1、利用Kettle工具实现JSON文件的抽取。提交：JSON抽取.ktr文件、构建的数据文件、生成的数据文件。

要求：

（1）构建源数据文件“JSON抽取.json”，数据内容：

{

"Class1":[

{

"name":"A1",

"age":20

},

{

"name":"A2",

"age":23

}

],

"Class1":[

{

"name":"B1",

"age":22

},

{

"name":"B2",

"age":20

}

]

}

(2) 结果保存为Excel文件“JSON抽取.xls”，抽取内容：

|  |  |  |  |
| --- | --- | --- | --- |
| name1 | age1 | name2 | age2 |
| A1 | 20 | B2 | 20 |

2、利用Kettle工具实现缺失值处理。提交：缺失值处用,ktr文件、构建的数据文件、生成的数据文件。

要求：

（1）使用制表分隔符构建文本文件“缺失值.txt”，数据内容：

|  |  |  |  |
| --- | --- | --- | --- |
| 姓名 | 性别 | 年龄 | 住址 |
| Aaaaaaa | 男 |  | 佛山 |
| Bbbbbb | 女 | 20 |  |
| Ccc | 女 | 30 | 深圳 |
| Dddd | 女 |  | 肇庆 |
| Eeeeeee | 男 | 40 |  |

（2）住址为缺失值的统一填允为：深圳；年龄为缺失值：男性填充为 23，女性填充为20。填充后的内容保存为“缺失值.xls”。

3、利用Kettle工具实现多数据源合并。提交：多数据源合并.ktr文件、构建的数据文

件、生成的数据文件。

(1) 分别构建“数据源合并.csv”、“数据源合并.txt"，数据内容：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 姓名 | 性别 |  | 姓名 | 性别 |
| A | 男 |  |
| B | 女 |  | C | 女 |
| C | 女 |  | D | 男 |

（2）使用“排序合并”控件，以“姓名”为升序，实现两种类型文件内容的合并，合并后的内容保存为“多数据源合并.xls”

4.利用Kettle工具实现字段合并。提交：宁段合并.ktr 文件、生成的数据文件。

要求：

（1）使用“自定义常量数据”控件，构建数据：

|  |  |
| --- | --- |
| FirstName | LastName |
| A | B |
| C | D |
| E | F |

（2）使用“JavaScript代码”控件合并字段：FirstName、LastName，合并后的字段为Name，结果保存为Excel文件“字段合并.xls”。“

5、利用Kettle工具实现缺失值处理。提交：缺失值处理.ktr 文件、构建的数据文件、生成的数据文件。

要求：

（1）使用制表分隔符构建文本文件“缺失值.txt”，数据内容：

|  |  |  |  |
| --- | --- | --- | --- |
| 姓名 | 性别 | 年龄 | 住址 |
| Aaaaaaa | 男 |  | 佛山 |
| Bbbbbb | 女 | 20 |  |
| Ccc | 女 | 30 | 深圳 |
| Dddd | 女 |  | 肇庆 |
| Eeeeeee | 男 | 40 |  |

（2）住址为缺失值的统一填充为：深圳；姓名字符长度范围（0-6），年龄填充为23，否则填充为 20。填充后的内容保存为“缺失值.xls”。

6、利用Kettle工具实现XML文件的抽取。提交：XML 抽取.ktr文件、构建的数据文件、生成的数据文件。

要求：

（1）构建源数据文件“XML抽取.xml”，数据内容：

<?xml version="1.0" encoding="UTF-8"?>

<AllRows description = "xml test ">

<Rows rowlD = "1"> first row chunk of data

<Row> <v1> 1.1.1 </v1> <v2> 1.1.2 </v2> </Row>

<Row> <v1> 1.2.1 </v1> <v2> 1.2.2 </v2> </Row>

</Rows>

<Rows rowlD = "2"> second row chunk of data

<Row> <v1> 2.1.1 </v1> <v2> 2.1.2 </v2> </Row>

<Row> <v1> 2.2.1 </v1> <v2> 2.2.2 </v2> </Row>

</Rows>

</AllRows>

（2）结果保存为Excel文件“XML 抽取.xls”，抽取内容：

|  |  |  |  |
| --- | --- | --- | --- |
| description | rowID | v1 | v2 |
| xml test | 1 | 1.1.1 | 1.1.2 |
| xml test | 1 | 1.2.1 | 1.2.2 |
| xml test | 2 | 2.1.1 | 2.1.2 |
| xml test | 2 | 2.2.1 | 2.2.2 |