

目录

目录	1
入门概述	2
登陆大数据云平台	2
添加子账号	3
选购产品	4
创建项目	4
添加数据源	4
新建库表	5
数据同步作业	5
配置调度	5
作业流发布	6

入门概述

本模块将指引您快速完成一个完整的数据同步和运维操作。

使用场景：将业务数据库MySQL的数据同步到数据仓库Hive的表中。

主要包括以下步骤：

1. [准备工作](#)
2. [元数据管理](#)
3. [创建作业](#)
4. [配置调度及发布](#)

登陆大数据云平台

1. 已有大数据租户账号，直接输入用户名和密码登录。还没有大数据云账号？请点击【注册有礼】！

金山云账号登录

登 录

[忘记账号 / 密码?](#) [注册有礼](#)

IAM用户/合作伙伴员工，请用子用户登录

欢迎注册金山云账号

用户名

设置密码

确认密码

手机号码

验证码 [获取手机验证码](#)

[《金山云云服务使用协议》](#) [《信息源入网安全责任书》](#) [《隐私政策声明》](#)

注册有礼

已有金山云账号? 请 [登录](#)>

2. IAM用户/合作伙伴员工, 请点击【子用户登录】, 输入主账号(用户账号) ID或用户名、子用户或合作伙伴员工账号的用户名、相应的登录密码, 即可登录。

子用户登录

主账号ID / 用户名

子用户 / 合作伙伴员工账号 用户名

子用户 / 合作伙伴员工账号 登录密码 

登录

添加子账号

1. 管理用户

(1) 点击【平台概览】页的【用户管理】或在左侧菜单【组织管理】目录下找到【用户管理】, 点击进入用户管理页。

2. 添加到群组

- (1) 点击【平台概览】页的【群组管理】或在左侧菜单【组织管理】目录下找到【群组管理】，点击进入群组管理页。
 - (2) 点击【新建群组】，输入对应信息点击【创建】完成群组的新建。
 - (3) 在【群组管理】页面，为刚才创建的群组添加用户，并进行角色授权
- ### 3. 分配角色
- (1) 在左侧组织管理目录下点击进入【角色管理】
 - (2) 添加用户：点击【添加用户】按钮进行用户的添加。选中需要添加到该角色的用户点击【确认】即可。

选购产品

1. 购买产品

从【大数据云】【数据管理】【数据集成】【离线计算】等产品官网主页点击购买按钮自助购买相应产品。公测阶段请联系服务人员开通。

2. 资源配置

- (1) 在左侧菜单栏点击【已购产品】——>【资源管理组】，进入资源管理组页面。
- (2) 点击【新建资源组】，为已购产品配置资源。也可以使用默认的资源分组。

参数	说明
资源组名称	尽量填写能唯一标识该资源组的名称
所属产品	选择资源组所属产品，其他产品不可用该资源组
资源组类型	CU、DCU，根据产品联动一个资源组只能包含一种类型的资源
所属环境	资源组对应的物理环境，可映射到一个物理环境（生产或测试），也可以同时映射到生产和测试
测试环境配额	该资源组在测试环境能使用的最个资源
生产环境配额	该资源组在生产环境能使用的最个资源
是否项目独享	开启项目独享的话，该资源组只能授权给一个项目使用
资源组描述	非必填，对该资源组的简单说明

创建项目

1. 新建项目

- (1) 在左侧菜单栏选择【项目空间】，进入【我的项目】页面。
- (2) 点击【新建项目】，填写相关信息。
 - 填写基本信息，点击下一步。
 - 选择该项目要包含的产品服务，未购买的不允许勾选。其中基础版数据管理、运维中心、项目管理为公共部分，默认包含在所有项目内。
 - 为项目选择的产品服务配置资源组。此处只能选择该管理人员有权限的资源组。
 - 确认基本信息，点击确认完成创建。

2. 添加成员

点击项目名称，跳转到【成员管理】页面，点击【添加成员】可向该项目添加项目成员。

3. 项目配置

点击【项目配置】可以对项目名称，责任人，项目描述，安全配置进行查看和修改。

添加数据源

1. 新增起始及目标数据源

- (1) 点击【数据管理】，进入【数据源管理】页面。

(2) 点击【新增数据源】，选择【MySQL】作为起始数据源。选择测试环境和测试环境，托管模式，填写相关信息，测试连通性，确定完成添加。

(3) 点击【新增数据源】，选择【Hive】作为目标数据源。选择测试环境和测试环境，托管模式，填写相关信息，测试连通性，确定完成添加。

2. 同步元数据

在【数据源管理】页面，选择刚才创建的MySQL数据源，点击【同步】，选择要同步的数据库，所属项目选择之前创建的项目，确定。将元数据同步到【库表管理】中。

新建库表

1. 新建数据库

(1) 点击【数据管理】——>【元数据管理】，进入【库表管理】页面。

(2) 选择【关系型数据库】，点击【新建数据库】。

(3) 填写相关信息，所属项目选择刚才创建的项目，确认创建成功。

(4) 在【库表管理】页面，选择刚才创建的数据库，点击【发布到测试】。数据源类型选择【Hive】，数据源名称选择刚才创建的Hive数据源。

(5) 确认，发布到测试环境。审核通过后，从测试环境发布到生产环境。

2. 新建数据表

(1) 选择【关系型数据库】，点击【数据表】Tab，进入数据表管理页面。

(2) 点击【新建数据表】，填写相关信息：

- 基本信息：所属项目选择刚才创建的项目。
- 字段设置：参照与MySQL中要同步的表。
- 设置分区

(3) 确认创建成功。

(4) 选择刚才创建的数据表，点击【发布到测试】。数据源类型选择【Hive】，数据库选择刚才发布到测试的数据库。

(5) 确认，发布到测试环境。审核通过后，从测试环境发布到生产环境。

数据同步作业

1. 创建作业

(1) 进入【项目空间】->【我的项目】，点击刚才创建的项目名称进入【项目空间】->【我的项目】，点击项目名称进入大数据开发套件。点击进入【数据开发】->【离线作业开发】。

(2) 选择【任务开发】，在左侧目录点击创建的作业流，新建一个作业流。

2. 编辑作业

(1) 双击作业流，进入作业流开发面板，拖拽数据同步插件，输入节点名称。

(2) 双击打开新建的同步任务，打开同步任务页面后整个同步任务分成三步：

- 选择数据源表 选择刚才创建的MySQL数据源中要同步的表。
- 选择数据目标表 选择刚才创建的Hive数据表，写入方式选择【insert overwrite】
- 设置数据源表和数据目标表的映射管理 在映射过程中左边字段信息来自源表，右边字段信息来自目标表。

3. 点击上方【运行】进行测试，点击【停止】停止运行。

4. 点击【前往运维】，查看测试环境的运行情况。

配置调度

1. 作业流调度信息配置

- (1) 在作业流面板，点击作业流右侧【调度配置】。
- (2) 填写调度的首次生效日期，一般指定当天既可。调度周期时区选择选择北京东八区，并设置是否节假日运行状态为否表示节假日不运行作业流，以及相应的节假日日历信息。
- (3) 配置作业流的时间依赖：分为指定时间范围和指定时间点两个时间范围选项。可以指定时间点为希望运行的时间。

2. 作业调度信息配置

- (1) 双击进入创建的数据同步作业，点击作业右侧【调度配置】。
- (2) 选择调度频度为周期类，且频度选择每一天。作业就依照作业流上选定的的周期日历每日运行。

作业流发布

1. 作业流提交

- (1) 点击作业流界面【提交】按钮，选择提交的作业节点和提交备注。提交之后生成一个作业流新版本。
- (2) 点击【发布】，选择刚刚生成的作业流新版本，点击确定进行发布。

2. 发布作业流审核

- (1) 切换到【发布管理】界面查看【发布列表】查看相应的发布任务，当作业流经项目管理员审批通过之后就上线了。
- (2) 切换到【运维中心】【生产实例】页面，查看作业周期性运行的实例状态。