密级：□一般☑重要□机密

**京柏医疗**

**妇幼信息集成共享平台**

**操作手册**

2017-1-10

**目 录**

[1 系统简介 4](#_Toc7905)

[2 集成门户部分 4](#_Toc20993)

[3 统一认证部分 4](#_Toc14120)

[4 ETL采集部分 4](#_Toc19979)

[4.1 资源库 4](#_Toc12424)

[4.1.1 资源库类型 4](#_Toc25873)

[4.1.2 资源库的创建 4](#_Toc32553)

[4.1.3 修改密码 11](#_Toc26298)

[4.1.4 导入导出资源库 12](#_Toc14252)

[4.1.5 删除、修改资源库转换、作业名称 15](#_Toc29268)

[4.1.6 运行日志 16](#_Toc11412)

[4.1.7 常见工具组件 17](#_Toc10008)

[4.2 数据采集 17](#_Toc17391)

[4.2.1 前置采集任务 17](#_Toc28594)

[4.2.2 缓存采集任务 19](#_Toc12102)

[4.2.3 增量采集任务 20](#_Toc2790)

[4.2.4 BI报表采集任务 20](#_Toc16417)

[4.3 采集任务日志 21](#_Toc17036)

[4.3.1 采集作业日志 21](#_Toc9638)

[4.3.2 采集转换日志 23](#_Toc2872)

[5 EMPI部分 25](#_Toc1041)

[5.1 合并患者 25](#_Toc23109)

[5.2 拆分患者 26](#_Toc7456)

[5.3 患者主索引配置 27](#_Toc11551)

[5.4 域信息查询 28](#_Toc14499)

[5.5 PIX查询 29](#_Toc21505)

[5.6 患者信息查询 29](#_Toc14037)

[5.7 日志查询 30](#_Toc25476)

[6 数据标准部分 31](#_Toc16931)

[6.1 标准化管理 31](#_Toc2988)

[6.1.1 基础数据管理 31](#_Toc1168)

[6.1.2 值域代码管理 32](#_Toc18519)

[6.1.3 数据元管理 34](#_Toc9433)

[6.1.4 数据集管理 35](#_Toc15935)

[6.1.5 数据集分类管理 38](#_Toc12796)

[6.1.6 数据批量导入 39](#_Toc28906)

[6.2 平台资源维护 39](#_Toc6110)

[6.2.1 项目字典维护 39](#_Toc8510)

[6.2.2 药品字典维护 40](#_Toc19989)

[6.2.3 诊断字典维护 41](#_Toc24983)

[6.2.4 手术字典维护 42](#_Toc14894)

[6.2.5 组织机构维护 43](#_Toc4997)

[6.2.6 人员信息维护 44](#_Toc18653)

[6.3 存储模型管理 45](#_Toc25769)

[6.3.1 数据源管理 45](#_Toc7677)

[6.3.2 主题库管理 47](#_Toc19269)

[6.3.3 存储模型管理 48](#_Toc31056)

[7 BI部分 50](#_Toc18554)

# 系统简介

信息共享平台包含集成门户、统一认证、ETL数据采集、患者主索引（EMPI）、BI（商业智能）、数据标准六部分，下面分述之。

# 集成门户部分

# 统一认证部分

# ETL采集部分

## 资源库

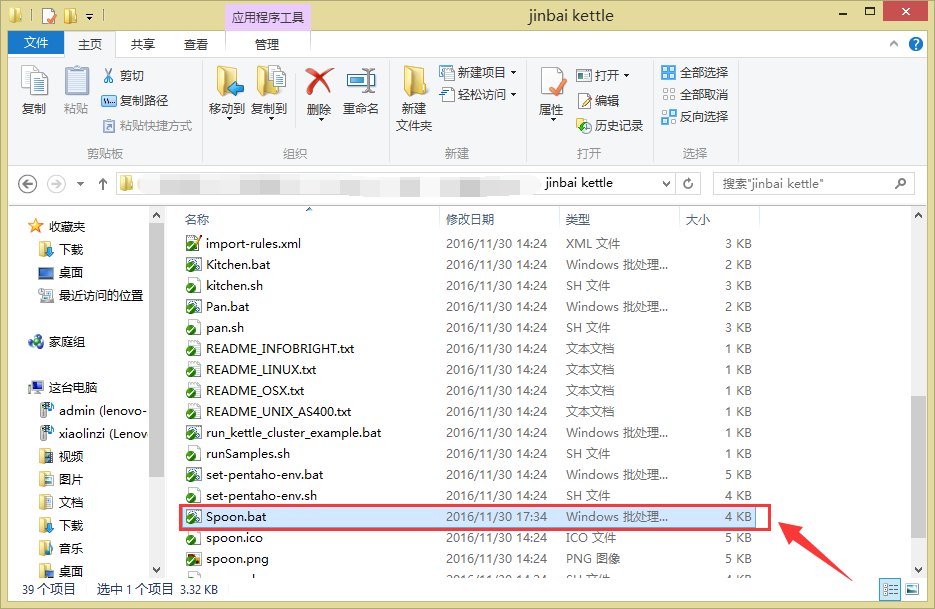
用户通过kettle图形界面创建的trans转换、job作业任务，可以保存在资源库当中。资源库的存在，一方面是能够共享转换任务给多个用户，一方面可以通过文件夹的形式进行分组管理，用户可以创建文件夹并且进行更改，方便管理。

### 资源库类型

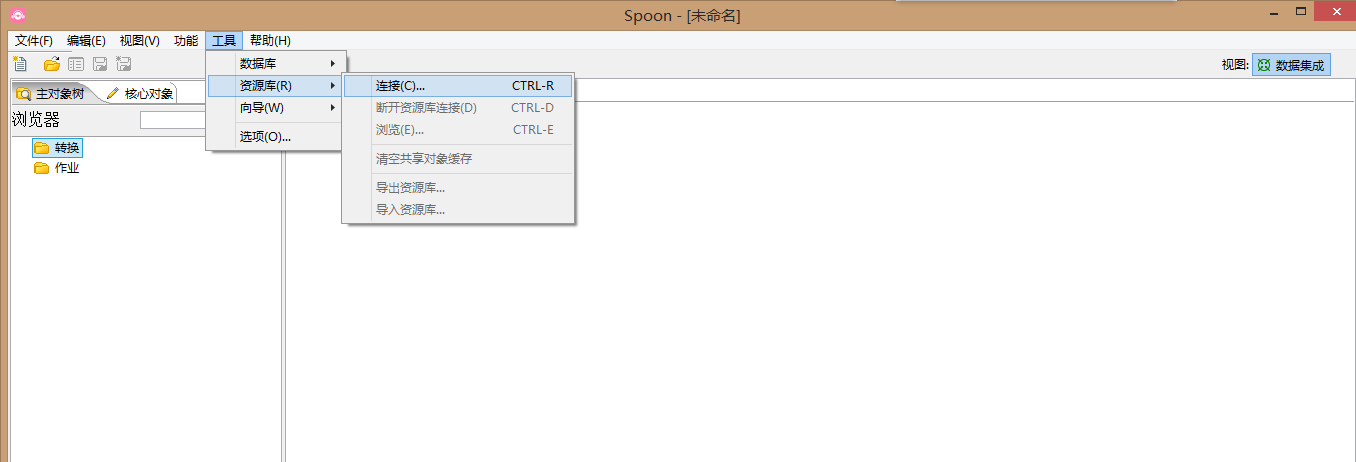
资源库有两种形式。常见的是保存在数据库的资源库类型（Kettle database repository），允许用户通过用户名、密码访问资源库的资源，默认用户名、密码是admin/admin和guest/guest；另一种保存在服务器硬盘文件夹的资源库类型（Kettle file repository），无需登录，可直接进行操作。

### 资源库的创建

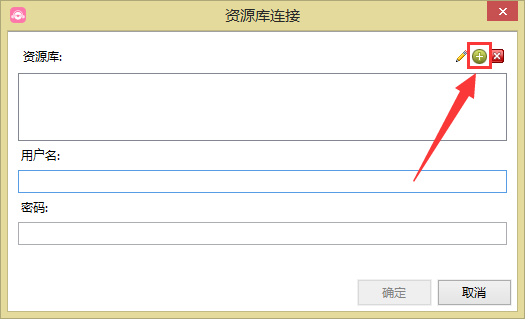
* 1. 启动kettle：选择执行【Spoon.bat】，如下图所示：



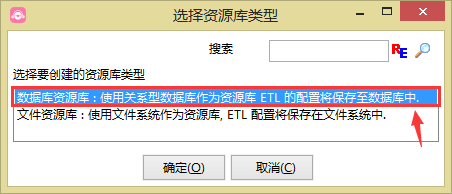
* 1. 连接资源库：
     1. 选择工具，数据库，连接，如下图所示：



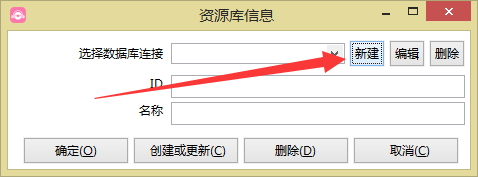
* + 1. 弹出“资源库连接”窗口，点击符号【+】，如下图所示：



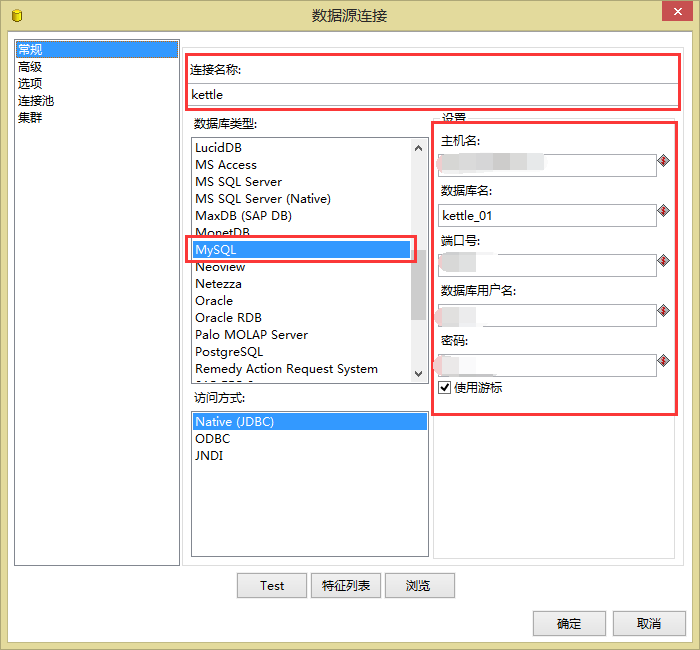
* + 1. 弹出“选择资源库类型”窗口，选择数据库资源库，点击【确定】。如下图所示：



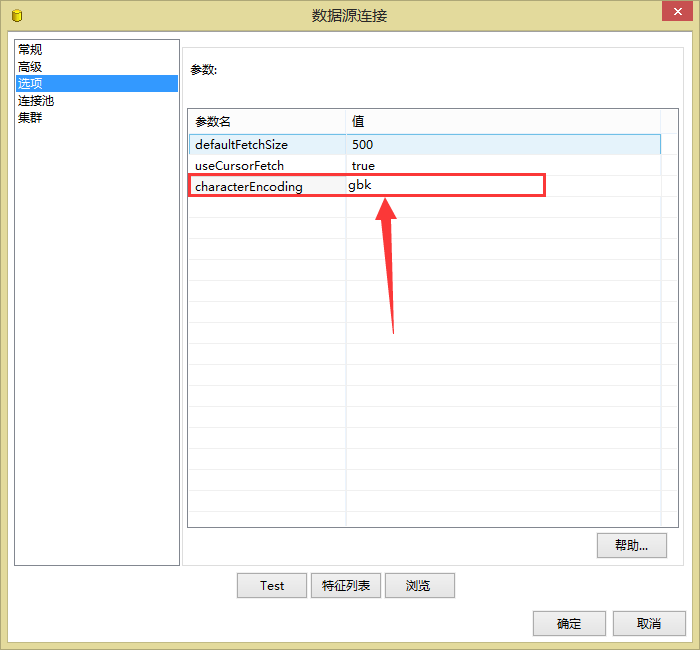
* + 1. 弹出“资源库信息”窗口，选择【新建】，如下图所示：



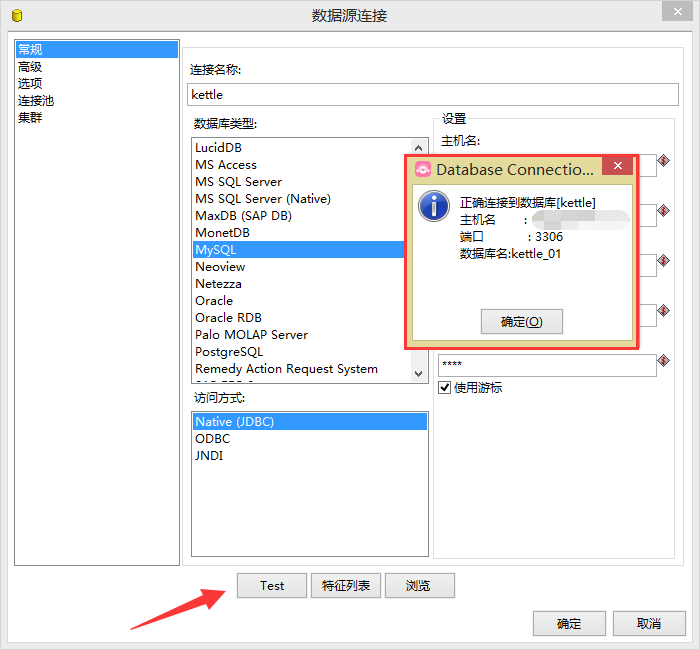
* + 1. 弹出“资源库信息”窗口，选择【新建】，如下图所示：即出现数据源连接弹窗，在弹窗输入对应信息（红色框部分）



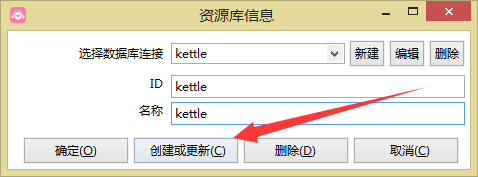
* + 1. 选择【选项】，新增一条参数，参数名为characterEncoding，值为gbk。如下图所示：



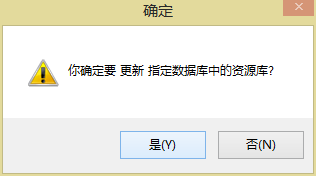
* + 1. 此时可点击【Test】进行测试，判断是否成功连接。成功即选择确认，如下图所示。否则检查对应配置信息是否正确。



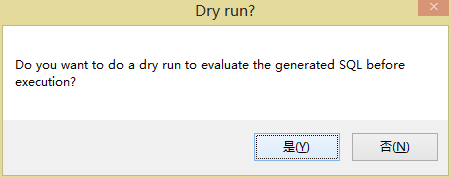
* + 1. 在资源库信息填入对应信息，如下图所示，点击【创建或更新】。



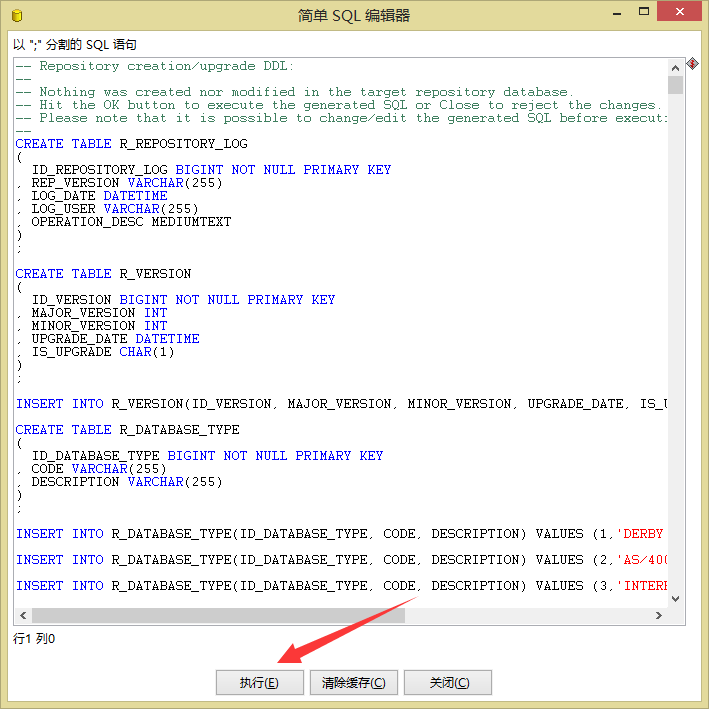
如下图所示，点击【是】



如下图所示，选择【是】：

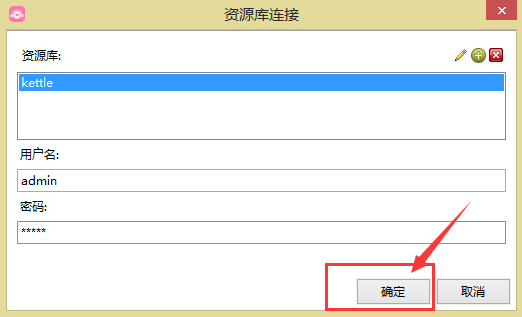


* + 1. 等待几秒后，点击执行，出现简单sql编辑器，如下图所示。点击【执行】。



当执行完成后，将返回执行结果。确定并关闭后，回到资源库信息，选择确定，返回资源库连接。

1. 进入采集工具管理界面
   1. 在资源库连接中填入用户名和密码（默认均为admin或guest），如下图所示，点击确认，即可进入采集工具主页面：



注意：SqlSerevr需要拷贝脚本到数据库上，替换bit类型的值‘Y’/‘N’后执行SQL语句。

其他支持‘Y’/‘N’数据库可以直接点击“执行”按钮执行。

在弹出的“简单SQL编辑器”窗口中，点击“执行”按钮，执行相应的资源库创建脚本。（用户可以查看这些脚本具体是进行了什么操作：创建了42张表，并插入了一些相关信息数据，如R\_USER表里插入了admin和guest两个用户的信息，这两个用户也是该资源库的登录用户，想深入研究资源库的用户可以并一步去了解相关内容）

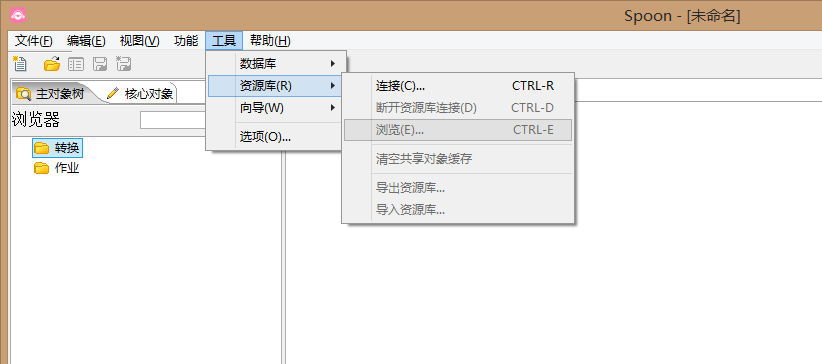
如下图所示，在“资源库信息”窗口，填写资源库的名称和描述，然后点击“确定”按钮。

温馨提示：建议使用英名，因为使用中文容易出现乱码，而使得退出后，下次登录资源库时无法找到相应的配置，遇到此类情况，可根据下一章节“资源库名称乱码问题解决方法”进行相关修改设置。



### 修改密码

为了资源库的安全，我们需要修改资源库用户的登录密码，“工具” (“资源库”(“浏览”，如下图所示。也可以使用快捷键CTRL + E。



选择“安全”选项卡，然后选择需要修改的用户再点击编辑按钮，如下图所示：

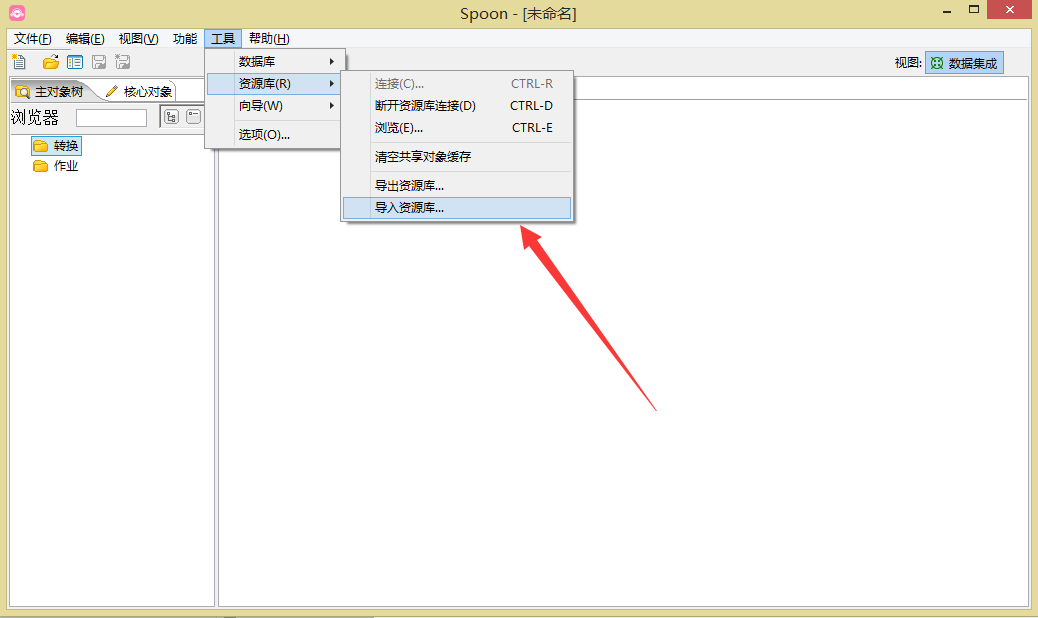


填写新密码，然后点击“确定”按钮即可完成用户密码的修改，如下图所示：

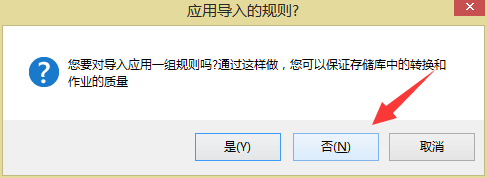


### 导入导出资源库

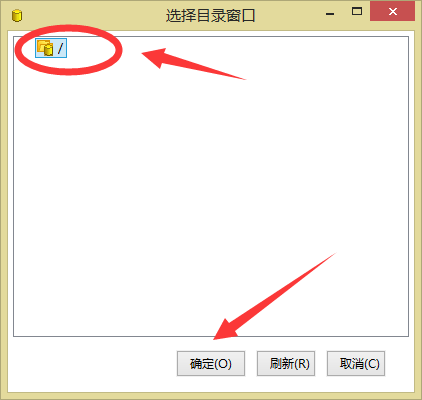
将本地（转换、作业或其他）文件导入到资源库里。选择“工具”-->“资源库”-->“浏览”，如下图所示：



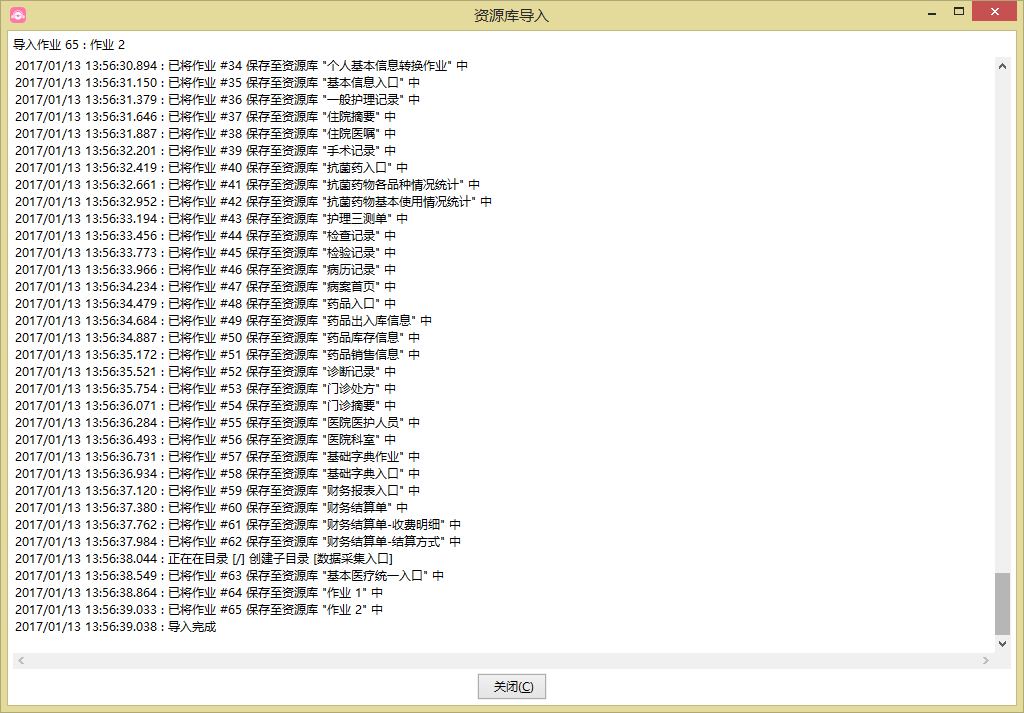
选择备份的资源库，出现弹窗【应用导入的规则？】，如下图所示，选择【否】：



选择目录窗口为第一级，如下图所示，点击确认：

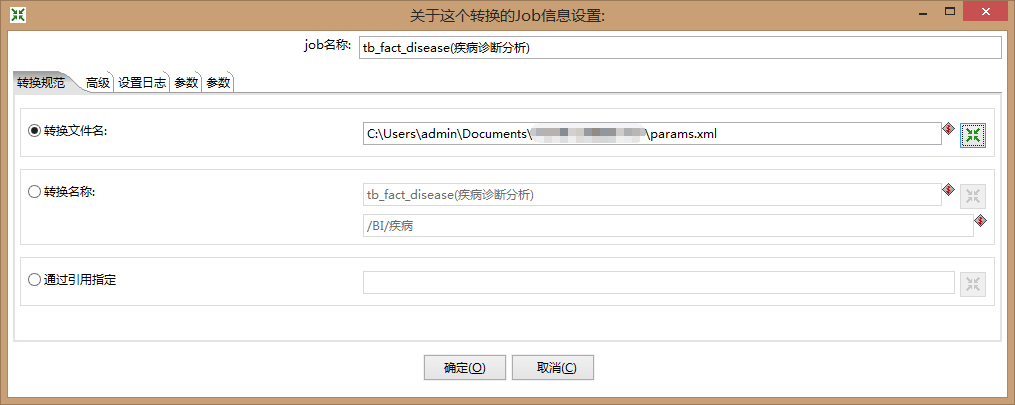


等待资源库导入。当显示导入完成，如下图所示，即完成导入工作。

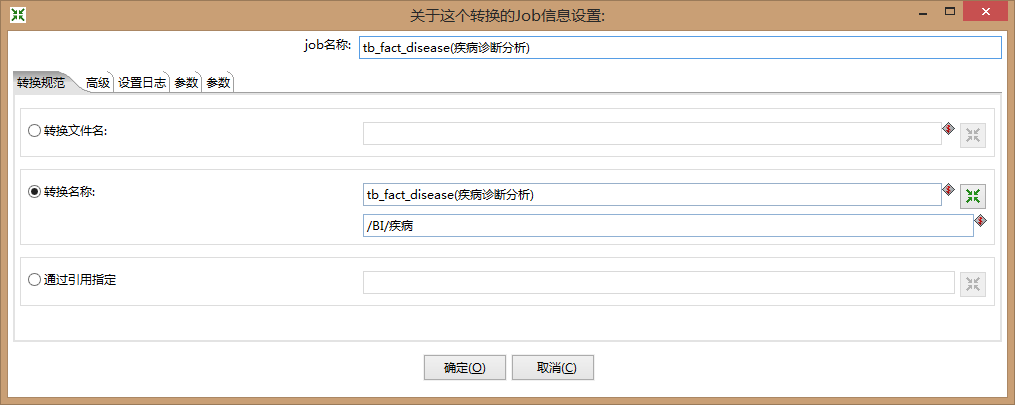


导入资源库不会自动改变转换的路径到资源库目录，需要手动设置，如下：

双击作业项打开转换设置界面，如下图所示：

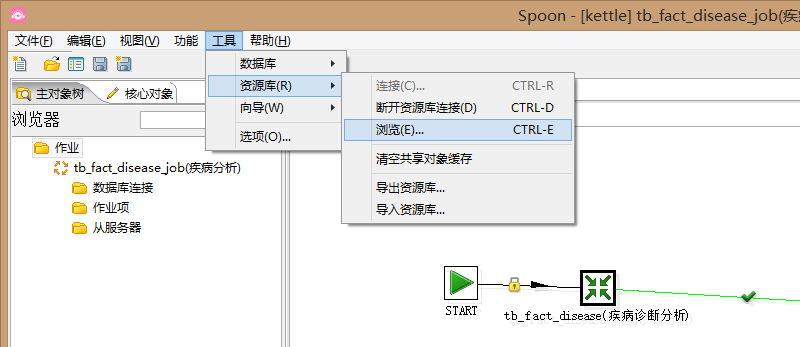


上图是引用本地文件的路径，需改为，如下图所示：



### 删除、修改资源库转换、作业名称

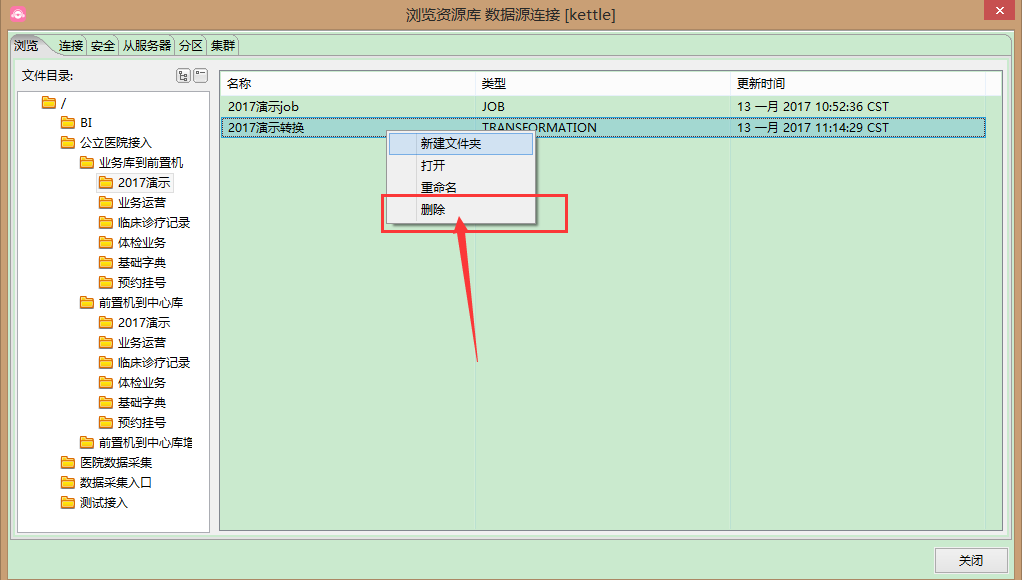
将本地（转换、作业或其他）文件从资源库里删除。选择“工具”-->“资源库”-->“浏览”，如下图所示：



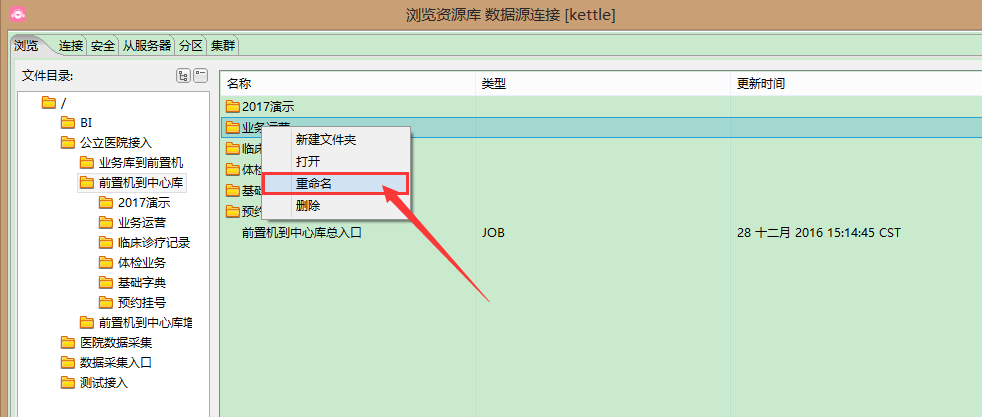
则进入浏览资源库 数据库连接界面，如下图所示：



选择作业或转换，右键，选择“删除”，如下图所示，即可删除对应作业或转换：



选择作业或转换，右键，选择“重命名”，如下图所示，即可修改名称：



### 运行日志

执行建日志脚本。

转换日志有2张表，JOB日志有1张表，具体功能如下表所示。

|  |  |  |
| --- | --- | --- |
| **日志类型** | **日志表名称** | **日志内容** |
| 转换 | TRANS\_STEP\_LOG | 转换的各个STEP步骤执行结果 |
| 转换 | TRANS\_LOG | TRANS转换的执行结果 |
| JOB | JOB\_LOG | JOB的执行结果 |

### 常见工具组件

ETL是基于kettle核心控制台，利用资源库中的作业与转换，利用kettle核心api研发任务调度。常用组件工具如下：

① 表输入：利用连接和SQL语句，读取某数据源信息。能根据数据源表结构自动生成基本SQL语句。

② 插入/更新：利用数据源关键字进行匹配搜索操作，若未匹配则直接插入。若匹配到对应数据，假如数据一致则不作操作；假如出现不同，则更新需要设置的字段。注意，步骤名称在单个转换中必须唯一。如果仅提供插入操作，需要勾选“不执行任何更新”。在“commit size”后文本框填写提交前要改变的数据量。

③ 数据库查询：在数据源中查找值。可以选择是否使用缓存。

④ 调用数据库存储过程：运行数据库的存储过程，并获取返回值。支持自动提交，支持带参数。

⑤ 过滤行：通过比较和条件过滤到记录。当之前有步骤连接到它时，简单的单击“<field>”，“=”和“<value>”区域来构建条件。

⑥ 空操作：什么都不做。测试时，充当一个占位符。

⑦ 设置变量：允许在任务或者虚拟机中设置变量。

⑧ 获取变量：获取一个变量，返回行或者附加值到输入行。

## 数据采集

数据采集，是通过资源库中的各个采集任务，即通过kettle资源库中的各个转换，实现从不同的业务数据源抽取数据到本地缓存库，然后进行清洗、整合，从而实现数据的集成和展示。

执行数据采集任务顺序依次为：前置采集任务，缓存采集任务，增量采集任务，BI报表采集任务。

### 前置采集任务

前置采集任务，也就是将业务数据从医院的数据源库采集到我们系统前置数据库。这一个采集任务的工作是考虑到两方面，一方面是集成系统在操作时不能影响医院业务数据库的正常使用，另一方面是集成系统需要对数据进行部分加工，且要求能够共享当前业务数据。

如果不进行前置任务采集，会影响多方的操作。所以对于采集工作来说，这是非常重要的。当然，建议前置任务设置为定时调度，以保证前置数据库的数据能尽可能保持每日更新。

以新增前置采集任务为例：







### 缓存采集任务

缓存采集任务，将数据从前置数据库抽取到缓存数据库。通过这一步骤，将当前的增量数据抽取到缓存库,既是起到抽取数据的作用，也是为他的功能如BI数据报表做数据基础铺垫。所以说，如果需要及时的抽取数据，只要保证前置业务库的数据是最新的，那么缓存任务调度必不可少。

以修改缓存采集任务为例：







### 增量采集任务

增量采集任务，是将数据从缓存库清洗到中心网关。这一步的工作是为了将增量数据抽取到中心网关数据库，其他功能需要更新数据时，仅需加载最新的数据，不需要将全部业务数据做更新处理，这也在效率上做了提高。



### BI报表采集任务

将数据从缓存库整合到BI数据库，这一步的工作是为了将BI的报表数据进行统一的整合。BI报表是以缓存数据库为基础数据，统计整合数据后存储到对应数据库后，再将数据从数据库中取出进行展示到对应网站。如果需要做BI报表的展示，BI任务调度是必不可少的一项准备工作。



## 采集任务日志

当执行任务调度开始，便开始记录日志对应的日志。日志存储在kettle工具所使用的同一个数据库中，这既是统一对日志的存储，同时也便于对日志的查询。

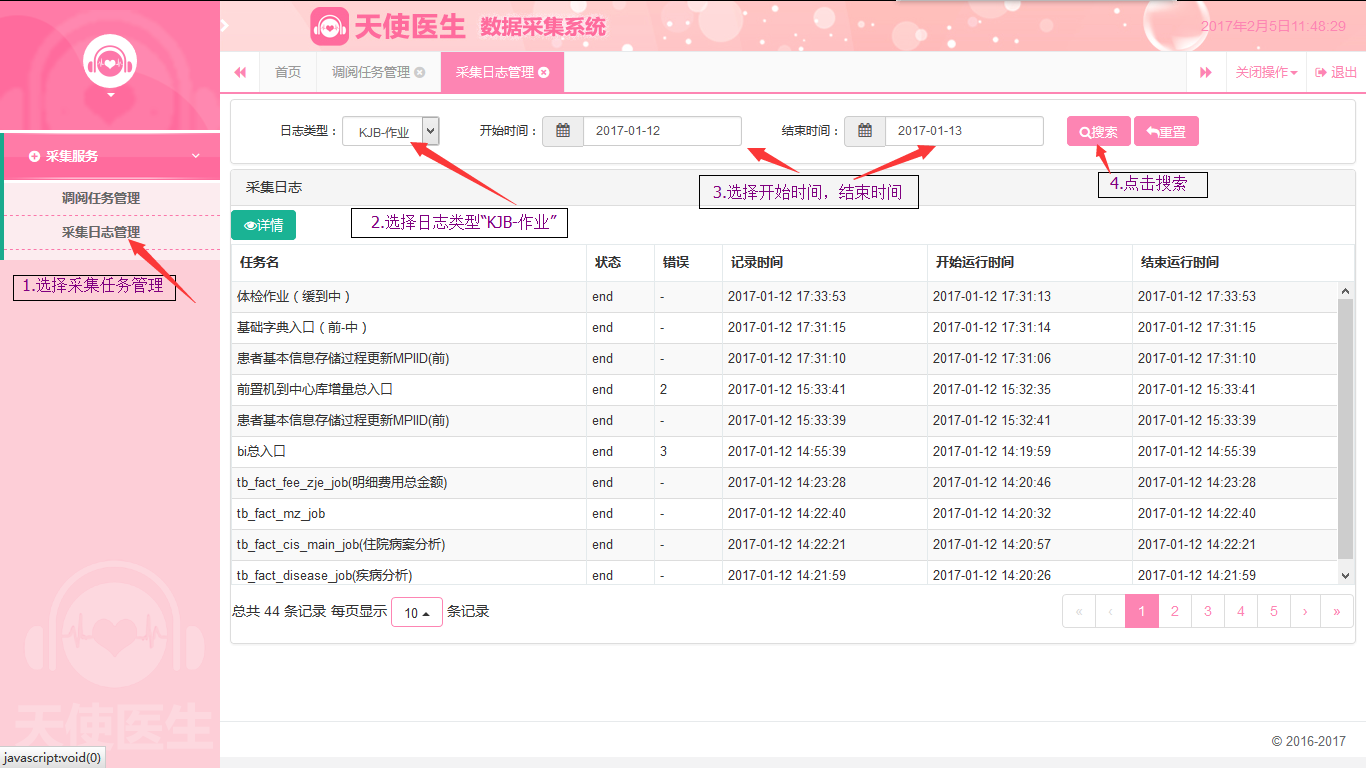
在执行任务调度执行完毕或者因其他原因停止采集任务，此时该任务的对应日志会做更新动作，将对应的日志结束运行时间、状态进行更改，同时在日志详情中更新各个数值。

采集任务日志分成两种类型，一种是JOB作业日志，另一种是TRANS转换日志。

### 采集作业日志

JOB作业日志，存储在表job\_log中。作业日志类型详细记录了任务名、状态、开始运行时间、结束运行时间、记录时间、错误数以及详细说明。

采集作业日志查询如下图所示：

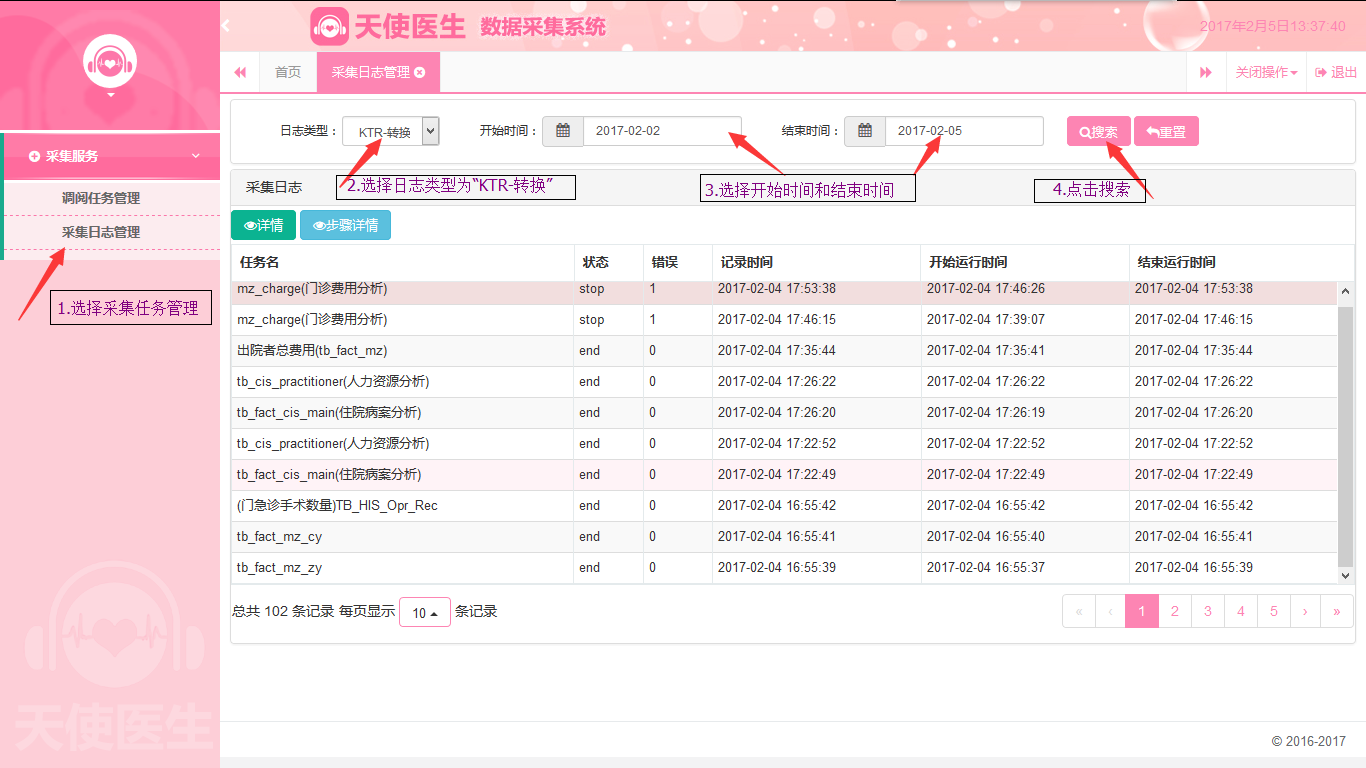




### 采集转换日志

TRANS采集转换日志，存储在表trans\_log中。这种日志类型详细记录了任务名、状态、开始运行时间、结束运行时间、记录时间、错误数以及详细说明。同时，为了更详细的查看转换日志的详情，我们将转换的step步骤详情也记录下来，存储在表trans\_step\_log中。在步骤详情表里，我们可以更清晰的看到，当前任务所包含的每一个步骤，在各自执行之后读取、写入、更新、输出、跳过、错误数量，根据这些数量我们能够得到一些重要的信息。

采集转换日志查询如下图所示：







# EMPI部分

EMPI为患者主索引管理，该部分提供了合并、拆分患者的功能，同时提供了患者主索引配置、域信息查询、合并拆分日志查询等功能，下面分述之。

## 合并患者

|  |
| --- |
|  |
|  |

## 拆分患者

|  |
| --- |
|  |
|  |

## 患者主索引配置

|  |
| --- |
|  |
|  |

## 域信息查询

|  |
| --- |
|  |
|  |

## PIX查询

|  |
| --- |
|  |

## 患者信息查询

|  |
| --- |
|  |
|  |

## 日志查询

|  |
| --- |
|  |

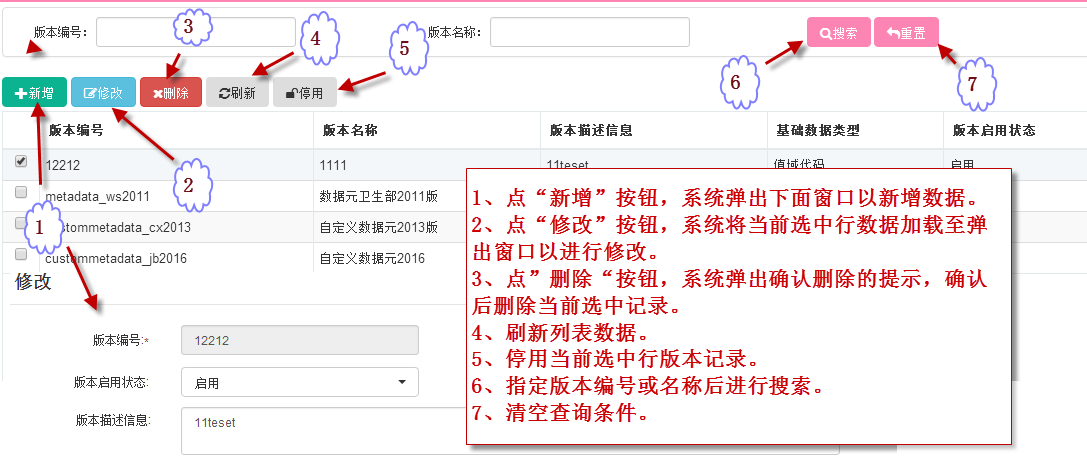
# 数据标准部分

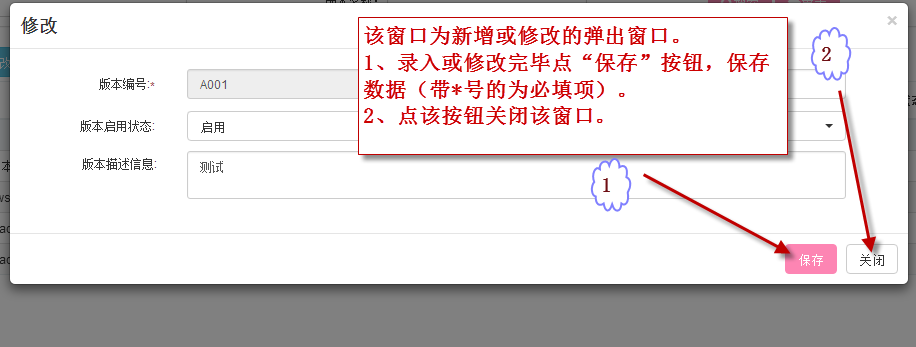
## 标准化管理

本部分对一些基础数据、值域代码、数据集分类、数据集等内容进行管理。

### 基础数据管理

1、选择标准化管理-基础数据管理菜单。





### 值域代码管理

1、选择标准化管理-值域代码管理菜单。

|  |
| --- |
|  |
|  |
|  |
|  |

### 数据元管理

1、选择标准化管理-数据元管理菜单。

|  |
| --- |
|  |
|  |
|  |
|  |

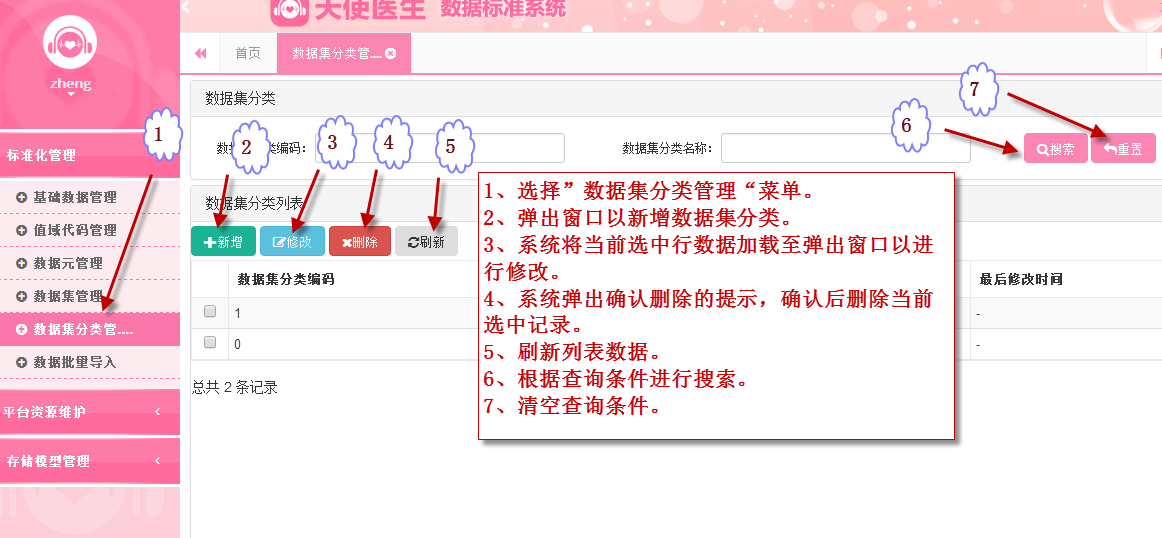
### 数据集管理

1、选择标准化管理-数据集管理菜单。

|  |
| --- |
|  |
|  |
|  |
|  |

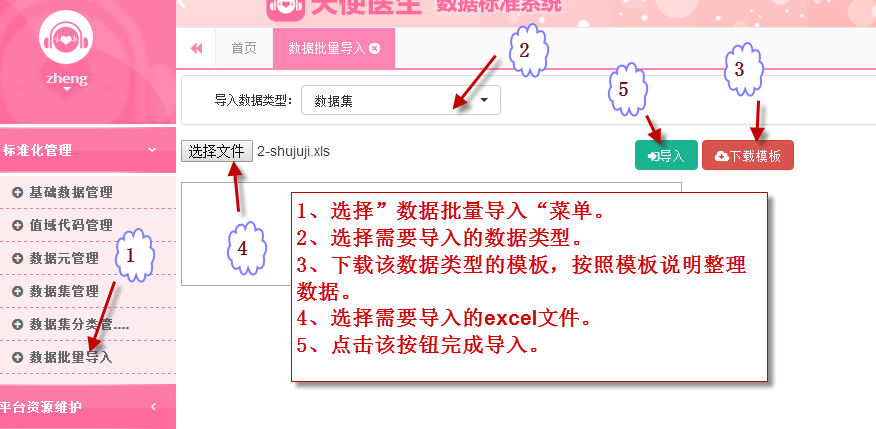
### 数据集分类管理

1、选择标准化管理-数据集分类管理菜单。





### 数据批量导入



## 平台资源维护

该部分对数据标准中所用的项目字典、药品字典、诊断字典等基础数据进行维护。

### 项目字典维护

|  |
| --- |
|  |
|  |

### 药品字典维护

|  |
| --- |
|  |
|  |

### 诊断字典维护

|  |
| --- |
|  |
| C:\Users\chenccp\AppData\Local\Temp\SNAGHTML8640d776.PNG |

### 手术字典维护

|  |
| --- |
|  |
|  |

### 组织机构维护

|  |
| --- |
|  |
|  |

### 人员信息维护

|  |
| --- |
|  |
|  |

## 存储模型管理

本部分对系统所用数据源、主题库、主题库下的存储模型及其字段进行管理。

### 数据源管理

|  |
| --- |
| C:\Users\chenccp\AppData\Local\Temp\SNAGHTML8656f19a.PNG |
|  |

### 主题库管理

|  |
| --- |
| C:\Users\chenccp\AppData\Local\Temp\SNAGHTML865793d5.PNG |
|  |

### 存储模型管理

|  |
| --- |
|  |
| C:\Users\chenccp\AppData\Local\Temp\SNAGHTML86725d22.PNG |
|  |

# BI部分