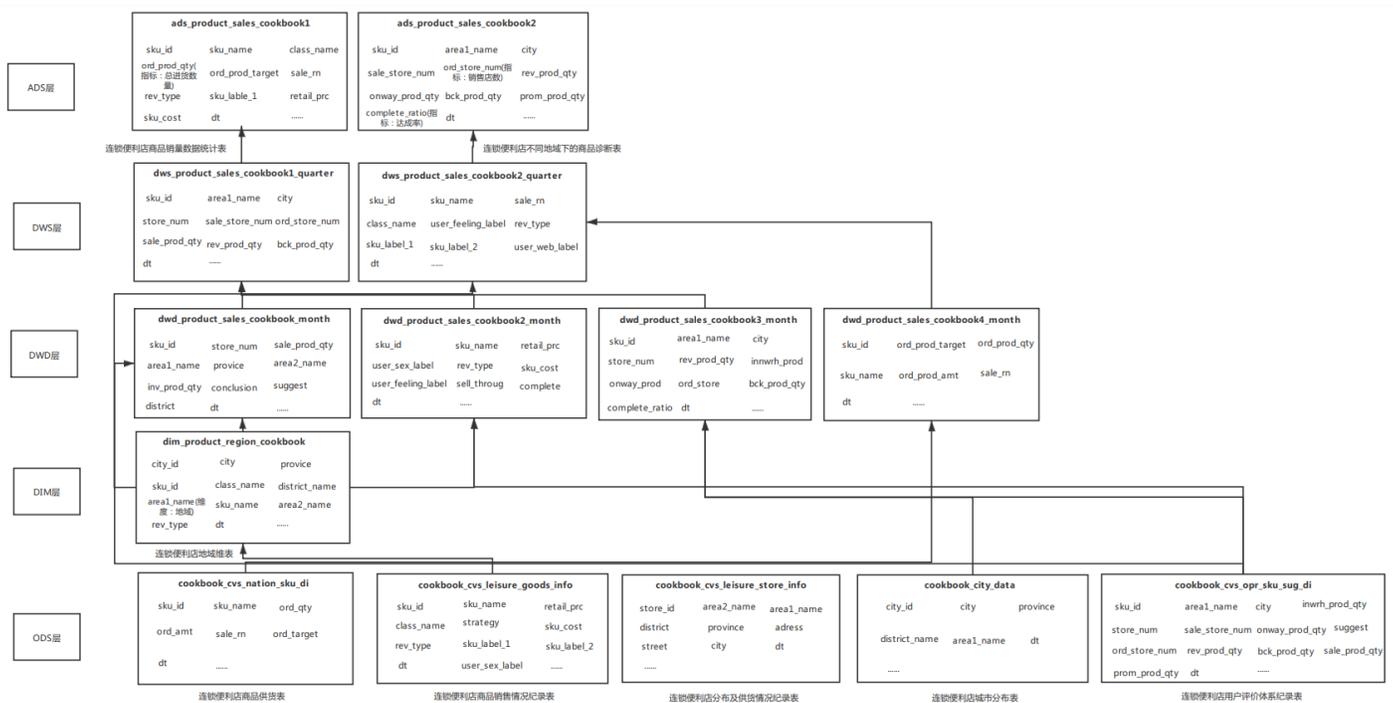


模型设计中心：组织和刻画数据，规范化建设数仓

模型设计中心主要提供组织数据、刻画数据、规范约束和效果衡量的能力，用于实现数仓的规范化建设，同时指引数仓建设的方向，提供主题域、表分层、字典集以及表设计规则管理等功能。

连锁便利店的数据已从业务库集成至平台中，也通过数据测试中心进行了数据表的形态探查。但想要获取销量榜单以及不同地域下的商品门店销售诊断报告等内容，首先需要通过模型设计中心产品，对主题域、表分层以及表设计规则进行组织，对数据进行刻画和规范约束，实现数仓的规范化建模，基于数仓建设，以提供规范化业务数据。

对于当前案例想要达到的目标，在数仓规范化建设过程中，对于DIM—DWD—DWS—ADS各个分层所要创建的规范化表，如下所示(根据不同的案例场景，首先需要确定好各个分层下的表字段内容，结合模型设计中心的产品功能，完成表的规范化设计)，如下图为数仓不同分层下的表字段设计：



在本案例中，结合模型设计中心的产品功能，最终可实现不同分层下，多张表的规范化设计，最终可构建如下表内容：

- dim层表——dim_product_region_cookbook(连锁便利店地域维表)
- dwd层表——dwd_product_sales_cookbook_month(连锁便利店商品供货表)
- dwd_product_sales_cookbook2_month(连锁便利店商品销售情况纪录表)
- dwd_product_sales_cookbook3_month(连锁便利店分布及供货情况纪录表)
- dwd_product_sales_cookbook4_month(连锁便利店用户评价体系纪录表)
- dws层表——dws_product_sales_cookbook1_quarter(连锁便利店商品销量和供货情况宽表)

- dws_product_sales_cookbook2_quarter(连锁便利店不同地域下商品销量和用户评价体系宽表)
- ads层表——ads_product_sales_cookbook1(连锁便利店商品销量数据统计表)
- ads_product_sales_cookbook2(连锁便利店不同地域下的商品诊断统计分析表)

具体实现

下面讲解如何通过模型设计的功能实现上述的规范化表设计目录。

点击左上角的logo，打开导航栏，点击数据治理模块的模型设计中心，进入产品界面(用户需为项目下**数仓开发**角色，才可以访问模型设计中心产品，对于申请开通的试用账号，管理员已默认将用户加入该角色，均可访问模型设计中心产品)。



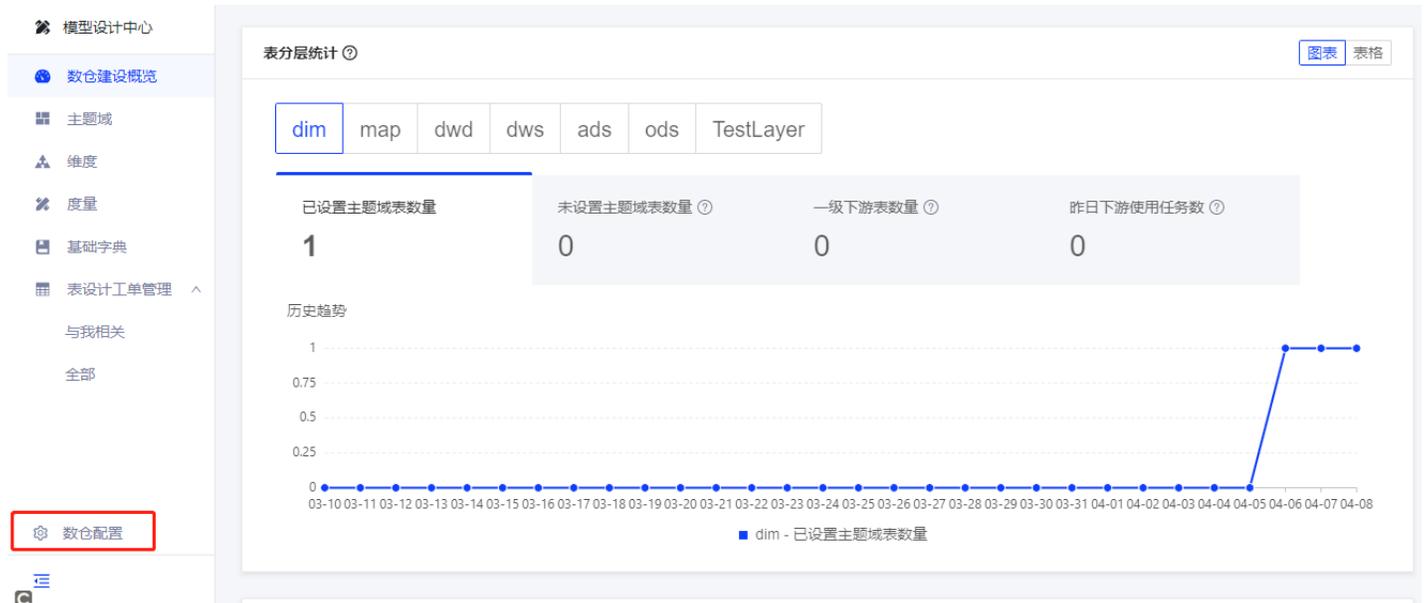
最初需求：

1. 制作**连锁便利店近30天内新品Top10销量榜单**报表，并且可以清楚的了解近30天内Top10中每件商品的具体销售量；
2. 制作不同地域下的商品门店销售诊断报告，可以根据地域、城市等进行筛选查找，了解不同地域下的销量、铺货率、售罄率、达成率等内容。

维度：地域(东北区、华北区、华东区、华南区、西北区和西南区)。

主题域：商品域—销量。

在进行表规则设计之前，需要将所需主题域、分层等内容提前创建好，点击产品界面左下角的**数仓配置**(仅当前项目组负责人或**数据团队管理者**角色有权限查看该功能按钮)，进入数仓配置界面。



在**主题域配置**下，新增一级主题域，弹框中填写主题域名称、英文标识，指定域负责人和域审批人，域负责人有权限在该一级域下创建二级，且对该域有编辑权限，域审批人用来对归属于该域下的工单进行审批。

新增主题域

* 主题域名称: 商品域

* 英文缩写: product

* 域负责人: [选择]

* 域审批人: [选择]

确定 取消

添加时间	操作
2021-03-25 15:32:19	编辑 删除 上移 下移
2021-03-25 15:52:09	编辑 删除 上移 下移
2021-03-26 13:38:02	编辑 删除 上移 下移
2021-04-06 09:55:55	编辑 删除 上移 下移
-	-

在主题域列表中，可以对创建好的主题域进行编辑、删除、上移和下移操作，上下移动操作会影响主题域在模型设计中心的展示顺序。

主题域名称	英文缩写	域责任人	域审批人	添加人	添加时间	操作
商品域	product	MURDO			2021-04-09 16:17:41	编辑 删除 上移 下移
订单主题	order				2021-03-25 15:32:19	编辑 删除 上移 下移
乘客主题	passager				2021-03-25 15:52:09	编辑 删除 上移 下移
保险	insurance				2021-03-26 13:38:02	编辑 删除 上移 下移
法务主题	FW				2021-04-06 09:55:55	编辑 删除 上移 下移

在**分层配置**下，系统提供预置分层，包括dim、map、dwd、dws、ads和ods，系统预置的分层可以选择启用或者不启用，也可以选择暴露或者不暴露(是否暴露并不影响模型设计中心的产品功能，仅针对数据地图产品，是否决定将当前分层对所有用户开放的设置项)，也可以开启自动审批，开启后，归属于该分层下的工单将自动通过。

分层名称	英文标识	描述	启用状态	显示控制	表抓取规则	允许建表的库	自动审批	操作
维表	dim	维度数据	启用	暴露	-	-	关闭	编辑 禁用 不暴露
桥接层	map	用于存放桥接维度的关系表	启用	暴露	-	-	关闭	编辑 禁用 不暴露
明细层	dwd	从ods层经过ETL得到的明细数据，表示具体的事实，主要由维度和度量等构成	启用	暴露	-	-	关闭	编辑 禁用 不暴露
汇总层	dws	由明细数据经过汇总得到的数据，主要由统计维度和指标构成	启用	暴露	-	-	关闭	编辑 禁用 不暴露
应用层	ads	由明细或汇总层加工的出来面向报表、服务、应用等使用的数据	启用	暴露	-	-	关闭	编辑 禁用 不暴露

若系统提供的预置分层不满足需求，也可以自定义分层。自定义分层时，可以设置抓取规则，对于没有设置分层的表，若符合抓取规则，则可以抓取归属到自定义的分层中，比如自定义分层为adm，抓取规则为bdms.adm_%，表示bdms库下以adm开头的表均归属于adm层，创建完成之后，可以在对应的产品界面查看抓取到的表信息(模型设

计中心—主题域—未分组列表，点击立即抓取，会呈现对应符合抓取规则的表分层下的表信息)。

自定义分层

+ 新建自定义分层

分层名称	英文标识	描述	启用状态	显示控制	表抓取规则	允许建表的库	自动审批	操作
测试层	TestLayer	test	启用	暴露	命中规则: db1.at _ % 删除规则:	-	关闭	编辑 禁用 不暴露 删除

在字典集Tab下，可以新建字典集，比如对日期的后缀统一标准，day表示每日，hour表示每小时等，点击+新建字典集，填写名称和描述等内容。字典集创建完成之后，在模型设计中心主页面的基础字典中可以针对该字典集添加字典内容。

新增字典集

字典集类型: 简单字典集

* 字典集名称: 日期后缀

描述: 对日期的后缀进行统一定义和管理

15/256

确定 取消

在**表设计规则管理**下，可以针对各个分层下表的设计规则进行设置，包括表的命名、表的分区生命周期、默认的存储格式等内容，若开启强制分区表，则归属于该分层下的表必须为分区表。

表分层	表名规则	表属性默认项配置	强制分区表
维表 - dim	无 设置	Impala同步：是 表类型：内部表 分区生命周期：永久 表生命周期：永久 存储格式：Parquet 编辑	<input checked="" type="checkbox"/>
桥接层 - map	无 设置	Impala同步：是 表类型：内部表 分区生命周期：永久 表生命周期：永久 存储格式：Parquet 编辑	<input checked="" type="checkbox"/>
明细层 - dwd	无 设置	Impala同步：是 表类型：内部表 分区生命周期：永久 表生命周期：永久 存储格式：Parquet 编辑	<input checked="" type="checkbox"/>
汇总层 - dws	无 设置	Impala同步：是 表类型：内部表 分区生命周期：永久 表生命周期：永久 存储格式：Parquet 编辑	<input checked="" type="checkbox"/>

点击表名规则下的**设置**，可以针对当前分层下的表命名规则进行设置，包括设置表的前缀和后缀，均可以选择固定字符串、主题域缩写、二级主题域缩写、字典集或是正则表达式，对于常见的配置，系统也给出参考**配置建议**，可以灵活调整各个分层下的表命名规则。在本示例中，设置ads应用层的表，前缀为固定字符串ads，后缀为二级主题域的缩写，系统会根据用户选择的二级主题域自动填充表名后缀。

编辑表名规则 - 应用层

- 1) 表名拼接规则为：f1_f2_f3_f4_f5_表名主体_s1_s2_s3。
- 2) 表名仅支持小写，对于正则表达式请按小写设置。
- 3) 字典集仅支持选择“简单字典集”。
- 4) 常见的配置可参考 [配置建议](#)

前缀 前缀 控制逻辑

f1 固定字符串 ads

+ 添加

后缀 后缀 控制逻辑

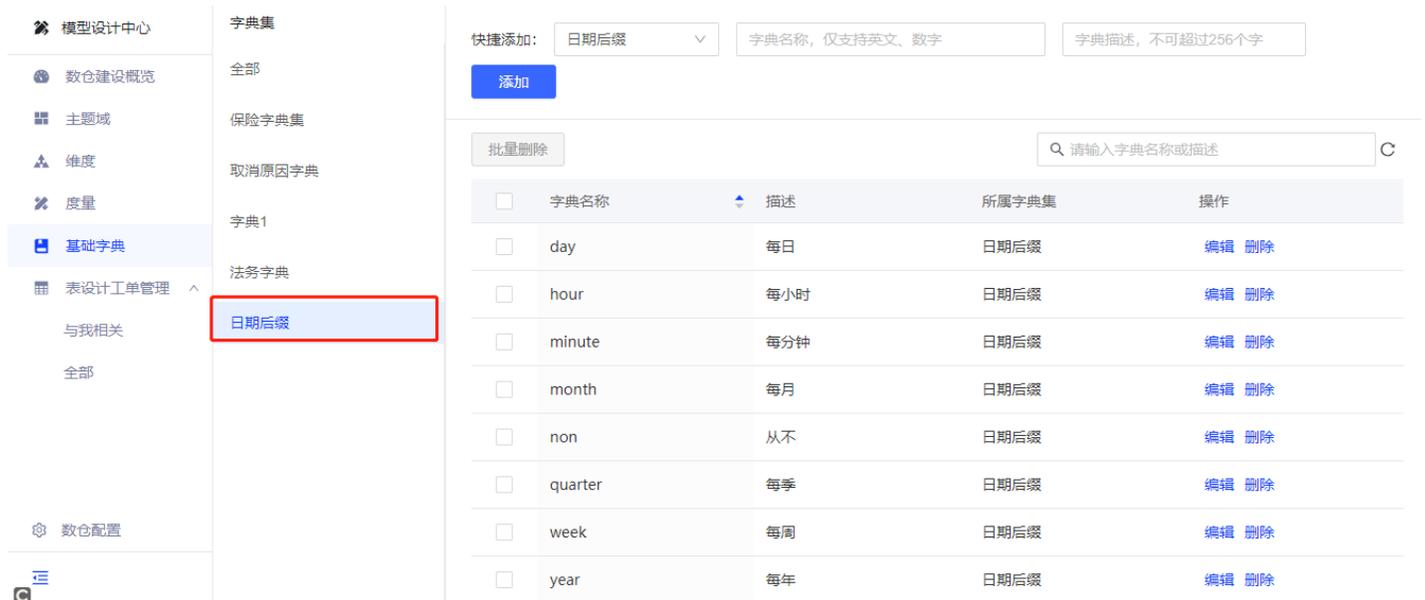
s1 二级主题缩写 表名会按照用户选择的二级主题自动替换

+ 添加

建表说明 用于展示给建表人员的说明性文字，比如未采用基础字典配置的后缀枚举值说明、表名主体的命名建议等，从而指导建表人员设置合理的表名

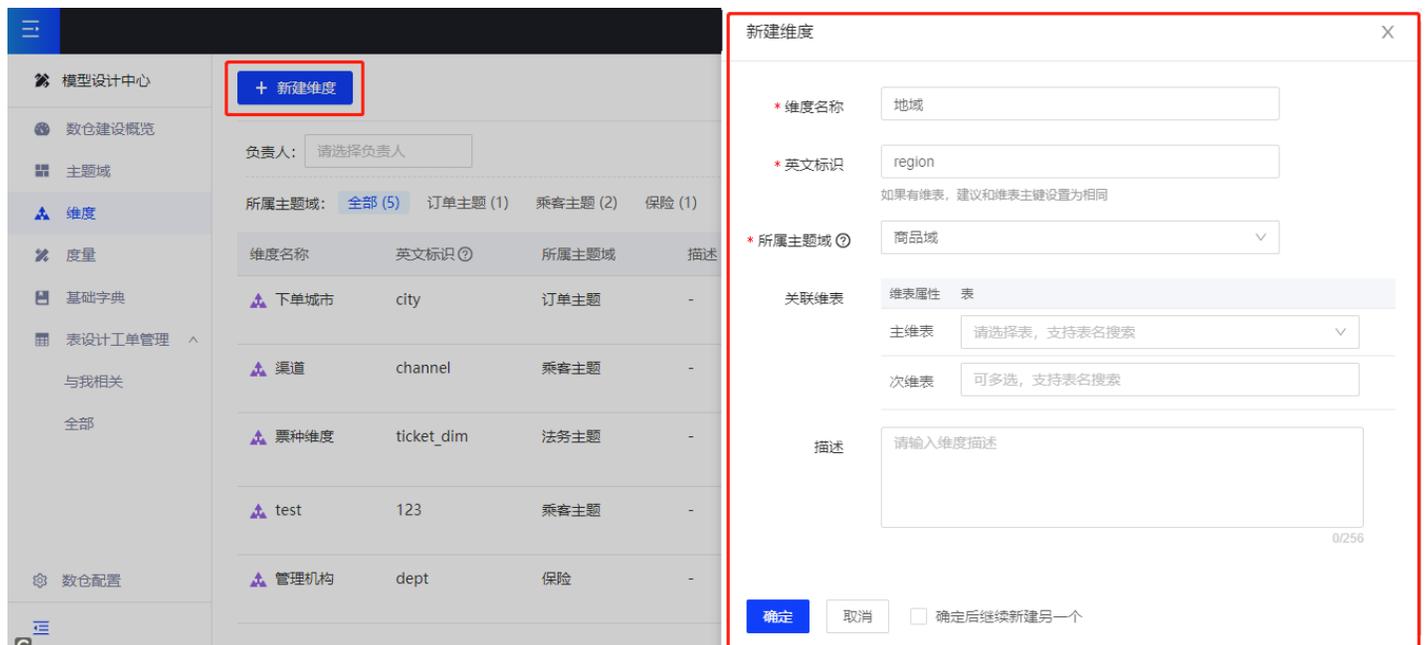
以上，以项目组负责人(或数据团队管理者)的身份创建完了一级主题域、分层、字典集以及各个分层下的表设计规则，下面以数仓开发的角色进入模型设计中心主页面。

在**基础字典**下，针对已经创建的**日期后缀**字典集添加字典，填写字典的名称和描述，点击添加，即在当前选中的字典集下快速创建字典，添加完成后，可以对字典进行编辑和删除等操作(当前只有项目组管理员或**数据团队管理者**角色或项目负责人或项目管理员可以添加字典内容，其他角色可查看并使用已创建好的字典集)。



字典名称	描述	所属字典集	操作	
<input type="checkbox"/>	day	每日	日期后缀	编辑 删除
<input type="checkbox"/>	hour	每小时	日期后缀	编辑 删除
<input type="checkbox"/>	minute	每分钟	日期后缀	编辑 删除
<input type="checkbox"/>	month	每月	日期后缀	编辑 删除
<input type="checkbox"/>	non	从不	日期后缀	编辑 删除
<input type="checkbox"/>	quarter	每季	日期后缀	编辑 删除
<input type="checkbox"/>	week	每周	日期后缀	编辑 删除
<input type="checkbox"/>	year	每年	日期后缀	编辑 删除

在**维度**下，针对本案例中的分析维度**地域**(东北区、华北区、华东区、华南区、西北区和西南区)进行创建。维度创建完成后，可以对维度进行筛选、编辑和删除等，若维度被表所引用，也可以在维度列表中查询当前维度的引用详情。



在**主题域**下，针对已经创建的一级主题域**商品域**创建二级主题域**销量**(只有一级主题域的域负责人才有权限在该域下创建子主题，填写主题名称和英文标识，可以选择主题负责人为同父级或者自定义选择其他成员为子主题的负责人)。



子主题域创建完成之后，可对主题域进行编辑、删除，也可以再创建二级主题域下的子主题域，最多可以创建至四级主题域。



以上有关主题域、维度、字典等内容创建完成后，就可以规范化设计表了。在**表设计工单管理**的**与我相关**下，可以创建建表工单，来规范化设计各个分层和主题域下的表。点击**+创建建表工单**，选择表分层、库和主题归属，根据在**数仓配置**的表设计规则管理中设计的表名规则选择或填写完善表名称。



基本信息配置完后，在下一步可以设计表的字段、分区、业务以及基础属性等内容。下图为本示例设计的dwd层表的基本属性的部分字段，其中，area1_name字段表示地域内容，和维度相关，在**标准化标签**一栏，可以选择维度，然后将该字段与之前创建好的“地域”维度关联起来。创建完成后，点击提交，需要该表所归属的**商品域**的域审批人进行工单审批(若在数仓配置中，分层配置中开启了自动审批，则归属于该分层下的工单会自动通过)。

模型设计中心 < 创建建表工单

表名 ———— 2 基础属性及字段

字段配置

+ 添加字段

#	字段名称	类型	描述	主键	标准化标签	操作
1	sku_id	STRING	商品id	<input type="checkbox"/>	维度 度量 请选择标准化标签	删除 上移 下移
2	area1_name	STRING	地域名称	<input type="checkbox"/>	维度 度量 地域	删除 上移 下移
3	city	STRING	城市	<input type="checkbox"/>	维度 度量 请选择标准化标签	删除 上移 下移
4	area2_name	STRING	地域别名	<input checked="" type="checkbox"/>	维度 度量 请选择标准化标签	删除 上移 下移

上一步 提交 保存

在**表设计工单管理**的**与我相关**页签下，可查看工单状态(在**数仓配置**页面中**主题域**配置下，指定域审批人，负责审批归属于该域下的工单)，支持快速查看表的DDL语句，也可以在**更多**下快速复制DDL语句。

模型设计中心 未提交 0 待审批 0 被驳回 0 待我审批 0 已通过 1 + 创建修改表工单 + 创建建表工单 不合理命名表清单

表名称 请输入工单编号、表名称

工单编号	工单类型	工单操作对象	对象配置模式	工单状态	工单信息	操作
9	新建Hive表	dwd.dwd_product_sale_s_cookbook_month	表单模式	已通过	创建人: y... 审批人: ... 提交备注: cookbook的dwd层表 审批意见: 自动审批通过	查看DDL 更多

与我相关 全部

在模型设计中心的建表工单通过之后，系统会自动在当前项目下，对应的库下创建表，比如本示例中在dwd库下申请创建的dwd_product_sales_cookbook_month表，工单通过之后，可以在**主题域**下，查看已经创建成功的表。

#	库.表名称	表负责人	表描述	主键	表分层	主题归属	操作
1	dwd.dwd_product_sales_cookbook2_month		cookbook示例中dwd层的表2	sku_id	dwd	商品域/销量	查看核心属性 变更主题归属 取消当前域关联
2	dwd.dwd_product_sales_cookbook2_month		cookbook中dwd层表的分析	area2_name store_id	dwd	商品域/销量	查看核心属性 变更主题归属 取消当前域关联
3	dwd.dwd_product_sales_cookbook3_month		cookbook示例下dwd层的表3	city city_id	dwd	商品域/销量	查看核心属性 变更主题归属 取消当前域关联
4	dwd.dwd_product_sales_cookbook4_month		cookbook示例中dwd层的表4	sku_id	dwd	商品域/销量	查看核心属性 变更主题归属 取消当前域关联

在本示例中，以上面同样的方式，设计并创建以下各个分层的表：

- dim层表——dim_product_region_cookbook;
- dwd层表——dwd_product_sales_cookbook2_month、dwd_product_sales_cookbook3_month、dwd_product_sales_cookbook4_month;
- dws层的表——dws_product_sales_cookbook1_quarter和表 dws_product_sales_cookbook2_quarter;
- ads层的表——ads_product_sales_cookbook1、ads_product_sales_cookbook2。

其中，在dwd层，可以将表字段与度量关联，在dws和ads层，对于表字段可以关联指标。本示例中在表ads_product_sales_cookbook1的ord_prod_qty、ord_prod_amt等字段中，关联了在指标系统已经创建的指标 总进货数量、实际商品销售量 等(对于已经创建完成的表，若也想将字段与标签，即指标、度量和维度关联，可以在数据地图

的表详情页面，针对字段进行编辑，关联标签化标签，前提是当前表已归属于某个主题域和表分层下)，如下图所示，创建表工单的第二步，选择指标，下拉框中选择对应的指标即可。



以上，完成了在模型设计中心对于各个分层的规范性表设计，基于系统已经创建成功的各个分层下的表，接下来就可以在离线开发中心构建数仓任务，对数据进行ETL加工和处理，完成任务的开发和提交上线等流程。