

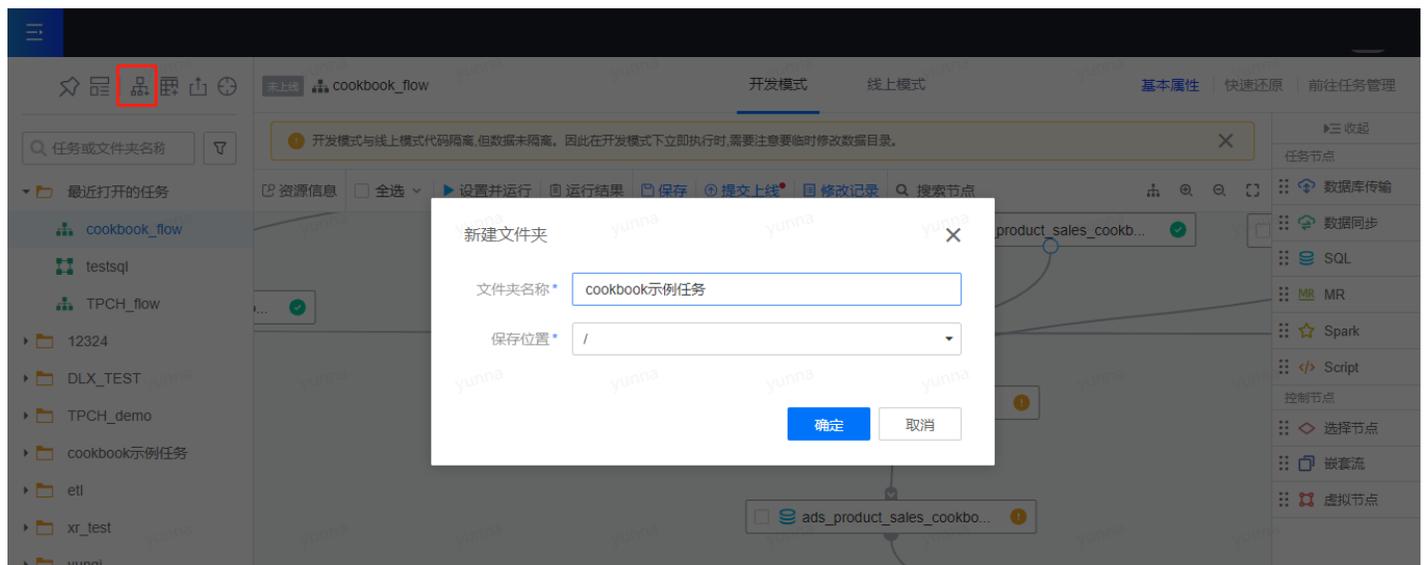
离线开发：构建工作流，完成敏捷性数据开发

离线开发通过拖拉拽的方式快速构建任务流，支持数据同步、SQL节点、MR、Spark、Script、AI等任务节点，支持选择节点、嵌套流、虚拟节点控制节点，实现敏捷性的数据开发。

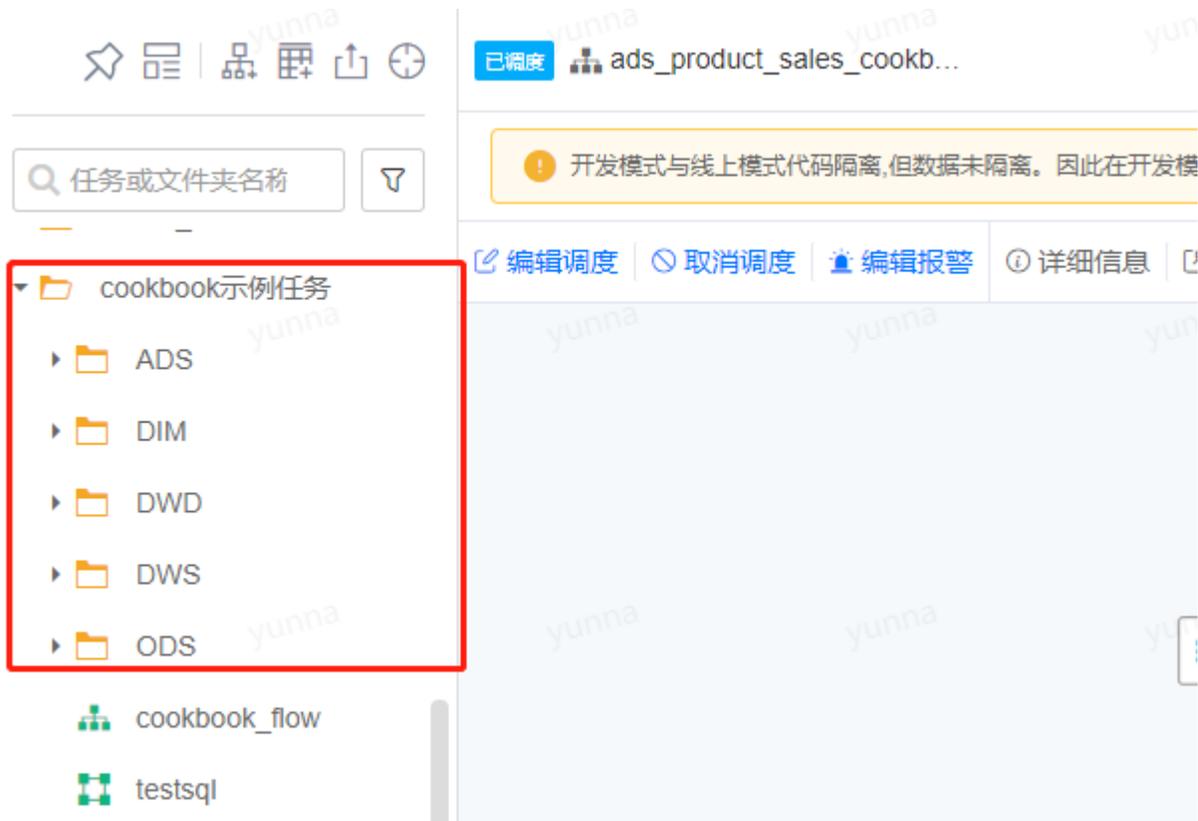
连锁便利店的数据已通过模型设计中心完成了对主题域、表分层和表设计规则的组织，生成了规范化设计的数据表，且也通过数据测试中心对数据进行了初步的探查工作，接下来可以基于已经生成的规范表，在离线开发中执行ETL加工过程。打开导航栏，点击**离线开发**进入产品页面。



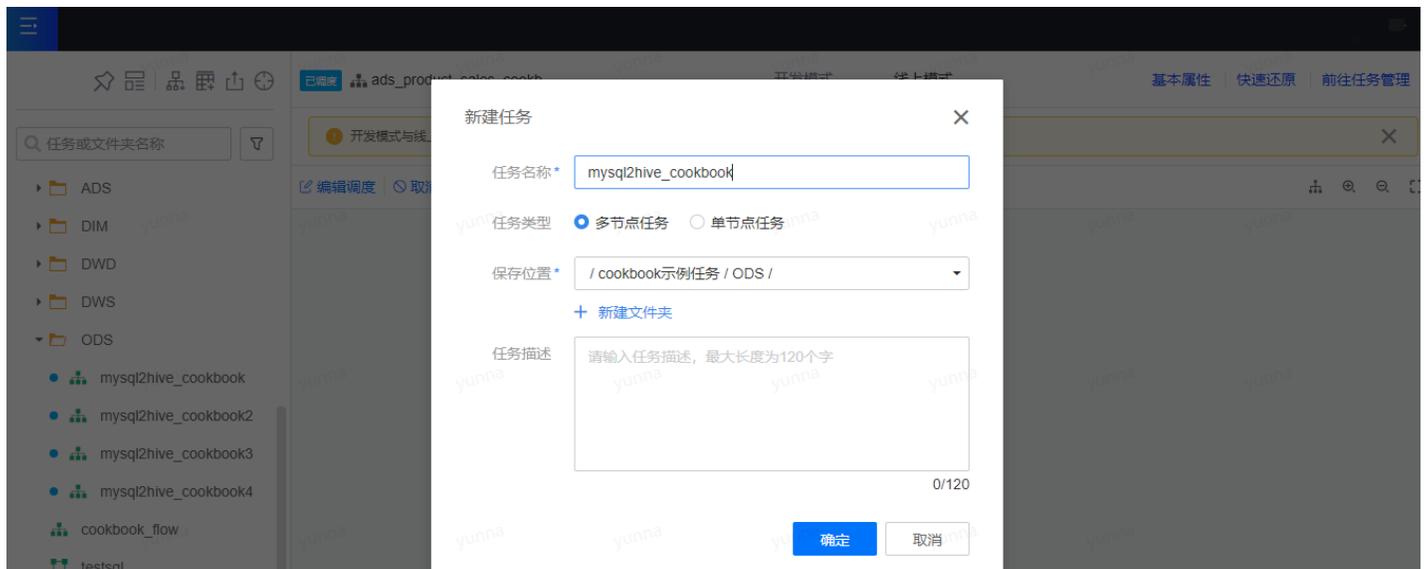
在左上角点击新建文件夹图标，本示例中文件夹名为**cookbook示例任务**，确认创建后，在离线开发页面左侧，呈现创建好的文件夹。



使用同样的方式, 创建ODS、DWD、DWS、DIM和ADS文件夹, 保存位置选择**cookbook示例任务**, 用于分类管理不同层级的数仓开发任务。



文件夹创建好之后, 点击新建任务, 填写任务名称, 以ODS层的任务为例(数据传输的任务命名, 建议采用**{源端}2{目的端}**的方式, 比如mysql2hive表示将Mysql库下的数据导入Hive数据源中), 本案例中任务名称命名为mysql2hive_cookbook, 保存位置选择**cookbook示例任务**下的**ODS**文件夹, 任务类型选择多节点任务。



任务创建成功后，在画布正中间会呈现游离节点mysql2hive_cookbook，从右侧选择数据同步节点(具体使用的节点类型根据实际情况有所变动，本示例中通过数据同步节点完成数据集成任务)，将节点从右侧拖拽到画布中间，选中节点，右键选择**重命名**，对节点进行命名(节点任务名称建议为ods_{source_table})。



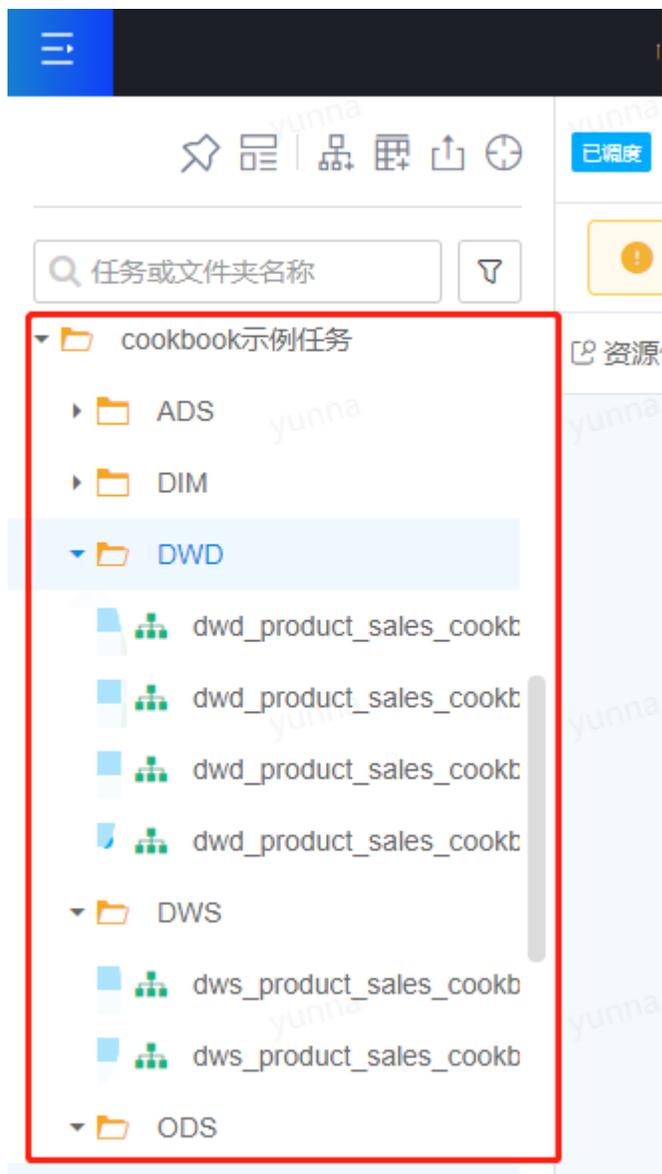
对于拖拽到画布中的节点，通过连接点与另外的节点关联，从而构建关联关系，以同样的方式，拖拽多个节点构建整个工作流。



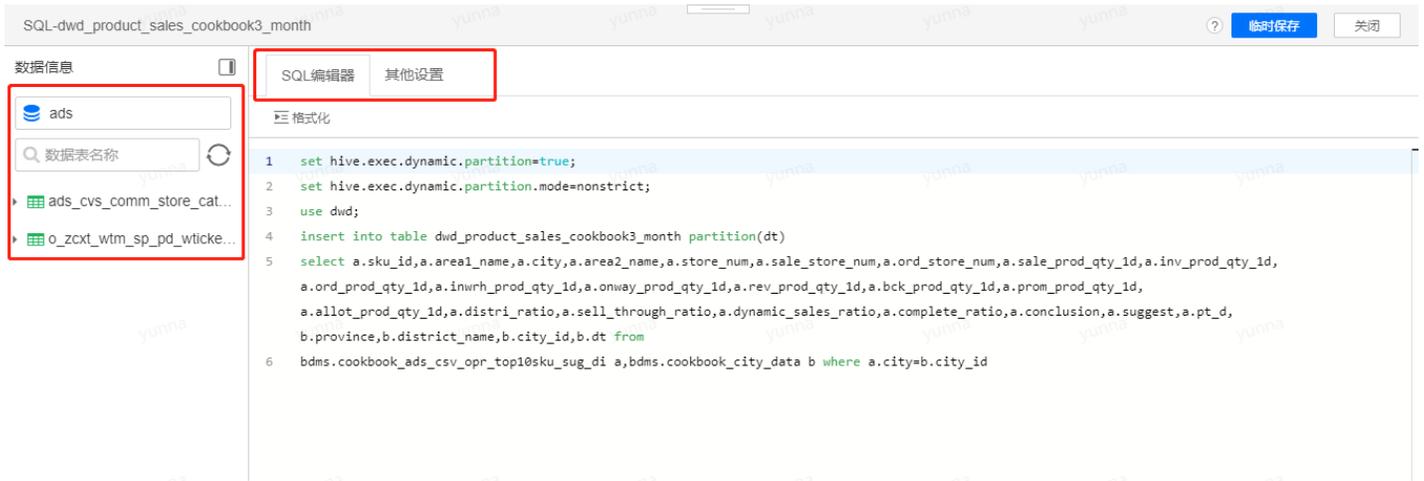
对于数据同步节点，双击打开节点，需要选择在**数据集成**模块创建的数据传输任务，比如ods_city_data该节点选择cookbook_city_data传输任务。



以同样的方式，创建ODS-DIM-DWD-DWS-ADS文件夹，在不同的文件夹下创建对应的工作流，在**cookbook示例任务**文件夹下呈现各个分层下的多个任务流。



对于SQL节点任务，双击打开节点，在SQL编辑器下书写语句，在**其他设置**页签下可以选择使用Hive或者Spark引擎执行。同时，在左侧还可以根据库表筛选定位数据表，查看表的字段信息，方便在写SQL语句时对表字段内容进行查询。



任务创建完成之后，可以选择任务流中某个节点或者选择全部，在开发模式下点击**设置并运行**，在未提交上线之前先验证任务是否可成功运行，点击**运行结果**，可查看选中的节点任务的执行情况，支持查看日志、运行结果等内容。



点击**修改记录**，可查看当前任务流在上次提交上线之后对于任务的修改详情(仅展示保存及对资源的相关操作)，包括修改人、修改时间和对应的修改内容。



详细信息

仅展示上次提交上线之后的保存及资源相关操作

序号	操作人	操作时间	操作类型	操作内容 ?
1	...	2021-04-13 13:47:56	保存	删除节点: 无 修改节点: dws_product_sales_cookbook2_quarter
2	...	2021-04-13 13:40:09	保存	删除节点: 无 修改节点: dws_product_sales_cookbook1_quarter
3	...	2021-04-13 11:39:48	保存	删除节点: 无 修改节点: dws_product_sales_cookbook1_quarter
4	...	2021-04-13 11:36:20	保存	删除节点: dws_1,dws_2 修改节点: dws_product_sales_cookbook1_quarter,ads_1,dws_product_sales_cookbook2_quarter,ads_2

若开发模式和线上模式下的任务流存在不一致，**提交上线**的功能按钮会有红点提示，建议将开发模式下的任务测试成功后，及时提交上线运行。



开发模式 线上模式

开发模式与线上模式代码隔离,但数据未隔离。因此在开发模式下立即执行时,需要注意要临时修改数据目录。

保存 提交上线 修改记录

当前开发模式和线上模式不一致,请及时提交上线

```

graph TD
    A[ods_cvs_leisure_goods_inf...] --> B[ods_cvs_opr_sku_sug_di]
    A --> C[mysql2hive_cookbook2]
    B --> C
  
```

点击**提交上线**，平台会针对当前任务流进行检测，给出结论，针对没有测试的任务节点也会罗列出来，每一次提交上线，均需要填写上线备注，便于数据开发后续查看每个上线版本所修改的内容(针对任务流是否需要强制走工单，可以在项目配置中的**离线开发提交控制**的强制走工单审批设置中进行配置)。



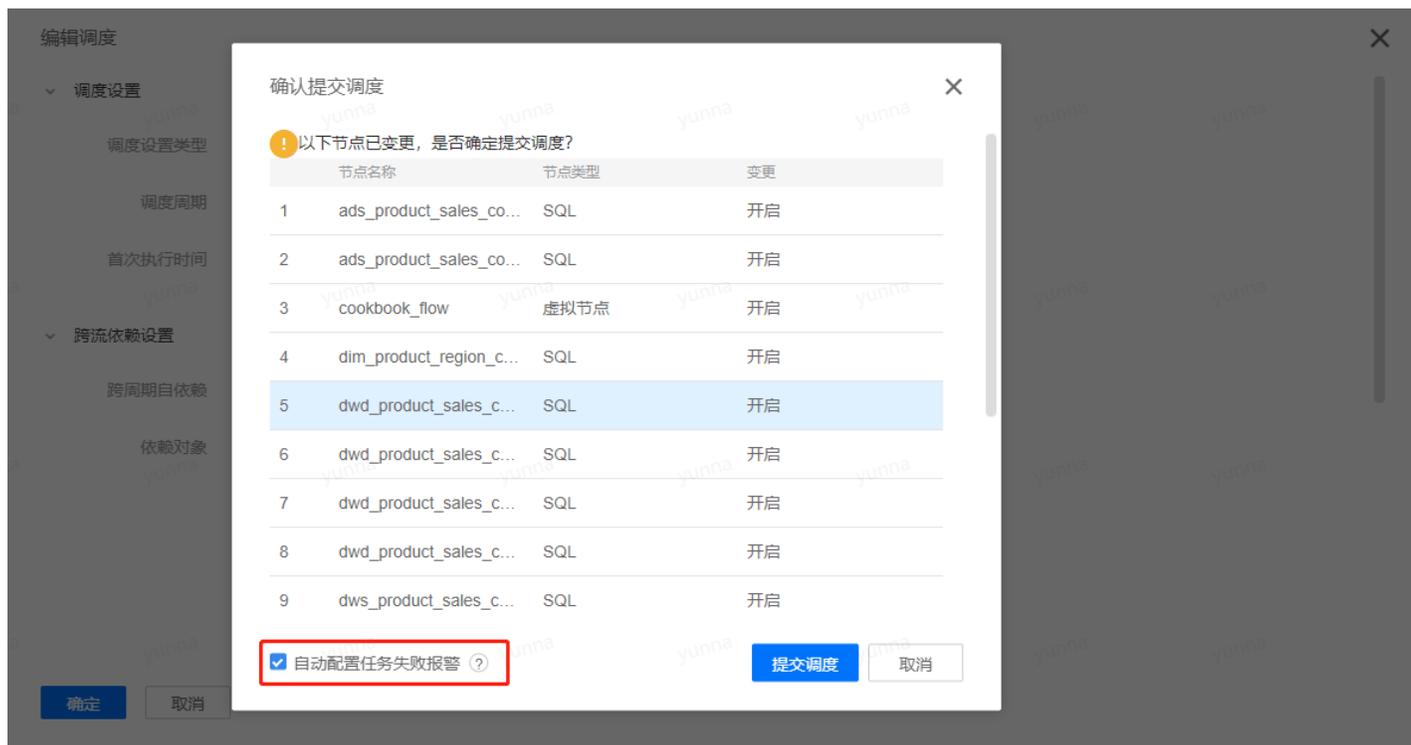
任务提交上线后，会从开发模式被提交至线上模式，点击**线上模式**，可以针对已经提交至线上模式的任务编辑调度，可以选择将该工作组所有节点任务均提交上线，也支持只选择部分节点提交上线。



点击**设置并提交调度**，会弹窗对任务提交上线的调度内容进行设置，包括设置调度周期、跨流依赖设置(是否依赖别的任务或者节点)以及任务执行设置(包括执行队列、失败设置和并发设置等内容)。调度设置中支持常规设置和crontab设置，选择crontab设置后，页面会切换为crontab语法，支持填写分钟、小时等灵活配置调度周期时间。



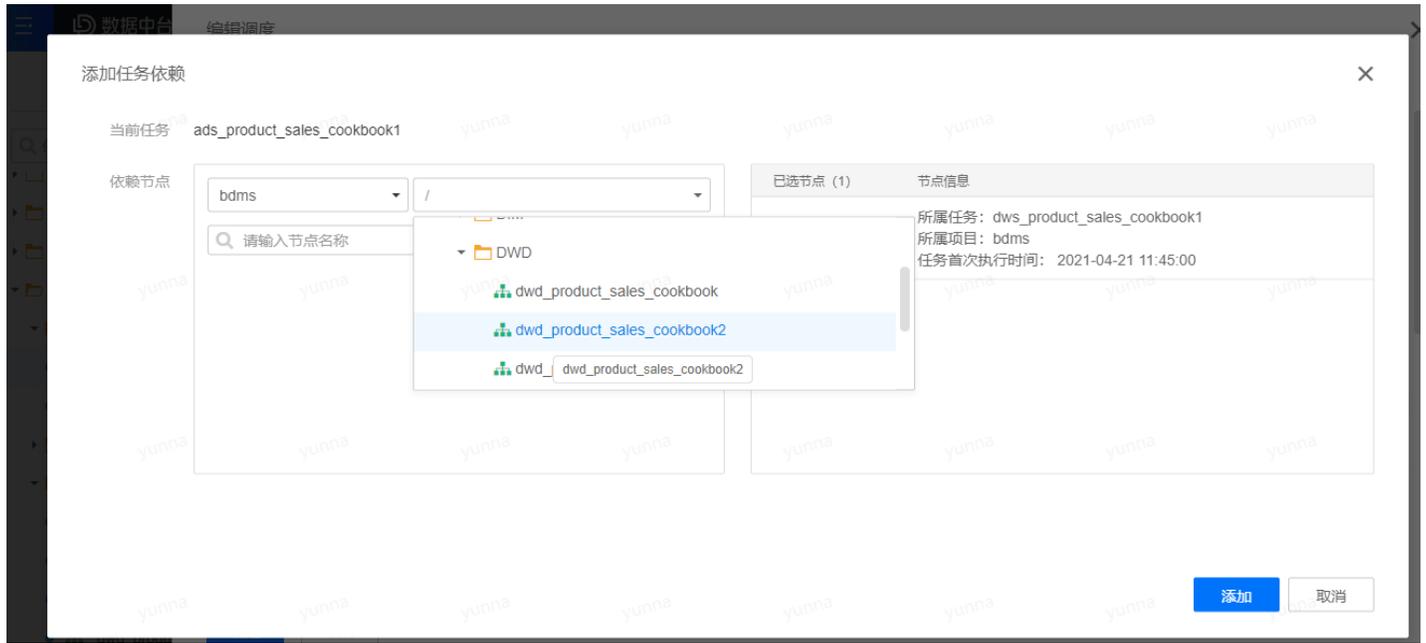
本示例中我们以常规设置为调度配置，调度周期为1天，选择首次执行时间后，点击**确定**。如果节点有变更，页面会弹窗提醒**是否需要继续提交调度**。同时，平台默认对提交上线的节点任务自动配置失败报警，若节点任务失败，会通知任务的创建人，可根据需求取消自动报警配置功能。



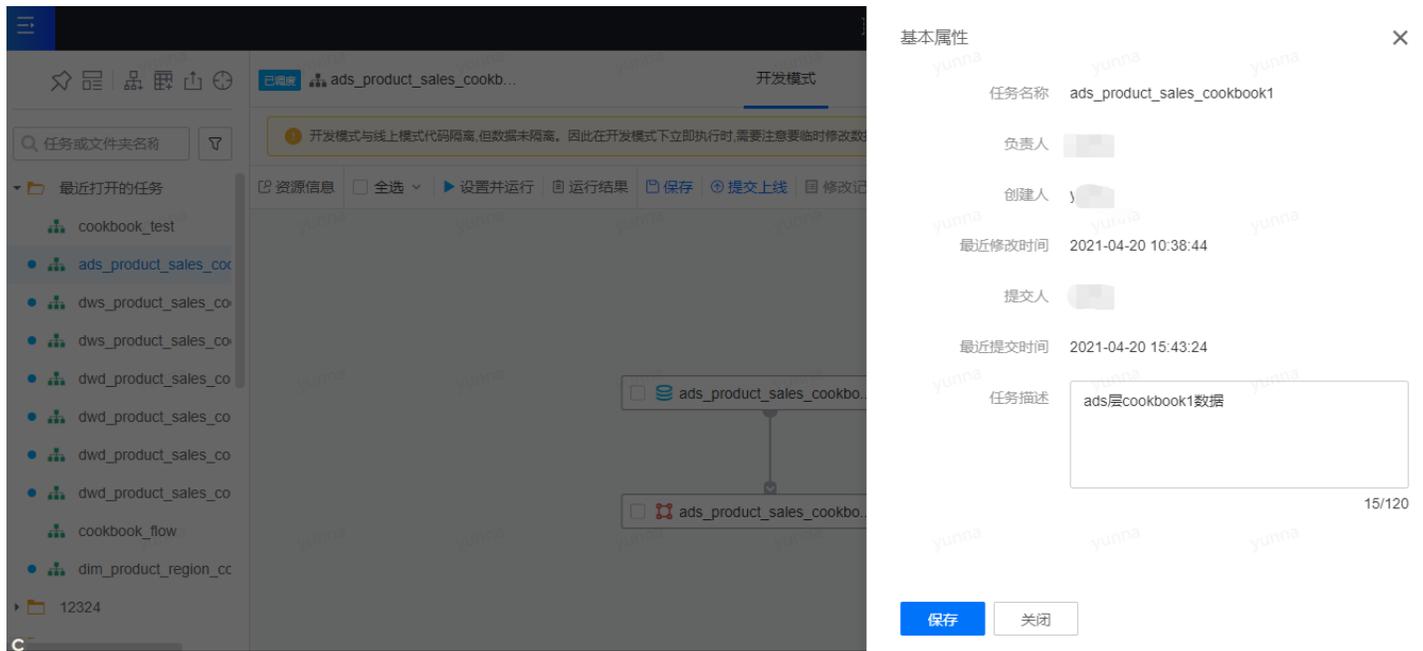
点击弹窗中的**提交调度**，任务被提交成功。至此完成了在离线开发产品中对任务的初步ETL加工以及任务调度，接下来可以在任务运维中心查看任务的执行情况。

除了ODS层的任务流，针对DIM-DWD-DWS-ADS层级的任务流，需要添加任务依赖，在设置调度的弹框中，**跨流依赖设置**下点击**添加任务依赖节点**，选择依赖节点的名称，点击**添加**，在调度设置中会呈现当前任务流的依赖对象（需要注意的是，添加的依赖节点，必须是已经提交上线的任务流，比如A任务依赖B任务，则对A进行配置时，B任

务需要已经提交上线，才可以被选择)。



在开发模式下，对于任务的基本属性，可以点击右上角**基本属性**弹窗查看相关的负责人、创建人等信息。



对于已经提交至线上的任务，切换到**线上模式**下。在编辑报警按钮旁边，点击**详细信息**，可弹框查看当前任务的调度信息、报警信息以及嵌套信息。

