620/900.html 1 POST /service/addViewTimes 544.htm 1 POST /service/addViewTimes 900.htm 1 (base) [root@ferry output]# 5. MR4 MR 调优 【案例 5.1】mapreducetopN 排名 一、项目准备阶段 需求: 第一个字段: 第二个字段: 第三个字段: 第四个字段: orderid 订单 ID userid 用户 ID payment 付款金额 productid 产品 ID 1) 付款金额最大的几笔订单 1, 9819, 100, 121 2, 8918, 2000, 111 3, 2813, 1234, 22 4, 9100, 10, 1101 5, 3210, 490, 111 6, 1298, 28, 1211 7, 1010, 281, 90 8, 1818, 9000, 20 100, 3333, 10, 100 101, 9321, 1000, 293 102, 3881, 701, 20 103, 6791, 910, 30 104, 8888, 11, 39







|   | S I PINMapper, ava I * IopNMeducer, Iava M   |                         |
|---|--|-------------------------|
| SAL   | ▼ Is interruptedexception 4<br>I4 for (int i = len: i > 0: i) {  | r.                      |
| DeduplicationInfo   | 15 context.write(new Text(String.valueOf(len - i + 1)), ne   | w Text(                 |
| EnglishWordsCour  | nt 16 String.valueOf(top[i])));  |                         |
| Adoop   | 18 }   | <b>v</b> (              |
| IntsortCode   | 19⊖@Override<br>420 protected void reduce(IntWritable key, Iterable <intwrit< td=""><td>able&gt; values,</td></intwrit<>   | able> values,           |
| <ul> <li>SecondarySort</li> </ul>   | 21 Context context) throws IOException, InterruptedExcepti   | ion {                   |
| StudentAvgScore   | <pre>22 for (Intwritable val : values) { 23 add(val.get());</pre>  |                         |
| StudentTotalScore   | 24 }   |                         |
| 🔻 🖾 TopN  | 26⊖ private void add(int val) {  |                         |
| V 🖉 Src   | 27 top[0] = val;<br>28 Arrays.sort(top):// 排席 从小到大順序 top 初始化值都是 0  |                         |
| topNMapper  | 29 }   |                         |
| TopNReducer.  | java | eption,                 |
| 🔻 🛋 JRE System Libra  | ary [jdk1.8 32 InterruptedException {  |                         |
| resources.jar -   | /root/simp<br>34 len = context.getConfiguration().getInt("N", 10);   |                         |
| rt.jar - /root/sin  | mple/jdk/jc 35 top = new int[len + 1];// 这个数组要是 TOP N+1  |                         |
| ice.jar - /root/s   | simple/jdk/  |                         |
| 4   | * (<br>20  |                         |
|   |  |                         |
| 2.7 在项目 src   | 目录卜指定的包名" com.simple.topn" 卜石  | 键点击,新建                  |
| 一个测试主类名   | 4为" TopNJob"并指定 main 主方法。  |                         |
|   |  |                         |
| Source folder:  | TopN/src   | Browse                  |
| Package:  | com.simple.topn  | Browse                  |
| Enclosing type  | E:   | Browse                  |
|   |  |                         |
|   |  |                         |
| Name:   | lobNjop  |                         |
| Modifiers:  | public Opackage Oprivate Oprotected  |                         |
|   | Dishetraet Difinal Distation   |                         |
|   | ans at the set   |                         |
|   |  |                         |
| Superclass:   | java.lang.Object   | Browse                  |
| Superclass:<br>Interfaces:  | java.lang.Object   | Browse                  |
| Superclass:<br>Interfaces:  | java.lang.Object   | Browse<br>Add           |
| Superclass:<br>Interfaces:  | java.lang.Object   | Browse<br>Add           |
| Superclass:<br>Interfaces:  | java.lang.Object   | Browse<br>Add<br>Remove |
| Superclass:<br>Interfaces:<br>Which method stu                                    | java.lang.Object   | Browse<br>Add<br>Remove |
| Superclass:<br>Interfaces:<br>Which method stu                                    | java.lang.Object   | Browse<br>Add<br>Remove |
| Superclass:<br>Interfaces:<br>Which method stu                                    | java.lang.Object   | Browse<br>Add<br>Remove |
| Superclass:<br>Interfaces:<br>Which method stu                                    | java.lang.Object   | Browse<br>Add<br>Remove |
| Superclass:<br>Interfaces:<br>Which method stu                                    | java.lang.Object  ibs would you like to create?  public static void main(String[] args)  Constructors from superclass  i Inherited abstract methods  | Browse<br>Add<br>Remove |
| Superclass:<br>Interfaces:<br>Which method stu                                    | java.lang.Object java.lang.Object ubs would you like to create? v public static void main(String[] args) Constructors from superclass v Inherited abstract methods dd comments? (Configure templates and default value here  | Browse<br>Add<br>Remove |
| Superclass:<br>Interfaces:<br>Which method stu<br>Do you want to ad               | java.lang.Object   | Browse<br>Add<br>Remove |
| Superclass:<br>Interfaces:<br>Which method stu<br>Do you want to ac<br>2. 8 测试代码如 | java.lang.Object<br>java.lang.Object<br>ibs would you like to create?<br>✓ public static void main(String[] args)<br>Constructors from superclass<br>✓ Inherited abstract methods<br>Id comments? (Configure templates and default value here<br>1下所示。   | Browse<br>Add<br>Remove |
| Superclass:<br>Interfaces:<br>Which method stu<br>Do you want to ac<br>2.8 测试代码如  | abstract I final I static<br>java.lang.Object<br>ibs would you like to create?<br>I public static void main(String[] args)<br>Constructors from superclass<br>I nherited abstract methods<br>Id comments? (Configure templates and default value here<br>1下所示。   | Browse<br>Add<br>Remove |
| Superclass:<br>Interfaces:<br>Which method stu<br>Do you want to ac<br>2.8 测试代码如  | abstract I minit I static<br>java.lang.Object<br>↓bs would you like to create?<br>✓ public static void main(String[] args)<br>↓ Constructors from superclass<br>✓ Inherited abstract methods<br>dd comments? (Configure templates and default value here<br>1下所示。  | Browse<br>Add<br>Remove |
| Superclass:<br>Interfaces:<br>Which method stu<br>Do you want to ac<br>2.8 测试代码如  | abstract I final I static<br>java.lang.Object<br>ibs would you like to create?<br>✓ public static void main(String[] args)<br>Constructors from superclass<br>✓ Inherited abstract methods<br>dd comments? (Configure templates and default value here<br>1下所示。  | Browse<br>Add<br>Remove |
| Superclass:<br>Interfaces:<br>Which method stu<br>Do you want to ac<br>2.8 测试代码如  | abstract I man I state<br>java.lang.Object<br>ibs would you like to create?<br>✓ public static void main(String[] args)<br>Constructors from superclass<br>✓ Inherited abstract methods<br>dd comments? (Configure templates and default value here<br>1下所示。   | Browse<br>Add<br>Remove |
| Superclass:<br>Interfaces:<br>Which method stu<br>Do you want to ac<br>2.8 测试代码如  | java.lang.Object<br>java.lang.Object<br>ibs would you like to create?<br>✓ public static void main(String[] args)<br>Constructors from superclass<br>✓ Inherited abstract methods<br>dd comments? (Configure templates and default value here<br>1下所示。   | Browse<br>Add<br>Remove |
| Superclass:<br>Interfaces:<br>Which method stu<br>Do you want to ac<br>2.8 测试代码如  | addate I mining State<br>java.lang.Object<br>↓ bs would you like to create?<br>↓ public static void main(String[] args)<br>↓ Constructors from superclass<br>☑ Inherited abstract methods<br>dd comments? (Configure templates and default value here<br>1下所示。   | Browse<br>Add<br>Remove |



| 起来,展示为每个公共服务号码对应多个,   | 用户号码。  |
|---|--|
| 原始数据:   |  |
| 13718855152 112   |  |
| 18610117315 110   |  |
| 89451849 112  |  |
| 13718855153 110   |  |
| 13718855154 112   |  |
| 18610117315 114   |  |
| 18910117315 114   |  |
| (base) [root@rerry output]<br>(base) [root@rerry simple]<br>13718855152 112<br>18610117315 110<br>89451849 112<br>13718855153 110<br>13718855154 112<br>18610117315 114<br>(base) [root@rerry simple]<br>二、程序编写<br>2.1 在 eclipse 中的项目列表中, 右键点<br>Project"新建一个项目"TelPubCount | # CO ~/SIMDLE<br>Cat source.txt<br>#<br>击,选择"new"—>" Java<br>" 。 |
| Use default location  |  |
| Location: //root/workspace/TelPubCount  | Browse   |
| JRE   |  |
| <ul> <li>Use an execution environment JRE:</li> </ul>   | JavaSE-1.8 Å   |
| Ose a project specific JRE:   | jdk1.8.0_241 *   |
| <ul> <li>Use default JRE (currently 'jdk1.8.0_241')</li> </ul>  | Configure JREs   |
| Project layout  |  |
| <ul> <li>Use project folder as root for sources and</li> </ul>  | class files  |
| Create separate folders for sources and cl  | ass files <u>Configure default</u>                               |
| 2.2 在项目 src 目录下,右键点击,选择<br>件名称为"TelPubCountMapper"并指定包  | "New"—"Class"创建一个类文<br>2名" tel.pub.count" 。                      |

| Source folder:                         | TelPubCount/src   | Browse        |
|--|---|---------------|
| Package:                               | tel.pub.count   | Browse        |
| Enclosing type:                        |   | Browse        |
|  |   |               |
| Name:                                  | TelPubCountMapper   | J             |
| Modifiers:                             | public O package O private O protected  | i             |
|  | □ abstract □ final □ static   |               |
| Superclass:                            | java.lang.Object  | Browse        |
| Interfaces:                            |   | Add           |
|  |   | -             |
| 1                                      |   | Remove        |
| Which method stub                      | s would you like to create?   |               |
|  | public static void main(String[] args)  |               |
|  | <ul> <li>Constructors from superclass</li> </ul>                              |               |
|  | Inherited abstract methods  |               |
| 2.3 在编写"lel<br>首先右击项目选择<br>(桌面 lib 文件夹 | PublountMapper" 尖乙前需要把 hadoop 相关<br>译"New"—"Folder"创建一个 lib 文件夹<br>)的包放入该文件中。 | 并把指定位置中       |
| Ψ.(                                    | 2 TelPubCount   |               |
|  | r 📇 src   |               |
|  | 🔻 🌐 tel.pub.count   |               |
|  | I TelPubCountMapper.javi  |               |
|  | IRE System Library [jdk1.8  |               |
|  | r 🖕 lib   |               |
|  | activation-1.1.jäř  |               |
|  | aopalliance-1.0.jar   |               |
|  | asm-3.2.jar   |               |
|  | avro 1.7.1.jar  |               |
| 2.4 把 1 ib 下所有                         | 有的 jar 包导入到环境变量,首先全选 lib 文  | <br>了件夹下的 iar |
| 包文件,右键点:                               | 告,选择"build path"> "add to build r   | path",添加后,    |
| 发现在项目下很多                               | 多奶瓶图表的 jar 包。   | ,, ,,         |
|  |   |               |
|  |   |               |
|  |   |               |
|  |   |               |
|  |   |               |
|  |   |               |





2.7 在项目 src 目录下指定的包名" tel.pub.count"下右键点击,新建一个测试主类名为" TelPubCountJob"并指定 main 主方法。

| Source folder:      | TelPubCount/src  | Browse |
|---------------------|--|--------|
| Package:            | tel.pub.count  | Browse |
| Enclosing type:     |  | Browse |
| Name:<br>Modifiers: | TelPubCountJob            • public         • package         • private         • protected         • abstract         • final         • static         • static |        |
| Superclass:         | java.lang.Object   | Browse |
| Interfaces:         |  | Add    |
|                     |  | Remove |
| Which method stui   | Swould you like to create?<br>Spublic static void main(String[] args)<br>Constructors from superclass<br>Inherited abstract methods  |        |
| 2.8 测试代码如           | 下所示。   |        |
|                     |  |        |
|                     |  |        |
|                     |  |        |



| 刘八 18  |   |
|--|---|
| 胡三 29  |   |
| <pre>(base) [root@ferry output]# (base) [root@ferry simple]# 张三 13 季四 28 王五 34 赵六 22 钱七 17 刘八 18 胡三 29 (base) [root@ferry simple]# 二、程序编写</pre>  | cd ~/simple<br>cat source.txt   |
| 2.1 在 eclipse 中的项目列表中,右键点  | 击,选择"new"—>"Java  |
| Project…"新建一个项目"StuAgeCata"  | 0   |
| Project name StuAgeCata  |   |
| 🗹 Use default location   |   |
| Location: /root/workspace/StuAgeCata   | Browse  |
| IRE  |   |
|  |   |
| <ul> <li>Use an execution environment JRE:</li> </ul>  | JavaSE-1.8  |
| Ose a project specific JRE:  | jdk1.8.0_241 *  |
|  | *0  |
| O Lise default IRE (currently 'idk1 8 0 241')  | Configure IREs  |
| <ul> <li>Use default JRE (currently 'jdk1.8.0_241')</li> </ul>   | Configure JREs  |
| <ul> <li>Use default JRE (currently 'jdk1.8.0_241')</li> <li>Project layout</li> </ul>   | Configure JREs  |
| O Use default JRE (currently 'jdk1.8.0_241') Project layout  | Configure JREs  |
| <ul> <li>Use default JRE (currently 'jdk1.8.0_241')</li> <li>Project layout</li> <li>Use project folder as root for sources and</li> </ul>   | Configure JREs  |
| <ul> <li>Use default JRE (currently 'jdk1.8.0_241')</li> <li>Project layout</li> <li>Use project folder as root for sources and</li> <li>Create separate folders for sources and cl</li> </ul>   | Configure JREs<br>class files<br>ass files <u>Configure default</u>   |
| <ul> <li>Use default JRE (currently 'jdk1.8.0_241')</li> <li>Project layout</li> <li>Use project folder as root for sources and</li> <li>Create separate folders for sources and cl</li> <li>2.2 在项目 src 目录下,右键点击,选择<br/>"StudentWritable"并指定包名" com.</li> <li>的封装,并作为 mapper 的输出键值对象</li> </ul> | Configure JREs<br>class files<br>ass files <u>Configure default</u><br>"新建"创建一个类文件名称为<br>stu",该类是对给定数据的三列值<br>。 |



2.3 在项目 src 目录下,右键点击包名" com. stu"选择" New" — "Class" 创建一个类文件名称为"StuPartitioner",该类是对数据处理后的结果进行分区设置。



2.4 在项目 src 目录下,右键点击包名" com. stu"选择" New"—"Class" 创建一个类文件名称为"StudentMapper"。

| Source folder:  | StuAgeCata/src   | Browse             |
|---|--|--------------------|
| Package:  | com.stu  | Browse             |
| Enclosing type:                                       |  | Browse             |
| Name:<br>Modifiers:                                   | StudentMapper  |                    |
|   | □ abstract □ final □ static  |                    |
| Superclass:   | java.lang.Object   | Browse             |
| Interfaces:   |  | Add                |
|   |  | Remove             |
| Which method stub                                     | s would you like to create?  public static void main(String[] args)  Constructors from superclass  Inherited abstract methods  |                    |
| 先右击项目选择<br>(桌面 1 ib 文件夹<br>2.6 把 1 ib 下所 <sup>7</sup> | <ul> <li>"New" — "Folder" 创建一个 lib 文件夹并把</li> <li>达山AgeCata</li> <li>StuAgeCata</li> <li>Studectua</li> <li>StudentMapper.java</li> <li>StudentWritable.java</li> <li>JRE System Library [jdk1.8</li> <li>activation-1.1.jar</li> <li>aopalliance-1.0.jar</li> <li>asm-3.2.jar</li> <li>auro-1 7.4 jar</li> <li>commons-beanutils-1.7.0</li> </ul> | 巴指定位置中<br>件夹下的 jar |
| 包文件,右键点;<br>发现在项目下很                                   | 击,选择"build path">"add to build pa<br>多奶瓶图表的 jar 包。   | ath",添加后,          |





2.9 在项目 src 目录下指定的包名" com. stu"下右键点击,新建一个测试 主类名为" TestStuMapReducer"并指定 main 主方法。

| Source folder:        | StuAgeCata/src  | Browse       |
|-----------------------|---|--------------|
| Package:              | com.stu   | Browse       |
| Enclosing type:       |   | Browse       |
| Name:                 | TestStuMapBeducer   |              |
| Modifiers:            | public Opackage Oprivate Oprotected                           |              |
| - autor at            | □ abstract □ final □ static                                   |              |
| Superclass:           | java.lang.Object  | Browse       |
| Interfaces:           |   | Add          |
|                       |   | Remove       |
| Which method stu      | bs would you like to create?                                  |              |
|                       | public static void main(String[] args)                        |              |
|                       | Constructors from superclass                                  |              |
|                       | Inherited abstract methods                                    |              |
| 2.10 最后在项目<br>则试主类名为" | 目 src 目录下指定的包名" com.stu"下右键点<br>TestStuMapReducer "添加测试代码如下所表 | 点击,新建−<br>示。 |
|                       |   |              |
|                       |   |              |
|                       |   |              |
|                       |   |              |



| 18610117315 500 700  |
|--|
| 13718855152 400 900  |
| 13121521297 100 800  |
| <pre>(base) [root@ferry output]#_cd ~/simole<br/>(base) [root@ferry simple]# cat source.txt<br/>18610117315 200 300<br/>13718855152 300 500<br/>18610117315 100 300<br/>18610117315 500 700<br/>13718055152 400 900<br/>13121521297 100 800<br/>(base) [root@ferry simple]#<br/>二、程序编写<br/>2.1 在 eclipse 中的项目列表中, 右键点击, 选择 "new "—&gt;" Java</pre> |
| Project… 新建一个项目"FlowAccount"。  |
| Project name: FlowAccount  |
| ☑ Use default location   |
| Location: /root/workspace/FlowAccount Browse   |
| JRE  |
| O Use an execution environment JRE: JavaSE-1.8   |
| Use a project specific JRE: jdk1.8.0_241   |
| O Use default JRE (currently 'jdk1.8.0_241') Configure JREs  |
| Project layout   |
| <ul> <li>Use project folder as root for sources and class files</li> </ul>   |
| Create separate folders for sources and class files     Configure default  |
|  |
| 2.2 在项目 src 目录下, 石键点击, 选择"New"—"Class"创建一个类文件名称为"Flow"并指定包名" com. flow", 该类是对给定数据的三列值的封装, 并作为 mapper 的输出键值对象 。   |



2.3 在项目 src 目录下,右键点击,选择"New"—"Class"创建一个类文件名称为"FlowMapper"并指定包名" com.flow" 。

| Source folder:                         | FlowAccount/src  | Browse             |
|--|--|--------------------|
| Package:                               | com.flow   | Browse             |
| Enclosing type:                        |  | Browse             |
| Name:                                  | FlowMapper   |                    |
| Modifiers:                             | public opackage oprivate oprotected     abstract final static              |                    |
| Superclass:                            | java.lang.Object   | Browse             |
| Interfaces:                            |  | Add                |
|  |  | Remove             |
| Which method stub                      | s would you like to create?  |                    |
|  | public static void main(String[] args)                                     |                    |
|  | Constructors from superclass   |                    |
|  | Inherited abstract methods   |                    |
| 2.4 在编写"Flo<br>击项目选择"New<br>lib 文件夹)的包 | owMapper"类之前需要把 hadoop 相关的 jar 包<br>v"—"Folder"创建一个 lib 文件夹并把指定<br>放入该文件中。 | 2导入,首先右<br>6位置中(桌面 |





| Source folder:   | FlowAccount/src  | Browse   |
|--|--|--|
| Package:   | com.flow   | Browse   |
| Enclosing type:  |  | Browse   |
| Name:<br>Modifiers:  | TestFlow            • public         • package         • private         • protected         • abstract         • final         • static         • |  |
| Superclass:  | java.lang.Object   | Browse   |
| Interfaces:  |  | Add  |
|  |  | Remove   |
| Do you want to add   | ☑ Inherited abstract methods<br>comments? (Configure templates and default value <u>here</u><br>□ Generate comments<br>下所示。  | 0  |
| <ul> <li>Package Explorer &amp;</li> <li>Package Explored Scout</li> <li>Package Explo</li></ul> | <pre>*FlowMapper.jav</pre>   | <pre>bit =  bit format; bit fo</pre> |
| 2.10 按照以上的<br>后,选中测试类"<br>Application",{   | 步骤,把 mapper 和 reducer 阶段以及测试代码<br>' TestFlow ",右键点击选择"Run as"<br>查看控制台显示内容查看是否正确执行   | 马编写完毕之<br>->"Java  |
| uce.Job ( <u>Job.java:mon</u><br>uce.Job ( <u>Job.java:mon</u><br>uce.Job ( <u>Job.java:mon</u><br>uce.Job ( <u>Job.java:mon</u>   | <pre>itorAndPrintJob(1355)) - Job job local285986574 0001 run<br/>itorAndPrintJob(1362)) - map 100% reduce 100%<br/>itorAndPrintJob(1373)) - Job Job_tocat285986574_0001 com<br/>itorAndPrintJob(1380)) - Counters: 33</pre>   | ning in uber mo  |

2.11 程序执行完毕之后,可以到输出信息目录/simple/output下,执行查看 命令:cat part-r-00000,查看对数据处理后产生的结果。

| (base) [root@ferry simpl<br>(base) [root@ferry outpu<br>part_r_00000 | .e]# <mark>cd output</mark><br>nt]# ls |
|--|--|
| (hase) [root@ferry_outou   | ıt]# cat part-r-00000                  |
| 13121521297:(100,800)<br>13718855152:(300,500)                       |  |
| 13718855152:(400,900)<br>18610117315:(200,300)                       |  |
| 18610117315:(100,300)<br>18610117315:(500,700)                       |  |
| L(base) Iroot@terry_outou  | it]#                                   |

## 【案例 5.5】mapreduce 电话号码流量统计并分区

一、项目准备阶段

需求:对所给的所有电话号码产生的流量记录进行按电话号码进行汇总,求 出所有相同电话号码产生的流量和。

原始数据:

| 18610117315 | 200 | 300 |  |
|-------------|-----|-----|--|
| 13718855152 | 300 | 500 |  |
| 18610117315 | 100 | 300 |  |
| 18610117315 | 500 | 700 |  |
| 13718855152 | 400 | 900 |  |
| 13121521297 | 100 | 800 |  |

| (base)  | [root@ferry output]#_cd ~/simple    |
|---------|-------------------------------------|
| (base)  | [root@ferry simple]# cat source.txt |
| 1861011 | 7315 200 300                        |
| 1371885 | 5152 300 500                        |
| 1861011 | 7315 100 300                        |
| 1861011 | 7315 500 700                        |
| 1371885 | 5152 400 900                        |
| 1312152 | 1297 100 800                        |
| (base)  | [root@ferry simple]#                |

二、程序编写

2.1 在 eclipse 中的项目列表中,右键点击,选择"new"—>" Java Project…"新建一个项目"FlowAccount"。

| Project name: FlowAccount<br>Vse default location<br>Location: /root/workspace/FlowAccount                          |  | Browse          |
|---|--|-----------------|
| JRE   |  |                 |
| O Use an execution environment JRE:   | JavaSE-1.8   | A<br>V          |
| <ul> <li>Use a project specific JRE:</li> </ul>   | jdk1.8.0_241   | 4               |
| O Use default JRE (currently 'jdk1.8.0_241')  | Confi  | gure JREs       |
| Project layout     O Use project folder as root for sources and o     O Create separate folders for sources and cla | class files Configur   | e default       |
| <pre>2.2 在项目 src 目录下,右键点击,选择'<br/>件名称为 "Flow"并指定包名" com.flow<br/>封装,并作为 mapper 的输出键值对象。</pre>                       | <pre>" New" — "Class" 包<br/>w",该类是对给定数据<br/>[ava<br/>t;<br/>;;<br/>p.io.Text;<br/>p.io.WritableComparable;<br/>arable<text>{<br/>int down, int up) {<br/>{<br/>{<br/>rng tel) {</text></pre> | ]建一个类文<br>的三列值的 |



2.4 在项目 src 目录下,右键点击,选择" New"—"Class"创建一个类文件名称为"FlowMapper"并指定包名" com. flow" 。(同上案例)

2.5 在编写"FlowMapper"类之前需要把 hadoop 相关的 jar 包导入,首先右 击项目选择"New"—"Folder"创建一个 lib 文件夹并把指定位置中(桌面 lib 文件夹)的包放入该文件中。(同上案例)

2.6 把 lib 下所有的 jar 包导入到环境变量,首先全选 lib 文件夹下的 jar 包文件,右键点击,选择 "build path" --> "add to build path",添加后,发现在项目下很多奶瓶图表的 jar 包。(同上案例)

2.7 让类 "FlowMapper"继承类 Mapper 同时指定需要的参数类型,根据业务 逻辑修改 map 类的内容如下。



深圳大学学生实验报告用纸




# 实验结论与分析:

### 1. 粘贴最终结果图,并对结果做分析说明,,,,....

#### 1.MR1-1 词频统计

首先对待处理的信息进行拆分,拆分之后在 map 阶段,把拆分的每个单词作为 map 方法的输出键,而 map 的方法输出的值设置为 1,最后在 reduce 阶段 对每个键的值集合进行遍历并把遍历的值进行相加,输出结果即可。



### 2. MR1-2 单词个数

首先对待处理的信息进行拆分,拆分之后在 map 阶段,拆分后计算出单词个数并作为 map 方法的输出值,而 map 的方法输出键作为 NullWritable 即可,最后在 reduce 阶段对每个键的值集合进行遍历并把遍历的值进行相加,输出结果即可。



## 3.MR2-2 成绩统计

Map 处理的是一个纯文本文件,文件中存放的数据时每一行表示一个学生的 姓名和他相应一科成绩。Mapper 处理的数据是由 InputFormat 分解过的数据 集,其中 InputFormat 的作用是将数据集切割成小数据集 InputSplit,每一 个 InputSplit 将由一个 Mapper 负责处理。此外, InputFormat 中还提供了 一个 RecordReader 的实现,并将一个 InputSplit 解析成<key, value>对提供 给了 map 函数。InputFormat 的默认值是 TextInputFormat,它针对文本文件, 按行将文本切割成 InputSlit,并用 LineRecordReader 将 InputSplit 解析 成<key, value>对,key 是行在文本中的位置,value 是文件中的一行。Map 的结果会通过 partion 分发到 Reducer, Reducer 做完 Reduce 操作后,将通 过以格式 OutputFormat 输出。Mapper 最终处理的结果对<key, value>, 会送 到 Reducer 中进行合并, 合并的时候, 有相同 key 的键/值对则送到同一个 Reducer 上。Reducer 是所有用户定制 Reducer 类地基础, 它的输入是 key 和这个 key 对应的所有 value 的一个迭代器, 同时还有 Reducer 的上下文。 Reduce 的结果由 Reducer. Context 的 write 方法输出到文件中。



4.MR3 MRShuffle

#### 【案例 4.1】mapreduce 整数排序并指定顺序编号

(base) [root@ferry output]#

本试验要求对一系列数据进行排序你,自然就会想到只要把这些数据作为 map 阶段的输出键即可,排序后的数据在输出时,在数值前面添加一个数字 的自然顺序编号,此时只要考虑在 reduce 阶段定义一个全局变量,每次执行 reduce 方法时,就是对 map 阶段的 key 进行处理,此时全局变量就可以作为 编号,每执行一次 reduce,全局变量就会增加 1。



在 MapReduce 操作时,我们知道传递的<key, value>会按照 key 的大小进行排序,最后输出的结果是按照 key 排过序的。有的时候我们在 key 排序的基础上,对 value 也进行排序。这种需求就是二次排序。二次排序是在框架在对 key2 排序后再对 reduce 输出结果的结果 value3 进行二次排序的需求。在 map 阶段,使用 job. setInputFormatClass 定义的 InputFormat 将输入的数据集分割成小数据块 splites,同时 InputFormat 提供一个 RecordReder 的实现。本例子中使用的是 TextInputFormat,他提供的 RecordReader 会将文本的字节偏移量作为 key,这一行的文本作为 value。



【**案例 4.3】mapreduce 多条数据去重处理** 数据去重的最终目标是让原始数据中出现次数超过一次的数据在输出文件中 只出现一次。具体就是 reduce 的输入应该以数据作为 key, 而对 value-list 则没有要求。当 reduce 接收到一个 < key, value-list > 时就直接将 key 复制 到输出的 key 中,并将 value 设置成空值。



#### 【案例 5.1】mapreducetopN 排名

在每一个 map 端,就求出开始 TopN,这样可以减少在 reduce 端的压力,即 在每一个 map 中先求出 Top10,然后将这 top10,传给 reduce,让后 reduce 把所有全部的传来的数据中找出 top10





## 【案例 5.5】mapreduce 电话号码流量统计并分区

原理: 首先按电话号码作为键进行排序,相同键的内容形成一个集合,然后把相同键的所有内容值进行流量相加,最后按照指定分区条件进行分区输出。



## 2. 谈谈自己对这个实验的体会感受 xxxxx

通过本次实验,我掌握了基本的 MapReduce 编程方法;掌握用 MapReduce 解 决一些常见的数据处理问题,包括数据去重、数据排序和数据挖掘等。本次 实验到此结束,但我知道对 MapReduce 的学习是没有尽头的。学习就是一个 不断积累的过程,学到东西,或多或少,以另一种方式将它记录下来。 总结了一下排除上述问题的一个自己的心得:

感觉刚开始 MapReduce 任务的坑比较多,很难定位到问题的所在,出现上述 类问题主要有以下几点: 1、集群环境处于不稳定状态; 2、代码问题,可能 是很小的一个细节; 3、mapreduce 任务配置参数的问题。

总之,遇到问题,不要怕麻烦,从源头开始一步步排查,在排查的过程中不断总结,能力也就一步步在此过程中会慢慢提升。其实大多数人遇到问题最怕没有思路,不知道从哪里下手,这样会耗费大量时间。遇到问题能有个自己的思路是关键,我们首先得明白有哪些可能会导致出现这些问题,然后开始排查,这样处理问题才能有条不紊的进行下去,也是一种解决问题的比较好的方式。