

金融大数据-作业4

0、阅读说明

本实验所有功能均在**伪分布模式**下运行，**开发环境为Windows+IntelliJ IDEA**。具体操作步骤为：命令行start-all.cmd启动hdfs，通过**9870端口上传各类input文件**。阅读指南包括每一条小问对应的**文件名，及参数配置**。

为了便于理解本实验所有的参数及输入输出路径,先看一下9870端口的input文件:

localhost:9870/explorer.html#/

Hadoop

Overview

Datanodes

Datanode Volume Failures

Snapshot

Startup Progress

Utilities

Browse Directory

/

Go!

Show 25 entries

Search:

		Permission		Owner		Group		Size		Last Modified		Replication		Block Size		Name	
		drwxr-xr-x		dr.who		supergroup		0 B		Oct 26 15:17		0		0 B		input1	
		drwxr-xr-x		dr.who		supergroup		0 B		Oct 26 15:17		0		0 B		input2	
		drwxr-xr-x		dr.who		supergroup		0 B		Oct 26 15:55		0		0 B		inputRa	
		drwxr-xr-x		dr.who		supergroup		0 B		Oct 26 16:57		0		0 B		inputRa1	
		drwxr-xr-x		dr.who		supergroup		0 B		Oct 26 18:49		0		0 B		inputRab	
		drwxr-xr-x		dr.who		supergroup		0 B		Oct 26 16:52		0		0 B		inputRb	
		drwx-----		yangfan		supergroup		0 B		Oct 25 15:50		0		0 B		tmp	

Showing 1 to 7 of 7 entries

Previous1Next

其中：

M_3_4和N_4_2分别在input1和input2中；

Ra.txt在inputRa中；

Rb.txt在inputRb中；

Ra.txt和Rb.txt均在inputRab中；

Ra1.txt和Ra2.txt均在inputRa1中；

1、矩阵乘法

项目名称：MatrixMultiply

Main Class选择Selection

Edit Configuration：

```
hdfs://localhost:9000/input1/M_3_4
hdfs://localhost:9000/input2/N_4_2
hdfs://localhost:9000/output/
```

运行结果：

```
1,1 6851
1,2 4947
2,1 13220
2,2 6935
3,1 12523
3,2 13426
```

2、关系代数

项目名称: RelationAlgebra

2.1 选择

在Ra.txt上选择age=18的记录；在Ra.txt上选择age<18的记录。

Edit Configuration:

Main Class选择Selection

Program Arguments

```
hdfs://localhost:9000/inputRa/
hdfs://localhost:9000/output/
2
18
```

age=18和<18的情况下，RelationA.java文件的判断条件也不同。

```
//line36
public boolean isCondition(int col, String value){
    if(col == 0 && Integer.parseInt(value) == this.id)
        return true;
    else if(col == 1 && name.equals(value))
        return true;
    else if(col ==2 && Integer.parseInt(value) == this.age)//age<18时请改成大
于符号
        return true;
    else if(col ==3 && Double.parseDouble(value) == this.weight)
        return true;
    else
        return false;
}
```

运行结果

```
//age=18
1,tom,18,60.0
4,tony,18,62.0
7,brown,18,65.0
12,ivy,18,58.0
13,sam,18,67.0
16,steven,18,60.0
19,coco,18,55.0
//age<18
2,jack,16,59.0
3,lily,17,58.0
```

```
8,candy,17,56.0
10,grace,16,56.0
11,henry,17,61.0
17,jimmy,16,62.0
18,lucas,17,59.0
20,zoey,17,56.0
```

2.2 投影

在Ra.txt上对属性name进行投影。

Edit Configuration:

Main Class选择Protection

Program Arguments

```
hdfs://localhost:9000/inputRa/
hdfs://localhost:9000/output/
1
```

运行结果

```
bill
bob
brown
candy
cici
coco
grace
henry
ivy
jack
jimmy
leon
lily
lucas
owen
sam
steven
tom
tony
zoey
```

2.3 并集

Edit Configuration:

Main Class选择UnionSet

Program Arguments

```
hdfs://localhost:9000/inputRa1/
hdfs://localhost:9000/output/
```

运行结果

```
1,tom,18,60.0
2,jack,16,59.0
3,lily,17,58.0
4,tony,18,62.0
5,bob,20,65.0
6,leon,19,58.0
7,brown,18,65.0
8,candy,17,56.0
9,cici,19,55.0
10,grace,16,56.0
11,henry,17,61.0
12,ivy,18,58.0
13,sam,18,67.0
14,owen,20,63.0
15,bill,19,62.0
16,steven,18,60.0
17,jimmy,16,62.0
18,lucas,17,59.0
19,coco,18,55.0
20,zoey,17,56.0
21,linda,19,60.0
22,tina,17,56.0
25,monica,20,61.0
```

2.4 交集

Edit Configuration:

Main Class选择InterSection

Program Arguments

```
hdfs://localhost:9000/inputRa1/
hdfs://localhost:9000/output/
```

运行结果

```
3,lily,17,58.0
8,candy,17,56.0
9,cici,19,55.0
10,grace,16,56.0
12,ivy,18,58.0
19,coco,18,55.0
20,zoey,17,56.0
```

2.5 差

Edit Configuration:

Main Class选择Difference

Program Arguments

```
hdfs://localhost:9000/inputRa1/  
hdfs://localhost:9000/output/  
Ra2.txt
```

运行结果

```
21,linda,19,60.0  
22,tina,17,56.0  
25,monica,20,61.0
```

2.6 自然连接

Edit Configuration:

Main Class选择NaturalJoin

Program Arguments

```
hdfs://localhost:9000/inputRab/  
hdfs://localhost:9000/output/  
0  
Ra.txt
```

运行结果

```
1,tom,18,1,60.0,178.0  
10,grace,16,0,56.0,170.0  
11,henry,17,1,61.0,173.0  
12,ivy,18,0,58.0,162.0  
13,sam,18,1,67.0,182.0  
14,owen,20,1,63.0,177.0  
15,bill,19,1,62.0,177.0  
16,steven,18,1,60.0,175.0  
17,jimmy,16,1,62.0,178.0  
18,lucas,17,1,59.0,183.0  
19,coco,18,0,55.0,160.0  
2,jack,16,1,59.0,175.0  
20,zoey,17,0,56.0,168.0  
3,lily,17,0,58.0,165.0  
4,tony,18,1,62.0,173.0  
5,bob,20,1,65.0,179.0  
6,leon,19,1,58.0,180.0  
7,brown,18,1,65.0,182.0  
8,candy,17,0,56.0,166.0  
9,cici,19,0,55.0,168.0
```