# SQL\_HEPPER 使用说明书

### 1. 总体结构

### 需求管理

(一个基于需求开发过程的代码、文档管理工具)

实现需求过程中产生的口径、表元数据、任务元数 据等信息,被元数据管理 整合

更快、更规范的开发出需 求需要的代码

### 元数据管理

(一个全局性不断成长的的关 系管理工具)

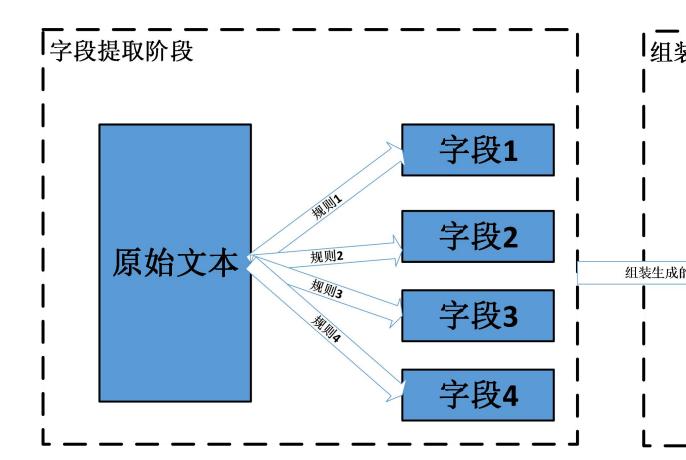
SQL\_Hepper核心主件 (一个基于规则的批量化生成工 具)

复用已有的口径、元数 据、规则,进一步提高开 发效率

## 2. 模块介绍

# 2.1 SQL\_HEPPER 核心主件

SQL\_HEPPER 核心主件是一个批量化的文本生成工具,可以根据用户提供的原始文本,基于自定义规则,一键生成目标文本;可以分为字段提取阶段和组装阶段两个环节,具体工作流程图如下:



■ SQL助手				
文件 元数据管理 关于				
需求管理 元数据管理	口径管理 表管理	字段管理 需求血缘		
当前工程: 思特奇	→ 需求:	~]		
,T2.REGION_CODE ,T2.CITY_CODE				
字段提取工作	X			
2				
- -				
2				
1				
2				
参数: RCSQ RCSQ	- 添加	删除	复制 分割换行	去除指定行  回退
- i				
history:			> 搜索模版:	
,history: 字段组	装工作区			
57				

# 2.1.1 字段提取阶段

字段提取阶段:根据用户提供的原始文本与编辑的文本处理规则,分别从原始文本的每行提取出目标字段,这里将每个变量提取的逻辑称为一个变量。

# 2.1.1.1 变量

#### 变量预处理:

- 分隔换行:根据指定分隔符,把原始文本分为多行
- 去除指定行:保留原始文本中包含指定字符串的行

#### 变量加工:

一个变量由 1 个或多个算子组成,算子可以叠加使用,下游算子的输入是上游算子的输出,根据序号排序算子的上下游关系,序号小的先执行。目前支持的算子如下:

算子名称	算子功能介绍	
常规	不进行任何操作	啥也不用写
正则表达式	以正则表达式方式从文本中提取目标文本	例子: reg=\'(.*?)\';index=1 实现了从文本 reg 后跟正则表达式, index 跟索引号,
字符串切割	通过简单的 python 语法,完成对目标字段的提取	例子 1: \${arg}.split('')[1].strip() 实现\${arg} 是算子中一个的内置变量,可以str 对象支持的方法,\${arg}都支持(有大小写转换等,把特定变量换成\${arg}例子 2: \${arg}[1:] if \${arg}[0]==',' else \$否则保持原样功能,字符串切割也支持容,写特定的语法。
轮询	根据传入参数,扩充目标文本行数	例子:传入一个整数,可以是正数也可 法
去除换行	将原始文本由多行转化为一行,可以指定分隔符	例子:, 将原始文本转化为1行,通流格式)
GROUP BY	根据原始文本行数,生成 Gbase Group by 写法中的 1,2,3,4,5,6	无
行分隔	根据参数选择原始文本中的特定行	例子: n 输出原始文本中第 n+1 行( -n 输出原始文本中倒数第 n 行 n:m 输出原始文本中底 n+1~第
元数据	给定表和字段,查询字段元数据信息	例子: TB:PMID.TB_MID_PAR_USER_DAY K1 TB:PDATA.TB_PDT_SERV_ATTR_INFO_DATB:PCDE.TB_CDE_SERV_ATTR_CODE T1 K1.USER_ID K1.CUST_ID K1.ACCT_ID K1.PHONE_NO K1.MAIN_SERV_ID K2.SVC_ID 从特定表中查询特定字段的类型和注释 K1 USER_ID DECIMAL(14,0) 用户 ID K1 CUST_ID DECIMAL(14,0) 密户 ID K1 ACCT_ID DECIMAL(14,0) 帐户 ID K1 PHONE_NO VARCHAR(15) 手机号码K1 MAIN_SERV_ID VARCHAR(20) 主体从K2 SVC_ID VARCHAR(40) 服务标识
自定义函数	传入一个 python 自定义函数,按行处理原始文本	例子: def fun1(a): if a.endswith('_DATE'):

```
return 'DATE'
                                                     if a.endswith('_PAY'):
                                                         return 'DECIMAL(18,4)'
                                                     if a.endswith(' FEE'):
                                                         return 'DECIMAL(18,4)'
                                                     if a.endswith('_CODE'):
                                                         return 'VARCHAR(10)'
                                                     if a.endswith('_RATE'):
                                                         return 'DECIMAL(5,4)'
                                                     if a.endswith('_NAME'):
                                                         return 'VARCHAR(500)'
                                                     return 'VARCHAR(50)'
                                                 实现:根据字段名称,一键生成字段多
                                                     函数名称随意,传入参数只能有一
                                                 出参数只能是文本类型
超级自定义 | 传入一个 python 自定义函数,直接处理整个原始
                                                 例子:
函数
           文本
                                                 def fun2(a):
                                                     tmp = [f"'{i}" for i in a]
                                                     res = ','.join(tmp)
                                                     return [res,]
                                                 实现: 加引号+换行, 场景 A,B,C -> 'A'
                                                    函数名称随意,传入参数只能有一
                                                 出参数只能是数组
```

### 2.1.1.2 变量分发

算子分发是指将一个变量处理的结果,分发给其他变量去处理,避免同一个文本被贴来 贴去。变量分发写法如下,传入参数是要分发给的变量名称。

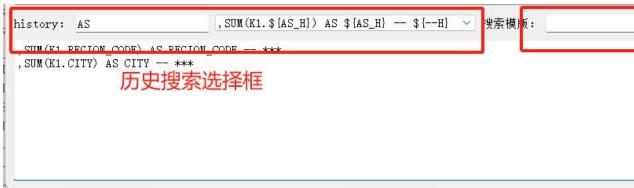


### 2.1.2 字段组装阶段

字段组装阶段是指将字段提取阶段生成的变量,按照特定的格式组装起来,写法如下:  $_{\text{SUM}(K1.\${AS\_H})}$  AS  $_{\text{AS\_H}}$  --  $_{\text{$\{--H\}}}$ 

其中\${} 包含的都是变量,其他的所见即所得。字段组装支持复用历史写过的拼装语句,

通过历史搜索选择框实现;也支持模版功能(模版主要是实现多段拼装语句),通过模版搜索选择框实现。



模版文件已 txt 格式存储在特定位置, 想要新加模版, 在特定位置创建 txt 文件就行, 点一下'帮助文档'按钮, 会自动把模版文件加载进来, 模版文件写法如下(例子):

```
CREATE TABLE STQ_REPORT.${JBYJ_5}(

,${JBYJ_O_1} ${JBYJ_O_2}
)

;

COMMENT ON TABLE STQ_REPORT.${JBYJ_5} IS '${JBYJ_9}';

COMMENT ON COLUMN STQ_REPORT.${JBYJ_5}.${JBYJ_O_1} IS '${JBYJ_8}';

DELETE FROM STQ_REPORT.${JBYJ_5} WHERE DEAL_DATE=DATE '${$}${<}TX_DATE_FORMAT${<}';

DELETE FROM STQ_REPORT.${JBYJ_5} WHERE DEAL_DATE= '${$}${<}TX_MONTH${<}';

ALTER TABLE STQ_REPORT.${JBYJ_5} ENABLE ROW MOVEMENT;

ALTER TABLE STQ_REPORT.${JBYJ_5} SHRINK SPACE;
```

### 2.2 需求管理

需求管理工具辅助管理需求开发过程中的全部代码和文档,其中代码支持版本管理,并 提供高效的检索功能,主要界面如下:



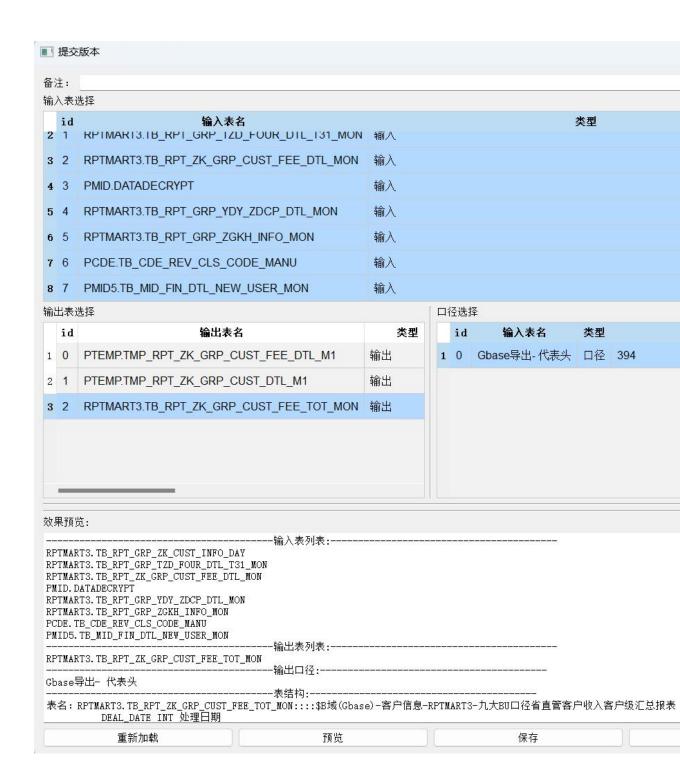
### 2.2.1 快速检索功能

需求、文档、代码支持"与"、"或"检索,例子如下(以检索需求为例,文档代码同): 工信部|BU|欠费&物联网

搜索需求名称包含"工信部",或者包含"BU",或者包含"物联网"且包含"欠费"的需求。

### 2.2.2 版本管理/版本提交

版本管理仅支持对 SQL 代码的版本管理,选择 SQL 代码文件后点击"提交版本"按钮, 弹出提交版本界面,点击保存后,可以在通过点击"显示历史版本"查看保存的历史版本(历 史版本为只读模式,不可编辑)。提交版本不仅可以实现对代码版本的管理,也同步完成了 对代码血缘、需求血缘、口径等元数据的录入。



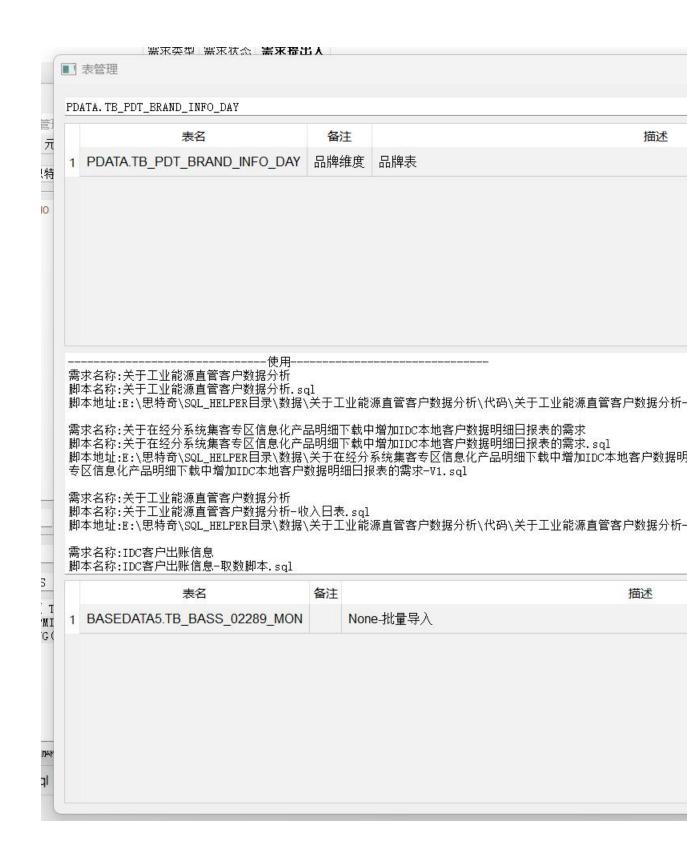
### 2.2.3 超级检索

超级检索实现了在多个需求之间根据关键字对代码的搜索(正则表达式),多用于一下想不起 SQL 在哪个需求里写的了的场景。点击需求管理界面上的"超级检索",输入关键字完成搜索。

# 2.3 元数据管理

# 2.3.1 表管理

表管理中导入了目前 GBASE 中常用库的大部分表结构信息,支持根据表名(英文)、表描述、表备注(结合自己开发实际做的笔记)对表进行检索,可以便捷的查询到表是哪个需求产生的,哪些需求中用到了这个表。



### 2.3.2 字段管理

字段管理展示了表的元数据信息,支持根据自己理解对每个字段添加备注。

	字段名	字段描述	友汁/自	今仍光刊	表名
10		状态变化时间	备注信息	字段类型	
19	STAT_CHG_TIME	1人心文化时间		INT(11)	PMID.TB_M
20	USER_TYPE_CODE	用户类型代码		VARCHAR(2)	PMID.TB_M
21	USER_STAT_CODE	用户状态代码		VARCHAR(2)	PMID.TB_M
22	BELONG_TYPE_C	所属类型代码	3是集团	VARCHAR(4)	PMID.TB_M
23	CRDT_GRADE	信用等级		VARCHAR(3)	PMID.TB_M
24	CUST_GRADE_C	客户等级代码		VARCHAR(3)	PMID.TB_M
25	CUST_TYPE_CODE	客户类型代码		VARCHAR(2)	PMID.TB_M
26	VIRTL_USER_CODE	虚拟用户标识		VARCHAR(1)	PMID.TB_M
27	GROUP_FLAG	群组实例标志		VARCHAR(1)	PMID.TB_M
28	ASSURE_FLAG	担保标志		VARCHAR(1)	PMID.TB_M
29	MAIN SERV ID	主休服各ID		VARCHAR(8)	PMID TR M

### 2.3.2.1 快速模版

字段管理支持基于元数据的快速代码生成,选择待选字段,右键,选择对应模版(模版支持自定义写法),会在下方展示界面生成对应的语句(并将内容直接放在剪切板中,直接Ctrl+V就可以贴在自己代码中),效果如下:

	字段名	字段描述	MA SALE	字段类型	表名	表描述
	DEAL_DATE	处理日期		DATE	PDATA.TB_PDT_B	品牌表
	BRAND_ID	品牌标识		VARCHAR(3)	PDATA TB_PDT_B	品牌表
	SUPR_BRAND_ID	上级品牌标识		VARCHAR(3)	PDATA TB_PDT_B.	品牌表
1	BRAND_TYPE	品牌类型	I-字段名称.txt	(ARCHAR(3)	PDATA TB_PDT_B	品牌表
5	BRAND_NAME	品牌名称	K心块速-字段注释.txt K2-快速-字段注释.txt	(ARCHAR(250)	PDATA.TB_PDT_B	品牌表
6	BRAND_DESCRIBE	品牌描述	K3-快速-字段注释.txt	ARCHAR(250)	PDATA.TB_PDT_B	品牌表
7	IDEA .	理念	K4-快速-字段注释.txt K5-快速-字段注释.txt	(ARCHAR(250)	PDATA.TB_PDT_B	品牌表
8	SLOGAN	口号	探查-DEAL_DATE重复性检查.txt	(ARCHAR(250)	PDATA.TB_PDT_B	品牌表
O,	CREATE_DATE	创建日期	探查-字段重复性检查.txt 探查-重复数据重复记录探查.txt	DATE	PDATA TB_PDT_B	品牌表
3	10 CREATE_TIME	创建时间	查询-※-不去重.txt	NT(11)	PDATA.TB_PDT_B	品牌表
	11 CREATE LOGIN		查询-按字段 - 去重.txt 查询-按字段-不去重.txt	VARCHARI201	POATA TR POT R	見機事
表信息:PDATA. TB_PDT_BRAND_INFO_DAY- 字段信息:BRAND_NAME-品牌名称 备注信息: 表信息:PDATA. TB_PDT_BRAND_INFO_DAY-			模版样例 txt 生成修改表字段语句 txt 统计 字段个数 txt 获取创表语句 txt			

	字段名	字段描述	备注信息	字段类型	表名
19	STAT_CHG_TIME	状态变化时间		INT(11)	PMID.TB_MID_PA
20	USER_TYPE_CODE	用户类型代码		VARCHAR(2)	PMID.TB_MID_PA
21		用户状态代码		VARCHAR(2)	PMID.TB_MID_PA
22		所属类型代码	3是集团	VARCHAR(4)	PMID.TB_MID_PA
23		信用等级		VARCHAR(3)	PMID.TB_MID_PA
24		客户等级代码		VARCHAR(3)	PMID.TB_MID_PA
25		客户类型代码		VARCHAR(2)	PMID.TB_MID_PA
26		虚拟用户标识		VARCHAR(1)	PMID.TB_MID_PA
27	GROUP_FLAG	群组实例标志		VARCHAR(1)	PMID.TB_MID_PA
28	ASSURE_FLAG	担保标志		VARCHAR(1)	PMID.TB_MID_PA
29	MAIN SERV ID	主休服各ID		VARCHAR(8)	PMID TR MID PA
表信 字段 备注	言急:PMID.TB_MID_PAR_US 设信息:BELONG_TYPE_CODE 注信息:3是集团 USER_TYPE_CODE AS USEI	SER_MON-用户基本信息: B-所属类型代码	类型代码	VARCHAR(8)	PMIN TR MIN PA

#### 自定义模版编辑

处规则 处规则

自定义模版是 txt 格式,放在指定位置,写好模版后直接放在特定位置,便可在字段管理中右键使用。

自定义模版例子:

- -- 关键字介绍
- -- ZDMC:选择的字段名称
- -- BM: 字段对应表名
- -- TX\_DATE: 业务日期 YYYYMMDD
- -- TX\_DATE\_FORMAT: 业务日期 YYYY-MM-DD
- -- TX\_MONTH:业务日期 YYYYMM
- -- SELECT\_ZDMC:选择了的字段名称,逗号分隔
- -- SELECT ZDMC RANGE: 根据选择字段个数返回 1,2,3,4,5, 一般用于 GROUP BY
- -- SELECT\_ZDMC\_MS: 选择了的字段名+注释,TN.CN AS CN -- 注释
- K5.\${SELECT\_ZDMC\_MS}

### 2.3.3 口径管理

口径管理用于记录开发过程中的可能会复用的口径,通过简单的标记后,可以在后续开

发中十分快捷的查询到。

### 2.3.3.1 口径管理标记方法

如图,通过"-->>"在代码中进行标记

```
95
96
97
      INSERT INTO PTEMP. TB RPT GRP DZXSZ FLAG MEB D2(
98
         DEAL_DATE
99
         ,USER_ID
         ,DX_CNT
100
101
      )
102
     SELECT
103
         K1.DEAL DATE AS DEAL DATE -- 处理日期
104
        ,K1.USER_ID AS USER_ID -- 用户ID
         ,SUM(K1.SEND SMS TIMES) AS DX CNT -- 发送短信次数
105
106 FROM PMID.TB MID BLL SMS USER MON K1
     WHERE K1.DEAL_DATE BETWEEN PERIOD_ADD(SUBSTRING('${TxDate}',1,6),-3) AND PERIOD_ADD
107
108 GROUP BY 1,2
109 ;
```

### 2.3.3.2 口径管理提交方法

在需求管理提交版本的时候, 提交口径。

	主: 入表设	±±×					
∠ ענהו	id						类型
13	3 12	PMID5.TB_MID_GRP_TAG_XSZ_REL_DAY	输入				
14	£ 13	PCDE.TB_CDE_LOC_REGION	输入				
15	5 14	PDATA.TB_PAR_LOGIN_INFO_DAY	输入				
16	15	PCDE.TB_CDE_RUN_DICT_CODE	输入				
17	7 16	PMID.TB_MID_BLL_SMS_USER_MON	输入				
18	3 17	PMID.TB_MID_PAR_USER_EXTEND_DAY	输入				
		PCDFTB CDF SITE CODF	输入	ĺ	Πź	조选排	×
HUL	id	输出表名	类型			id	* 输入表名
+	5	I TEIWII .TD_IWIID_OIXI _IAO_AOZ_IWIED_DTE_DO	4B) LLI		1	0	用户语音使用情况(月表
5	4	PTEMP.TB_MID_GRP_TAG_XSZ_MEB_DTL_D4	输出		2	1	用户发送短信次数(月表
6	5	PTEMP.TB_MID_GRP_TAG_XSZ_MEB_DTL_D5	输出	_1	3	2	用户流量
7	6	PTEMP.TB_MID_GRP_TAG_XSZ_MEB_DTL_D06	输出		4	3	渠道对应网格
8	7	PTEMP.TB_MID_GRP_TAG_XSZ_MEB_DTL_D6	输出	_	5	4	主副卡对应关系
9	8	PMID5.TB_MID_GRP_TAG_XSZ_MEB_DTL_DAY	输出	-	6	5	用户状态标识解析
						- F	7107 100-10-2004-101
PC PD PC PM PM	ATA. DE.T ID.T ID.T	u: B_CDE_LOC_REGION TB_PAR_LOGIN_INFO_DAY B_CDE_RUN_DICT_CODE B_MID_BLL_SMS_USER_MON B_MID_PAR_USER_EXTEND_DAY 					
用	户发	送短信次数(月表)					
表	名: I	ASA19: PMID5.TB_MID_GRP_TAG_XSZ_MEB_DTL_DAY::::\$B域(Gb DEAL_DATE DATE 业务日期 JZ_USER_ID VARCHAR(25) 用户ID CUST_ID VARCHAR(25) 客户ID JZ_PHONE_NO VARCHAR(15) 手机号码 REGION_CODE VARCHAR(2) 地市代码	ase)-集团用	J戸-PMII	)5-	【存量	<b>量标签】电子学生证成员类</b>

# 2.3.3.3 口径管理查询方法

在口径管理中,输入关键字查询口径。



### 2.3.4 需求血缘

需求血缘用于记录该需求的对应 SQL 代码用了哪些表,产出了哪些表。

#### 11分皿塚

#### 关于上线2024年物联网收入总报表及直管和省内报表的需求(特战队)

	脚本编号	需求名称	需求类型	需求提出人	
1	717	关于上线2024年物联网收入总报表及直管和省内报表的需求(特战队)	任务	特战队	创;
2	726	关于上线2024年物联网收入总报表及直管和省内报表的需求(特战队)	任务	特战队	月:
3	727	关于上线2024年物联网收入总报表及直管和省内报表的需求(特战队)	任务	特战队	收
4	728	关于上线2024年物联网收入总报表及直管和省内报表的需求(特战队)	任务	特战队	收
5	737	关于上线2024年物联网收入总报表及直管和省内报表的需求(特战队)	任务	特战队	贵
6	744	关于上线2024年物联网收入总报表及直管和省内报表的需求(特战队)	任务	特战队	活
7	746	关于上线2024年物联网收入总报表及直管和省内报表的需求(特战队)	任务	特战队	收
8	747	关于上线2024年物联网收入总报表及直管和省内报表的需求(特战队)	任务	特战队	收
9	748	关于上线2024年物联网收入总报表及直管和省内报表的需求(特战队)	任务	特战队	月
10	749	关于上线2024年物联网收入总报表及直管和省内报表的需求(特战队)	任务	特战队	收
11	752	关于上线2024年物联网收入总报表及直管和省内报表的需求(特战队)	仟条	特战队	活品

PMID5.TB\_MID\_FIN\_OWE\_DTL\_USER\_MON PDATA.TB\_BASS\_RES\_91017\_ALL\_DAY

PCDE.TB\_CDE\_BOSS\_ERP\_RLT\_MANU

PMID5. TB\_MID\_CMIOT\_CUST\_MAGR\_LOC\_ZK\_INFO\_DAY

PMID5.TB\_MID\_GRP\_USER\_MAGR\_LOC\_FOR\_FEE\_INFO\_DAY

PCDE.TB\_CDE\_ERP\_NAME\_MANU

PMID. TB\_MID\_PAR\_USER\_MON

PCDE.TB\_CDE\_CMIOT\_ERP\_RLT\_MANU

PCDE.TB\_CDE\_WLW\_GK\_CODE\_MANU

PCDE.TB\_CDE\_RUN\_DICT\_CODE

PDATA.TB\_IOT\_INVOICE\_DETAIL\_OWE\_MON

·-----输出表------RPTMART3.TB\_RPT\_GRP\_TZD\_FOUR\_DTL\_T32\_MON

### 3. 安装配置

#### 步骤一:解压发布版 SQL\_助手.zip,解压后会有三个文件

□ 初始化文件 2024/3/29 10:12 文件夹 2024/3/29 10:16 YML 文件 2 KB 2024/3/29 10:07 应用程序 72,434 KB

- SQL\_HELLPER\_R.exe : 应用程序
- conf.yml : 配置文件
- 初始化文件:程序运行所需的模版、数据库信息

步骤二:将 "SQL\_HELLPER\_R.exe"和配置文件 "conf.yml" 放在同一个目录下

修改配置文件,根据自己需要,修改配置文件的路径。

```
□conf path:
                           主路径, 只要设置这个就行
     GEN PATH: E:\\思特奇\\
2
     XUQIU_MOBAN_PATH: SQL_HELPER目录\\模版\\ #文件模版
GONGZUOQU_PATH: 学习笔记\\数仓任务\\ #工作区代码路径
3
4
5
     DATA PATH: SQL HELPER目录\\数据\\ # 数据区代码路径
6
     HSZ PATH: SQL HELPER目录\\回收站\\ # 数据区代码路径
     VS PATH: D:\\Users\\86158\\AppData\\Local\\Programs\\Microsoft VS Code\\Code.exe
     #打开代码的工具,建议使用VSCODE
8
     WPS PATH: D:\\Users\\86158\\AppData\\Local\\Kingsoft\\WPS Office\\ksolaunch.exe #
9
     TXT_PATH: D:\\Program Files\\Notepad++\\notepad++.exe # TXT文件打开程序路径
     EDGE PATH: C:\\Program Files (x86)\\Microsoft\\Edge\\Application\\msedge.exe #浏览
10
     VSDX PATH: C:\\Program Files\\Microsoft Office\\root\\Office16\\VISIO.EXE # visio
11
     DB PATH: SQL HELPER目录\\ # 数据库存储路径
     MOBAN_PATH: SQL_HELPER目录\\模板目录\\ #代码模版路径 组装区模版路径
13
     DB NAME: SQL Helper.db
14
     SJTC_MOBAN_PATH: SQL_HELPER目录\\模板目录\\数据探查模版\\ #数据探查脚本模版目录 字
15
     LSJS PATH: E:\\临时接收\\ # 临时接收路径,导入文档打开的初始目录
16
     XQLX: 任务,取数,其他 # 任务类型自定义
17
     ZDGL MAX CNT: 10 #同时打开字段管理界面的个数
18
     TXSJ: 5 # 提醒周期,每几分钟提醒一次
19
     TXHS: 3 #喝水提醒(每几个周期提醒一次)
20
     TXYD: 6 # 运动提醒 (每几个周期提醒一次)
21
     ZDXQ: 图书馆 | 周报 | 口径 #需要置顶的需求
                                        需要始终展示的需求的名称
```

#### 应用配置与打开文件的对应关系:

配置路径	打开的文件
WPS_PATH	打开文档文件,包括 word,excel,ppt
TXT_PATH	后缀是 txt 的文件
EDGE_PATH	pdf 文件
VSDX_PATH	visio 文件
VS_PATH	SQL 文件

注:推荐使用 VSCOD 和 WPS 这些都做了适配,其他的可以修改成对应路径试一下,看可不可以。

程序界面

■ SQL助手									
文件 元数据管理 关于 需求管理 元数据管理	口径管理 表管理 字段	管理 需求血缘							
当前工程: 思特奇	∨ 需求:	~							
,T1.PHONE_NO AS '手机号'						T: *** PI *** ** ** T: **	( ************************************	*******JIE	EMI_1***********  EMI_2*********  EMI_3*******  EMI_4*******  EMI_4***********  EMI_5************
						1		号	模式字符串切割
						2	2		字符串切割
						3	3		字符串切割
						4	999		变量分发
参数: JI JIEMI	~ 添加	删除 复制	分割换行	去除指定行	回退		添加	保存	预览
, \$ {AS_H}									
history: AS	, \$ {AS_H}		> 搜索模版:					选择模片	版 DS-关联-地市区县

, SUBSTRING(T1.PHONE\_NO,1,LENGTH(T1.PHONE\_NO)-4)||F.LAST\_NUM AS'手机号'LEFT JOIN PMID.DATADECRYPT F
ON SUBSTRING(T1.PHONE\_NO,LENGTH(T1.PHONE\_NO)-3,4)=TRIM(F.DECR\_STR)