

ALIBABA CLOUD

# 阿里云

## 智能数据构建与管理 Dataphin 准备工作

文档版本：20220420

 阿里云

## 法律声明

阿里云提醒您 在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。
<b>粗体</b>	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>

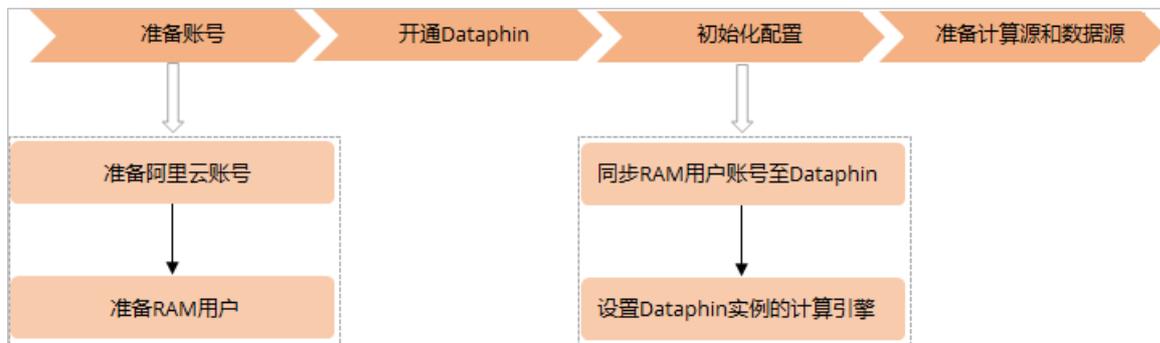
# 目录

1.准备工作简介	05
2.准备账号	06
2.1. 准备阿里云账号	06
2.2. 准备RAM用户	07
3.开通Dataphin	09
4.初始化配置	13
4.1. 同步RAM用户账号至Dataphin	13
4.2. 设置Dataphin实例的计算引擎	14
5.准备数据源和计算源	16

# 1.准备工作简介

您需要完成注册阿里云账号、购买Dataphin实例、初始化配置Dataphin等一系列准备工作，才能开始Dataphin的正式使用。

Dataphin准备工作流程简介。



准备账号

开通Dataphin

准备数据源和计算源

初始化配置

准备阿里云账号

同步RAM用户账号至Dataphin

准备RAM用户

设置Dataphin实例的计算引擎

## 1. 准备账号

### i. 准备阿里云账号

只能使用阿里云账号开通Dataphin。

### ii. 准备RAM用户

为保证项目数据安全，建议您创建RAM用户并交由其他用户使用，实现对参与Dataphin项目的人员权限进行严格把控。

## 2. 开通Dataphin

您需要开通Dataphin服务，才可以使用Dataphin。

## 3. 初始化配置

### i. 同步RAM用户账号至Dataphin

如果您使用RAM用户对Dataphin进行精细的权限管理，则需要将其他人员使用的RAM用户账号同步至Dataphin中，才可以使用Dataphin。

### ii. 设置Dataphin实例的计算引擎

在开始使用Dataphin前，您需要完成Dataphin实例计算引擎的设置。

## 4. 准备数据源和计算源

在创建项目空间前，您需要准备用于读取和写入数据的数据源、用于提供计算资源的计算源。

## 2. 准备账号

### 2.1. 准备阿里云账号

在使用Dataphin产品前，您需要准备好阿里云账号。本文为您介绍如何创建阿里云账号并完成账号相关准备工作。

#### 步骤一：注册阿里云账号

1. 打开[阿里云官网](#)。
2. 在阿里云官网右上角，单击立即注册。
3. 在注册账号页面，按照操作提示完成账号注册。

创建成功的阿里云账号，会作为阿里云系统识别的资源消费账户，阿里云账号拥有该账户的所有权限。

 **注意** 为保证账号和密码的安全性，请勿借给他人使用，并定期更新密码。

#### 步骤二：阿里云账号实名认证

为保证后续操作顺利进行，请务必完成实名认证操作。阿里云账号需要进行实名制认证后，才可以购买和使用阿里云的产品。

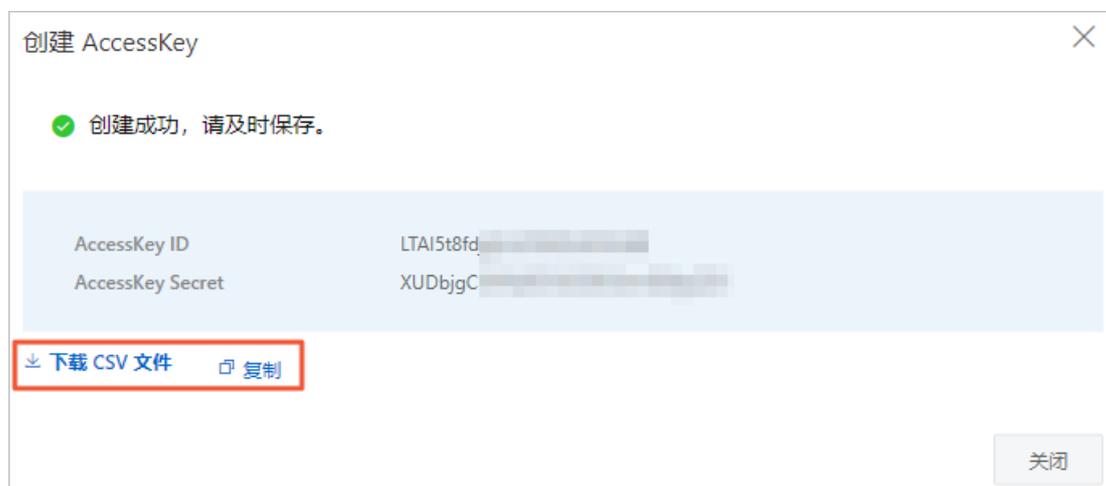
1. 进入[账号中心](#)页面。
2. 在左侧导航栏，单击实名认证，按照操作提示完成账号实名认证。

如果您是企业用户，推荐进行企业认证，以便获取更多便利。更多实名认证操作信息，请参见[实名认证](#)。

#### 步骤三：创建访问密钥AccessKey

为保证后续作业可以顺利运行，您需要创建一个运行密钥AccessKey。该密钥与注册阿里云账号时设置的账号、密码不同，主要用于在阿里云各产品间互相认证时使用权限。AccessKey包括AccessKey ID和AccessKey Secret。

1. 进入[AccessKey](#)页面。
2. 在AccessKey页面，单击创建AccessKey，系统会自动创建AccessKey。
3. 在创建AccessKey对话框中，单击下载CSV文件或复制，保存创建的AccessKey信息，并单击关闭。



4. 在AccessKey页面，查看已经创建的AccessKey的状态，并对其可以执行禁用、删除操作。

 注意

- 为保证AccessKey ID和AccessKey Secret的安全性，请勿借给他人使用，一旦有泄漏的风险，请及时禁用或更新AccessKey。
- AccessKey只在创建时显示AccessKey ID和AccessKey Secret，不提供查询功能，请妥善保管。
- 禁用AccessKey后，使用该AccessKey的服务将运行失败并报错，请谨慎操作。如果AccessKey状态变更，需要及时关注使用该AccessKey的产品和服务。

### 后续步骤

- 如果需要添加其他人员共同使用Dataphin，您还需要创建RAM用户进行安全管控。更多创建RAM用户信息，请参见[准备RAM用户](#)。
- 如果仅您本人使用Dataphin，请开通Dataphin。更多开通Dataphin信息，请参见[开通Dataphin](#)。

 说明 推荐您创建RAM用户，并通过RAM用户使用Dataphin，为处理因业务增长可能会产生的权限管控问题提供便利。

## 2.2. 准备RAM用户

为保证项目数据安全，推荐您创建RAM用户并交由其他用户使用，实现对参与Dataphin项目的人员权限进行严格把控。本文为您介绍如何创建RAM用户。

### 背景信息

在创建RAM用户前，请您了解阿里云账号和RAM用户及其在Dataphin中的用户角色及其权限，并结合实际需求进行选择。

账号	对应到Dataphin中的用户角色及其权限说明
阿里云账号	即阿里云用户的主账号，默认会成为Dataphin的超级管理员账号。通常，企业管理者使用阿里云主账号，由于其权限范围宽泛，不推荐企业中开发者、运维者或分析师等角色的员工使用阿里云账号。
RAM用户	<p>从属于阿里云账号的子账号。您可以将RAM用户同步至Dataphin后，添加为项目空间的成员并授予项目管理员、开发者、分析师、运维者或访客等角色，以实现对Dataphin进行精细的权限管理。RAM用户在Dataphin中可以被授予的角色及权限，请参见<a href="#">用户角色和权限</a>。</p> <p>为避免产生项目数据安全性问题，推荐您创建RAM用户并交由其他用户使用，这样可以：</p> <ul style="list-style-type: none"> <li>● 基于企业的业务组织，在您的阿里云账号中，给企业中不同职能部门的员工创建RAM用户，让员工拥有唯一安全凭证，并使用Dataphin。</li> <li>● 基于企业用户的职能，设置不同的访问权限，以实现用户间的权限隔离。</li> </ul> <p>例如，在使用Dataphin过程中，某企业规划不同的员工负责数据开发、数据运维、数据分析阶段的工作，并且企业管理者对员工间的权限控制和数据安全有要求。因此，您可以创建3个RAM用户，将RAM用户同步至Dataphin后，添加为项目空间的成员并分别授予开发者、运维者和分析师的角色。</p>

## 前提条件

在开始创建RAM用户前，请确认您已完成阿里云账号的创建。如何创建阿里云账号，请参见[准备阿里云账号](#)。

## 注意事项

- RAM用户归属于阿里云账号，本身不拥有资源，也没有独立的计量计费机制。
- RAM用户在阿里云各产品中操作产生的费用，将在RAM用户归属的阿里云账号下统一计费。

## 步骤一：创建RAM用户

1. 使用阿里云账号登录[RAM控制台](#)。
2. 在左侧导航栏，选择身份管理 > 用户。
3. 在用户页面，单击创建用户。
4. 在创建用户页面的用户账号信息区域，输入登录名称和显示名称。

 说明 单击添加用户，可一次性创建多个RAM用户。

5. 在访问方式区域，选择访问方式。
  - 控制台访问：设置控制台登录密码、重置密码策略和多因素认证策略。

 说明 自定义登录密码时，密码必须满足密码复杂度规则。关于如何设置密码复杂度规则，请参见[设置RAM用户密码强度](#)。

- OpenAPI调用访问：自动为RAM用户生成访问密钥（AccessKey），支持通过API或其他开发工具访问阿里云。

 说明 为了保障账号安全，建议仅为RAM用户选择一种登录方式，避免RAM用户离开组织后仍可以通过访问密钥访问阿里云资源。

6. 单击确定。
7. 在用户信息页面，单击下载CSV文件或操作列的复制，保存RAM用户的登录名称和登录密码。

## 步骤二：将RAM用户账号交由其他用户使用

将创建的RAM用户账号交由其他用户使用时，需要提供如下信息：

- RAM用户登录链接。

打开[RAM控制台](#)，在概览页面右上角的账号管理区域中，用户登录地址即为RAM用户的登录链接。
- RAM用户的账号和密码，即[步骤一](#)中保存的RAM用户登录名称和登录密码。

## 后续步骤

准备好RAM用户后，即可开通Dataphin服务，请参见[开通Dataphin](#)。

## 3. 开通Dataphin

您需要开通Dataphin服务，才可以使用Dataphin。本文为您介绍如何开通Dataphin。

### 前提条件

请确认您已完成阿里云账号的创建，并完成账号的实名认证。

如何创建阿里云账号及进行实名认证，请参见[准备阿里云账号](#)。

### 购买限制

暂不支持金融云账号购买Dataphin实例。

### 注意事项

开通Dataphin前，请您了解以下几点：

- 购买前建议您先联系阿里云[售前咨询](#)并提供企业数据建设需求及背景信息。待阿里云售前咨询帮您确认Dataphin当前版本及功能是否符合您的需求后，会为您开通购买权限。
- Dataphin不支持无条件退款，购买前请您仔细确认购买的Dataphin版本。

购买Dataphin后，如果因特殊情况需要申请退款，请您提[工单](#)并联系对应客户经理进行沟通，确认符合退款条件后（非产品问题不予退款），系统将根据您的使用情况扣除对应费用，并后台执行退款操作。

- Dataphin支持包年包月购买模式。
- Dataphin仅支持新用户购买Dataphin实例。Dataphin实例到期停机且在14天的保留期内时，仅支持续费操作。

### 购买智能研发版的Dataphin实例

1. 使用主账号登录[阿里云官网](#)。
2. 在阿里云官网，鼠标悬停至[产品](#)后，再将鼠标悬停至左侧导航栏的[大数据](#)上，单击[数据开发和治理](#)列下的[智能数据构建与管理Dataphin](#)。
3. 在Dataphin产品详情页，单击[立即购买](#)，进入Dataphin购买页面。
4. 在购买页面，选择[地域](#)、[版本](#)、[计算引擎](#)、[数据处理单元](#)和[资产质量基本配置](#)及[购买数量](#)和[购买时长](#)，同时您可以根据业务需求选择[增值功能包](#)。

地域
华东2 (上海) | 华南1 (深圳) | 华北2 (北京) | 华东1 (杭州)

不同地域之间网络互不连通。建议选择与您的业务数据所在地相同的地域，可降低网络延迟，提高您的访问速度。服务开通后不支持更改地域。您可购买多个地域的不同实例。

版本
智能研发版 | 基础研发版

请选择您要使用的计算引擎。服务开通后不支持切换。请注意，此处的选择仅影响系统初始化时的配置，您仍需要单独购买计算引擎。[点击购买MaxCompute>>](#)

计算引擎
MaxCompute

请选择您要使用的计算引擎。服务开通后不支持切换。请注意，此处的选择仅影响系统初始化时的配置，您仍需要单独购买计算引擎。[点击购买MaxCompute>>](#)

数据处理单元
200 | 500 | 1000 | 1500 | 2000 | 2500 | 3000

数据处理单元代表您可以生产的数据对象数量，包括集成任务、离线计算任务、实时计算任务、明细数据表及派生指标。详情见 [Dataphin计费说明>>](#)

资产质量
500规则(免费) | 1000规则 | 1500规则 | 2000规则 | 3000规则 | 5000规则

资产质量支持从物理视角和逻辑视角对数据资产进行质量监控，帮助您高效配置质量规则并执行校验任务，同时为您提供清晰全面的数据质量报告和质量概览，帮助您快速了解质量规则覆盖程度并识别数据质量风险。[了解更多>>](#)

说明

Dataphin智能研发版，提供标准化的数据构建与管理功能，支持一站式数据引入、建模研发、调度运维、资产查看与权限管理，支持生产开发隔离。

您购买的数据处理单元规格越高，则分配的默认调度资源越多，免费赠送500条数据质量监控规则的配置额度。

如遇到问题，详细请参考[产品帮助文档](#)，或加入钉钉咨询群：23381533。

---

数据萃取

标准版

增值功能包，可选。需叠加版本购买。面向各行各业数据价值深度挖掘的需求，萃取为您提供以目标对象为中心的数据打通和深度挖掘，并自动生成代码与调度任务，完成实体对象识别、连接及标签生产，帮助您提升可直接应用的高价值标签数据，助力您实现业务的融会贯通并驱动创新。[了解更多>>](#)

数据服务

api base (最大500 QPS)

增值功能包，可选。需叠加版本购买。数据服务可以通过标准化、配置化方式，帮助您统一数据主题单元，并进一步支持面向主题的自助API配置、调试、研发、上线及调用监控管理，做到数据使用权限字段粒度控制，有效降低业务的数据消费门槛，同时为您保障数据开放的安全性。[了解更多>>](#)

资产管理

标准版

增值功能包，可选。需叠加版本购买，包括从元数据的采集、元数据统一模型建立、治理项为核心的驱动闭环等，以支撑数据资产管理应用层能力，服务于资产管理与开发者，构建可信可控的数据资产健康体系。[了解更多>>](#)

资产安全(公测)

标准版

数据安全支持定义数据的业务分类和安全等级并构建数据识别规则，同时支持针对敏感数据设置脱敏规则，以保障您的数据资产安全性。该功能公测期间特别推出优惠价，1000元/月。[了解更多>>](#)

购买数量

1

购买时长

1个月 | 3个月 | 6个月 | 9个月 | 1年 | 2年 | 更多时长

总配置费用 ¥ 0.00

[立即购买](#) [加入购物车](#)

区域	参数	描述
基本配置	地域	选择购买Dataphin实例的地域。系统支持选择的地域包括： <ul style="list-style-type: none"> <li>华东2 (上海)</li> <li>华南1 (深圳)</li> <li>华北2 (北京)</li> <li>华东1 (杭州)</li> </ul>
	版本	系统默认选中智能研发版。
	计算引擎	系统默认选中MaxCompute，不支持修改。
	数据处理单元	系统默认选中500，支持选择的数据处理单元数包括： <ul style="list-style-type: none"> <li>200</li> <li>500</li> <li>1000</li> <li>1500</li> <li>2000</li> <li>2500</li> <li>3000</li> </ul>

区域	参数	描述
	数据质量	系统默认选中 <b>500规则（免费）</b> ，支持选择的数据质量包括： <ul style="list-style-type: none"> <li>◦ 500规则（免费）</li> <li>◦ 1000规则</li> <li>◦ 1500规则</li> <li>◦ 2000规则</li> <li>◦ 3000规则</li> <li>◦ 5000规则</li> </ul>
增值功能包	数据萃取	系统默认将 <b>数据萃取</b> 选择为 <b>标准版</b> 。如果您暂时不需购买增值功能包，单击 <b>数据萃取</b> 后的下拉箭头选择 <b>暂不选择</b> 。
	数据服务	系统默认将 <b>数据服务</b> 选择为 <b>api.base（最大500 QPS）</b> 规格。如果您需要选择 <b>数据服务</b> 的更高规格，则单击 <b>api.base（最大500 QPS）</b> 后的下拉箭头选择规格。系统支持的数据服务规格包括： <ul style="list-style-type: none"> <li>◦ api.base（最大500 QPS）</li> <li>◦ api.s1.small（最大2500 QPS）</li> <li>◦ api.s1.emdium（最大5000 QPS）</li> <li>◦ api.s2.large（最大10000 QPS）</li> </ul> 如果您暂时不需购买数据服务增值功能包，单击 <b>数据服务</b> 后的下拉箭头选择 <b>暂不选择</b> 。
	资产治理	系统默认将 <b>资产治理</b> 选择为 <b>标准版</b> 。如果您暂时不需购买增值功能包，单击 <b>资产治理</b> 后的下拉箭头选择 <b>暂不选择</b> 。
	资产安全	当前，资产安全处于公测期。系统默认将 <b>资产安全</b> 选择为 <b>标准版</b> 。如果您暂时不需购买增值功能包，单击 <b>资产安全</b> 后的下拉箭头选择 <b>暂不选择</b> 。
购买量	购买数量	系统默认选中 <b>1</b> 。当前 <b>1</b> 个地域（Region）仅支持购买 <b>1</b> 个Dataphin实例。
	购买时长	系统默认选中 <b>1个月</b> 。系统支持选择的购买时长包括： <ul style="list-style-type: none"> <li>◦ 1个月</li> <li>◦ 3个月</li> <li>◦ 6个月</li> <li>◦ 9个月</li> <li>◦ 1年</li> <li>◦ 2年</li> <li>◦ 3年</li> </ul>

5. 请您仔细核对购买信息，确认无误后单击**立即购买**。
6. 在**确认订单**页面，请您确认购买的Dataphin实例规格。确认无误后，单击**服务协议**后的**智能数据构建与管理服务协议**，请仔细阅读服务协议内容。确认服务协议内容后，选中**我已阅读并同意智能数据构**

建与管理服务协议，单击去支付。

确认订单

参数配置

智能数据构建与管理

配置详情	付费方式	数量	时长	优惠金额	价格
地域: 华东2 (上海) 版本: 智能研发版 计算引擎: MaxCompute 数据处理单元: 200 资产质量: 500规则(免费) 数据摄取: 标准版 数据服务: api.base (最大500 QPS) 资产管理: 标准版 资产安全(公测): 标准版	包年包月	1	1个月	¥	配置费用: ¥

服务协议  我已阅读并同意智能数据构建与管理服务协议

应付款: ¥

去支付

购买Dataphin成功后，通常需要5分钟~10分钟开通服务，请您耐心等待。如果开通服务失败，系统会自动为您退款。

### 后续步骤

完成Dataphin开通后，您需要将阿里云账号下的RAM用户添加为Dataphin成员进行后续的数据开发。具体操作，请参见[同步RAM用户账号至Dataphin](#)。

## 4. 初始化配置

### 4.1. 同步RAM用户账号至Dataphin

在添加其他人员共同使用Dataphin前，您需要将其他人员使用的RAM用户账号同步至Dataphin中。本文为您介绍阿里云账号如何通过配置AccessKey同步RAM用户账号至Dataphin中。

#### 前提条件

在开始执行操作前，请确认您已具备以下条件：

- 完成阿里云账号及其访问密钥AccessKey的创建。具体操作，请参见[准备阿里云账号](#)。
- 完成阿里云账号下RAM用户的创建。具体操作，请参见[准备RAM用户](#)。

#### 操作步骤

1. 使用阿里云账号登录[Dataphin控制台](#)。
2. 在Dataphin控制台首页，单击Access Key配置。

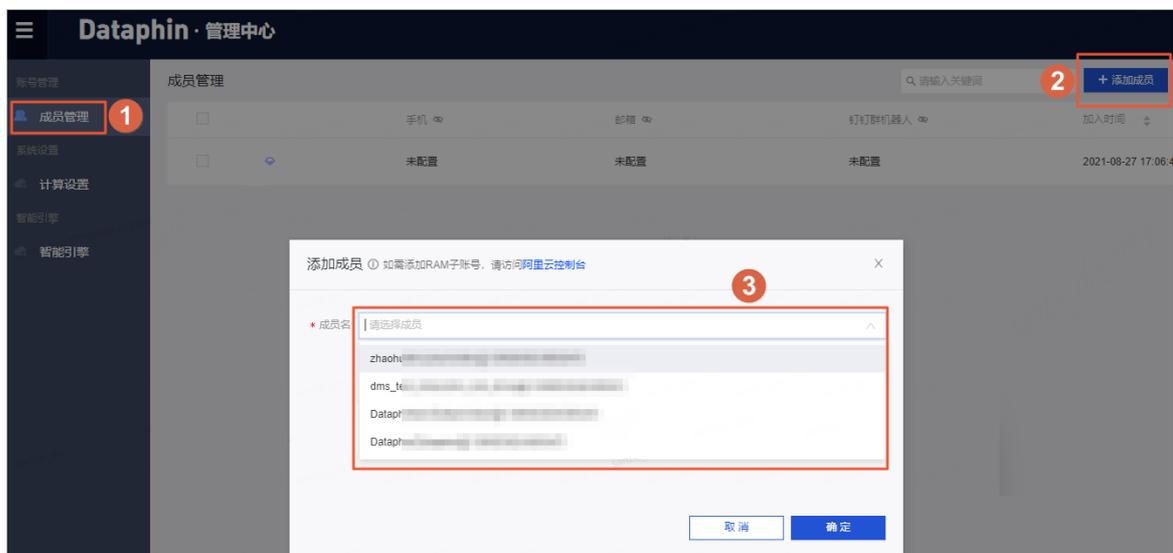


3. 在AccessKey配置对话框，单击修改后，按照操作提示配置需要同步至Dataphin的RAM用户账号所属的阿里云账号信息，包括主账号名称、AccessKey ID和AccessKey Secret。

您可以在[AccessKey](#)页面，获取AccessKey ID和AccessKey Secret；在[账号中心](#)页面，查看阿里云账号名称。

4. 单击校验。
5. 返回校验成功后，单击确定。

完成配置AccessKey后，您可以在下图的添加成员对话框的成员名下拉列表中查看RAM用户账号。



## 后续步骤

同步阿里云账号下的RAM用户至Dataphin后，需要先添加RAM用户为Dataphin成员，再将该RAM用户添加至Dataphin的项目空间进行研发数据。如何添加RAM用户为Dataphin成员，请参见[添加成员](#)。

## 4.2. 设置Dataphin实例的计算引擎

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。本文为您介绍如何设置Dataphin实例的计算引擎。

### 背景信息

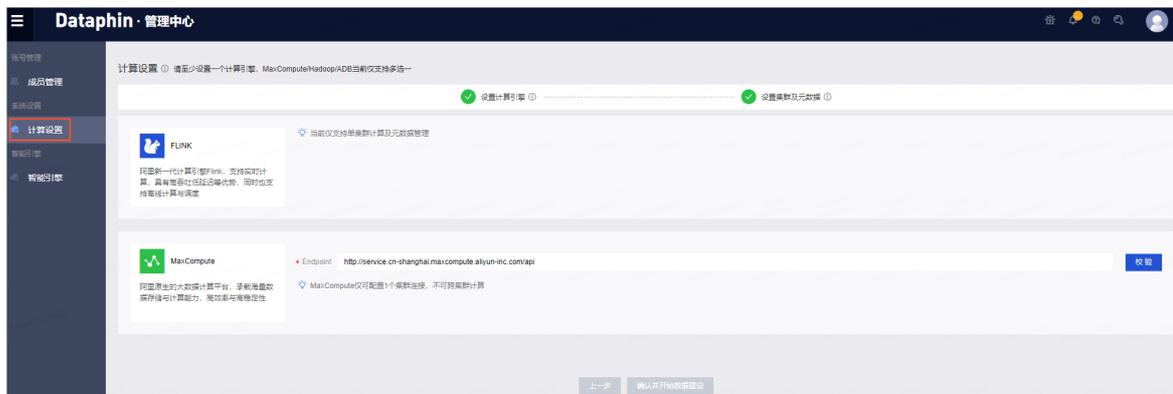
当前，Dataphin计算引擎包括MaxCompute和Flink，默认支持Flink且已内置了Flink的计算集群地址。因此，在设置Dataphin实例计算引擎的过程中，您只需要配置MaxCompute的集群地址。

### 前提条件

在开始执行操作前，请确认您已开通Dataphin。如何开通Dataphin请参见[开通Dataphin](#)。

### 操作步骤

1. 使用阿里云账号登录[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产品首页，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**计算设置**。



5. 在计算设置页面的**MaxCompute**区域，根据Dataphin实例所在的地域，配置对应地域MaxCompute的**Endpoint**。

MaxCompute不同地域的**Endpoint**，请参见下表。

地域	Endpoint
华东2（上海）	http://service.cn-shanghai.maxcompute.aliyun-inc.com/api
华北2（北京）	http://service.cn-beijing.maxcompute.aliyun-inc.com/api
华南1（深圳）	http://service.cn-shenzhen.maxcompute.aliyun-inc.com/api
华东1（杭州）	http://service.cn-hangzhou.maxcompute.aliyun-inc.com/api

6. 单击**校验**。Dataphin会自动校验您所配置的Endpoint所属的地域是否和当前Dataphin实例在同一个地域。
7. 校验成功后，单击**确认并开始数据建设**，即可开始数据建设。

## 后续步骤

完成设置Dataphin实例的计算引擎后，您就可以准备后续数据研发过程中用于读取和写入数据的数据源、用于提供计算资源的计算源。具体操作，请参见[准备数据源和计算源](#)。

## 5. 准备数据源和计算源

在使用Dataphin前，您需要选择符合业务场景需求的数据库或数据仓库作为数据源，用于读取原始数据和写入数据建设过程中的数据；同时您需要准备MaxCompute项目、Flink项目作为Dataphin的计算源，为数据的研发提供计算资源。

### 准备数据源

Dataphin支持对接的数据源包括大数据存储型数据源、文件数据源、消息队列数据源、关系型数据源和NoSQL数据源，各模块支持对接的数据源类型请参见[Dataphin支持的数据源](#)。

确定了业务数据源类型后，您需要创建对应类型的数据源使其与Dataphin间网络连通，以实现Dataphin读取该数据源的原始数据及向该数据源写入Dataphin加工后的数据。如何创建数据源，请参见[概述](#)。

### 准备计算源

在创建项目空间前，您需要完成项目空间计算源的准备。计算源包括离线计算源（MaxCompute）和实时计算源（Flink）。您可以根据业务情况进行准备对应的计算源：

- 如何创建MaxCompute项目，请参见[创建MaxCompute项目](#)。

Dataphin支持为MaxCompute计算引擎的离线数仓添加加速计算源，以满足对查询速度有要求的业务场景。如果需要加速业务数据的查询速度，则您需要完成Hologres的准备。如何开通Hologres，请参见[购买Hologres](#)。

- 如何创建Flink项目，请参见[独享模式开通流程](#)。

 **注意** 当前，Dataphin仅支持Blink独享集群版的Flink作为Dataphin的计算源。

### 后续步骤

完成数据源和计算源准备后，您就可以创建用于研发数据的项目空间。如何创建项目空间请参见[创建项目空间](#)。