

数仓建模介绍

目录

- 1. 概述
- 2. 界面概览
- 3. 数仓建模
 - 3.1 创建业务域
 - 3.2 操作数据仓库
 - 3.2.1 创建维度文件夹
 - 3.2.2 创建维度表
 - 3.2.3 创建日期维度表
 - 3.2.4 维度表字段设置
 - 3.2.5 创建业务过程文件夹
 - 3.2.6 创建事实表
 - 3.2.7 事实表字段设置
 - 3.3 操作分析指标
 - 3.3.1 分析指标维度表
 - 3.3.2 创建原子指标
 - 3.3.3 创建派生指标
 - 3.3.4 创建业务限定
 - 3.3.5 创建宽表

1. 概述

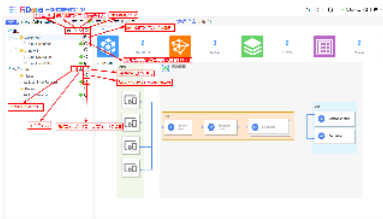
数仓建模功能提供可视化的界面，方便用户通过建模的方法更好的组织、存储数据，以便在性能、成本、效率和数据质量之间找到最佳平衡点

2. 界面概览

入口



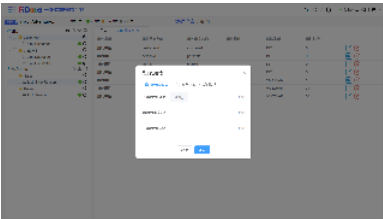
数据仓库



维度文件夹



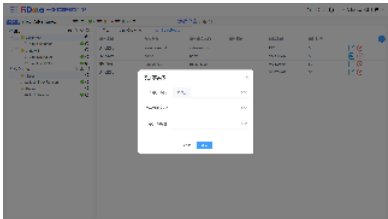
维度表



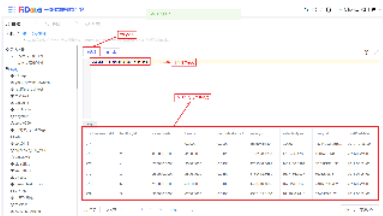
业务过程文件夹



事实表



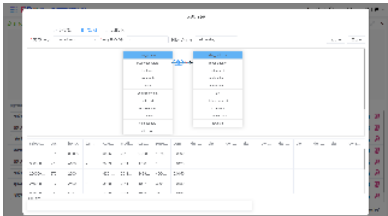
维度字段设置



维度字段转换



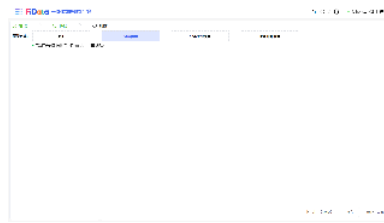
添加维度键



退化维度&度量字段



覆盖方式



保存&发布



分析指标



原子指标



派生指标



业务限定



时间函数



宽表



3. 数仓建模

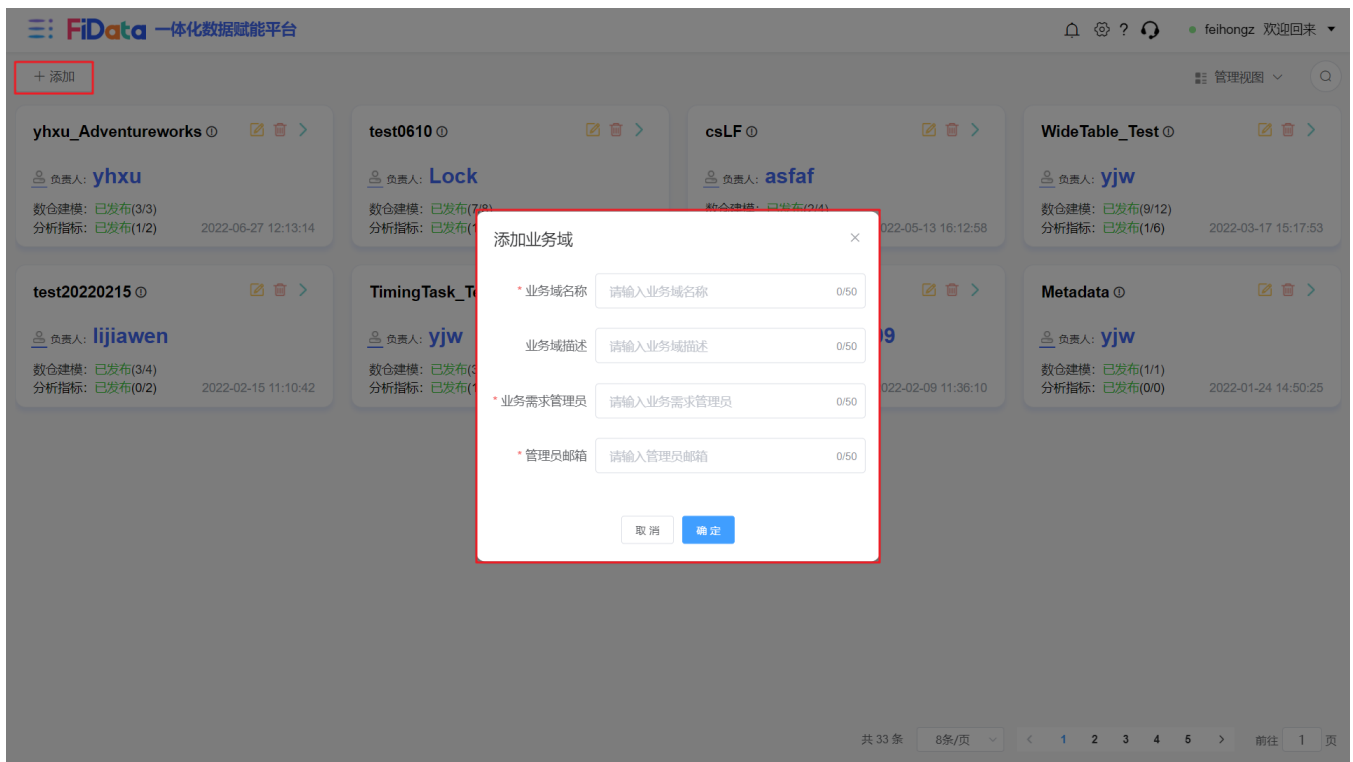
3.1 创建业务域

创建业务域： 点击添加，输入业务域基本信息，即可创建业务域

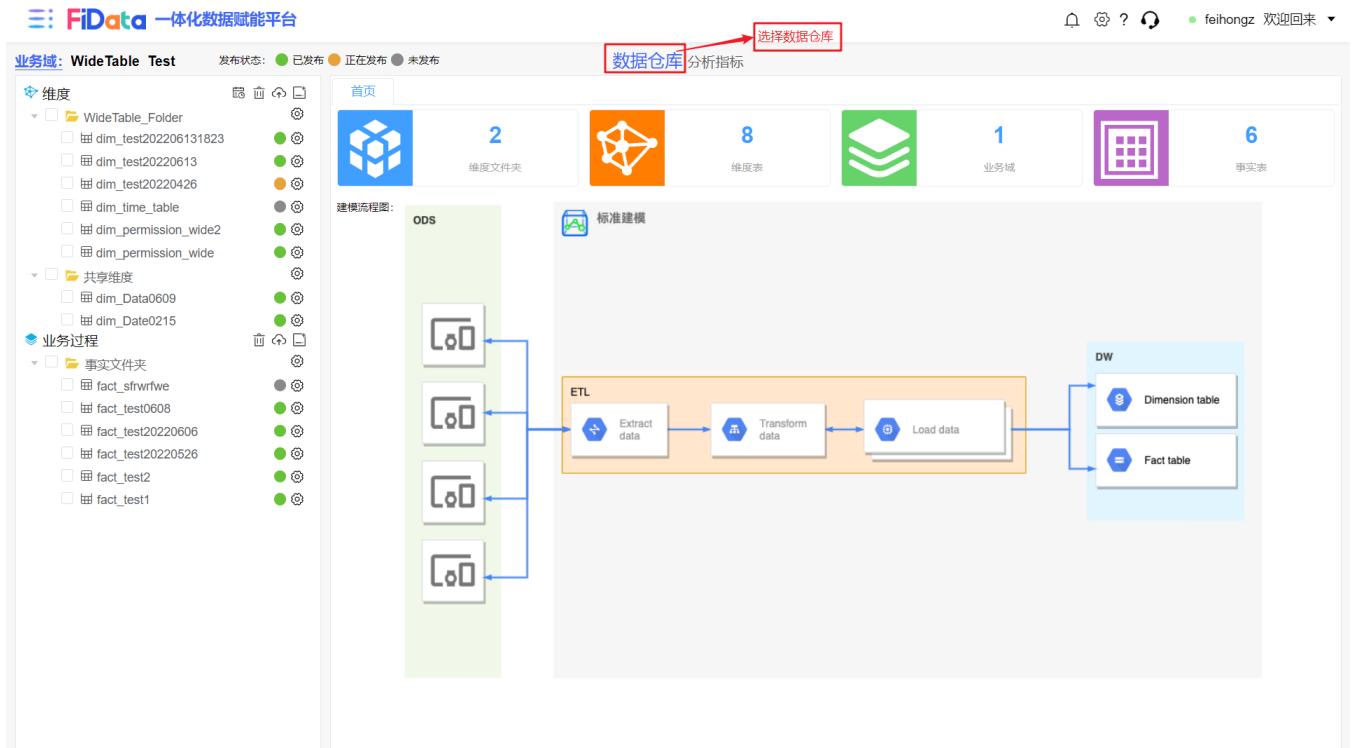
业务域名称： 中英文名称皆可

业务需求负责人： 该系统的IT负责人

管理员邮箱： 该系统的IT负责人邮箱



3.2 操作数据仓库

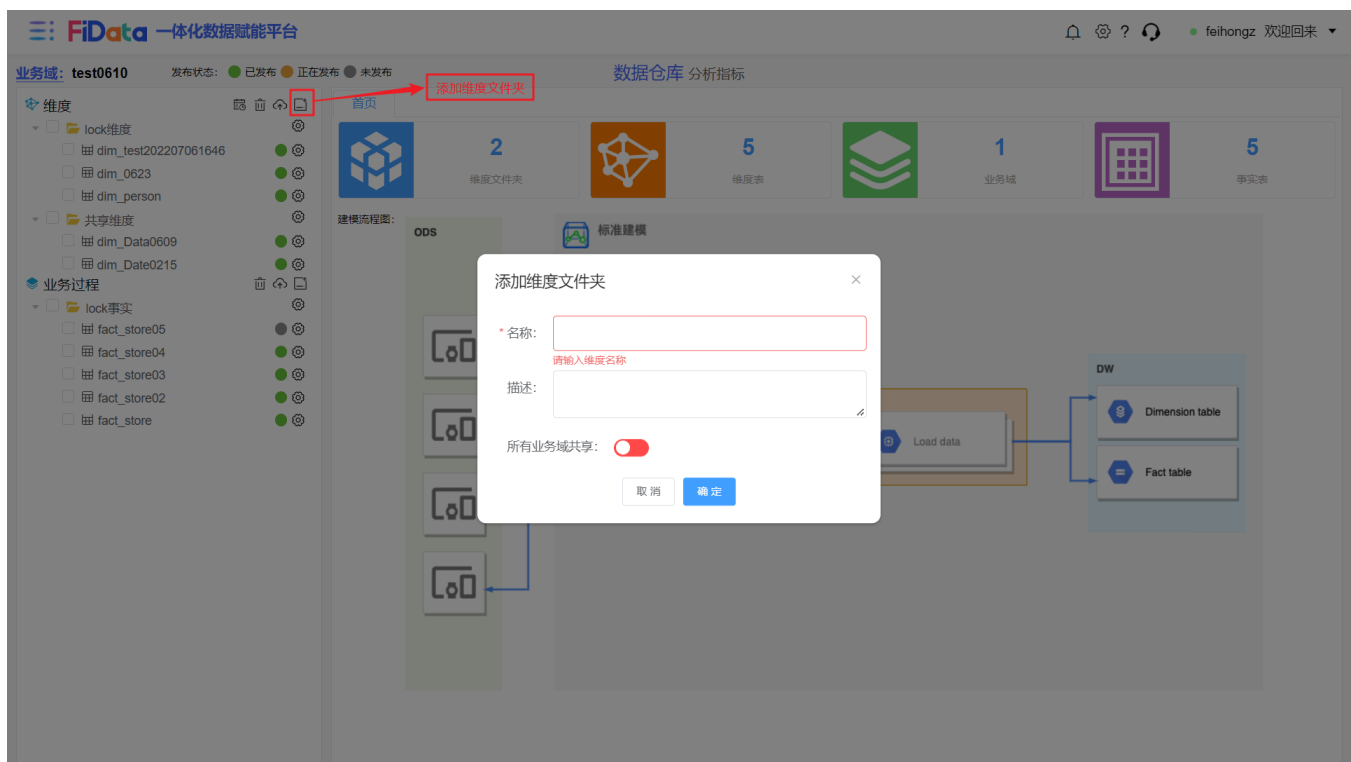


3.2.1 创建维度文件夹

创建维度文件夹：点击添加，输入基本信息，即可创建维度文件夹

名称：中英文名称皆可

所有业务域共享：开启业务域共享，意味着其他业务域也可以使用共享维度文件夹及其下面的所有维度表。反之，只能当前业务域允许使用

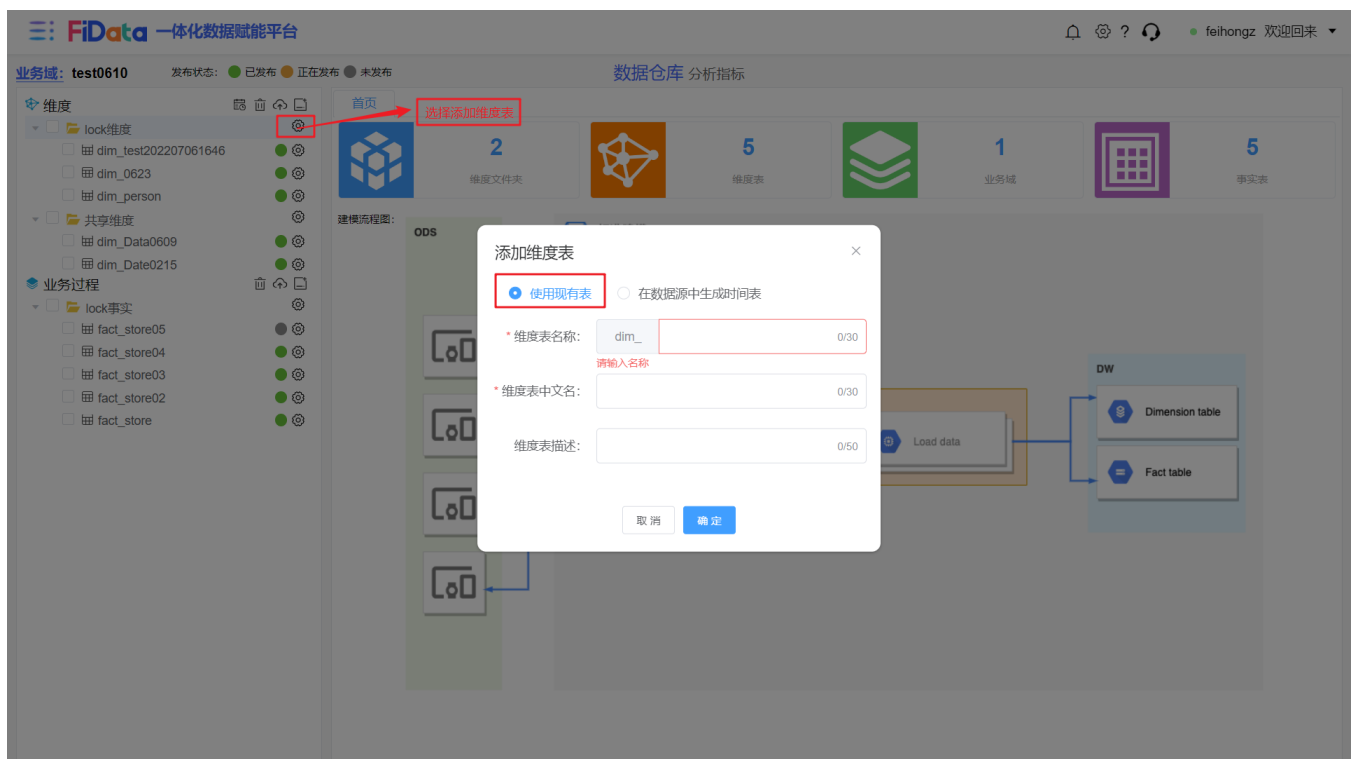


3.2.2 创建维度表

创建维度表： 点击添加，选择“使用现有表”，输入维度表基本信息，即可创建维度表

维度表名称：一般是英文名称

维度表中文名：一般是中文名称



3.2.3 创建日期维度表

创建日期维度表： 点击添加，选择“在数据源中生成时间表”，输入维度表基本信息，即可创建日期维度表

时间范围：用于生成时间范围的维度表

维度表名称：一般是英文名称

维度表中文名：一般是中文名称



3.2.4 维度表字段设置

1.选择新增的维度表右边的设置logo → 2.选择字段设置；

FiData 一体化数据赋能平台

业务域: test0610 发布状态: 已发布 正在发布 未发布 数据仓库 分析指标

维度

lock维度

dim_test202207061646

编辑

删除

字段设置

共享维度

dim_Data0609

dim_Date0215

业务过程

lock事实

fact_store05

fact_store04

fact_store03

fact_store02

fact_store

属性类型

属性英文名称

属性中文名称

属性描述

数据类型

属性长度

操作

维度属性

businessentityid

businessentityid

INT

0

操作

维度属性

personypecc

personypecc

VARCHAR

2

操作

维度属性

namestylev

namestylev

VARCHAR

20

操作

维度属性

titleq

titleq

VARCHAR

8

操作

维度属性

firstnameeq

firstnameeq

VARCHAR

50

操作

维度属性

middlenameq

middlenameq

VARCHAR

50

操作

维度属性

lastnameq

lastnameq

VARCHAR

50

操作

维度属性

suffixq

suffixq

VARCHAR

10

操作

维度属性

emailpromotionq

emailpromotionq

INT

0

操作

维度属性

additionalcontactinfoq

additionalcontactinfoq

VARCHAR

5000

操作

维度属性

demographicsq

demographicsq

VARCHAR

5000

操作

维度属性

rowguidq

rowguidq

VARCHAR

36

操作

维度属性

modifieddateq

modifieddateq

VARCHAR

50

操作

维度属性

personkey

关联key

VARCHAR

200

操作

3.编写业务SQL → 4.进入字段映射

FiData 一体化数据赋能平台

抽取 转换 加载

来源类型: 自定义脚本

系统变量

上次同步开始时间

上次同步结束时间

数据库

File-app

yhxu_Adventureworks

realTime_api_test

test0524

test0516

cs_lfasdasd

dsgafafd

test_0509

非实时示例(MDM接口)

we

test0418

sqlserver_DW_MDS

testviews

工时测试

实时测试API

gy-test

TimingTask_Test2

GYTest

非实时测试

test1111

Adventureworks

执行SQL, 预览数据

编写业务SQL

SQL查询出的数据

结果

business...	person...	namest...	titleq	firstna...	middlen...	lastnam...	suffixq	emailpr...	addition...	demogr...	rowguidq	modifie...
11385	IN	false		Jasmine	T	Williams	1			<Individ...	C5FAAB...	2014-02...
11386	IN	false		Christy	C	Xie	1			<Individ...	78EBBD...	2014-02...
11387	IN	false		Jasmine	J	Jones	1			<Individ...	FCFA5E...	2013-07...
11388	IN	false		Jasmine		Brown	0			<Individ...	FD09FE...	2013-11...
11389	IN	false		Jasmine	M	Davis	2			<Individ...	74A22C...	2012-12...
11390	IN	false		Christy	R	Tang	0			<Individ...	1972DB...	2013-02...

共 509300 条 10条/页 1 2 3 4 5 6 ... 50930 前往 1 页

下一步: 字段映射

5.操作字段 → 6.进入覆盖方式

数据类型:

- 整型:
 - 存放数值类型, 无小数类型
- 日期类型:
 - 存放日期类型数据。

- 精确数值型：
 - 存放小数类型数据。
 - 设置精度
- 字符串：
 - 存放字符串类型数据。
 - 设置长度

FiData 一体化数据赋能平台

feihongz 欢迎回来

抽取

转换

加载

源: SQL输入

businessentityid
territoryid
salesquota
bonus
commissionpct
salesytd
saleslastyear
rowguid
modifieddate

目标: dim_0623

businessentityid
territoryid
salesquota
bonus
commissionpct
salesytd
saleslastyear
rowguid
modifieddate

属性类型	源字段	目标中文名称	目标英文名称	目标字段描述	目标字段类型	目标字段长度	业务主键	操作
维度属性	businessentityid	businessentityid	businessentityid		整型	0	☒	编辑当前行字段的所有属性 删除当前行字段
维度属性	territoryid	territoryid	territoryid		整型	0	☐	
维度属性	salesquota	salesquota	salesquota		FLOAT	0	☐	
维度属性	bonus	bonus	bonus		FLOAT	0	☐	
维度属性	commissionpct	commissionpct	commissionpct		FLOAT	0	☐	
维度属性	salesytd	salesytd	salesytd		FLOAT	0	☐	
维度属性	saleslastyear	saleslastyear	saleslastyear		FLOAT	0	☐	
维度属性	rowguid	rowguid	rowguid		字符串型	36	☐	

添加字段

上一步: 来源对象

下一步: 覆盖方式

7.选择同步数据的覆盖方式 → 8.保存&发布维度表

FiData 一体化数据赋能平台

feihongz 欢迎回来

抽取

转换

加载

覆盖方式:

追加

全量覆盖

业务主键覆盖

业务时间覆盖

每次同步时系统将清空原来的表, 并重新插入

选择覆盖方式

上一步: 字段映射

保存

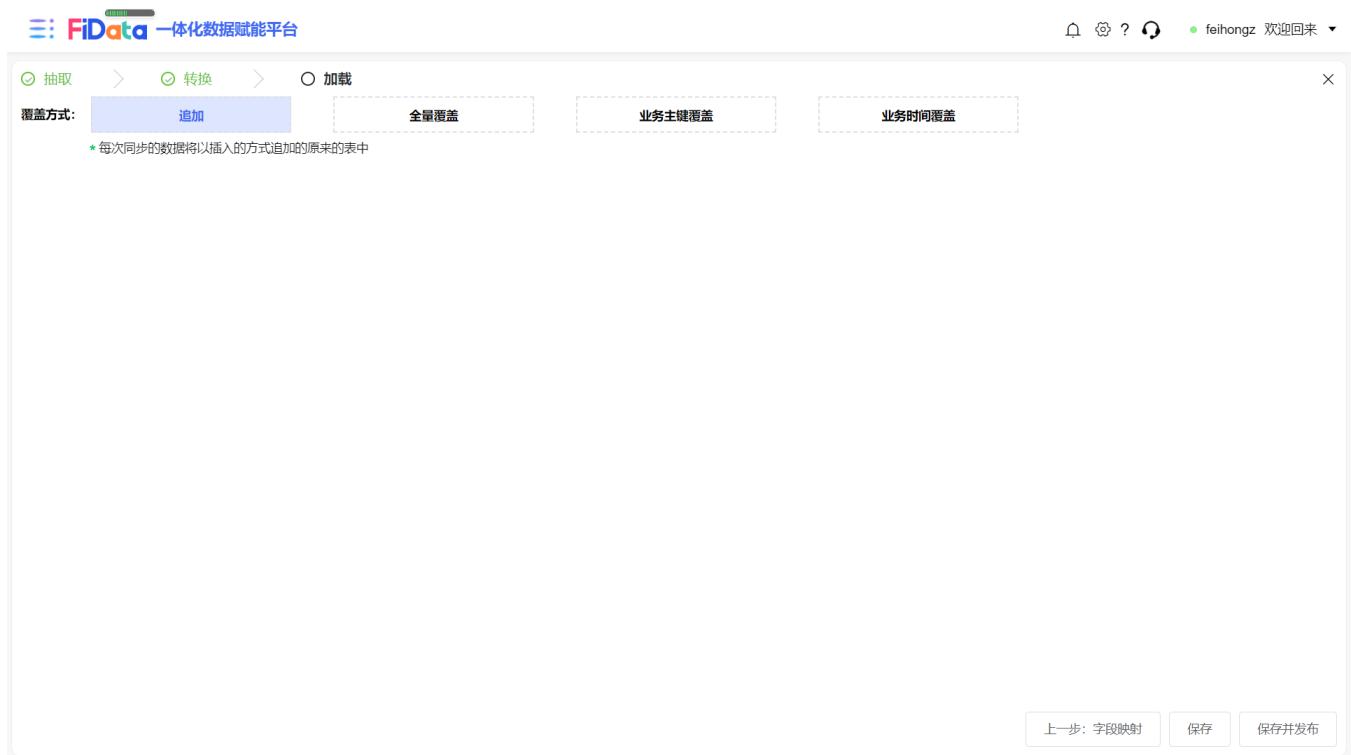
保存并发布

9.填写本次发布描述 → 10.发布



同步方式:

- 追加: 每次同步的数据将以插入的方式追加的原来的表中
- 全量: 每次同步时系统将清空原来的表, 并重新插入
- 业务主键覆盖: 系统将根据当前设定的业务主键进行数据覆盖
- 业务时间覆盖: 系统将根据以下选择的条件先从表中删除满足条件的数据, 然后重新从插入满足条件的数据



抽取 > 转换 > 加载

覆盖方式:

追加

全量覆盖

业务主键覆盖

业务时间覆盖

* 每次同步时系统将清空原来的表，并重新插入

上一步: 字段映射

保存

保存并发布

抽取 > 转换 > 加载

覆盖方式:

追加

全量覆盖

业务主键覆盖

业务时间覆盖

* 系统将根据当前设置的业务主键: 【businessentityid】进行数据覆盖

上一步: 字段映射

保存

保存并发布

FiData 一体化数据赋能平台

通知

帮助

头像

feihongz 欢迎回来

抽取

转换

加载

覆盖方式:

追加

全量覆盖

业务主键覆盖

业务时间覆盖

* 系统将根据以下选择的条件先从表中删除满足条件的数据，然后重新从插入满足条件的数据

覆盖

请选择

请选择

T +

请输入内容

请选择

的数据

假如在

每年

每月

每天

覆盖

请选择

请选择

T +

请输入内容

请选择

的数据

否则

覆盖

请选择

请选择

T +

请输入内容

请选择

的数据

* 备注: T代表当前时间

上一步: 字段映射

保存

保存并发布

3.2.5 创建业务过程文件夹

点击添加，输入基本信息，即可创建业务过程文件夹

名称：中英文名称皆可

FiData 一体化数据赋能平台

通知

帮助

头像

feihongz 欢迎回来

业务域: test0610

发布状态:

已发布

正在发布

未发布

数据仓库 分析指标

维度

lock维度

dim_test202207061646

dim_0623

dim_person

共享维度

dim_Data0609

dim_Date0215

业务过程

lock事实

fact_store05

fact_store04

fact_store03

fact_store02

fact_store

属性类型

属性英文名称

属性中文名称

属性描述

数据类型

属性长度

维度属性

businessentityid

businessentityid

INT

0

编辑

删除

维度属性

territoryid

territoryid

INT

0

编辑

删除

维度属性

salesquota

salesquota

FLOAT

0

编辑

删除

维度属性

bonus

bonus

FLOAT

0

编辑

删除

维度属性

FLOAT

0

编辑

删除

维度属性

FLOAT

0

编辑

删除

维度属性

FLOAT

0

编辑

删除

维度属性

VARCHAR

36

编辑

删除

维度属性

VARCHAR

50

编辑

删除

添加业务过程文件夹

* 名称:

描述:

取消

确定

3.2.6 创建事实表



创建事实表：点击添加，输入事实表基本信息，即可创建事实表

事实表名：一般是英文名称

事实表中文名：一般是中文名称

3.2.7 事实表字段设置

1.选择新增的事实表右边的设置logo → 2.选择字段设置



3.编写业务SQL → 4.进入字段映射

FiData 一体化数据赋能平台

来源类型: 自定义脚本

系统变量

- 上次同步开始时间
- 上次同步结束时间

数据库

- File-app
- yhxu_Adventureworks
- realTime_api_test
- test0524
- test0516
- cs_ifasdasd
- dsgafafd
- test_0509
- 非实时示例(MDM接口)
- we
- test0418
- sqlserver_DW_MDS
- testviews
- 工时测试
- 实时测试API
- gy-test
- TimingTask_Test2
- GYTest
- 非实时测试
- test1111
- Adventureworks

执行SQL, 预览数据

3 编写业务SQL

SQL查询出的数据

businessentityid	name	salespersonid	demographics	rowguid	modifieddate
292	Next-Door Bike Store	279	<StoreSurvey xmlns="htt...	A22517E3-848D-4EBE-B...	2014-09-12 11:15:07.497
294	Professional Sales and S...	276	<StoreSurvey xmlns="htt...	B50CA50B-C601-4A13-B...	2014-09-12 11:15:07.497
296	Riders Company	277	<StoreSurvey xmlns="htt...	337C3688-1339-4E1A-A...	2014-09-12 11:15:07.497
298	The Bike Mechanics	275	<StoreSurvey xmlns="htt...	7B94F278-F0C8-4D16-B...	2014-09-12 11:15:07.497
300	Nationwide Supply	286	<StoreSurvey xmlns="htt...	C3FC9705-ABC4-4F3A-9...	2014-09-12 11:15:07.497
302	Area Bike Accessories	281	<StoreSurvey xmlns="htt...	368BE6DD-30E5-49BB-9...	2014-09-12 11:15:07.497

共 701 条 10条/页 < 1 2 3 4 5 6 ... 71 > 前往 1 页

4 下一步: 字段映射

5.添加字段:

- (1)选择字段的属性类型, 其中退化维度和度量字段配置一样;
- (2)设置字段的其他属性
- (3)保存当前字段设置

其他操作同3.4维度表字段设置一致

抽取 > 转换 > 加载

源: SQL输入

- businessentityid
- name
- salespersonid
- demographics
- rowguid
- modifieddate

目标: fact_store05

- businessentityid
- name
- salespersonid
- demographics
- rowguid
- modifieddate

属性类型	源字段	目标中文名称	目标英文名称	目标字段描述	目标字段类型	目标字段长度	操作
退化维度	businessentityid	businessentityid	businessentityid		整型	0	
退化维度	name	name	name		字符串型	50	
退化维度	salespersonid	salespersonid	salespersonid		整型	0	
退化维度	demographics	demographics	demographics		字符串型	5000	
退化维度	rowguid	rowguid	rowguid		字符串型	36	
退化维度	modifieddate	modifieddate	modifieddate		字符串型	50	

添加字段
上一步: 来源对象
下一步: 覆盖方式

删除当前行字段

编辑当前行字段的所有属性

此按钮用于自定义字段

5

下一步: 覆盖方式

FiData 一体化数据赋能平台

抽取 > 转换 > 加载

源: SQL输入

- businessentityid
- name
- salespersonid
- demographics
- rowguid
- modifieddate

目标: fact_store05

- businessentityid
- name
- salespersonid
- demographics
- rowguid
- modifieddate

添加字段

5.1

☒ 退化维度
 ☐ 维度键
 ☐ 度量字段

5.2

* 源字段
请选择字段
* 目标中文名称
目标英文名称
* 目标字段类型
字符串型
目标字段长度
200
目标字段描述

取消
确定
5.3

添加字段
上一步: 来源对象
下一步: 覆盖方式



3.3 操作分析指标



3.3.1 分析指标维度表



3.3.2 创建原子指标

1.添加事实表的原子指标 → 2.填写原子指标的基本信息 → 3.保存基本信息



4.选择配置公式模式 → 5.选择具体的度量字段，配置聚合类型 → 6.保存所有配置

聚合类型 说明：

1. SUM：求和，多行的 Value 进行累加

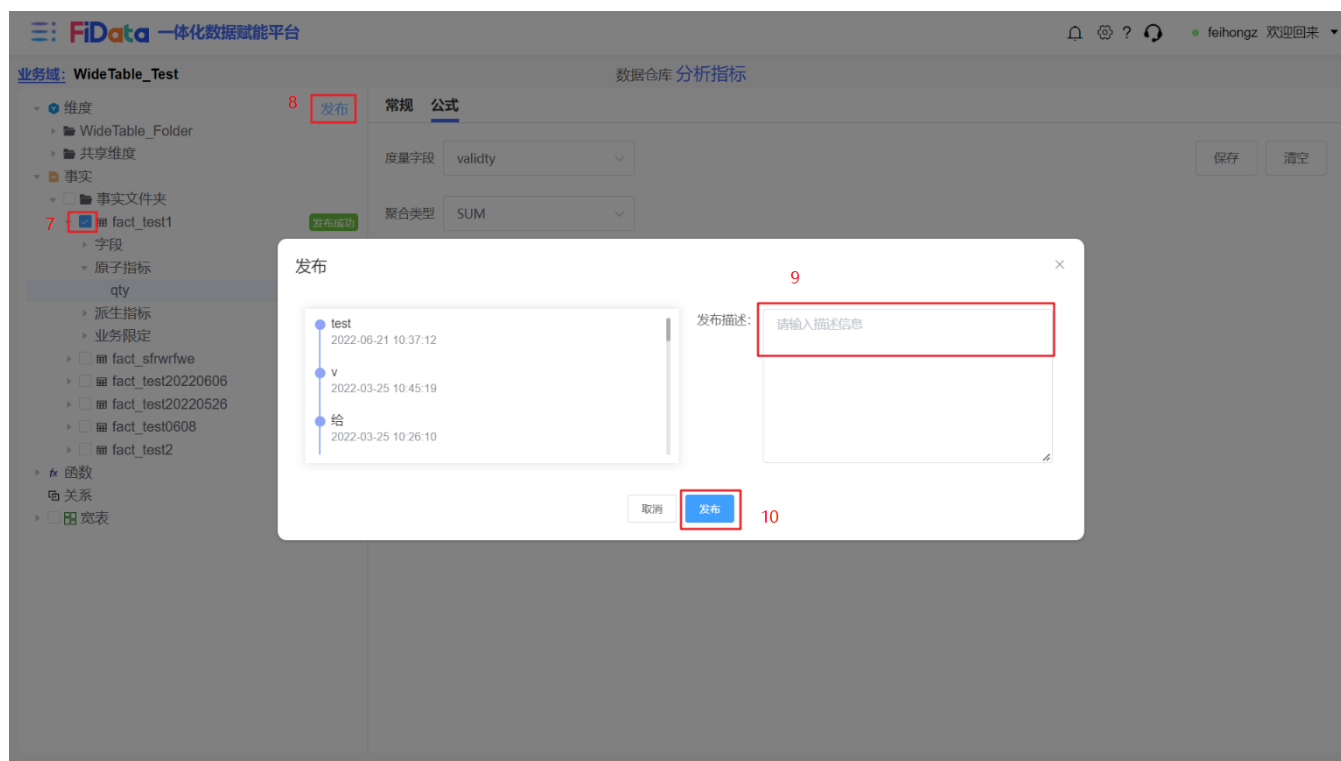
2. REPLACE：替代，下一批数据中的 Value 会替换之前导入过的行中的 Value

3. MAX：保留最大值

4. MIN：保留最小值

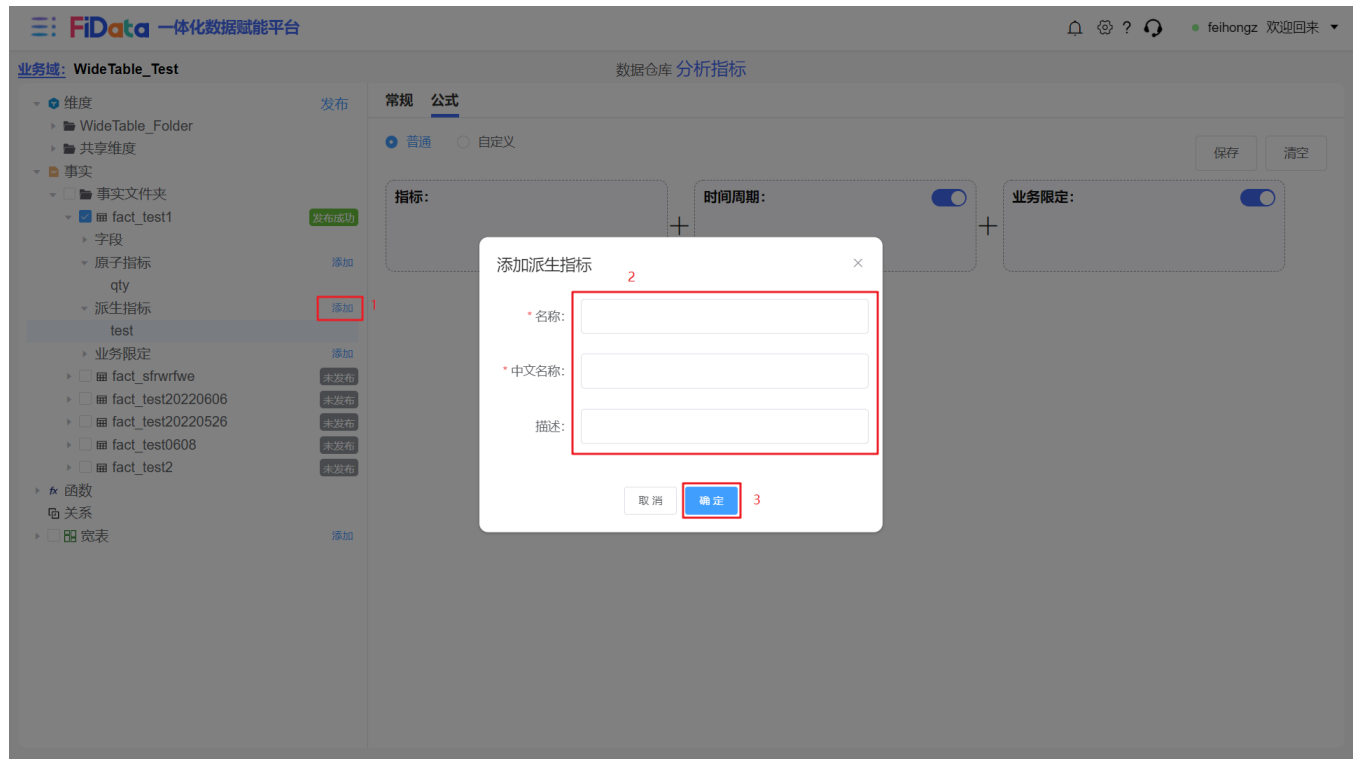


7.选择需要发布的事实表 → 8.选择发布按钮 → 9.填写本次发布的描述 → 10.发布



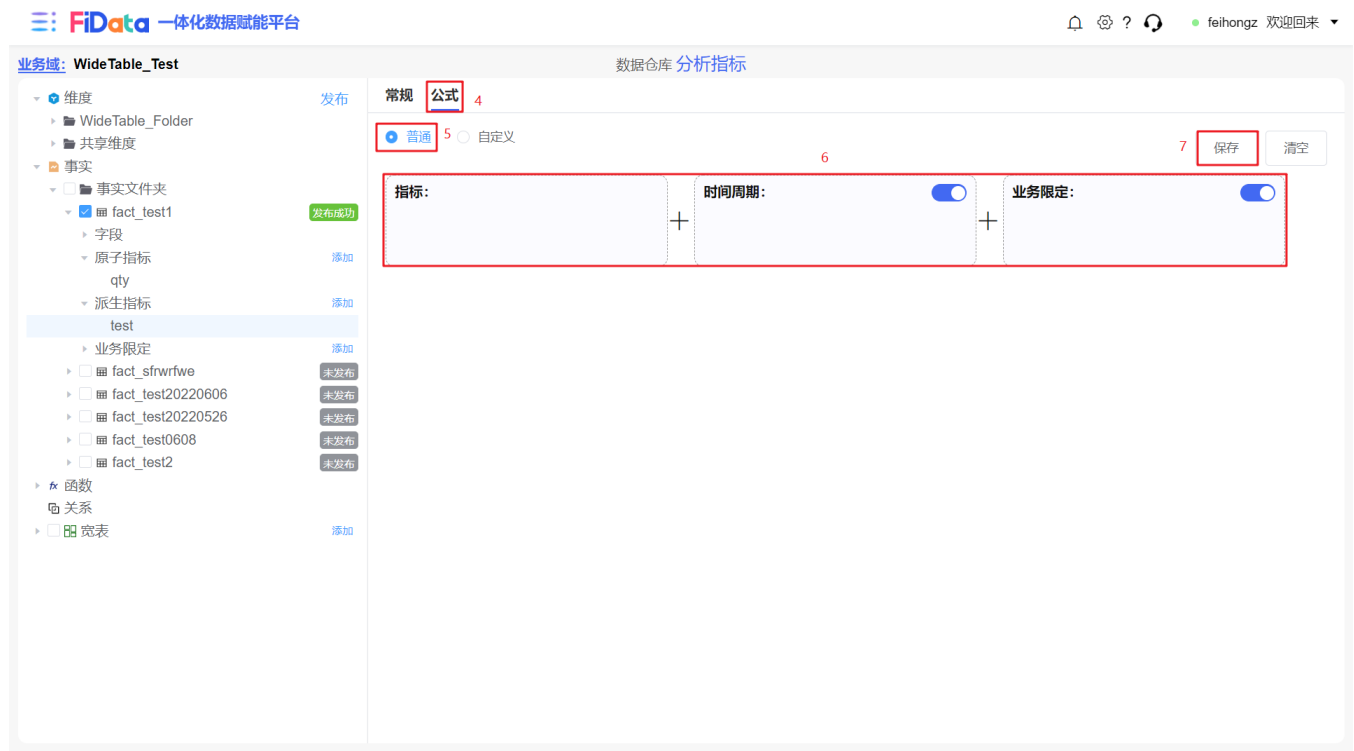
3.3.3 创建派生指标

1.添加事实表的派生指标 → 2.填写派生指标的基本信息 → 3.保存基本信息



4.选择配置公式模式 → 5.选择普通公式or自定义公式 → 6.配置公式

其他发布操作同3.3.2一样



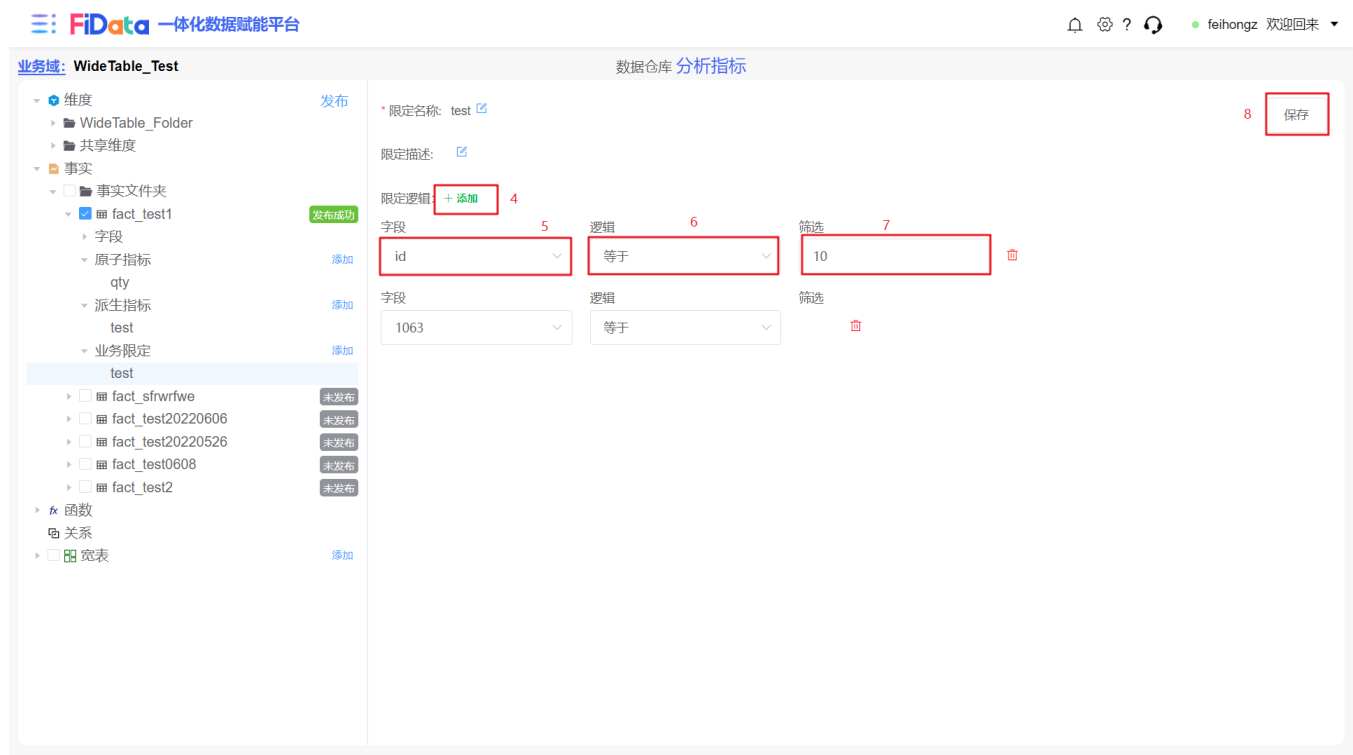


3.3.4 创建业务限定

1.添加事实表的业务限定 → 2.填写业务限定的基本信息 → 3.保存基本信息；



4.添加限定逻辑 → 5.下拉选择需要限定的字段 → 6.选择字段的限定逻辑：等于、大于、小于、包含 → 7.选择字段的筛选条件 → 8.保存所有配置
其他发布操作同3.3.2一样。



3.3.5 创建宽表

1. 添加宽表

2. 填写宽表名称，一般是英文名称

3. 从左边维度or事实中拖拽多张表到中间的编辑框

4. 设置连接关系：内连接、左连接、右连接、全连接、联合

5. 执行多表的连接关系

6. 保存所有配置

其他发布操作同3.3.2一样

业务域: WideTable_Test

数据仓库 分析指标

维度

- WideTable_Folder
 - dim_test202206131823 未发布
 - dim_test20220613 未发布
 - dim_test20220426 未发布
 - dim_time_table 3 未发布
 - dim_permission_wide2 未发布
 - dim_permission_wide 未发布
- 共享维度

事实

- 事实文件夹
 - fact_test2 未发布
 - fact_sfrwrfwe 未发布
 - fact_test0608 未发布
 - fact_test20220526 未发布
 - fact_test1 发布成功
 - fact_test20220606 未发布

函数

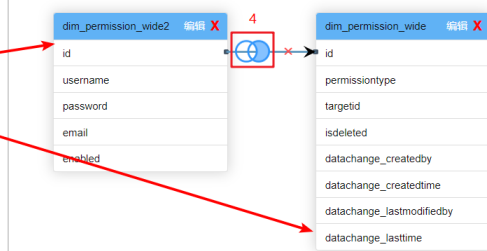
- 时间周期
 - MTD
 - QTD
 - YTD
 - 上一期
 - 去年同期

关系

- 宽表
 - dws_wideTableTest1 未发布
 - dws_wideTableTest3 未发布
 - dws_wideTableTest4 未发布

1 添加

请从左侧拖入维度表/事实表到此区域:



预览数据

id	userna...	passw...	email	enabled	dim_p...	permis...	targetid	isdeleted	datach...	datach...	datach...	datach...
1	apollo	\$2a\$1...	apollo...	1	1	Create...	Sampl...	true		16286...	apollo	16287...
2	yhxu	\$2a\$1...	yhxu...	1	2	Create...	Sampl...	true		16286...	apollo	16287...
					3	Assign...	Sampl...	true		16286...	apollo	16287...
					4	Modify...	Sampl...	true		16286...	apollo	16287...
					5	Relea...	Sampl...	true		16286...	apollo	16287...
					6	Create...	Syste...	false	apollo	16287...	apollo	16287...

2 wideTableTest5 14/30 5 6 执行 保存