

# 快速入门

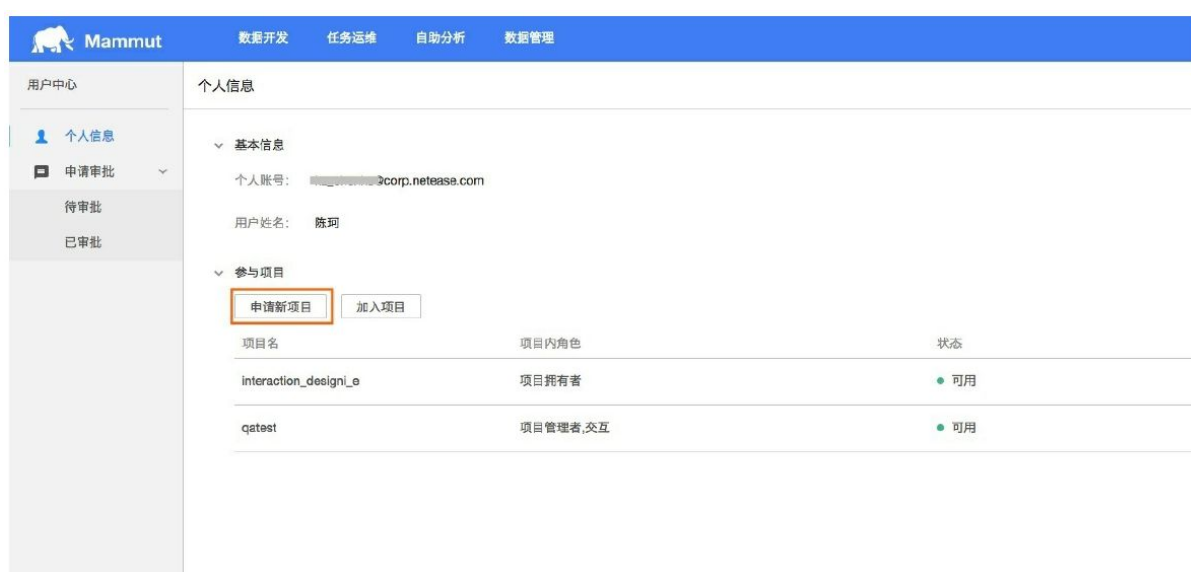
在阅读本指南之前，请确保您拥有了项目空间，可通过下面的步骤拥有项目空间：

## 1、被邀请加入项目

其他项目空间的管理者邀请您加入项目，并赋予了相应的权限，更多关于权限的问题，请参考“用户手册”的“权限管理”部分。

## 2、申请自己的项目空间

- 登陆猛犸平台，鼠标放置在右上角您的用户名处，出现悬浮框，选择“用户中心”，进入“个人信息”页面，可以看到有“申请新项目”和“加入项目”两个选项，选择“申请新项目”。



- 填写项目的有关信息以及申请的原因。

← 申请新项目

▼ 项目信息

应用业务线\*

1-25个字，可输入中文，大小写字母及其他字符

项目名称\*

1-64个字符，可输入小写字母、下划线

集群\*

滨江三期

存储主路径\*

/user/\*

存储配额\*

1T

Hive库\*

库名	路径
1-64个字符，可输入小写字母、下划线	/user/*hive_db/*db

队列名称\*

1-64个字符，可输入小写字母、下划线

资源配置\*

默认策略

方案	CPU配额	内存配额	最大并行任务数
<input checked="" type="radio"/> 方案1	100 核	300 G	20 个
<input type="radio"/> 方案2	300 核	900 G	20 个
<input type="radio"/> 方案3	500 核	1500 G	20 个
<input type="radio"/> 自定义	10核	10G	20个

弹性策略

☐ 同一时间系统将优先采用弹性策略配置

申请原因

预估并发任务数

请输入数字个

单表最大数据量

请输入数字G

描述\*

输入描述

0/120

主管邮箱\*

申请

- 上述信息填写之后，点击“申请”，新项目需要平台管理员的审批，在申请审批的“待审批”页面中可以看到审批的进度。



- 管理员审批之后，在“个人信息”中可以查看自己新申请的项目。

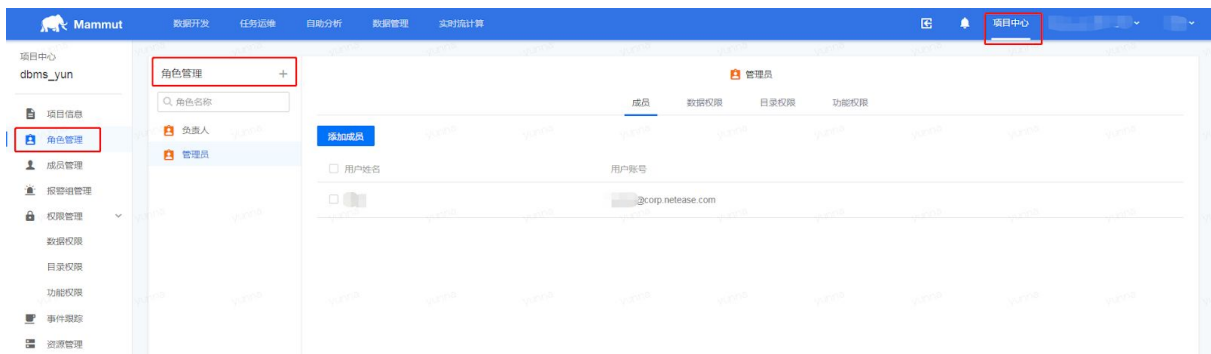


注：您一旦成为了项目空间的owner，意味着该项目内您拥有最高的权限，在没有给别的用户赋予权限之前，任何人都没有权利访问您自己的项目空间。

### 3、作为管理者，邀请用户，授予权限

如果您作为项目的管理者，计划邀请其他用户使用您创建的项目空间，可继续学习如何为其赋予相应的权限。

- 首先创建匹配被邀请人的角色信息。点击顶部菜单栏的“项目中心”，左侧菜单栏中选择“角色管理”，左上角的“角色管理+”用来创建新的项目角色。



- 填写对应的角色信息，“确定”之后，可以在“角色管理”界面看到新增的角色信息。



添加角色

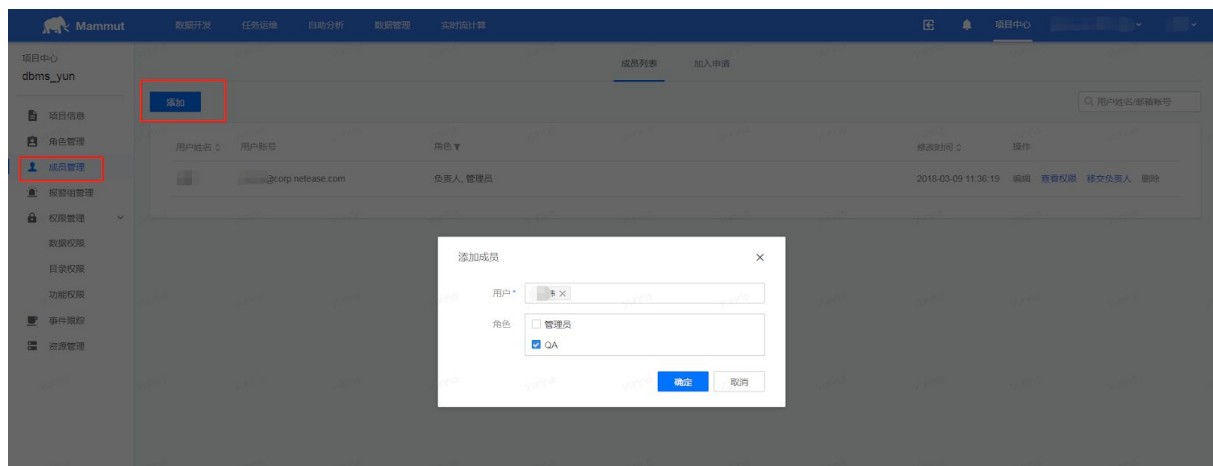
角色名称\* QA

角色描述 测试人员

4/120

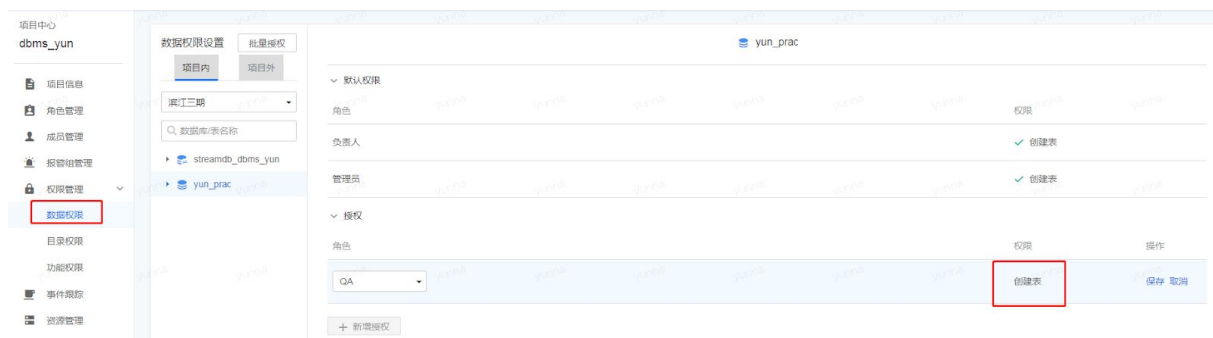
确定 取消

- 左侧菜单栏中选择“成员管理”，单击左上角的“添加”，填写被邀请人的账号信息，并赋予其对应的项目角色。



- 选择左侧菜单栏中的“权限管理”，对新增的项目角色赋予对应的权限。

针对“数据权限”，赋予角色某数据库中创建表的权限：



针对“目录权限”，授予角色某目录只读或者读写的权限：



针对“功能权限”，默认有“方案一”和“方案二”，可自己添加针对特定角色的权限信息。点击“添加”，填写授权的信息，指定授权的对象。



#### 4、申请新队列

若您认为创建项目时分配的队列资源过于紧张，您可以继续阅读该小节，学习如何申请新队列资源。

- 选择“项目中心”下的左侧菜单栏中“资源管理”，可以在您当前的集群资源信息中看到有“申请新队列”选项。



- 填写您想要申请的新的队列信息。

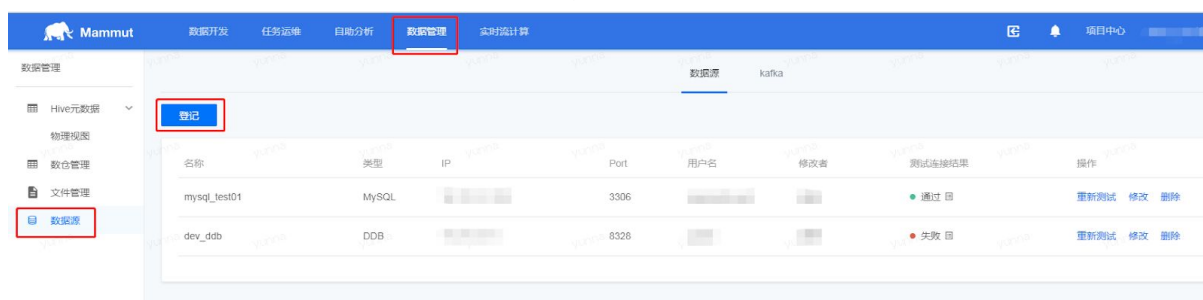


完成上述操作之后，您已经拥有了可操作的项目空间，并且学会了如何申请新的资源；若您是管理员，也知道了如何邀请别人加入到当前的项目空间中，可以继续进行下面的学习了。

在您拥有了项目空间的基础上，本指南将指引您快速完成一个完整的数据传输和运维操作。本文以MySQL数据源为例，说明如何利用数据管理功能将猛犸平台汇总的数据导出到MySQL数据源中，其他数据源的使用详解请参阅“用户手册”的“数据管理”部分。

## 1、新增数据源

- 进入顶部菜单栏中的“数据管理”页面，单击左侧导航栏中的“数据源”。



- 单击左上角的“登记”，类型选择“MySQL”，填写相关信息。

登记数据源

数据源信息

名称\*

1~64个字符

类型\*

MySQL

JDBC URL\*

格式：jdbc:mysql://ServerIP:Port/Database

用户名\*

密码\*

测试连接信息

权限\*

系统将测试所选集群对登记数据源的连通性，请根据线下申请情况进行选择

集群

权限

☐ 滨江三期

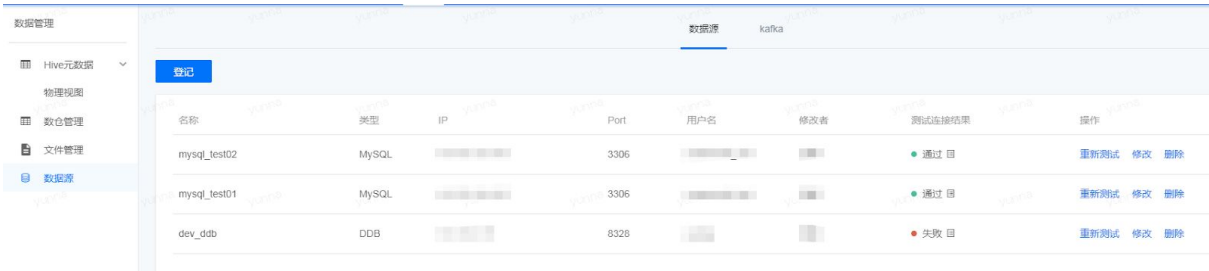
☒ 读 ☒ 读写

注意：系统不验证所选权限，仅做记录。

登记

取消

- 点击“测试连接”，测试数据源的连通性，若测试连通性成功，可进行下面的步骤，若测试连通性失败，请根据自身情况进行调整，检查数据源信息或网络状况等。



名称	类型	IP	Port	用户名	修改者	测试连接结果	操作
mysql_test02	MySQL		3306			通过 0	重新测试 修改 删除
mysql_test01	MySQL		3306			通过 0	重新测试 修改 删除
dev_ddb	DDB		8328			失败 0	重新测试 修改 删除

## 2、创建Hive元数据离线表

- 进入顶部菜单栏中的“数据管理”界面，点击左侧“Hive元数据”下面的“物理视图”。



- 左上角可以看到有“新建离线表”和“新建流表”，本案例中使用“新建离线表”操作，有关于“新建流表”的具体使用步骤可以参阅“用户手册”的“数据管理”章节。



新建离线表

表单模式SQL模式

数据库\*

yun\_prac

表名称\*

test\_test

类型

☐ 外部表 ☒ 内部表

表生命周期\*

☒ 永久 ☐ 自定义

天

模式

☒ 手动添加字段 ☐ 从内容中解析

普通字段\*

列名	描述	类型	
col	字段1的描述	STRING	✕
col2	字段2的描述	STRING	✕

+ 添加

分区表

☐

存储格式\*

TEXTFILE

序列化格式\*

JSON

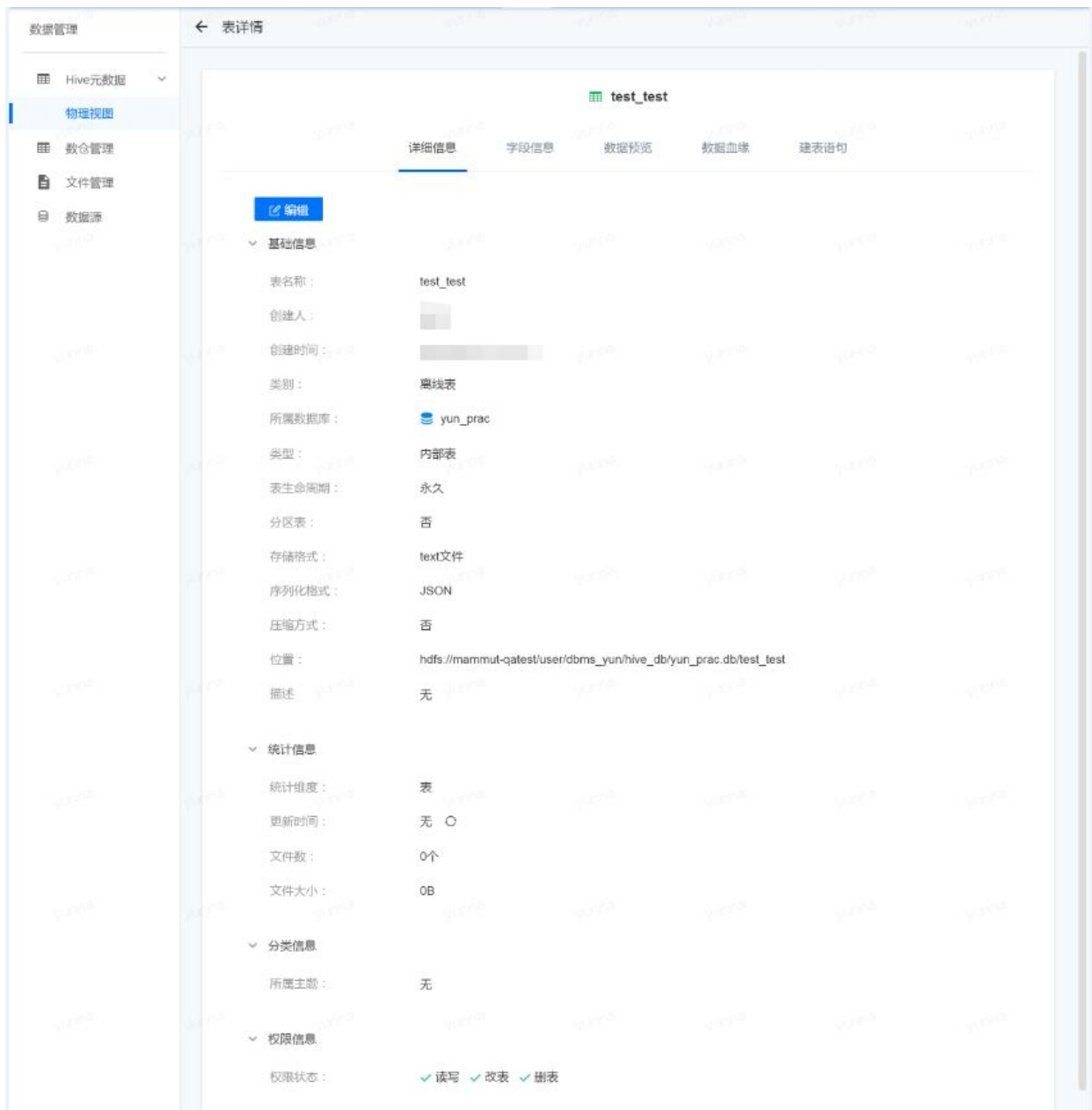
预览

新建

取消

对于离线表的创建，本平台提供两种创建方式：“表单模式”和“SQL模式”。在“表单模式”下，可以使用手动添加字段或者从内容中解析的形式创建表；在“SQL模式”下，可以书写SQL语句来创建表。如上所示，创建了一个test\_test的离线表，保存在yun\_proc的数据库中。

- 离线表创建完成之后，在“Hive元数据”的“物理视图”对应的数据库下，可以看到创建的离线表，右侧的“详情”可查看创建的离线表的具体信息。



### 3、新建文件夹，新建数据库传输的节点任务

- 进入顶部菜单栏的“数据开发”界面，单击左上角的“+新建”操作，选择“新建文件夹”选项，可自定义文件夹名称，例如my\_test，默认保存在根目录下。

新建文件夹

文件夹名称 \* my\_test

保存位置 \* /

确定 取消

- 同样单击左上角的“+新建”选项，创建一个数据库传输的节点任务，任务可自定义名称，例如hive\_to\_mysql，选择保存在my\_test文件夹下。

新建任务

任务名称 \* hive\_to\_mysql

任务类型 ☐ 作业流 ☒ 节点 ☐ 任务组

节点类型 \* 数据库传输

保存位置 \* / my\_test /

+ 新建文件夹

任务描述 \* 数据库传输实例

7/120

确定 取消

- 节点任务创建成功之后，在“数据开发”的界面新增了“hive\_to\_mysql”的数据传输节点，右键选择“编辑”，对源数据和目标数据进行配置。

来源类型选择“Hive”,来源库以及表选择上面创建的yun\_proc数据库下的test\_test离线表；目的类型选择“MySQL”，目的库选择在“数据管理”的“数据源”中测试连接成功的MySQL数据库的名称，表选择MySQL中存在的表。

注：在导出之前，要确保Hive中表的字段类型和MySQL中的表字段类型的一致性。

- 节点信息编辑完成之后，就可以选择该节点“设置并运行”，有关于“运行设置”如下图所示。

其中的“队列”指的是在申请项目空间时所命名的队列，用于设置作业流执行使用。有关于其他字段的详细说明请参阅“用户手册”中的“运行设置”选项。

- 点击“运行”之后，可以看到该节点任务的具体执行状况，同时对应有“节点日志”选项，可在任务失败的情况下查看报错信息，定位失败原因。

运行结果-717770 成功 查看日志

筛选 (按开始时间) 重置

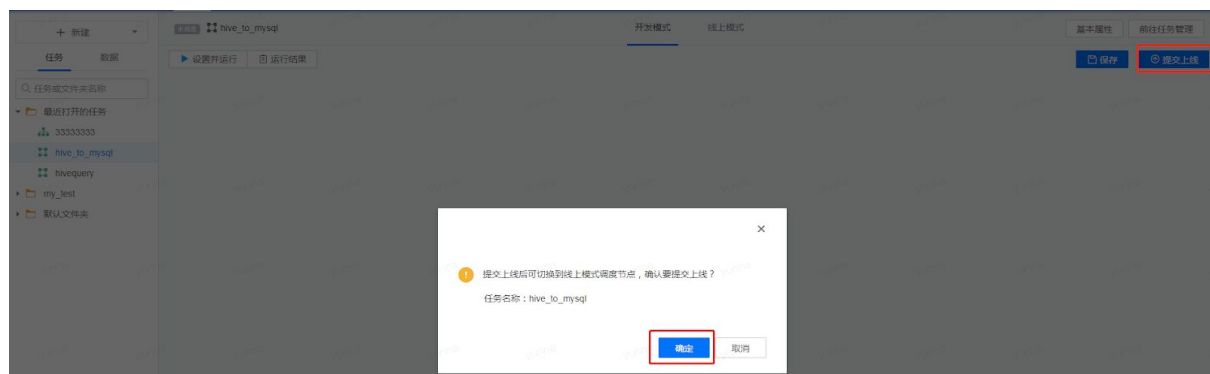
2018-03-13 00:00 2018-03-14 00:00

717770 成功  
开始时间 2018-03-13 14:43

节点名称	节点类型	时间轴	开始时间 ~ 结束时间	运行时长 (s)	状态	操作
hive_to_mysql	数据库传输	<div></div>	2018-03-13 14:43:31 2018-03-13 14:43:53	22.533	成功	节点日志

#### 4、提交上线，设置调度周期

- 对已经测试成功的节点任务，可点击“提交上线”。



- 在界面上由“开发模式”切换到“线上模式”，可对提交上线的节点任务进行调度的编辑操作。选择“设置并提交调度”选项，对调度信息进行配置。

提交调度

调度设置

调度周期1天

首次执行时间2018-03-15 16:15

跨流依赖设置

依赖节点添加节点

报警设置

报警规则

作业流/作业flows:hive\_to\_mysql

报警规则失败报警

+ 添加

报警通知

报警接收人

报警人

报警接收方式

☒ 邮件☒ 短信☒ 语音☒ popo☒ stone

+ 添加

执行设置

队列dbms\_yun

高优先级

自动重试

重试次数\*0

重试间隔\*0秒


提交调度取消

注：满足任一规则即报警

本平台支持天、小时、分钟、月以及星期的周期选项，有关于更详细的配置说明，请参阅“用户手册”中的“提交调度”章节。

注：新版本的猛犸平台区分了测试环境(开发模式)和线上环境(线上模式)，两个环境中的代码隔离，但是数据没有隔离，故在开发模式下立即执行的时候，需要注意临时修改数据目录。

- 任务提交之后，可前往顶部菜单栏“任务运维”的“实例运维”中进行相关信息的查看。

Mammut



数据开发

任务运维

自助分析

数据管理

实时计算

  项目中心

任务运维

今天

全部提交人

全部运行方式

全部状态

重置

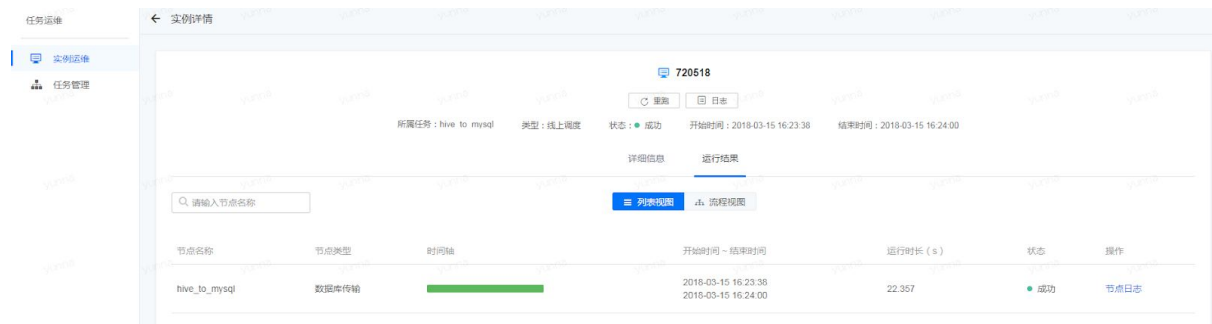
批量停止

实例运维

任务管理

实例ID	任务名称	提交人	状态	运行方式	开始时间	结束时间	运行时长(s)	计划执行时间	操作
720510	hive_to_mysql		成功	线上调度	2018-03-15 16:23:30	2018-03-15 16:24:00	22.4	2018-03-15 16:23:00	详情 重跑 停止

- 针对调度任务，可以进行“重跑”和“详情”操作。点击“详情”选项，可以从列表和视图两种形式查看实例的具体运行信息。

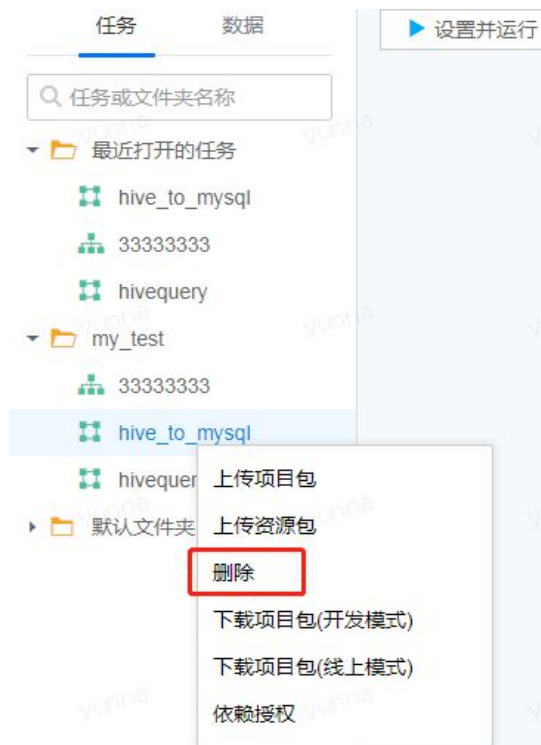


## 5、取消调度，删除节点任务，删除文件夹

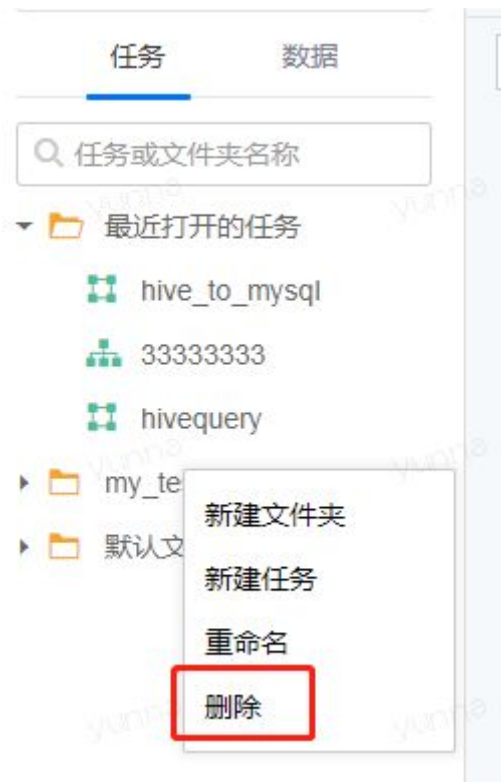
- 回到顶部菜单栏中的“数据开发”界面，针对该节点任务，切换到“线上模式”，选择“取消调度”。



- 选中节点任务，右键选择“删除”，对节点任务进行删除操作。



- 选中对应的文件夹，右键选择“删除”，对文件夹进行删除操作。



现在，完成上面的操作之后，您已经学会了如何在该平台上将Hive离线表中的数据导入到MySQL数据源中，同时提交上线，设置调度周期，查看运维实例，取消调度和删除这一系列的操作，当然这仅仅是我们平台小小的一个功能实现，有关于更多的使用详情，请参阅“用户手册”。