## 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E TL加工过程。打开导航栏,点击**离线开发**进入产品页面。

≣													
数	S S S S S S S S S S S S S S S S S S S	数据服务	6000 5及应用				开发植	ma 武    线	上模式		基本属性	快速还	原前往任务管理
Δ	数据传输	① 流	程协作与通知中心	离,但数据未隔离	5。因此在开发	模式下立即	丸行时,需要注意專	临时修改数据目录	<b>己</b>			×	▶三收起
数	居开发	7 问4	答中心										任务节点
E	实时计算	0标	签工厂	设置并运行	🖻 运行结果	🗅 保存	⑦ 提交上线	目 修改记录	Q 搜索节点		₼ ⊕	Θ. []	👬 🈯 数据库传输
~	Yunna	VUI											🔛 🎓 数据同步
_	数据测试中心	in sxi	后期又为			⊇ ods cv	s leisure goo	ds inf 🕕		ds cvs opr sku sud	a di 🕴	0	👯 😂 SQL
- 44	离线开发	数据治理	Ŧ			-		_			-		II MR MR
C	自助分析	》模	型设计中心										🗄 😭 Spark
任約	务运维	頁 指	际系统	ookb 🥑				号 dwd_prod	uct_sales_cookb	. 😗 yunt@	🛛 🨂 dwd	_product_s	
Ø	任务运维中心 (旧)	② 数	居管理						T			7	控制节点
~		****											👬 🔷 选择节点
197	11方运维中心(新)	U 201	店 <u>顶</u> 里中心										🔡 🗇 嵌套流
		12 数	居资产中心										:: 🞇 虚拟节点
		<ol> <li>数1</li> </ol>	居地图						yunna	dws_1		9	

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。

≘									
ى ₪ 🖪 🛱 🗘	未上线 品 cookbook_flow			开发模式	线上模式	yunna E	本属性 1		ā 原 前往任务管理
	① 开发模式与线上模式代码	闷隔离,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	数据目录。		×		▶三 收起
• 🖿 最近打开的任务	じ资源信息 🗌 全选 🗸	▶ 设置并运行 🔋 🗎	运行结果 🕒 保存 🕥 打	是交上线 📒 🛽 修改	记录 Q 搜索节点	-	h. @. @.	0	: 🗇 数据库传输
a cookbook_flow	yunna	新建文件夹			×	product_sales_cookb	0		🔛 字 数据同步
testsql						9			II 🛢 SQL
TPCH_flow		文件夹名称*	cookbook示例任务			/			II MR MR
• 🛅 12324		保存位置*	1		•				🗄 🟠 Spark
• 🛅 DLX TEST and Da	Minna					Junna			🕂 > Script
						0			
TPCH_demo				确定	取消				: 🔷 选择节点
▶ 📩 cookbook示例任务		_		_					<b>…</b> 同 嵌套流
🔸 🛅 etl					<u>d</u>				:: 😫 虚拟节点
→ 📩 xr_test				ads_pr	oduct_sales_cookbo				
> m vungi									

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。

∽ 础 础 ↔	已调度 晶 ads_product_s	ales_cookb	
Q、任务或文件夹名称 7	1) 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
── -	ピ 编辑调度 │ ○ 取消调度	≦ 编辑报警	② 详细信息 🛛 🖓
<ul> <li>ADS</li> </ul>	yunna		
DIM			
• 🗖 DWD			
• 🗖 DWS			
🕨 🗖 ODS yunna	yunna		J.
🚠 cookbook_flow			
testsql			

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mysql2h ive 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。

$\equiv$							
☆ 品 晶 酉 む ⊕	已调度 晶 ads_prod	just sales sookh	yunna	onnna 工 <del>生模式</del>	谷 日時子	基本属	性 - 快速还原 - 前往任务管理
Q.任务或文件夹名称 7	1 开发模式与线	新建任务			×		×
• 🛅 ADS	ピ 编辑调度 ○ 取	任务名称*	mysql2hive_cookb	ook			± € € [
• 🖿 ом учала		任务类型	○ 多节点任务 〇	单节点任务	yunna		
• 🖿 DWD		保存位置*	/ cookbook示例任领	务 / ODS /	•		
• 🛅 DWS			+ 新建文件夹				
- 🗁 ODS		任务描述	请输入仟务描述, 1	最大长度为120个字			
• 🚠 mysql2hive_cookbook		yunna			yunna		
mysql2hive_cookbook2							
• 🚠 mysql2hive_cookbook3							
• 🏦 mysql2hive_cookbook4					0/120		
a cookbook_flow		vunna		「い」確定	取消。000		
testsql							

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。

☆ 品 ■ 品 ⊕	已编度 🚠 m	ysql2hive_cookbook			开发模式	线上	iggnna 模式		基本属	性 快速	008 还原 前往任务管理
	开发楼	载式与线上模式代码隔离,但数据未隔	裔。因此在开发	模式下立即拼	1/5时,需要注意要	副临时修改数据目:	灵.			×	▶三 收起
											任务节点
ADS	2 资源信息	□ 全选 ~ ▶ 设置并运行	🔋 运行结果	🗅 保存	⑦ 提交上线	■ 修改记录	Q 搜索节点		ф.	) Q [	3 👬 🍞 数据库传输
🕨 🛅 DIM yunna											10 🔡 🍄 数据同步
DWD											👯 🤤 SQL
🕨 🛅 DWS											₩ MR
- 🗁 ODS											🗄 🔂 Spark
• 🚠 mysql2hive cookbook		10000-	sama-		- anna		aunna	omna			
A mysql2hive cookbook2		ods_cvs_	leisure_store	_info		🔤 🗘 🖓 🖓	_opr_sku_sug_di	34			II 🚳 AI
· ····································											控制节点
mysql2hive_cookbook3											👬 🔷 选择节点
mysql2hive_cookbook4				🗂 mysal	Phive cookboo	)k	1				:: 🗇 嵌套流
🚠 cookbook_flow			yunna		yunna		yunna				11 : 二 虚拟节点
testsql											

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。

<b>Ξ</b>						
☆ 品 品 囲			开发模式	线上模式	yunna 基本属	<b>属性</b> 快速还原 前往任务管理
Q.任务或文件夹名称	♥ 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。		と三 收超           任务节点
数据同步-ods_city_data						? 临时保存 关闭
数据传输2.0任务*	cookbook_city_data	前往数据传输				
节点属性	cookbook_city_data	值				
	cookbook_ads_csv_opr_top10sku_sug_di					
	<pre>cookbook_dim_cvs_leisure_store_info</pre>					
	cookbook_dim_cvs_leisure_goods_info					
	cookbook_ads_cvs_nation_top10sku_di					

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。

	I
☆ 品   器 匣 凸 ⊕	已调度
Q、任务或文件夹名称 V	•
▼ 🗁 cookbook示例任务	12 资源
🕨 🗖 ADS yunna	yunna
DIM	
🕶 🛅 DWD	
dwd_product_sales_cookb	
dwd_product_sales_cookb	vunna
dwd_product_sales_cookb	200
dwd_product_sales_cookb	
🗝 🛅 DWS	
dws_product_sales_cookb	unna
dws_product_sales_cookb	Annu
- 🗁 ODS	

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。

SQL-dwd_product_sales_cookboo	ok3_month					? 临时保存	关闭
数据信息 🔲	SQL编辑器 其他设	置					
🛢 ads	▶ 相式化						
Q、数据表名称	1 set hive.exec.dyn	amic.partition=true;					
▶ mas_cvs_comm_store_cat	2 set hive.exec.dyna 3 use dwd;	amic.partition.mode=nonstrict;					
Image: mage: ma	4 insert into table	dwd_product_sales_cookbook3_mont	h partition(dt)				
	5 select a.sku_id,a	.area1_name,a.city,a.area2_name,a	.store_num,a.sale_store_n	um,a.ord_store_num,a.s	ale_prod_qty_1d,a.in	v_prod_qty_1d,	
	a.allot prod gty :	,a.inwrn_prod_qty_id,a.onway_prod 1d,a.distri ratio,a.sell through	ratio,a.dynamic sales rat:	io,a.complete ratio,a.	conclusion,a.suggest	,a.pt d,	
	b.province,b.dist	rict_name,b.city_id,b.dt from	yunna	yunna	yunna	yunna	
	6 bdms.cookbook_ads	_csv_opr_top10sku_sug_di a,bdms.c	cookbook_city_data b where	a.city=b.city_id			

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。

$\equiv$	详细信息	1							×
	仅展示上	一次提交上线之后的保	存及资源相关操作						
Q、任务或文件夹名称	序号	操作人	操作时间	操作类型	操作内容 ?				
<ul> <li>▶ 最近打开的任务</li> <li>▲ cookbook_flov</li> <li>■ testsql</li> </ul>	1	а	2021-04-13 13:47:56	保存	删除节点:无 修改节点:dws_product	t_sales_cookbook2_quart	yunna er		
TPCH_flow     12324	2	2	2021-04-13 13:40:09	保存	删除节点:无 修改节点:dws_product	t_sales_cookbook1_quart	er		
<ul> <li>DLX_TEST your</li> <li>TPCH_demo</li> <li>Cookbook示例任</li> </ul>	3	yunna	2021-04-13 11:39:48	保存	删除节点:无 修改节点:dws_product	t_sales_cookbook1_quart	er		
<ul> <li>ett</li> <li>xr_test</li> <li>yunqi</li> <li>默认文件夹</li> </ul>	4	samna L	2021-04-13 11:36:20	www.ma 保存	删除节点: dws_1,dws_ 修改节点: dws_product er,ads_2	2 t_sales_cookbook1_quart	er,ads_1,dws_pro	duct_sales_cookbook2_qua	rt

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。

☆ 〒 品 鹿 む ⊕	已開展 盐 mysql2hive_cookbook2	yunna	开发模式    线上模式	yunna	基本属性	快速还加	а 亰 │ 前往任务管理
	① 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。			×	▶三 收起
<sup>130</sup> 又件夹名称 U			<u> </u>	任务节点			
	2 资源信息 □ 全选 ∨ ▶ 设置并运行 🗈	回 运行结果 ビ 保存 回 携	腔上线 <sup>●</sup> 目修改记录 Q 搜索节点		.∄ ⊕	Q ()	🗄 🏠 数据库传输
awd_product_sales_cookc							👯 🎓 数据同步
DWS		致, 明及时5	ex1.m				🗄 😑 SQL
dws_product_sales_cookb							I MR MR
dws_product_sales_cookb							🗄 🟠 Spark
ODS	unna conna						:: > Script
t mysql2hive_cookbook	ods_cvs_lei	isure_goods_inf 0	ods_cvs_opr_sku_sug_	_di yu			🗄 🖾 Al
mysql2hive_cookbook2							控制节点
mysql2hive_cookbook3			Y				\rm 😯 选择节点
tookbook4		🔲 🞇 mysql2hive_	_cookbook2				🗄 🗇 嵌套流
cookbook flow		vunna	vunna vunna				:: 📫 虚拟节点

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)。

	提交上线					×
	✓ 任务提交校验					
	✓ 检测完成,结论如卜:					
	检测—:是否必须开发模式测 结论:非必须,但建议测试	l试 ?				
• 🚠 dwd_produ						
- 🖿 DWS	检测二: 是否需要走工单审批	5				
• 🏭 dws_produ	结论: 不需要					
• 🚠 dws_produ						
- D ODS	~ 待测试节点					
🔹 🚠 mysql2hive	任务名称			节点名称		
• 🚠 mysql2hive	mysql2hive_cookbook2			ods_cvs_leisu	re_goods_info	
mysql2hive						
🔹 👬 mysql2hive	→ 提交信息					
cookbook_flov	yon yon					
testsql	提交选项 ○ 仅提交上线	● 提交上线并取消调	度 ?			
• 🛅 etl	确定 取消 ☑ 强制	提交上线, 我已知晓风)	睑			

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。

☆ 品 ■ む ⊕	← 编辑调度					
Q 任务或文件夹名称 V	开发模式与线	上模式代码隔离,但数据;	k隔离。因此在开发模式下 <u>。</u>	2即执行时,需要注意要临时修改	攻数据目录。	×
12324	✓ 全部开启 ∨	▶ 放弃编辑 ① 设置	并提交调度 Q 搜索节	点		# @ @ []
DLX_TEST     Yunna     TPCH_demo	Yuuno	γαριτο	yunnia			
▼ 🗁 cookbook示例任务						
- 🗁 ADS						
• 🚠 ads_product_sales_cookb					inna	
ads_product_sales_cookb			Sealer ads_p	roduct_sales_cookbook		
DIM				T		
- 🗁 DWD						
dwd_product_sales_cookt     dwd_product_sales_cookt     dwd_product_sales_cookt			ads_p	roduct_sales_cookbook1	yunna	

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。

⇒	编辑调度					×
☆ 記 ■ 路 囲	✓ 调度设置 vunna					
Q. 任务或文件夹名称	调度设置类型	<ul> <li>● 常规设置</li> <li>○ crontab设置</li> </ul>				
▼ □ 最近打开的任务	调度周期	1	天 •			
🚠 cookbook_flov	首次执行时间	2021-04-13 22:59				
testsql State						
TPCH_flow	◇ 跨流依赖设置					
• 🛅 12324	跨周期自依赖	?				
• 🛅 DLX_TEST	次部⊽‡鱼	仁久戊的				
• TPCH_demo	Yunna	YUNNA				
▶ 🛅 cookbook示例任		添加任务依赖节点				
🔸 🛅 eti		节点依赖				
► 📩 xr_test		添加节点依赖节点				
🔸 🛅 yunqi 🔤 yum <sup>a</sup>						
▶ 🛅 默认文件夹	确定 取消					

## 编辑调度

a č	调度设置				
	调度设置类型	○ 常规设置   ○ crontab设置			
	分钟	57	(〕输	入说明	
	小时 Vunna	22 50000	1.3 2.3	輸入框内仅允许輸入数字及*, *任意一个值 制字————————————————————————————————————	-/
	天	*	4 5. /	, 前定   91%20国, 914 1,2 - 表示一个整数范围, 例如"2-4 / 时间的间隔频率, 例如"0-23/	.,3 4"表示"2,3,4" 12"表示每两个小时执行一次
	月	*			
	星期	*			
	设置预览	57 22 *** <sup>yunna</sup>			
		接下来5次调度执行时间为:			
		2021-04-13 22:57:00			
		2021-04-14 22:57:00 2021-04-15 22:57:00			
	确定 取消	7			

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。

2022/7/30

编辑调度					×
→ 调度设置	确认提交调度		×		
日本ノマンギー		₩UNNA ₩		yunne	
响反反直失坐	市点名称 节	夏: 這些			
调度周期	1 ads_product_sales_co S	QL 开启			
首次执行时间	2 ads_product_sales_co S	QL 开启			
a yunna	3 cookbook_flow 虚	割 初 ガ 点		yunna	
→ 跨流依赖设置	4 dim_product_region_c S	QL 开启			
跨周期自依赖	5 dwd_product_sales_c S	QL 开启			
依赖对象	6 dwd_product_sales_c S	QL 开启		winna	
ý ar	7 dwd_product_sales_c S	QL 开启		Year	
	8 dwd_product_sales_c S	QL 开启			
	9 dws_product_sales_c S	QL 开启			
e yunne	☑ 自动配置任务失败报警 ⑦	yunna	是交调度取消	yunna	
确定取消					

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)。

<u> </u>	ら 数据中台	编辑调度									
	添加任务依赖							×			
	当前任务	ads_product_sales_cookbook1									
•	依赖节点	bdms -	1	•	已选节点 (1)	节点信息					
• == • ==		Q、请输入节点名称	→ DWD			所属任务:dws_pro 所属项目:bdms 任务首次执行时间:	oduct_sales_cookbook1 2021-04-21 11:45:00				
- 🗅			yun 🚠 dwd_pro	oduct_sales_cookbook							
-			🚠 dwd_pro	oduct_sales_cookbook2							
			👬 dwd_ d	wd_product_sales_cookbook2	2						
•											
	yunna	yunna	yunna	yunna	yunna	yunna	yunna	添加取消			
~	跨流依赖设置										
	跨周期自依棘	ĝ <b>()</b> ?									
	依赖对象	象 任务依赖 <sup>100003</sup>									
		添加任务依赖节点									

任务名称

dws\_product\_sales\_cookbook1

当前任务: ads\_product\_sales\_cookbook1

项目名称

节点依赖

确定

取消

添加节点依赖节点

bdms

编辑

已选节点

dws\_product\_sales\_cookbook1\_quarter X

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。

=				]	基本属性			×
今日日田市の		iles cookb	yunna	开发模式				
		-			任务名称	ads_product_sales_cook	.book1	
Q 任务或文件夹名称 🛛 🔽	① 开发模式与线上模式代	品隔离,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	负责人			
▼ ▶ 最近打开的任务	28资源信息 🗌 全选 🗸	▶ 设置并运行	果日保存	⑦ 提交上线 国 修改记	创建人	3		
cookbook_test				yunna	yunna			
• 🚠 ads_product_sales_cod					最近修改时间	2021-04-20 10:38:44		
• 🊠 dws_product_sales_co					提交人			
• 🏦 dws_product_sales_co					最近提交时间	2021-04-20 15:43:24		
• 🏭 dwd_product_sales_co			Sads	product sales cookbo	yunna	Anung		
• 🚠 dwd_product_sales_co					任劳油还	ads层cookbook1数据		
• 🏭 dwd_product_sales_co								
• 🏭 dwd_product_sales_co			C P ada					15/120
cookbook_flow			vanna	product_sales_cookbo				
• 🚠 dim_product_region_cc								
• 🛅 12324					保存关闭			
c					700			

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。

⇒	详细信息							×
御話一郎父			调度信息	报警信息	嵌套信息			
Q. 任务或文件夹名称	∨ 调度设置							
• 🗅 最近打开的任务	周期	1天						
tes:	首次执行时间	2021-04-21 12:00						
• 🏭 ads_product_s	✓ 跨流依赖设置							
• 🚠 dws_product_	····································							
• 🚠 dws_product_	运向期日怅颗							
• 🚠 dwd_product_	依赖对象	任务依赖						
• 🏭 dwd_product_		当前任务: ads_product_sales_cook	book1					
• 🚠 dwd_product_		项目名称	任务	名称		已选节点		
• 🚠 dwd_product_		bdms	dws	_product_sales_cookb	oook1	dws_product_sales_cook	book1_quarter	
cookbook_flov	Minia	而日夕我 yunna	红灰交势		首次执行时间	vunna		
• 🏦 dim_product_r	時にため、「「」」に見ていた。	bdms	dws product sales	cookbook1	2021-04-21 11:45:00	0		
• 🛅 12324	<b>确定</b> 取消					-		