

# lab1说明文档

## 使用方法

使用JDK17

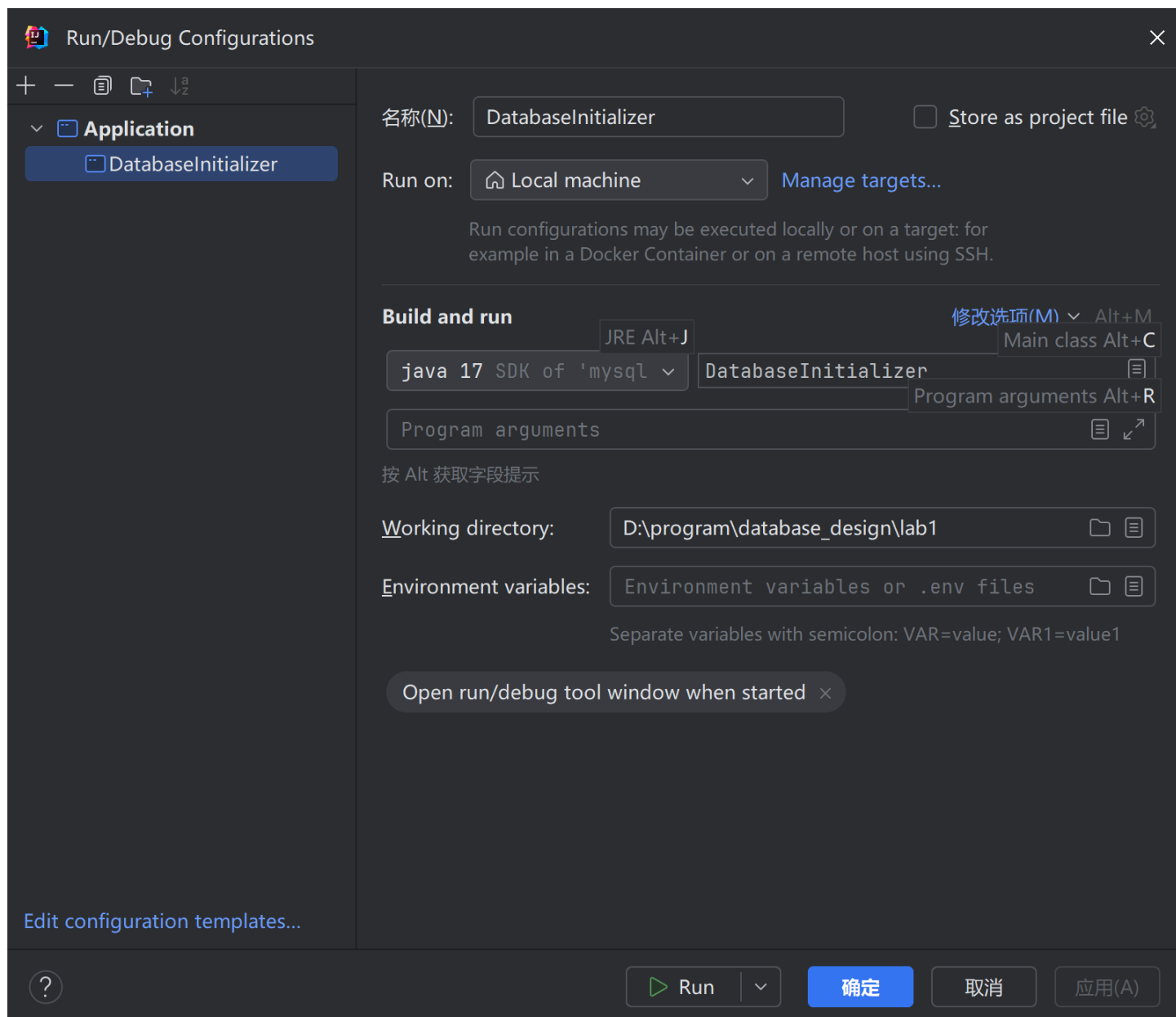
用IDEA打开文件夹 lab1

maven下载依赖

将DatabaseInitializer类里的csv\_to\_database函数的第5、6、7行的for\_database\_lab1、root、123456改成符合自己本地database\_name,root,password。

```
String dbUrl = "jdbc:mysql://localhost:3306/for_database_lab1?useUnicode=true&characterEncoding="+charsetName;
String dbUser = "root";
String dbPassword = "123456";
```

然后run就好啦



得到结果：

1 ✓

2 ✓

select \* from room;

select \* from student;

数据库应用接口.d

输出

for\_database\_lab1.room

for\_database\_lab1.student

Tx

	"registno" ▾	"name" ▾	"kdno" ▾	"kcno" ▾	"ccno" ▾	"seat" ▾
1	"0313016"	"苏雪菲"	"1101"	"2"	"1"	"0"
2	"0314026"	"徐文燕"	"1101"	"2"	"1"	"0"
3	"0332013"	"吴春华"	"1101"	"2"	"1"	"0"
4	"0342019"	"杭燕婷"	"1101"	"2"	"1"	"0"
5	"0342026"	"顾凡"	"1101"	"2"	"1"	"0"
6	"0342032"	"莫康妮"	"1101"	"2"	"1"	"0"
7	"0342033"	"吴嘉蕾"	"1101"	"2"	"1"	"0"
8	"0342034"	"陆瑛"	"1101"	"2"	"1"	"0"
9	"0342057"	"陈岑"	"1101"	"2"	"1"	"0"
10	"0342059"	"王展"	"1101"	"2"	"1"	"0"
11	"0357005"	"杨佩"	"1101"	"2"	"1"	"0"
12	"0357006"	"郑磊"	"1101"	"2"	"1"	"0"
13	"0357009"	"查成军"	"1101"	"2"	"1"	"0"
14	"0357021"	"白鸽"	"1101"	"2"	"1"	"0"
15	"0357023"	"黄山"	"1101"	"2"	"1"	"0"
16	"0357024"	"王婧"	"1101"	"2"	"1"	"0"
17	"0357025"	"鲍静"	"1101"	"2"	"1"	"0"
18	"0361051"	"陈程"	"1101"	"2"	"1"	"0"
19	"0361059"	"张锦"	"1101"	"2"	"1"	"0"
20	"0361060"	"邢智杰"	"1101"	"2"	"1"	"0"
21	"0361063"	"张斌"	"1101"	"2"	"1"	"0"
22	"0361067"	"余骏"	"1101"	"2"	"1"	"0"

	kdno	kcno	ccno	kdname	exptime	papername
1	1101	1	1	复旦大学	2004-06-10 8:00	
2	1101	1	2	复旦大学	2004-06-10 8:20	
3	1101	1	3	复旦大学	2004-06-10 8:40	
4	1101	1	4	复旦大学	2004-06-10 9:00	
5	1101	1	5	复旦大学	2004-06-10 9:20	
6	1101	1	6	复旦大学	2004-06-10 9:50	
7	1101	1	7	复旦大学	2004-06-10 10:10	
8	1101	1	8	复旦大学	2004-06-10 10:30	
9	1101	1	9	复旦大学	2004-06-10 10:50	
10	1101	1	10	复旦大学	2004-06-10 11:10	
11	1101	1	11	复旦大学	2004-06-10 12:00	
12	1101	1	12	复旦大学	2004-06-10 12:20	
13	1101	1	13	复旦大学	2004-06-10 12:40	
14	1101	1	14	复旦大学	2004-06-10 13:00	
15	1101	1	15	复旦大学	2004-06-10 13:20	
16	1101	1	16	复旦大学	2004-06-10 13:40	
17	1101	1	17	复旦大学	2004-06-10 14:00	
18	1101	1	18	复旦大学	2004-06-10 14:20	
19	1101	1	19	复旦大学	2004-06-10 14:40	
20	1101	1	20	复旦大学	2004-06-10 15:20	
21	1101	1	21	复旦大学	2004-06-10 15:40	
22	1101	1	22	复旦大学	2004-06-10 16:00	

## 思考题

如果外部数据（原始数据表）数据不完整（例如某个不应该为空的字段缺失数据）或不一致（例如本应有外键关系的数据并没有保持引用完整性），有哪些方法可以处理？

- 填充缺失的数据，可以使用默认值进行填充。
- 建立异常处理机制来处理无法清洗或转换的异常情况。
- 记录并报告无法处理的数据，以便进一步的分析和解决。

你能总结出哪些处理原始数据的原则？

原始数据的类型、格式、模式、范围要求可能都不一样，建议数据预处理的时候写一个标准使得处理过后的数据都有一样的标准。

处理数据的时候要保证数据不乱码，表示准确的完整意思。

原始数据需要保留。

## debug

1. 因为发现两个文件编码类型不一样，所以用一个检测编码类型的函数来得出编码类型

```
1 String charsetName = isUTF8(path);
```

然后在写入的时候和读入的时候都使用这个编码

```
1 //以这种编码类型写入
2 String dbUrl = "jdbc:mysql://localhost:3306/for_database_lab1?
  useUnicode=true&characterEncoding="
3           +charsetName;
4
5 //以这种编码类型读入
6 try (BufferedReader br = new BufferedReader(new InputStreamReader(new
  FileInputStream(path), charsetName))) {
```

2. 在建表时，列名要用 ‘来包括否则如果遇到 “...” 的会报错

```
1 for (String column : columns) {
2     //这里要加单引号，否则会报错：“通常使用反引号 (``) 来引用表名和列名，而不是双引号。”
3     createTableSQL += "``" column + "` VARCHAR(255), ";
4 }
```

3. 在读入一行数据的时候不能放过任何一个数据否则会报错

```
1 String[] data = line.split(",", -1); //TODO:这里要加-1，否则不能拿到最有一个空元素
```