Do\_Tool生成工具使用说明

北京先进数通信息技术有限公司

编写说明

中文标题：大数据文档、SQL、DLF生成工具使用说明

文档类型：管理

版本历史：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **版本** | **作者** | **修改日期** | **修改内容** | **审核** | **批准** |
| V1.0.0 | 郭宇航、任艺 | 2021-08-30 | 初稿 |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

目 录

[编写说明 1](#_Toc6873)

[目 录 i](#_Toc22709)

[1. 引言 2](#_Toc29080)

[1.1. 背景 2](#_Toc30713)

[1.2. 定义 2](#_Toc22633)

[1.3. 参考 2](#_Toc26707)

[2. 功能说明 3](#_Toc16663)

[2.1. 源系统表来源分析 3](#_Toc2223)

[2.1.1第一种：连接数据库导出表结构 3](#_Toc14828)

[2.1.2第二种：通过api返回的数据结构 4](#_Toc21782)

[2.2. 相关文档设计 4](#_Toc25127)

[2.2.1 贴源层文件设计 4](#_Toc7612)

[2.2.2 主题层文件设计 6](#_Toc1872)

[2.2.3 专题层文件设计 7](#_Toc26504)

[2.3. 生成相关TABLE\_SQL 7](#_Toc23090)

[2.3.1源系统SQL 7](#_Toc18865)

[2.3.2主题层SQL 8](#_Toc24691)

[2.3.3专题层SQL 9](#_Toc31074)

[2.4. 生成相关DLF 9](#_Toc12327)

[2.4.1源系统到主题层的DLF 9](#_Toc31633)

[2.4.2主题层到专题层的DLF 10](#_Toc11479)

# 引言

## 背景

## 定义

介绍生成文档表格、创建表SQL和数据清洗过程DLF的使用方法。

## 参考

# 功能说明

## 源系统数据结构分析

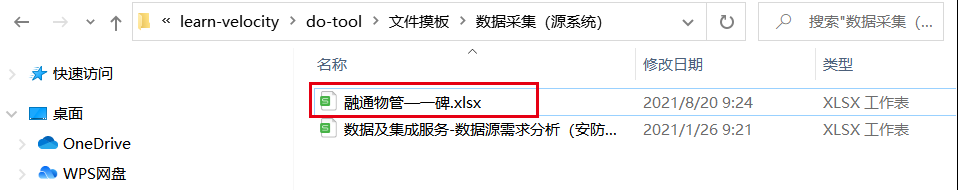
本教程只介绍第一种：分析源系统数据库表结构

### **2.1.1第一种：源系统数据库表结构**

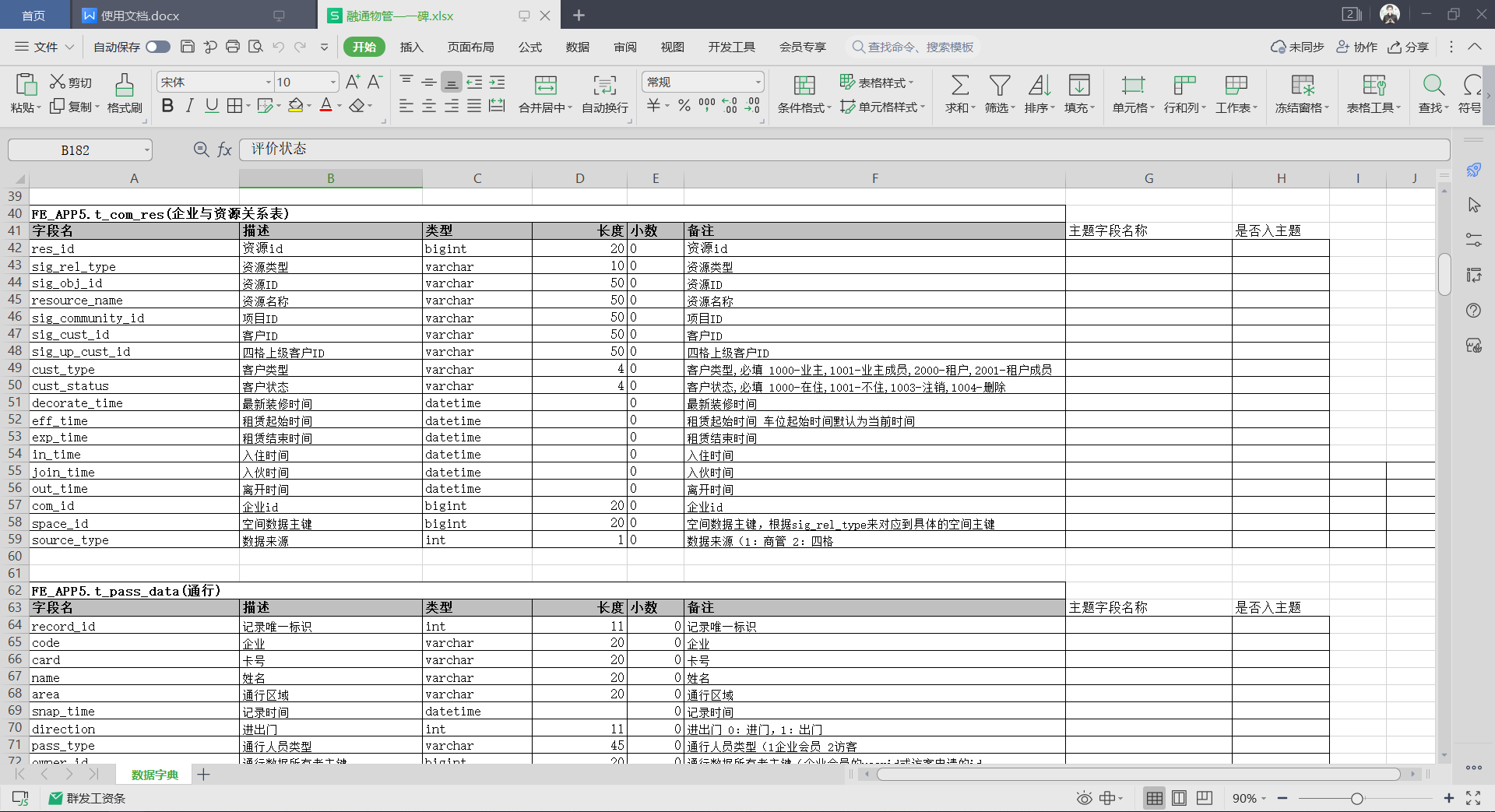
湖表只是增加了审计、批处理字段。

根据表结构整理表结构文档“融通物管—一碑.xlsx”

1、需要准备一个源文件Excel文件（文件摸板\数据采集（源系统）\融通物管—一碑.xlsx）、放到D盘根目录下。



1. 文件内容要求



第一行“ID”行是必须存在的，剩下的行是从数据库表中数据段移动过来，主题字段名称可以自定义主题表字段名称，是否入主题列表填写n后可以不在主题建立该字段。

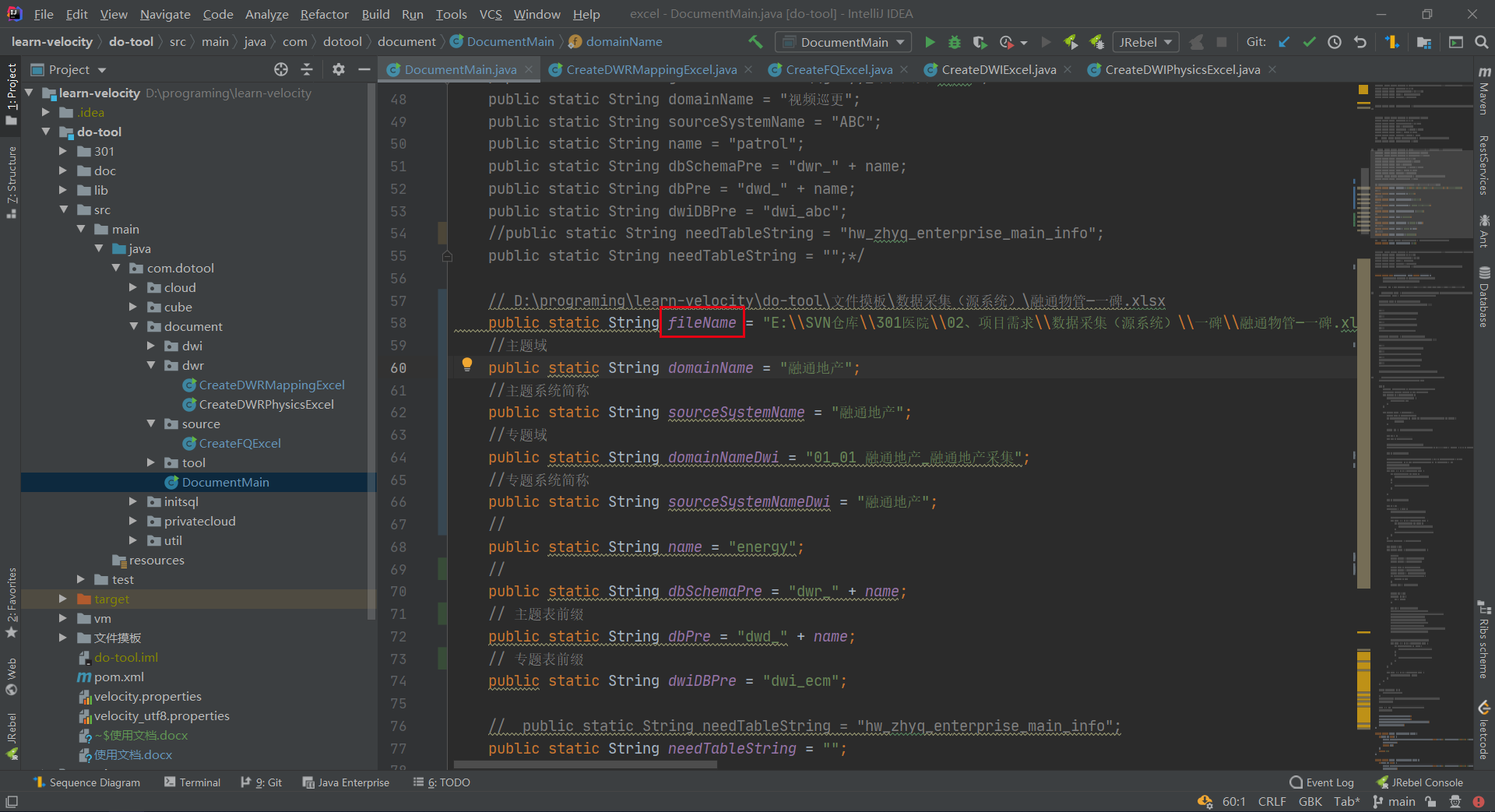
### **2.1.2第二种：源系统接口数据结构**

根据对接系统的API接口返回的数据分析并整理数据结构表。

## 相关文档设计

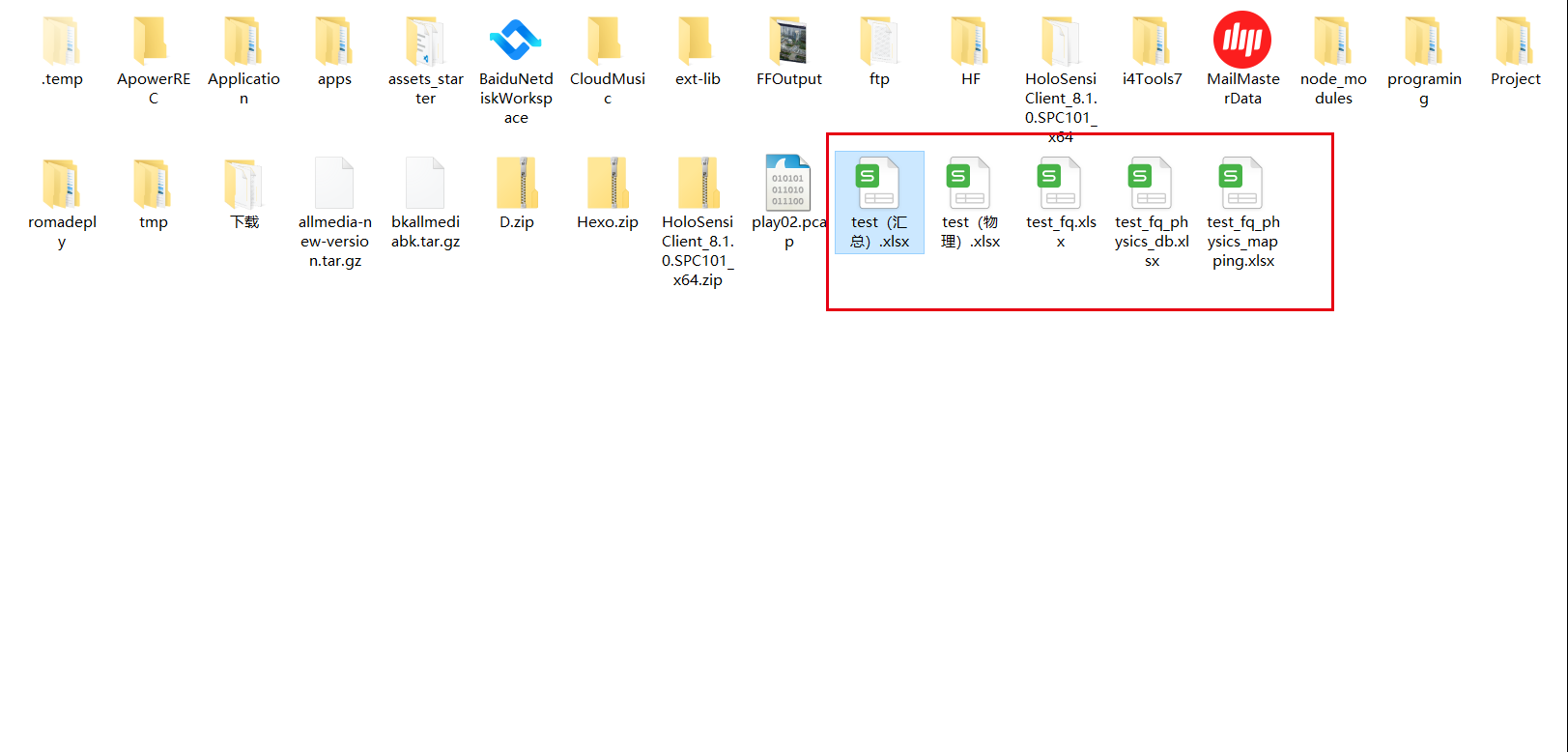
### **2.2.1 贴源层文件设计**

#### **2.2.1.1运行程序：**



修改fileName变量为源文件Excel路径“D:\programing\learn-velocity\do-tool\文件摸板\02、项目需求\数据采集（源系统）\商管\北京华贸-商管.xlsx。

生成5个文件，路径在“D:\”，3个是贴源层的，2个是主题层的。



#### 2.2.1.2 test（汇总）.xlsx：

源系统到贴源层湖的mapping，该文件中的表结构会增加4个字段，为入湖记录标识字段：

|  |  |
| --- | --- |
| dw\_creation\_by | 数据创建者 |
| dw\_creation\_date | 数据创建时间 |
| dw\_last\_update\_by | 数据最后更新者 |
| dw\_last\_update\_date | 数据最后更新时间 |
| dw\_batch\_number | 数据批次号 |

1）拷贝到“03、系统设计\详细设计\贴源层\商管\02-Mapping”下的“DWI-商管-Mapping分析表-V1.0.xlsx”中，并整理数据资源sheet。

2）拷贝到“D:\programing\learn-velocity\do-tool\文件摸板\03、系统设计\详细设计\贴源层\商管”下的“北京华贸-数据及集成服务-DWI设计汇总表-V1.0.xlsx”中。并设计“数据资源”的相关信息：业务分类(数据来源)、是否入湖、不入湖原因、采集频率、数据库表名、贴源层-sechma(根据源系统设计)、贴源层-表名(自己设计，没info加info后缀)、贴源层-表中文名、贴源层-FDI任务名称、贴源层-任务描述。

后续生成湖表结构SQL时使用该表。

**问题**： mapping里的对应源系统表字段的character varying类型不对，应该整理的那个表类型一致，修改。

#### 2.2.1.3 **test（物理）.xlsx：**

贴源层的物理表,会增加4个字段：

拷贝到“D:\programing\learn-velocity\do-tool\文件摸板\03、系统设计\详细设计\贴源层\商管\01-模型设计”下的“DWI-商管-物理模型设计V1.0.xlsx”中。

#### 2.2.1.4 **test\_fq.xlsx:**

源系统的物理表，没有4个字段。

拷贝到“D:\programing\learn-velocity\do-tool\文件摸板\02、项目需求\数据采集（源系统）\商管”下的“数据及集成服务-数据源数据分析(商管)-V1.0.0.xlsx”表格中，并填写数据资源中的：是否入湖、采集策略、采集频率、采集策略、数据库表名等。

也可在没有提供数据库的时候用来做假源系统数据库，需要手写创建表sql。

### **2.2.2 主题层文件设计**

使用刚才在2.2.1.1中生成的5个文件的2个文件： test\_fq\_physics\_mapping.xlsx、test\_fq\_physics\_db.xlsx，根据贴源层的表分析主题要那几块内容，然后编写物理表和mapping表(先哪个都可以)。

华贸的分为：工单、空间、能耗、企业、人员、通行、组织。

#### 2.2.2.1 **test\_fq\_physics\_db.xlsx：**

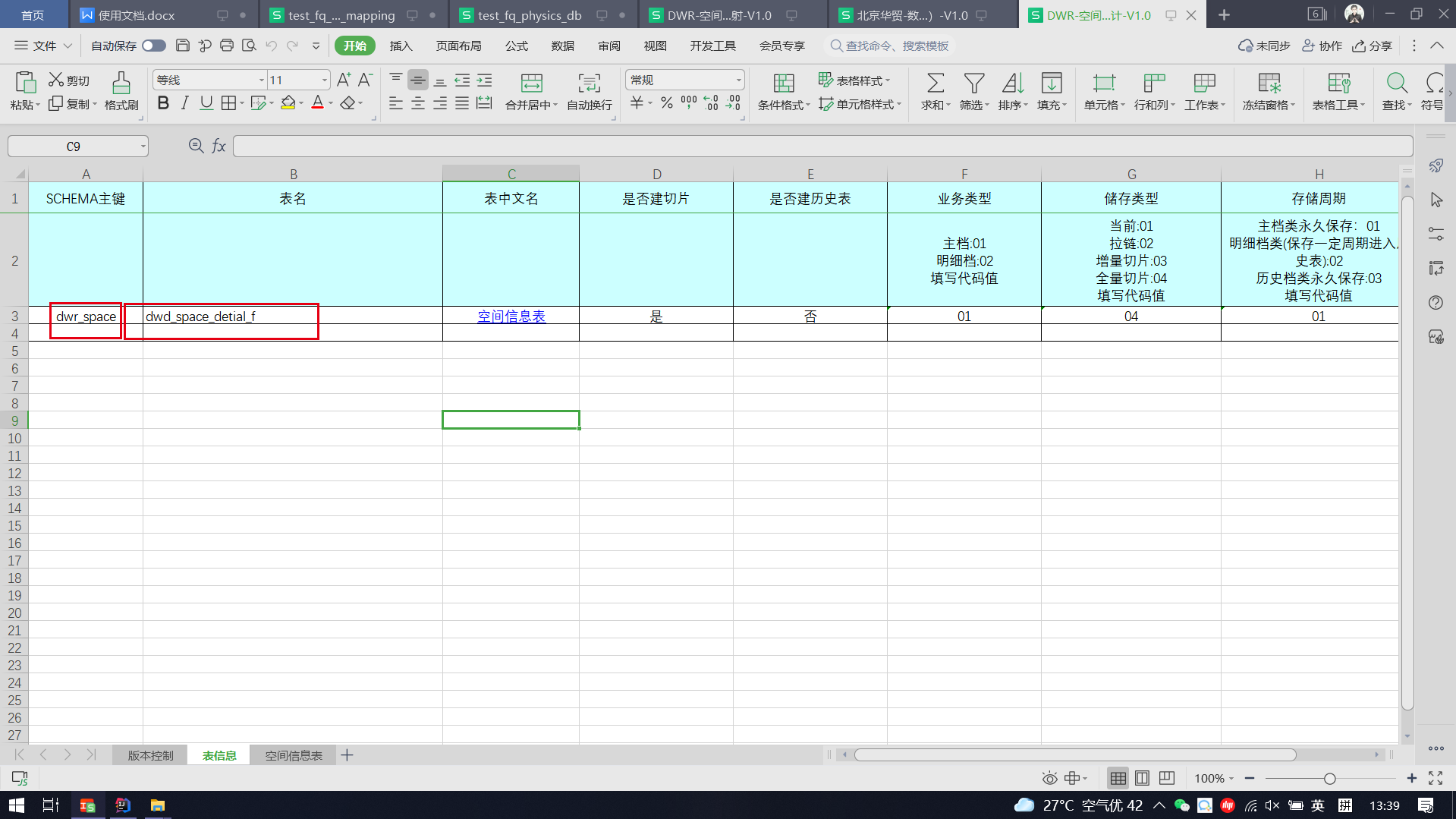
主题的物理表：

放到“D:\programing\learn-velocity\do-tool\文件摸板\系统设计\详细设计\主题层\工单\02物理模型设计”下的DWR-工单-物理表设计-V1.0.xlsx文件。

生成主题表sql时用。

#### 2.2.2.1 **test\_fq\_physics\_mapping.xlsx：**

主题表mapping，湖表到主题的1比1，又多加了1个系统id字段。在这里设计表，在表信息中设计schma和表名(加\_f后缀)。



放到“D:\programing\learn-velocity\do-tool\文件摸板\系统设计\详细设计\主题层\空间\03Mapping”下的DWR-空间-MAPPING映射-V1.0.xlsx表中。

#### 2.2.2.1 北京华贸-数据及集成服务-主题设计（空间）-V1.0.xlsx**：**

这张表是集成的主题中的物理表。

位置：D:\programing\learn-velocity\do-tool\文件摸板\系统设计\详细设计\主题层\空间

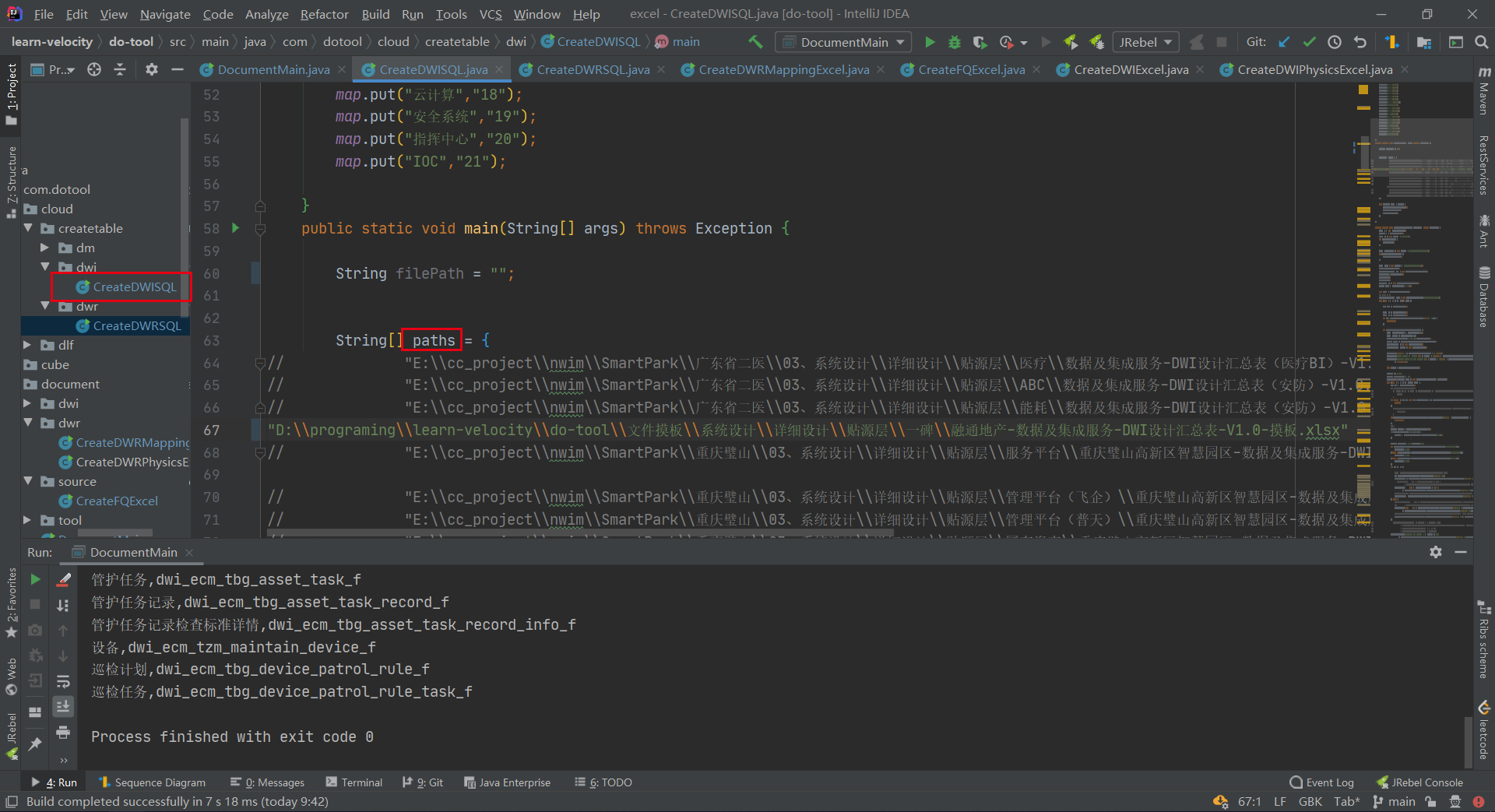
### **2.2.3 专题层文件设计**

根据主题表设计出表字段，并在表摸板文件中设计出专题表。

## 生成相关TABLE\_SQL

### **2.3.1源系统SQL**

运行程序：

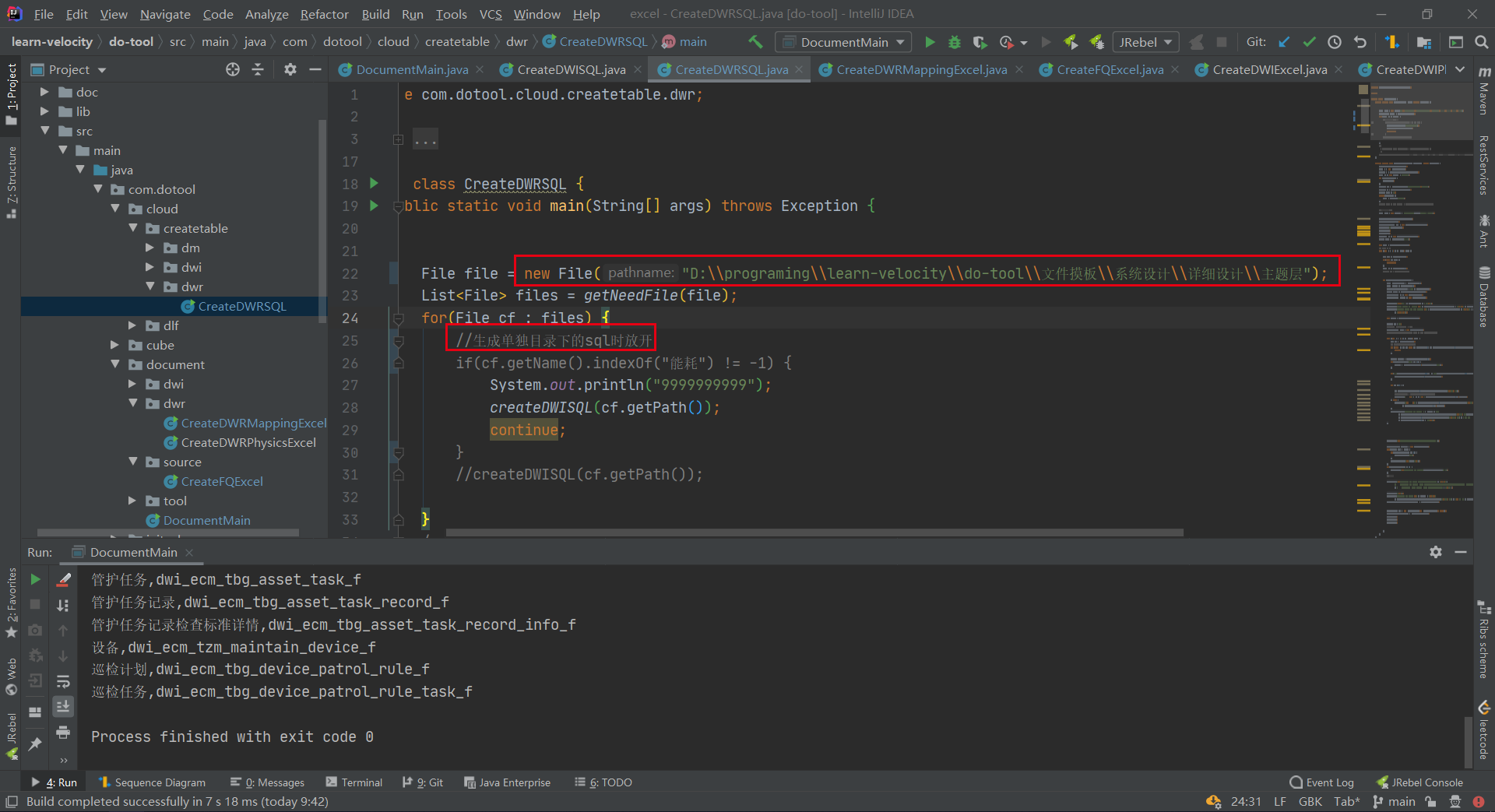


路径“D:\programing\learn-velocity\do-tool\文件摸板\系统设计\详细设计\贴源层\一碑\融通地产-数据及集成服务-DWI设计汇总表-V1.0-摸板.xlsx”。

“dwi\_sql.txt”文件为dayu中创建表的sql。

### **2.3.2主题层SQL**

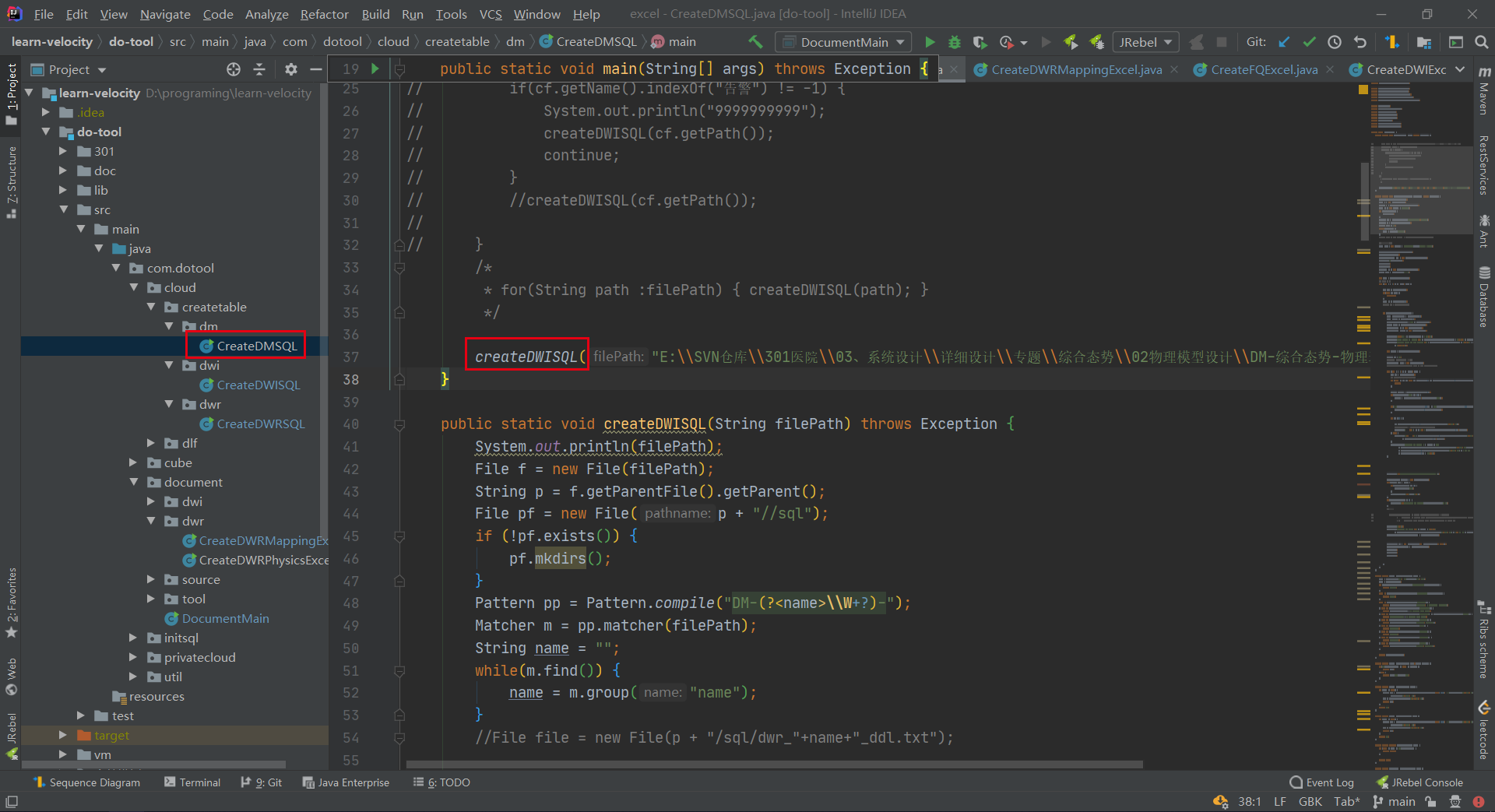
运行程序：



路径：“D:\programing\learn-velocity\do-tool\文件摸板\系统设计\详细设计\主题层”。

### **2.3.3专题层SQL**

运行程序：

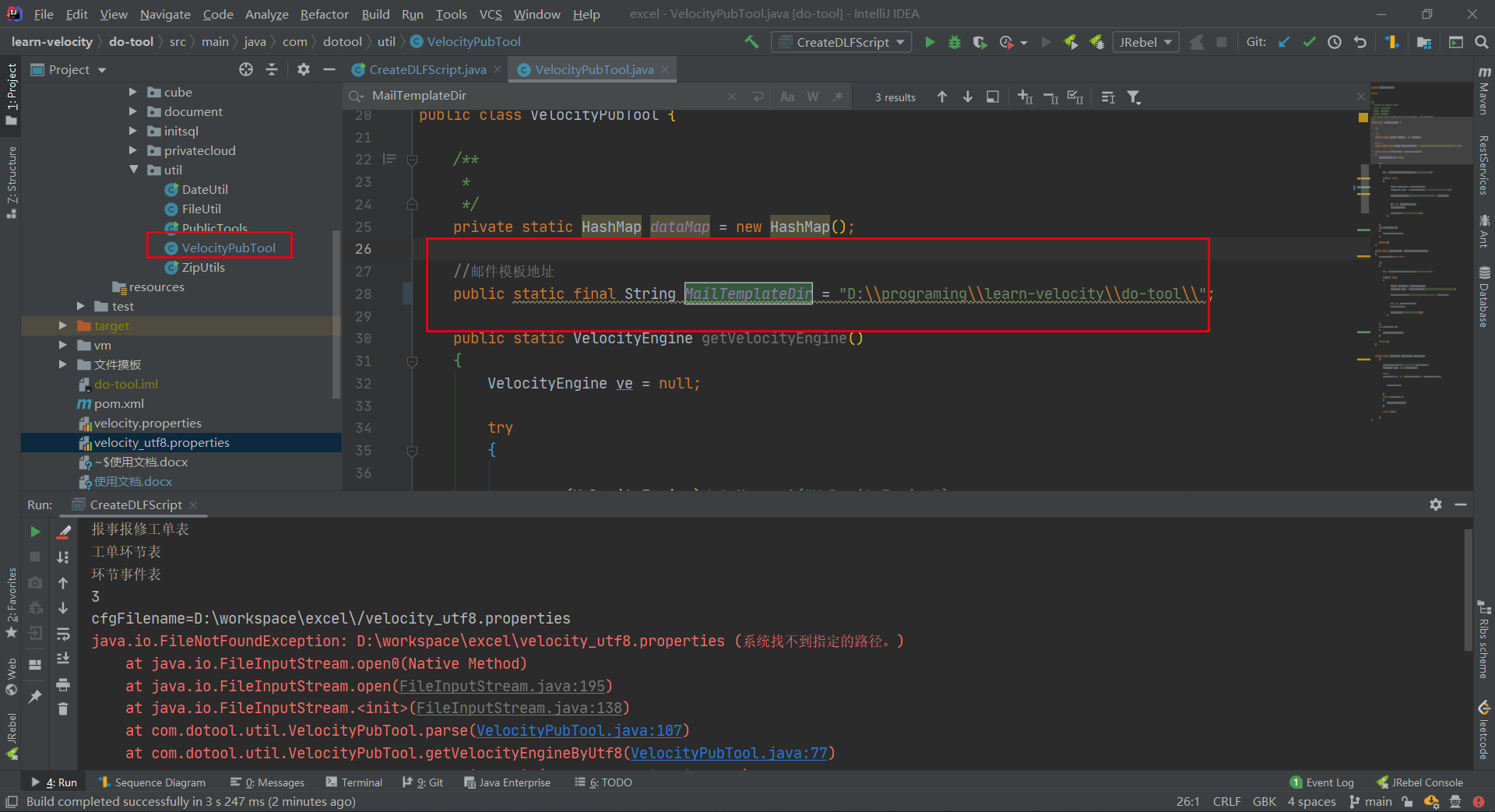


路径：“E:\SVN仓库\301医院\03、系统设计\详细设计\专题\综合态势\02物理模型设计\DM-综合态势-物理表设计-V1.0.xlsx”。

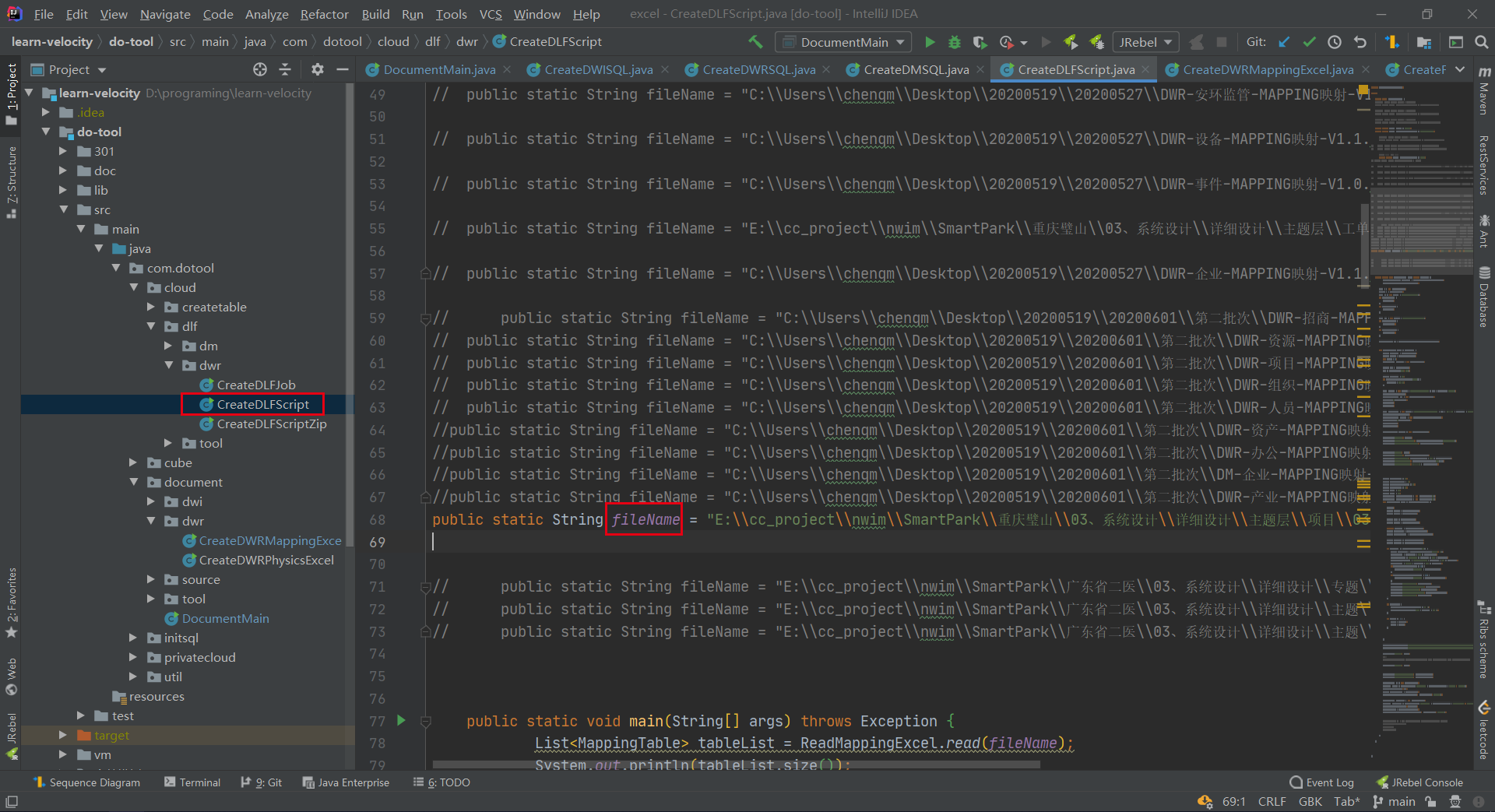
## 生成相关DLF

### **2.4.1源系统到主题层的DLF**

修改路径

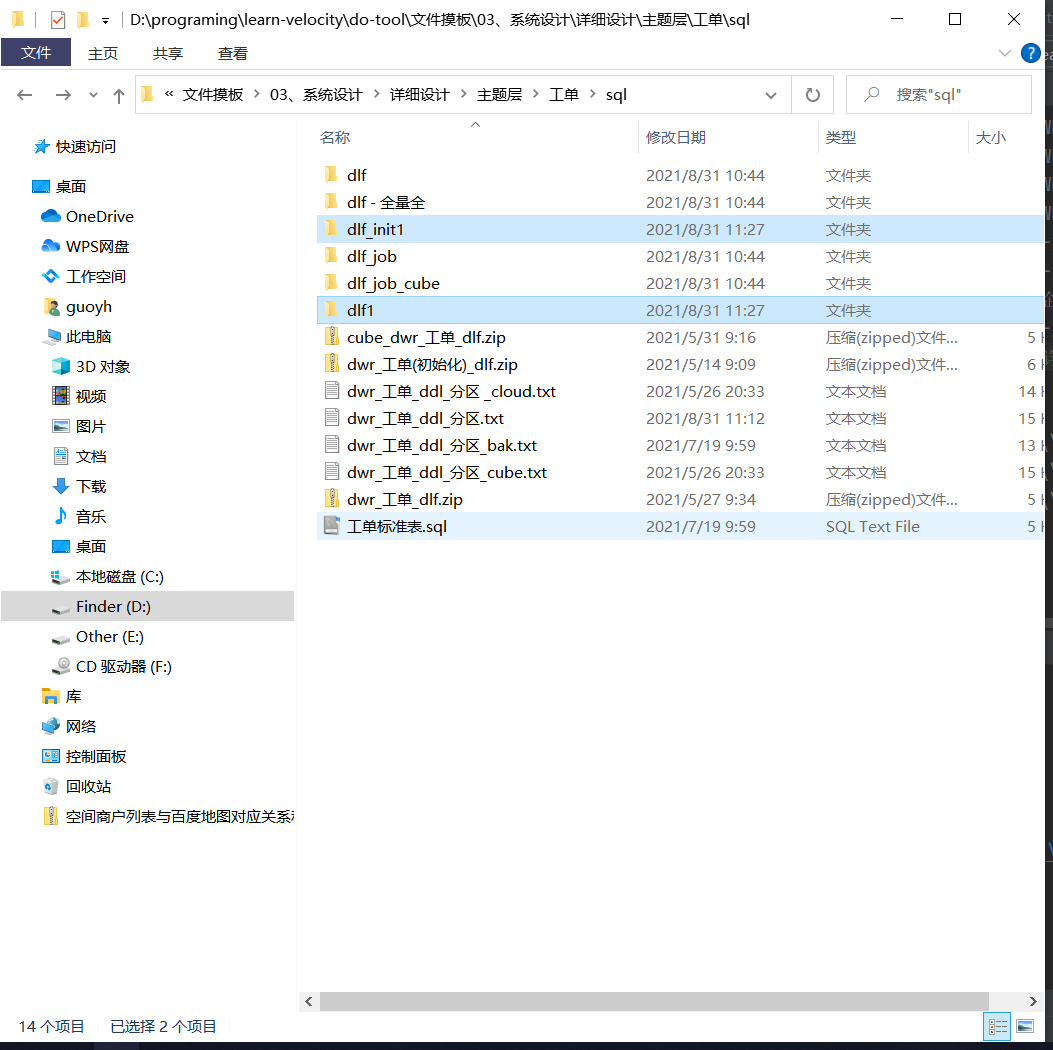


运行程序：



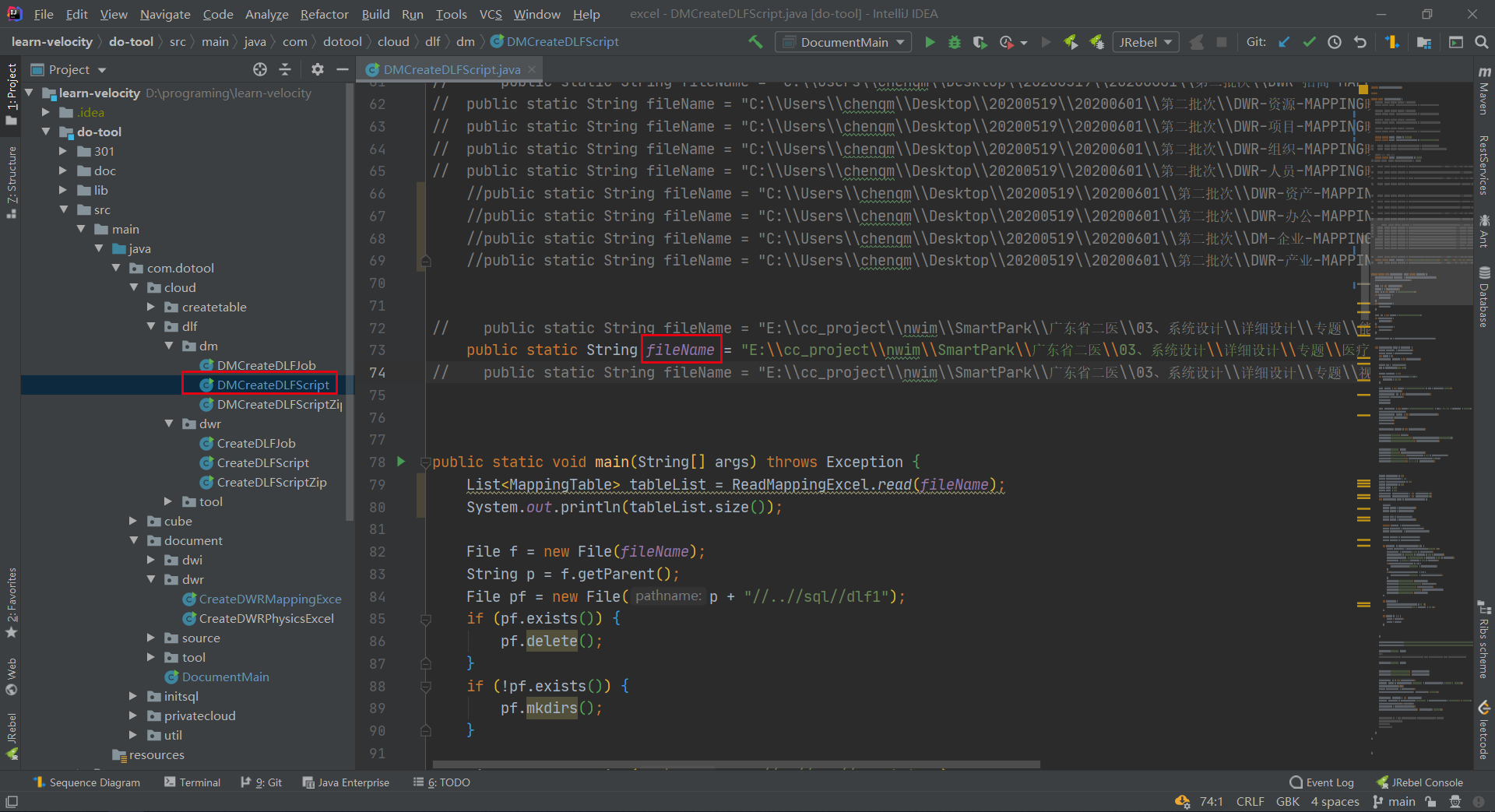
fileName = "E:\\cc\_project\\nwim\\SmartPark\\重庆璧山\\03、系统设计\\详细设计\\主题层\\项目\\03Mapping\\DWR-项目-MAPPING映射-V1.1.xlsx"。

文件输出路径：dlf1、dlf\_init1



### **2.4.2主题层到专题层的DLF**

运行程序：



fileName = "E:\\cc\_project\\nwim\\SmartPark\\广东省二医\\03、系统设计\\详细设计\\专题\\医疗\\03Mapping\\DM-医疗-MAPPING映射-V1.0.xlsx";

文件输出路径：dlf1、dlf\_init1