

补数据实例运维

当前任务运维中心支持**单个任务补数据**和**批量补数据**功能。单个任务补数据功能可在**任务列表**针对具体任务进行补数据操作。**补数据实例运维**页面主要提供**批量补数据**功能。

批量补数据

补数创建时间: 全部时间 | 计划执行时间: 全部时间 | 提交人: 全部 | 补数规模: 全部 | 批量

任务名称:

补数状态: 全部(5) | 运行中(0) | 运行中异常(0) | 正在终止(0) | 终止(1) | 暂停中(0) | 已暂停(0) | 成功(4) | 未开始(0)

补数任务ID	补数任务名称	补数状态	补数时间范围	任务及实例统计	运行时间信息	提交人	操作
100000108	批量补数任务_20211215_18014 7	终止	2021-12-14 00:00:00 2021-12-14 23:59:59	涉及任务数: 1 (1) 实例总数: 1 实例完成率: 0.00% 失败实例未处理数: 1	执行方式: 立即执行 开始时间: 2021-12-15 18:01:53 结束时间: 2021-12-15 18:02:00 补数据总时长: 00:00:07		暂停 终止 恢复补数据
100000107	批量补数任务_20211215_18003 6	成功	2021-12-15 00:00:00 2021-12-15 18:00:36	涉及任务数: 1 (1) 实例总数: 0 实例完成率: 100.00% 失败实例未处理数: 0	执行方式: 立即执行 开始时间: 2021-12-15 18:00:49 结束时间: 2021-12-15 18:00:49 补数据总时长: 00:00:00		暂停 终止 恢复补数据
100000040	批量补数任务_20211104_17192 5	成功	2021-11-03 00:00:00 2021-11-03 23:59:59	涉及任务数: 1 (1) 实例总数: 1 实例完成率: 100.00% 失败实例未处理数: 0	执行方式: 立即执行 开始时间: 2021-11-04 17:19:32 结束时间: 2021-11-04 17:21:00 补数据总时长: 00:01:28		暂停 终止 恢复补数据
100000014	批量补数任务_20211014_14043 4	成功	2021-10-14 00:00:00 2021-10-14 14:04:26	涉及任务数: 6 (3) 实例总数: 5 实例完成率: 100.00% 失败实例未处理数: 0	执行方式: 立即执行 开始时间: 2021-10-14 14:04:42 结束时间: 2021-10-14 14:26:51 补数据总时长: 00:22:08		暂停 终止 恢复补数据

说明:

1. **涉及任务数**的第一个数字表示需要补数据的任务总数，括号中的数字表示源头任务数量。
2. **实例完成率**表示计划补数据的实例总数的完成率（直接运行成功或最近一次重跑成功即算成功）。
3. 补数据任务支持暂停，补数据总时长会计算暂停的时间。

使用准备

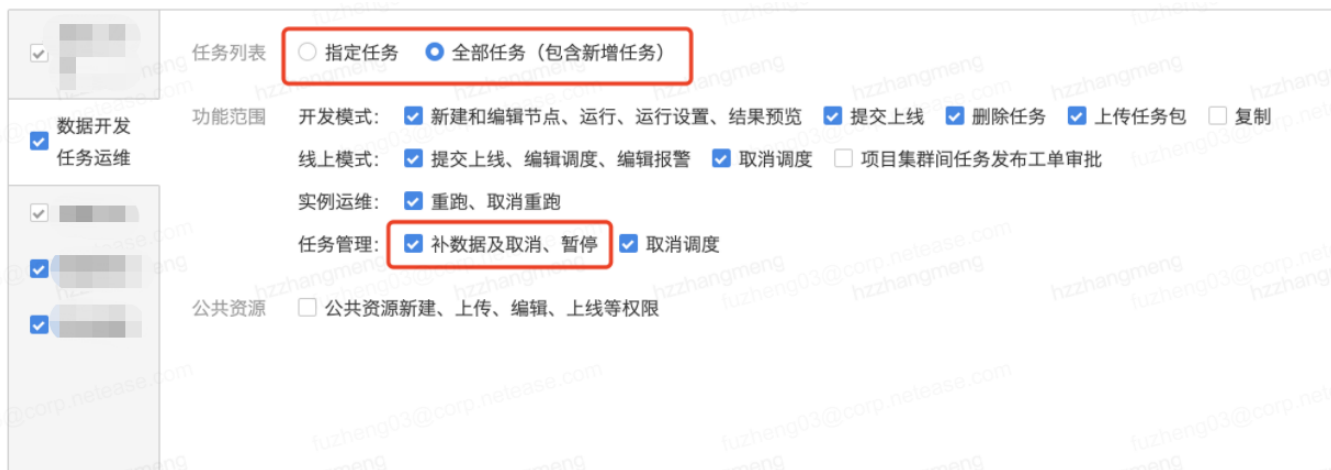
功能使用场景:

- 当任务新建后希望对历史数据进行加工，可通过补数据功能选择历史某个时间段来执行任务；
- 当任务某个时间段内数据有问题或者任务异常，可通过补数据对历史异常的数据或任务进行重新生成实例并执行，达到修复历史数据的目的。

操作权限说明

补数据操作需要具备以下两个权限:

- 在**项目中心的功能权限**模块中指定任务并勾选补数据权限，授权后才可进行补数据操作；



- 在**项目配置的补数据时间窗口**中配置**允许补数据时间窗口**。暂停补数据时间窗口期之外的时间段才可进行补数据操作。



功能使用流程

批量补数据共包含三个步骤：**选择源头任务**、**确定下游任务**以及**配置补数据设置**。

选择源头任务

在**补数据实例运维**页面单击**批量补数据**按钮，进入配置界面。在配置页面中，首先要选择源头任务，选中一个或多个任务后移入右边已选任务框中，也可以通过批量导入功能添加任务。源头任务和已选任务都支持通过**负责人**和**任务所属基线**进行筛选。

任务选择完成后，可点击**下一步**，进入下一环节。

1 选择源头任务 2 确定下游任务 3 配置补数据设置

选择源头任务

负责人: 傅正 所属基线: 全部

任务(组)名称	负责人	所属基线
<input type="checkbox"/> Des_encrypt		-
<input type="checkbox"/> Des_hash		-
<input type="checkbox"/> Des-mysql2hive03		-
<input type="checkbox"/> Des-mysql2hive		-
<input type="checkbox"/> mysql_task		-
<input type="checkbox"/> training1.dwd_pos_pos_record		12点基线
<input checked="" type="checkbox"/> training_dim_common_employee_info		-
<input type="checkbox"/> test12		-
<input type="checkbox"/> training1.ads_adsdata_canteen_dangkou_time_1d		数仓14点基线

下一步 取消

已选任务 共3个源头任务 批量导入

负责人: 全部 所属基线: 全部 批量设置 一键清空

任务默认选中所有已启用的节点, 可点击“设置”修改

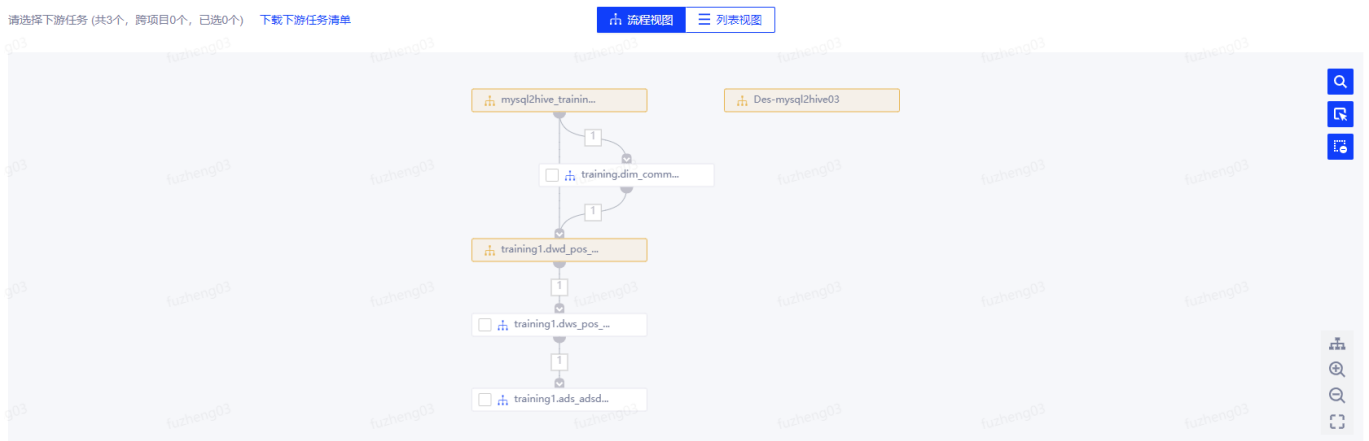
任务(组)名称	负责人	所属基线	操作
<input type="checkbox"/> Des-mysql2hive03		-	设置
<input type="checkbox"/> mysql2hive_training_canteen		-	设置
<input type="checkbox"/> training1.dwd_pos_pos_record		12点基线	设置

说明: 此处支持任务批量导入功能, 单击**批量导入**按钮, 输入任务名称, 分隔符支持空格、换行、“/”、“;”, 最多500个。对于任务组内任务, 按照**任务名称 (任务组名称)**的格式进行填写, 例如my_task1 (task_group_1)。

确定下游任务

在**确定下游任务**页面, 用户可根据实际需求勾选需要补数据的**下游任务**, 支持**流表视图**和**列表视图**两种展示方式。

- **流程视图**



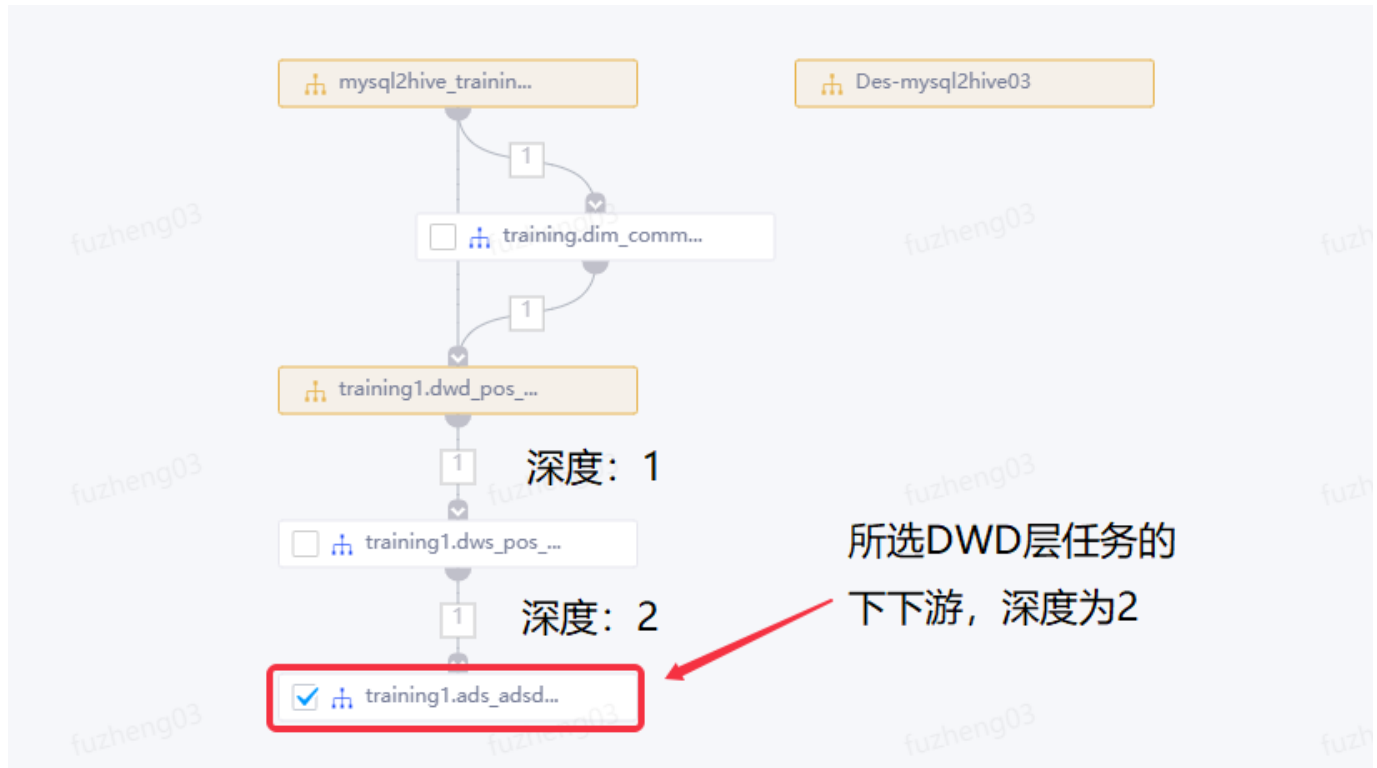
• 列表视图

请选择下游任务 (共3个, 跨项目0个, 已选0个) [下载下游任务清单](#)

任务名称	深度	调度周期	任务负责人	提交人	所属项目
<input type="checkbox"/> training_dim_common_employee_info	1	1天			
<input type="checkbox"/> training1.dws_pos_pospay_employee_1d	1	1天			
<input type="checkbox"/> training1.ads_adsdata_canteen_dangkou_time...	2	1天			

共 3 条记录 < 1 > 25 条/页

此处，**深度**表示当前任务与上游任务之间的层数关系，1表示为直接相连的下游节点，2表示下游第2层节点。



注意:

- 暂不支持对其它项目的任务执行补数据，可通过下载下游任务清单，通知对应项目的人员执行补数据。
- 针对不同调度周期的下游，确定是否需要选择执行补数据。比如上游是天调度任务，存在下游是小时调度任务，是否需要小时任务执行补数据等。
- 针对有配置调度日历的任务（包括下游任务），在补数据时仅会对计划执行时间范围内且调度日历中已配置“调度日”的日期才会生成实例，不在调度日的日期不会生成实例。

配置补数据设置

配置补数据设置包括三块配置内容：**基础设置**、**执行设置**以及**通知设置**。

基础设置包括**补数据任务名称**和**计划执行时间范围**，可通过日历设定补数据时间范围。此外，**计划执行时间范围**还支持通过**今日截止此刻**和**昨日**两个选项快速设置补昨日数据和补今日数据。

基础设置

* 补数据任务名称

* 计划执行时间范围 今日截止此刻 昨日

执行设置包括**执行方式**、**跨周期自依赖**、**补数据时间顺序**、**并发设置**、**队列**、**高优先级**、**自动重试**、**失败设置**以及**自定义参数**。

- **执行方式**：支持**立即执行**和**延后执行**两种方式，默认为**立即执行**，点击**延后执行**，可设置开始执行时间。

执行方式 立即执行 延后执行

- **跨周期自依赖**：**开启**表示某任务的实例依赖上一周期实例成功后，才可运行；**关闭**表示某任务的实例不依赖上一周期，各周期期间实例相互独立。开启后不支持补数据根据时间倒序。
- **补数据时间顺序**：当**跨周期自依赖**处于关闭状态，支持**正序**和**倒序**两种方式。**正序**表示补数据日期从最早开始执行；**倒序**表示补数据日期从最近开始执行。

补数据时间顺序 正序 倒序

- **并发设置**：表示当连续补多个周期的数据时，同时满足并发数，则可以同时运行多个周期的实例，这样可以提升补数据的效率，当然也会消耗更多的资源。

点击**展开更多设置**，可以继续配置更多执行参数。

收起更多设置

队列

高优先级

自动重试 若运行的节点开启了自动重试，按照节点设置的规则优先

失败设置

自定义参数

参数可能存在重复设置的问题，其优先级为：节点内单独设置的参数优先级最高，此处配置的参数次之，任务内引用参数组的最低。

参数名称	参数值
	无数据

+ 添加

- **队列**：默认为任务自身的队列，可切换选择其他具有权限的队列。
- **高优先级**：支持**开启**和**关闭**两种状态。开启后在动态资源分配中将抢占更多资源。
- **自动重试**：开启后支持**重试次数**和**重试间隔**的设置。
- **失败设置**：同在离线开发中进行任务设置一样，支持三种模式：**仅完成执行中的节点**、**继续完成其他分支节点**以及**停止所有节点**。
- **自定义参数设置**：支持自定义参数的设置，需要注意的是参数可能存在重复设置的问题。节点内单独设置的参数优先级最高，此处配置的参数次之，任务内引用参数组的最低。

通知设置包括**通知条件**、**通知对象**以及**通知方式**。

- **通知条件**：包括**补数据任务成功**、**补数据任务终止**以及**实例出现失败**三种条件类型。
- **通知对象**：包括**当前用户**、**任务负责人**和**其他**。当选择**其他**时，可输入邮箱地址，通过“,”进行分隔。
- **通知方式**：支持邮件、短信、POPO等。

通知设置

通知条件

- 补数据任务成功 补数据任务终止 实例出现失败

通知对象

- 当前用户 任务负责人 其他

请输入邮箱地址，多个用“,”分隔

通知方式

- 邮件 短信 POPO Stone