

ETL自动化工程

- 功能入口
- 操作流程
- 操作详情
 - 配置模板
 - 上传模板
 - 保存
 - 校验Excel
 - 执行自动化生成
 - 参数配置

为了简化数据库表的迁移操作，使项目实施流程标准化，降低项目实施门槛，提高项目实施效率。产品中提供了ETL自动化工具，它能够自动生成自助ETL和物理表，用户可以执行生成的ETL实现物理表的灌数操作。同时，它还允许用户手动修改对应的Excel规则模板，实现更细致的配置。

功能入口

入口一：在 **数据准备 > 自助ETL** 中，选择 **ETL自动化**。

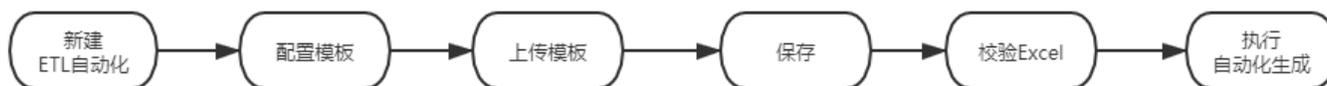


入口二：在“数据准备”界面左侧资源树的“ETL自动化”节点上，右键选择 **新建 > ETL自动化**。



操作流程

ETL自动化工程的操作流程如下：



各流程节点的说明如下：

- 新建ETL自动化：指创建新建ETL自动化，入口参见：[功能入口](#)。
- 配置模板：用于配置及下载Excel模板。详情请参见：[配置模板](#)。
- 上传模板：用于上传配置及下载好的Excel模板。详情请参见：[上传模板](#)。
- 保存：用于自动生成保存自助ETL的目录。详情请参见：[保存](#)。
- 校验Excel：按照内置规则校验Excel模板内容。详情请参见：[校验Excel](#)。
- 执行自动化生成：自动生成自助ETL，并且执行该ETL生成相应的物理表。详情请参见：[执行自动化生成](#)。

操作详情

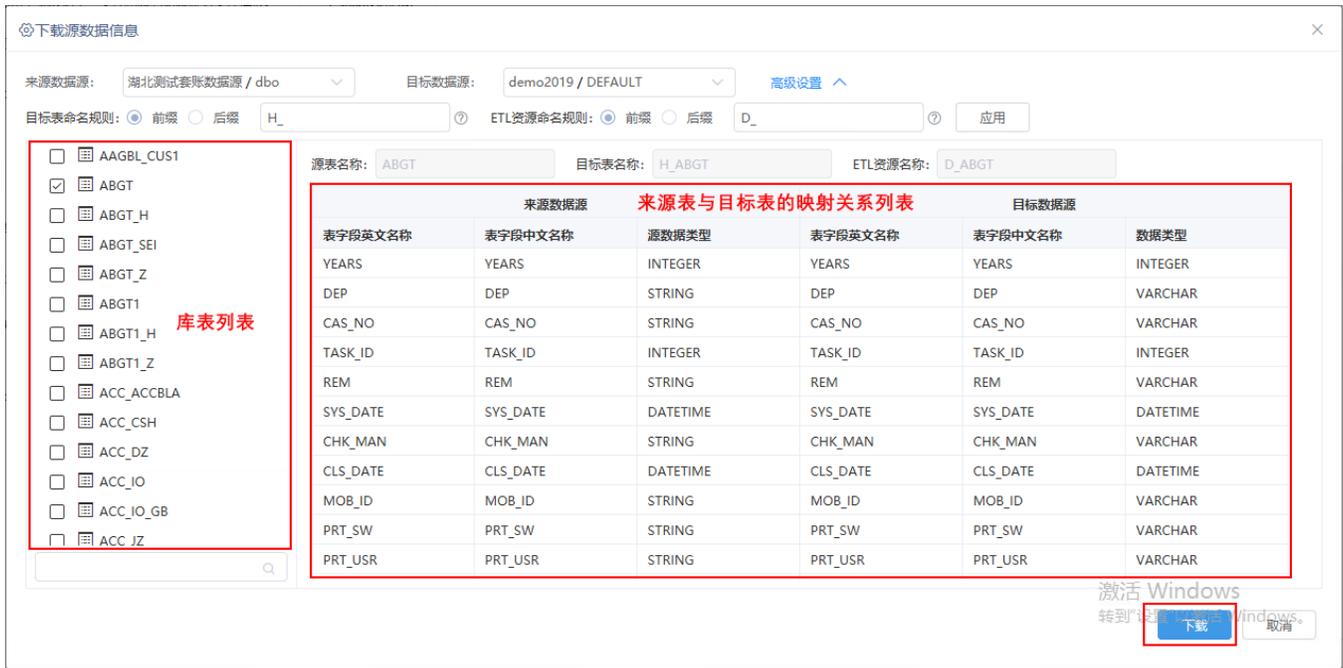
下面以示例的方式对操作详情进行说明。

配置模板

1、在“新建ETL自动化”界面单击 **下载源数据信息** 按钮：



2、弹出“下载源数据信息”窗口，在该窗口中进行模板配置：



该窗口中各配置项说明如下：

配置项	说明
来源数据源	选择需要迁移的数据库表，该项一确定后，“库表列表”则列出所选库下的所有表。
目标数据源	选择数据库表迁移的目标库。
目标表命名规则 ETL资源命名规则	用于分别设置目标表或ETL资源的命名规则，支持前缀或后缀。 要求：只能输入1~20个字符，可使用数字，字母和下划线或者它们的组合，需以字母开头，后缀不能以下划线结尾。 需要单击 高级设置 ，才显示该设置项。 这两项设置完后，单击 应用 按钮，命名效果即显示在“源表名称”、“目标表名称”、“ETL资源名称”。
源表名称 目标表名称 ETL资源名称	命名规则应用后，显示的命名效果。
库表列表	显示用户所选的来源数据源的库表列表，支持通过表名关键字搜索。
来源表与目标表的映射关系列表	显示来源数据源中库表字段与目标数据源库表字段的映射关系。

3、以上配置完成后，单击右下角 **下载** 按钮，将配置好的内容以Excel模板的介质存储到本地。示例文件如下：

[ExcelTemplate.xlsx](#)

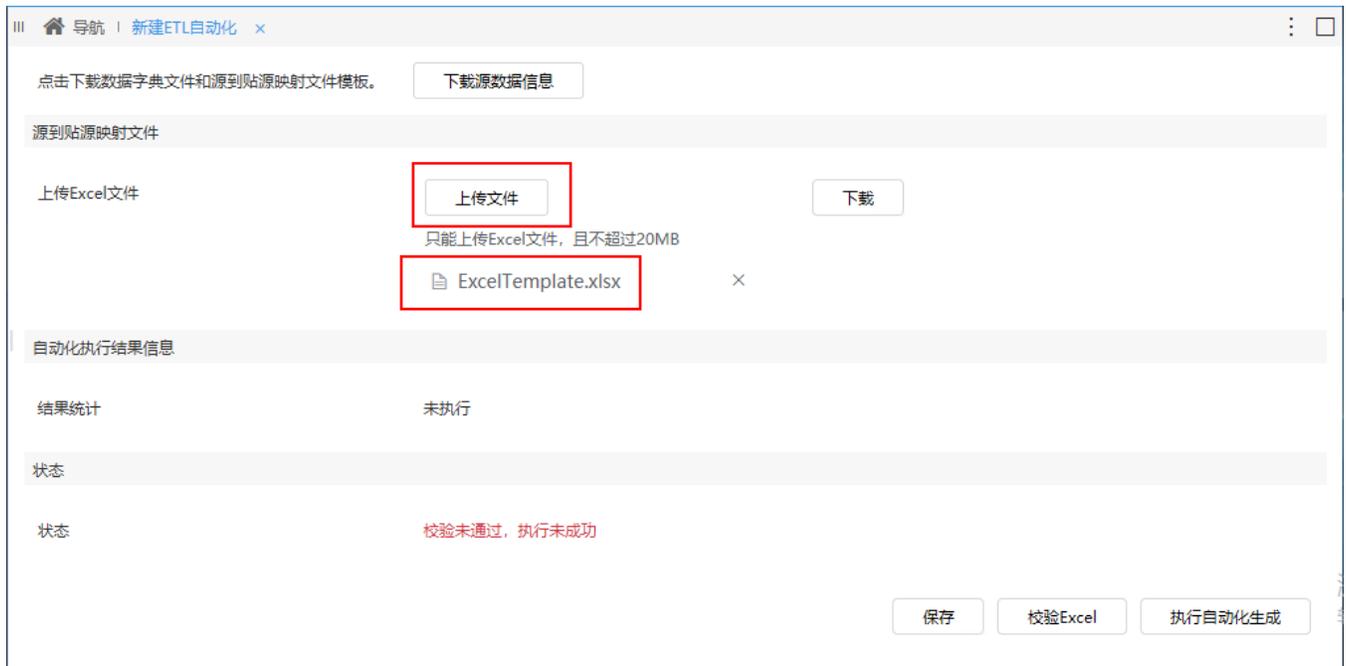
Excel中各主要标签页说明如下：

标签页	说明
系统清单	列举所有数据源（来源、目标）的名称、数据库信息等
变量与参数	为了简化表述而使用的变量或参数。该表可以为空
目标系统数据字典	以 数据源 → schema → 表名 → 字段的结构排列 列举目标数据源的对应目标表中所有写入的字段名称、字段类型、主键信息等
表清单	列举本次ETL自动化中创建的所有ETL资源，及其中分别用到的数据源和目标源信息
字段映射清单	列举来源表和表名之间的所有字段映射关系

详细说明见模板中的 **0-使用说明** 和 **模板检查规则** 标签页

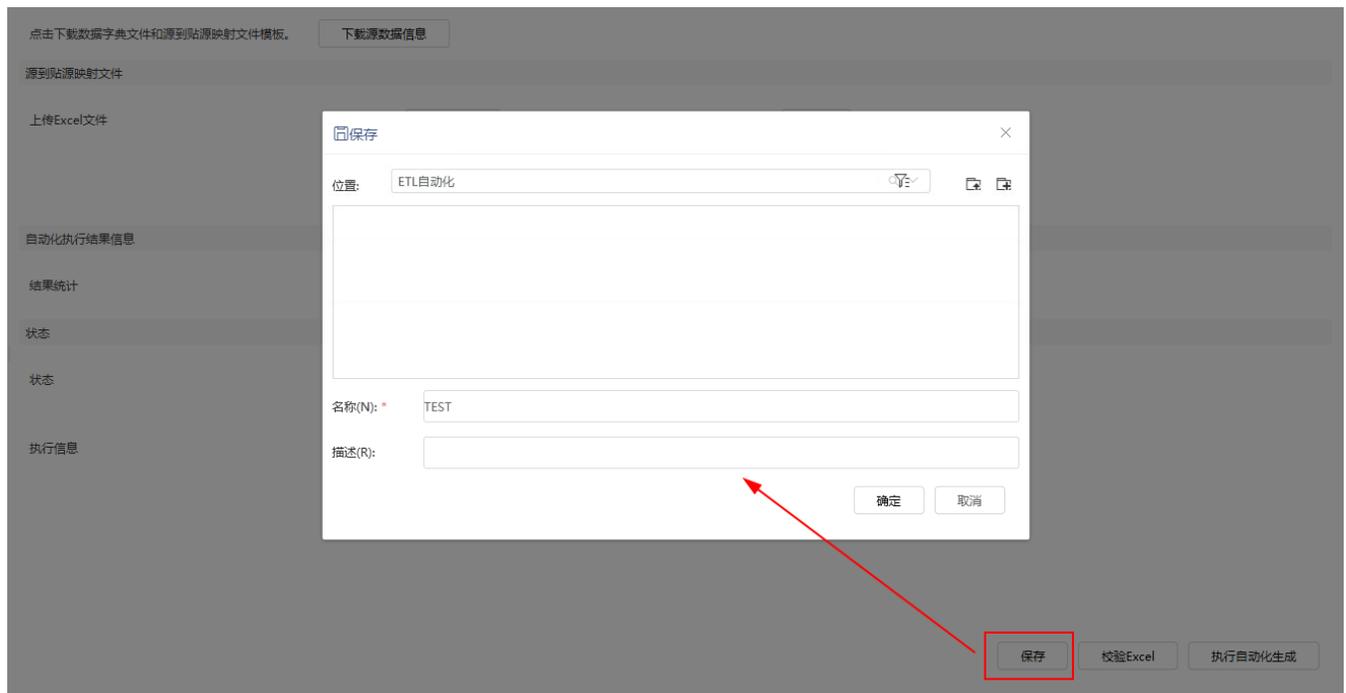
上传模板

将以上导出的Excel模板上传：

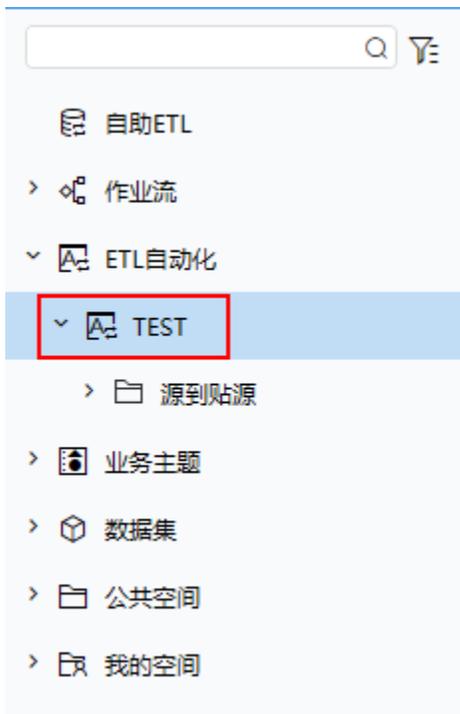


保存

单击右下角的 **保存** 按钮，弹出“保存”窗口，设置名称为“TEST”。



“保存”窗口单击 **确定**，提示保存成功后，系统在“ETL自动化”节点下自动生成“TEST”目录：



校验Excel

单击右下角 **校验Excel** 按钮，只有校验通过才能执行下一步操作。



校验检查规则可以在模板“ExcelTemplate”中查看：

	A	B	C	D	E	G
1	1-系统清单					
2		1 星号必填检查				
3		序号、系统名称、数据连接名称、数据库 必填		报错		
4						
5		2 信息准确性验证				
6		数据连接名称 必须与产品中“数据连接”配置一致		报错		
7		下拉选择项填写准确性检查		报错		
8						
9	2-变量与参数					
10		1 星号必填检查				
11		序号、类型、名称		报错		
12						
13		2 名称				
14		支持字母、数字、下划线，且开头只能为字母		报错		
15						
16		3 变量表达式				
17		当“变量表达式”列非空时，检查对应类型是否为变量；不是则报错		报错		
18		检查“变量表达式”中引用的参数名称或变量名称是否准确，不准确则报错		报错		
19		被引用的参数或变量，必须在比本行序号小的行中定义过，否则报错		报错		
20		表达式的引用格式需与“使用方法”列描述一致，否则警告，将当成常量处理		警告		
21						
22	3-数据字典					
23		1 星号必填检查				
24		序号、系统名称、库名或Schema、表英文名称、字段英文名称、字段数据类型		报错		
25						

执行自动化生成

检验通过后，单击 **执行自动化生成** 按钮，执行最后一步操作。

点击下载数据字典文件和源到贴源映射文件模板。
下载源数据信息

源到贴源映射文件

上传Excel文件
上传文件
下载

只能上传Excel文件，且不超过20MB

ExcelTemplate.xlsx

×

自动化执行结果信息

结果统计
创建0个表，1个ETL实验

状态

状态
校验通过，执行成功

执行信息

校验通过！计划创建1个表
执行结果信息

表'h_abgt'已存在，跳过该表创建过程

2021-09-10 11:04:45: 创建0个表

开始创建，计划创建1个ETL实验

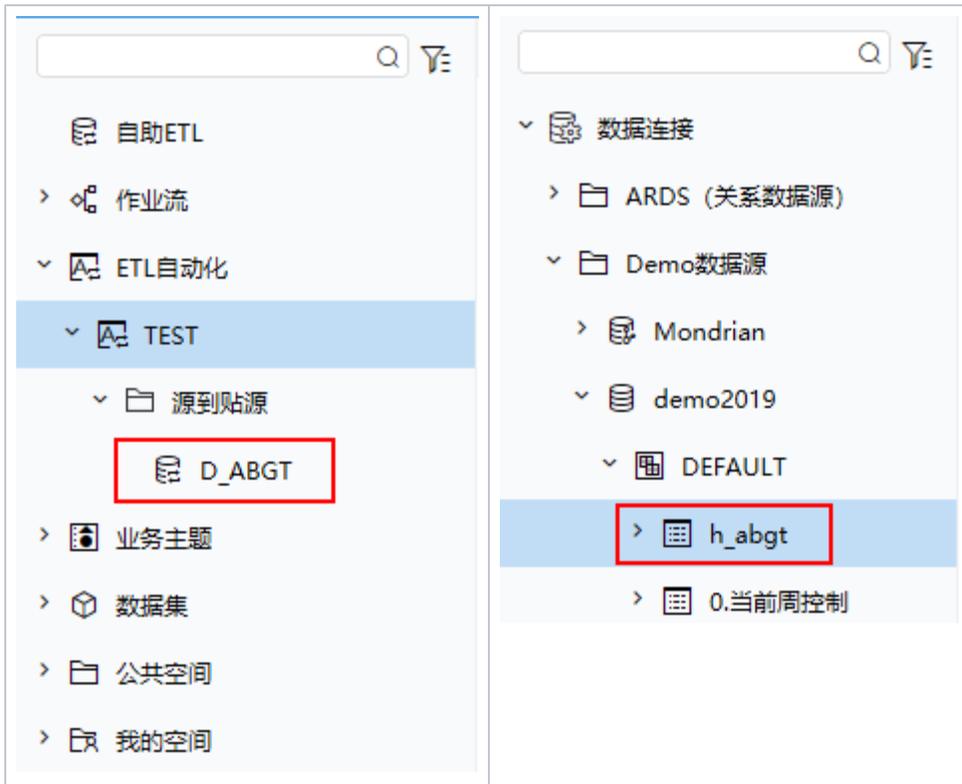
ETL实验'D_ABGT'创建成功！

2021-09-10 11:04:45: 创建1个ETL实验

保存
校验Excel
执行自动化生成

该操作完成后，系统自动生成如下两类资源，实现了库表迁移：

按照规则命名的“ETL资源”	按照规则命名的目标库表
----------------	-------------



提示

- 模板填写中，需留意字段的类型以及长度，避免ETL自动化生成的ETL资源因字段长度为空，导致运行失败，例如：字符串类型
- 执行成功后，系统会自动创建ETL资源和表，此时表数据为空，需运行其相关的ETL资源。

参数配置

某些场景还需要配置参数，目前参数配置有如下两个方法：

第一种：在模板上手工配置：



第二种：执行自动化生成后，在ETL资源编辑界面进行配置：

