

ALIBABA CLOUD

阿里云

智能数据构建与管理 Dataphin
管理中心

文档版本：20220519

 阿里云

法律声明

阿里云提醒您阅读或使用本文档之前仔细阅读、充分理解本法律声明各条款的内容。如果您阅读或使用本文档，您的阅读或使用行为将被视为对本声明全部内容的认可。

1. 您应当通过阿里云网站或阿里云提供的其他授权通道下载、获取本文档，且仅能用于自身的合法合规的业务活动。本文档的内容视为阿里云的保密信息，您应当严格遵守保密义务；未经阿里云事先书面同意，您不得向任何第三方披露本手册内容或提供给任何第三方使用。
2. 未经阿里云事先书面许可，任何单位、公司或个人不得擅自摘抄、翻译、复制本文档内容的部分或全部，不得以任何方式或途径进行传播和宣传。
3. 由于产品版本升级、调整或其他原因，本文档内容有可能变更。阿里云保留在没有任何通知或者提示下对本文档的内容进行修改的权利，并在阿里云授权通道中不时发布更新后的用户文档。您应当实时关注用户文档的版本变更并通过阿里云授权渠道下载、获取最新版的用户文档。
4. 本文档仅作为用户使用阿里云产品及服务的参考性指引，阿里云以产品及服务的“现状”、“有缺陷”和“当前功能”的状态提供本文档。阿里云在现有技术的基础上尽最大努力提供相应的介绍及操作指引，但阿里云在此明确声明对本文档内容的准确性、完整性、适用性、可靠性等不作任何明示或暗示的保证。任何单位、公司或个人因为下载、使用或信赖本文档而发生任何差错或经济损失的，阿里云不承担任何法律责任。在任何情况下，阿里云均不对任何间接性、后果性、惩戒性、偶然性、特殊性或刑罚性的损害，包括用户使用或信赖本文档而遭受的利润损失，承担责任（即使阿里云已被告知该等损失的可能性）。
5. 阿里云网站上所有内容，包括但不限于著作、产品、图片、档案、资讯、资料、网站架构、网站画面的安排、网页设计，均由阿里云和/或其关联公司依法拥有其知识产权，包括但不限于商标权、专利权、著作权、商业秘密等。非经阿里云和/或其关联公司书面同意，任何人不得擅自使用、修改、复制、公开传播、改变、散布、发行或公开发表阿里云网站、产品程序或内容。此外，未经阿里云事先书面同意，任何人不得为了任何营销、广告、促销或其他目的使用、公布或复制阿里云的名称（包括但不限于单独为或以组合形式包含“阿里云”、“Aliyun”、“万网”等阿里云和/或其关联公司品牌，上述品牌的附属标志及图案或任何类似公司名称、商号、商标、产品或服务名称、域名、图案标示、标志、标识或通过特定描述使第三方能够识别阿里云和/或其关联公司）。
6. 如若发现本文档存在任何错误，请与阿里云取得直接联系。

通用约定

格式	说明	样例
 危险	该类警示信息将导致系统重大变更甚至故障，或者导致人身伤害等结果。	 危险 重置操作将丢失用户配置数据。
 警告	该类警示信息可能会导致系统重大变更甚至故障，或者导致人身伤害等结果。	 警告 重启操作将导致业务中断，恢复业务时间约十分钟。
 注意	用于警示信息、补充说明等，是用户必须了解的内容。	 注意 权重设置为0，该服务器不会再接受新请求。
 说明	用于补充说明、最佳实践、窍门等，不是用户必须了解的内容。	 说明 您也可以通过按Ctrl+A选中全部文件。
>	多级菜单递进。	单击设置> 网络> 设置网络类型。
粗体	表示按键、菜单、页面名称等UI元素。	在结果确认页面，单击 确定 。
Courier字体	命令或代码。	执行 <code>cd /d C:/window</code> 命令，进入Windows系统文件夹。
斜体	表示参数、变量。	<code>bae log list --instanceid</code> <i>Instance_ID</i>
[] 或者 [a b]	表示可选项，至多选择一个。	<code>ipconfig [-all -t]</code>
{ } 或者 {a b}	表示必选项，至多选择一个。	<code>switch {active stand}</code>

目录

1.概述	05
2.配置Dataphin成员	06
3.安全设置	11
4.智能引擎	13
4.1. 数据倾斜优化	13
4.2. 配置自定义任务参数	14
4.3. 设置即席查询的引擎	15

1.概述

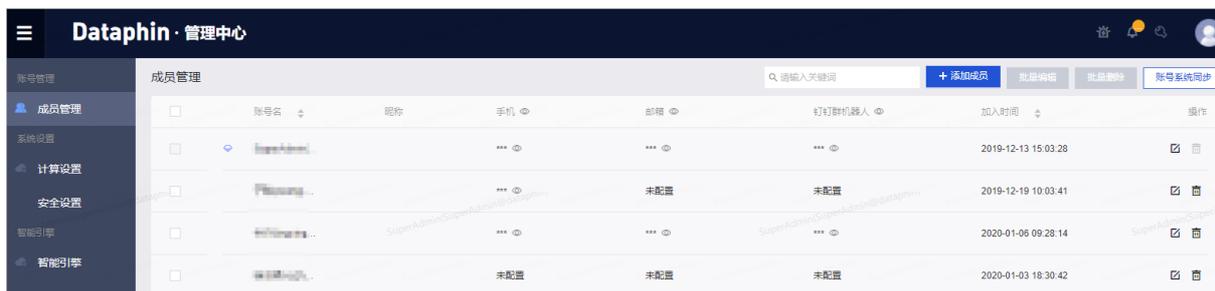
在使用Dataphin前，超级管理员需要通过管理中心添加Dataphin成员及设置Dataphin实例的计算引擎。同时您也可以配置计算和存储资源的规则，优化逻辑表任务、SQL查询任务及设置查询引擎，以提高数据构建效率。

使用限制

仅支持超级管理员在管理中心管理成员、设置Dataphin计算引擎、配置智能引擎。

功能介绍

管理中心支持管理Dataphin成员、设置Dataphin实例的计算引擎、配置Dataphin实例的智能引擎。



功能	功能介绍
配置Dataphin成员	<ul style="list-style-type: none"> 查看当前Dataphin成员列表信息。 添加RAM用户至Dataphin成员列表。 配置成员联系方式。 删除Dataphin成员列表中的成员。 如果RAM用户账号信息发生变更，您可以通过账号系统同步更新Dataphin成员列表中。
设置Dataphin实例的计算引擎	<p>在开始使用Dataphin前，您需要完成Dataphin实例计算引擎的设置，即配置Dataphin实例的计算集群地址。在后续数据研发过程中，计算引擎用于采集、连接及管理元数据。</p>
智能引擎	<p>配置计算和存储资源的参数规则，优化逻辑表任务、SQL查询任务及设置查询引擎，以提高数据构建效率。</p>

2.配置Dataphin成员

在您开始使用Dataphin构建智能数据前，需要将您企业的开发人员加入至Dataphin系统后，再将Dataphin成员加入至相应的项目空间并赋予不同的开发角色。本文为您介绍如何添加Dataphin成员、配置成员的联系方式及账号系统同步等操作。

页面介绍

1. 使用超级管理员账号（阿里云主账号）登录Dataphin控制台。
2. 在Dataphin首页，单击顶部菜单栏的**管理中心**。
3. 在**管理中心**页面，单击左侧导航栏的**成员管理**。
4. 在**成员管理**页面，不同角色查看的信息及支持的操作不同。

角色	描述
非超级管理员	查看系统内所有成员的 账号名、昵称、加入时间 信息。但仅能查看和编辑成员自己的 手机、邮箱、钉钉群机器人 信息。
超级管理员	<ul style="list-style-type: none"> 查看系统内所有成员的账号名、昵称和加入时间。您可以单击手机、邮箱和钉钉群机器人列下的  图标，查看系统内所有已配置联系方式的成员手机、邮箱、钉钉群机器人信息。 查看信息后，单击手机、邮箱、钉钉群机器人地址后的  图标，可以隐藏手机、邮箱、钉钉群机器人信息。同时您也可以编辑所有成员的手机、邮箱、钉钉群机器人的信息。 单击成员管理页面首行手机、邮箱、钉钉群机器人后的  图标，批量查看所有配置了联系方式成员的信息。 查看信息后，单击手机、邮箱、钉钉群机器人地址后的  图标，可批量隐藏手机、邮箱、钉钉群机器人地址信息。

添加成员

说明

- 系统仅支持超级管理员添加Dataphin的成员。
- 配置主账号的Access Key，详情请参见[同步RAM用户账号至Dataphin](#)。

1. 在**成员管理**页面，单击页面上方的**添加成员**。
2. 在**添加成员**对话框，单击成员名显示框。
3. 选择成员后，单击下拉列表和成员名显示框之外的区域。

系统支持选择多个成员。您可以在搜索框输入成员名关键字，搜索成员。



4. 单击**确定**，完成Dataphin成员的添加。

配置成员联系方式

说明

- 超级管理员可以配置所有成员的联系方式。
- 非超级管理员仅限配置自己的联系方式。

超级管理员批量配置成员的联系方式：

1. 在**成员管理**页面，选中多个成员后，单击页面上方的**批量编辑**。
2. 在**编辑联系方式**对话框，配置参数。



参数	描述
手机	填写手机号码。 <div style="border: 1px solid #ccc; background-color: #e6f2ff; padding: 5px; margin-top: 10px;"> ? 说明 系统仅支持添加中国内地手机号码。 </div>
邮箱	填写邮箱地址。
钉钉群机器人	配置钉钉群机器人。鼠标悬停至图标，查看添加钉钉群机器人的方法。 <div style="background-color: #1a202c; color: white; padding: 10px; margin-top: 10px;"> <p>如何通过钉钉群机器人接收消息?</p> <ol style="list-style-type: none"> 1) 在PC版中打开您要接收报警通知的钉钉群 2) 单击右上角的群设置图标，打开群设置弹窗 3) 单击智能助手-添加机器人，打开群机器人窗口，单击自定义，创建一个用于接收报警和消息通知的钉钉机器人 4) 在机器人详情窗口，单击添加，进入添加机器人窗口 5) 输入机器人名字，例如Dataphin报警通知 6) 进行安全设置，选择自定义关键词方式，并输入关键词“Dataphin”，单击完成即可 7) 单击复制，复制webhook地址，单击完成，即可完成添加机器人 8) 将上述创建好的钉钉机器人webhook地址添加到报警联系人中，该联系人即可通过钉钉群接收报警通知 </div>

3. 单击确定，完成多个成员联系方式的配置。

配置单个成员的联系方式：

1. 在成员管理页面，单击需要配置联系方式的成员操作列下图标。
2. 在编辑联系方式对话框，配置参数。

编辑联系方式
✕

账号名
昵称 -

邮箱

国家/地区 中国内地 (+86)

手机

钉钉群机器人

如何通过钉钉群机器人接收消息?

取消

确定

参数	描述
----	----

参数	描述
手机	填写手机号码。 <div style="border: 1px solid #ccc; background-color: #e6f2ff; padding: 5px; margin-top: 10px;"> ? 说明 系统仅支持添加中国内地手机号码。 </div>
邮箱	填写邮箱地址。
钉钉群机器人	配置钉钉群机器人。鼠标悬停至  图标，查看添加钉钉群机器人的方法。 <div style="background-color: #2c3e50; color: white; padding: 10px; margin-top: 10px;"> <p>如何通过钉钉群机器人接收消息?</p> <ol style="list-style-type: none"> 1) 在PC版中打开您要接收报警通知的钉钉群 2) 单击右上角的群设置图标，打开群设置弹窗 3) 单击智能群助手-添加机器人，打开群机器人窗口，单击自定义，创建一个用于接收报警和消息通知的钉钉机器人 4) 在机器人详情窗口，单击添加，进入添加机器人窗口 5) 输入机器人名字，例如Dataphin报警通知 6) 进行安全设置，选择自定义关键词方式，并输入关键词“Dataphin”，单击完成即可 7) 单击复制，复制webhook地址，单击完成，即可完成添加机器人 8) 将上述创建好的钉钉机器人webhook地址添加到报警联系人中，该联系人即可通过钉钉群接收报警通知 </div>

3. 单击**确定**，完成单个成员联系方式的配置。

账号系统同步

? **说明** 只有配置了Access Key，才可以同步账号信息。配置主账号的Access Key，详情请参见[同步RAM用户账号至Dataphin](#)。

如果账号信息发生了变更，则单击**成员管理**页面上方的**账号系统同步**，同步刷新账号信息。

删除成员

对于已添加的Dataphin成员，系统同时支持删除成员。

? **说明**

- 超级管理员账号（即主账号）默认不支持被删除，只有子账号可以被删除。
- 删除成员只是将子账号从Dataphin系统成员中移除，并没有删除该子账号。
如果您已确认不再使用该子账号开发数据，您可以使用主账号登录[RAM控制台](#)，删除该子账号。
- 系统仅支持超级管理员删除Dataphin的成员。

批量删除成员：

1. 在**成员管理**页面，选中多个成员后，单击页面上方的**批量删除**。
2. 在删除确认对话框中，单击**确定**，完成多个成员的删除。

删除单个成员：

1. 在**成员管理**页面，单击需要删除的成员操作列下的图标。

2. 在删除确认对话框中，单击**确定**，完成单个成员的删除。

3.安全设置

您可以通过设置项目安全模式，保障数据的安全性。本为您介绍如何设置项目的安全模式。

背景信息

Dataphin管理中心支持项目安全模式设置功能，在未设置的情况下，会存在以下安全问题：

- 不同项目，在项目管理员不知情的情况下，用户可以随意跨项目进行DDL操作。

说明 DDL语句详情，请参见[DDL语句](#)。

- 同一项目，用户可以在Dev环境直接操作Prod环境的数据，造成Prod环境数据泄漏或意外变更。

下表以开发者角色为例，介绍安全模式开启与关闭的情况下，对项目的操作权限。

操作	默认（权限模式均未开启）	跨项目安全模式（开启）	生产数据安全模式（开启）
当前项目Dev环境的所有表操作	支持	支持	支持
当前项目Prod环境的所有表操作	支持	支持	不支持
跨项目查询	支持	支持	支持
跨项目DDL	支持	不支持	支持

注意 设置安全模式后，项目数据有更高的安全性，但是会导致系统使用流畅程度下降，建议您根据业务需求进行设置。

操作步骤

1. 登录Dataphin控制台。
2. 在Dataphin控制台页面，选择工作区地域后，单击进入Dataphin>>。
3. 在Dataphin页面，单击顶部菜单栏的管理中心。
4. 在管理中心页面，按照下图指引设置相应的安全模式。



安全模式	说明
跨项目安全模式	<ul style="list-style-type: none"> ◦ 关闭：在具备操作目标对象权限的前提下，允许跨项目执行DDL任务。 ◦ 开启：禁止跨项目执行和提交DDL任务。
生产数据安全模式	<ul style="list-style-type: none"> ◦ 关闭：允许在Dev环境修改Prod环境数据。 ◦ 开启：对于Dev-Prod 模式的项目，在Dev环境将不能修改Prod环境的数据。 <div style="border: 1px solid #ccc; background-color: #e6f2ff; padding: 5px; margin-top: 10px;"> <p> 说明 开启后，如需修改Prod环境的数据，需要走发布流程，详情请参见发布中心概述。</p> </div>

4. 智能引擎

4.1. 数据倾斜优化

本章节为您介绍如何通过Dataphin数据倾斜优化功能对调度中触发的异常状态任务做出优化。

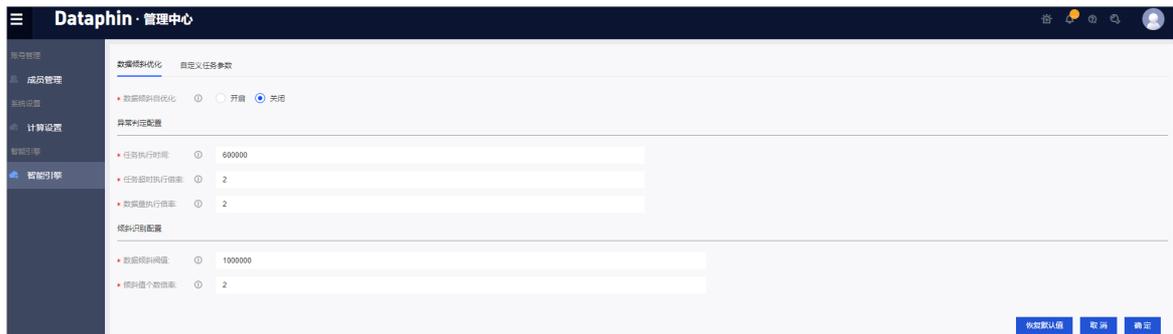
数据倾斜优化简介

数据倾斜优化在Dataphin调度中的应用如下：

1. 如果已开启数据倾斜优化功能，在您首次提交任务中，当执行时间、任务超时倍率、数据量膨胀倍率超过设定的阈值时，会触发异常判定功能，完成初步的判断。
2. 任务执行的过程中，会同步调度数据探查，针对关联的字段，定位到出现频次过高的热点数据。根据该探查信息，对执行中的任务代码进行重构、规避数据倾斜，同时终止原正在执行中的任务，重新提交新的代码，实现对逻辑表任务的优化。
3. 倾斜状态会被记录下来，作为次日数据的探查触发点，如果次日探查结果仍旧有数据倾斜，则继续使用优化后的代码，如果探查结果无数据倾斜，则使用非数据倾斜优化的代码。
4. 对于重跑的逻辑表任务：
 - 如果引用到已经探查过的数据表及对应字段，且探查结果为该字段有倾斜，则会根据探查信息，直接对提交的SQL进行优化，并使用优化后的代码进行执行。
 - 如果重跑任务中使用的数据表未被探查过，则与已开启数据倾斜功能触发的异常判断方式相同，调度探查任务，根据探查结果重组代码。智能数据引擎自动生成的任务和模型在生产过程中遇到了数据倾斜的问题，由智能数据引擎自行处理。

数据倾斜优化配置

1. 登录Dataphin控制台。
2. 在Dataphin控制台页面，选择工作区地域后，单击进入Dataphin>>。
3. 在Dataphin首页，单击顶部菜单栏的管理中心。
4. 在左侧菜单栏中选择智能引擎，默认进入数据倾斜优化配置页签。
5. 根据页面提示配置参数，详细内容请参见下表。



参数名称	描述	配置示例
数据倾斜自优化	全局配置智能数据引擎是否开启数据倾斜自优化功能。	开启

参数名称	描述	配置示例
任务执行时间	任务执行时间阈值，是触发对任务进行数据探查的基础条件。即当任务时间超过设定值时，会触发异常分析及判定机制，单位ms。	600000
任务超时执行倍率	某任务执行时间倍率阈值，是触发对任务进行数据探查的基础条件。即当某任务执行时间超过平均执行时间该倍率时，会触发异常任务分析与判定机制。	2
数据量执行倍率	某任务处理数据量倍率阈值，是触发对任务进行数据探查的基础条件。即当某任务处理的数据量超过所有任务平均处理数据量该倍率时，会触发异常任务分析与判定机制。	2
数据倾斜阈值	倾斜值数据量阈值设置，是判定数据集中倾斜的规则。即当数据集的某值大于该阈值时，判定该计算任务为倾斜。	1000000
倾斜值个数倍率	倾斜值倾斜倍率设置，是判定数据集中倾斜值的规则。即当数据集中某值个数大于所有值平均个数的该倍率时，判定为计算任务为倾斜。	2

6. 单击**确定**，完成参数的配置。

您也可以在**数据倾斜优化**页面，单击**恢复默认值**。在确认框中，单击**确定**，恢复默认值。

7. 在确认框中，单击**确定**，完成数据倾斜优化参数的配置。

4.2. 配置自定义任务参数

本文为您介绍如何配置自定义任务参数，防止系统分配资源时出现不合理的现象。

背景信息

自定义任务参数用于指定计算引擎的某些参数，避免由于企业的计算引擎特异性导致的引擎适配问题，同时提高了逻辑表任务的效率。

操作步骤

1. 登录**Dataphin控制台**。
2. 在Dataphin控制台页面，选择**工作区地域**后，单击**进入Dataphin>>**。
3. 在Dataphin首页的顶部菜单栏，单击**管理中心**。
4. 在管理中心页面，单击左侧导航栏中的**智能引擎**。
5. 在智能引擎页面，单击**自定义任务参数**页签。
6. 根据**自定义任务参数**页面提示，配置参数。

数据倾斜优化 自定义任务参数

自定义参数生效: 开启 关闭

自定义参数配置

参数配置: 示例: set oneservice.optimize.skewjoin = true/false;
set hive.exec.dynamic.partition.mode=nonstrict; (动态分区开启/关闭)
set hive.vectorized.execution.enabled = false; (向量查询开启/关闭)
set hive.vectorized.execution.reduce.enabled = false;

0/2048

恢复默认值 取消 确定

参数	描述
自定义参数生效	选择自定义任务参数需要开启还是关闭。
自定义参数配置	配置自定义任务参数，参数需要与Dataphin系统绑定的计算引擎源的特性一致，示例如下。 <pre>set odps.sql.mapper.cpu=100 set odps.sql.mapper.split.size=256 set odps.sql.reducer.cpu=100 set odps.sql.joiner.cpu=100</pre> 说明 自定义参数配置采用优先级策略，任务节点的物理化配置较智能引擎的自定义任务参数优先级高。

7. 参数配置完后，单击**确定**。
8. 在确认对话框中，单击**确定**，完成自定义任务参数的配置。

您也可以恢复计算引擎的默认值，在自定义任务参数页面，单击**恢复默认值**。在确认对话框中，单击**确定**，完成计算引擎的默认值的恢复。

4.3. 设置即席查询的引擎

系统根据汇总逻辑表的来源表和统计周期，会生成多张物化表，当使用 `select *` 语法查询汇总逻辑表时，底层物化SQL会将所有物化表全部关联，造成大量资源的浪费，您可以通过设置即席查询的引擎，规避资源浪费的问题。本文为您介绍如何设置即席查询引擎。

操作步骤

1. 登录Dataphin控制台。
2. 在Dataphin控制台页面，选择工作区地域后，单击**进入Dataphin>>**。
3. 在Dataphin首页，单击顶部菜单栏的**管理中心**。
4. 在管理中心页面，单击左侧导航栏的**智能引擎**。
5. 在智能引擎页面，单击**查询引擎设置**页签。
6. 在查询引擎设置页签，选择是否允许在即席查询中使用 `select*` 语法查询汇总逻辑表。
 - 单击关闭后，在确认对话框中，单击**确定**，关闭使用 `select *` 语法查询汇总逻辑表的权限。

 **注意** 因为关闭使用 `select *` 语法的权限，可能会造成已经使用了 `select *` 和 `SELECT *` 语法的生产任务运行失败，因此需确保生产任务中没有 `select *` 和 `SELECT *` 语法。您可以在全局代码搜索入口中，输入 `select *` 和 `SELECT *` 查询，详情请参见[概述](#)。

- 单击开启，开启使用 `select *` 语法查询汇总逻辑表的权限。

 **说明** 系统默认支持使用 `select *` 语法查询汇总逻辑表。

7. 单击确定。

如果您需要恢复系统默认值，则单击**恢复默认值**。系统自动开启使用 `select *` 语法查询汇总逻辑表的权限。