数据传输: 连通业务库, 完成数据集成

数据传输解决异构数据源之间的数据交换问题,是数据进出数仓的桥梁。

连锁便利店的业务数据存储在MySQL数据库中,首先需要将业务库中的数据集成至大数据开发及管理平台中,以便 于后续的进一步处理,通过数据传输功能完成数据同步任务。

点击左上角的Logo,打开产品导航栏,点击数据传输产品,进入产品页面。

							-
数据集成	数据服务及应用	yunna	yunna	yunna	yunna	yunna	yunna
● 数据传输	⑥ 流程协作与通知中心						
数据开发	1 问答中心						
会けけ算 いいの ¹⁰							
数据测试中心	面 数据服务						
🛔 离线开发	数据治理						
自助分析	淤 模型设计中心						
任务运维	口 指标系统						
🖗 任务运维中心(旧)	🔊 数据管理						
🙆 任务运维中心 (新)	数据质量中心						
	2 数据资产中心						
	② 数据地图						

在数据源管理Tab下,点击数据源登记,选择MySQL数据源类型,填写数据源的基本信息,包括数据源连接串、用 户名和密码,登记完成之后,可以在数据源列表中对连通性进行测试,保证元数据中心服务可正常访问登记的数据 源信息,正确获取数据库表内容。

产品功能详述 - 数据传输:连通业务库,完成数据集成 - 《有数数据开发及管理平台用户手册 V6.5.0》

⇒				n	登记数据源 - N	MySQL		Х
≙ ₩	数据传输 离线同步任务		l) D	DB(QS) Oracle	* 数据源名称	1-64个字符		
	数据源管理				* 版本	请选择版本	\vee	
		数据源名称	版本	数据源连接	* 数据源连接	格式: jdbc:mysql://ServerlP:Port/Database		
		cookbook数据源	5	jdbc:mysql://59.111.93.226 7/youdata	* 用户名	请输入用户名		
		kis_test	5	jdbc:mysql://rm- bp1fi0kd6bpzp6r61mo.mys	* 密码	请输入密码 该项为必填项		
		finance_pro	8	jdbc:mysql://59.111.211.44: 6/finance_pro				
c							上一步	确定

₫	数据传输	数据源登记	申请权限			
-	离线同步任务	< M	连通性测试 - cookb	ook数据源	Х	MongoDB RocketMQ >
8	数据源管理					
				and the second second		
		数据源	测试项	测试结果	最近测试时间	操作
		cookbc	元数据中心	● 测试通过	2021-04-01 10:51:15	编辑 测试连接 更多 >
		kls_tes		had fold the month of the second	取 消 测 鼠	编辑 测试连接 更多 >
				DD THUKOBDDZD616 THO. HIVSOL		

数据源测试成功后(本案例中使用的MySQL数据源名称为"cookbook数据源"),点击离线同步任务,选择新建任务,选择单个任务创建模式。

≕		中心创建任务
۵	数据传输	後上任祭 任名警理 第个任务创造
	离线同步任务	
	数据源管理	创建猛犸任务 批量删除 批量任务创建
		创建人:全部 ✓ 支持批量数据传输任务的创建
		SQL模式任务创建
		一 市点名称 任务状态 T 数据来源关型 来源信 ^{支持直接使用SOL从源数据库该取数据}

新建任务分为三步骤:

1. 第一步需要填写同步任务的基本信息,包括任务名称和基本描述;

Ξ.			
● 数据传输	离线同步任务 / 新	建任务	
₩ 离线同步任务	1 基本信息 —	(2)选择数据来源与去向	(3) 字段映射
数据源管理			Ŭ
	* 同步任务名称	中英文开头,支持连字符(-)或下划线(_),长度不超过128个字符]
		建议使用数据来源表名作为同步任务名称	
	描述	输入同步任务描述,最长不超过128个字符	
=			下一步

2. 第二步中选择数据来源和数据去向,本案例中选择cookbook数据源下的库表,选择Hive类型的bdms库作为数据去向,若Hive库下没有合适的存储表,数据传输服务支持快速创建表,点击快速创建表后,平台基于来源表的字段内容自动生成DDL语言,只需更换table名称即可(在对应库下快速新建表,前提是对该库有创建表的权

产品功能详述 - 数据传输: 连通业务库, 完成数据集成 - 《有数数据开发及管理平台用户手册 V6.5.0》

限,可在**项目配置-权限管理-数据权限**,查看对应库下,当前用户对应的角色是否有创建表的权限)。

▲ 数据传输		
高线同步任务	数据来源: MySQL ∨	数据去向: Hive ∨
2 数据源管理	童找方式 💿 库表选择 💿 正则匹配	* 数据库 bdms V
	* 数据源 cookbook数据源 V youdata V	* 表 请选择表 V
	* 表 youdata_ ×	快速创建表
	+ 跨库表	写入规则 Insert overwrite <>
	导出方式 💿 常规型 🔘 流水型	高级配置 属性 值
		点击下方的添加按钮新增配置
	过滤条件 where过滤语句(不含where关键字);通常用作增量同步	+ 添加
	* 流量控制 (④) 不限速 () 限速	
● 数据传输		
■ 高线同步任务	数据来源: 快速新建目标Hive表	×
2 数据源管理	重线方:	×
	数据库 bdms * 数据	
	* 建表语句 1 请修改输入框中的 'your_table_name'	
	* CREATE TABLE IF NOT EXISTS 'your_table_name '日期' STRING COMMENT '', '商品小类' STRING COMMENT '', '南品小类' STRING COMMENT '', '单款策略' STRING COMMENT '', '收益类型' STRING COMMENT '',	·() · · · · · · · · · · · · · · · · · ·

`组成元素1` STRING COMMENT ",

`组成元素2` STRING COMMENT ", `销量` INT COMMENT ",

*建议零售价、INT COMMENT ", *地区、STRING COMMENT ", *城市、STRING COMMENT ", *采购数量、INT COMMENT ",

`入库数量` INT COMMENT ",

`门店接收` INT COMMENT ",

过滤条

* 流量控

C

添加按钮新增配置

产品功能详述 - 数据传输:连通业务库,完成数据集成 - 《有数数据开发及管理平台用户手册 V6.5.0》

3. 第三步,平台会展示来源表和目标表的字段的映射关系,可以选择某几个字段不导入,或者调整字段之间的映射关系。确认好映射关系之后,点击**确定**,同步任务即创建完成。

● 数据传输	离线同步任务 / 新建任务											
 高线同步任务 数据源管理 	 ✓ 基本 重置 	信息			✓ 选择数据来源	与去向 ———			- 3 字段映射			
	序号	来源表字段		类型	描述		目标表字段	类型				
	1	store_id v		INT(10, 0)	-		store_id	int				
	2	cate_id \lor		INT(10, 0)	-		cate_id	int				
	3	avail_skunum <		INT(10, 0)	-		avail_skunum	int				
	4	avail_inv_amt v		DOUBLE(10, 2)	-		avail_inv_amt	double				
	5	sale_prod_amt_tot_1d \lor		DOUBLE(10, 2)	-		sale_prod_amt_tot_1d	double				
	6	sale_profit_tot_1d V		DOUBLE(10, 2)	-		sale_profit_tot_1d	double				
۵	7	ord_prod_amt_tot_1d \lor		DOUBLE(10, 2)	-		ord_prod_amt_tot_1d	double				

同步任务创建成功之后,在任务列表中可点击**运行**,支持直接跳转到日志中查看任务运行情况,任务运行成功之后,点击**提交上线**,可将该同步任务提交至离线开发的线上模式下(离线开发分为开发模式和线上模式,只有将同步任务提交上线,在线上模式中才可以引用该同步任务,执行任务调度等)。

● 数据传输	线上任务 任务管理	新建任务
器 离线同步任务		
数据源管理	创建猛犸任务 批量删除	
	创建人: 全部 ∨	任务名称 ∨ Q、请输入任务名称 Q
	节点名称 任务状态 👻 数据来源类型 👻 来源信息	数据去向类型 👕 去向信息 操作
	ods_2_cookb 已提交 MySQL youdata:ads_cvs_opr_	_top10sku_ Hive bdms:ods_2_cc 编辑 删除
	ook sug_di,dim_cvs_leisur	re_store 运行 日志
		提交上线
	□ ods_cookboo 已提交 MySQL youdata:ads_cvs_opr_	_nation_top Hive bdms:ods_cool 编辑 删除
	K 105KU_al,alm_cvs_leis	sure_goo 运行 日志
		提交上线

产品功能详述 - 数据传输:连通业务库,完成数据集成 - 《有数数据开发及管理平台用户手册 V6.5.0》

提交上线之后的任务,可以在**线上任务**列表中查看,若离线同步任务已经被任务引用,可以点击查看详情,快速了 解同步任务被引用的详细信息。

● 数据传输	线上任:	务 任务管理	≞				新建任务
■ 离线同步任务	批量下	引用详情				×	
■ 数据源管理	创建人:	离线任务名称	mysql2hive_ndi_ads_all	power		入请输	
		引用的节点	节点名称	所属任务	创建人		操作
			ads_all_power	测试任务6 (开发模式)	liut23	_	下线
							运行结果 引用详情

数据传输—功能扩展介绍

在**任务管理**下,平台支持基于已经创建的数据集成任务,快速构建离线开发任务,生成的任务默认存放在**默认文件 来**。首先选中已经创建成功的数据集成任务,点击**创建离线开发任务**,弹窗中填写待创建的任务流的名称,任务流 中各个节点任务的名称也可以进行预置,如果前后缀均未填写,则默认以source_table的名称命名节点任务的名 称;如果选了多个数据集成任务,可设置分组数量,即离线开发生成的任务中,连接线分支的数量。

● 数据传输	线上任务	任务管理				新	建任务	批量任务创建	
器 离线同步任务		批量创建任务	ξ.	Х					
数据源管理	创建猛犸任务								
	创建人:全部 >	* 任务流名称	cookbook_test		称 >	Q、请输入日			0
	■ 任务名称	任务名称	cookbook \${source_table}		Ψ :	去向信息	操作		
		* 分组数量	5				提交上线	ŧ	
	cookbook _csv_opr_		✔ 将选中数据传输任务提交到线上		ł	odms:cookbook 10sku_sug_di	编辑 运行 提交上约	删除 日志	
	Cookbookdata		取消	创建	I	odms:cookbook	编辑 运行	删除日志	

点击创建之后,在离线开发的默认文件夹下可查看快速创建的任务,画布中与虚拟节点之间的连接线数量即为设置 的**分组数量**。该功能适用于批量快速基于已有的集成任务创建离线开发任务,简化数据集成任务运行链路,提高数 据同步任务提交上线的效率。

产品功能详述 - 数据传输: 连通业务库, 完成数据集成 - 《有数数据开发及管理平台用户手册 V6.5.0》

☆品】番車む⊕	未上线 🚠 cookbo	pok_test		开发模式 线	上模式	yunna 基本属性	快速还	》 原 前往任务管理
Q 任务或文件夹名称 ♀	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	1行时,需要注意要临时修改数据目:	큧.		×	▶三 收起
	2 资源信息	全选 ~ ▶ 设置并运行 । 自	运行结果 🕒 保存	⑦ 提交上线 [●] 目 修改记录	Q 搜索节点	#.⊛	Q []	社会で見 ひましん ひょう
Vunna					yunna	yunna		・ 😭 数据同步
								👯 😂 SQL
•								II MR MR
• F	pr. nati		vs. opr. top1		tv data		kdim cvs	_ 🗄 🏠 Spark
B	yunna	yunna	yunna	yunna	yunna	yunna	Zan	Script
• 🖿								. :: (A) AI 控制节点
· 🖿 💷				🗌 🞇 cookbook t	test			: 🔷 选择节点
▼ 🗅 默认文件夹								🔡 🗇 嵌套流
🚠 cookbook_test								: 📫 虚拟节点
🚠 测试1								
🏭 测试任务6								