蚂蚁平台用户指南

ETL 数据开发平台

通用约定

格式	说明	样例
	该类警示信息将导致系统重大变更甚至故障，或者导致人身伤害等结果。	**危险**:重置操作将丢失用户配置数据。
	该类警示信息可能导致系统重大变更甚至故障，或者导致人身伤害等结果。	**警告**:重启操作将导致业务中断，恢复业务所需时间约 10 分钟。
	用于补充说明、最佳实践、窍门等，不是用户必须了解的内容。	**注**:您也可以通过按 Ctrl + A 选中全部文件。
>	多级菜单递进。	**设置** > **网络** > **设置网络类型**
粗体	表示按键、菜单、页面名称等 UI 元素。	单击 **确定**。
courier 字体	命令。	执行 cd /d C:/windows 命令，进入 Windows 系统文件夹。
斜体	表示参数、变量。	bae log list --instanceid *Instance_ID*
[]或者 [a|b]	表示可选项，至多选择一个。	ipconfig *[-all|-t]*
{}或者 {a|b}	表示必选项，至多选择一个。	swich *{stand | slave}*

通用约定 2
一、文件操作 6
1、新建文件 6
2、私有文件转移到公有文件夹 6
3、查看文件基本信息 6
4、新建文件夹 6
5、移动文件夹 6
6、删除文件夹 7
7、复制文件 7
7、移动文件 7
8、删除文件 7
9、搜索文件 8
10、加锁文件 8
11、解锁文件 8
12、查看文件版本 9
13、比对版本 9
14、回滚版本 10
15、导出文件 11
16、导入文件 11
二、文件编辑 12
1、开始 12
1.1 新建文件 12
1.2 编辑公共文件 12
1.3 编辑私有文件 12
1.4 最大打开文件数 12
1.5 保存文件 13
1.6 格式化文件 13
1.7 语法检查 13
2 工具 14
2.1 文件比较 14
2.2 时间函数 15
3、编辑 16
3.1 剪切 16
3.2 复制与粘贴 16

3.3 撤销与重做 16
3.4 缩进和取消缩进..... 17
3.5 全选 17
3.6 转到行 17
3.7 查找 18
3.8 替换 19
3.9 代码注释 19
3.10 代码折叠 19
3.11 全屏显示 20
三、变量设置..... 21
1、自定义变量替换 21
2、时间变量替换 21
2.1 根据业务日期进行加减计算..... 21
2.2 根据当前时间进行加减运算..... 22
四、运行与结果..... 24
1、提交 SQL 运行 24
2、常用数据源选择 24
3、终止 SQL 运行 25
4、SQL 变量替换 25
5、运行日志 25
6、运行结果 26
7、结果集操作 26
五、创建 UDF 27
1、ODPS 的 UDF 简介 27
2、创建 UDF 27
3、修改 UDF 28
4、查看 UDF 28
5、删除 UDF 28
6、使用 UDF 29
六、数据字典..... 30
1、Project 选择 30
2、浏览表 30
3、搜索表 31
4、查看表详细信息 32
5、收藏表 34

七、运行记录..... 35
1、查看运行记录 35
2、搜索运行记录 35
3、打开运行记录 35
4、刷新运行记录 36
5、删除运行记录 37
6、复制运行记录 37
八、管理设置..... 39
1、编辑区设置 39
2、脚本格式化 39
3、个人偏好 39

一、文件操作

1、新建文件

数据工厂中目前支持 ODPS 脚本的编写和执行，可以从工具栏和浏览窗口新建文件。

1. 通过以下方式之一，新建 ODPS 文件：
   - 在菜单详情区，右键点击目标位置，选择 **新建** > **新建 ODPS 文件**，输入文件名，点击 **确定**。
   - 在工具栏，单击工具栏的添加图标（ ），选择 **新建 ODPS 文件**。
     注意：此时文件并未指定位置，需在保存时选择目标位置并填写文件名。
   - 在文件标签栏，点击加号图标（ ）快速创建 ODPS 文件。
     注意：此时文件并未指定位置，需在保存时选择目标位置并填写文件名。

2、私有文件转移到公有文件夹

在菜单详情区，您可以看到以下文件类型的根目录：

- 公有文件：租户内部所有人都可以看到和编辑的文件。
- 私有文件：只有自己才能看到和编辑的文件。
  限制：私有文件可以移动到公有文件夹中，但是公有文件不能移动到私有文件夹中。

3、查看文件基本信息

在菜单详情区，右键点击目标文件，选择 **详细信息**，该文件的详细信息窗口。

4、新建文件夹

1. 在菜单详情区，右键点击目标位置，选择 **新建文件夹**。
2. 填写文件夹名
3. 点击 **确定**。

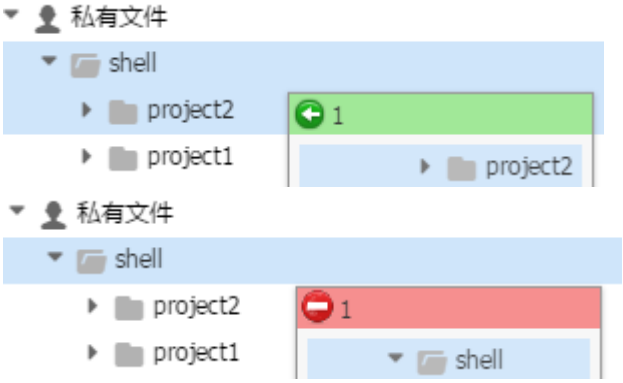
5、移动文件夹

选中要移动的文件夹，拖动文件夹至目标位置。

提示： 拖动时，鼠标右侧会显示文件夹的可移动状态，如下图所示：

- 绿色表示可以移动到当前位置。

- 红色表示不可以移动到当前位置。



6、删除文件夹

1. 在菜单详情区，选中准备删除的文件夹，点击 **删除**。
2. 点击 **确定**。

7、复制文件

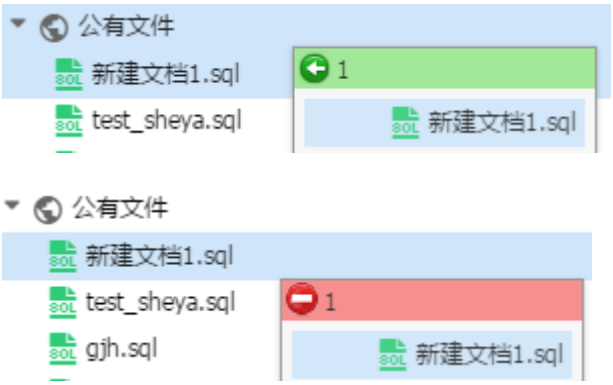
1. 右键点击要复制的文件，点击 **复制**。
2. 右键点击目标位置单击右键，点击 **粘贴**。
   *注意：您也可以右键点击要复制的文件，选择 **复制并粘贴**。*

7、移动文件

选中要移动的文件，拖动文件至目标位置。

提示：拖动时，鼠标右侧会显示文件的可移动状态，如下图所示：

- 绿色表示可以移动到当前位置。
- 红色表示不可以移动到当前位置。



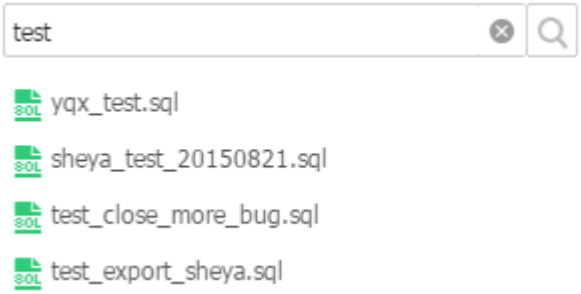
8、删除文件

1. 右键点击要删除的文件，点击 **删除**。

2. 点击 **确定**。

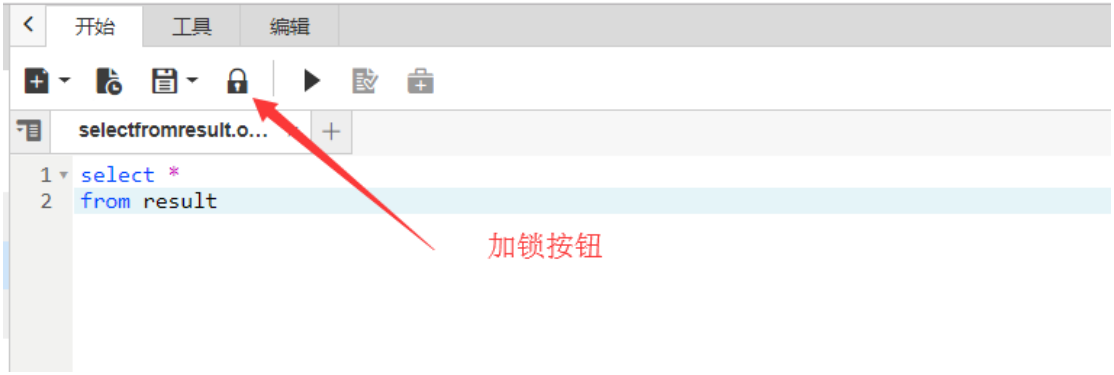
9、搜索文件

在菜单详情区的搜索框中，输入文件名、关键字进行查询。



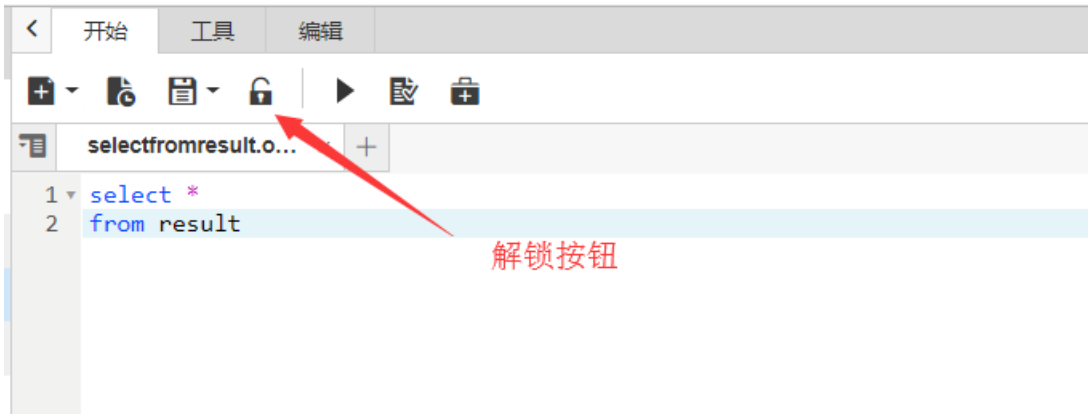
10、加锁文件

对于公有文件夹中的文件，只有经操作者加锁后才允许编辑，因此编辑公有文件前需要进行加锁操作。打开公有文件夹中未加锁的文件，在工具栏中点击“加锁”按钮。



11、解锁文件

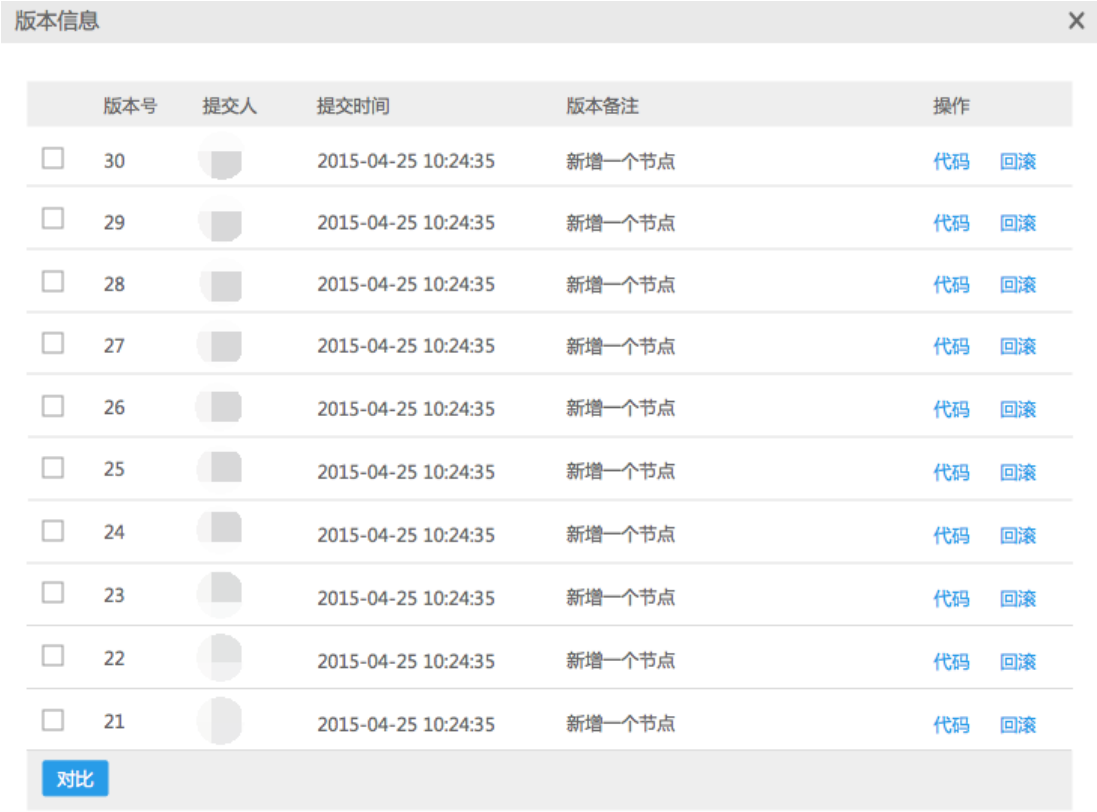
对于公有文件夹中由当前用户加锁的文件，经当前用户编辑完成后，为避免资源占用，应当及时进行解锁操作。打开公有文件夹中已加锁的文件，在工具栏中，点击 **解锁** 按钮。



12、查看文件版本

公有文件可以多人协同工作，同一个文件可能由于被不同成员或不同时间的多次修改，从而存在多个版本。

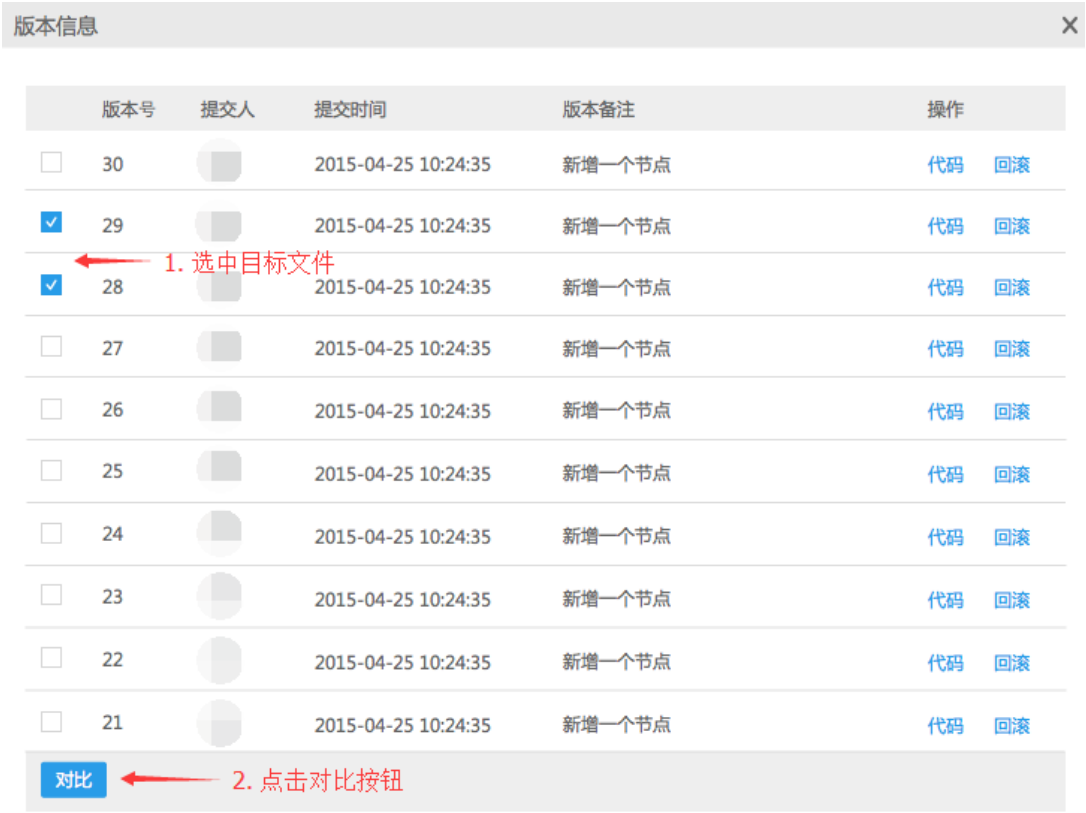
右键点击目标文件，选择 **代码对比**。在弹出窗口中可以查看文件的版本信息，包括版本号、提交人、提交时间和版本备注等。



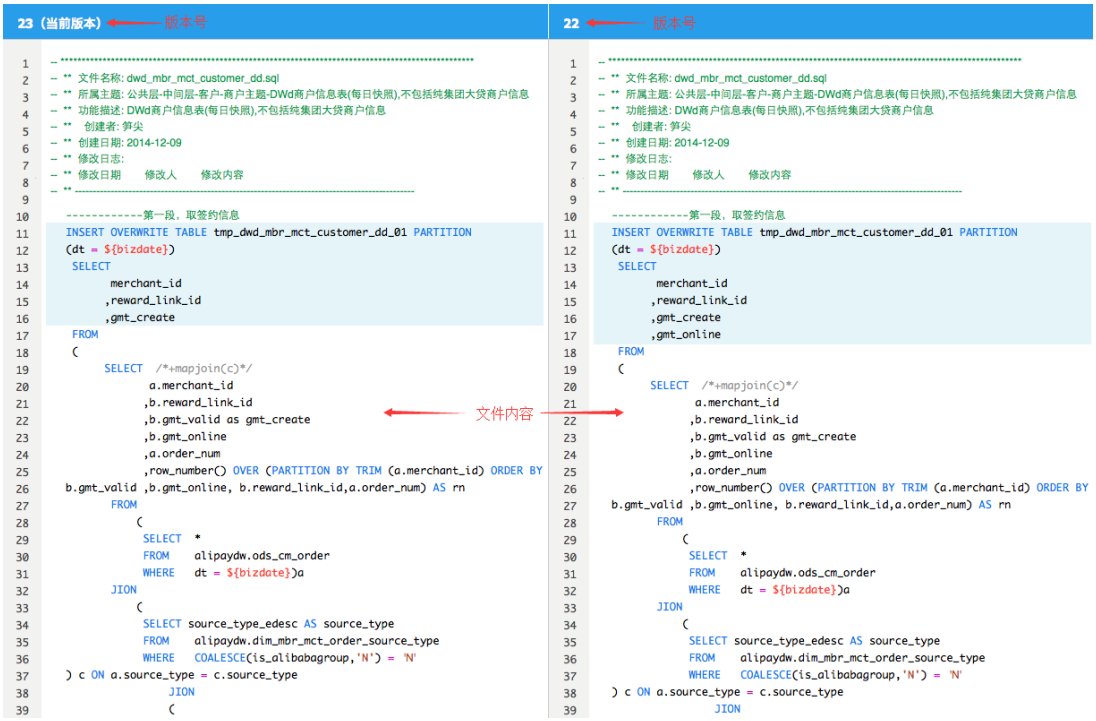
13、比对版本

1. 在 **版本信息** 窗口中选择需要对比的文件，点击 **对比** 按钮。

- **注意：**目前仅支持两个版本的文件对比。



2. 进行文件对比。



14、回滚版本

在 **版本信息** 窗口中，点击目标版本右边的 **回滚** 按钮可以执行版本回滚操作。

15、导出文件

在菜单详情区，右键点击要导出的文件，点击 **导出**。文件将以 `.zip` 压缩包的形式导出到本地机器。

16、导入文件

1. 右键点击目标文件夹，选择 **导入**。
2. 选择由文件导出功能导出的 `.zip` 压缩文件，点击 **打开**。文件导入到当前文件夹。
   注意： 暂不支持其他类型的文件。

二、文件编辑

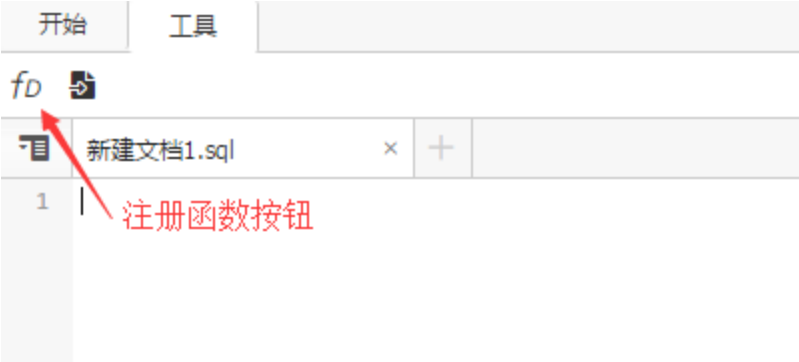
1、开始

1.1 新建文件

参见 文件操作 > 新建文件。

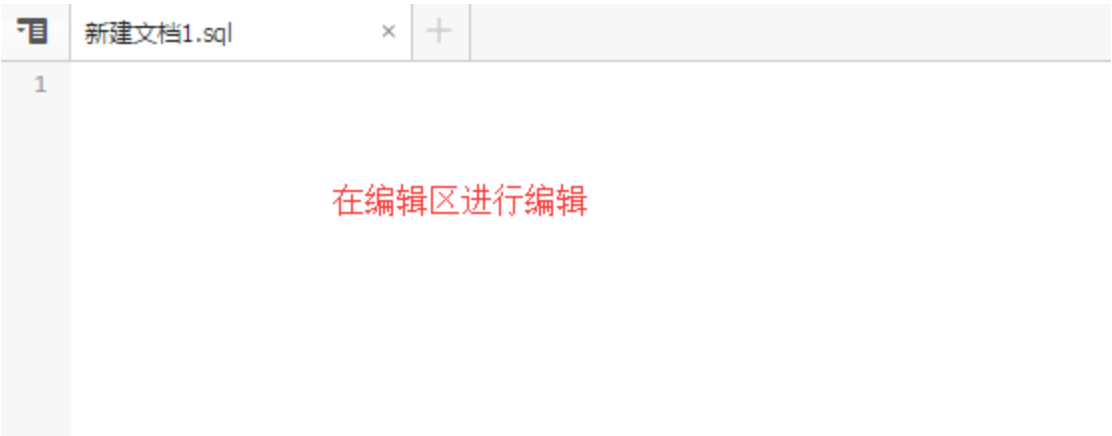
1.2 编辑公共文件

打开公有文件夹中的文件，在编辑区中进行编辑。在编辑公有文件前，操作者需对文件加锁，否则无法编辑文件。当前用户未加锁或已有他人加锁的文件，编辑区会弹出 **没有文件编辑权限，请先解锁文件** 的提示。完成编辑后，需及时解锁文件，以便他人使用。



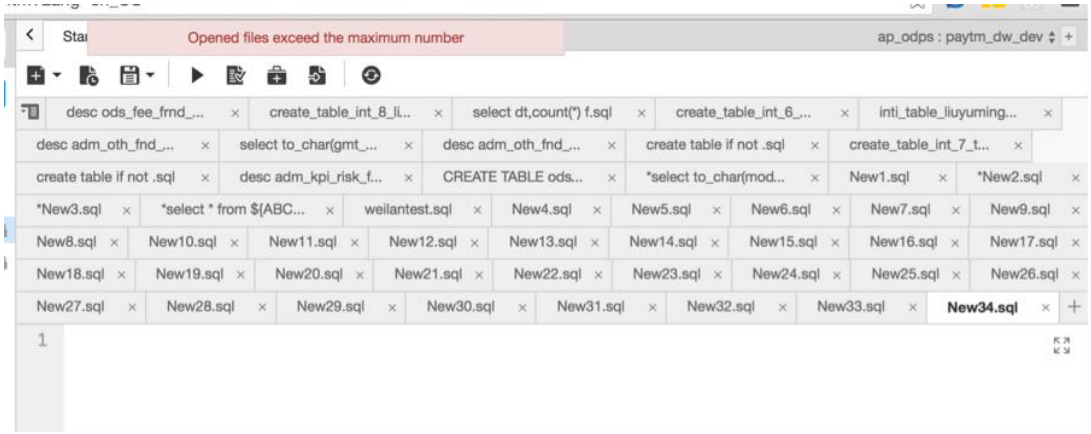
1.3 编辑私有文件

打开私有文件夹中的文件，在编辑区中进行编辑。私有文件无加锁解锁概念，当前用户可以随时编辑。



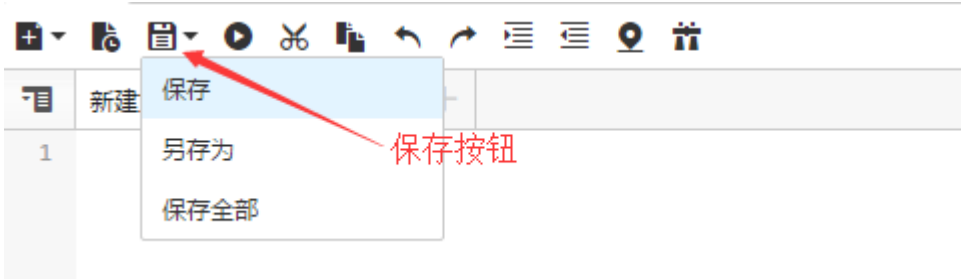
1.4 最大打开文件数

编辑区最多可以打开 15 个文件，超过该数目，编辑区会弹出 **打开的文件超过上限** 的提醒。



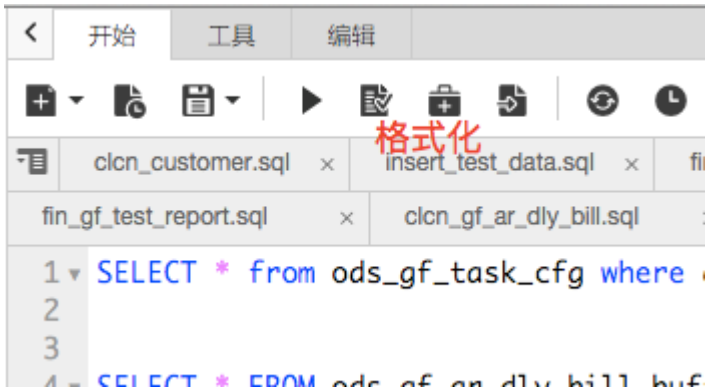
1.5 保存文件

在工具栏中点击 **保存** 按钮，选择 **保存** 、**另存为** 或 **保存全部**。



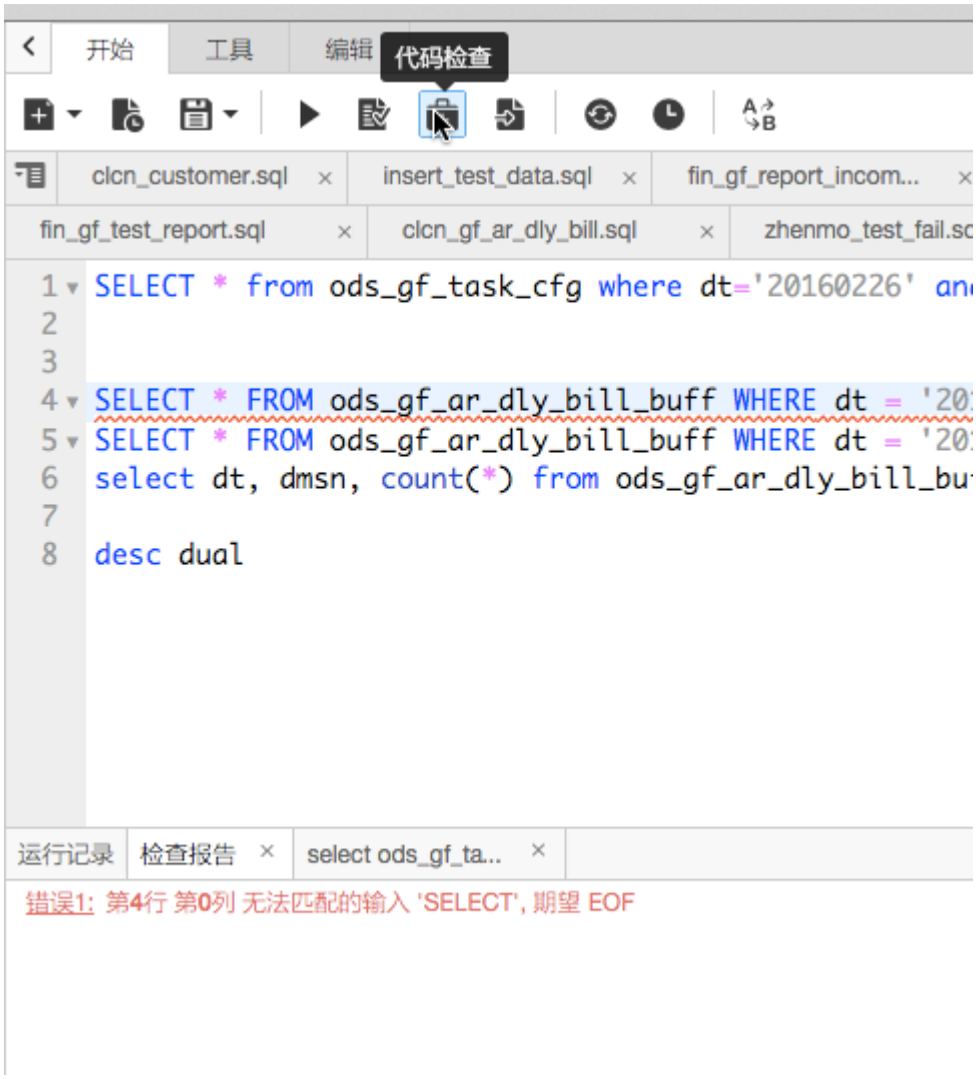
1.6 格式化文件

在工具栏中点击 **格式化** 按钮，可以格式化当前的 SQL 语句，关键字会高亮显示，会自动换行。



1.7 语法检查

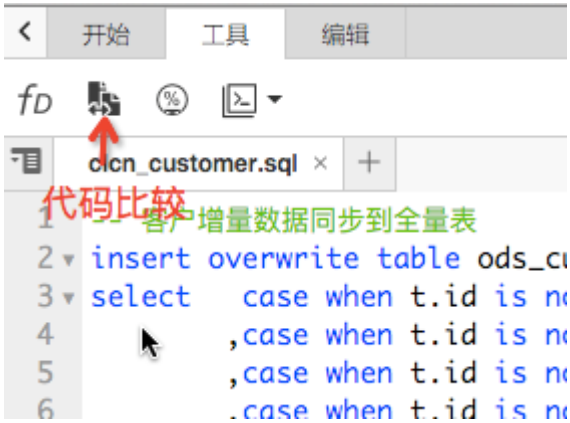
在工具栏中点击 **语法检查** 按钮，可以检查当前的 SQL 语句的语法是否正确。



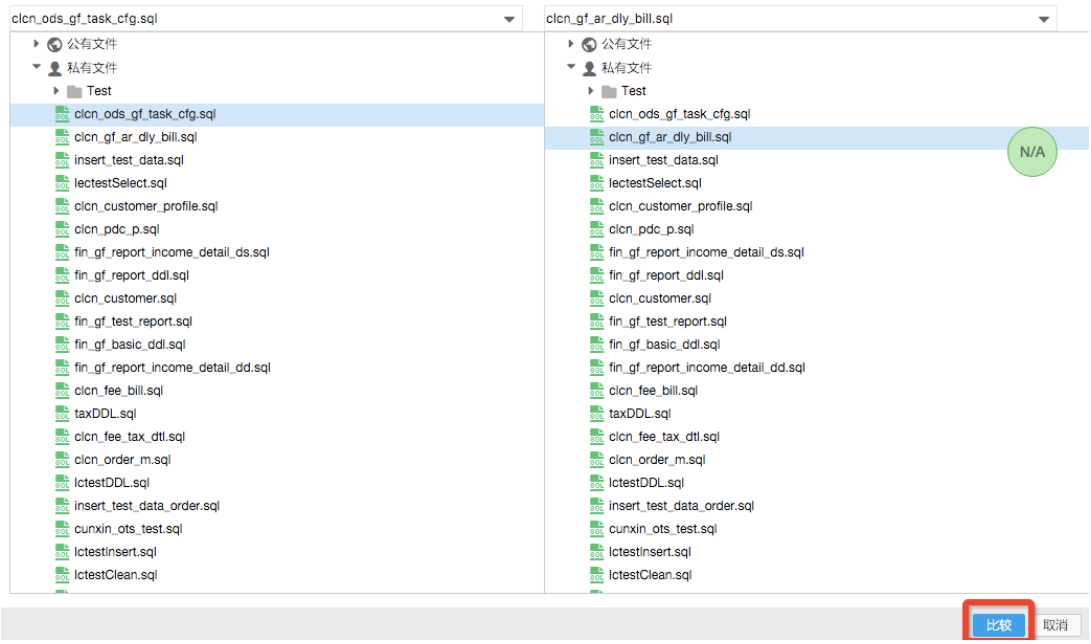
2 工具

2.1 文件比较

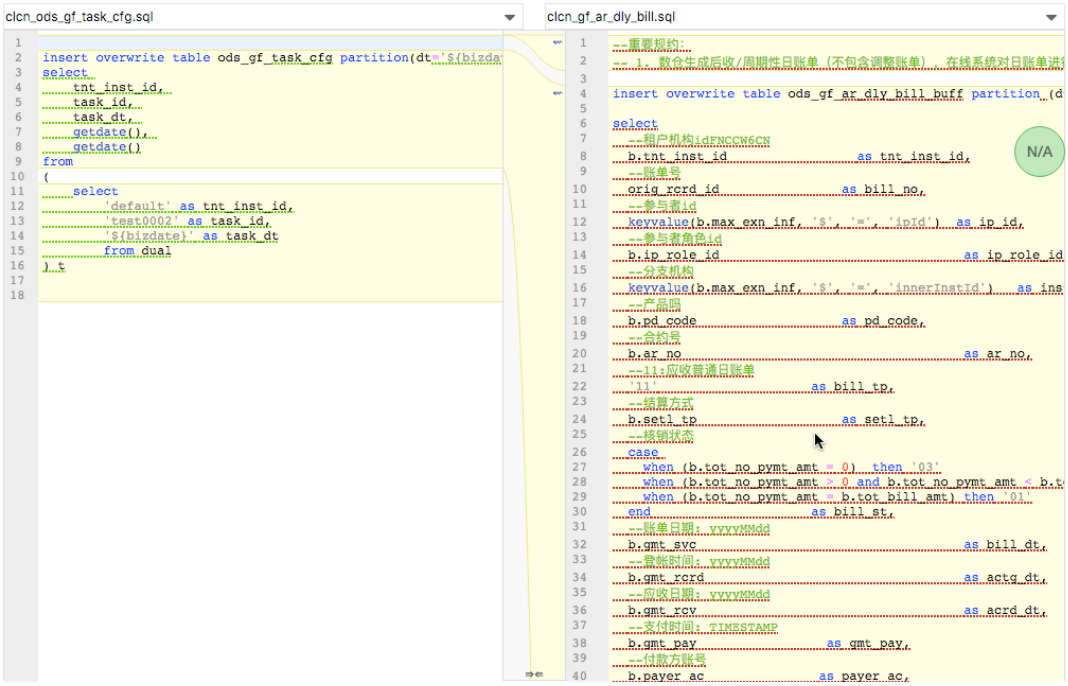
在工具栏中点击 **文件比较** 按钮，可以选择任意两个文件，进行内容比较：



在文件浏览器中，选择要比较的文件，点击 **比较** 按钮：

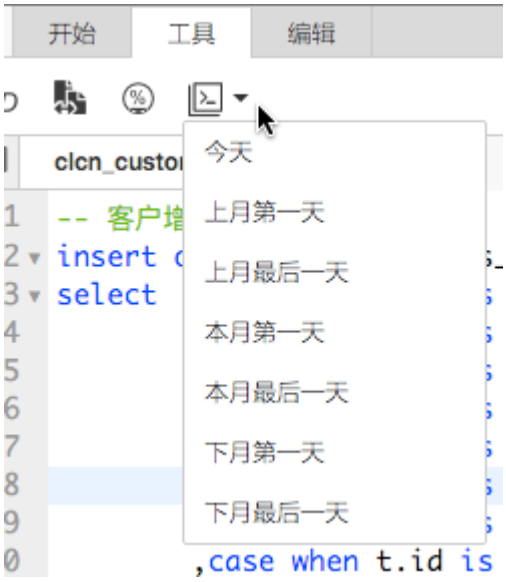


在比较结果页面，可以看到两个文件的差异点：



2.2 时间函数

系统支持常见的时间变量的替换：



3、编辑

3.1 剪切

选中需要剪切的内容，点击工具栏上的 **剪切** 按钮或由键盘键入 CTRL+X。

3.2 复制与粘贴

选中需要复制的内容，点击工具栏上的 **复制** 按钮或由键盘键入 CTRL+C。

将光标移动目标位置，由键盘输入 CTRL+V，完成粘贴。

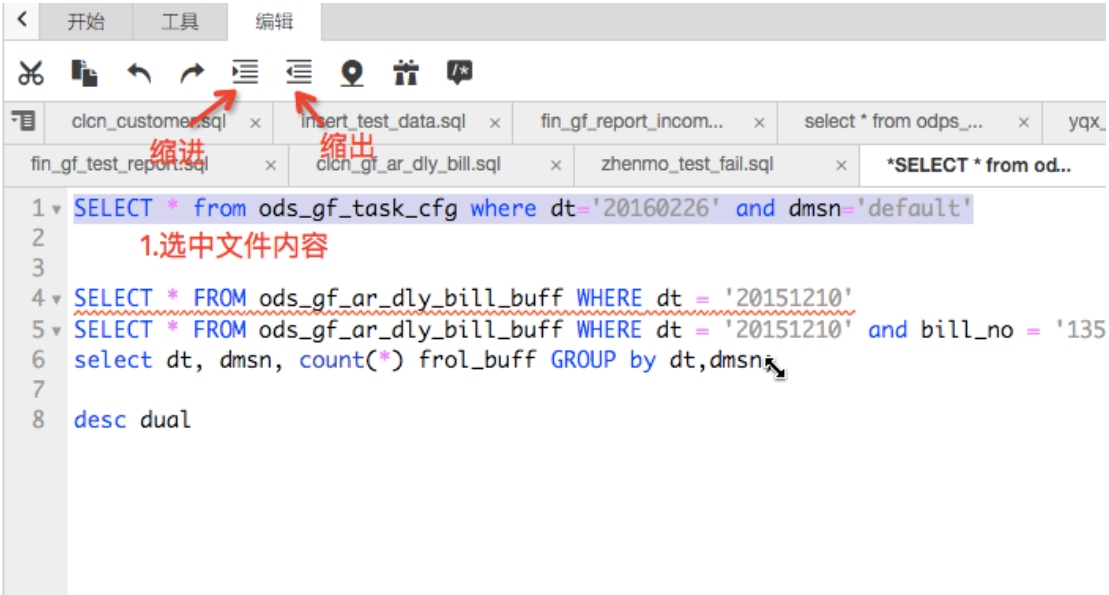
3.3 撤销与重做

对已执行的操作，可以进行撤销和重做。在工具栏上点击 **撤销** 或 **重做** 按钮即可。

3.4 缩进和取消缩进

为了代码排版美观、便于阅读，可以在数据工厂中对代码进行缩进和取消缩进。

选中需要操作的内容，点击工具栏上的 **缩进** 按钮或由键盘键入 TAB 进行缩进；点击工具栏上的 **取消缩进** 按钮或由键盘键入 SHIFT+TAB 键进行取消缩进。

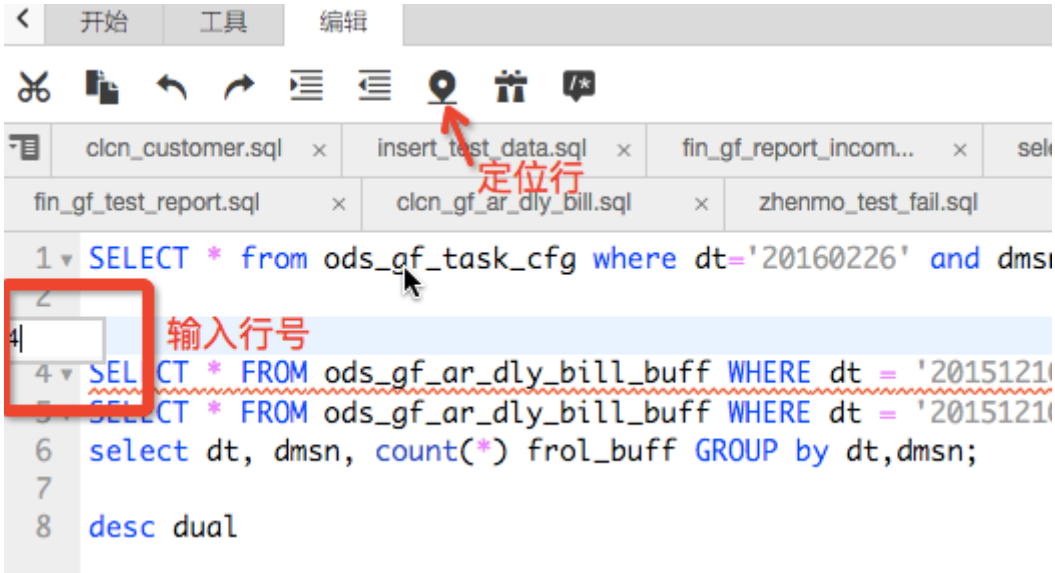


3.5 全选

将光标移至编辑区，由键盘键入 CTRL+A，编辑区内所有文字即被选中。

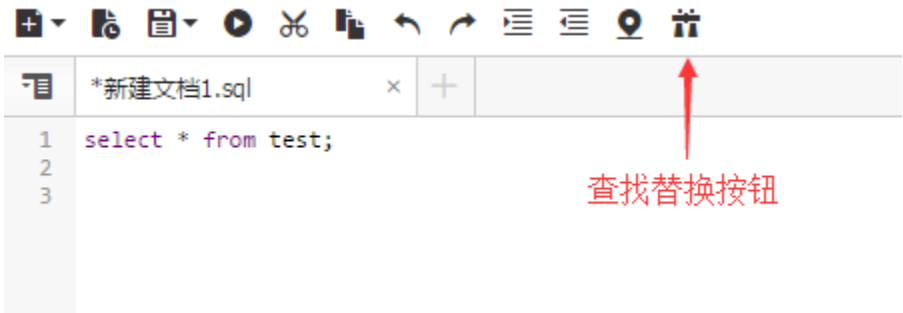
3.6 转到行

为了快速定位到某一行代码，可以点击工具栏的 **转到行** 按钮。在弹出的行号输入框中输入目标行号，点击回车，光标将跳转到输入的行数处。

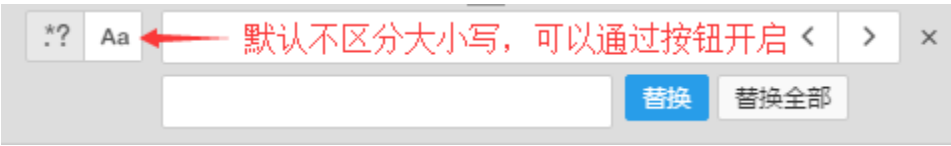


3.7 查找

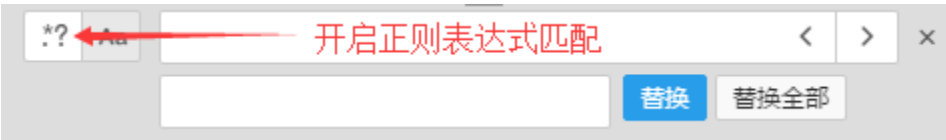
点击工具栏的 **查找替换** 按钮，或由键盘键入 CTRL+F，调出查找框。在查找框中输入查找条件。匹配的内容会在编辑区内背景高亮显示。



默认不区分大小写，可以点击左侧 **大小写** 按钮进行区分。



查找功能支持正则表达式（仅支持 JavaScript 中正则表达式语法），可以点击左侧的 **正则** 按钮进行开关。

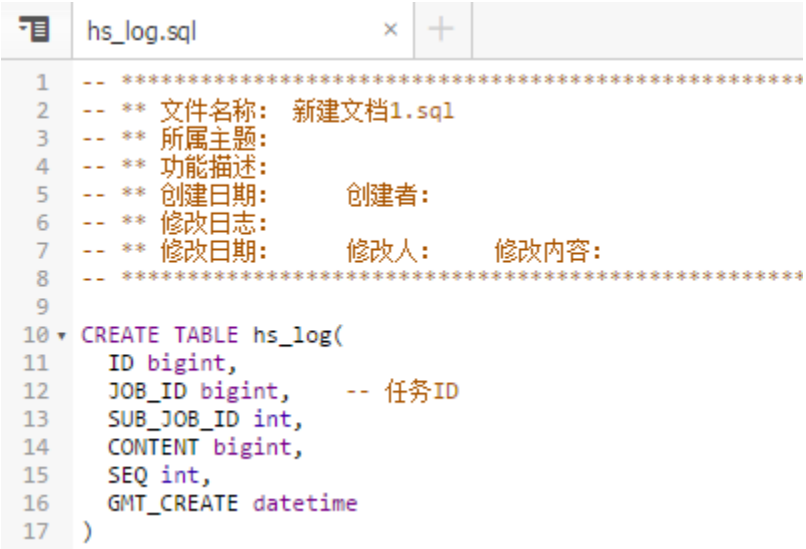


3.8 替换

点击工具栏的 **查找替换** 按钮，在查找框中输入查找条件，在替换框中输入替换内容。查找条件同样支持大小写区分和正则表达式。待完成查找条件和替换内容的填写，点击 **替换** 或 **替换全部** 按钮。

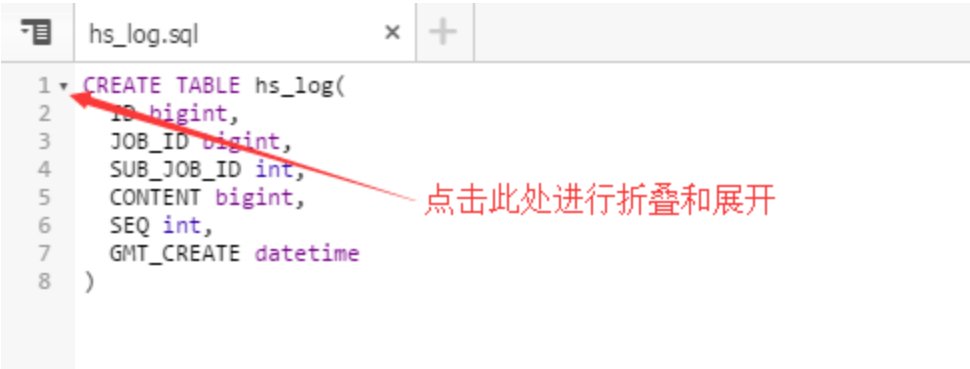
3.9 代码注释

数据工厂目前支持一种简易模式的注释，即 --。虽然支持类似 Oracle 的标准注释，形如 /* …. . */，但并不推荐，因为 /* …. . */ 也为 ODPS 的关键字。



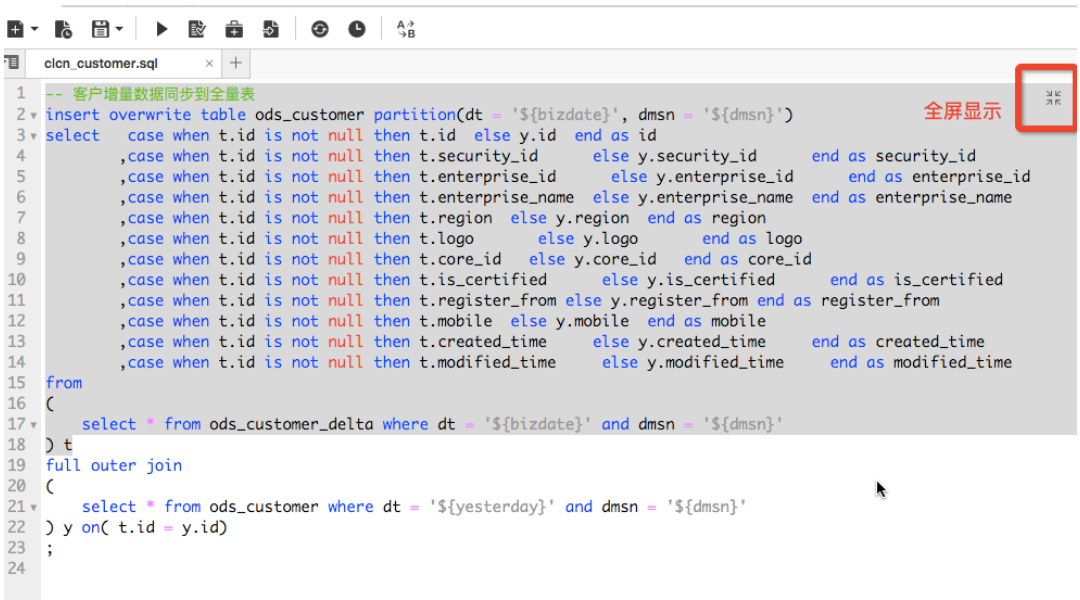
3.10 代码折叠

为方便代码阅读，可以将代码块进行折叠。点击编辑区左侧的 **折叠** 或 **展开** 按钮即可。



3.11 全屏显示

点击 **全屏切换** 按钮，可以全屏化编辑区，方便脚本编写和查阅。

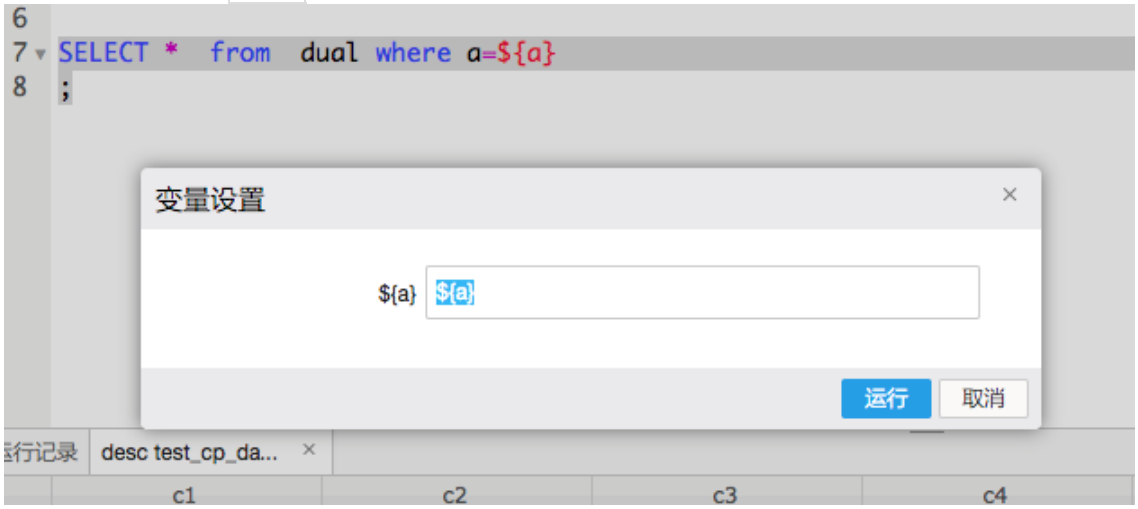


三、变量设置

数据工厂支持变量替换，方便在脚本执行时，替换成实际值运行。

1、自定义变量替换

数据工厂支持 `${a}` 的方式定义变量，在运行时，会提示输入变量值：



2、时间变量替换

2.1 根据业务日期进行加减计算

数据开发中往往会使用 T+1 的数据，所以使用业务日期来进行计算，一般而言，这个业务日期就是昨天。

业务代码中可以根据 **业务日期** 进行加减计算，精确到天，一般形式如：

```
1. @@[yyyyMMdd +/- Ny]
```

- 说明
  - @@{}：表示以数据日期为基准进行计算
  - yyyyMMdd：表示日期的格式，还可以为 yyyy、yyyyMM、MM、dd 等
  - +/-：表示进行加或减计算
  - Ny：N 表示需要加减的数值；y 表示单位：年，其中包括如下的单位
    - y：年
    - m：月
    - d：天

- 例子

  假设当前的数据日期为 20140603 ，则

  - @@{yyyy-MM-dd + 2y}：2016-06-03
  - @@{yyyyMMdd - 2m}：20140403
  - @@{yyyyMMdd - 2d}：20140601
  - @@{yyyy - 2y}：2012
  - @@{yyyyMM + 3m}：201409

  代码示例：

```
1. SELECT * FROM table_a WHERE t_date = @@{yyyyMMdd - 1d}
```

上例中 @@{yyyyMMdd - 1d} 即为数据日期 -1 天，如果数据日期为 20140603 ，则 @@{yyyyMMdd - 1d} 表示 20140602 ，那么最终执行的代码为：

```
1. SELECT * FROM table_a WHERE t_date = 20140603
```

2.2 根据当前时间进行加减运算

可以根据当前时间进行加减计算，精确到秒。一般形式如：

```
1. @@[yyyyMMdd HH:mm:ss +/- Ny]
```

- 说明
  - @@[]：表示以当前时间为基准进行计算
  - yyyyMMdd HH:mm:ss：表示时间的格式，还可以为 yyyy、yyyyMM、MM、dd、HH:mm:ss 等
  - +/-：表示进行加或减计算
  - Ny：N 表示需要加减的数值；y 表示单位：年，其中包括如下的单位
    - y：年
    - m：月
    - d：天
    - h：小时
    - mi：分

- 例子

  以当前时间 20140603132000 为例，则

  - @@[yyyy-MM-dd HH:mm:ss + 2y]：2016-06-03 13:20:00
  - @@[yyyyMMdd - 2m]：20140403
  - @@[yyyyMMdd - 2d]：20140601
  - @@[yyyy - 2y]：2012
  - @@[yyyyMM + 3m]：201409
  - @@[HH:mm:ss + 3mi]：13:23:00

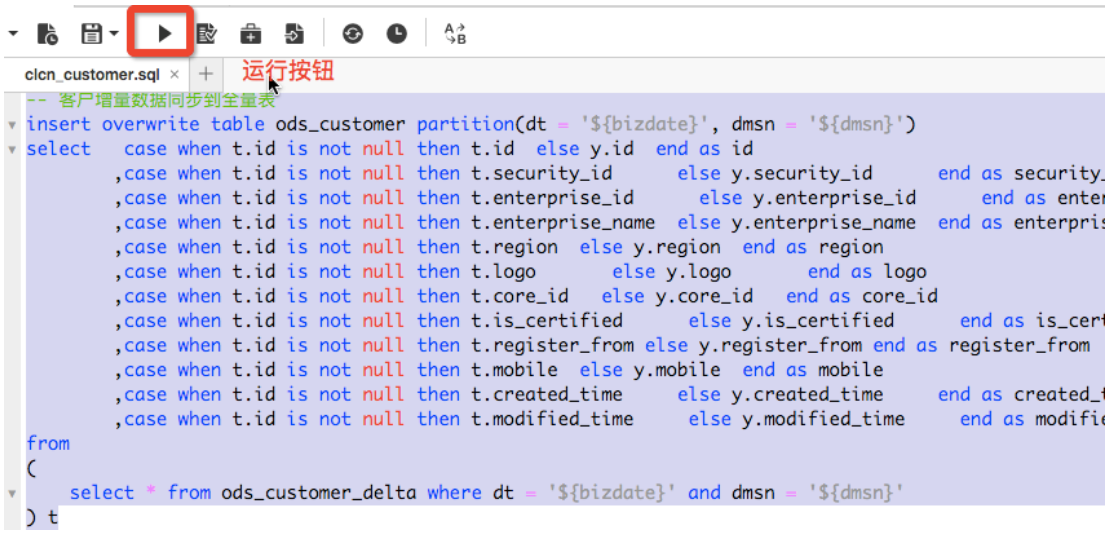
  代码示例：

```
1. SELECT * FROM table_a WHERE t_date = '@@[yyyy-MM-dd HH:mm:ss - 1d]'
```

四、运行与结果

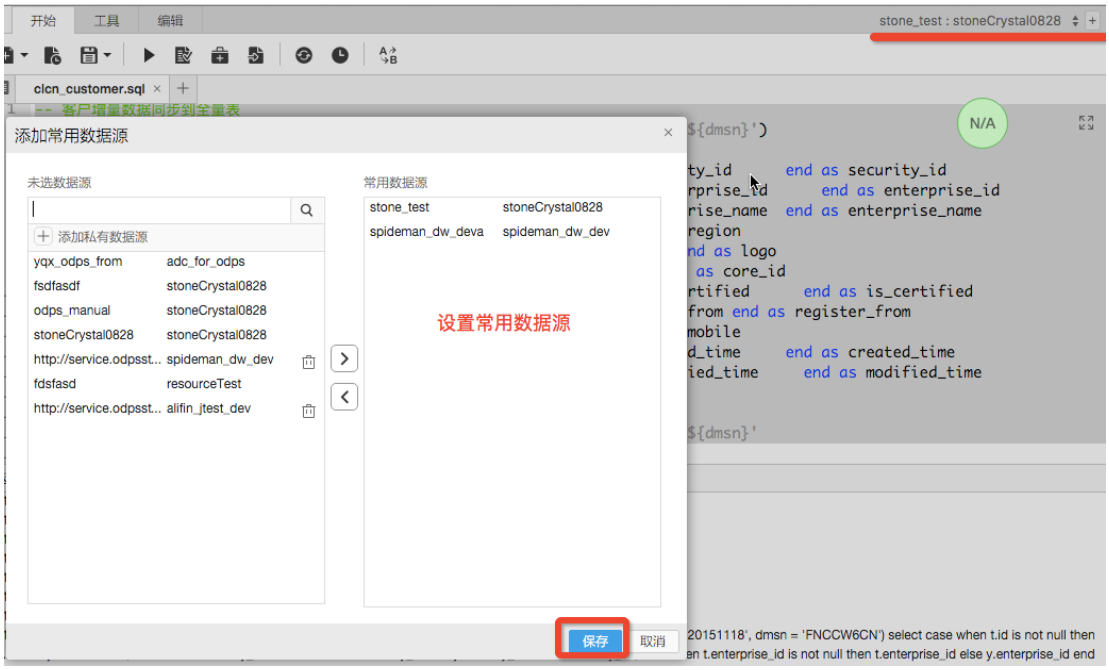
1、提交 SQL 运行

选中需要运行的 SQL 代码行，点击工具栏的 **运行** 按钮。没有选中编辑区中的任何代码时，点击运行后将执行当前文件中的所有代码。



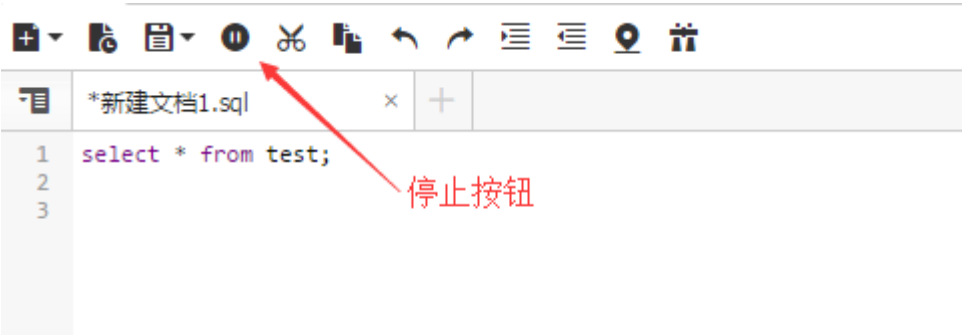
2、常用数据源选择

若当前没有选择 Project ，或选择了错误的 Project，代码执行将会出错。可以通过工具栏选择需要使用的 Project。



3、终止 SQL 运行

对于已经运行但尚未完成的任务，可以在工具栏点击 **停止** 按钮，此时处于等待中的任务将不会被执行。



4、SQL 变量替换

为减少 SQL 编码，方便执行，可以设置变量替换。需要替换的变量以 $ 符号开始，并紧跟 {} ，变量名置于花括号中。

点击执行时，若语句中有变量替换，则需在弹出窗口中填写变量替换的值。

5、运行日志

运行脚本后，日志区将呈现当前执行语句的 Log，包括当前运行的语句、执行的日志以及执行的状态。

6、运行结果

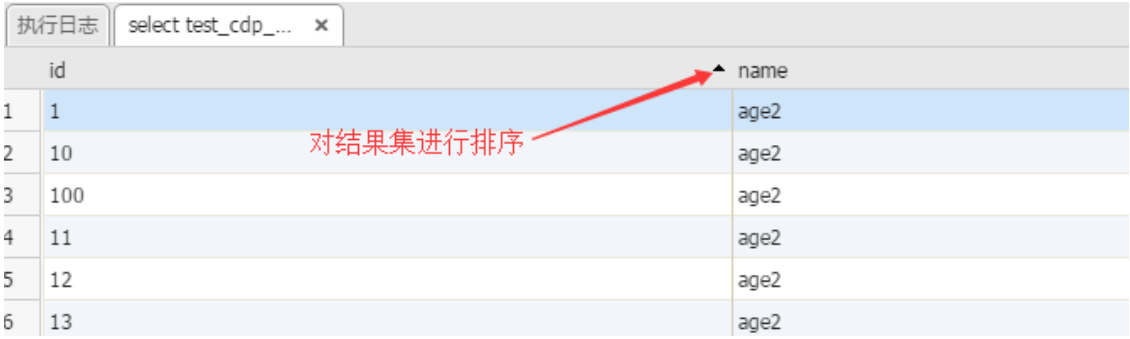
查询语句执行成功后，将会在日志区开启一个新的标签页展示查询结果。

每个文件都会产生与之对应的运行结果标签页，当切换文件标签页时，结果集也将随之切换。

此外，结果集展示的默认数据量可以由设置里的 **数据工厂最大显示行数** 进行设置（最多不超过 5000 行）。执行的 select 语句将会自动加上 `limit <数据工厂最大显示行数>`，您无需自行编写 limit 命令。

7、结果集操作

在查询结果的标签页中，可以对结果集进行排序等操作。



五、创建 UDF

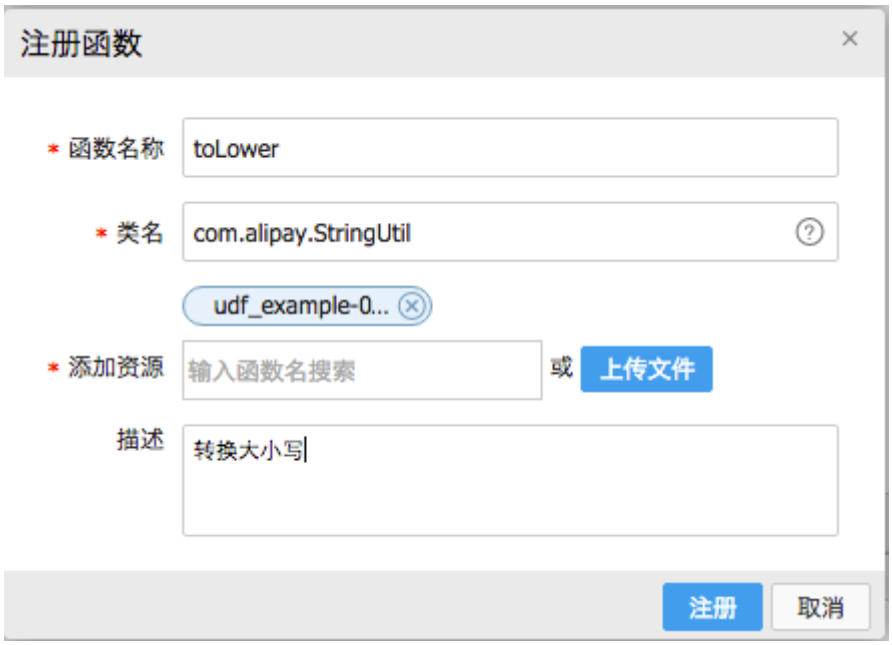
1、ODPS 的 UDF 简介

用户可以根据一定规则，注册自定义函数 UDF。经过注册的 UDF 可以在脚本文件中使用，减少重复编码，提高开发效率。

2、创建 UDF

在工具栏选择工具标签页，点击 **注册函数** 按钮。

在弹出窗口中填写相应内容，点击 **注册**。

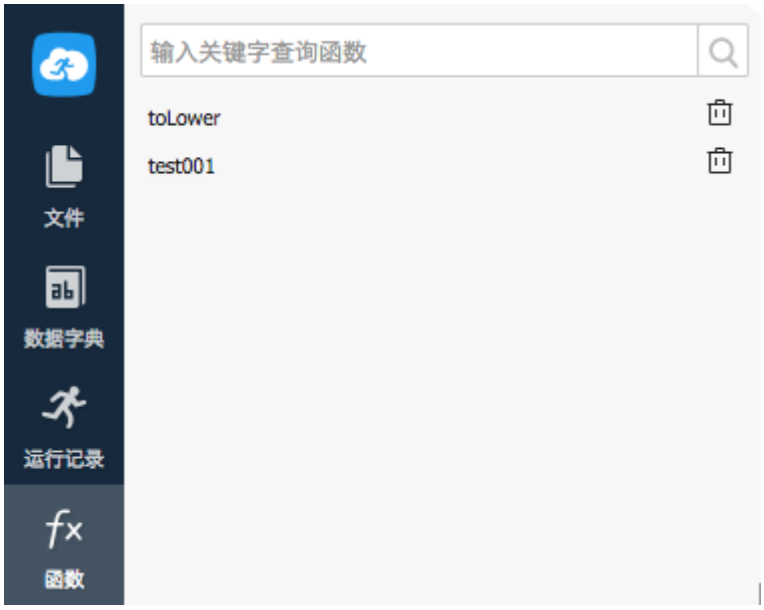


3、修改 UDF

如需修改 UDF，创建一个同名的 UDF。同名 UDF 会覆盖旧版本。

4、查看 UDF

在数据工厂侧边栏点击 **函数** 按钮，进入函数浏览窗口。在函数列表中查看已注册的 UDF。

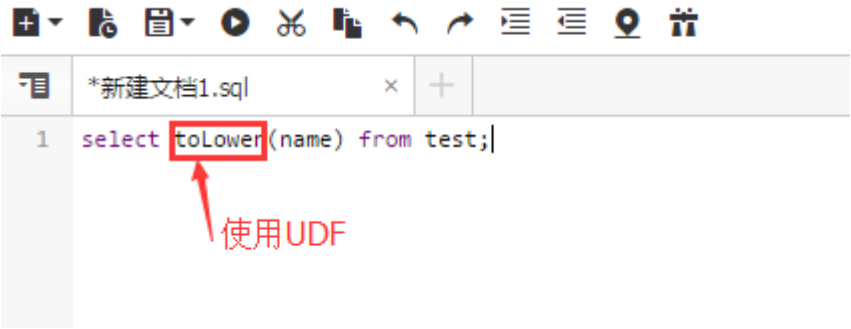


5、删除 UDF

在数据工厂侧边栏点击 **函数** 按钮，进入函数浏览窗口。在函数列表中点击要删除的 UDF 右侧的 **删除** 按钮。在弹出窗口中点击 **确认** 或 **取消**。

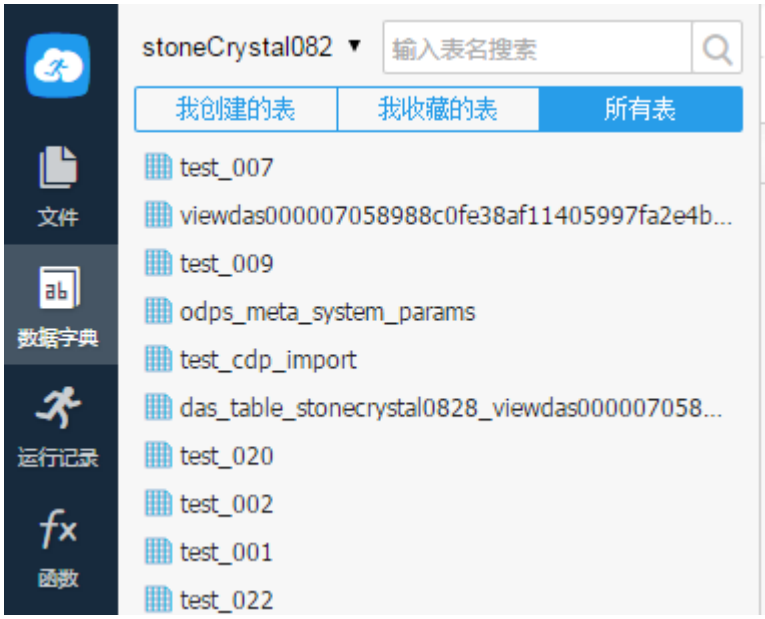
6、使用 UDF

在脚本中直接使用已注册的 UDF 即可。



六、数据字典

数据工厂提供数据字典功能以便查看 Project 下表结构，或对表进行收藏。点击左侧侧边栏中的 **数据字典** 按钮，浏览窗口中的内容相应变化为下图所示内容：

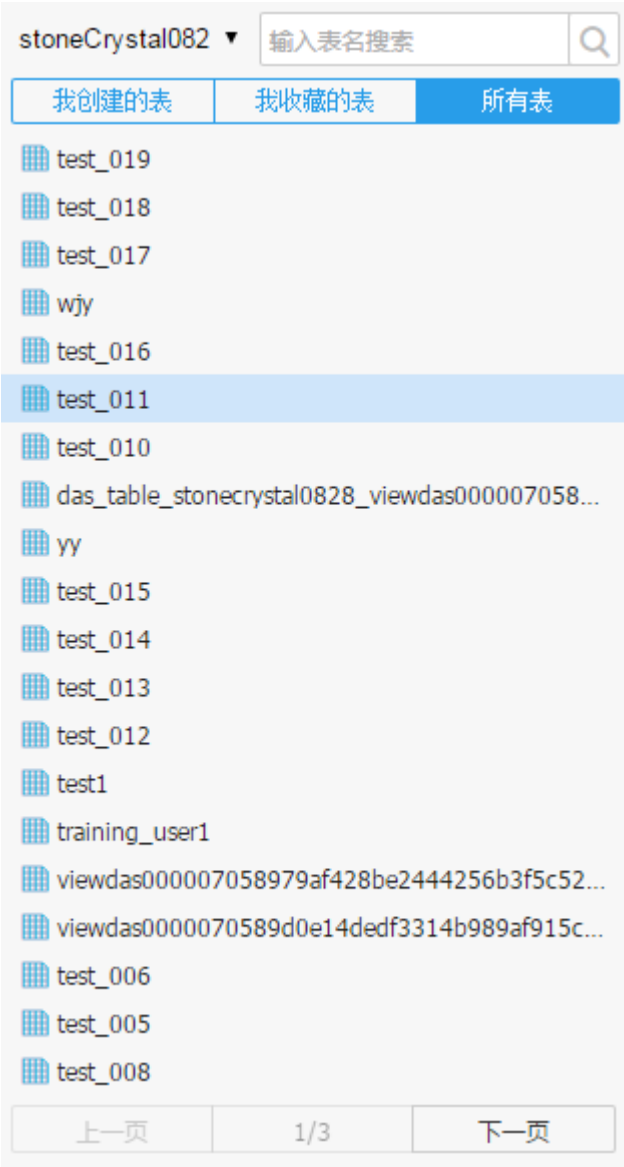


1、Project 选择

用户可以对不同 Project 中的表进行操作。浏览窗口左上角展示当前 Project，可以点击其后的三角形按钮展开 Project 列表，从列表中选择一个 Project。

2、浏览表

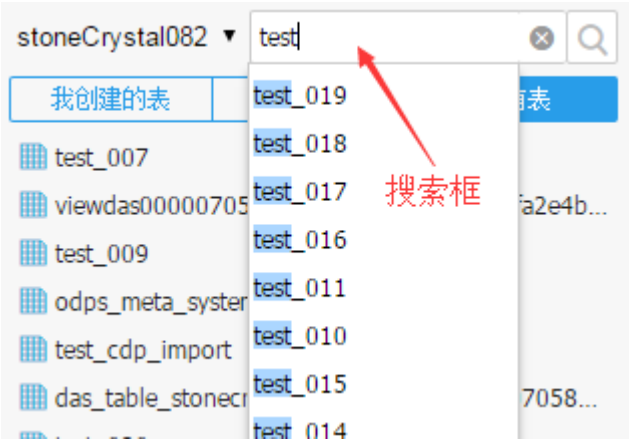
选择 Project 后，浏览窗口的列表将按照 **我创建的表** 、**我收藏的表** 和 **所有表** 对当前 Project 中的表进行展示，用户可以按照不同的需要选择不同的标签页进行查看。



当列表中的表数量过大时，将自动分页。列表下方展示了当前页码和所有页数。可以通过 **上一页** 和 **下一页** 按钮进行翻页。

3、搜索表

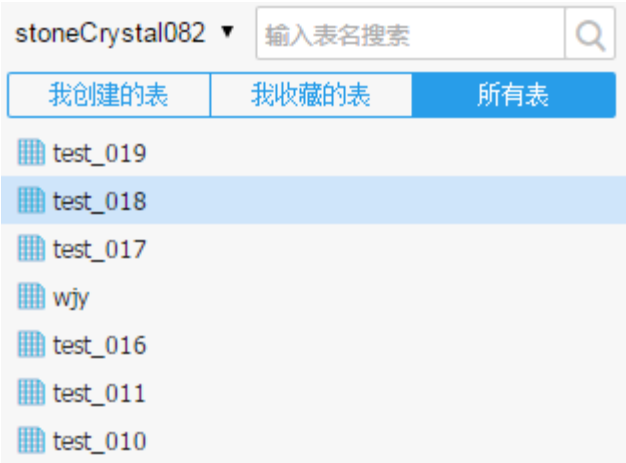
当表的数量过多时，难以从浏览窗口中快速找到目标表，为提高工作效率，数据字典提供表的搜索功能。在浏览窗口右上角的搜索框中输入搜索关键字，将会列出匹配到的表名候选项，以供选择。



4、查看表详细信息

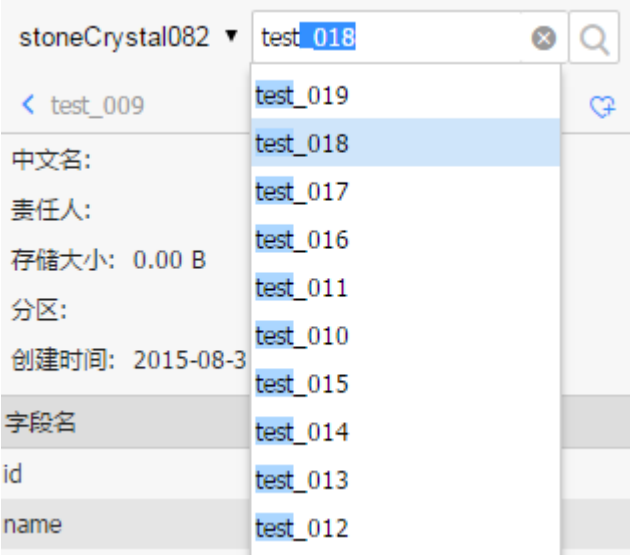
可通过如下三种方式进入表的详细信息页面：

1. 在浏览窗口中双击目标表。



2. 在浏览窗口中选中目标表，点击右键，在右键菜单中选择 **详细信息**。

3. 在搜索给出的候选项中通过上下移动光标，选中目标表，然后键盘上敲击回车键。



在表的详细信息页面展示了表的基本信息（包括中文名、责任人、存储大小、分区、创建时间等）和表结构信息（包括字段名、字段类型和描述）。

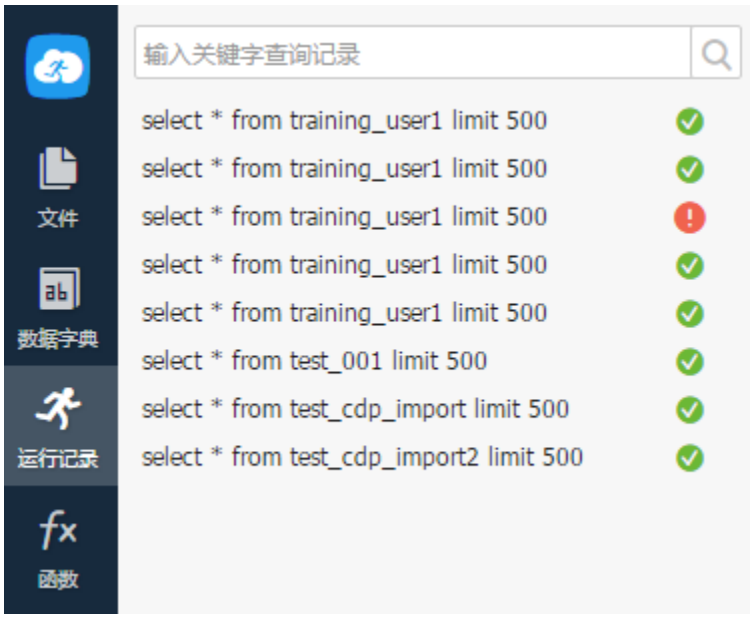
5、收藏表

为方便表的查找，可以将常用表进行收藏。在表的详细信息页面点击右上角的 **收藏** 按钮进行收藏或取消收藏。收藏完成后，表将会出现在 **我收藏的表** 标签页中。

七、运行记录

1、查看运行记录

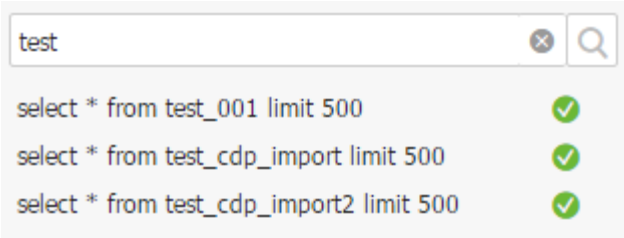
点击左侧导航栏中的 **运行记录**，用户可以查看当前 Project 中已执行的所有 SQL 语句的记录和执行状态。切换不同的 Project ，运行记录会相应切换。



浏览窗中语句右侧的绿色或红色图标分别表示该语句执行成功或失败。

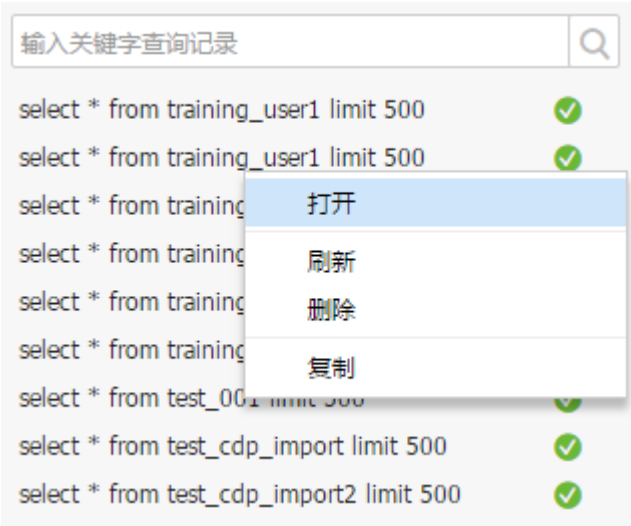
2、搜索运行记录

在浏览窗口上方的搜索框中输入搜索关键字，敲击回车键，浏览窗口将会展示符合查询条件的所有执行记录。

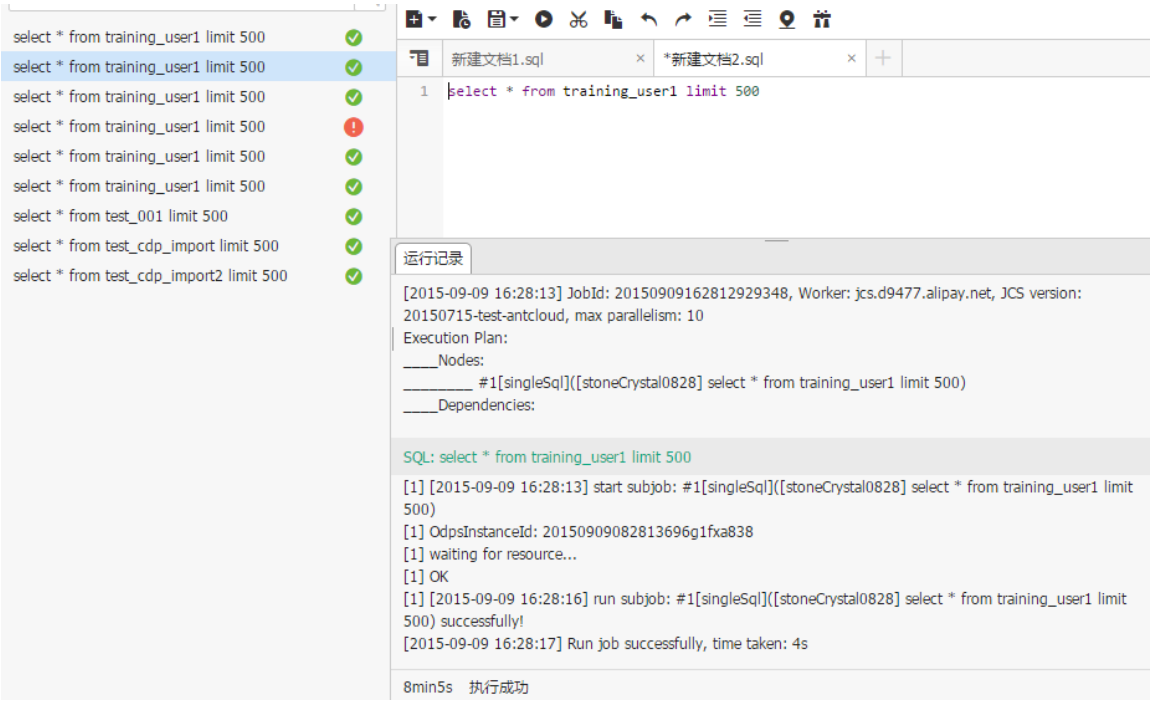


3、打开运行记录

在浏览窗口中双击目标运行记录或点击右键，在右键菜单中选择 **打开**。

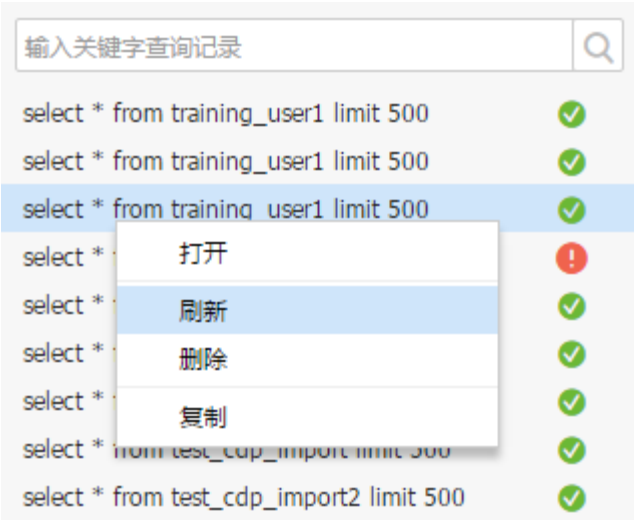


运行记录中的 SQL 语句将会在新的文件标签页中打开。语句运行时的详细信息将会展示在运行记录标签页中。



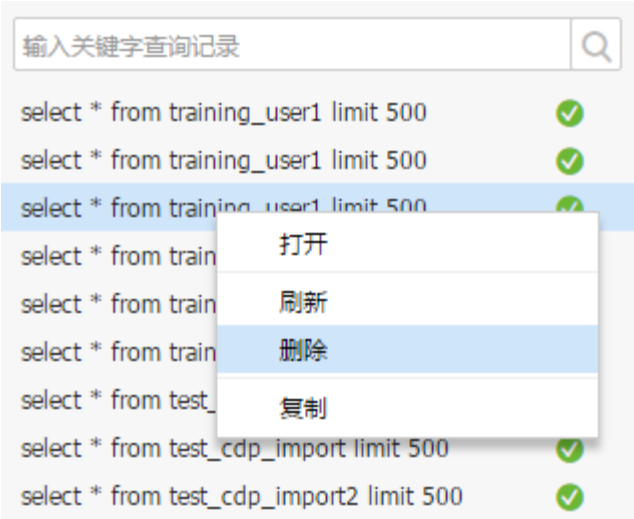
4、刷新运行记录

在浏览窗口中点击右键，在右键菜单中选择 **刷新**，将当前已执行完成，但尚未显示的运行记录展示在浏览窗口中。



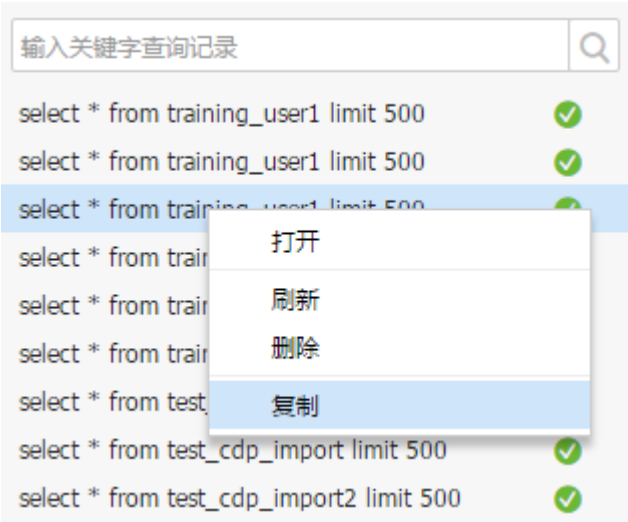
5、删除运行记录

在浏览窗口中选中目标运行记录，点击右键，在右键菜单中选择 **删除**。在弹出的对话框中选择 **确认** 或 **取消**。若点击 **确认**，被选中的运行记录将会被删除。



6、复制运行记录

在浏览窗口中选中目标运行记录，点击右键，在右键菜单中选择 **复制**。被选中的 SQL 语句将会被复制到粘贴板，可以粘贴到其他应用程序中。



复制成功后，工具栏将出现“复制成功！”的提示

八、管理设置

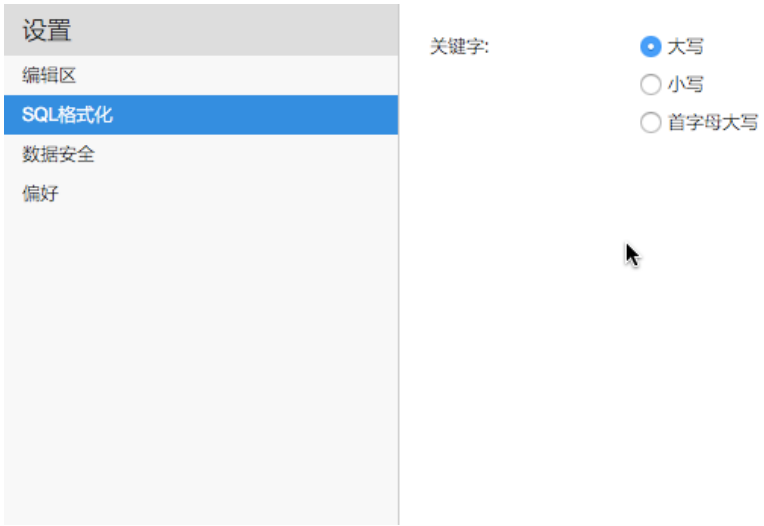
1、编辑区设置

进入采云间的管理面板，点击 **编辑区**，可以设置脚本的如下属性：

- 字体：设置显示的字体
- 字号：设置默认显示的字号
- 换行：是否自动换行
- 自动提示：是否启用自动提示，启用后，可以做脚本提示
- 编辑风格：系统预置了 18 种编辑风格，用户可以配置自己喜欢的风格

2、脚本格式化

设置格式化时，关键字的格式规则



3、个人偏好

进入采云间的管理面板，点击 **偏好**。

设置
编辑区
SQL格式化
数据安全
偏好
保留Tab窗口的历史结果集：
是
运行时提示变量替换：
是