

任务 1：统计所在国家标识符相同所有客户的账户余额

一、设计思路

(1) Mapper 类

(a) 输入

- Key (键): LongWritable (行偏移量)
- Value (值): Text (一行客户记录)

(b) 输出

- Key (键): IntWritable (客户国家键) 代表的是 c_nationkey
- Value (值): DoubleWritable (客户余额账户) 代表的是 c_acctbal

思路

1. 每一行输入数据表示一个客户的记录，字段之间使用 ' | ' 分隔。
2. 通过 `split("\\|")` 将每一行数据分割成多个字段，并检查字段数量是否为 8 (c_custkey, c_name, c_address, c_nationkey, c_phone, c_acctbal, c_mktsegment 和 c_comment) 以确保数据格式正确。
3. **提取相关字段**: 第四个字段 (`fields[3]`) 表示客户的国家键 (c_nationkey) 和第六个字段 (`fields[5]`) 表示客户的账户余额 (c_acctbal)。
4. 将解析出的国家键和账户余额分别存入 IntWritable 和 DoubleWritable 对象中，然后通过 `context.write(nationKey, accountBalance)` 输出键值对。

(2) Reducer 类

(c) 输入

- Key (键): IntWritable 类型，表示 Mapper 输出的国家键 (c_nationkey)。
- Value (值): Iterable<DoubleWritable> 类型，表示 Mapper 输出的对应于该国家键的所有客户账户余额的集合。

(d) 输出

- Key (键): IntWritable 类型, 表示客户所属的国家键 (c_nationkey)。
- Value (值): DoubleWritable 类型, 表示该国家的所有客户账户余额的总和。

思路:

1. 定义一个 sum 变量, 用于累加每个国家的账户余额。
2. 对于每一个国家键, 遍历所有对应的账户余额, 将其累加到 sum。
3. 将累加的总和存入 DoubleWritable 对象中, 然后通过 context.write(key, result) 输出最终的键值对。

整个程序的目的是在 Map 阶段将输入的数据拆分成 8 个字段, 只提取 4 和 6 个字段, 生成输出键值对 (国家键和账户余额); 在 Reduce 阶段, 分组相同国家键, 并累加相同国家的余额, 输出每一个国家的账户余额总和。

(3) 伪代码

Map 和 Reduce 的伪代码。

```
public static class CustomerMapper extends Mapper<LongWritable, Text, IntWritable, DoubleWritable> {
    private IntWritable nationKey = new IntWritable();
    private DoubleWritable accountBalance = new DoubleWritable();

    @Override
    protected void map(LongWritable key, Text value, Context context) throws IOException, InterruptedException {
        String line = value.toString();
        String[] fields = line.split(regex:"\\|"); //用|区分
        if (fields.length == 8) {
            int c_nationkey = Integer.parseInt(fields[3]); //选择第四字段
            double c_acctbal = Double.parseDouble(fields[5]); //选择第六字段
            nationKey.set(c_nationkey);
            accountBalance.set(c_acctbal);
            context.write(nationKey, accountBalance); //输出键值对
        }
    }
}
```

```
public static class CustomerReducer extends Reducer<IntWritable, DoubleWritable, IntWritable, DoubleWritable> {
    private DoubleWritable result = new DoubleWritable();

    @Override
    protected void reduce(IntWritable key, Iterable<DoubleWritable> values, Context context) throws IOException, InterruptedException {
        double sum = 0.0; //创建一个变量

        for (DoubleWritable val : values) {
            sum += val.get(); //将每个账户余额加到sum里
        }

        result.set(sum);
        context.write(key, result); //输出键值对
    }
}
```

二、实验结果

输出文件路径：/lab3_task1

(1) 输出结果文件的部分截图

```
time taken: 0.000 seconds
hive> create table Lab3Tasklres(c_nationkey string, c_acctbal_sum string) row format delimited fields terminated by ' ' location 'lab3_task1';
OK
Time taken: 0.125 seconds
hive> SELECT * FROM Lab3tasklres LIMIT 20;
0      927882.0300000003      NULL
1      931041.4300000002      NULL
2      873369.01      NULL
3      956474.2100000002      NULL
4      908891.6399999995      NULL
5      929062.3300000003      NULL
6      775537.5199999997      NULL
7      925470.0000000003      NULL
8      852511.1900000003      NULL
9      861944.9300000006      NULL
10     1006397.4300000002      NULL
11     830376.9799999999      NULL
12     909859.1699999995      NULL
13     854463.7999999996      NULL
14     851283.0800000003      NULL
15     966222.8699999998      NULL
16     990429.0999999992      NULL
17     893393.5900000003      NULL
18     906726.3699999996      NULL
19     835428.9100000005      NULL
Time taken: 0.904 seconds, Fetched: 20 row(s)
hive> \q
```

OK

南京理工大学

(2) Yarn Resource Manager 的 WebUI 执行报告内容

在线实验平台

Application application_1678703

在线实验平台

114.212.190.95:8088/cluster/app/application_1678703754602_10358

南京大学教务系统 南京大学新选课系统 EDU-CODER 南京大学计算机系... ICS Lembaga Hasil Dala... 计算机系统基础 (... 计算机网络: 自顶... Computer Networks chatpt 智慧树网 飞书

hadoop

Cluster

Tools

Application Overview

Application Metrics

Application Details

Application Overview

Application Metrics

Application Details

Application Overview

Application Metrics

Application Details