Alibaba Cloud
DataWorks

Preparation

Legal disclaimer

Alibaba Cloud reminds you to carefully read and fully understand the terms and conditions of this legal disclaimer before you read or use this document. If you have read or used this document, it shall be deemed as your total acceptance of this legal disclaimer.

1. You shall download and obtain this document from the Alibaba Cloud website or other Alibaba Cloud-authorized channels, and use this document for your own legal business activities only. The content of this document is considered confidential information of Alibaba Cloud. You shall strictly abide by the confidentiality obligations. No part of this document shall be disclosed or provided to any third party for use without the prior written consent of Alibaba Cloud.

2. No part of this document shall be excerpted, translated, reproduced, transmitted, or disseminated by any organization, company, or individual in any form or by any means without the prior written consent of Alibaba Cloud.

3. The content of this document may be changed due to product version upgrades , adjustments, or other reasons. Alibaba Cloud reserves the right to modify the content of this document without notice and the updated versions of this document will be occasionally released through Alibaba Cloud-authorized channels. You shall pay attention to the version changes of this document as they occur and download and obtain the most up-to-date version of this document from Alibaba Cloud-authorized channels.

4. This document serves only as a reference guide for your use of Alibaba Cloud products and services. Alibaba Cloud provides the document in the context that Alibaba Cloud products and services are provided on an "as is", "with all faults " and "as available" basis. Alibaba Cloud makes every effort to provide relevant operational guidance based on existing technologies. However, Alibaba Cloud hereby makes a clear statement that it in no way guarantees the accuracy, integrity , applicability, and reliability of the content of this document, either explicitly or implicitly. Alibaba Cloud shall not bear any liability for any errors or financial losses incurred by any organizations, companies, or individuals arising from their download, use, or trust in this document. Alibaba Cloud shall not, under any circumstances, bear responsibility for any indirect, consequential, exemplary, incidental, special, or punitive damages, including lost profits arising from the use

or trust in this document, even if Alibaba Cloud has been notified of the possibility of such a loss.

5. By law, all the content of the Alibaba Cloud website, including but not limited to works, products, images, archives, information, materials, website architecture, website graphic layout, and webpage design, are intellectual property of Alibaba Cloud and/or its affiliates. This intellectual property includes, but is not limited to, trademark rights, patent rights, copyrights, and trade secrets. No part of the Alibaba Cloud website, product programs, or content shall be used, modified , reproduced, publicly transmitted, changed, disseminated, distributed, or published without the prior written consent of Alibaba Cloud and/or its affiliates . The names owned by Alibaba Cloud shall not be used, published, or reproduced for marketing, advertising, promotion, or other purposes without the prior written consent of Alibaba Cloud. The names owned by Alibaba Cloud include, but are not limited to, "Alibaba Cloud", "Aliyun", "HiChina", and other brands of Alibaba Cloud and/or its affiliates, which appear separately or in combination, as well as the auxiliary signs and patterns of the preceding brands, or anything similar to the company names, trade names, trademarks, product or service names, domain names, patterns, logos, marks, signs, or special descriptions that third parties identify as Alibaba Cloud and/or its affiliates).

6. Please contact Alibaba Cloud directly if you discover any errors in this document.

Generic conventions

Table -1: Style conventions

Style	Description	Example
⛔	This warning information indicates a situation that will cause major system changes, faults, physical injuries, and other adverse results.	⛔ Danger: Resetting will result in the loss of user configuration data.
⚠️	This warning information indicates a situation that may cause major system changes, faults, physical injuries, and other adverse results.	⚠️ Warning: Restarting will cause business interruption. About 10 minutes are required to restore business.
📋	This indicates warning information, supplementary instructions, and other content that the user must understand.	① Notice: Take the necessary precautions to save exported data containing sensitive information.
	This indicates supplemental instructions, best practices, tips, and other content that is good to know for the user.	📋 Note: You can use Ctrl + A to select all files.
>	Multi-level menu cascade.	Settings > Network > Set network type
Bold	It is used for buttons, menus, page names, and other UI elements.	Click **OK**.
`Courier font`	It is used for commands.	Run the `cd / d C :/ windows` command to enter the Windows system folder.
Italics	It is used for parameters and variables.	`bae log list -- instanceid` *`Instance_ID`*
[] or [a|b]	It indicates that it is a optional value, and only one item can be selected.	`ipconfig` `[-all|-t]`

Style	Description	Example
{} or {a|b}	**It indicates that it is a required value, and only one item can be selected.**	`swich` *{stand | slave}*

Contents

1 RAM User Operations

If this is your first time using DataWorks and logging on to the DTplus platform with a RAM user account, you must have the following information:

· The enterprise alias of the RAM user primary account .

· The user name and password of the RAM user.

· The AccessKey ID and AccessKey Secret of the RAM user.

· Confirm the primary account has granted the RAM user Enable Console Logon permission.

· Confirm the primary account has granted the RAM user Manage AccessKeys permission.

If you do not have the preceding information, verify the following information from the primary account:

· The primary account user has prepared all configurations and information required by the RAM user according to #unique_4 and #unique_5.

· The primary account user has created a project and configured the appropriate member roles for your RAM user account, by following the instructions mentioned in #unique_6 .

Log on to DataWorks

1. Log on to the DTplus console by using the RAM user logon link provided by the primary account.

2. Go to the DataWorks > Workspace List page, and click Enter Project in the corresponding project to enter the data development page.
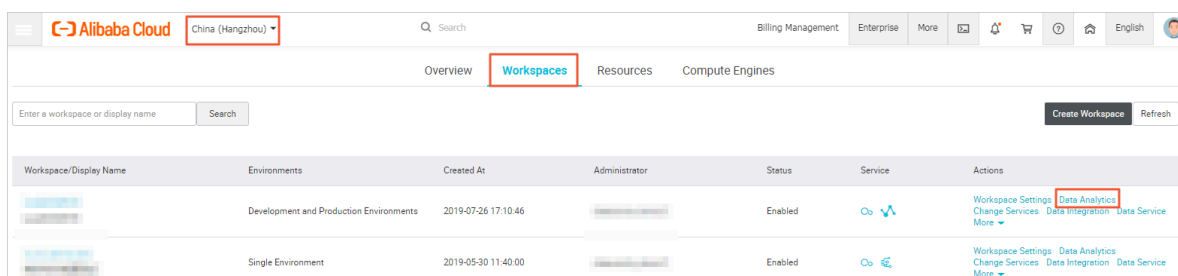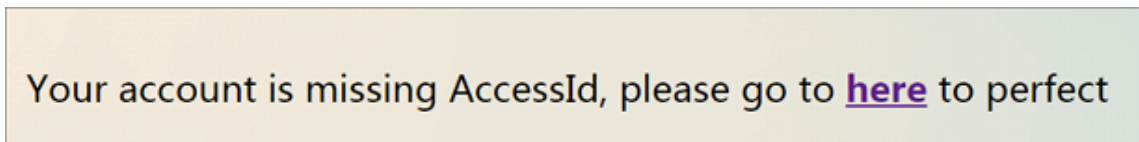

Update personal information

Before you start DataWorks,you must enter the AccessKey manually.

1. **Enter the personal information page.**

   You can use any of the following methods to enter the personal information page.

   · When you log on to DataWorks dashboard for the first time, you will be prompted to enter the AccessKey. Click the link in the red box and enter the personal information page to update your AccessKey information.

   · If you have used DataWorks dashboard before, you can enter the Overview page and click Personal Information.
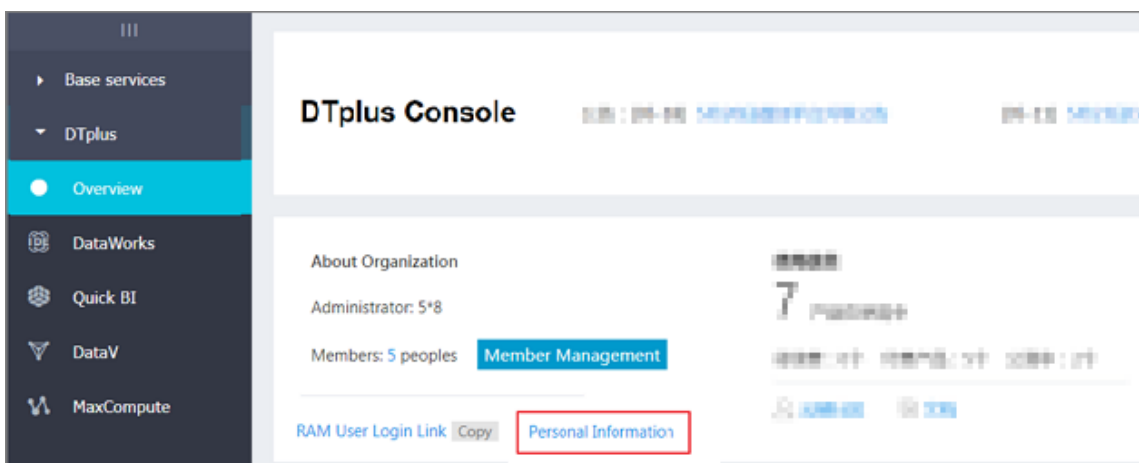
2. **Click Modify AccessKey Information to enter AccessKey information obtained from the primary account.**
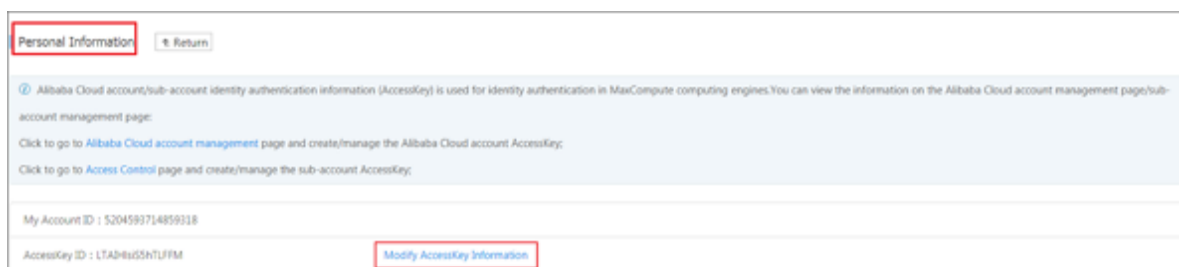
3. **Enter the DTplus console again or refresh the page. Now, you can start data development with DataWorks.**

2 Administrator Operations

2.1 Alibaba Cloud Account Preparations

Alibaba Cloud DTplus products uses RAM primary account and RAM user logon system.

- · Alibaba Cloud account (or the primary account) has ownership of Alibaba Cloud resources and is the main account for metering and billing resource usage. The primary account allows you to create Resource Access Management (RAM) users for your enterprise to manage and authorize these RAM user accounts.
- · Created and managed by the RAM primary account, the RAM user has no resources , and cannot perform metering and billing. All RAM users are controlled and paid by their primary account.

Therefore, you must have an Alibaba Cloud account to manage RAM users through RAM before using DTplus products.

Register an Alibaba Cloud account

If you have not registered an Alibaba Cloud account yet, go to the Alibaba Cloud Official Website and click Free Account. This takes you to the Alibaba Cloud account registration page, where you can create a new Alibaba Cloud account.

> **Note:**
> The Alibaba Cloud system recognizes the primary account you created is a resource-consuming account with high-level permission. Keep your account name and password safe. Change your password periodically and do not lend your account to others.

Perform real-name registration on the new Alibaba Cloud account

You must perform real-name registration on your Alibaba Cloud account before you purchase and use Alibaba Cloud products. If you have not completed real-name registration yet, go to Real-name Registration to go through the process. To save you any trouble in the future, be sure to complete the real-name registration.

If you are an enterprise user, we strongly recommend you complete the enterprise-level authentication for greater convenience.
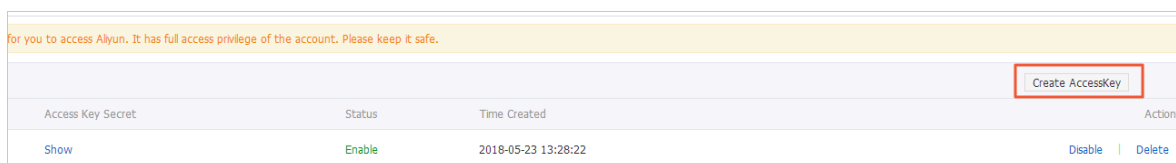
Create an AccessKey

To guarantee smooth processing of tasks in DataWorks, you must create an AccessKey
. Unlike the account name and password you enter to log on, an AccessKey is
primarily used for access permission verification between various Alibaba Cloud
products. An AccessKey comprises an AccessKey ID and an AccessKey Secret.

1. Log on to the Alibaba Cloud official website, and click AccessKeys under the
   Username in the upper-right corner to enter the AccessKey Management page.
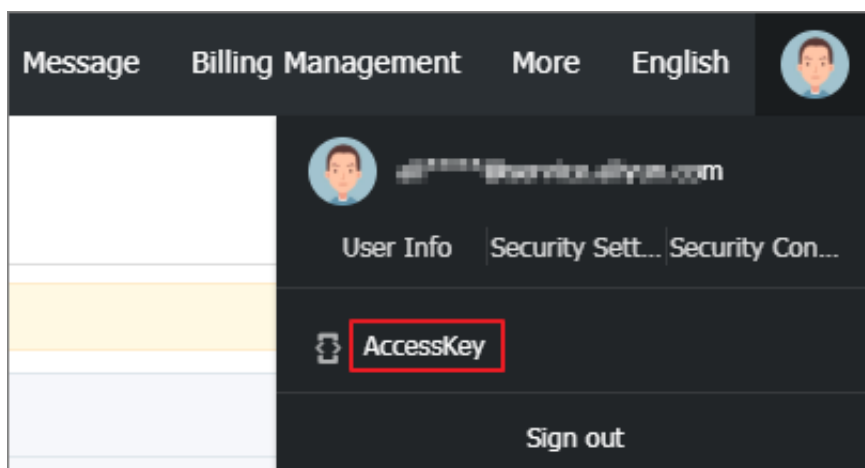
2. Click Create an AccessKey in the upper-right corner, and click Agree and Create in
   the dialog window that appears to create an AccessKey.

3. When an AccessKey is created, the system automatically jumps to the AccessKey
   Management page, where you can view the status of the AccessKey and disable or
   delete it.
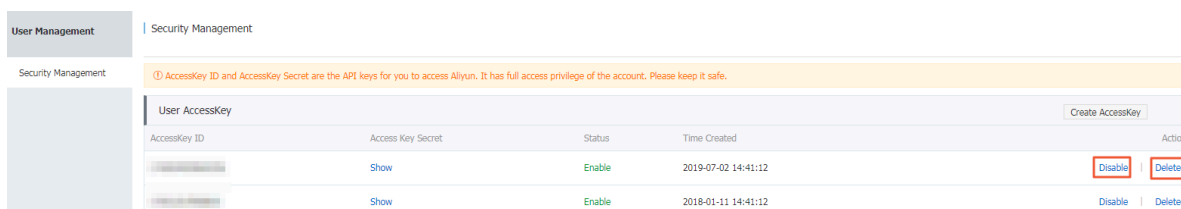
A disabled AccessKey will affect services that use this AccessKey and cause
errors. Therefore, if you change your AccessKeys, you must promptly make

the corresponding changes for the products and services that use the affected AccessKeys.

> 📋 **Note:**
>
> The AccessKey is critical to the account. Keep the AccessKey ID and AccessKey Secret safe. Do not reveal this information to others. If the information may have been leaked, immediately disable and update your AccessKey.

2.2 Create a RAM user

If you are the only user of this product, skip this section. To create your own account, see #unique_4. To invite other users to use this product, you can create and manage RAM user accounts through the Resource Access Management (RAM) console.

Create a sub-account

1. Log on to the Alibaba Cloud DTplus Console with the primary account. Choose RAM > Indentities > Users.
2. Click Create User.
3. Enter the user information.
4. Click OK to complete the user creation.

Enable a RAM user to log on the console

1. After the RAM user is created, logs on to the RAM > Users page with your primary account, and click RAM user account.
2. Click Modify Logon Settings and Console Password Logon, enter the logon password of this RAM account.

Create an AccessKey for a RAM user

The AccessKey facilitates smooth operation of the tasks created in DataWorks. Therefore, a primary account must create an AccessKey for the RAM user, or allow RAM account users to create and manage their own AccessKey information.

· Create an AccessKey for a RAM user

1. Use your primary account to log on to the RAM > Indentities > Users page, and click RAM user account.

2. Click Create AccessKey, and enter the verification code sent to your mobile phone. The AccessKey is created for the RAM user account.

> 📋 Note:
>
> The AccessKeySecret of the RAM user will appear only once. Please record the AccessKeySecret and keep it confidential. If you lose it, create a new one.

· Authorize the RAM users to create and manage their own AccessKeys

1. The main account enters the RAM > Identity Management > Settings page, click update RAM User Security Settings from the Actions column.

2. Select RAM User Security Settings tab, and select Manage AccessKey > allowed, click OK to complete settings.

   If this option is not selected, by default the RAM user cannot create nor manage AccessKey. After allow the user to manage their own AccessKey, RAM user account logon to Alibaba cloud, you can create the AccessKey in the console.

> 📋 Note:
>
> It is important to run the private AccessKey. Keep the created AccessKey of the primary account or the RAM account secure and ensure its scope of use. If the AccessKey is leaked, timely disable it and create a new one.

Authorize a RAM account

If you want to create projects with your RAM account, you may need AliyunDataWorksFullAccess policy permission.
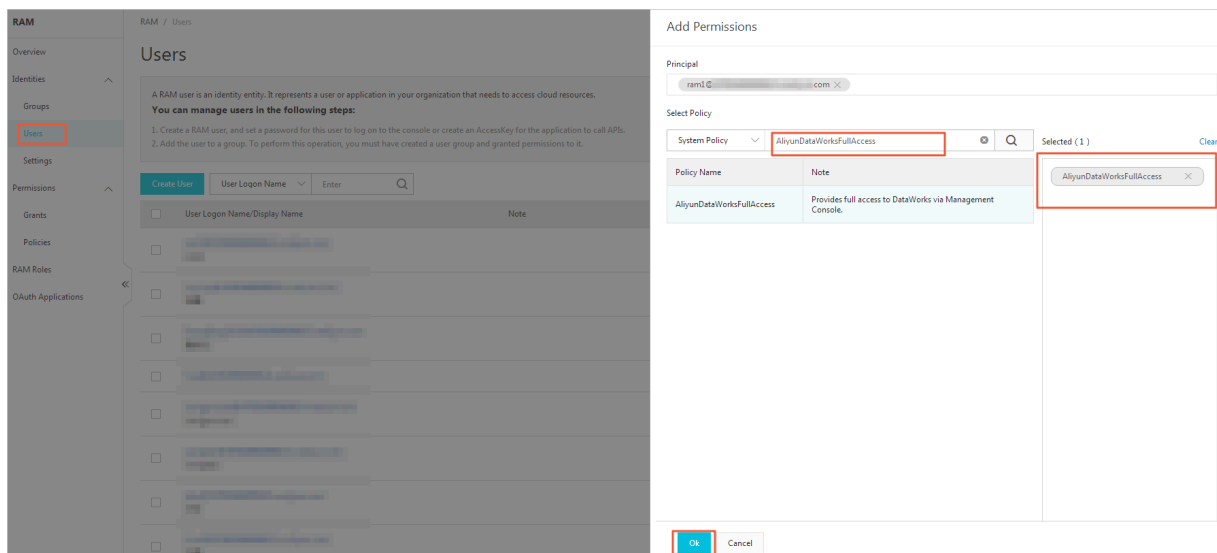
Delivery RAM account to other users

> ### Note:
>
> RAM users belong to a primary account. Thus, RAM users do not have resources and billing systems of their own. The operation cost of a DTplus RAM user is borne by the primary account. Fees incurred by operations of RAM users in Alibaba Cloud products are uniformly paid through the primary account to which these RAM users belong. Therefore, the primary account views the RAM user log on link and its own enterprise alias (default domain), and provide this information to a RAM user.

Log on to RAM > Overview, and click Update default domain to change the enterprise alias (default domain).

When a RAM account is delivered to its users, you must provide the following details:

· Logon link of RAM users.

· Enterprise alias (default domain) of primary account.

· User name and password of RAM users.

· AccessKey ID and AccessKeySecret of a RAM user.

· Confirm the primary account authorized your RAM user Enable Console Logon permission.

· Confirm the primary account authorized your RAM user Manage AccessKeys permission.

2.3 Beginner's guide

If you are using DTplus products for the first time, log on to DTplus with your Alibaba Cloud account (primary account), and check whether the account is available by following the steps listed in Prepare Alibaba Cloud account topic. If the account is available, follow these steps to create a project.

Procedure

1. Create an AccessKey.

   To assure smooth processing of data tasks in DataWorks, you must create an AccessKey. Similar to logging on with an account name and password, an AccessKey is primarily used to access permission verification between various Alibaba Cloud products and services. For more information, see #unique_4.

   An AccessKey comprises of an AccessKey ID and an AccessKeySecret. Generally, the AccessKey is secured and kept confidential. Once the AccessKey is disabled, services that use this AccessKey fails to run and reports exceptions to users. If an AccessKey information changes, it can affect products and services that use this AccessKey. Therefore, you must keep a close eye on such changes and take appropriate actions timely.

   > **Note:**
   > The AccessKey is critical to your account. After creating the AccessKey, you must secure the AccessKey ID and AccessKeySecret. Do not reveal the AccessKey information to others under any conditions. If you suspect an information leak or other risks, you must promptly disable or modify your AccessKey.

2. Select region and services.

   If a region does not contain any projects, it automatically shows you have no projects. You can proceed with create a project. By default, data development, O&M center, and data management are selected.

3. Create a project.

   Only a primary account that has the data integration function can create a project, all other RAM users are treated only as users in the project.

   > **Note:**

> This makes it necessary for a primary account to prepare a project environment for other RAM users.

- Workspace name: The project name must be 3 to 27 characters in length.
- Display name: The display name must not exceed 27 characters in length.
- Enable scheduling frequency: Controls scheduling system frequency in the current project. If it is disabled, it cannot schedule tasks periodically.
- Allows editing tasks and code directly in this project: the current project member has permission to create or edit code files in this project. If this function is disabled, the project member cannot create or edit code files.
- Enable select result downloads in this project: Determines whether data results from the selected statements can be downloaded in this project. If it is disabled, it cannot download the data query results from a selected statement.
- MaxCompute project name: You can create a MaxCompute project with the same name as the project name.
- MaxCompute access identity: The personal account and the system account.
- Quota group: Quota is used to implement disk quotas.

4. Click Create Workspace.

   After a workspace(project) is created, the project appears on the project list page . This page usually displays the project according to the time of creation or latest time of use.

2.4 Create Workspace

Check whether the primary account is available according to the #unique_4 account. If the account is available, perform the following steps to create a workspace(project).
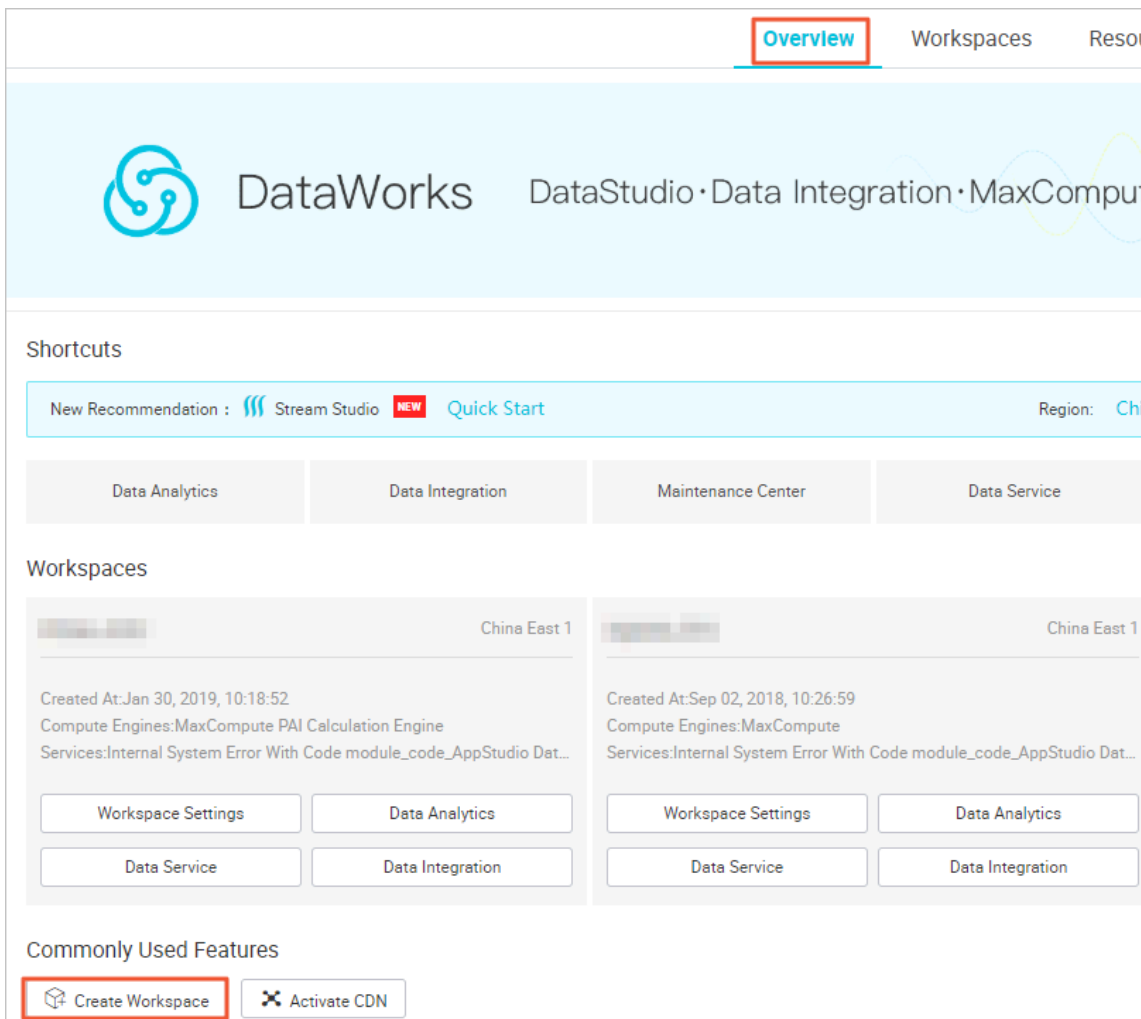
> 📋 Note:
>
> Currently, the RAM user can also create a workspace, following the same procedures as the primary account.
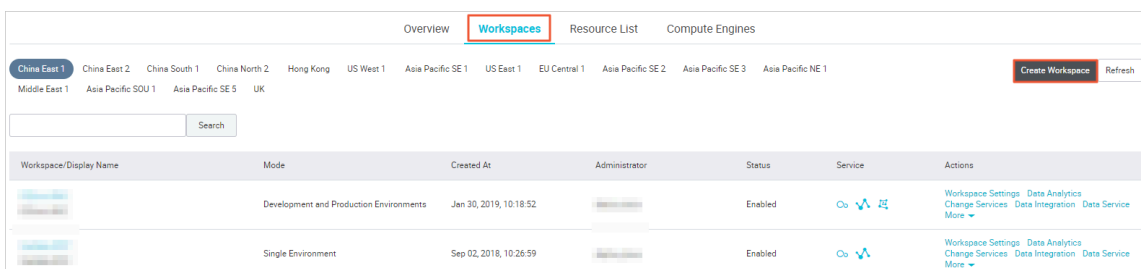
Procedure

1. Log on to the DataWorks console by using a primary account.

2. **You can create the MaxCompute project with the following two methods.**

   · **Click Create Workspace in Common Functions area.**

   · **Go on Workspace List page, click Create Workspace.**

3. Enter the configuration items in the Create Workspace dialog box. Select a region and a calculation engine service.

   If you have not purchased the relevant services, the Region section will directly display it as no service is available . Data development, O&M center, and data management are selected by default..

Choose calculation Engine services:

· MaxCompute: MaxCompute is a big data processing platform developed by Alibaba Cloud. It is mainly used for batch structural data storage and processing

, and can provide massive data warehouse solutions and big data modeling service.

· Machine learning PAI: Machine learning refers to a machine that uses statistical algorithms to learn a large amount of historical data to generate empirical models for business references.

Choose DataWorks service:

· Data integration: A data synchronization platform that provides stable, efficient, and elastically scalable services. Data integration is designed to implement fast and stable data movement and synchronization between multiple heterogeneous data sources in complex network environments. For more information, see #unique_12.

· Data development: Data development helps you design data computing processes based on business requirements and make task dependencies automatically run in the scheduling system. For more information, see Data development overview.

· O&M center: O&M center is a place where tasks and instances are displayed and operated. You can view all your tasks in Task List and perform such operations on the displayed tasks. For more information, see #unique_14.

· Data management: Data management in Alibaba Cloud DTplus platform displays the global data view and metadata details of an organization, and enables operations such as separate permission management, data lifecycle management, and approval and management of data table, resource, or function permissions. For more information, see Data management overview.

4. **Enter project basic information and set advanced settings.**

Basic Information

· Workspace name: The workspace name must be between 3 and 27 characters in length.

· Display name: The display name must be less than 27 characters in length.

· Mode: The project mode is a new feature introduced by the new DataWorks version. It is separated into simple and standard mode. For differences between the two project development patterns, see #unique_16.

    - Simple mode: A DataWorks project corresponding to a MaxCompute project, cannot set the development and production environment. Under this mode,

you can only perform simple data development, and cannot control the data development process and table permissions.

- Standard mode: A DataWorks project corresponding to two MaxCompute projects can set both the development and production environments , improve code development specifications, and strictly control table permissions, prohibits arbitrary operation of the production environment table, and ensure the data security of the production table.

Advanced Settings

· Enable scheduling frequency: Controls the current project scheduling system .If it is disabled, it cannot periodically schedule tasks.

· Enable select result downloads in this project: Whether data results from select statement can be downloaded in this project, and if it is disabled, it cannot download data query results from the select statement.

· For MaxCompute

- Development Environment MaxCompute Project name:The default name is the project name + "_ dev" suffix, and can be modified.

- Development Environment MaxCompute access identity: The default identity is a personal account.

- Production Environment MaxCompute Project name: The default name is the same as DataWorks project.

- Production environment MaxCompute access identity: The default name is the production account. We do not recommend modifying this name.

- Quota group: Quota is used to implement disk quotas.

5. When the configuration is complete, click Create Workspace.

After the workspace is created successfully, you can view content on the Workspace List page.

> 📋 Note:
>
> · Once you become a project owner, you have full ownership over all project content, and anyone that requires project access must apply for permission. first .
>
> · General users do not have to create a workspace(project.) If you are added to a project, you can use MaxCompute.

Next step

> Now that you have learned about project creation, you can move on to the next topic. In the next topic, you will learn how to quickly complete data development operations and data O&M. For more information, see Quick Start.

2.5 Add users and roles

> You can skip this article if DTplus is for your personal use. If DTplus is used for collaboration projects, you must create a Resource Access Management (RAM) user according to #unique_5 and #unique_6, then complete configurations as follows.
>
> Alibaba Cloud DTplus products adopt the RAM primary account or RAM user logon system.

RAM user billing

> Alibaba Cloud account (or the primary account) represents the ownership of Alibaba Cloud resources and is the main account for metering and billing resource usage. The primary account allows you to create RAM users for your enterprise to manage, and authorize permissions.
>
> Created and managed by the primary account in RAM , the RAM user has no resources, and therefore cannot be metered and billed separately. All RAM users are controlled and paid by their primary account.
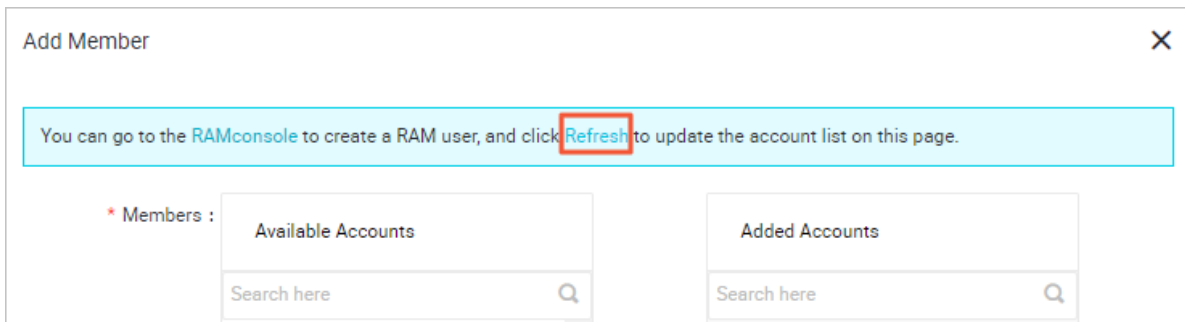
Procedure

1. Go to the Workspace List page of DataWorks, and click Enter Project after the corresponding project list.
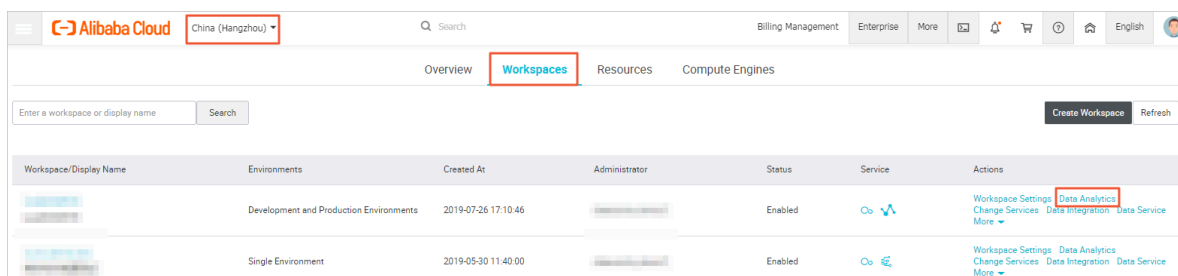
2. Click Project Management to go to the project management page.

3. Click Project Member Management in the left-side navigation pane.

4. **Click Add Members to display the following dialog box.**

   Click Refresh to synchronize the RAM user under the current Alibaba Cloud account in the RAM console to the selected list.


📋 **Note:**

To add a new RAM user, click RAM Console in the preceding figure to open the Alibaba Cloud RAM page. For more information about adding and delivering a RAM user, see #unique_4.

5. Check RAM users in the New Members dialog box, select a role permission for these RAM users in batches, and click OK to add members.
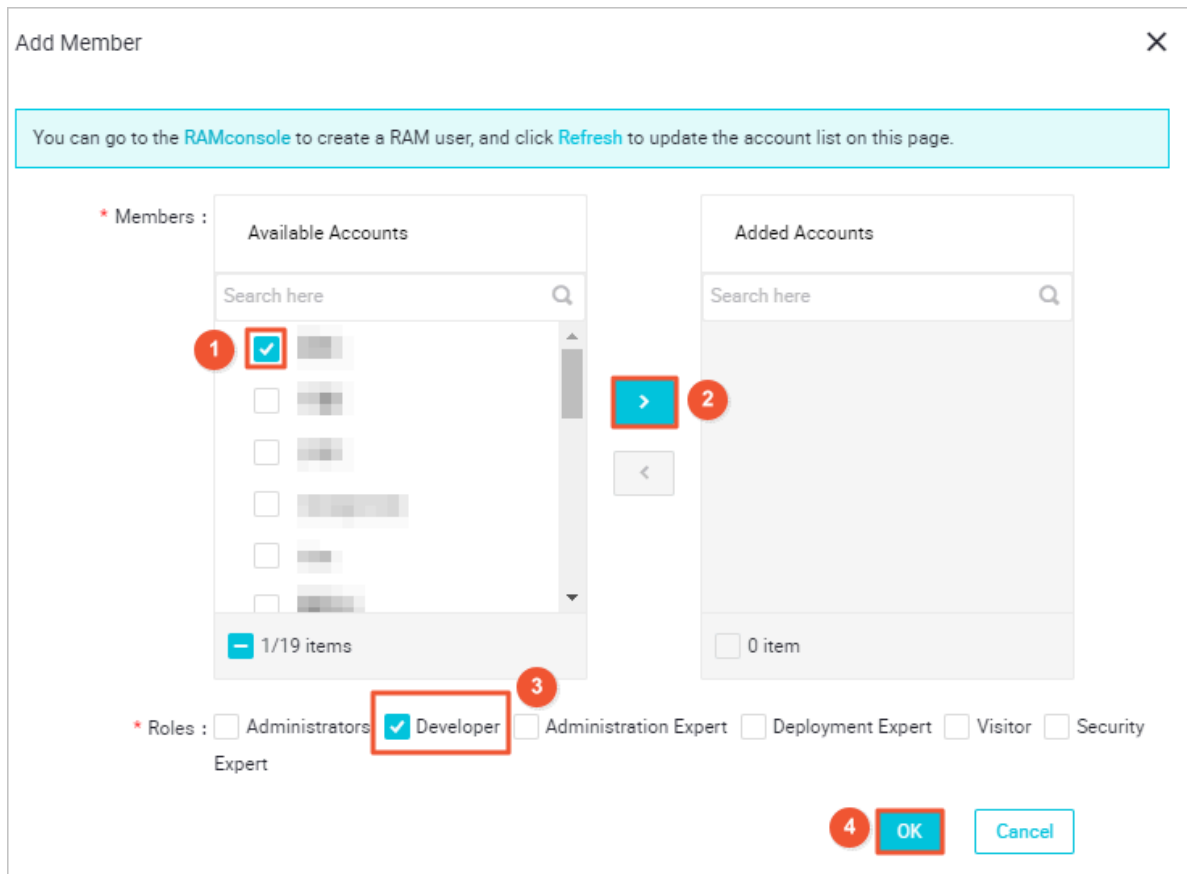
The existing members and their roles in the project can be viewed in the list and modified as required. You can also remove RAM users from the project.
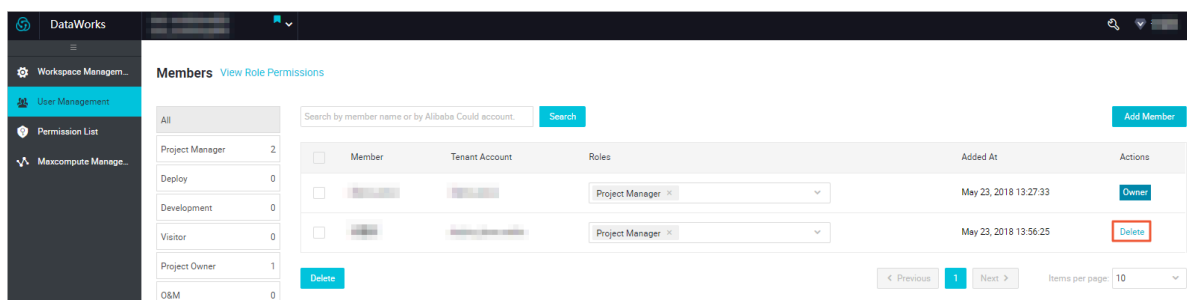
The project members can be assigned five roles namely, project administrator , deployer, developer, guest, and maintainer. The project creator is the project administrator by default.

A brief description about the roles is as follows.

· Development: Manages the design of data development page and workflow maintenance.

- O&M: Manages the running conditions of all tasks in the O&M center page and handles them accordingly.

- Project administrator:

  In addition, to having full permissions as to that of developer and maintainer, the project administrator can perform project level operations, such as adding or removing project members, granting roles, and creating custom resource groups.

- Deploy: Reviews the task code and decides whether to submit it to the maintainer under the multi-project mode.

- Visitor: The visitor can view the workflow design and code content on the data development page with read-only permissions.

- Safety Manager: Only has s Data Security Guard permissions.

> 📋 **Note:**
>
> If you already added the RAM user to the project, you can use the RAM user to log in to DataWorks. After logging in, you need to update the personal AccessKey information of the RAM user in the number plus console. For more information, see **#unique_19**.

2.6 Update Account Information

If this is your first time logging on to DataWorks, you must complete the configurations in the following procedure in #unique_4 and #unique_6, then update your account information as instructed.

The critical AccessKey information of the primary account is associated with the project member management along with most of the operations. Therefore, the primary account user must ensure the AccessKey information in DataWorks is consistent with the valid AccessKey value in the RAM management page.
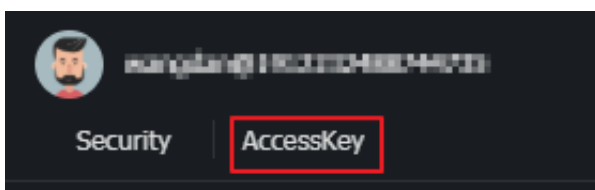
Procedure

1. Log on to the DTplus consoleand click Personal Information.
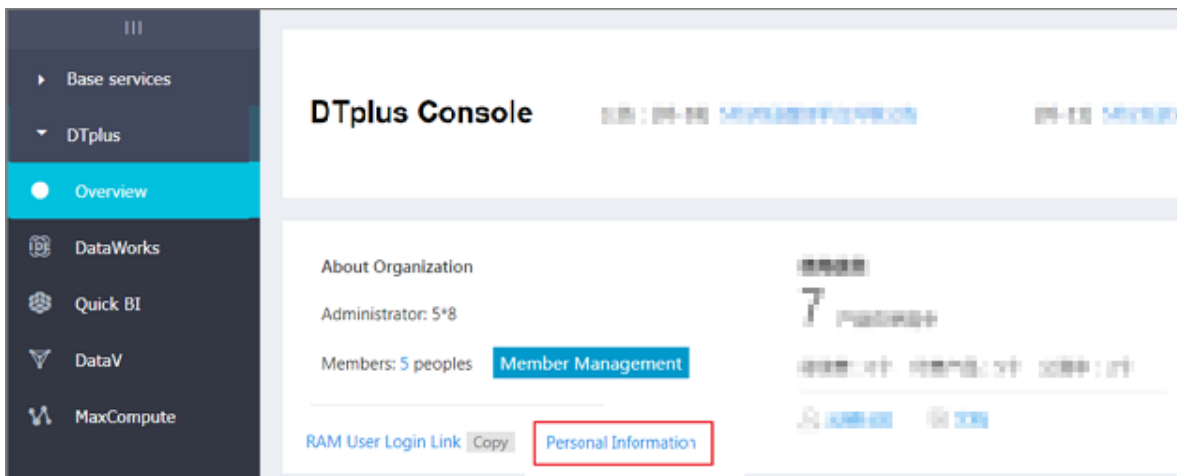
2. On the Personal Information page, you can view and modify the AccessKey information.
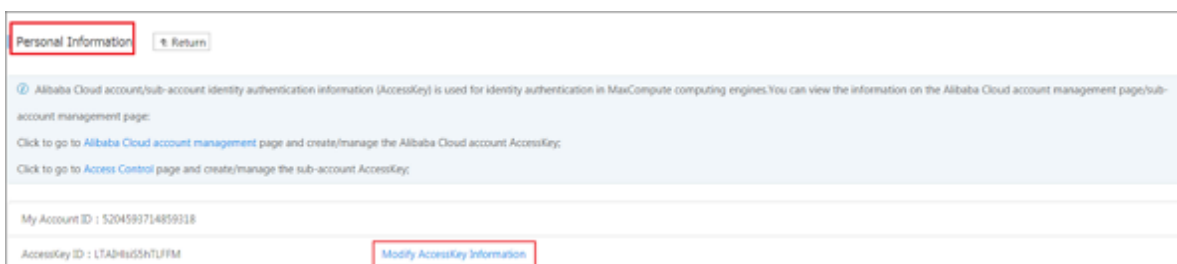
3. You must confirm the preceding AccessKey is consistent with the valid AccessKey, and is enabled and displayed in the DTplus console.