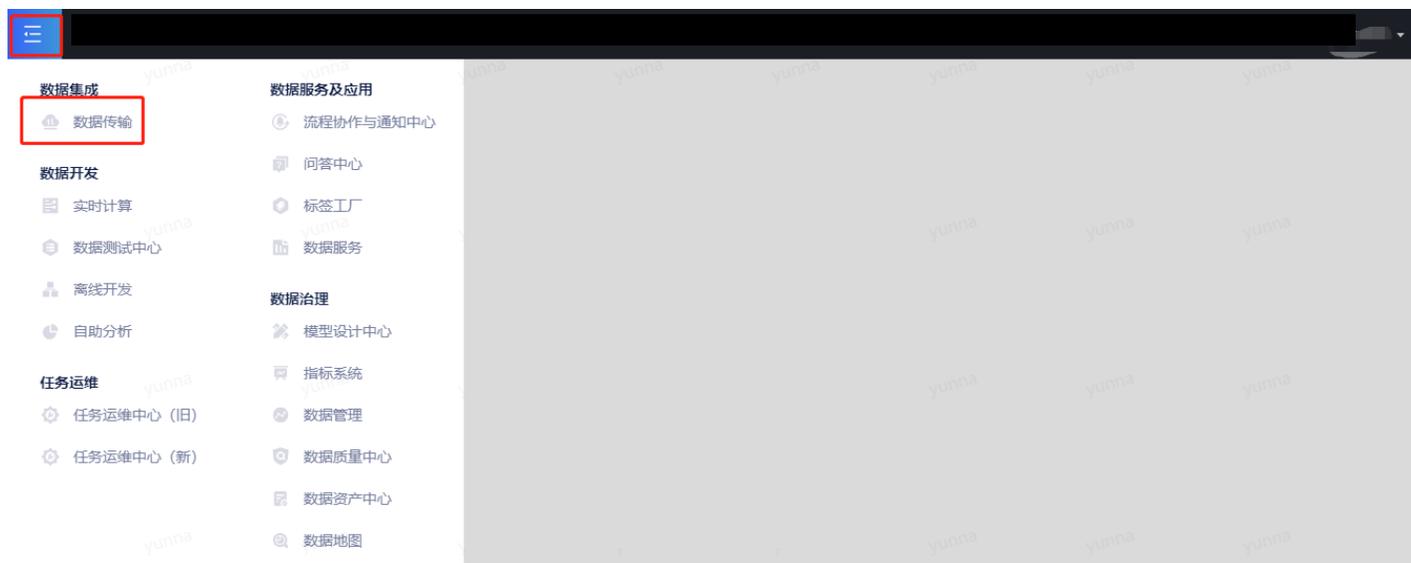


# 数据传输：连通业务库，完成数据集成

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 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。



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。



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**。



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；

数据传输

离线同步任务 / 新建任务

1 基本信息

2 选择数据来源与去向

3 字段映射

\* 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

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。



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。

高线同步任务 / 新建任务

基本信息 | 选择数据来源与去向 | 3 字段映射

重置

序号	来源表字段	类型	描述	目标表字段	类型
1	store_id	INT(10, 0)	-	store_id	int
2	cate_id	INT(10, 0)	-	cate_id	int
3	avail_skunum	INT(10, 0)	-	avail_skunum	int
4	avail_inv_amt	DOUBLE(10, 2)	-	avail_inv_amt	double
5	sale_prod_amt_tot_1d	DOUBLE(10, 2)	-	sale_prod_amt_tot_1d	double
6	sale_profit_tot_1d	DOUBLE(10, 2)	-	sale_profit_tot_1d	double
7	ord_prod_amt_tot_1d	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

离线同步任务

数据源管理

线上任务 任务管理 新建任务

创建猛犸任务 批量删除

创建人: 全部 任务名称 请输入任务名称

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_cookbook	已提交	MySQL	youdata.ads_cvs_opr_top10sku_sug_di,dim_cvs_leisure_store...	Hive	bdms:ods_2_cookbook	编辑 删除 运行 日志 提交上线
ods_cookbook	已提交	MySQL	youdata.ads_cvs_opr_nation_top10sku_di,dim_cvs_leisure_goo...	Hive	bdms:ods_cookbook	编辑 删除 运行 日志 提交上线

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—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**默认文件夹**。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。



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。

The screenshot displays a data integration platform interface. At the top, there are tabs for '未上线' (Not Online), 'cookbook\_test', '开发模式' (Development Mode), and '线上模式' (Online Mode). A yellow warning banner at the top center states: '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,需要注意临时修改数据目录。' (Code is isolated between development and online modes, but data is not. Therefore, when executing immediately in development mode, you need to pay attention to temporarily modifying the data directory.)

The main area shows a workflow diagram with several nodes: 'cookbooktest', 'cookbookcity\_data', 'cookbookdim\_cvs', 'cookbookads\_cvs\_opr\_top1...', and 'cookbook\_test'. A red box highlights the connections between these nodes. On the left, a file explorer shows a folder structure under '默认文件夹' (Default Folder), with 'cookbook\_test' highlighted in a red box. The right sidebar contains a list of task types: '数据库传输', '数据同步', 'SQL', 'MR', 'Spark', 'Script', 'AI', '控制节点', '选择节点', '嵌套流', and '虚拟节点'.