Informatica TDM 961

使用手册

V1.0

修订记录		
2014.12	创建	任梓珩
2015.1	修订/配图	贾珂
2015.1	添加概述并更新目录	任梓珩

目录

1. 概述.....4

2. 环境描述.....4

3. 后台级管理.....4

3.1. 平台访问入口.....5

3.2. 平台初始配置.....5

3.2.1. 配置平台级的数据域敏感级别.....5

3.2.2. 配置数据连接.....6

3.2.3. 配置字典表.....8

3.3. 平台业务操作.....10

3.3.1. 创建脱敏项目.....11

3.3.2. 导入数据源.....13

3.3.3. 添加安全策略.....15

3.3.4. 敏感对象发现.....16

3.3.5. 确认敏感发现的结果.....21

3.3.6. 脱敏规则的绑定.....22

3.3.7. 脱敏作业的创建.....24

3.3.8. 脱敏作业的初始化与执行.....25

4. 平台设计与扩展.....31

4.1. 创建脱敏规则.....31

4.2. 创建数据域.....33

4.3. 创建安全策略.....37

5. 平台安全管理.....40

6. 风险规范管理.....41

7. 数据申请规范.....43

8. 集团数据安全需求.....46

9. 数据访问安全制度.....49

9.1. 访问隔离.....50

9.2. 脱敏处理.....51

1.概述

本文以某农商银行真实实施环境为依据编写 ,其中的环境描述及所有相关截图均以实际情况为准。管理及操作步骤描述则可供通用参考

2.环境描述

PWC&TDM 服务器配置

主机 ip 地址 192.168.169.156

主机名：infatdm

用户名：sjtm 密码：sjtm

应用安装路径：/home/sjtm/Informatica/9.6.1

应用服务启停脚本位置：

/home/sjtm/Informatica/9.6.1/server/tomcat/bin/infaservice.sh

（启动： ./infaservice.sh startup 停止： ./infaservice.sh shutdown）

知识库数据库名：infarep 数据库用户：sjtminst 密码：sjtminst

3.后台级管理

TDM 的后台级别管理通过基于网页的 Informatica Administration Console（以下简称 AdminConsole）来进行，具体访问地址为 https://infatdm:8443

平台业务使用

3.1. 平台访问入口

Informatica TDM 平台(以下简称 Workbench)的主要操作入 http://infatdm:6605/

也可在AdminConsole中选中左侧服务导航栏中的"TDM_Service",并通过显示于右侧概览页中的"Service URL"进行访问

缺省情况下 Workbench 的访问口令为 PWC 的域管理用户口令，后续可以通过添加用户并限制特权实现真正意义上的分权管理，本文档描述中均以缺省管理用户为例

Administrator/Administrator

3.2. 平台初始配置

3.2.1. 配置平台级的数据域敏感级别

该敏感级别主要用于标识被纳入整个 TDM 平台管理中的敏感数据域的不同的安全级别。敏感级别可以在创建敏感数据域定义时与特定数据域绑定。目前系统中包含低，中，高三种不同的敏感对象可供直接使用，若需要添加，删除，修改敏感级别可以在 Workbench 中的"**Administrator 标签**"-->"**Preference 标签**"-->"**Data Domain Sensitive Level**".

注 1：后期频繁修改敏感级别定义，容易造成历史数据域的敏感定义发生偏离，因此建议在平台构建初期建立合理规范的敏感级别。

3.2.2. 配置数据连接

此处需要配置的链接主要用于访问源，目标以及字典表，可以通过 workbench 中"**Administrator 标签**"-->"**Connections 标签**"中右上方之"**Action**"-->"**New Connection**"进行添加

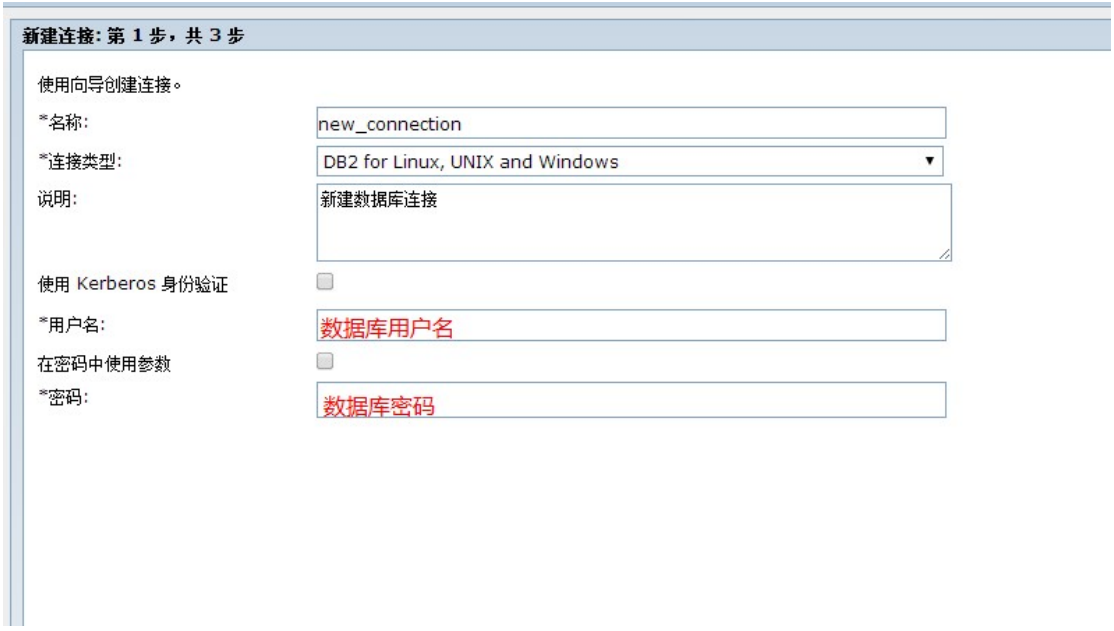
注 1：此处有系统创建的系统用 TDM_CONNECTION，请勿擅动

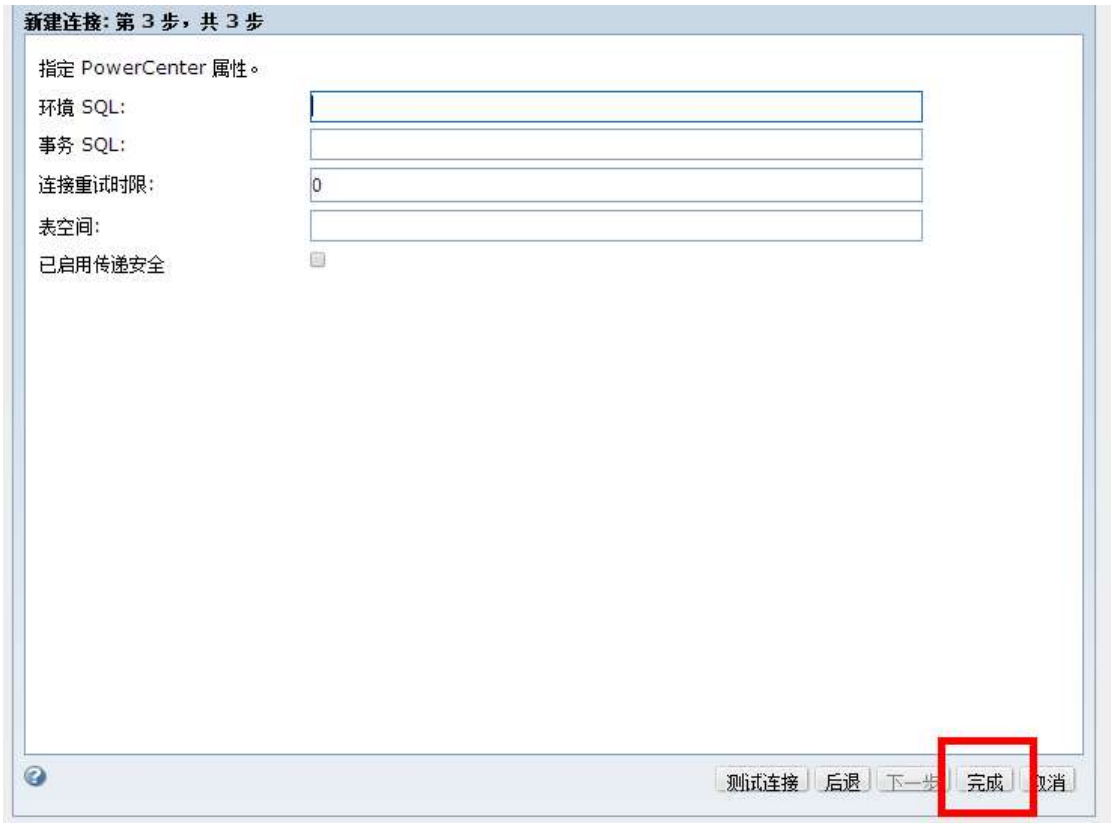
注 2：创建过程中涉及到关系型数据库的连接均包含"Metadata Properties"与"Data Access Porperites"前者采用 JDBC 连接（数据库或服务名为源端名称），主要用于元数据倒入，后者采用 Native Client（直接使用本地服务名），用于实际作业中的数据读取以及写入

注 3：创建连接时需要明确访问数据源所需要的字符集

注 4：字典表连接主要是用于创建字典表

注 5：也可通过"**Action**"-->"**Import Connection**"，将在 PWC Repository 中已有的连接导入到 TDM 平台中





3.2.3. 配置字典表

字典表主要用于脱敏规则中的"置换规则"，字典表可以保存于数据库中或者 Flat File，通过 workbench 中"**Administrator 标签**"-->"**Dictionaries 标签**"中右上方之

"**Action**"-->"**New Dictionary**"进行创建，若使用数据库保存字典表，则创建过程中需要用到此前在 2.2 中创建完毕的字典表，若使用 Flat File 则需要选择字典表文件，并明确字符集。字典表的结构通常包含两个字段，字段 A 用于"置换规则"中进行查找时的匹配值，字段 B 则包含了具体用于替换的拟真值，需要明确这两个字段的用途，用于后续的规则创建。当前平台下包含两个关系型字典表，分别为中文姓(surname)与中文名(givenname)的字典表

姓字典表中字段为：ID，SURNAME

名字典表中字段为：ID，GIVENNAME

注 1：原则上一个字典表可以包含多个匹配字段或替换字段，用于适配不同的置换需求

3.3. 平台业务操作

TDM 平台的使用遵循"定义"-->"发现"-->"绑定"-->"执行"四步骤的闭环实现方法

定义：主要用于创建可能使用的各种脱敏规则"Rule"，以及数据域"Data Domain"

发现：用于发现需要进行保护的具体的数据库中的字段，其发现依据基于此前环节中声明的"数据域"中的模板

绑定：对发现结果进行必要的人工干预，并将数据字段与预设的脱敏规则进行绑定

执行：创建具体的脱敏执行计划，并生成可执行的工作流"Workflow"

系统目前有部分可供使用的的脱敏规则与数据域定义。本章节以现有的资源对整个脱敏的业务操作环节进行说明。规则与数据域的定义方法将在次章进行说明

3.3.1. 创建脱敏项目

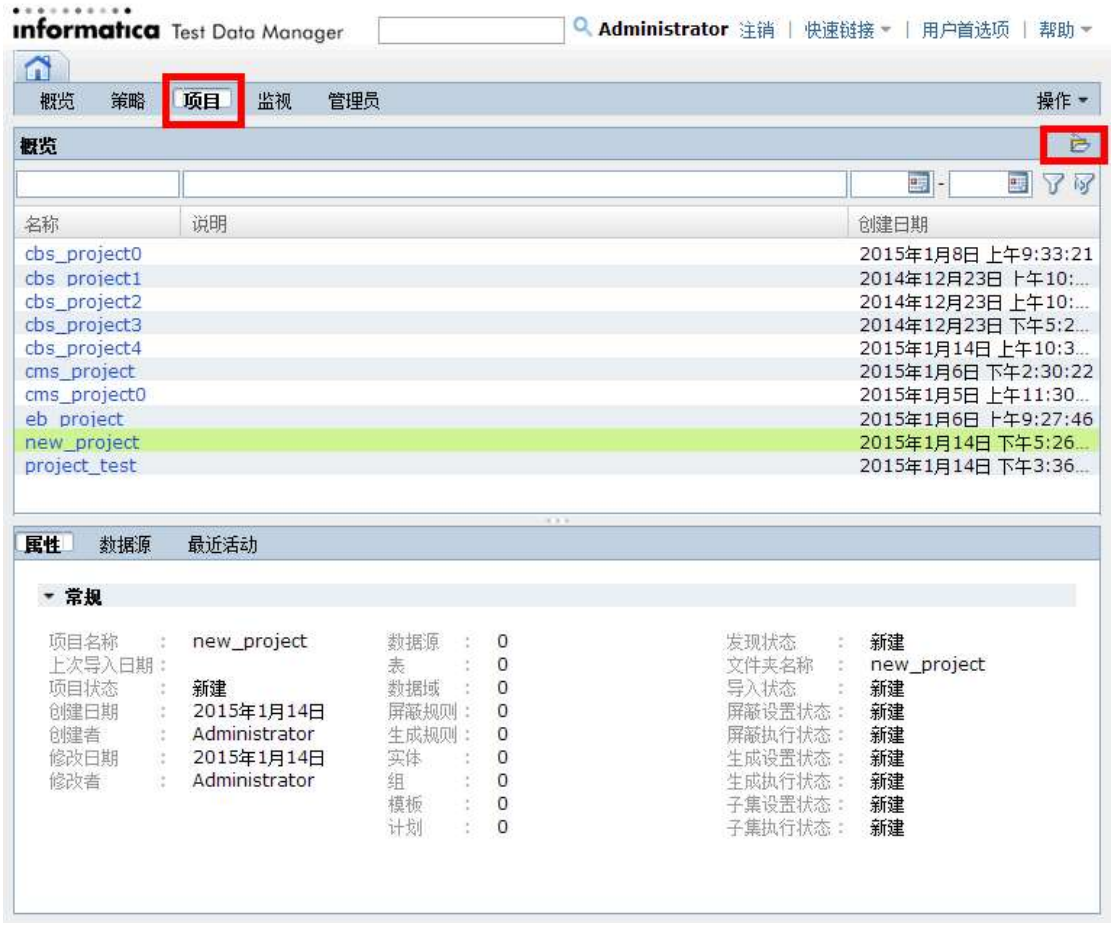
TDM 中的项目可以被理解为一个具体需要脱敏的业务系统对象，也可以被看作是针对数据的某一业务需求 ,比如特定业务模块的开发和测试等。项目的创建可以通过 workbench 中"**Project 标签**"-->"**Actions**"-->"**New**"来实现。

注 1：此处的项目最终会对应 PWC Repository 中的一个目录（Folder），因此 Folder 名称等同于项目名称

注 2：建议采用英文作为项目名称，也会以英语命名，也可项目名称为英文，手动设定英文 Folder 名称

注 3：该目录的所有者(owner)缺省为当前的用户，如无必要，不建议对其进行改动

注 4：也可通过"**Actions**"-->"**Export/Import**"功能将既有的项目成果导出成 XML 文件或将导出文件导入现有平台


informatica Test Data Manager
Administrator 注销 | 快速链接 | 用户首选项 | 帮助
概览 策略 项目 监视 管理员 操作
概览
名称 说明 创建日期
cbs_project0 2015年1月8日 上午9:33:21
cbs_project1 2014年12月23日 下午10:...
cbs_project2 2014年12月23日 上午10:...
cbs_project3 2014年12月23日 下午5:2...
cbs_project4 2015年1月14日 上午10:3...
cms_project 2015年1月6日 下午2:30:22
cms_project0 2015年1月5日 上午11:30...
eb_project 2015年1月6日 下午9:27:46
new_project 2015年1月14日 下午5:26...
project_test 2015年1月14日 下午3:36...
属性 数据源 最近活动
常规
项目名称 : new_project
上次导入日期 :
项目状态 : 新建
创建日期 : 2015年1月14日
创建者 : Administrator
修改日期 : 2015年1月14日
修改者 : Administrator
数据源 : 0
表 : 0
数据域 : 0
屏蔽规则 : 0
生成规则 : 0
实体 : 0
组 : 0
模板 : 0
计划 : 0
发现状态 : 新建
文件夹名称 : new_project
导入状态 : 新建
屏蔽设置状态 : 新建
屏蔽执行状态 : 新建
生成设置状态 : 新建
生成执行状态 : 新建
子集设置状态 : 新建
子集执行状态 : 新建

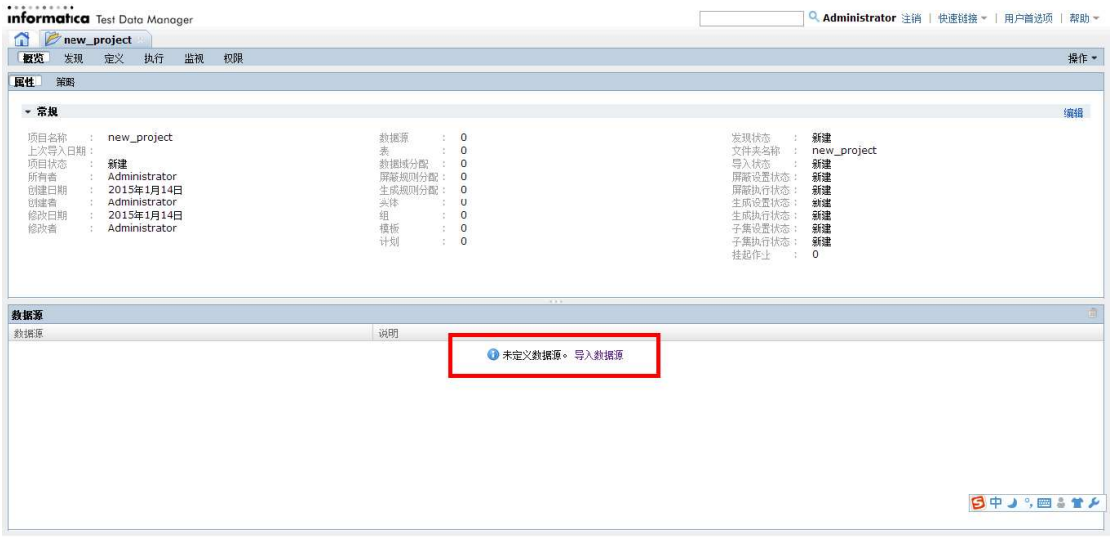
新建项目
定义项目属性。
*名称 new_project
说明
*PowerCenter 存储库 rep_centos
*文件夹 new_project 更改文件夹
*所有者 Administrator 更改所有者
确定 取消

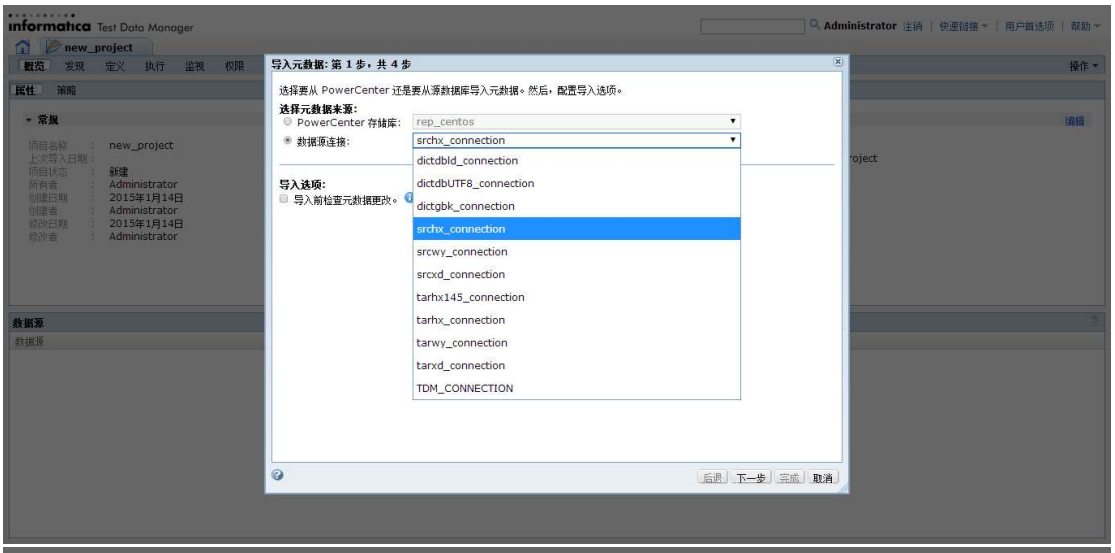
3.3.2. 导入数据源

TDM 可以通过导入功能将需要进行隐私保护的（脱敏）的业务系统的元定义导入平台中，点击打开此前创建的项目，在"Overview"标签页面下方的“**Data Source**”中通过"**Import Data Source**"功能进行

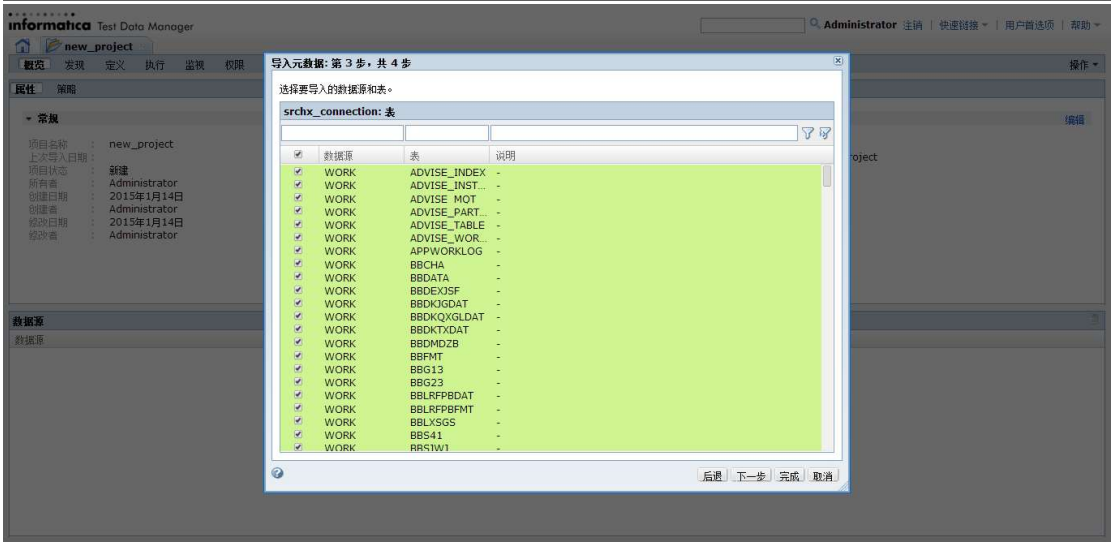
注 1：数据源导入功能仅将源数据库中的表及字段的结构定义导入，并不包括实际的数据

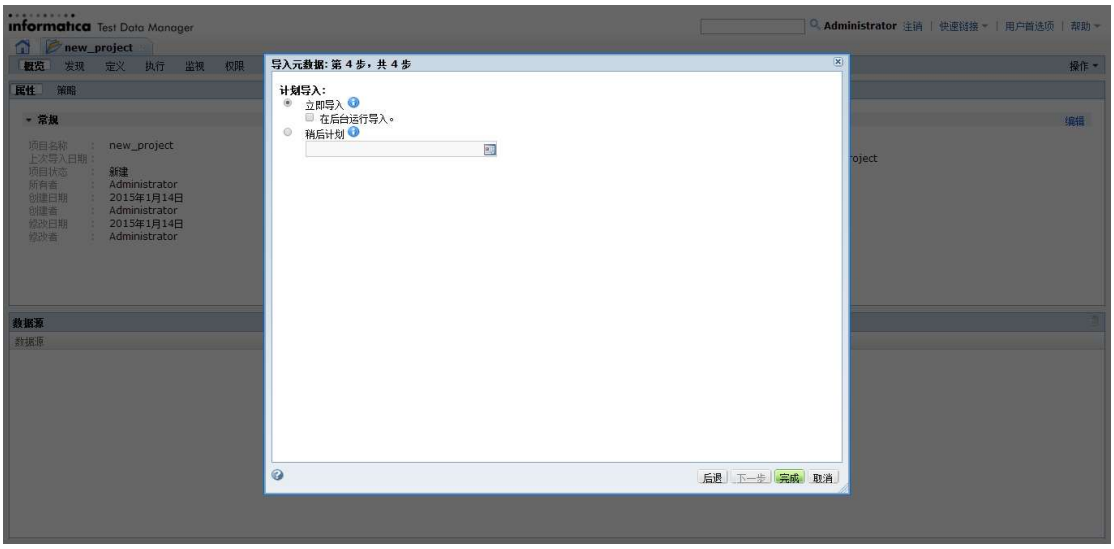
注 2：可以选择从 PWC Repository 中导入，也可通过 2.2 中创建的数据连接进行（通常采用此种方式）




informatica Test Data Manager
new_project
导入元数据: 第 1 步，共 4 步
选择要从 PowerCenter 还是要从源数据库导入元数据。然后，配置导入选项。
选择元数据来源:
PowerCenter 存储库:
rep_centos
数据源连接:
srchx_connection
dictdbld_connection
dictdbUTF8_connection
dictgbk_connection
srchx_connection
srcwy_connection
srcxd_connection
tarhx145_connection
tarhx_connection
tarwy_connection
tarxd_connection
TDM_CONNECTION
导入选项:
导入前检查元数据更改。
后退
下一步
完成
取消

informatica Test Data Manager
new_project
导入元数据: 第 2 步，共 4 步
选择要导入的架构。
连接: srchx_connection
数据源
ASN
ASNCAP
CICS
CQ
ICS
NULLID
SQLJ
SYSCAT
SYSFUN
SYSIBM
SYSIBMADM
SYSIBMINTERNAL
SYSIBMTS
SYSPROC
SYSPUBLIC
SYSSTAT
SYSTOOLS
WORK
WORKDB
WORKINST
WORKVIEW
后退
下一步
完成
取消

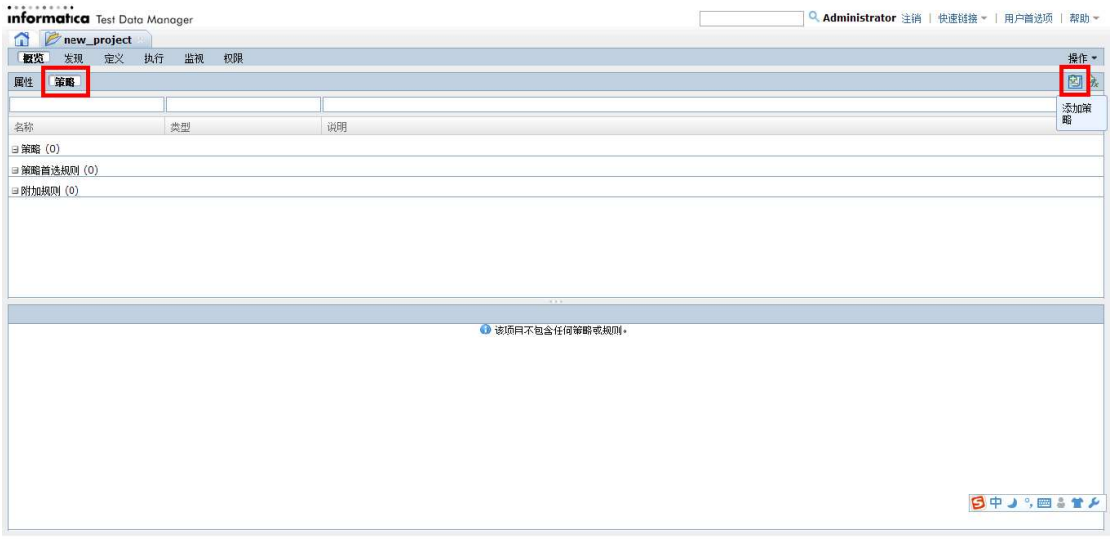

informatica Test Data Manager
new_project
导入元数据: 第 3 步，共 4 步
选择要导入的数据源和表。
srchx_connection: 表
数据源
表
说明
WORK ADVISE_INDEX -
WORK ADVISE_INST... -
WORK ADVISE_MQT -
WORK ADVISE_PART... -
WORK ADVISE_TABLE -
WORK ADVISE_WOR... -
WORK APPWORKLOG -
WORK BBCHA -
WORK BBDATA -
WORK BBDEXJSF -
WORK BBDKJGDAT -
WORK BBDKQXGLDAT -
WORK BBDKTXDAT -
WORK BBDMDZB -
WORK BBFMT -
WORK BBG13 -
WORK BBG23 -
WORK BBLRFPBDAT -
WORK BBLRFPBFMT -
WORK BBLXSGS -
WORK BBS41 -
WORK BBSIW1 -
后退
下一步
完成
取消

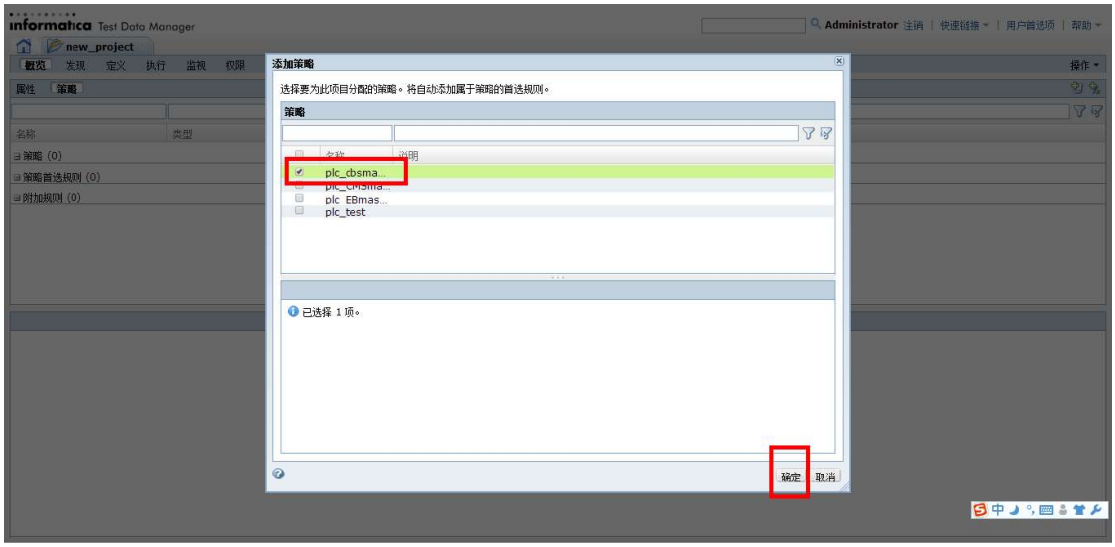


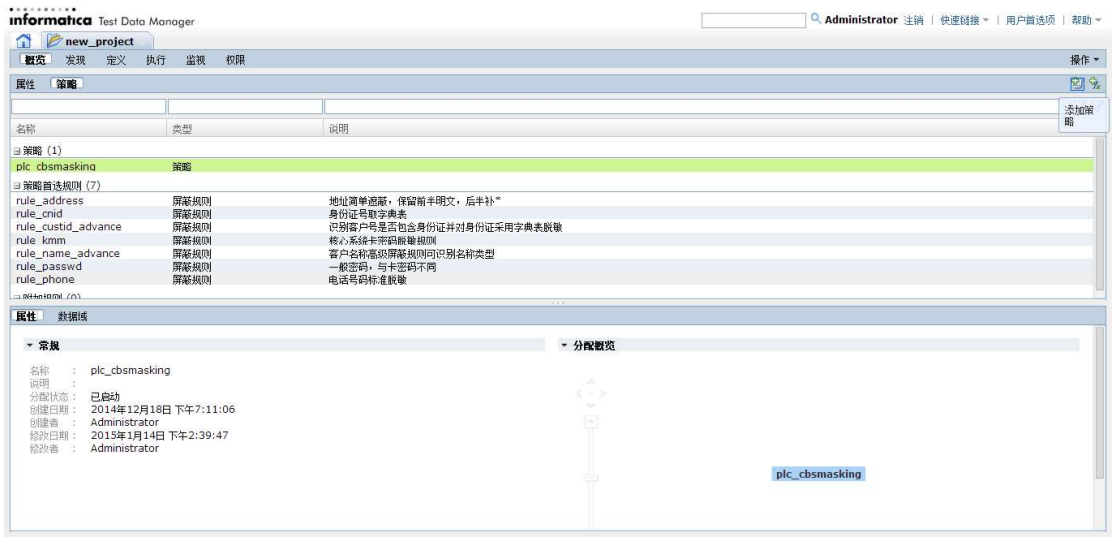
3.3.3. 添加安全策略

安全策略是 TDM 平台中的一个集合类概念，包含了一个或多个数据域以及其所属的脱敏规则。通过"**Overview 标签**"-->"**Policies 标签**"-->'**Actions**'-->"**Add Policies**"进行添加。

注 1：也可通过'**Actions**'-->"**Add Additional Rules**"添加单一的附加脱敏规则







3.3.4. 敏感对象发现

TDM 平台可以通过基于业务模型创建的数据域模板定义对数据库中的表进行扫描，从而返回可能的敏感字段。通过项目中的"**Discover 标签**"-->"**Profile 标签**"-->'**Actions**'-->"**New Profile**"创建发现作业

注 1：发现作业除了可以用于发现数据域(Data Domain)外，还可以发现表的主键(PK)以及表间逻辑关系(Entity)

注 2：对于脱敏而言，主要的发现作业对象是 Data Domain,PK 与 Entity 用于子集创建

注 3：基于 Data Domain 的发现又分为元数据发现与数据发现，分别依赖于 Data Domain 中 MetaData Pattern 与 Data Pattern 定义

注 4：发现对象可以是此前添加进来的全部表或部分表

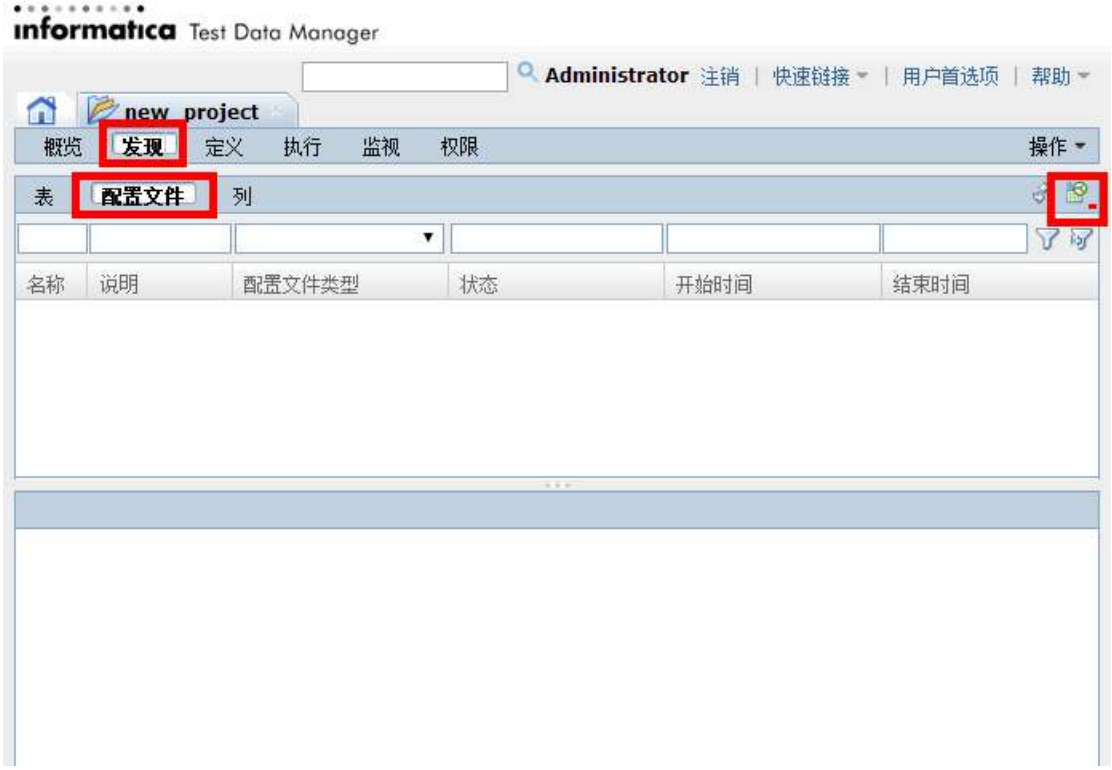
注 5：建议优先使用基于字段名称的元数据发现，基于数据本身的发现会根据数据库的规模产生较大的运算负载

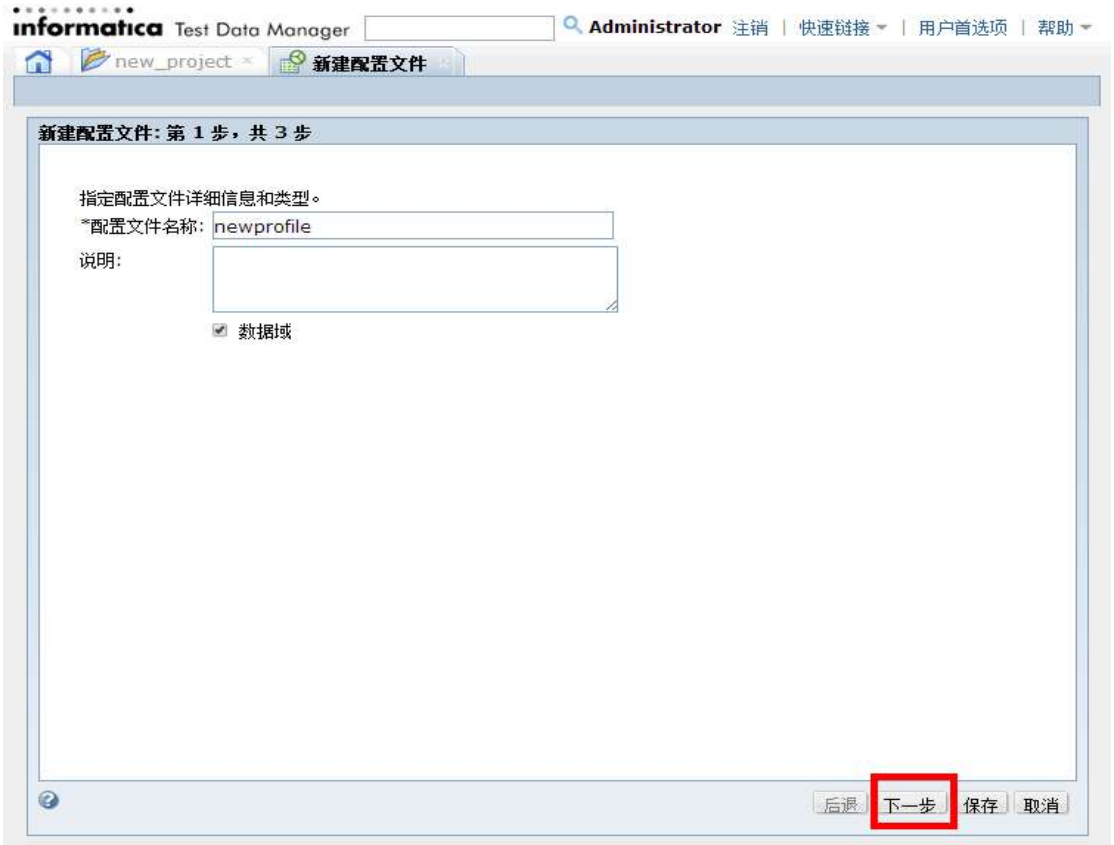
注 6：基于数据本身的发现，需要通过配置"Max Rows to profile""Minimum confirmance percent"两个参数来限制发现作业的运作规模及命中率阀值，从而提高作业效率

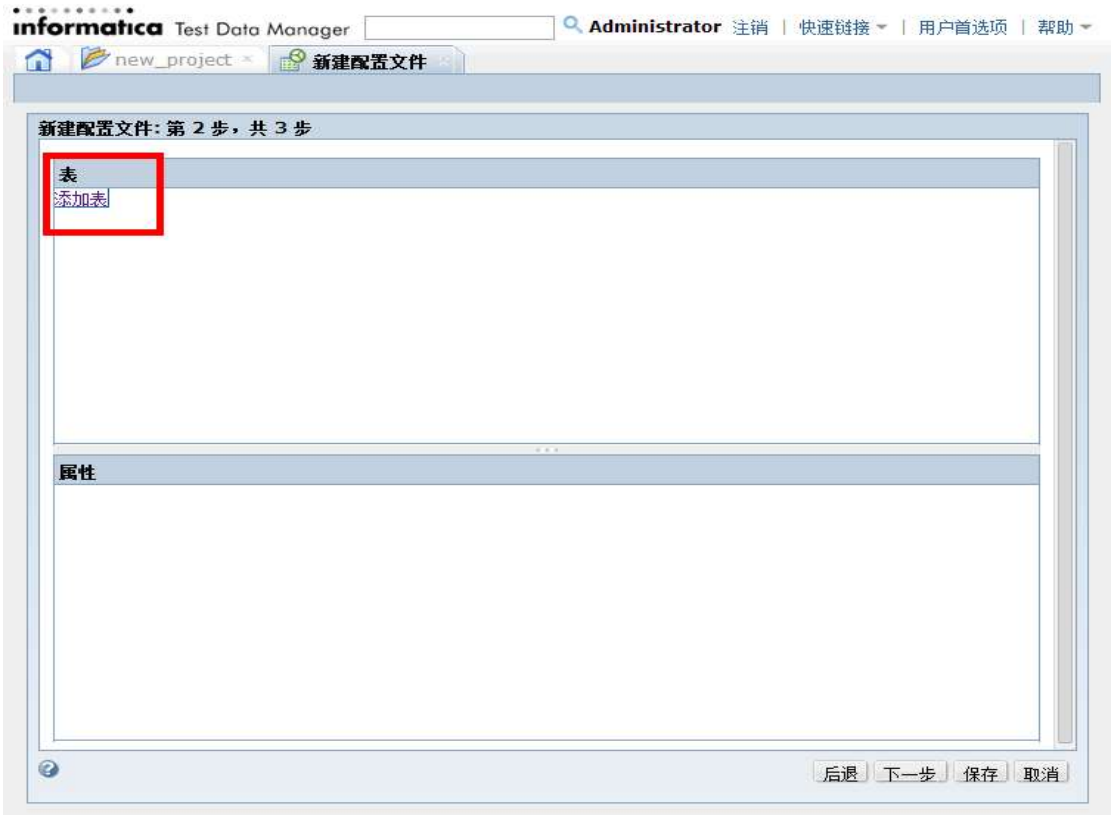
创建完成发现作业后，在"**Discover 标签**"-->"**Profile 标签**"中选中相关作业后，通过右上方"**Actions**"-->"**Run Profile**"功能进行发现作业的执行。通过"Monitor 标签"可以监控当前正在执行或执行完毕以及在队列中的作业。

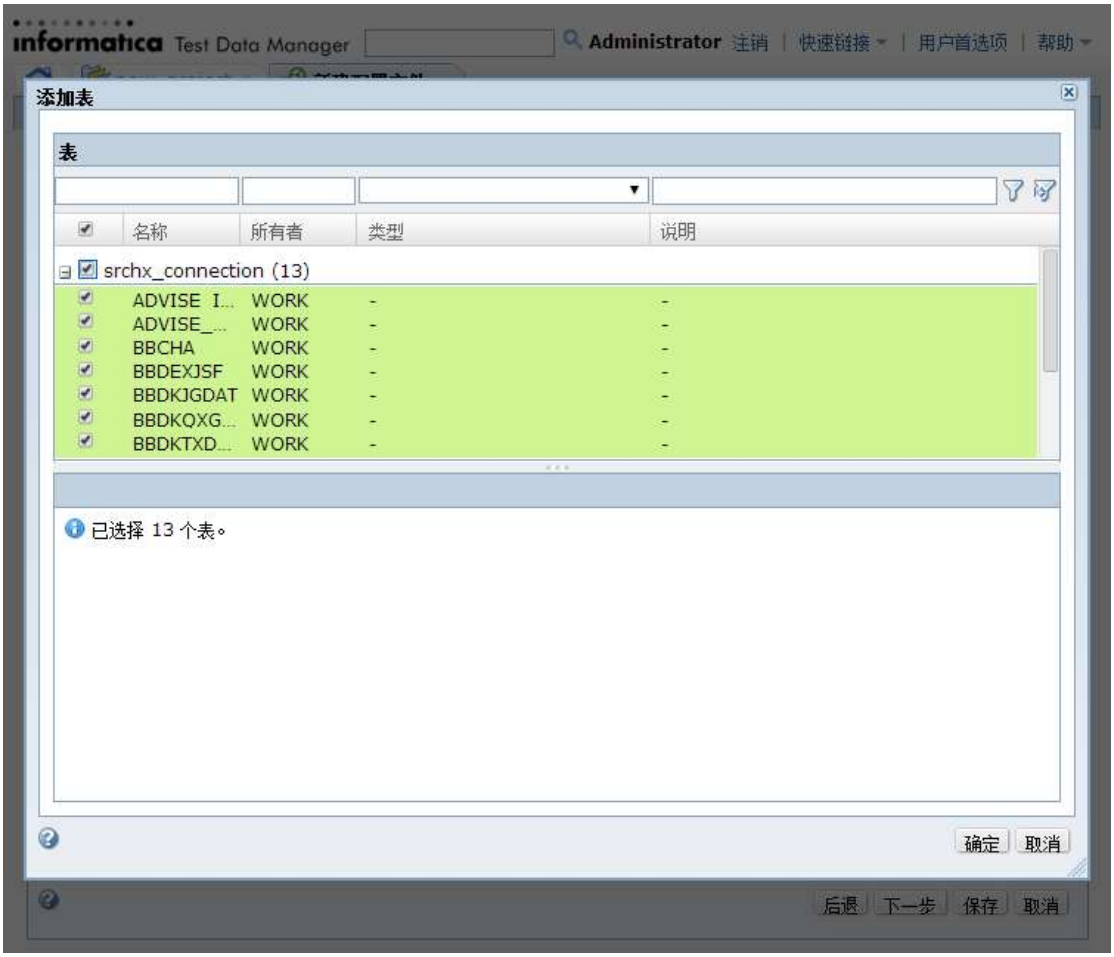
当发现作业在"**Monitor 标签**"中显示"**Success**"后可以打开所执行的发现作业，在"Data Domain 标签中"可以看到发现的结果

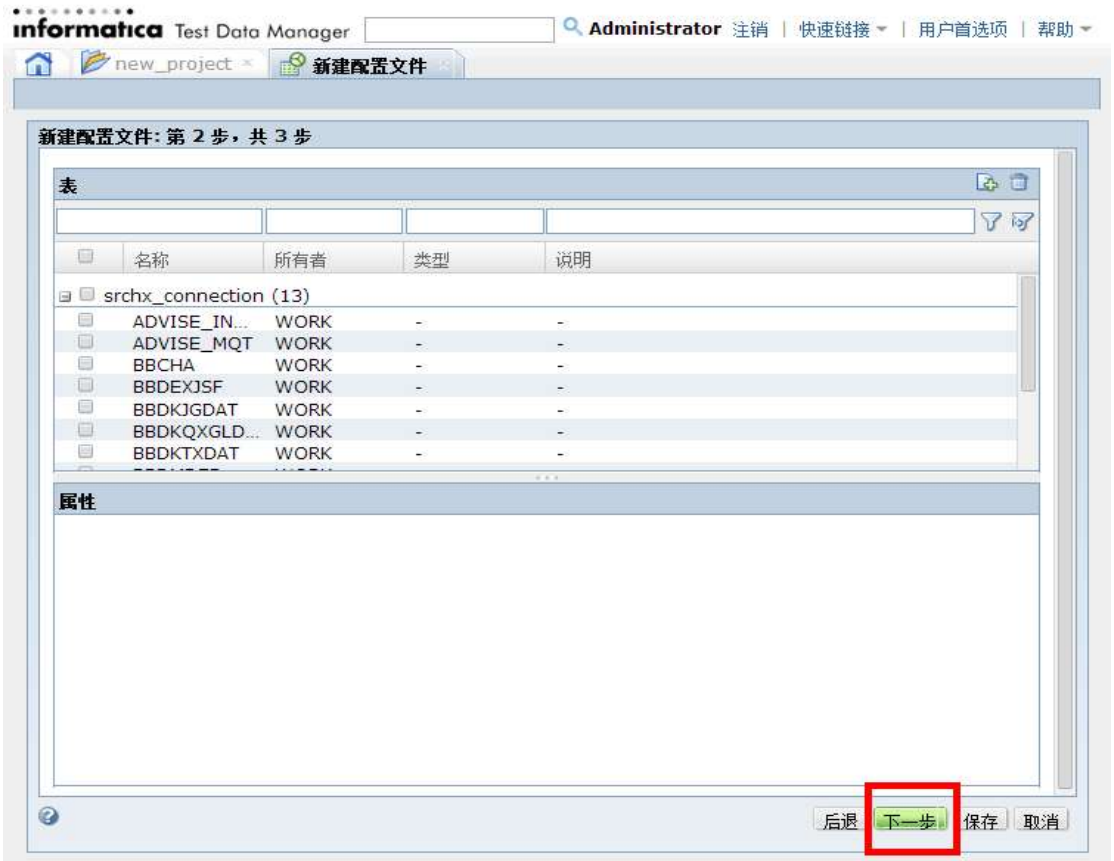
注 1：由于 web 显示的问题，完成发现作业通常需要关闭该作业的 Overview 标签并重新打开，以便正常显示"Data Domain"

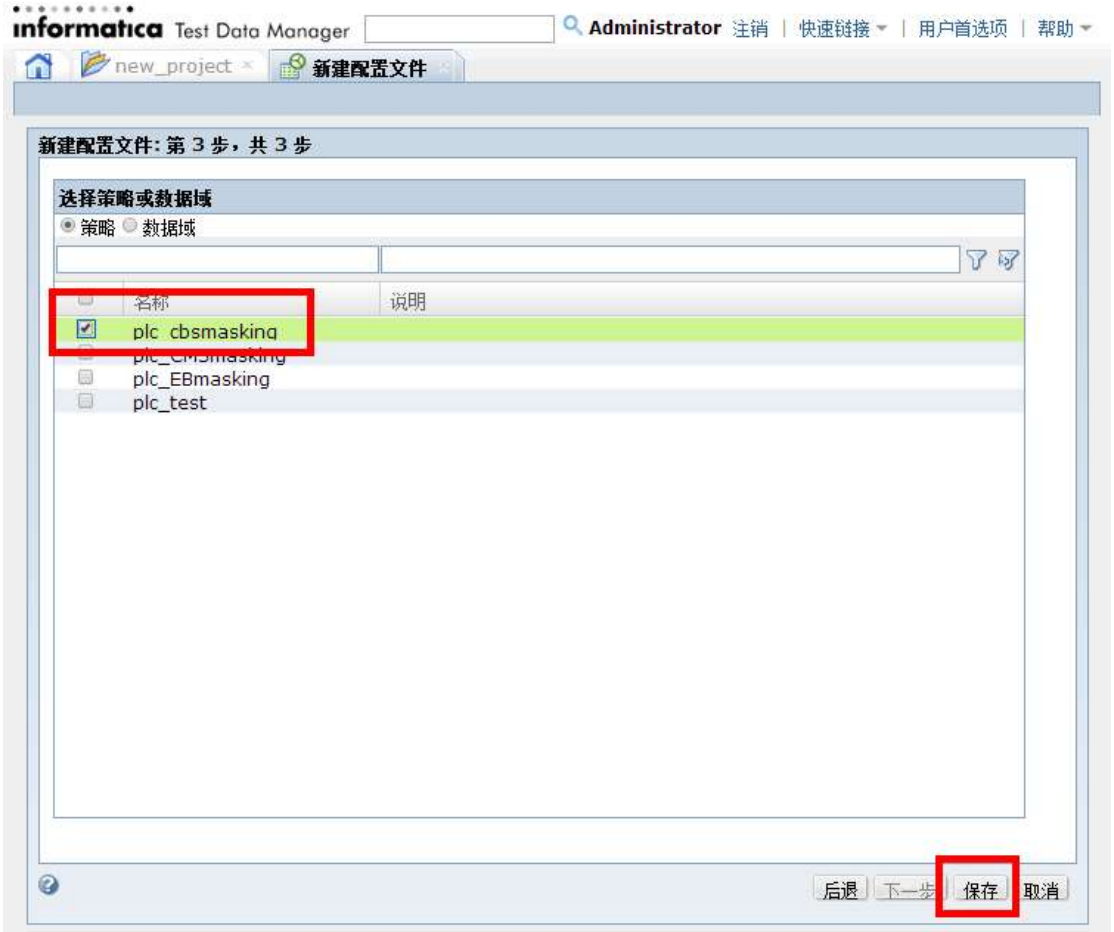


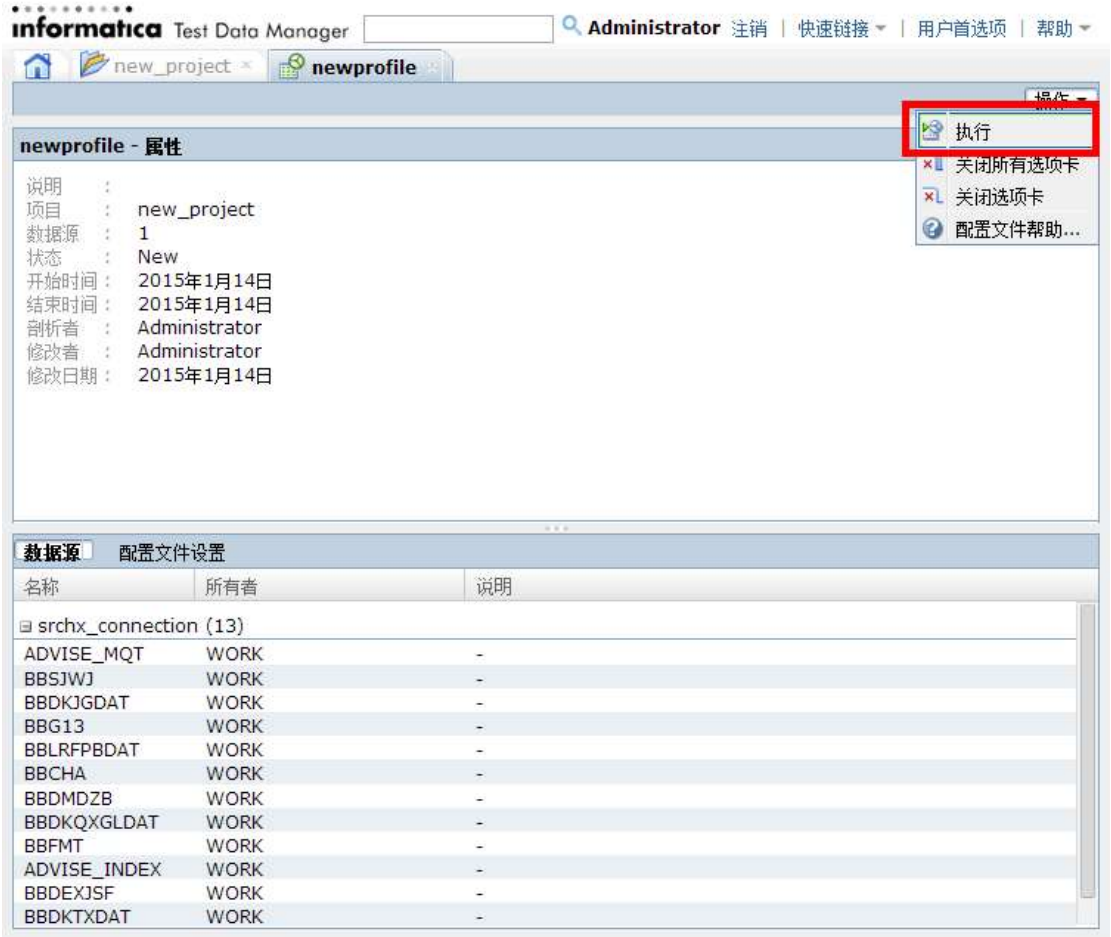



informatica Test Data Manager
Administrator 注销 | 快速链接 | 用户首选项 | 帮助
new_project
新建配置文件
新建配置文件: 第 2 步，共 3 步
表
添加表
属性
后退 下一步 保存 取消


informatica Test Data Manager
Administrator 注销 | 快速链接 | 用户首选项 | 帮助
添加表
表
名称 所有者 类型 说明
srchx_connection (13)
ADVISE_I... WORK - -
ADVISE_... WORK - -
BBCHA WORK - -
BBDEXJSF WORK - -
BBDKJGDAT WORK - -
BBDKQXG... WORK - -
BBDKTXD... WORK - -
已选择 13 个表。
确定 取消
后退 下一步 保存 取消


informatica Test Data Manager
Administrator 注销 | 快速链接 | 用户首选项 | 帮助
new_project
新建配置文件
新建配置文件: 第 2 步，共 3 步
表
名称 所有者 类型 说明
srchx_connection (13)
ADVISE_IN... WORK - -
ADVISE_MQT WORK - -
BBCHA WORK - -
BBDEXJSF WORK - -
BBDKJGDAT WORK - -
BBDKQXGLD... WORK - -
BBDKTXDAT WORK - -
属性
后退
下一步
保存
取消


informatica Test Data Manager
Administrator 注销 | 快速链接 | 用户首选项 | 帮助
new_project
新建配置文件
新建配置文件: 第 3 步，共 3 步
选择策略或数据域
策略 数据域
名称 说明
plc_cbsmasking
plc_EBmasking
plc_test
后退
下一步
保存
取消



3.3.5. 确认敏感发现的结果

TDM 基于 Pattern 的发现功能能够简化数据梳理的工作量，但其发现结果依旧需要必要的人工干预来予以确认。在"Data Domain 标签"中可以看到所有疑似敏感的数据表以及字段，甚至一个字段被初步标记为完全不同的多个敏感域定义。可以通过字段列表右侧"status"栏位,并从中选择"**approve/reject**"对发现结果进行确认

注 1：一个字段被标记为多个敏感域，往往是因为该字段同时符合多个 Data Domain 中的不同的 pattern，精炼 pattern 定义在一定程度上可以提高命中率

注 2：选择"**approve/reject**"后务必需要点击绿色对勾予以确认

3.3.6. 脱敏规则的绑定

通过 TDM 发现功能以及必要的人工干预后，数据库中的相关字段已经被成功表示为符合其业务含义的敏感对象域。由于一个 Data Domain 中可以包含多个不同的脱敏规则。因此基于发现结果需要将明确的脱敏规则与字段进行绑定。

打开相关的项目，在"**Define 标签**"-->"**Data Masking 标签**"中进行规则的绑定，分别在列表中的"Policy 栏位"，"**Masking Rule**"栏位，为字段指定相关的规则

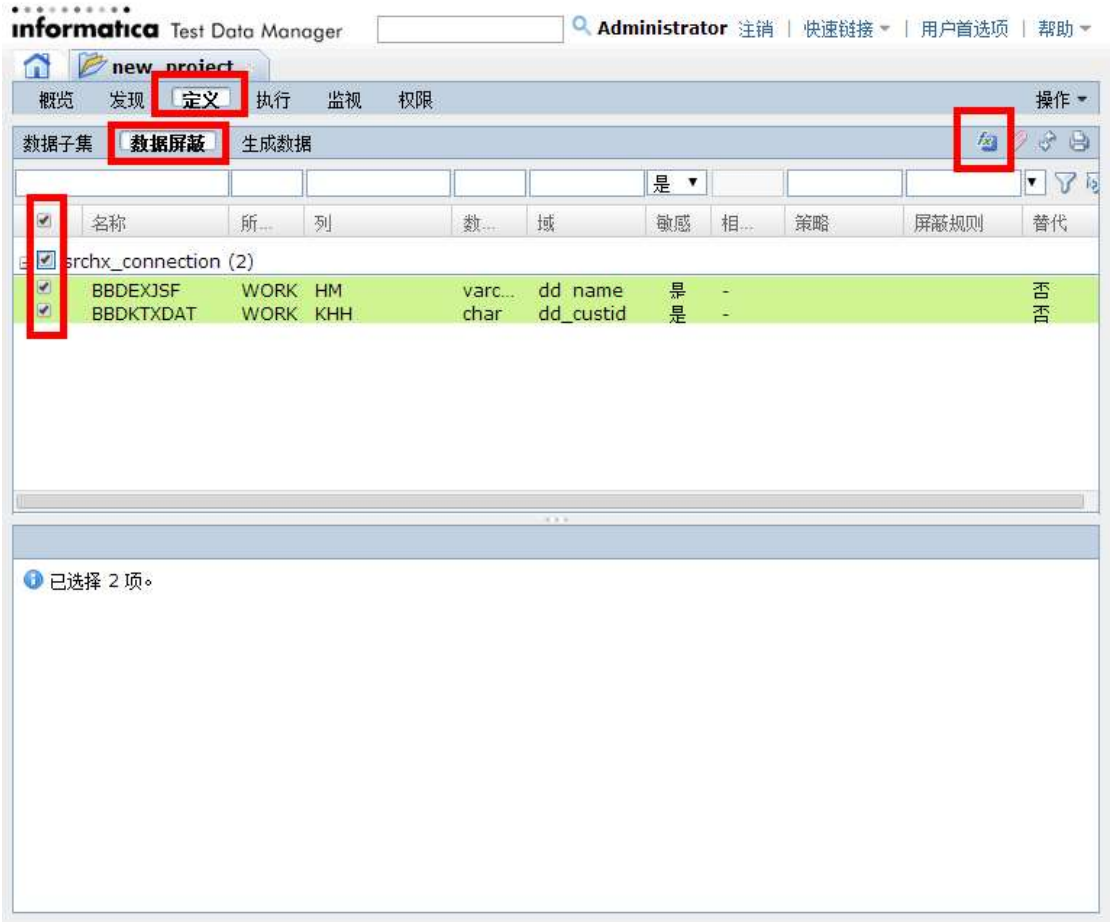
注 1：只有包含于安全策略(详见 3.3)中的规则才有可能被用于绑定

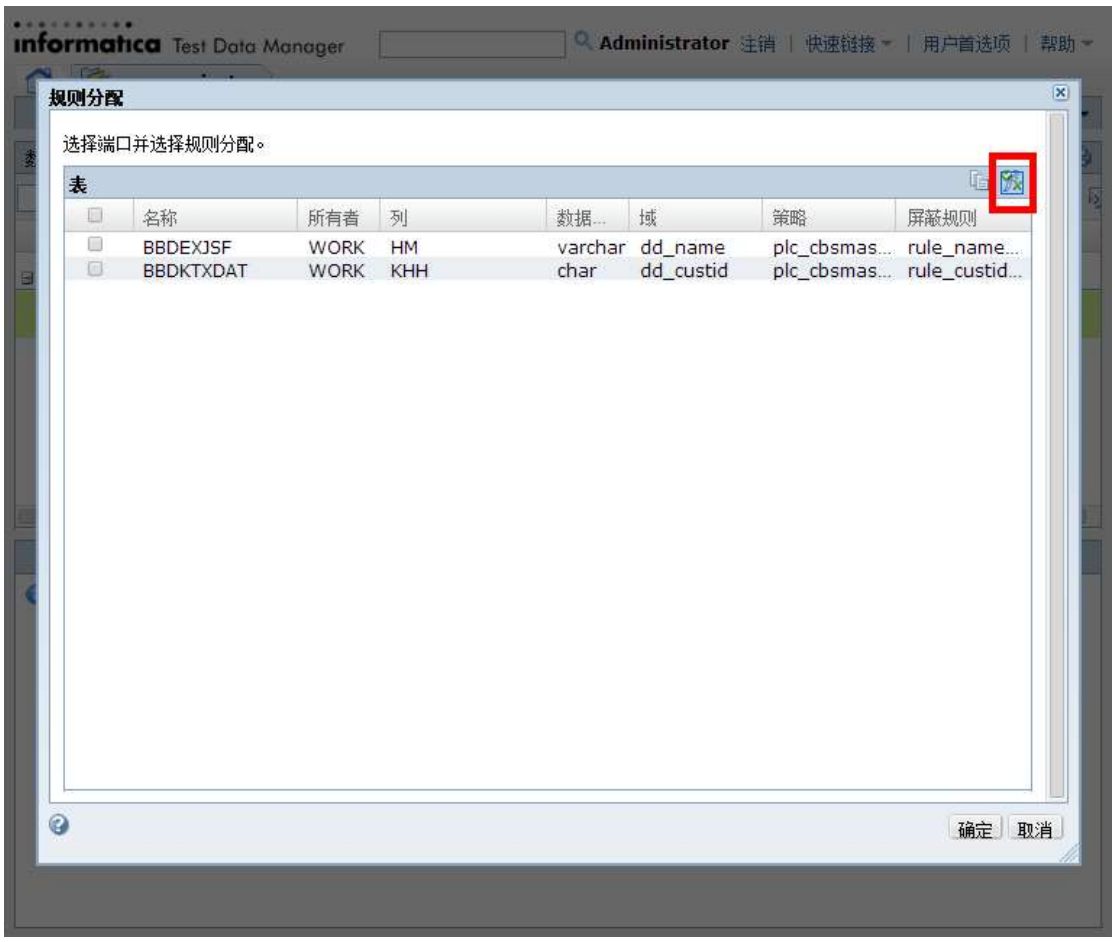
注 2：如果 Data Domain 中的某一个脱敏规则被设置为 Default，那么可以通过批量选中所有的相关字段，通过"**Rule Assignment**"按钮打开窗口，并选择

"**Default Assignment**"将缺省脱敏规则批量的与相关的字段进行绑定，可以极大地提高工作效率

注 3：由于此前对字段进行了敏感数据域的确认，因此在字段列表中的"**Sensitive 栏位**"可以选择"Yes"并刷新筛选条件，此操作可以把非敏感字段从列表中隐藏

注 4：本环节中另有"**Subset**"与"**Data Generation**"两个功能标签，本文档暂不对其进行描述





3.3.7. 脱敏作业的创建

打开具体的项目，通过"**Execute 标签**"-->"**Actions**"-->"**New Plan**"创建具体的执行计划。

在创建向导中的 Step.2 中通过"**Add Masking Components**"将此前的 Policy 加入执行计划，"**Add Subset Component**"以及"**Add Generation Components**"用于子集与测试数据创生在本文中暂时予以忽略跳过。

在 Step.5 中可以对添加的组件进行最后的适用范围配置，在右侧窗口选中特定的字段，并通过右上方"On"和"Off"开关对该字段在本次执行计划中是否需要进行最终的脱敏处理进行调整。

在 Step.7 中可以对执行计划的数据库连接属性进行配置，TDM Workbench 支持"库-->库"，"库-->文件"流向的脱敏作业，并可以选择是否在执行作业前 truncate 表，禁用索引等。在"**Error and Recovery**"栏可以对异常处理机制进行预设，"**Adavnced Settings**"则可以对包括 Schema 等相关的细节进行配置

3.3.8. 脱敏作业的初始化与执行

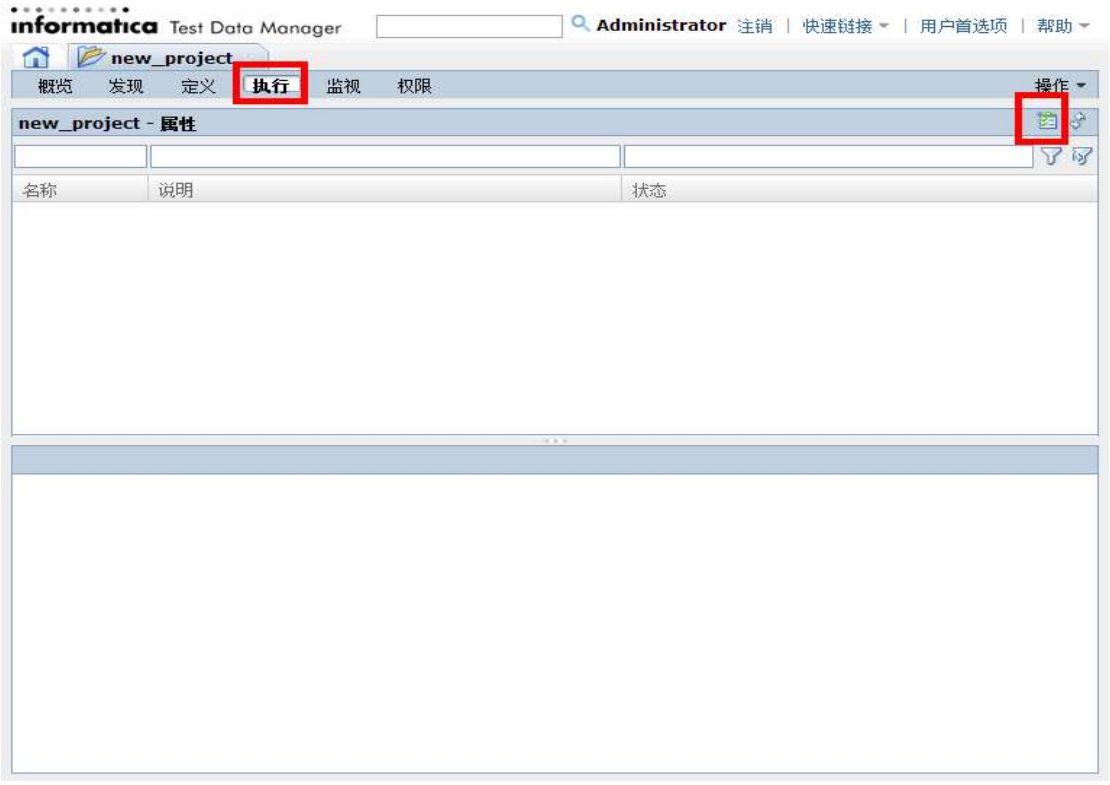
在完成创建执行计划的工作后，可在自动打开的执行计划(Plan)页面中通过"**Actions**"-->"**Generate Workflow**"对执行计划仅习惯初始化

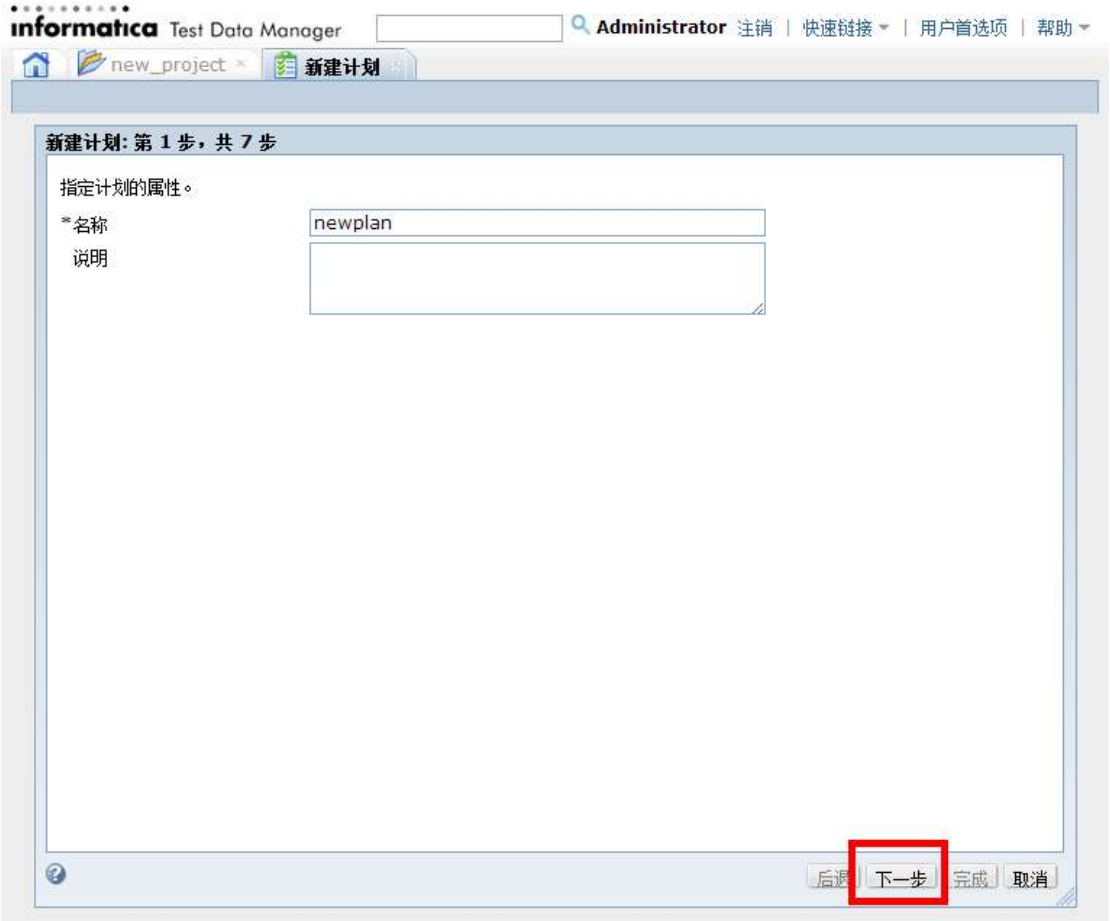
注 1：Generate Workflow 动作会将此前执行计划的各项配置转换为 PWC 可以识别并执行的映射及工作流，并上传至 PWC Repository 中去

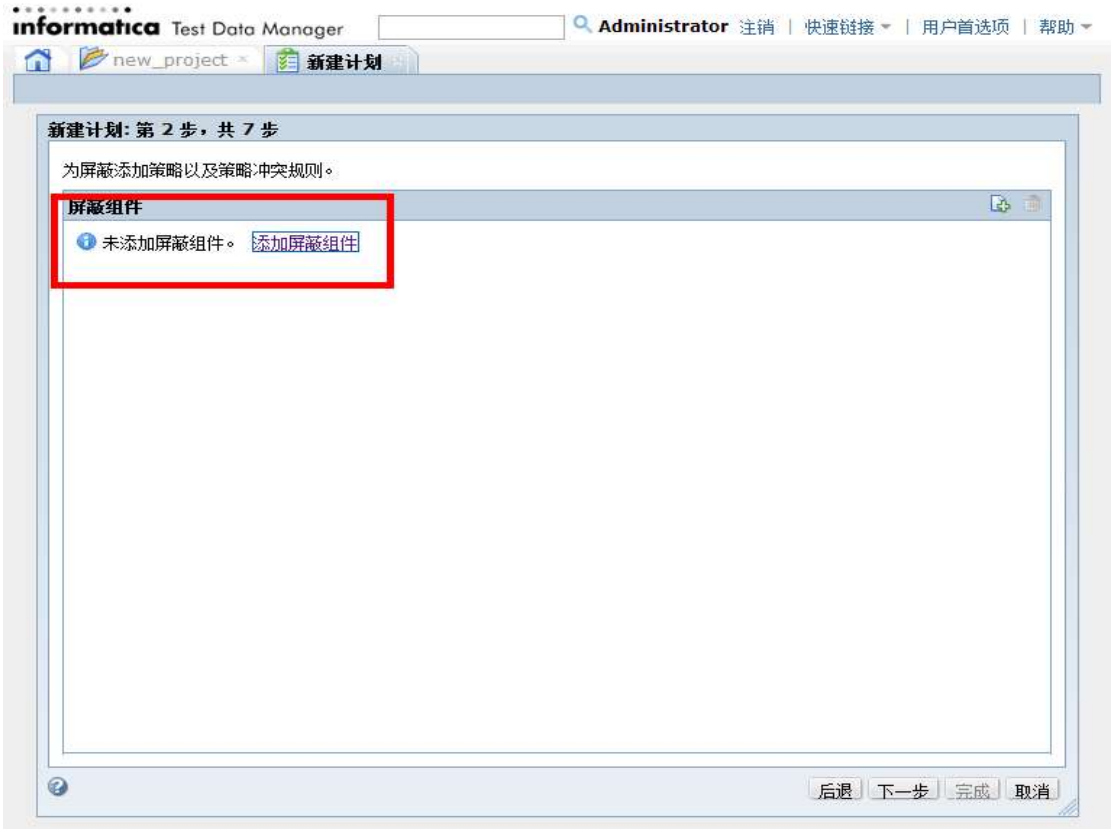
注 2：初始化作业，同样可以在"**Monitor 标签**"中进行监控

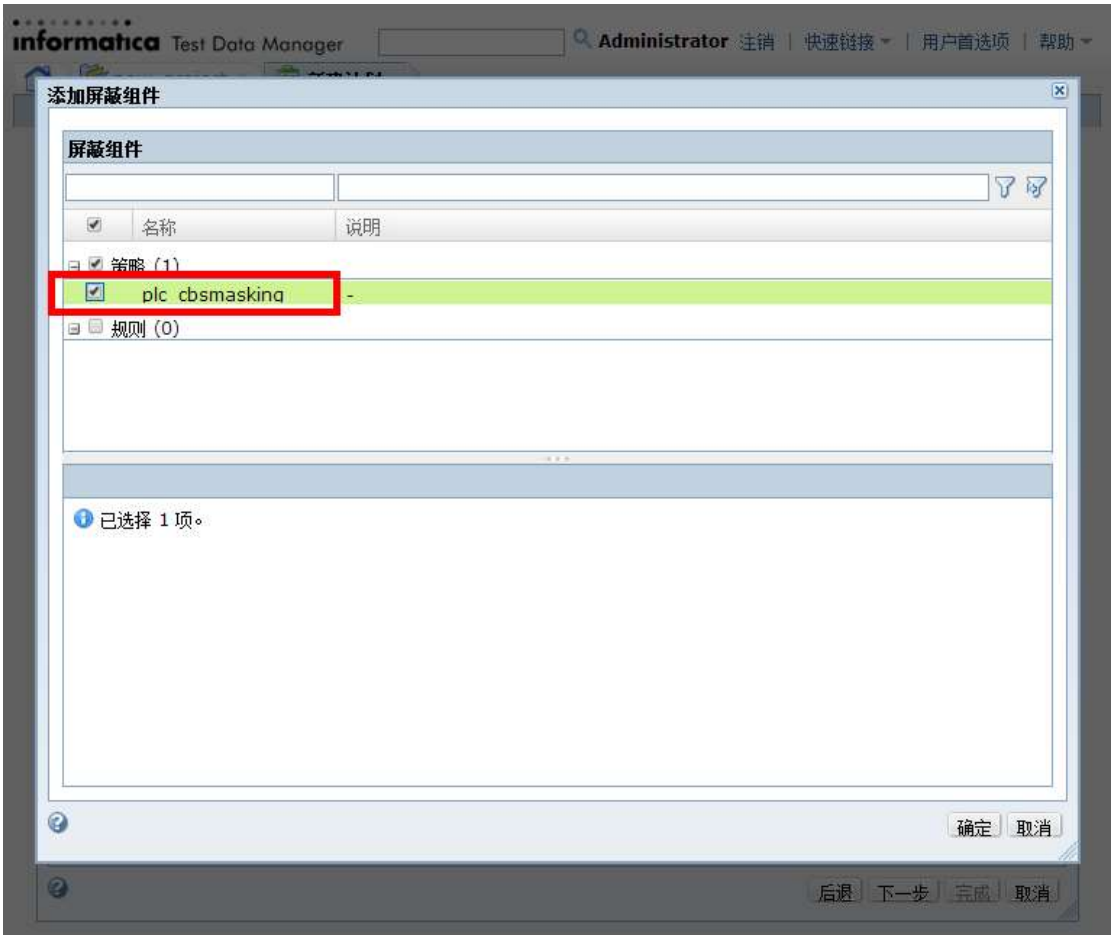
注 3：每当执行计划中的配置发生变化，均需要重新进行 Generate Workflow 动作

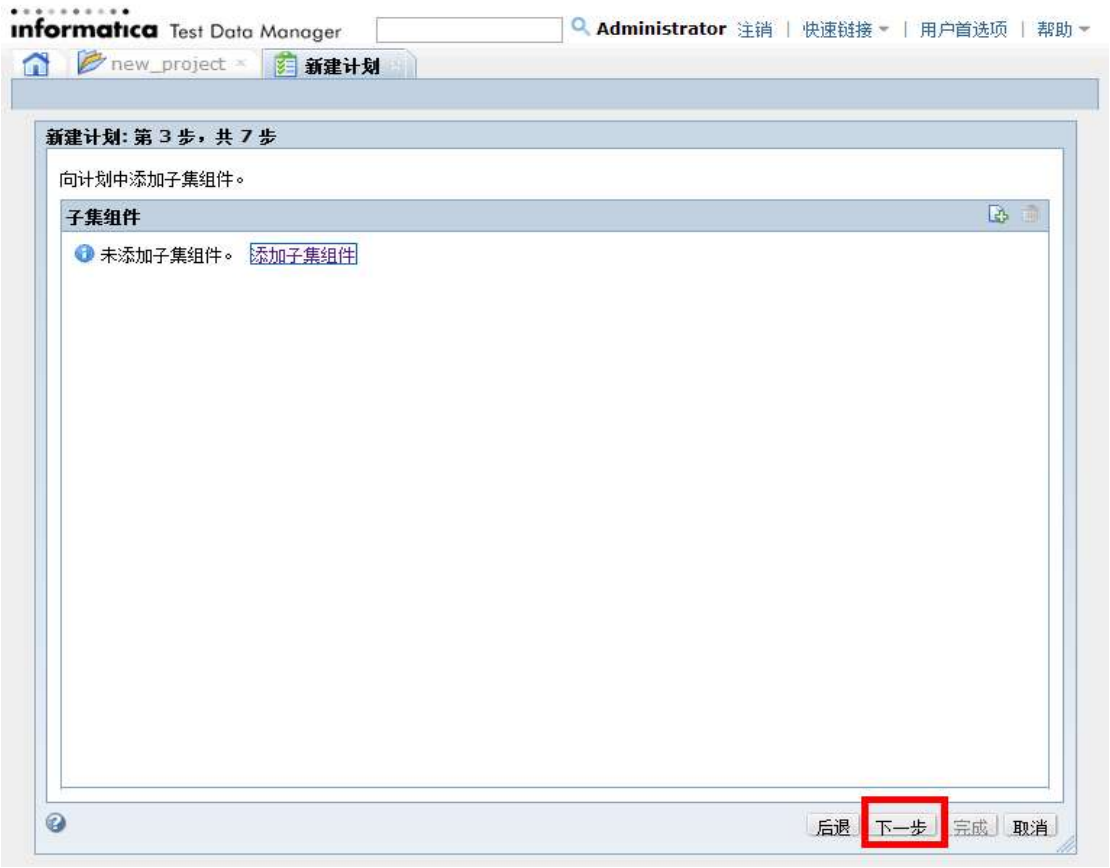
当 Workflow 成功生成并上传后，可以在执行计划的"**Workflow 标签**"中看到所生成的工作流，通过"**Action**"-->"**Execute Workflow**"正式执行脱敏作业，当"**Monitor 标签**"中的执行动作标识为"**Success**"后，意味着整个脱敏逻辑在技术上已经完成，此后所需要做的仅是在目标端对脱敏结果进行验证。

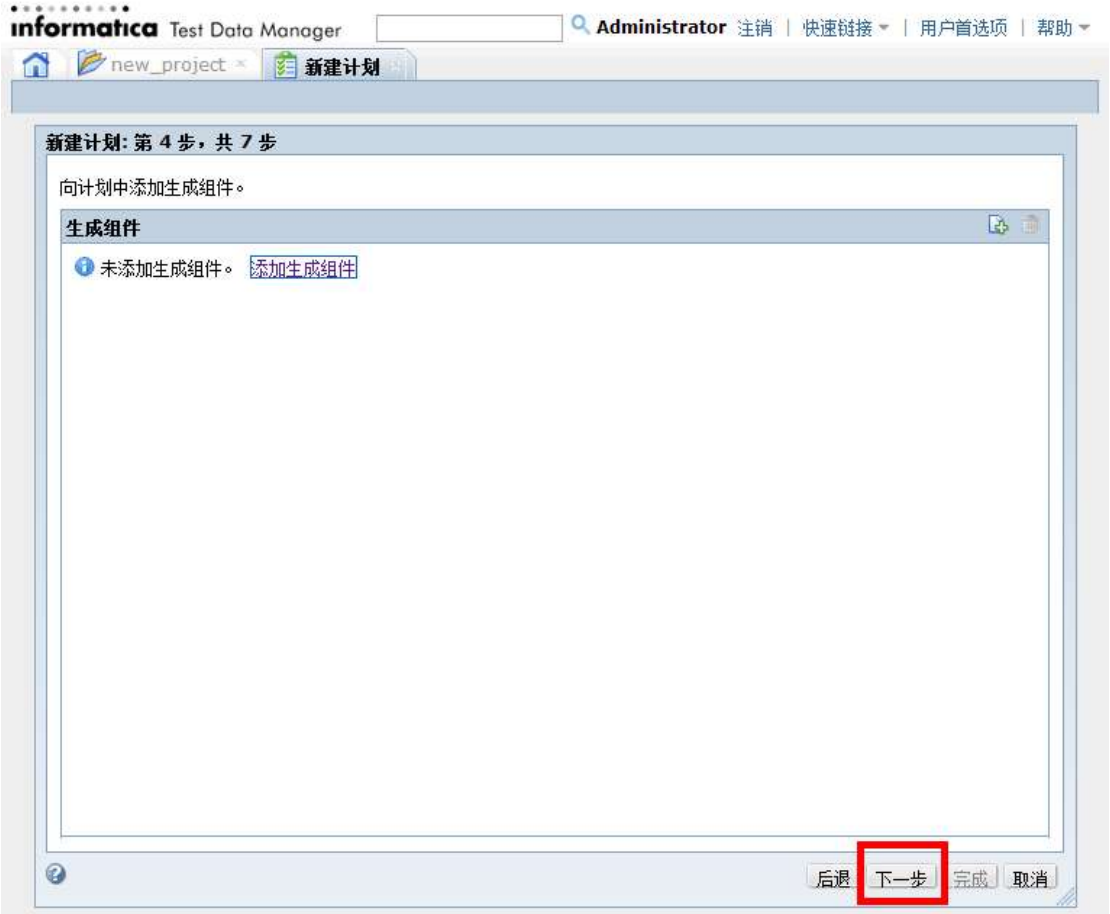

informatica Test Data Manager
Administrator 注销 | 快速链接 | 用户首选项 | 帮助
new_project
概览 发现 定义 执行 监视 权限 操作
new_project - 属性
名称 说明 状态

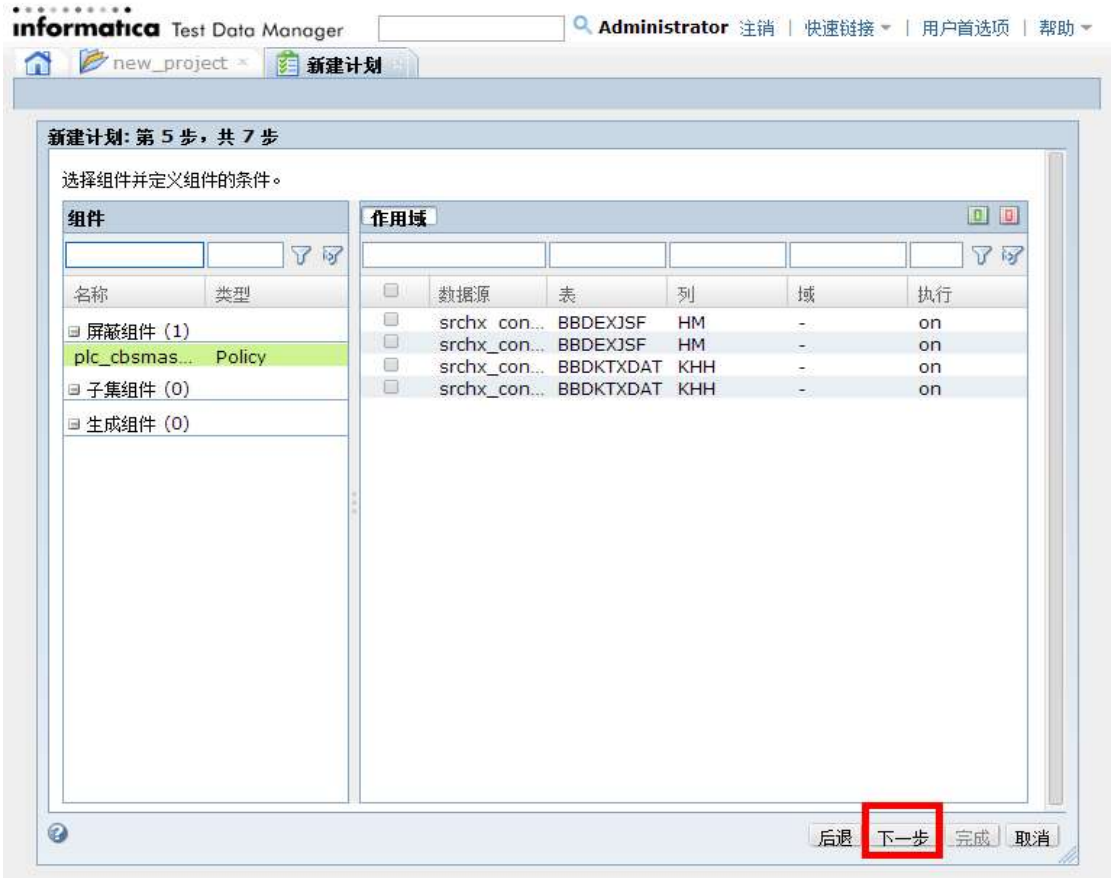

informatica Test Data Manager
Administrator 注销 | 快速链接 | 用户首选项 | 帮助
new_project
新建计划
新建计划: 第 1 步，共 7 步
指定计划的属性。
*名称
newplan
说明
后退 下一步 完成 取消

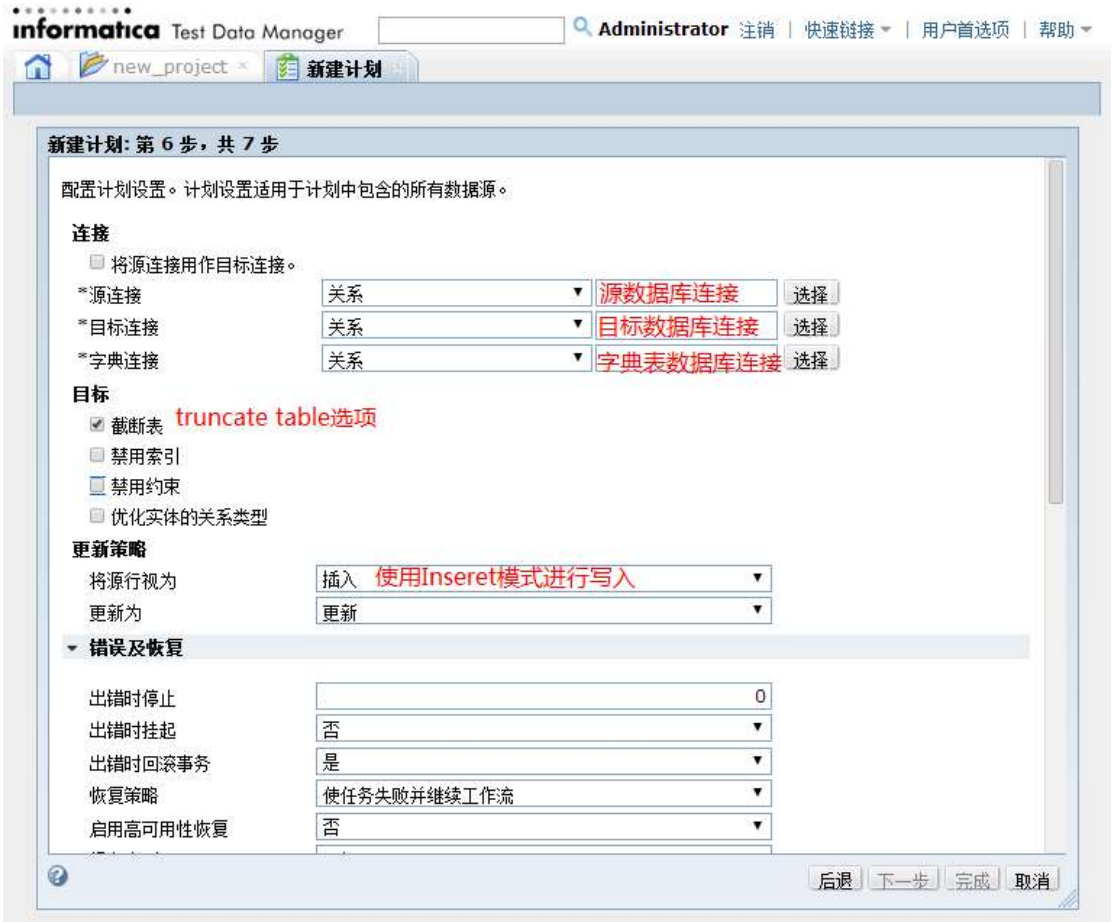

informatica Test Data Manager
Administrator 注销 | 快速链接 | 用户首选项 | 帮助
new_project
新建计划
新建计划: 第 2 步，共 7 步
为屏蔽添加策略以及策略冲突规则。
屏蔽组件
未添加屏蔽组件。 添加屏蔽组件
后退 下一步 完成 取消

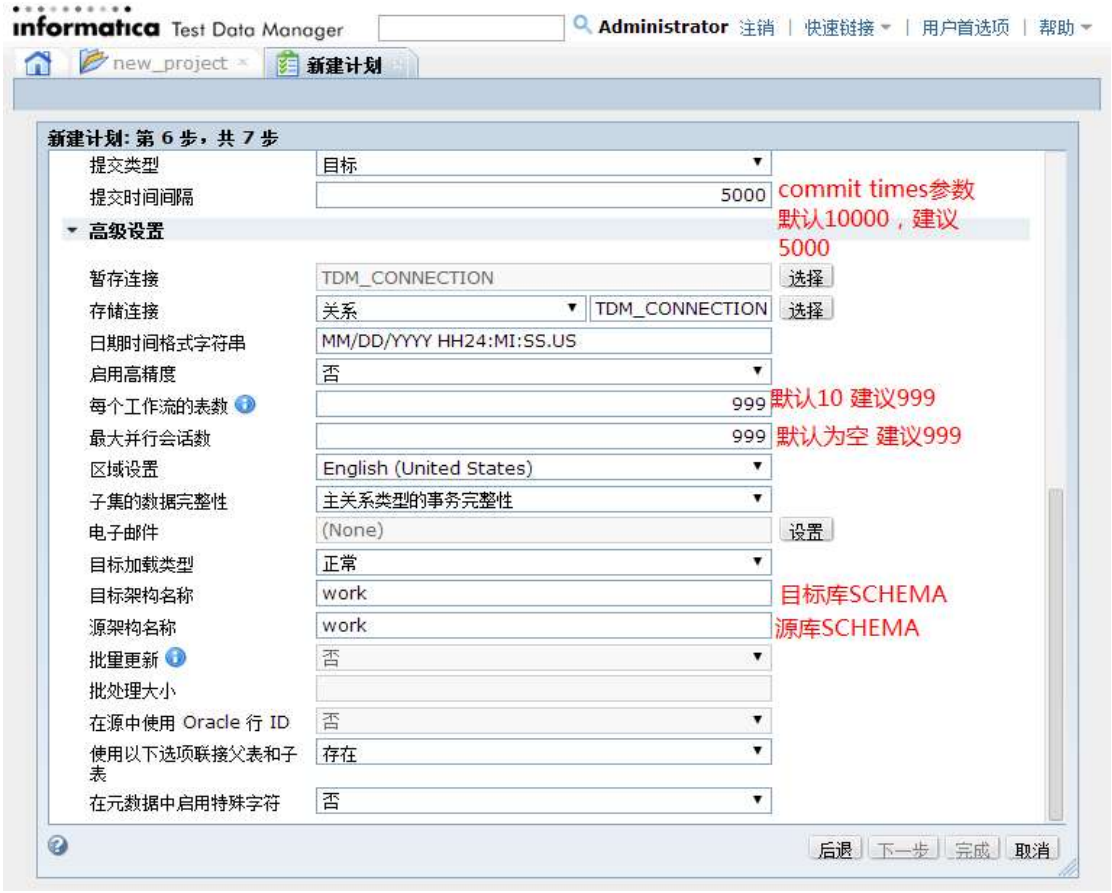

informatica Test Data Manager
Administrator 注销 | 快速链接 | 用户首选项 | 帮助
添加屏蔽组件
屏蔽组件
名称
说明
策略 (1)
plc_cbsmasking
规则 (0)
已选择 1 项。
确定 取消
后退 下一步 完成 取消

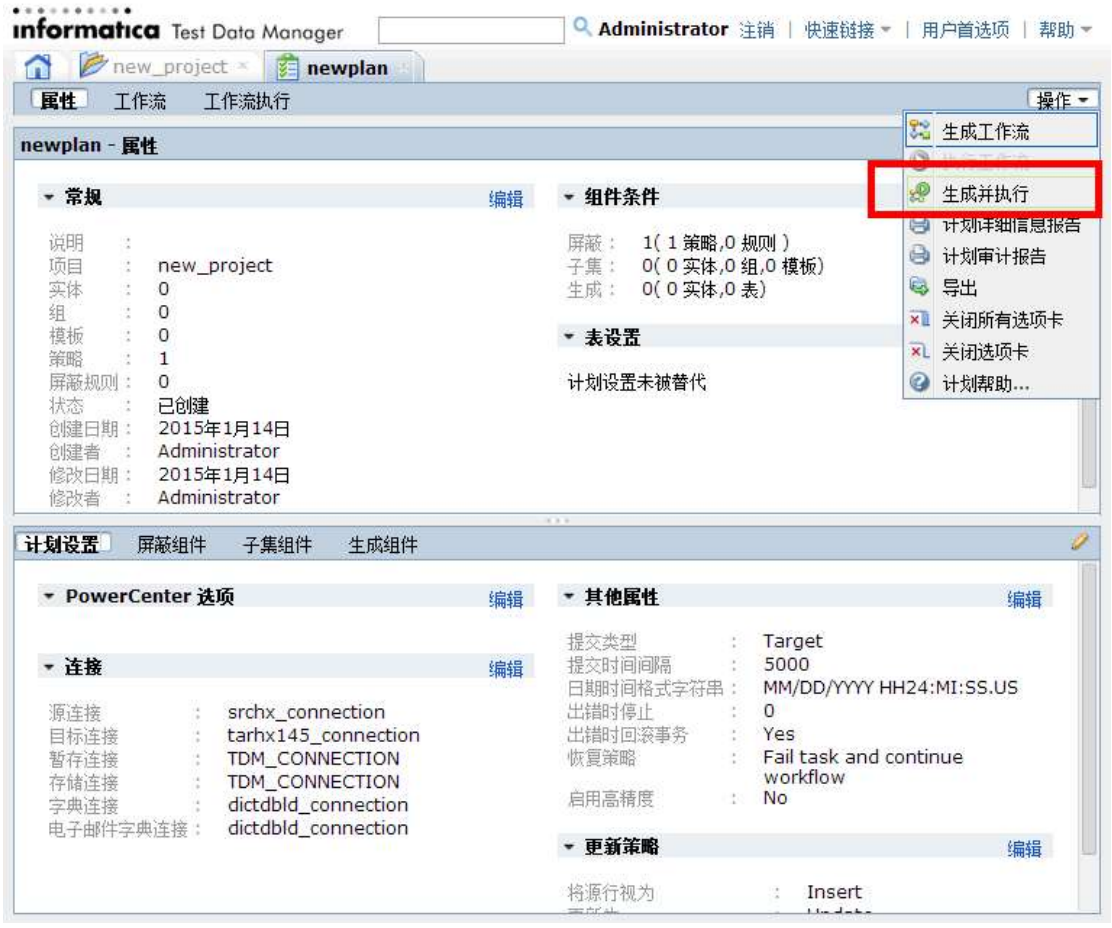

informatica Test Data Manager
Administrator 注销 | 快速链接 | 用户首选项 | 帮助
new_project
新建计划
新建计划: 第 3 步，共 7 步
向计划中添加子集组件。
子集组件
未添加子集组件。 添加子集组件
后退
下一步
完成
取消


informatica Test Data Manager
Administrator 注销 | 快速链接 | 用户首选项 | 帮助
new_project
新建计划
新建计划: 第 4 步，共 7 步
向计划中添加生成组件。
生成组件
未添加生成组件。 添加生成组件
后退
下一步
完成
取消


informatica Test Data Manager
Administrator 注销 | 快速链接 | 用户首选项 | 帮助
new_project
新建计划
新建计划: 第 5 步，共 7 步
选择组件并定义组件的条件。
组件
作用域
名称
类型
屏蔽组件 (1)
plc_cbsmas... Policy
子集组件 (0)
生成组件 (0)
数据源
表
列
域
执行
srchx_con... BBDEXJSF HM - on
srchx_con... BBDEXJSF HM - on
srchx_con... BBDKTXDAT KHH - on
srchx_con... BBDKTXDAT KHH - on
后退
下一步
完成
取消


informatica Test Data Manager
Administrator 注销 | 快速链接 | 用户首选项 | 帮助
new_project
新建计划
新建计划: 第 6 步，共 7 步
配置计划设置。计划设置适用于计划中包含的所有数据源。
连接
将源连接用作目标连接。
*源连接
关系
源数据库连接
选择
*目标连接
关系
目标数据库连接
选择
*字典连接
关系
字典表数据库连接
选择
目标
截断表
truncate table选项
禁用索引
禁用约束
优化实体的关系类型
更新策略
将源行视为
插入
使用Inseret模式进行写入
更新为
更新
错误及恢复
出错时停止
0
出错时挂起
否
出错时回滚事务
是
恢复策略
使任务失败并继续工作流
启用高可用性恢复
否
后退
下一步
完成
取消


informatica Test Data Manager
Administrator 注销 | 快速链接 | 用户首选项 | 帮助
new_project
新建计划
新建计划: 第 6 步，共 7 步
提交类型 目标
提交时间间隔 5000
commit times参数 默认10000，建议5000
高级设置
暂存连接 TDM_CONNECTION 选择
存储连接 关系 TDM_CONNECTION 选择
日期时间格式字符串 MM/DD/YYYY HH24:MI:SS.US
启用高精度 否
每个工作流的表数 999 默认10 建议999
最大并行会话数 999 默认为空 建议999
区域设置 English (United States)
子集的数据完整性 主关系类型的事务完整性
电子邮件 (None) 设置
目标加载类型 正常
目标架构名称 work 目标库SCHEMA
源架构名称 work 源库SCHEMA
批量更新 否
批处理大小
在源中使用 Oracle 行 ID 否
使用以下选项联接父表和子表 存在
在元数据中启用特殊字符 否
后退 下一步 完成 取消


informatica Test Data Manager
Administrator 注销 | 快速链接 | 用户首选项 | 帮助
new_project
newplan
属性 工作流 工作流执行
操作
生成工作流
生成并执行
计划详细信息报告
计划审计报告
导出
关闭所有选项卡
关闭选项卡
计划帮助...
newplan - 属性
常规 编辑
说明 :
项目 : new_project
实体 : 0
组 : 0
模板 : 0
策略 : 1
屏蔽规则 : 0
状态 : 已创建
创建日期 : 2015年1月14日
创建者 : Administrator
修改日期 : 2015年1月14日
修改者 : Administrator
组件条件
屏蔽 : 1(1 策略,0 规则)
子集 : 0(0 实体,0 组,0 模板)
生成 : 0(0 实体,0 表)
表设置
计划设置未被替代
计划设置 屏蔽组件 子集组件 生成组件
PowerCenter 选项 编辑
连接 编辑
源连接 : srchx_connection
目标连接 : tarhx145_connection
暂存连接 : TDM_CONNECTION
存储连接 : TDM_CONNECTION
字典连接 : dictdbld_connection
电子邮件字典连接 : dictdbld_connection
其他属性 编辑
提交类型 : Target
提交时间间隔 : 5000
日期时间格式字符串 : MM/DD/YYYY HH24:MI:SS.US
出错时停止 : 0
出错时回滚事务 : Yes
恢复策略 : Fail task and continue workflow
启用高精度 : No
更新策略 编辑
将源行视为 : Insert

4.平台设计与扩展

本章节将对脱敏规则及数据与的创建进行说明

4.1. 创建脱敏规则

在 Workbench 中的"Policies 标签"中，通过右上方"**Actions**"-->"**New**"-->"**Masking Rules**"开启规则创建向导。创建向导共提供了三种类型的脱敏规则可供创建。

Standard 规则

是 TDM 平台内嵌的脱敏函数，针对 String,Numeric,Date,Binary 四种数据类型，提供共计 14 种内置函数，其中 CreditCard，Phone，key，Expression，Substitution 均为常用函数，通过向导中选择"**Repeatable Output**"并给出"**Seed**"值可以使得相同数据重复脱敏可以出现唯一结果，此功能可以用于实现表间关联关系的保持

注 1：并非所有的函数都支持"**Repeatable Output**"，需要根据需要选择对应的函数

注 2：Substition 函数即为替换规则，需要用到 2.3 中创建的字典表数据，若想要实现替换规则的输出值得唯一性 ,则需要确保脱敏表的条数小于字典表的记录条数

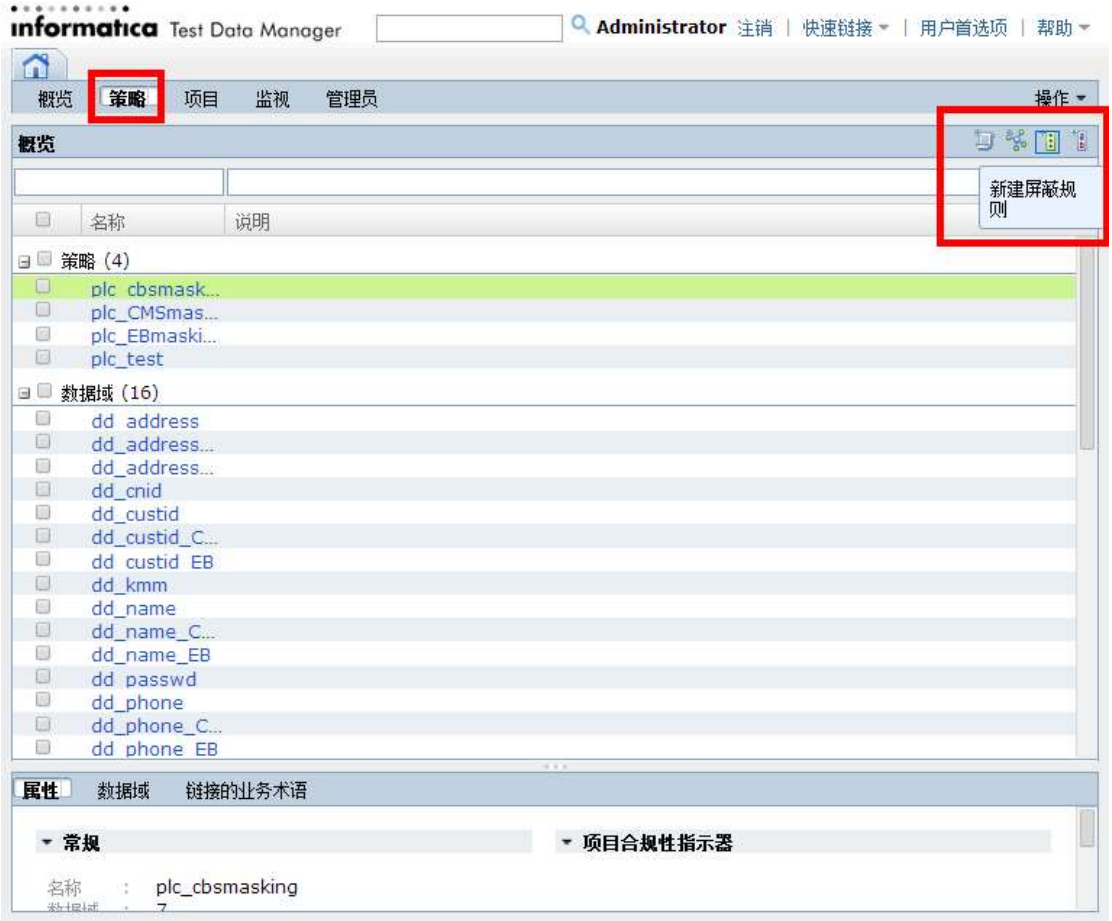
Mapplet 规则

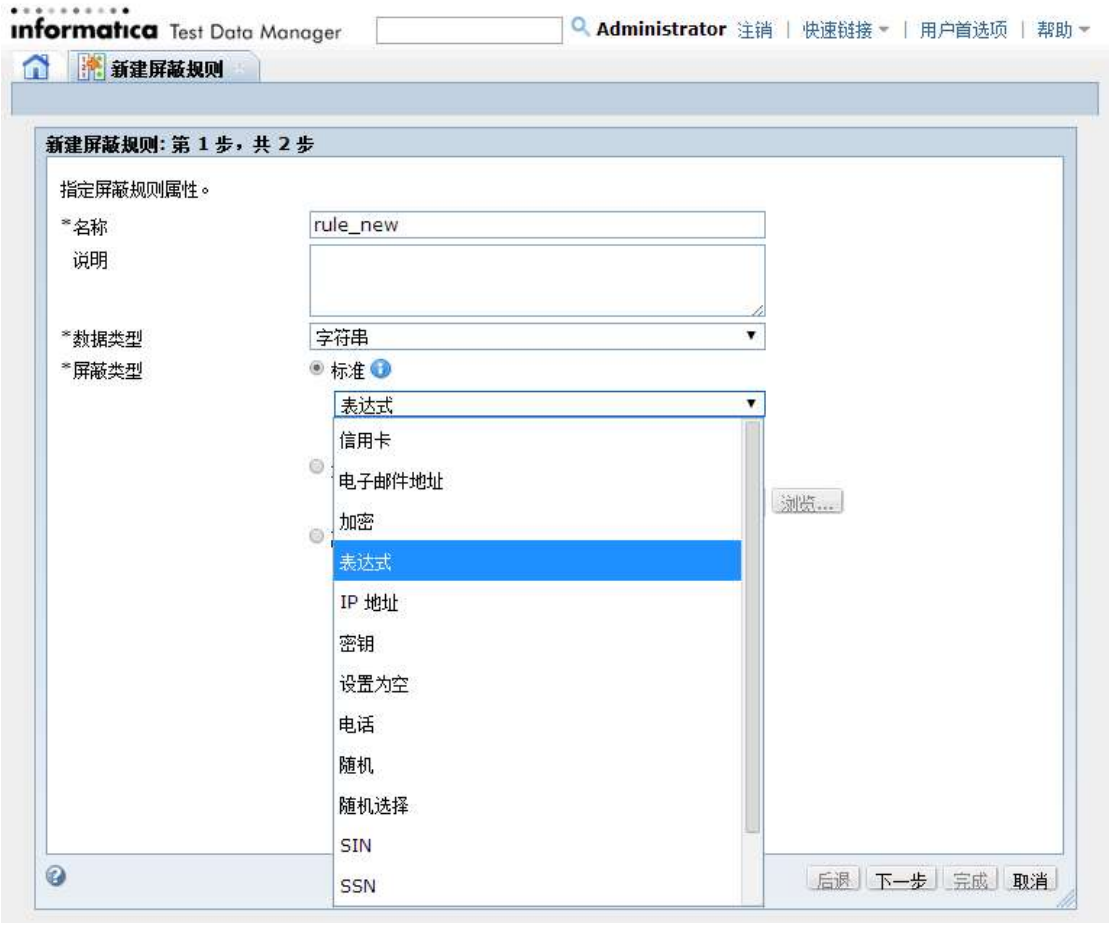
可以将通过 PWC Designer 设计的 mapplet 以 xml 文件的方式予以导入 ,此种类型可以实现更为复杂的脱敏逻辑，具体 Mapplet 设计参考 PWC 手册

Advanced 规则

提供一个基于 web 的可视化开发界面，在该向导中可以创建，一对一，一对多，多对一，多对多的脱敏规则，并可以嵌套调用其它的"原子"脱敏规则

创建后的规则，均可在"Policies 标签"中显示，点开具体的规则，可以在下方"**Masking Properties**"中看到其概况，并可在"Rule Simulator"对 Standard 规则进行模拟预览



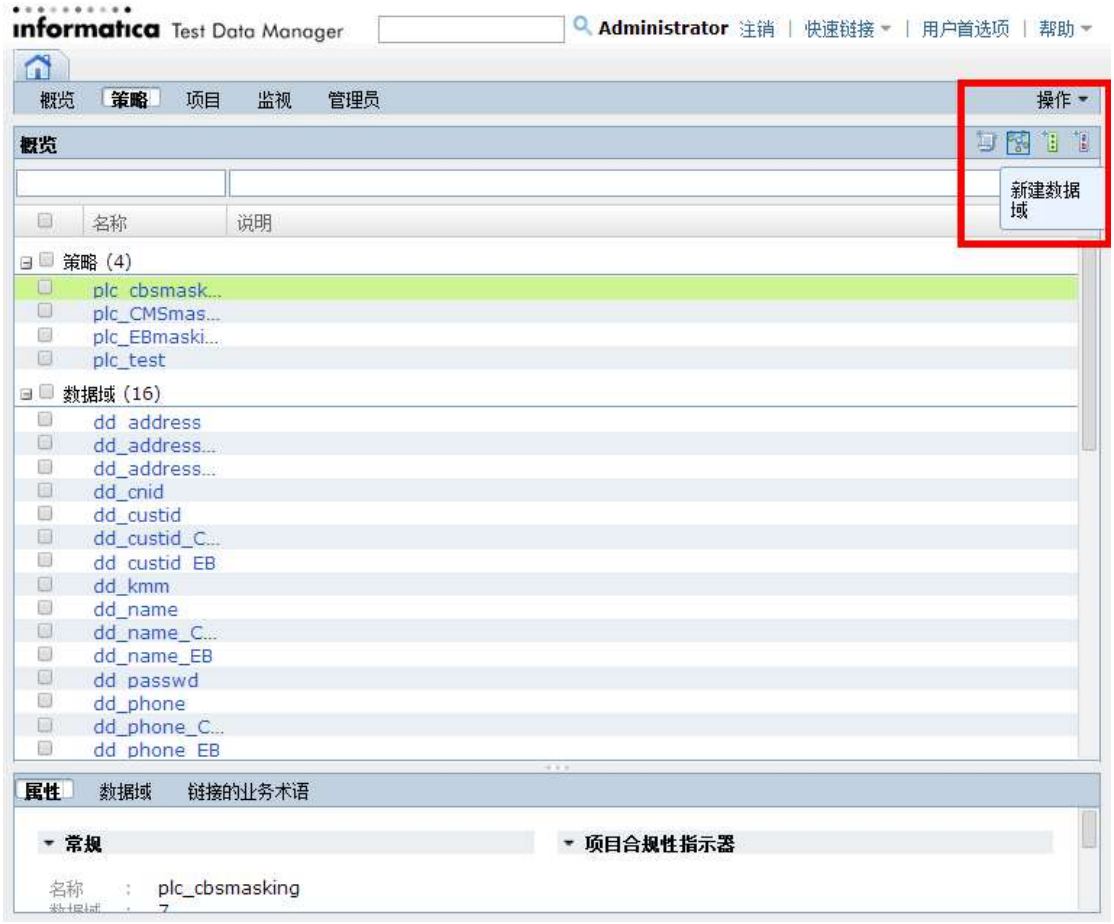


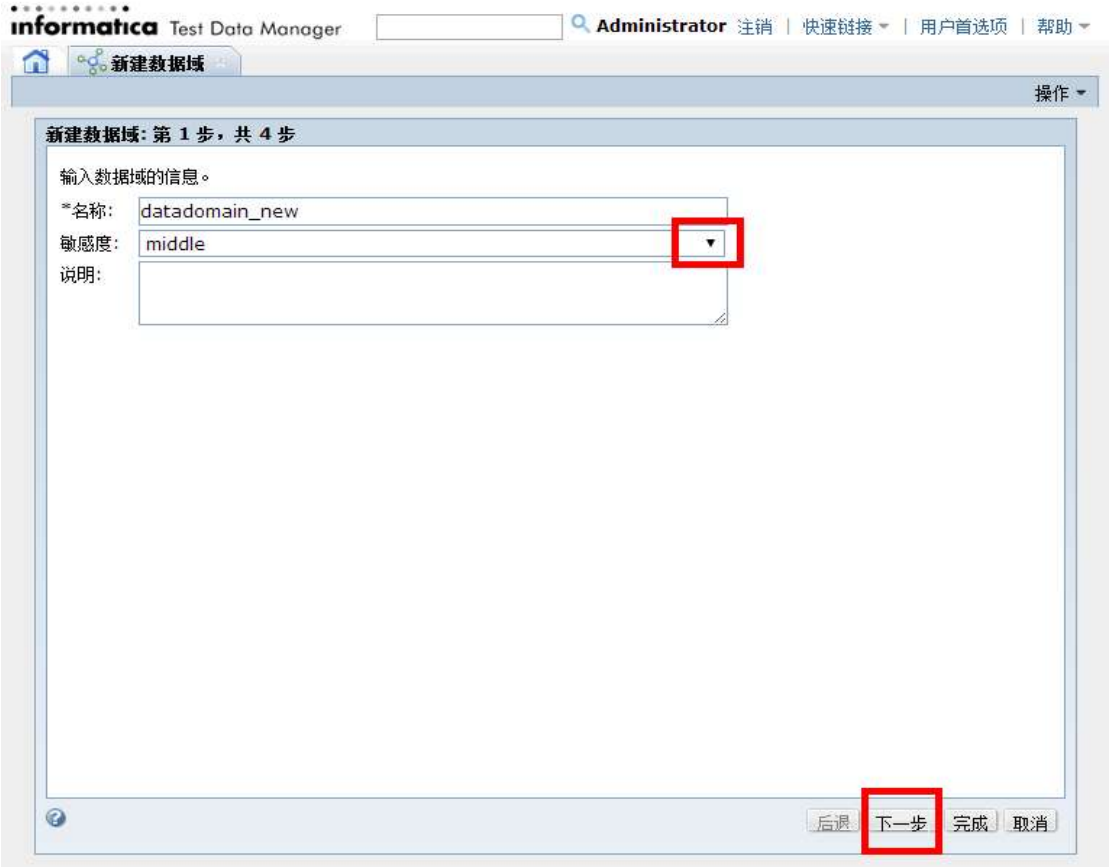
4.2. 创建数据域

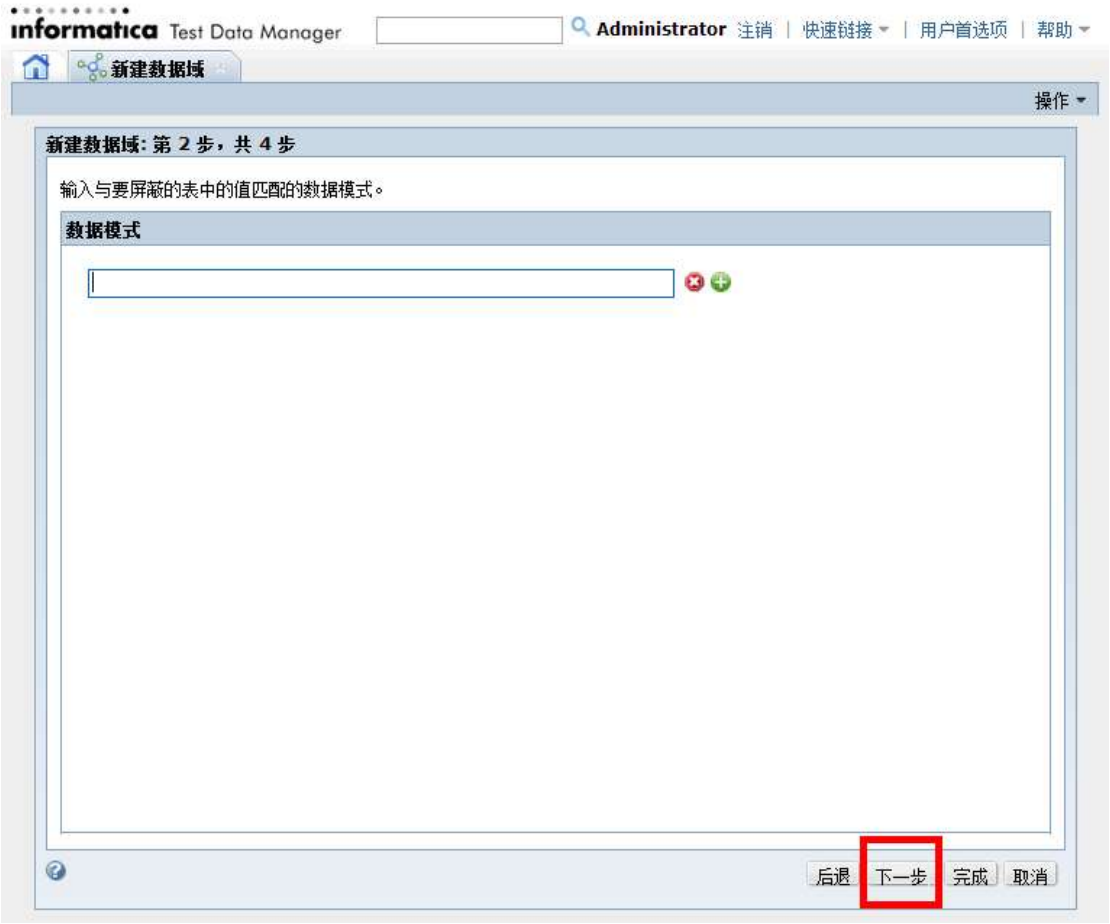
在 Workbench 中的"**Policies 标签**"中，通过右上方"**Actions**"-->"**New**"-->"**Data Domain**"开启数据域创建向导。在向导中给出数据域的名称与描述，此处的"**Sensitive Level**"即是指 2.1 中进行预设的敏感级别，可以根据业务需要为不同的数据域指派敏感级别。

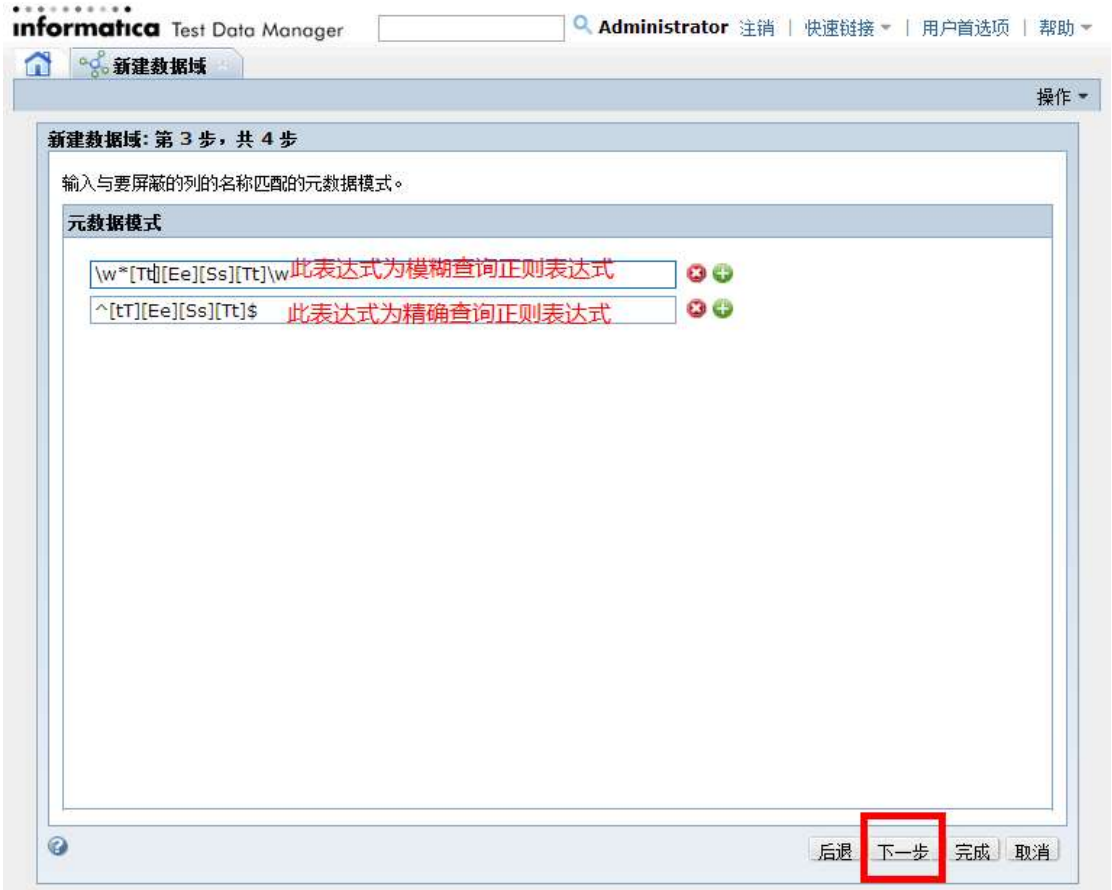
在"Data Domain"的定义中包括两种 Pattern,"**Data Pattern**"与"**Metadata Pattern**"，两者均通过正则表达式来匹配目标，前者主要匹配字段下的数据形态，例如卡号字段下的数据通常为 16 位数字，此时的"Data Pattern"便是"16 位数字"的正则表达式，而"**Metadata Pattern**"则用于匹配字段名称，例如卡号字段名通常包括 card 字串，因此其定义就应该是包含"card"字串的正则表达式。

在完成 Pattern 的定义后，需要将该数据域可能用到的脱敏规则添加进来(也可后续添加)，通过向导 Step.4 中的"**Add Rules**"可以实现




informatica Test Data Manager
Administrator 注销 | 快速链接 | 用户首选项 | 帮助
新建数据域
操作
新建数据域: 第 1 步，共 4 步
输入数据域的信息。
*名称: datadomain_new
敏感度: middle
说明:
后退 下一步 完成 取消

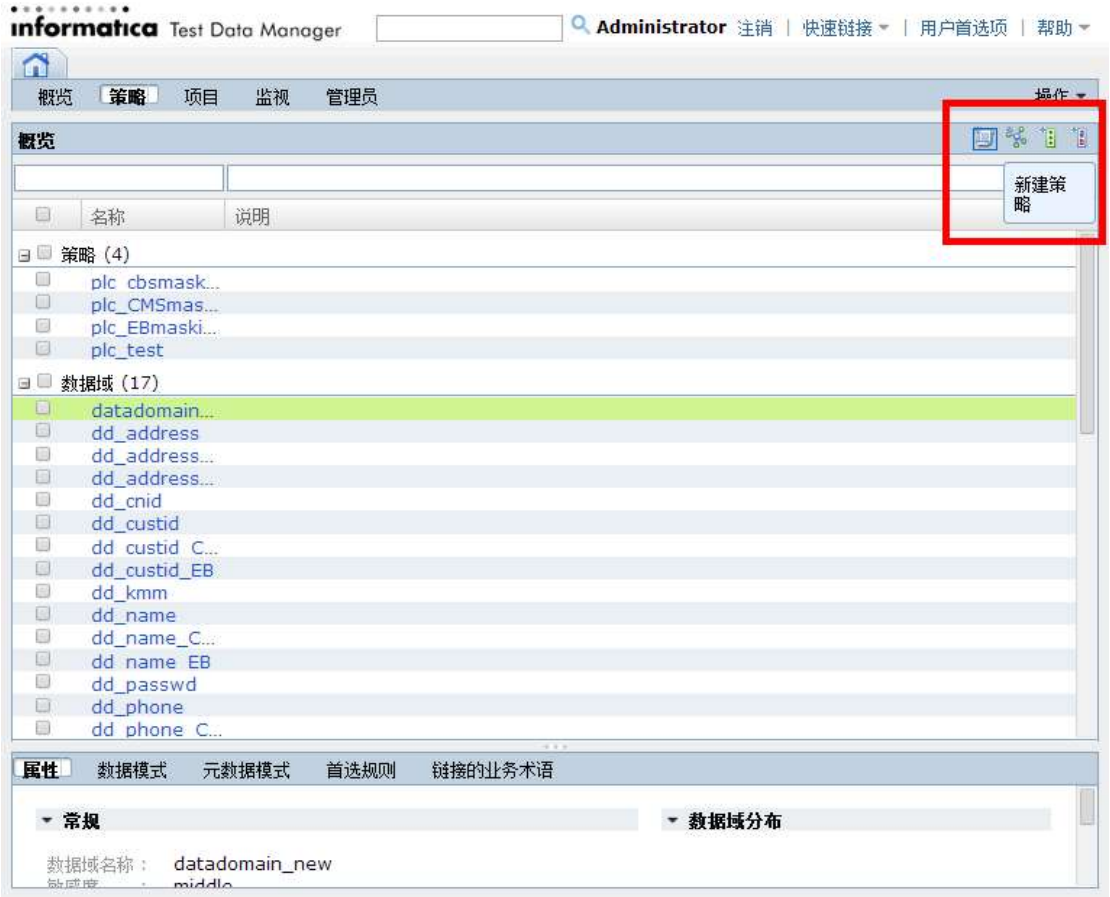

informatica Test Data Manager
Administrator 注销 | 快速链接 | 用户首选项 | 帮助
新建数据域
操作
新建数据域: 第 2 步，共 4 步
输入与要屏蔽的表中的值匹配的数据模式。
数据模式
后退 下一步 完成 取消

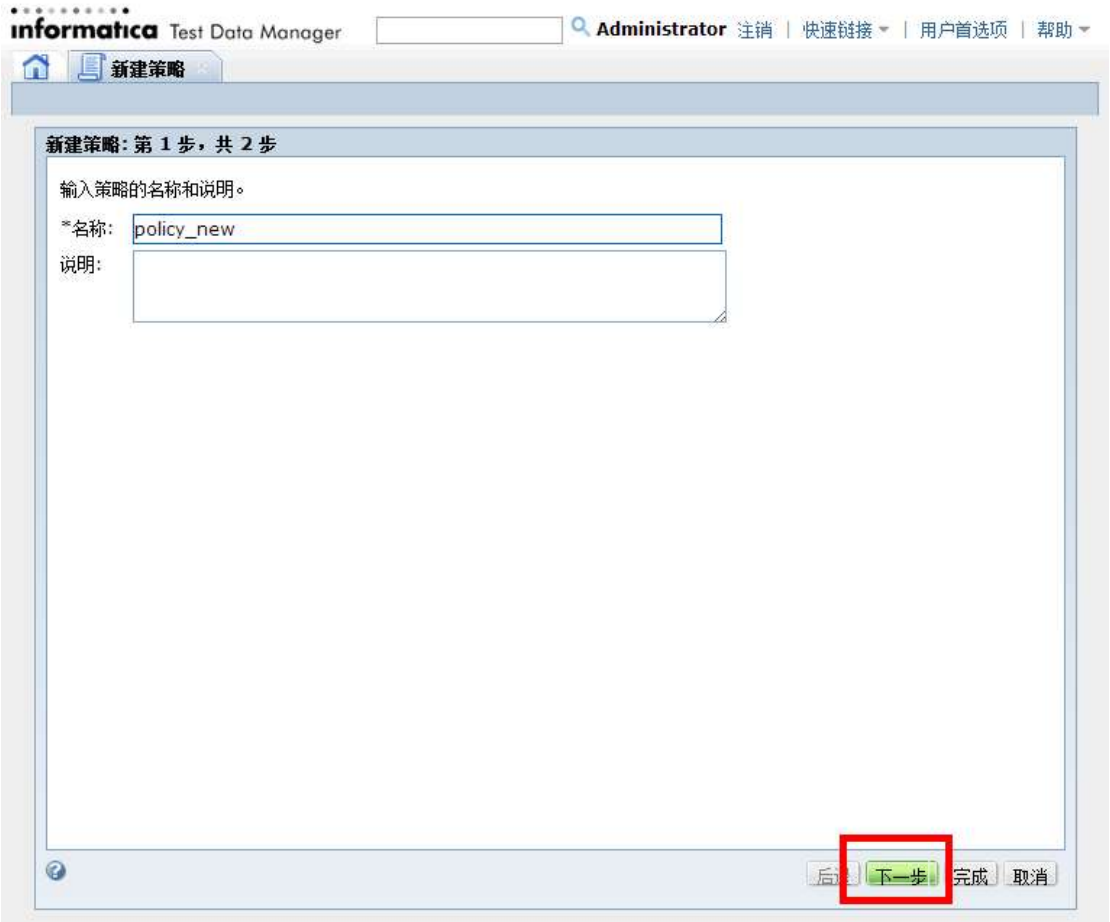

informatica Test Data Manager
Administrator 注销 | 快速链接 | 用户首选项 | 帮助
新建数据域
操作
新建数据域: 第 3 步，共 4 步
输入与要屏蔽的列的名称匹配的元数据模式。
元数据模式
\w*[Tt][Ee][Ss][Tt]\w 此表达式为模糊查询正则表达式
^[tT][Ee][Ss][Tt]$ 此表达式为精确查询正则表达式
后退
下一步
完成
取消

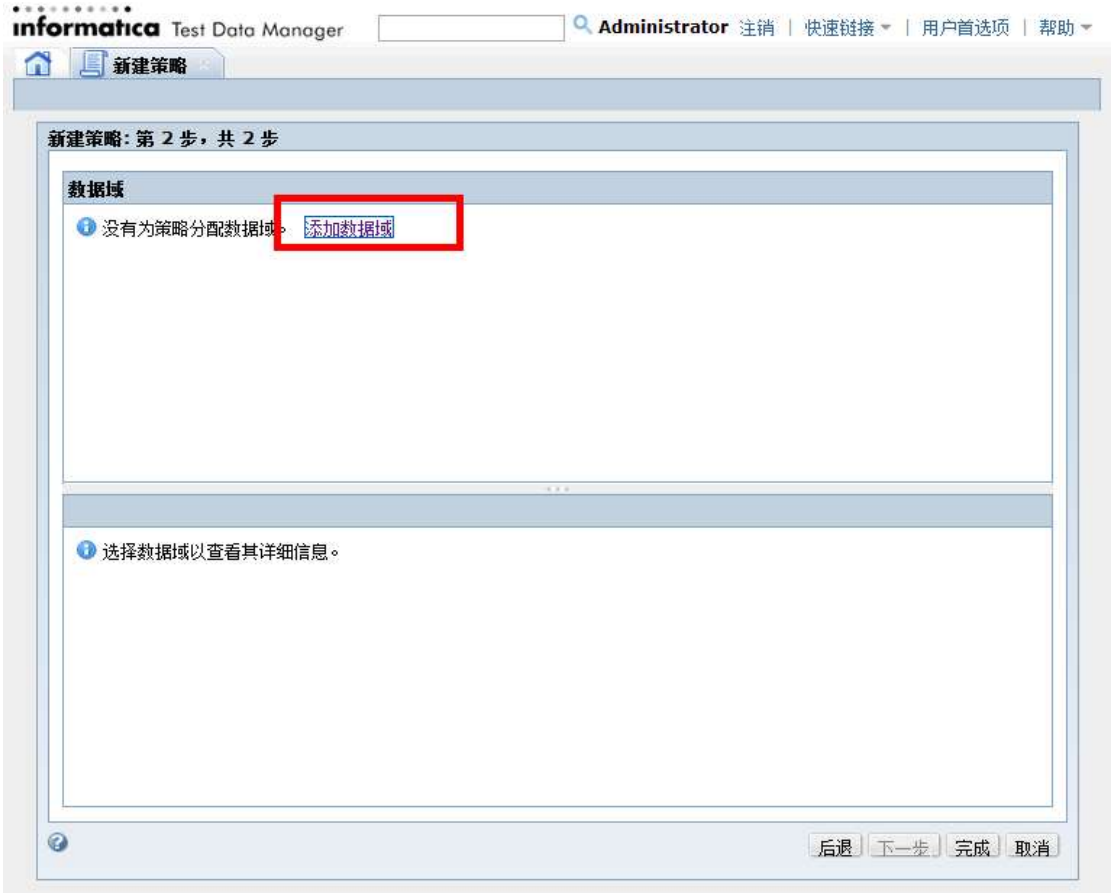
informatica Test Data Manager
Administrator 注销 | 快速链接 | 用户首选项 | 帮助
新建数据域
操作
新建数据域: 第 4 步，共 4 步
规则
未添加规则。
添加规则
属性
选择要查看其详细信息的规则。
后退
下一步
完成
取消

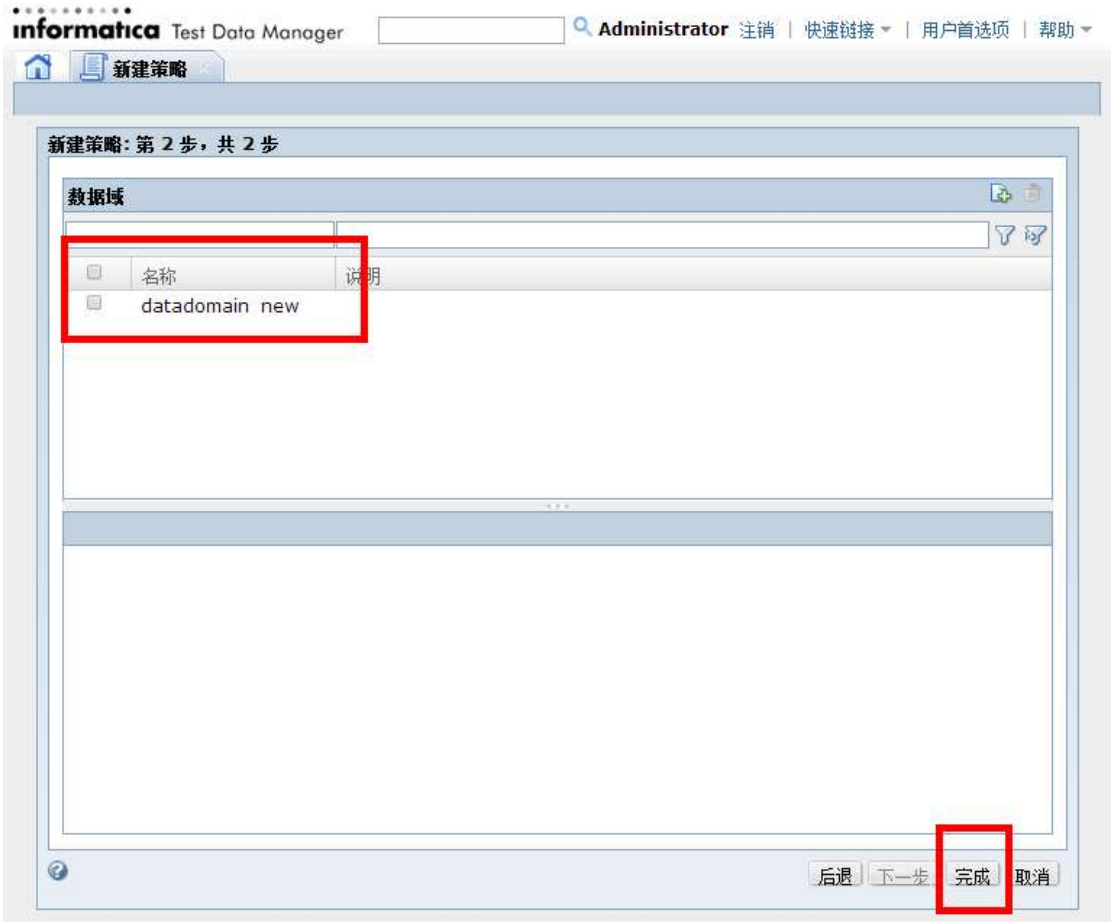
4.3. 创建安全策略

所谓安全策略"Policy"是复数数据域"Data Domain"的集合，主要是将一系列属于同一主题的数据域进行归纳，例如可以为个人信息创建相应的安全策略，其中包括涉及到个人隐私的各种敏感数据域，比如姓名，身份证，电话，住址等。安全策略本身并没有特殊的配置需求，只是数据域的主题化分类


informatica Test Data Manager
Administrator 注销 | 快速链接 | 用户首选项 | 帮助
概览 策略 项目 监视 管理员
操作
概览
新建策略
名称 说明
策略 (4)
plc_cbsmask...
plc_CMSmas...
plc_EBmaski...
plc_test
数据域 (17)
datadomain...
dd_address
dd_address...
dd_address...
dd_cnid
dd_custid
dd_custid_C...
dd_custid_EB
dd_kmm
dd_name
dd_name_C...
dd_name_EB
dd_passwd
dd_phone
dd_phone_C...
属性 数据模式 元数据模式 首选规则 链接的业务术语
常规
数据域分布
数据域名称： datadomain_new


informatica Test Data Manager
Administrator 注销 | 快速链接 | 用户首选项 | 帮助
新建策略
新建策略: 第 1 步，共 2 步
输入策略的名称和说明。
*名称: policy_new
说明:
后退 下一步 完成 取消


informatica Test Data Manager
Administrator 注销 | 快速链接 | 用户首选项 | 帮助
新建策略
新建策略: 第 2 步，共 2 步
数据域
没有为策略分配数据域。 添加数据域
选择数据域以查看其详细信息。
后退 下一步 完成 取消


informatica Test Data Manager
Administrator 注销 | 快速链接 | 用户首选项 | 帮助
新建策略
新建策略: 第 2 步，共 2 步
数据域
名称
说明
datadomain_new
后退 下一步 完成 取消

5.平台安全管理

整个 TDM 平台的业务管理都通过 Workbench 中'**Administrator 标签**'进行配置，但是其后台管理，包括用户及其权限则需要通过 Admin Console 来完成。目前系统中仅有 Administrator(Infa 域管理用户)，该用户拥有整个 PWC 体系的最高权限，因此在实际的使用过程中考虑到有外包人员等参与整个平台的使用 ,因此需要对平台的权限管理进行细分化。其配置入口位于 PWC Admin Console 中的"**Security 标签**"。

在该功能标签中，可以通过"**Action**"-->"**create User/Group/Role**"来创建新的平台内用户，也可导入 LDAP 配置。在左侧导航栏选中某一用户，在右侧的纵览中可以看到该用户的相关信息。

在"Overview 标签"

可以浏览该用户的基本属性，也可编辑，启用，禁用该用户。

在"Privillege 标签"

可以查看并编辑该用户在不同 Infa Service 上所拥有的功能权限。

在"Permission 标签"

可以查看并编辑该用户对不同的"对象的访问权限"，例如是否能够使用某些数据连接等

注 1：为用户授权可以基于特权"Privillege",也可基于角色"Role",其中"Role"是一系列特权的组合，其存在的目的是为了便于授权，系统已经为 TDM 平台预设了部分的"Role",可以在"**Security 标签**"-->"**Role 标签**"中看到所有预制的角色，其中涉及到 TDM 平台的主要包括

- ✧ Test Data Administrator
- ✧ Test Data Developer
- ✧ Test Data Project DBA
- ✧ Test Data Project Developer
- ✧ Test Data Project Owner
- ✧ Test Data Risk Manager
- ✧ Test Data Specialist

6.风险规范管理

TDM 平台提供了足够精细化的平台权限管理，但实际操作上需要和企业内部的的人员配置，流程规范进行配合。此处仅对上述内容进行建设性描述

整个 TDM 平台涉及到的人员角色主要包括如下

平台运维及管理员：

该角色主要负责对 TDM 及后台 PWC 的平台进行日常的运维调控，包括启停服务，用户管理。通常该角色拥有对整个平台较高的管理权限，因此该角色建议有集团内部工作人员担当，样例角色为 Test Data Administrator

业务规则开发人员：

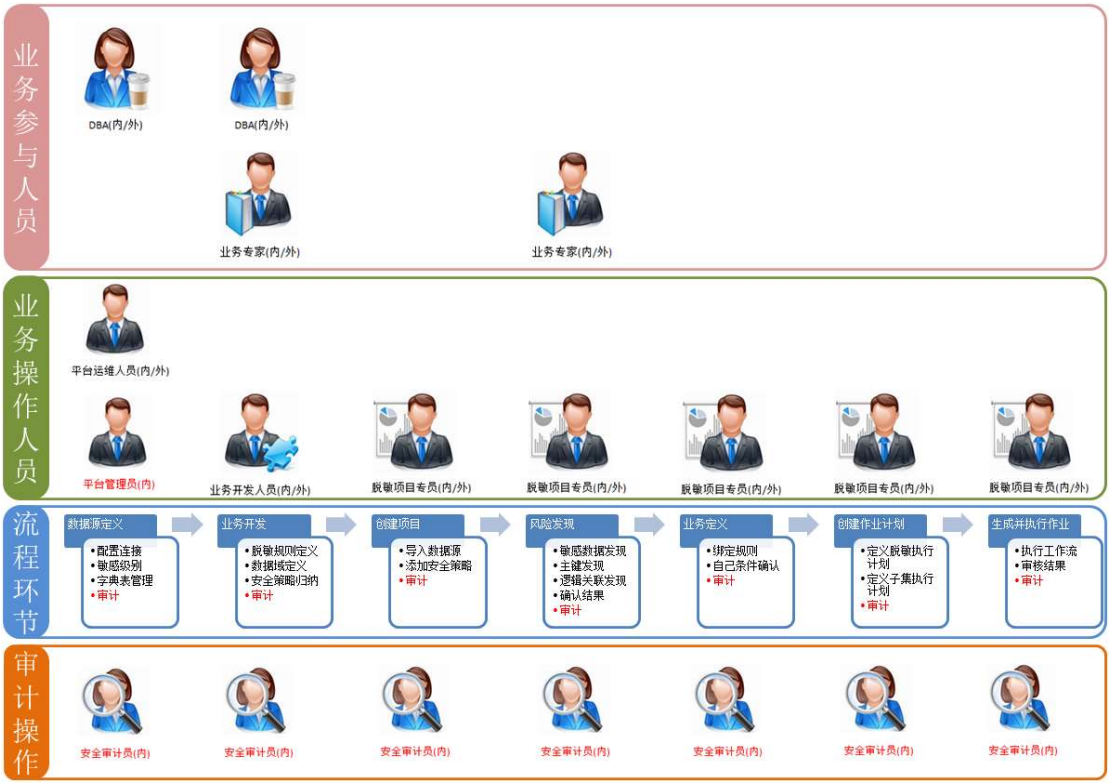
该角色主要负责脱敏规则开发，敏感数据域定义。该角色定位为开发人员，因此可以允许外包技术人员介入，因此在角色的权限分配上要求其仅能掌握规则，数据域的相关权限，而无法实现具体的脱敏作业。避免出现其创建"空白规则"后将原始数据原样导出的风险发生。样例角色为 Test Data Developer

作业执行人员：

该角色可以利用开发人员涉及的脱敏规则和数据域来创建具体的执行计划 ,此类角色建议有集团内部人员担任，其可以确保脱敏作业中不存在所谓的"空白"规则，其样例角色为 Test Data Project DBA

扩展开发人员：

TDM 的大量扩展调优工作需要在 PWC 中进行，因此需要一个额外的角色来承担这一任务，由于 PWC 的使用偏向技术，且本项目中涉及安全，应此不建议由外包人员出任，其样例角色为 PowerCenter Developer



7.数据申请规范

TDM 的核心价值在于以平台工具的方式为集团管理分发测试数据提供技术保障，在实际的使用过程中，还需要与集团内部数据申请分发的流程规范相结合，本章节对集团内数据安全使用的规范制度作简单的建议，这个流程可以分以下几个阶段

申请阶段

首先，整个流程制度的起点在于有相关的业务部门，例如开发组，测试组，培训组需要获取一批数据便于开展工作，因此需要由相关业务部门的负责人向”数据安全专员”发起申请，其申请的方式可以通过邮件或书面报告的方式进行提交。申请报告的内容包括但不限于

- ✧ 数据使用方
- ✧ 业务目的

- ✧ 涉及数据范围
- ✧ 使用起始时间

注 1：由于测试数据的生成分发需要时间，因此数据使用者需要根据自己所请求的数据集的规模，复杂程度等方面有一个预判，以便在提前提交申请，避免延误数据的使用。

注 2：对于所需的数据的涵盖范围，要求申请方能够尽可能的进行详细描述，详尽准确的描述有助于加快数据脱敏处理的进程。

审批阶段

“安全专员”接获申请后需要对相关的申请进行评估，评估内容包括所申请的的数据集规模的合理性，可能涉及的敏感信息的范围，以及整个数据使用的目的。数据使用方属于集团内部还是集团外部也是需要考虑的因素之一。根据评估的结果，”安全专员”需要给出相应审核结果，审核结果除了是否同意该次数据申请业务之外，最核心的内容就是需要应用的脱敏强度。同样的敏感数据在不同的应用场景需要通过套用不同强度的规则，以便在安全性，可用性以及高效率之间取得平衡。

注 1：审核意见还可包括需要多长的的处理周期才能完成数据分发的工作

审核结果会回复前一阶段的申请者以及”脱敏项目专员”

建项阶段

“脱敏项目专员”，在接到批复后，需要寻求 DBA 等相关人员的配合，创建具体的脱敏项目，其工作内容主要包括

- ✧ 项目创建
- ✧ 数据源导入
- ✧ 策略添加

- ✧ 敏感发现（主键，关联）与确认
- ✧ 规则绑定
- ✧ 创建，生成并执行脱敏任务

“脱敏项目专员”完成上述工作后，需要阶段性回复安全专员其工作成果。

注 1：事实上平台内既有的脱敏规则未必能够满足当前项目的需要，此时“项目专员“需要发起一个额外子流程，申请并要求”平台开发人员”开发适配的脱敏规则

审计结项阶段

当脱敏作业完成后，“安全专员”需要通知“审计员”对整个数据脱敏处理流程进行审计，审计内容包括但不限于下列

- ✧ 脱敏规则定义的正确性
- ✧ 字段与数据域绑定的正确性
- ✧ 脱敏作业配置的完整性
- ✧ 脱敏作业结果的正确性

“审计员”需要根据审计结果出据审计报告，特别是当有疑问时，需要明确描述其问题所在

注 1：审计动作可以是每个脱敏项目的必经阶段，也可以采用随机巡查的模式，定期对平台内的各种规则定义，数据域定义进行审查

注 2：审计动作也可安排于建项阶断中“执行脱敏计划”之前，可以避免滥用行为发生（推荐）“”

注 3：审计动作主要是查看 TDM 平台本身是否被滥用，审计平台中是否出现诸如“空白规则”之类的违规设计

额外扩展阶段

本阶段并非每次流程作业必经的环节，仅当平台当前脱敏规则或数据域定义不足并需要扩展开发时出现。该阶段由“脱敏项目专员”发起申请，由“安全专员”审批通过并转发“开发人员”实现，并报请“审计专员”审查开发结果。

可能的业务流程如下

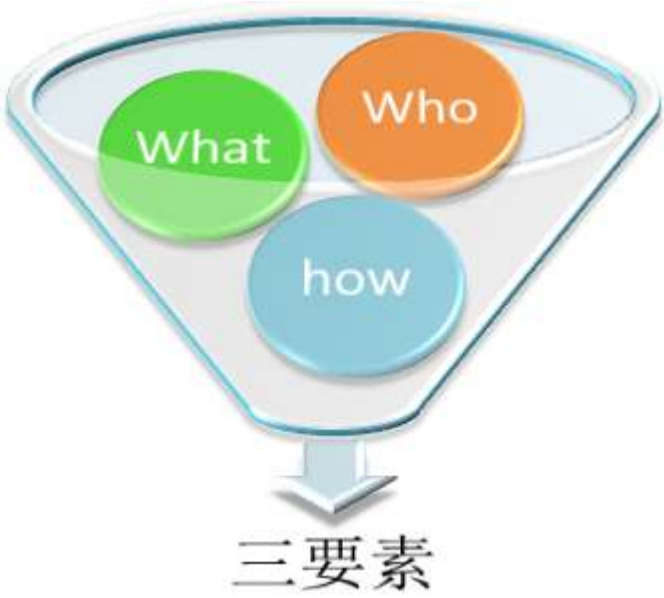
8.集团数据安全需求

随着集团业务系统数量的增多，规模的增大。海量且多样的业务关键数据经过长期的运营后积淀于各种业务数据库中。这些关键数据包括大量的敏感数据，已经以信息的方式成为

集团内部资产的重要组成部分，需要以对待资产的态度来看待其安全访问。然而虚拟数字资产的保护方式与传统资产的保护完全不同，不能简单通过物理安防的手段予以实现。

就目前而言，集团的 IT 建设存在大量的外包服务，无论是开发，测试，甚至系统维护都有大量的外包服务人员参与，其组织架构多样且复杂，在涉及到数据访问的每个环节都是数据安全的风险点。每个环节都会由于管理制度不规范以及技术手段不到位而发生可能的数据安全事故，为集团带来可能的直接或间接损失。因此数据安全已经逐渐成为为集团 IT 部门，特别是其安全部门在日常运作过程中必须考虑的重要需求。

从安全部门的角度来看，所有对于数据的访问都要遵循合理可控的大原则。所谓合理可控其实可以归纳为三大要素，即“谁”，“什么”，“如何”



“**谁**”即是指数据使用者，主要可以从两种维度来考虑

- ✧ 组织：内部和外部人员之分

  显然集团内部人员与外部人员在获取数据是所拥有的“权限”理应有所区别，内部人员属于集团编制，在监管和审计方面相对可控，从职权方面来看也有更多触碰敏感书局的必要。而外部人员组织分散，可管理性较差，因此对其数据访问的需求需要更多的限制

- ✧ 职能：以业务条线来区分

  可以触及到数据的人员按照只能划分又包括，维护组，开发组，测试组，分析组等，个业务职能体对数据本身的要求使不同的。例如：运维人员从职权划分来看仅对数据库具有管理权限，其不应该对数据本身具有访问权限。开发与测试人员需要有数据样本来开展工作，但如何防止其触碰乃至滥用真实数据是安全部门需要考虑的重点。同样的分析组对数据样本则更为关注起统计意义，其对数据的要求也不同于开发测试组。

“**什么**”即是指访问的数据对象

集团内业务数据非常之多，从安全角度来看，数据分为敏感与非敏感，显然对于非敏感数据，在访问控制上可以予以忽略，然而数据的敏感与否并非一成不变，同一数据对于不同的数据使用者而言，其敏感属性是可变的。例如员工薪酬对于一般员工来说是敏感数据，对其访问需要遵循相应的规定，各人只能访问自己的薪酬数据，他人的薪酬数据对其来说就是敏感数据。然而对于 HR 部门来说，由于工作需要，此类数据便不再是需要在 IT 上特殊对待的敏感数据了。

“**如何**”泛指访问数据的方式

访问数据的方式即是指访问途径，通常安全部门可以通过控制访问途径来对数据访问加以限制。但在控制访问方式的时候需要考虑到数据访问这得使用体验。

简单而言就是在数据使用的过程中安全部门可以从上述的的三大要素对整个数据访问的链路加以区分对待，根据不同的访问者，访问对象，访问用途为其设计合理合规的访问链路，从而实现数据安全访问。

数据安全需求实现的最终目的是协助集团实现内外合规，并加强数据访问的管控，是整个集团 IT 安全的重要组成部分。然而，从全局来看，数据的存在意义在于使用和访问，因

此在强调数据安全访问的需求的同时，也需要考虑到数据使用者的使用体验，在不降低数据的可用性的前提下，切实的实现安全保护。这要求在实现需求的时候能够兼顾安全性与可用性，在两者间取得合理的平衡。

基于以上的需求，IT 部门除了协助安全部门建立相应的数据访问和管控制度外需要寻求一种技术手段和平台进行支持，是的制度能够落于实处。

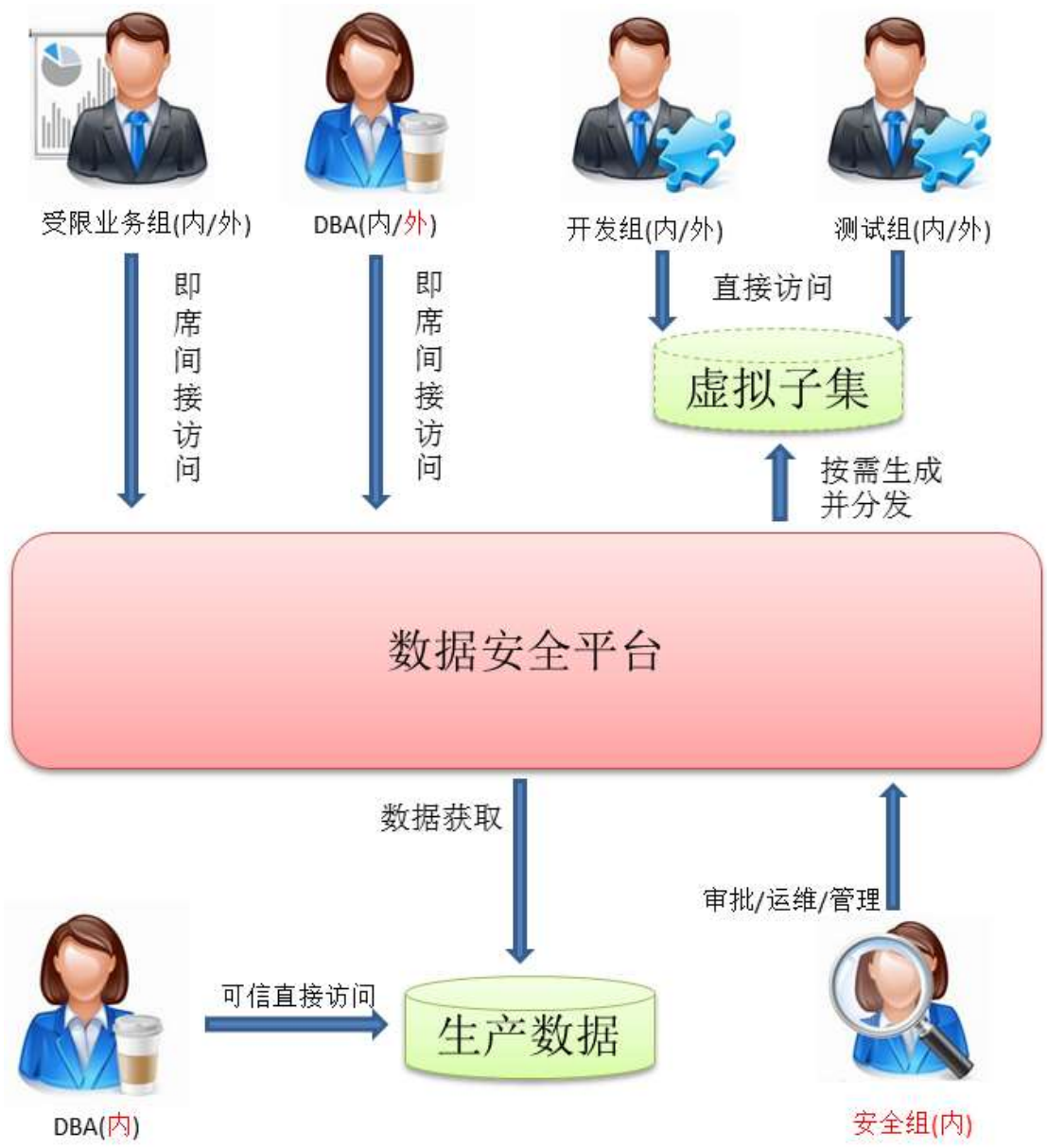
9.数据访问安全制度

部署并结合 Infa TDM&DDM 平台后可以制定相应的数据访问安全制度，TDM&DDM 为制度的落实提供了可行的技术支持。

建议的安全制度可从如下方面实现

9.1. 访问隔离

所谓访问隔离，即是指除了由内部高级可信人员负责的核心运维外，其他业务组，无论是开发组、测试组乃至分析组和外包的运维组均不能直接访问生产数据。其必须通过访问高拟真子集数据或者通过数据安全平台间接访问数据这两种方式开展业务



9.2. 脱敏处理

数据安全平台在为前端应用提供访问隔离的同时，需要为各种访问返回经过加工处理后的去敏感化数据。

对于面向运维分析组类别的间接访问方式，脱敏后的数据，仅需对真实数据进行显示遮蔽性处理，其拟真要求较低

而对于面向开发测试组的数据则要求高拟真，及需要保证数据可读性，关联性，正确性，因此需要通过多种脱敏规则想组合的方式进行加工处理，其要求包括

- 近乎真实
- 保持关联关系
- 保持业务特征
- 灵活、可定制化

通过不同规则的组合可以满足面向内部与外部，以及不同业务组的实际使用需求