

# Alibaba Cloud DataWorks

Preparation

Issue: 20181008

# 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. |  <b>Danger:</b><br>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.  |  <b>Warning:</b><br>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.                           |  <b>Note:</b><br>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.                       |  <b>Note:</b><br>You can use <b>Ctrl + A</b> to select all files.   |
| >   | Multi-level menu cascade.  | <b>Settings &gt; Network &gt; Set network type</b>   |
| <b>Bold</b>   | It is used for buttons, menus, page names, and other UI elements.  | Click <b>OK</b> .  |
| Courier font  | It is used for commands.   | Run the <code>cd /d C:/windows</code> command to enter the Windows system folder.  |
| <i>Italics</i>  | It is used for parameters and variables.   | <code>bae log list --instanceid Instance_ID</code>   |
| [] or [a b]   | It indicates that it is a optional value, and only one item can be selected.   | <code>ipconfig [-all -t]</code>  |
| { } or {a b}  | It indicates that it is a required value, and only one item can be selected.   | <code>swich {stand   slave}</code>   |

# Contents

---

|   |          |
|---|----------|
| <b>Legal disclaimer</b> .....           | <b>I</b> |
| <b>Generic conventions</b> .....        | <b>I</b> |
| <b>1 Sub-account operations</b> .....   | <b>1</b> |
| <b>2 Administrator operations</b> ..... | <b>3</b> |
| 2.1 Prepare Alibaba Cloud account.....  | 3        |
| 2.2 Create a sub-account.....           | 5        |
| 2.3 New user guide.....                 | 7        |
| 2.4 Create a project.....               | 9        |
| 2.5 Add users and roles.....            | 14       |
| 2.6 Update account information.....     | 16       |

# 1 Sub-account operations

If this is your first time to use DataWorks and to work on the DTplus platform with your sub-account, you must make sure that you are equipped with the following information:

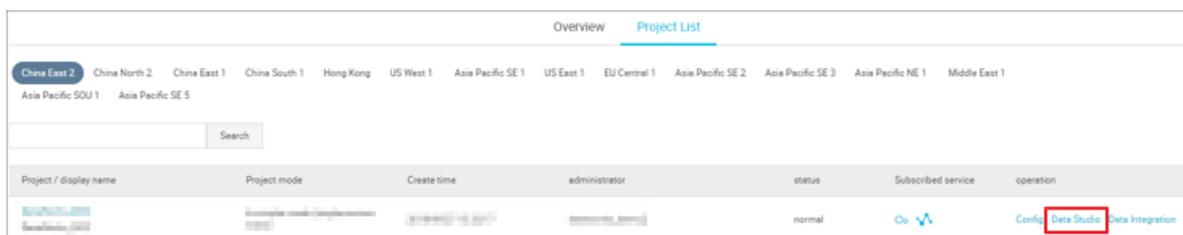
- The enterprise alias for the primary account to which your sub-account subordinates.
- The user name and password of your sub-account.
- The AccessKey ID and AccessKey Secret of your sub-account.
- Confirm that the primary account has allowed your sub-account to **Enable Console Logon**.
- Confirm that the primary account has allowed your sub-account to **Manage AccessKeys**.

If you are not sure about this information, verify the following with your primary account:

- The primary account user has prepared all configurations and information required by the sub-account according to [Prepare Alibaba Cloud account](#) and [Create a sub-account](#).
- The primary account user has successfully created a project and configured the appropriate member roles for your sub-account, by following the instructions mentioned in [Create a project](#) and [Create a project](#).

## Log on to DataWorks

1. Log on to the DTplus console by using the RAM user logon link provided by the primary account.
2. Navigate to the **DataWorks > Project List** page, and click **Enter Project** in the corresponding project to enter the data development page.



## Update personal information

Because your AccessKey is a private key, you must enter it manually before you start working on DataWorks.

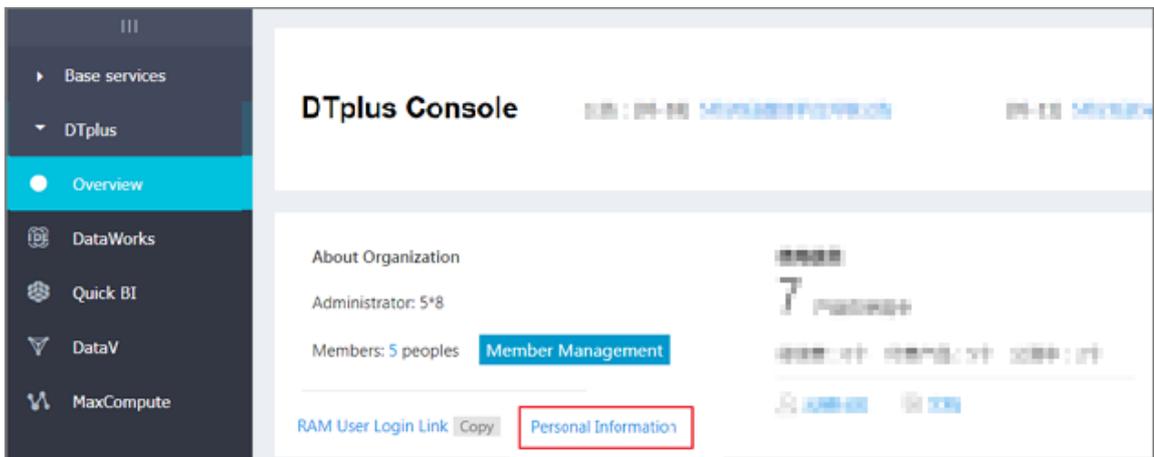
1. Enter the personal information page.

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

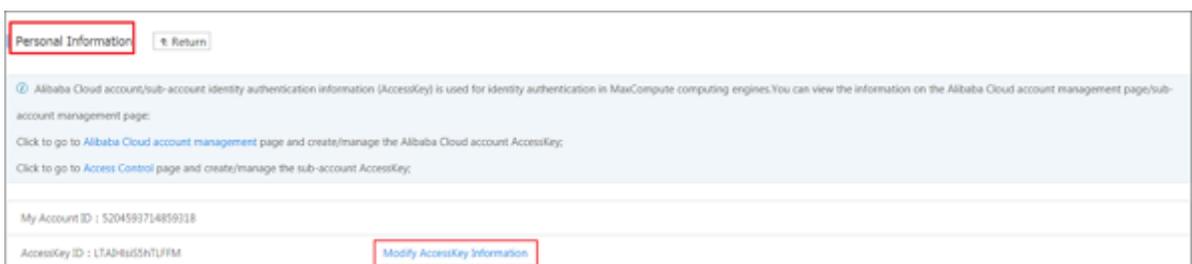
- If it is your first time to enter the DataWorks dashboard, you will be prompted to enter your AccessKey. Click the link in the red box and enter the personal information page to update your AccessKey information.

Your account is missing AccessId, please go to [here](#) to perfect

- 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 you have obtained from the primary account.



3. Enter the DTplus console again or refresh the page. Now, you can start your data development journey with DataWorks.

## 2 Administrator operations

---

### 2.1 Prepare Alibaba Cloud account

Like Alibaba Cloud, Alibaba Cloud DTplus products adopt the [RAM](#) primary account/subaccount logon system.

- Alibaba Cloud account (or the primary account) represents the ownership of Alibaba Cloud resources and is the basic subject for metering and billing resource usage. The primary account allows you to create sub-accounts for your enterprise and manage and authorize these sub-accounts.
- Created and managed by the primary account in the RAM system, the sub-account has no resources, and cannot perform metering and billing independently. All sub-accounts are controlled and paid by their primary account.

Therefore, you must have an Alibaba Cloud account and manage your sub-accounts by using the RAM before you use Alibaba Cloud DTplus products.

#### Register an Alibaba Cloud account

If you have not yet registered an Alibaba Cloud account, 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 create as a resource-consuming account with high-level permissions. Be sure to 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

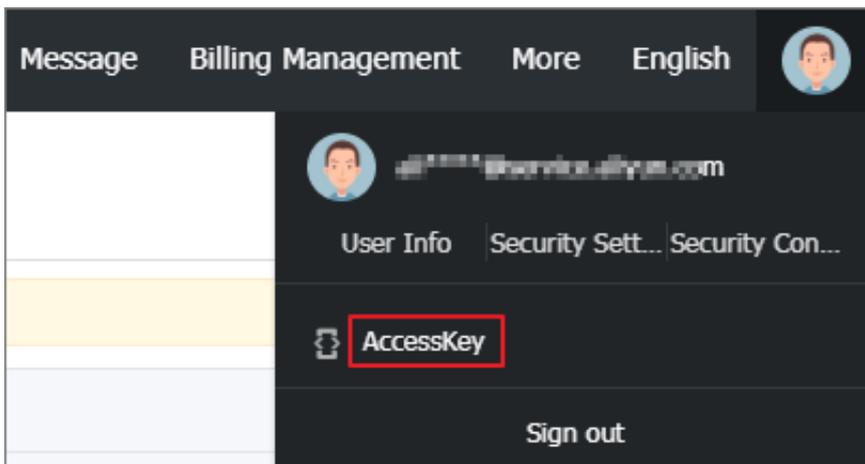
You must perform real-name registration on your Alibaba Cloud account before you purchase and use Alibaba Cloud products. If you have not yet completed real-name registration, go to **Real-name Registration** to go through the process. To save you any trouble in the future, be sure to complete the real-name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

If you are an enterprise user, we strongly recommend that you to 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 successfully created, the system automatically jumps to the AccessKey Management page, where you can view the status of the AccessKey and disable or delete it.

| Security Management  |                      |        |                     |                         |                                   |
|--|----------------------|--------|---------------------|-------------------------|-----------------------------------|
| ⓘ Access Key ID and Access Key Secret are the API keys for you to access Aliyun. It has full access privilege of the account. Please keep it safe. |                      |        |                     |                         |                                   |
| User Access Key  |                      |        |                     |                         | <a href="#">Create Access Key</a> |
| AccessKey ID   | Access Key Secret    | Status | Time Created        | Action                  |                                   |
| LTAI5E3v6i0R3FE  | <a href="#">Show</a> | Enable | 2018-02-26 15:49:00 | <a href="#">Disable</a> | <a href="#">Delete</a>            |
| LTAI5E3v6i0R3FE  | <a href="#">Show</a> | Enable | 2018-01-24 15:59:23 | <a href="#">Disable</a> | <a href="#">Delete</a>            |

Once an AccessKey is disabled, services that use this AccessKey encounter operational failures and report 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 an account. After creating an AccessKey, be sure to keep the AccessKey ID and AccessKey Secret safe. Do not reveal this information to others. If this information may have been leaked, immediately disable and update your AccessKey.

## 2.2 Create a sub-account

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

### Create a sub-account

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

### Enable a sub-account to log on the console

1. After the sub-account is successfully created, the main account enters the **RAM > Users** page, click **Manage** from the Actions column.
2. Click **Enable Console Logon** and reset the password, fill in the logon password of this sub-account.

### Create an AccessKey for a sub-account

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

- **Create an AccessKey for a sub-account**

1. The main account enters the **RAM > Users** page, click **Manage** from the Actions column.
2. Click **Create AccessKey**, and fill the verification code sent to your mobile phone. The AccessKey is created for the sub-account.



**Note:**

The AccessKeySecret of the sub-account appears only once. You must record and keep it confidential for further use. In case you lose it, create a new one.

- **Allow sub-account users to create and manage their own AccessKeys**

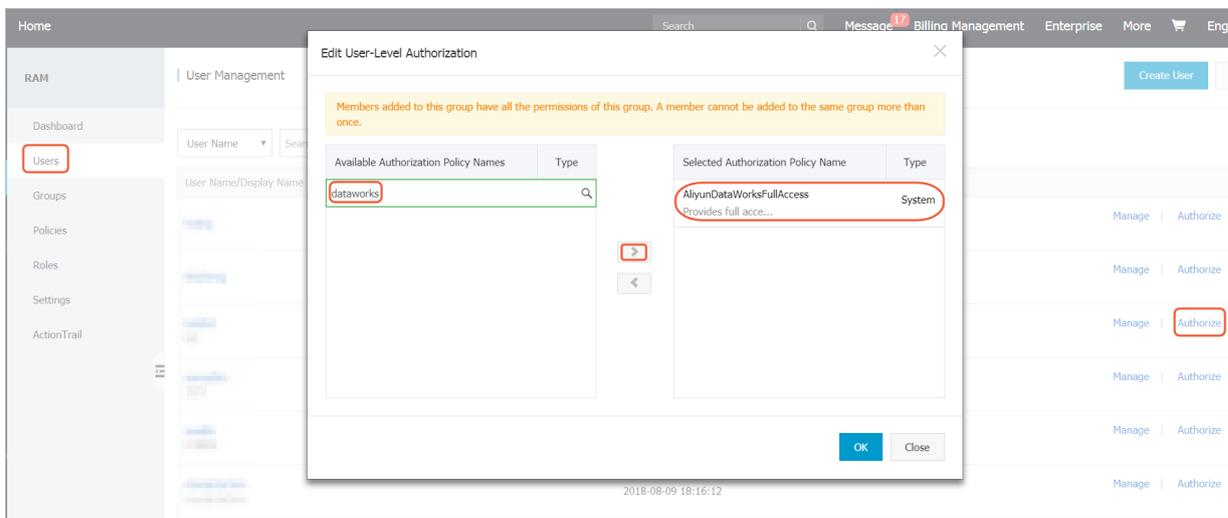
1. The main account enters the **RAM > Settings** page, click **User Security Settings** from the Actions column.
2. Select **User Security Settings** tab, and select **allow the user to manage their own AccessKey**, Click **Save Changes** to complete the settings.

This option is not checked by default, and the subaccount user cannot create and manage his own AccessKey. After **allow the user to manage their own AccessKey**, sub-account login to the Alibaba cloud, you can create the AccessKey on your own in the console.

 **Note:**  
It is very important to run the key AK. Whether it is the AK of the primary account or the AK of the sub-account, once it is created successfully, be sure to ensure its security and scope of use. If the AccessKey is leaked, timely prohibit it and create a new one.

### Authorize a sub-account

If you want to create projects using your sub-account, you may need you authorize it with **AliyunDataWorksFullAccess** policy.



### Delivery sub-accounts to other users

 **Note:**  
Sub-accounts belong to a primary account. Thus, sub-accounts do not have resources and billing system of their own. The operation cost of a DTplus sub-account is borne by the primary account. Fees incurred by operations of sub-accounts in Alibaba Cloud products shall be uniformly paid through the primary account to which these subaccounts belong. Therefore, the primary account

views the RAM user log on link and its own enterprise alias, and provide this information to a sub-account.

Log on to **RAM > Settings**, and click **Enterprise Alias Settings** to view enterprise alias.

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

- Logon link of RAM users.
- Enterprise alias of primary account.
- User name and password of sub-account.
- AccessKey ID and AccessKeySecret of a sub-account.
- Confirm that the primary account has allowed your sub-account to Enable Console Logon.
- Confirm that the primary account has allowed your sub-account to Manage AccessKeys.

## 2.3 New user guide

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

### Procedure

#### 1. Real-name registration.

To make purchases and use cloud products, your Alibaba Cloud account must be registered with a real-name. Log on to the **Real-name registration** page to complete the registration steps.



#### Note:

We recommend completing Real-name registration process in advance to assure smooth operation of the subsequent processes.

#### 2. Create an AccessKey.

To assure smooth processing of tasks of data in DataWorks, you must create an AccessKey. Similar to the way you use the account name and password to log on, an AccessKey is primarily used to access permission verification between various Alibaba Cloud products and services. For more information, see [Prepare Alibaba Cloud account](#).

An AccessKey comprises of an AccessKey ID and an AccessKeySecret. Generally, the AccessKey is secured and kept confidential. Once the AccessKey is disabled, the services that use this AccessKey fail to run and report exceptions to users. If an AccessKey information

changes, it can impact the products and services that use this AccessKey. Therefore, you must keep a close eye on such changes and take appropriate actions on time.

**Note:**

The AccessKey is highly critical to your account. After creating the AccessKey, you must secure the AccessKey ID and AccessKeySecret. Do not reveal this information to others on any condition. If you suspect of any information leak or other risks, you must promptly disable or modify your AccessKey.

**3. Select region and services.**

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

**4. Create a project.**

Only a primary account, which has the data integration function can create a project, and the other sub-accounts are treated only as users in the project.

**Note:**

This makes it necessary for a primary account to prepare a project environment for other sub-accounts.

- **Project name:** The length of the project name is at least 3 and must not exceed 27 characters.
- **Display name:** The length of the display name must not exceed 27 characters.
- **Enable scheduling frequency:** Control the current project whether to enable or disable the scheduling system. If it is disabled, it cannot schedule tasks periodically.
- **Allows editing of tasks and code directly in this project:** the current project member has permission to create/edit code files in this project, and if closed, cannot create/edit code files.
- **Enable select result downloads in this project:** 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:** A same project name of MaxCompute can be created when you create a project.
- **MaxCompute access identity:** The personal account and the system account.

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

5. Click **Create Project**.

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

## 2.4 Create a project

Check whether the primary account is available according to the [Prepare Alibaba Cloud account](#) account. If the account is available, perform the following steps to create a project.

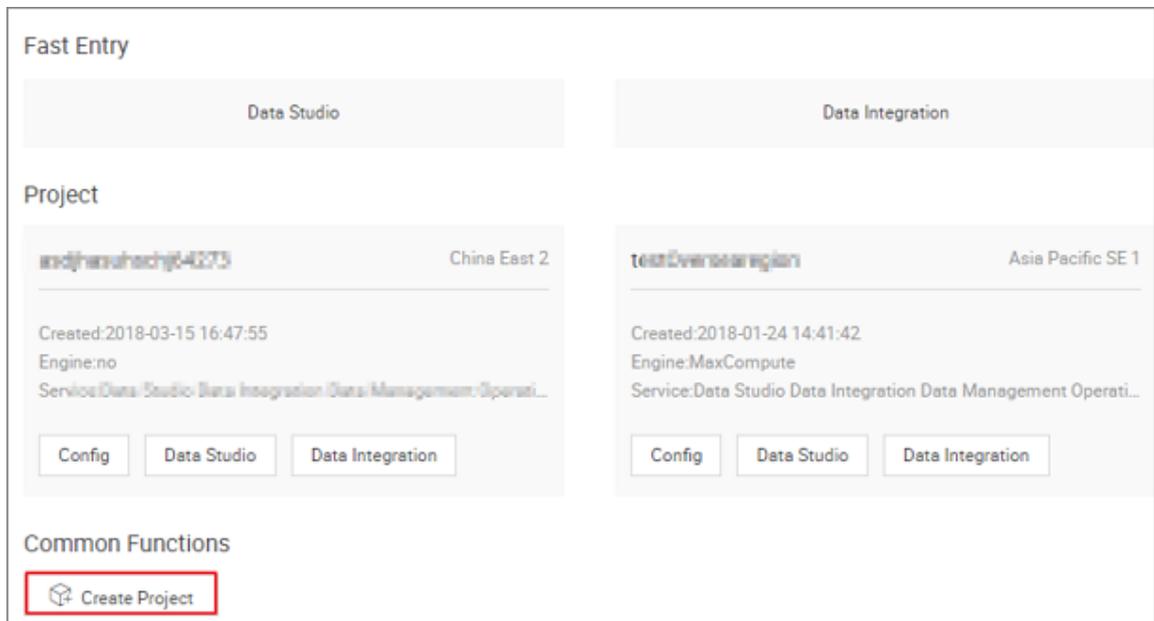


**Note:**

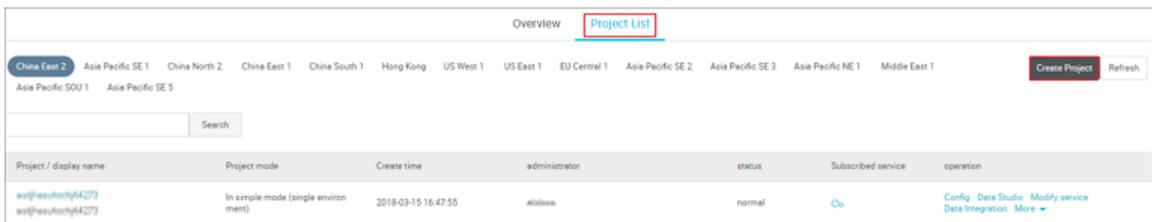
Currently, the sub-account can also create a project, and the procedure is the same as main accounts.

### Procedure

1. Log on to the [DataWorks console](#) by using a primary account.
2. You can create the MaxCompute project in the following two ways.
  - Click **Create Project** in **Common Functions** area.

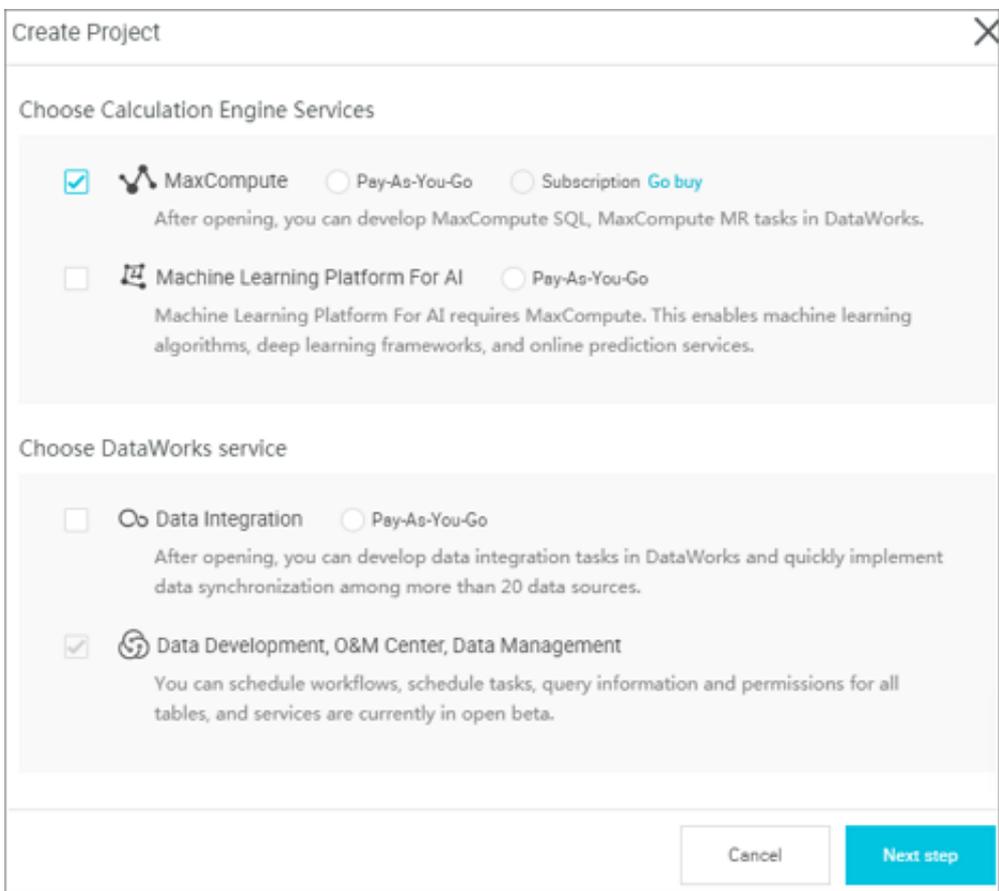


- Go on **Project List** page, click **Create Project**.



3. Fill in the configuration items in the **Create Project** dialog box. Select a region and a calculation engine service.

If you have not purchase the relevant services, it is directly display that there is no service available under the Region. The data development, O&M center, and data management are selected by default.



**Choose Calculation Engion Services:**

- **MaxCompute:** MaxCompute is a big data processing platform developed by Alibaba independently. It is mainly used for batch structural data storage and processing, which can provide massive data warehouse solution 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, and use empirical models to guide businesses.

**Choose DataWorks service:**

- Data integration: A data synchronization platform that provides stable, efficient, and elastically scalable services. The 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 [Data Integration overview](#).
  - Data development: The Data Development helps you to design data computing processes according to your business demands and make mutually dependent tasks be automatically run in the scheduling system. For more information, see [Data development overview](#).
  - O&M center: The 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 [Operation center overview](#).
  - Data management: The Data Management module of the Alibaba Cloud DTplus platform displays the global data view and metadata details of an organization, and enables operations such as divided permission management, data lifecycle management, and approval and management of data table/resource/function permissions. For more information, see [Data management overview](#).
4. Enter project basic information and set advanced settings.

**Create Project**

**Basic Information**

\* project name :

Display name :

project description :

**Advanced Settings**

\* Enable scheduling Frequency :  on ?

\* Enable Select result downloads in this project :  on ?

project :

for MaxCompute

\* MaxCompute Project Name:  ?

\* MaxCompute access identity: Project Leader Account

\* Quota group: Pay per view default resource ... ▾

Previous create

### Basic Information

- **Project name:** The length of the project name is between 3 and 27 characters.
- **Display name:** The length of the display name is not more than 27 characters.
- **Project mode:** The project mode is a new feature introduced by the new version of DataWorks. It is divided into simple mode and standard mode. for the difference between two project development patterns, see [The differences between a simple pattern and a standard pattern](#).
  - Simple mode: One DataWorks project corresponding to a MaxCompute project, can not set development and production environment, you can only do simple data development , can not control the data development process and table permissions.
  - Standard mode: One DataWorks project corresponding to two MaxCompute projects , can set development and production dual environment, improve code development specifications, and can strictly control the table permissions, prohibit the arbitrary

operation of the production environment table, to ensure the data security of the production table.

### Advanced Settings

- **Enable scheduling frequency:** Control the current project whether to enable or disable the scheduling system, and if it is disabled, it can not 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 the data query results from select statement.
- **For MaxCompute**
  - Development Environment MaxCompute Project name: the default is the project name + "\_dev" suffix, which can be modified.
  - Development Environment MaxCompute access identity: the default is a personal account.
  - Production Environment MaxCompute Project name: the default name is the same as the DataWorks project.
  - Production environment MaxCompute access identity: The default is the production account, it is not recommended to modify.
  - Quota group: Quota is used to implement disk quotas.

### 5. When the configuration is complete, click **Create project**.

After the project created successfully, you can view the content on the **Project List** page.

| Project / display name           | Project mode                        | Create time         | administrator | status | Subscribed service | operation   |
|----------------------------------|-------------------------------------|---------------------|---------------|--------|--------------------|---|
| aa@hahad2y4[2]<br>aa@hahad2y4[2] | In simple mode (single environment) | 2018-03-15 16:47:55 | alibcc        | normal | Cis                | Config Data Studio Modify service Data Integration More |



#### Note:

- Once you become a project owner, it means that everything in the project is yours, and no one has permission to access your project before the authentication.
- For general users, it is not necessary to create a project. If you are added to a project, you can use the MaxCompute.

## Next step

Now, you have learned how to create a project, and you can continue to learn the next tutorial. In next tutorial, you can learn how to quickly complete the operations of data development and data O&M. For more information, see [Quick Start](#).

## 2.5 Add users and roles

If the DTplus is for your personal use, you may ignore this article. If the DTplus is for your collaboration with other users, first you must create a sub-account according to [Create a sub-account](#) and [Create a project](#), then complete configurations as follows.

Like Alibaba Cloud, Alibaba Cloud DTplus products adopt the [RAM](#) primary account/sub-account logon system.

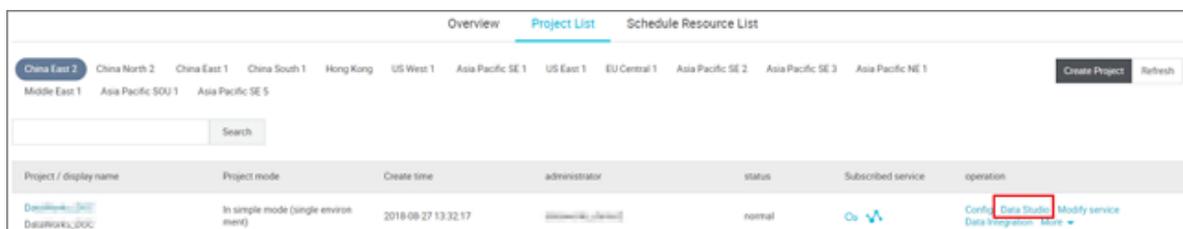
### Sub-account billing

Alibaba Cloud account (or the primary account) represents the ownership of Alibaba Cloud resources and is the basic specification for metering and billing resource usage. The primary account allows you to create sub-accounts for your enterprise, manage, and authorize these sub-accounts.

Created and managed by the primary account in the RAM system, the sub-account has no resources, and therefore cannot perform metering and billing independently. All sub-accounts are controlled and paid by their primary account.

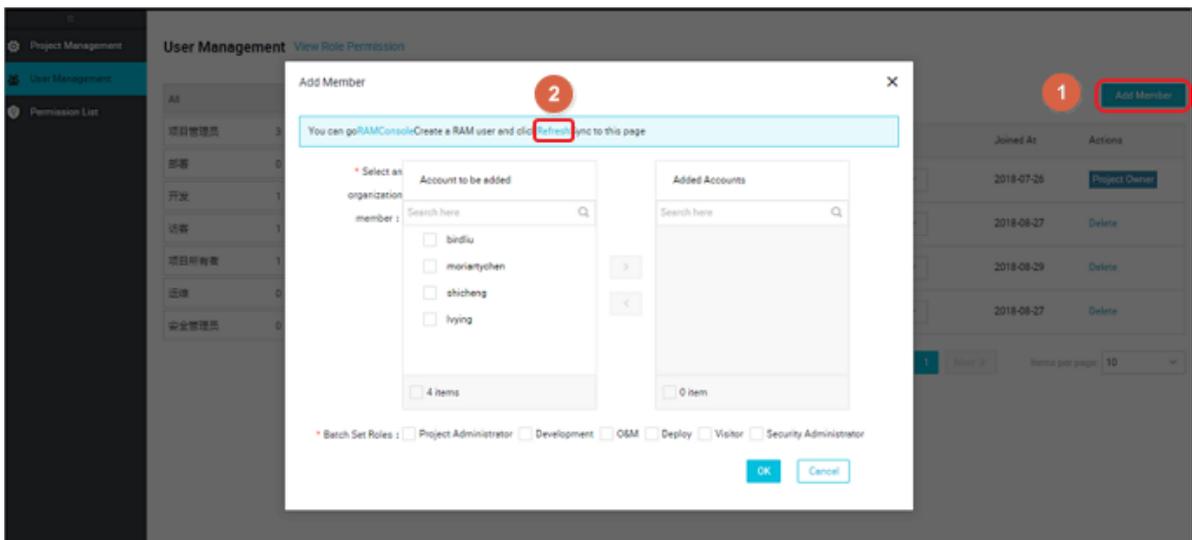
### Procedure

1. Go to the **project list** page of the DataWorks, and click the **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.

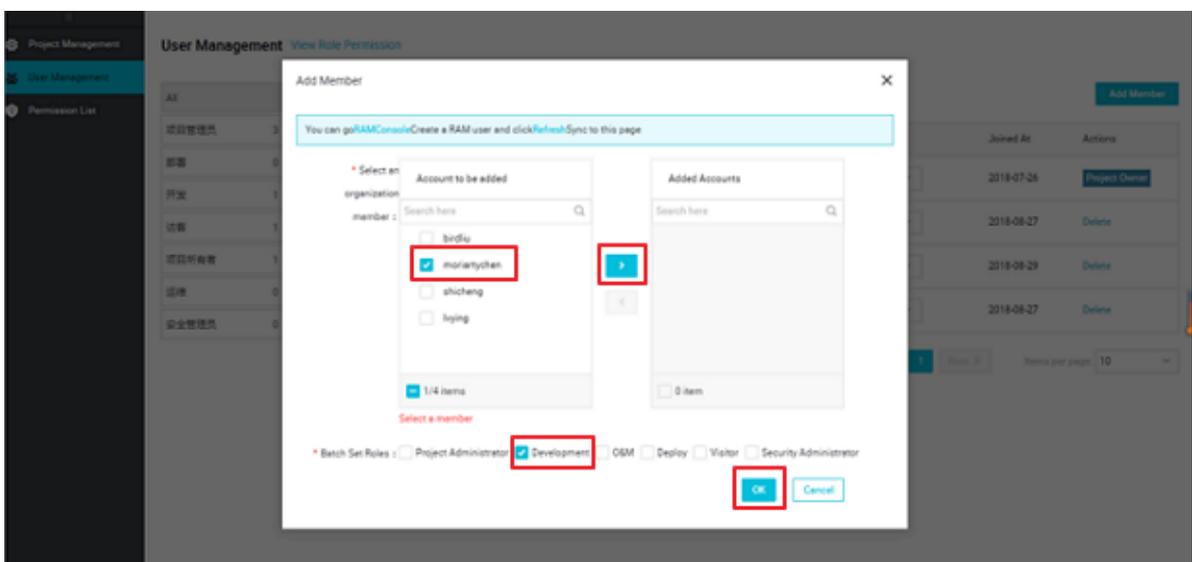
Click **Refresh** to synchronize the sub-account under the current Alibaba Cloud account in the RAM console to the list to be selected.



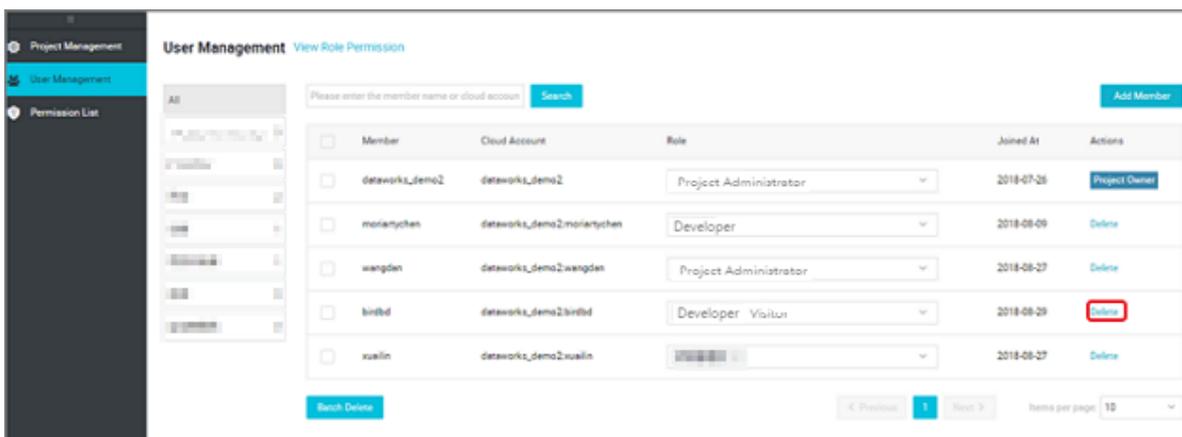
**Note:**

To add a new sub-account, click **RAM Console** in the preceding figure to open the Alibaba Cloud Resource Access Management page. For specific procedures to add a sub-account and deliver it for use, see [Prepare Alibaba Cloud account](#).

5. Check sub-accounts in the **New Members** dialog box, select a role to be granted to these sub-accounts 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 needed. You can also remove sub-accounts 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.

- **Developer:** Manages the design of data development page and the maintenance of workflow.
- **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 of the developer and the maintainer, the project administrator can also perform operations on the project level, such as adding/removing the project members, granting roles, and creating custom resource groups.
- **Deployer:** Reviews the task code and decide whether to submit it to the maintainer only in the multi-project mode.
- **Visitor:** The visitor can view the workflow design and code content on the data development page with the read-only permissions.



#### Note:

If you have already added the sub-account to the project, you can use the sub-account to log in to DataWorks. After logging in, you need to update the personal AK information of the sub-account in the number plus console. For more information, see [Sub-account operations](#).

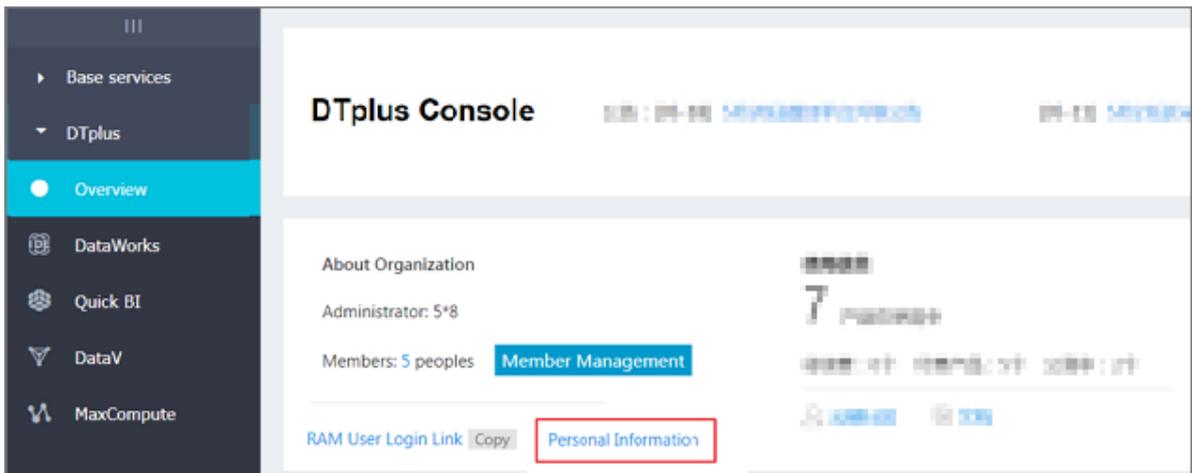
## 2.6 Update account information

If this is your first time to work with DataWorks, you must complete the configurations by following the procedure in [Prepare Alibaba Cloud account](#) and [Create a project](#), then update your account information as instructed.

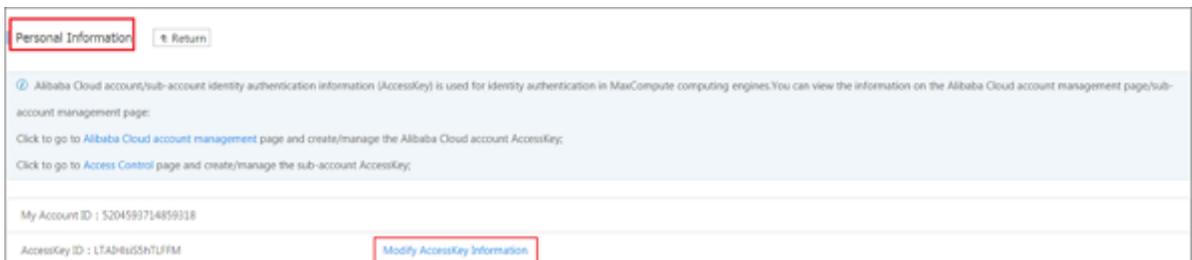
The critical AccessKey information of the primary account is associated with the project member management along with most of the operations. Therefore, the primary account user must make sure that the AccessKey information in DataWorks is consistent with the valid AccessKey value in the RAM management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

## Procedure

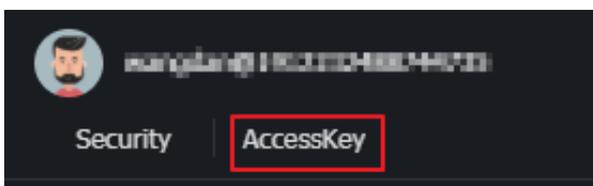
1. Log on to the **DTplus console** and click **Personal Information**.



2. On the **Personal Information** page, you can view and modify the AccessKey information.



3. You must confirm that the preceding AccessKey is consistent with the valid AccessKey, and it is enabled and displayed in the DTplus console.



Security Management

ⓘ Access Key ID and Access Key Secret are the API keys for you to access Aliyun. It has full access privilege of the account. Please keep it safe.

User Access Key Create Access Key

| AccessKey ID   | Access Key Secret | Status | Time Created        | Action           |
|----------------|-------------------|--------|---------------------|------------------|
| LTAI5T1Jo6R2Tl | Show              | Enable | 2018-02-26 15:49:00 | Disable   Delete |
| LTAI5T1Jo6R2Tl | Show              | Enable | 2018-01-24 15:59:23 | Disable   Delete |