

DataWorks（数据工场）

准备工作

准备工作

管理员使用云账号

准备阿里云账号

阿里云数加产品的账号登录体系与阿里云保持一致，统一采用RAM主子账号登录的方式：

阿里云账号（即主账号）是阿里云资源的归属及使用计量计费的基本主体，负责生成本企业组织下的子账号，并对子账号进行管理、授权等操作。

子账号由主账号在RAM系统中创建并进行管理，本身不拥有资源，也没有独立的计量计费，这些账号由所属主账号统一控制和付费。

综上所述，使用阿里云数加前，您需要准备一个阿里云账号并使用RAM管理子账号。

操作步骤

注册阿里云账号

如果您还没有注册过阿里云账号，请进入[阿里云官网](#)，单击**免费注册**，即可进入阿里云账号注册页面创建新的阿里云账号。

注意：

主账号创建成功后，作为阿里云系统识别的资源消费账号，会拥有非常高的权限，请您尽可能保证账号和密码的安全，切勿借给他人使用或定期更新密码。

阿里云账号实名认证

阿里云账号需要进行实名制认证后，才能购买和使用阿里云上的各种产品。如果您还没有实名认证，请进入实

实名认证页面对账号进行实名认证。为保证后续操作顺利进行，请务必完成实名认证操作。

对于企业级用户，最好可以进行企业级认证，以获取更多的便利。更多详情请参见会员账号&实名制认证。

创建运行密钥accesskeys

为了保证您在DataWorks（数据工场，原大数据开发套件）中的任务顺利运行，需要为您创建一个运行密钥。该密钥区别于您登录时填写的账号和密码，主要用于在阿里云各产品间互相认证使用权限。运行密钥AK包括AccessKey ID和AccessKey Secret两部分。具体操作如下：

登录阿里云官网，在右上角的用户名下单击accesskeys进入AccessKey管理页面。



单击右上角的创建AccessKey，单击弹出框中的同意并创建，即可成功创建。如下图所示：



成功创建AccessKey后，便自动跳转至AccessKey管理页面，您可查看相应AK的状态并对其进行禁用、删除等操作。如下图所示：



Access Key管理 (2)				
① Access Key ID和Access Key Secret是您访问阿里云API的密钥，具有该账户完全的权利，请您妥善保管。				
Access Key ID	Access Key Secret	状态	创建时间	操作
LTASB0C256569TKXW0	显示	启用	2017-07-14 10:32:26	禁用 删除
LTASB0C256569TKXW0	显示	启用	2017-02-16 17:14:49	禁用 删除

accesskeys一旦被禁用，使用该accesskeys的服务将运行失败并报错，因此如有变更，需要及时关注使用该accesskeys的产品和服务。

注意：

账号的AccessKeykeys非常重要，创建成功后，请您尽可能保证AccessKey ID和AccessKey Secret的安全，切勿让他人知晓，一旦有泄漏的风险，请及时禁用和更新。

准备RAM子账号

如果您只打算本人使用数加产品，请根据 [准备阿里云账号](#) 的操作准备好您的账号，并 [跳过此章节](#) 进行后续操作。

如果您计划邀请其他用户协作使用数加产品，请您通过 [访问控制](#) 创建和管理子账号。

操作步骤

创建子账号

主账号导航至 [访问控制](#) > [用户管理](#) 页面。如下图所示：



单击右上角的 **新建用户**，如下图所示：



填写弹出框的各配置项，如下图所示：



The 'Create User' dialog box contains the following fields and options:

- * 用户名:** Input field with value 'RAM_demo_1'. Below it, a note states: '长度1-64个字符, 允许输入大小写英文字母、数字、"."、"_"或"'.
- 显示名:** Input field with value 'RAM_demo'. Below it, a note states: '长度1-12个字符或汉字, 允许输入英文字母、数字、"@",".","_"或"'.
- 邮箱:** Input field with value 'RAM_demo@demo.com'.
- 国家/地区:** Dropdown menu with '中国大陆(+86)' selected.
- 电话:** Input field with value '01012345678'.
- 备注:** Input field with value 'RAM_1'.
- ☒ 为该用户自动生成AccessKey

At the bottom right, there are two buttons: '确定' (Confirm) and '取消' (Cancel).

单击 **确定**，即可创建子账号。

允许子账号登录

成功创建子账号后，主账号进入 **访问控制 > 用户管理** 页面，单击对应子账号后的 **管理**，进入用户详情页面。

单击 **启用控制台登录** 并重置密码，填写该子账号用于登录的密码。如下图所示：



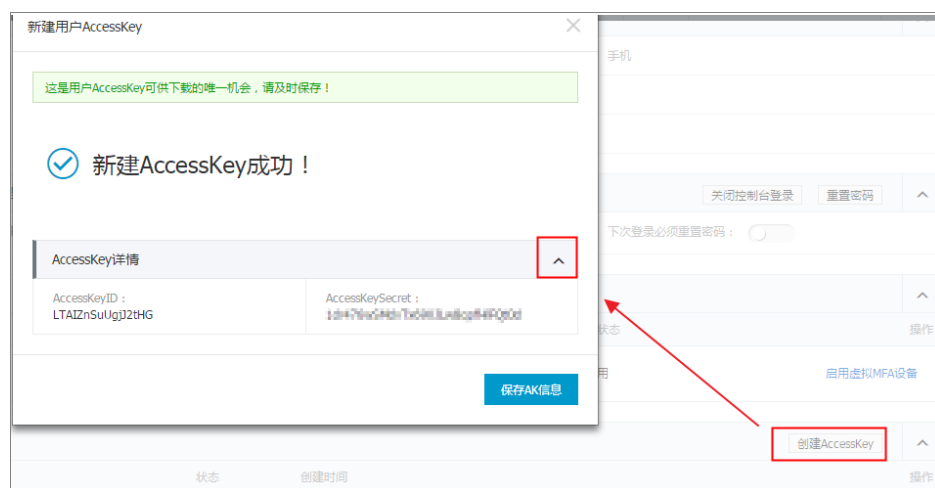
创建子账号的运行密钥

运行密钥对您在 DataWorks（数据工场，原大数据开发套件）中创建的任务的顺利运行非常重要，因此主账号需要为子账号创建 AK，或者允许子账号用户自行创建和管理自己的 AK 信息。具体操作如下：

为子账号创建 AK

主账号进入 访问控制 > 用户管理 页面，单击对应子账号后的 **管理**，进入用户详情页面。

单击 **创建 Accesskey**，填写手机收到的验证码，即可为子账号创建新的 AK。



注意：

子账号的 Access Key Secret 只会出现一次，请在创建的时候妥善保管。如有遗失，只能重新创建。

允许子账号自行创建 AK 并进行管理

主账号进入 访问控制 > 设置 页面，单击 **子用户安全设置**。

勾选 **子用户安全设置** 页面中的 **允许自主管理 AccessKey**，单击 **保存修改** 即可生效。如下图所示：



该选项默认不勾选，故使用子账号的用户默认无法为自己创建 AK 并进行管理。启用 **允许自主管理 AccessKey** 后，子账号登录阿里云官网后即可在管理控制台自行创建 AK。

注意：

运行密钥 AK 非常重要，无论是主账号的 AK 还是子账号的 AK，一旦创建成功，请您务必保证它的安全和使用范围。一旦有泄漏的风险，请及时禁用和更新。

将子账号交付其他用户使用

注意：

子账号归属于主账号，本身不拥有资源，也没有独立的计量计费。子账号在阿里云各产品中操作产生的费用，将由创建这些子账号的主账号统一付费。因此主账号需要查看 RAM 用户登录链接和自己的企业别名，并告知子账号。

主账号进入 **访问控制 > 设置** 页面，单击 **企业别名设置**，即可看到自己的企业别名，并可编辑进行修改。如下图所示：



交付子账号给其他用户使用时，需要提供如下信息：

RAM 用户登录链接。

所属主账号的企业别名。

该子账号的用户名和密码。

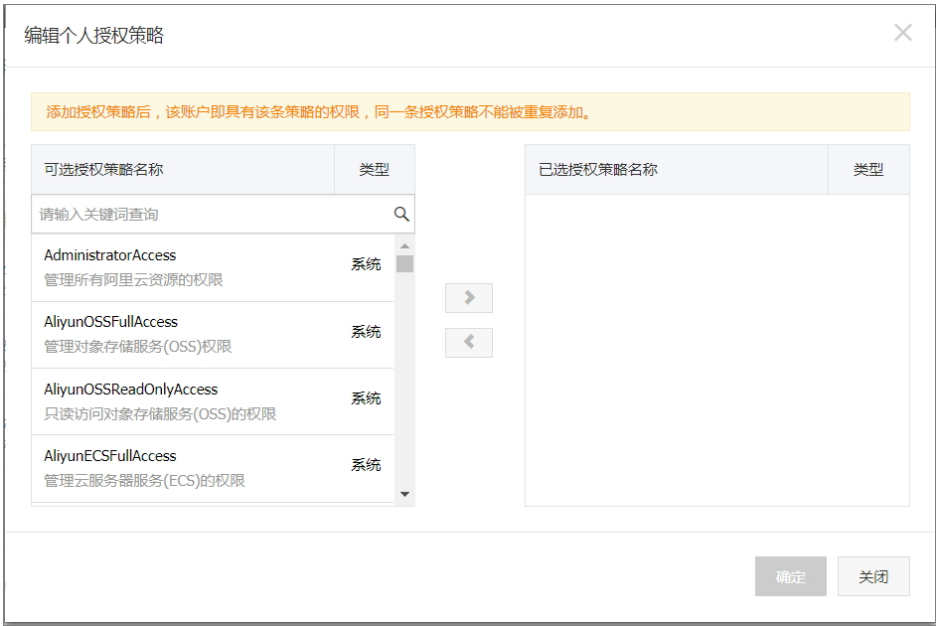
该子账号的 Access Key ID 和 Access Key Secret。

确认已经允许子账号 **启用控制台登录**。

确认已经 **允许子账号自主管理 accesskeys**。

注意：创建完了子账号，接下来您可能更关心的是子账号授权问题，一般是指两个方面：

1、是否需要访问控制RAM授权，这个是不需要的进行相应的授权。如下面的界面：



2、Dataworks是根据项目授权，可将子账号加入项目后再授予子账号角色权限，详细操作可以参考文档：[添加项目成员和角色](#)

后续步骤

现在，您已经学习了如何准备 RAM 子账号，您可以继续学习下一个教程。在该教程中您将学习如何为子账号配置项目成员和角色。详情请参见 [创建项目并添加成员](#)。

新手引导

如果您是第一次使用数加产品，请使用阿里云账号（即主账号）进行登录，并根据准备阿里云账号中的步骤检查账号是否已可用。如果验证无误，根据下文操作创建项目。

操作步骤

实名认证。

阿里云账号需要进行实名制认证才能购买和使用阿里云上的各种产品。如果您还没有实名认证，请进入[实名认证](#)页面对账号进行实名认证。**为保证后续步骤顺利进行，请务必完成实名认证操作。**

创建AccessKey。

为保证您在DataWorks中的任务顺利运行，需要为您创建一个运行密钥。该密钥区别于您登录时填

写的账号和密码，主要用于在阿里云各产品间互相认证使用权限。关于创建AccessKey的详情请参见准备阿里云账号。



运行密钥AK包括AccessID和AccessKey两部分，一般AccessKey为隐藏状态。Accesskeys一旦被禁用，使用该Accesskey的服务将运行失败并报错，因此如有变更，请及时关注使用该Accesskey的产品和服务。



注意：

账号的Accesskey非常重要，创建成功后，请您尽可能保证AccessID和AccessKey的安全，切勿让他人知晓，一旦有泄漏的风险，请及时禁用和更新。

选择Region及服务。

如果Region没有购买相关的服务，会直接显示该Region下暂无可用服务，数据开发、运维中心、数据管理默认是选中的。



新用户引导

1 实名认证 2 创建AccessKey 3 选择Region及服务 4 创建项目

选择region

华东2 华北1 华北2 华东1 华南1 香港 美西1 亚太东南 1 华北3 欧洲中部 1 亚太东南 2 亚太东南 3 亚太东北 1

选择计算引擎服务

☒ MaxCompute 按量付费 包年包月 去购买

开通后，您可在DataWorks里进行ODPS SQL，ODPS MR任务的开发。

☐ 机器学习PAI 按量付费 去购买

开通后，您可使用机器学习算法、深度学习框架及在线预测服务。使用机器学习PAI，需要使用MaxCompute

选择DataWorks服务

☒ 数据集成

开通后，您可在DataWorks里进行数据集成任务的开发，快捷实现二十多种数据源之间的数据同步。

☒ 数据开发、运维中心、数据管理

您可以进行工作流编排、周期调度任务、查询所有表的信息和权限，相关服务目前处于公测阶段。

下一步

创建项目。

只有开通数据集成的阿里云账号（即主账号）才可创建项目，其他子账号仅作为使用者在项目中操作。因此作为主账号，您需要为其他用户准备好项目环境。

项目名：项目名长度需要在3到27个字符之间。

显示名：显示名不得超过27个字符。

启用调度周期：控制当前项目是否启用调度系统，如果关闭则无法周期性调度任务。

允许在本项目中直接编辑任务和代码：当前项目成员在本项目中新建/编辑代码文件的权限，如果关闭则无法新建/编辑代码文件。

本项目中能下载select结果：控制数据开发中select出的数据结果是否能够下载，如果关闭无法下载select的数据查询结果。

MaxCompute项目名称：创建项目过程中可以创建MaxCompute相同的名称项目。

MaxCompute访问身份：个人账号，系统账号 推荐安全要求高的企业用户使用个人账号。

Quota组切换：Quota用来实现磁盘配额。

单击**创建项目**。

成功创建项目后，在项目列表和预览页面可看到创建项目的记录，预览页面一般是根据您创建的最新时间和使用的最近的时间来显示项目。



创建项目

并根据准备阿里云账号中的步骤检查账号是否已可用，如果验证无误，根据下文操作创建项目。

操作步骤

主账号登录DataWorks管理控制台创建MaxCompute项目，您可通过以下两种方式进行创建：

单击控制台概览页面**常用功能**下的**创建项目**，如下图所示：



导航至控制台**项目列表**页面，单击**创建项目**，如下图所示：

概览 项目列表 调度资源列表					
【通知】数加计划于4月20日 12:00 ~ 12:30对数加所有Gateway从升级jdk1.7至jdk1.8，不影响任务正常执行，如有异常请联系售后同学。					
华东2 华北1 华北2 华东1 华南1 香港 美西1 亚太东南1 华北3 欧洲中部1 亚太东南2 亚太东南3 创建项目 刷新列表					
亚太东北1					
搜索					
项目/显示名	创建时间	管理员	状态	开通服务	操作
eee eee	2018-01-10 15:38:54	354081ac@sohu.c om	正常	 	项目配置 进入工作区 修改服务 更多

填写**创建项目**弹出框中的各配置项，选择Region及服务，如Region没有购买相关的服务，会直接显示**该Region下暂无可用服务**，数据开发、运维中心、数据管理默认是选中的。如下图所示：

创建项目

选择region

华东2 华北1 华北2 华东1 华南1 香港 美西1 亚太东南1 华北3 欧洲中部1 亚太东南2 亚太东南3 亚太东北1

选择计算引擎服务

☐  MaxCompute ☐ 按量付费 ☐ 包年包月 [去购买](#)
开通后，您可在DataWorks里进行ODPS SQL，ODPS MR任务的开发。

☐  机器学习PAI ☐ 按量付费 [去购买](#)
开通后，您可使用机器学习算法、深度学习框架及在线预测服务。使用机器学习PAI，需要使用MaxCompute

选择DataWorks服务

☐  数据集成
开通后，您可在DataWorks里进行数据集成任务的开发，快捷实现二十多种数据源之间的数据同步。

☒  数据开发、运维中心、数据管理
您可以进行工作流编排、周期调度任务、查询所有表的信息和权限，相关服务目前处于公测阶段。

取消

下一步

选择计算引擎服务：

MaxCompute：大数据计算服务（MaxCompute，原名ODPS）是一种快速、完全托管的TB/PB级数据仓库解决方案，能够更快速为您解决海量数据计算问题，有效降低企业成本，并保障数据安全。详情请参见MaxCompute文档。

注意：完成创建Dataworks项目后，需要关联MaxCompute项目，否则现执行命令可能会出现报错“project not found”。

机器学习PAI：机器学习指的是机器通过统计学算法，对大量的历史数据进行学习从而生成经验模型，利用经验模型指导业务。详情请参见机器学习PAI。

选择DataWorks服务：

数据集成：数据集成是阿里集团对外提供的稳定高效、弹性伸缩的数据同步平台。致力于提供复杂网络环境下、丰富的异构数据源之间数据高速稳定的数据移动及同步能力。详情请参见数据集成概述。

数据开发：该页面是您根据业务需求，设计数据计算流程，并实现为多个相互依赖的任务，供调度系统自动执行的主要操作页面。详情请参见数据开发概述。

运维中心：该页面可对任务和实例进行展示和操作，您可以在这里查看您所有任务的实例。详情请参见运维中心概述。

数据管理：数据管理模块中可以看到组织内全局数据视图、分权管理、元数据信息详情、数据生命周期管理、数据表/资源/函数权限管理审批等操作。详情请参见数据管理概述。

配置新建项目的基本信息和高级设置。

创建项目

基本信息

* 项目名称：

haha01

显示名：

haha01

项目描述：

高级设置

* 启动调度周期：

开

?

* 本项目中能下载select结果：

开

?

面向MaxCompute

MaxCompute项目名称：

haha01

?

* MaxCompute访问身份：

☒ 个人账号

☐ 系统账号

?

* Quota组切换：

默认后付费资源组

▼

上一步

创建项目

基本配置：

项目名：项目名长度需要在3到27个字符之间。

显示名：显示名不能超过27个字符。

高级配置：

启用调度周期：控制当前项目是否启用调度系统，如果关闭则无法周期性调度任务。

允许在本项目中直接编辑任务和代码：当前项目成员在本项目中新建/编辑代码文件的权限，如果关闭则无法新建/编辑代码文件。

本项目中能下载select结果：控制数据开发中select出的数据结果是否能够下载，如果关闭无法下载select的数据查询结果。

MaxCompute项目名称：创建项目过程中可以创建MaxCompute相同的名称项目。

MaxCompute访问身份：个人账号，系统账号。推荐安全要求高的企业用户使用个人账号。

Quota组切换：Quota用来实现磁盘配额。

项目创建成功后，项目列表即会显示相应的内容，如下图所示：

概览项目列表调度资源列表

【通知】数加计划于4月20日 12:00 ~ 12:30对数加所有Gateway从升级dk1.7至dk1.8，不影响任务正常执行，如有异常请联系售后同学。

华东2华北1华北2华东1华南1香港美西1亚太东南1华北3欧洲中部1亚太东南2亚太东南3

创建项目刷新列表

亚太东北1

搜索

项目/显示名	创建时间	管理员	状态	开通服务	操作
haha01 haha01	2018-01-10 16:24:19	354989bac@scdhu.c	正常	 	项目配置 进入工作区 修改服务 更多

注意：

您一旦成为项目空间Owner，意味着该项目内**所有东西**都是您的，在给别人赋权之前，任何人无权限访问您的空间。

对于一般用户，并非一定要创建项目空间，只要被加入到某个项目便可使用MaxCompute。

后续步骤

现在，您已经学习了如何创建项目，您可以继续学习下一个教程。在该教程中您将学习如何快速完成一个完整的数据开发和运维操作。详情请参见快速入门。

添加项目成员和角色

如果您只打算本人使用数加产品，请跳过本文操作。

如果您计划邀请其他用户协作使用数加产品，请先根据准备RAM子账号的操作准备必要的子账号，并创建项目，然后根据下文进行配置。

阿里云数加产品的账号登录体系与阿里云保持一致，统一采用RAM主子账号登录的方式。

子账号计费

阿里云账号（即主账号）是阿里云资源的归属及使用计量计费的基本主体，负责生成本企业组织下的子账号，并对子账号进行管理、授权等操作。

子账号由主账号在RAM系统中创建并管理，本身不拥有资源，也没有独立的计量计费，这些账号由所属主账号统一控制和付费。

操作步骤

进入DataWorks（数据工场，原大数据开发套件）的项目列表页面，单击相应项目列表后的**项目配置**，如下图所示：



单击右侧弹窗中的**更多设置**。



注意：

如果您需要添加新的子账号，请单击上图中的**访问控制**，打开阿里云访问控制页面。添加子账号并交付使用的具体步骤请参见**准备阿里云账号**。

勾选**添加成员**对话框中的子账号，并选择批量授予这些子账号的角色，单击**确定**即可添加成功。



项目中已有的成员和对应角色可在列表中查看并酌情修改，也可将子账号从项目中移除。



项目成员可被分配的角色共有5种：项目管理员、部署、开发、访客和运维。项目创建者默认为项目管理员角色。具体角色的权限说明请参见项目管理手册>项目管理>项目成员管理。

角色说明：

开发：负责数据开发页面设计和维护 workflow。

运维：负责在运维中心页面管理全部任务的运行情况并做相应处理。

项目管理员：除了拥有开发角色和运维角色全部权限外，还可以进行添加 / 移出项目成员并授予角色，创建自定义资源组等项目级别的操作。

部署：仅在多项目模式时审核任务代码并决定是否提交运维。

访客：仅有只读权限，可查看数据开发页面的 workflow 设计和代码内容。

温馨提示：已经将子账号加入到项目中，可以使用子账号登录 DataWorks，登录完毕后需要在数加控制台更新子账号的个人 AK 信息，详细请参考子账号登录及更新信息

更新账号信息

如果您是第一次使用 DataWorks（数据工场，原大数据开发套件），请根据准备阿里云账号和创建项目的内容逐步配置，然后根据下文的操作更新账号信息。

主账号的AK信息非常重要，主要影响项目中的成员管理，有时也会影响任务运行等操作，因此请主账号用户务必保证在DataWorks中的AK信息是RAM管理页面中有效的AK值。

操作步骤

登录数加控制台概览页面，单击**个人信息**，如下图所示：



进入个人信息页面后，即可对AccessKey信息进行查看和修改，如下图所示：



定期确认上述AK是管理控制台中显示的已启用的有效AccessKey，如下图所示：



Access Key管理 (2) 刷新 创建Access Key

① Access Key ID和Access Key Secret是您访问阿里云API的密钥，具有该账户完全的权限，请您妥善保管。

Access Key ID	Access Key Secret	状态	创建时间	操作
LTMEKJ3E9LQMTW6Q4	显示	启用	2017-07-14 10:32:26	禁用 删除

用户使用子账号

如果您是第一次使用子账号登录数加平台和使用DataWorks，您需要获知以下内容：

- 该子账号所属主账号的企业别名。

该子账号的用户名和密码。

该子账号的AccessKey ID和AccessKey Secret。

确认主账号已允许子账号**启用控制台登录**。

确认主账号已允许子账号**自主管理accesskeys**。

如果您不确定上述信息，请向所属主账号确认如下两件事：

确认主账号用户已根据准备阿里云账号和准备RAM子账号准备好子账号所需的全部配置和信息。

确认主账号用户已参见创建项目和添加项目成员和角色成功创建项目并为您的子账号配置合适的成员角色。

登录DataWorks

根据主账号提供的**RAM用户登录链接**登录到控制台页面。

导航至**DataWorks>项目列表**页面，单击对应项目中的**进入工作区**，即可进入数据开发页面。

概览		项目列表				
华东2	输入项目名称进行搜索	搜索				
项目名/显示名	创建日期	项目管理员	状态	计费类型	操作	
mydemo mydemo	2017-05-17 16:39:04	ali_demo@yun-inc.com	正常	I/O后付费	项目配置	进入工作区

更新个人信息

由于AK信息属于个人密钥，需要您手动输入AK后，DataWorks才能正常使用。

进入个人信息页面。

您可通过以下两种方式进入个人信息页面。

如果您是第一次进入DataWorks的工作区，将看到如下页面提醒您填写AK。您可直接单击页面红框中的链接，前往个人信息页面更新AccessKey信息。



如果您不是第一次进入DataWorks的工作区，可进入数加控制台概览页面，单击个人信息，如下图所示：



单击**修改AccessKey**信息，填写您从主账号获得的AK信息。如下图所示：



再次进入工作区页面或刷新已经打开的工作区页面，即可使用DataWorks开始数据开发工作。