# 数据结构

本模板的运行依赖于ETL配置管理数据库中下列表中的数据：

|  |  |  |  |
| --- | --- | --- | --- |
| 表名 | 中文名称 | 是否导入 | 说明 |
| src\_table\_analysis | 源系统表级分析结果表 | 是 | 存放源系统各表的表级属性。其中必填项包括：  schema\_name  table\_name  sys\_name  cn\_name  is\_inc\_ext  stbl\_name  各字段的说明参见模型文件。 |
| src\_column\_analysis | 源系统字段级分析结果表 | 是 | 存放源系统各表的字段级属性。其中必填项包括：  column\_name  sys\_name  schema\_name  table\_name |
| etl\_tasks | ETL任务表 | 否 | 存放ETL映射任务属性，一般每一张需要加载的仓库表对应一个ETL映射任务。其中必填项包括：  task\_name  schema\_name  table\_name  sys\_name  etl\_dvlpr\_name  serial\_no  利用admin用户执行刷新和保存ETL任务列表为这个表填入数据。 |
| dw\_tables | 数据仓库表 | 是 | 定义数据仓库中的实体。必填项包括：  schema\_name  table\_name  sys\_name  phy\_name  load\_mode  subject\_name  各字段的说明参见模型文件。 |
| dw\_table\_mapping | 数据仓库表级映射 | 否 | 定义源系统表和数据仓库表之间的映射关系。必填项包括：  load\_batch  schema\_name  src\_schema  table\_name  src\_table\_name  sys\_name（固定为’DW’）  src\_sys\_name  join\_order  table\_alias  join\_type  各字段的说明参见模型文件。 |
| dw\_column\_mapping | 数据仓库字段级映射 | 否 | 定义数据仓库基础层实体的字段级映射。这个表的数据来自用户在映射模板中填写的映射内容。 |
| dw\_subjects | 数据仓库主题表 | 是 | 定义数据仓库的主题。必填项包括  subject\_name |
| dw\_columns | 数据仓库实体属性表 | 是 | 定义数据仓库实体的属性。必填项包括：  column\_name  schema\_name  table\_name  sys\_name  column\_id  data\_type  phy\_name  is\_pk  chain\_compare  is\_partition\_key |
|  |  |  |  |

# ETL作业定义工作表

## 系统参数页

在“系统参数”页进行参数设置。其中需要设置的参数包括：

* 开发人员名称：参与ETL映射开发的人员名称，需要与ETL配置管理数据库中的 etl\_tasks.etl\_dvlpr\_name字段相对应。admin表示管理员用户，管理员用户可以为未分配开发人员的ETL任务指定开发人员。
* 数据源名称：在运行本ETL模板的机器上配置的ODBC数据源的名称，该数据源指向ETL配置管理数据库。

## ETL作业列表页

本页显示数据仓库基础层ETL加载作业列表，仓库基础层每一个实体对应一个加载作业。用户点击“刷新”按钮显示当前的数据仓库基础层实体所对应的ETL加载作业，标题为蓝色的列允许用户修改。用户修改完毕需要点击“保存”按钮将数据保存到ETL配置数据库中。用户可以点击“创建ETL作业表”按钮为列表中每一个ETL作业创建一个Excel Worksheet，用于编辑该作业的映射。新创建的Worksheet按照ETL作业列表中的序号字段进行命名，命名规则为“ETL\_作业序号”。如果同名的Worksheet已经存在将不会重复创建新的Worksheet。

## ETL作业模板页

本页是用于创建ETL作业Worksheet的模板，用户不应该修改本页的任何内容。

## ETL作业页

命名为“ETL\_作业序号”的页面用于编辑具体ETL加载作业的映射参数。本页可以分为以下几个部分：

## 目标表ETL属性

包含本作业需要加载的目标仓库实体的和ETL有关的属性，用户可以在这里指定ETL作业属性，包括加载模式、积数计算、删除模式、保留首次加载日期、备注。其中积数计算、删除模式、保留首次加载日期尚未实现。

### 加载模式

支持的模式包括：

* 追加——适用于全量抽取。当前加载的数据将作为一个新的分区追加到目标表中，分区字段为data\_dt，表示数据加载的工作日。
* 更新——适用于增量抽取的情况。当前加载的数据将与目标表中原有数据进行主键FullOuterJoin产生当天的全量数据，关联结果中如果新加载记录不为空，就保留新加载记录，否则使用目标表原有记录。新的全量数据被加载到以当前工作日标记的新分区中（分区字段为data\_dt）。