Alibaba Cloud

MaxCompute Tutorials

Document Version: 20220610

(-) Alibaba Cloud

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.

- You shall download and obtain this document from the Alibaba Cloud website or other Alibaba Cloudauthorized 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 because of product version upgrade, adjustment, or other reasons. Alibaba Cloud reserves the right to modify the content of this document without notice and an updated version of this document will be released through Alibaba Cloud-authorized channels from time to time. You should 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 this document based on the "status quo", "being defective", and "existing functions" of its products and services. 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 take legal responsibility for any errors or lost profits incurred by any organization, company, or individual arising from download, use, or trust in this document. Alibaba Cloud shall not, under any circumstances, take responsibility for any indirect, consequential, punitive, contingent, 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 contents in Alibaba Cloud documents, including but not limited to pictures, architecture design, page layout, and text description, 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 this document 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 directly contact Alibaba Cloud for any errors of this document.

Document conventions

Style	Description	Example
<u>Nanger</u>	A danger notice 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.
<u> </u>	A warning notice 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 restart an instance.
Notice	A caution notice indicates warning information, supplementary instructions, and other content that the user must understand.	Notice: If the weight is set to 0, the server no longer receives new requests.
? Note	A note indicates supplemental instructions, best practices, tips, and other content.	? Note: You can use Ctrl + A to select all files.
>	Closing angle brackets are used to indicate a multi-level menu cascade.	Click Settings> Network> Set network type.
Bold	Bold formatting is used for buttons , menus, page names, and other UI elements.	Click OK.
Courier font	Courier font is used for commands	Run the cd /d C:/window command to enter the Windows system folder.
Italic	Italic formatting is used for parameters and variables.	bae log listinstanceid Instance_ID
[] or [a b]	This format is used for an optional value, where only one item can be selected.	ipconfig [-all -t]
{} or {a b}	This format is used for a required value, where only one item can be selected.	switch {active stand}

> Document Version: 20220610

Table of Contents

I.Build an online operation analysis platform 0)5
1.1. Business scenarios and development process 0)5
1.2. Activate services 0)5
1.3. Build data models and analyze data 1	3
1.3.1. Create tables 1	3
1.3.2. Design workflows2	20
1.3.3. Configure nodes2	21
1.3.4. Commit and test nodes2	23
1.4. Display data on dashboards 2	27

1.Build an online operation analysis platform

1.1. Business scenarios and development process

This tutorial describes an end-to-end solution for you to use big data services of Alibaba Cloud to build an online operational analytics platform based on the basic requirements for such a platform in the big data era. This solution allows you to write data to a storage system at high concurrency, process data efficiently, and analyze and display data.

Business scenarios

This tutorial uses HTTP access logs from a real website as the source data. You can analyze the access logs to obtain the following statistics:

- Page views (PVs) and unique visitors (UVs) of the website in terminal types, such as Android, iPad, iPhone, and PC.
 - Note PV and UV are two basic metrics for measuring the traffic of a website. A PV is recorded each time a user visits a web page. If a user visits the same web page multiple times, each view is added as a PV. An UV refers to a unique visitor to a website in one day. If a user visits the website multiple times in one day, the user is counted as only one UV.
- Source regions of website traffic.

Development process

The development process in this tutorial consists of the following steps:

- Step 1: Activate services.
- Step 2: Create tables.
- Step 3: Design workflows.
- Step 4: Configure nodes.
- Step 5: Commit and test nodes.
- Step 6: Display data on dashboards.

1.2. Activate services

This topic describes how to activate Tablestore, MaxCompute, DataWorks, and Quick BI before you build an online operational analytics platform.

Prerequisites

- You have an Alibaba Cloud account. If you do not have an Alibaba Cloud account, go to the International site (alibabacloud.com) and click Free Account. On the page that appears, create an Alibaba Cloud account.
- The Alibaba Cloud account completes real-name verification. Otherwise, go to the Real-name

Verification page to complete real-name verification.

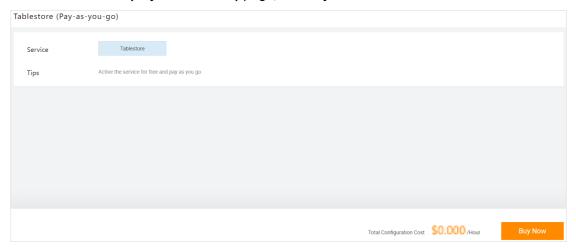
Context

The following Alibaba Cloud services must be activated:

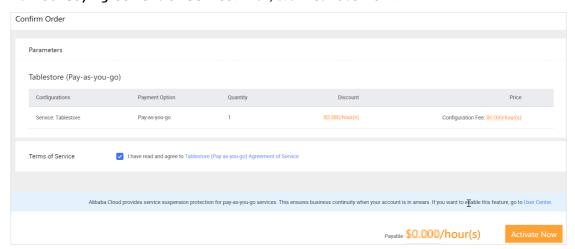
- Tablestore Tablestore
- MaxCompute MaxCompute
- DataWorks DataWorks
- Quick BI Quick BI

Procedure

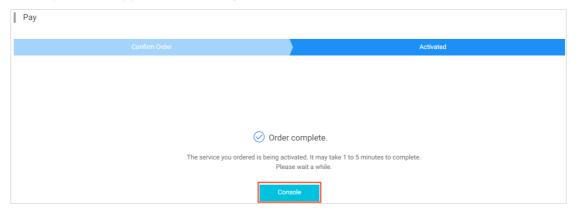
- 1. Activate Tablestore and create a Tablestore instance.
 - i. On the product page of Tablestore, click Get it Free.
 - ii. On the Tablestore (Pay-As-You-Go) page, click Buy Now.



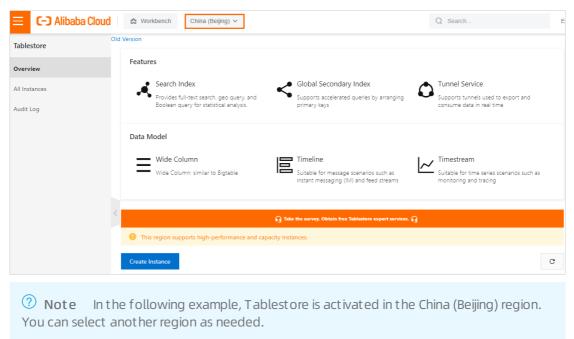
iii. On the Confirm Order page, read and select I have read and agree to Tablestore (Pay-As-You-Go) Agreement of Service. Then, click Activate Now.



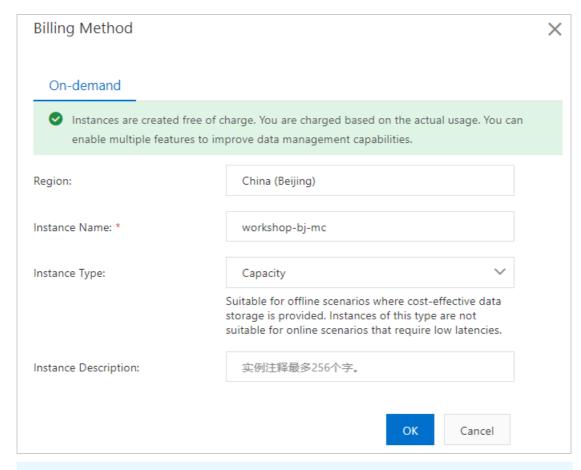
iv. On the page that appears, click Management Console.



v. On the **Overview** tab of the **Console** page, select **China (Beijing)** in the top navigation bar and click **Create Instance**.



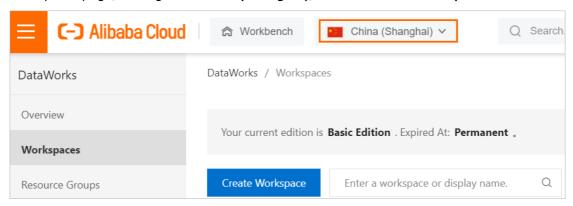
vi. In the **Billing Method** dialog box, click the **On-demand** tab, and enter a name in the **Instance Name** field. Select **Capacity** from the **Instance Type** drop-down list, and click **OK**.



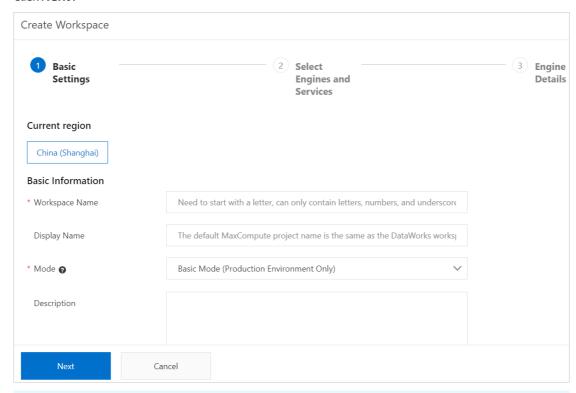
? Note

- The name of a Tablestore instance must be unique in each region. We recommend that you enter a name that is easy to identify and complies with the naming conventions.
- The instance name is required when you synchronize data from MaxCompute to the Tablestore instance. In this topic, the instance is named workshop-bj-mc. For more information about Tablestore instances, see Instance.
- vii. In the left-side navigation pane, click **All Instances**. On the All Instances page, view the created instance, which is in the **Running** state.
- 2. Activate MaxCompute.
 - i. On the product page of MaxCompute page, click Buy Now.
 - ii. On the page that appears, set **Product Type** to **Pay-As-You-Go**, set **Region** to **China** (**Shanghai**), and then click **Buy Now**.
 - **? Note** If you activate MaxCompute in the same region as Tablestore, the data transfer cost is reduced. Therefore, you can set Region to China (Beijing). In this topic, the region of MaxCompute is set to China (Shanghai) to demonstrate the use of external tables across regions.
- 3. Activate DataWorks.
 - i. On the product page of DataWorks page, click Buy Now.

- ii. Set Region to China (Shanghai), and click Buy Now.
 - Note If you activate DataWorks in the same region as Tablestore, the data transfer cost is reduced. Therefore, you can set Region to China (Beijing). In this topic, the region of DataWorks is set to China (Shanghai) to demonstrate data access across regions.
- 4. Create a DataWorks workspace.
 - i. Log on to the DataWorks console. In the left-side navigation pane, click Workspaces. On the Workspaces page, set Region to China (Shanghai), and click Create Workspace.



ii. In the **Create Workspace** pane, configure the parameters in the **Basic Settings** step and click **Next**.

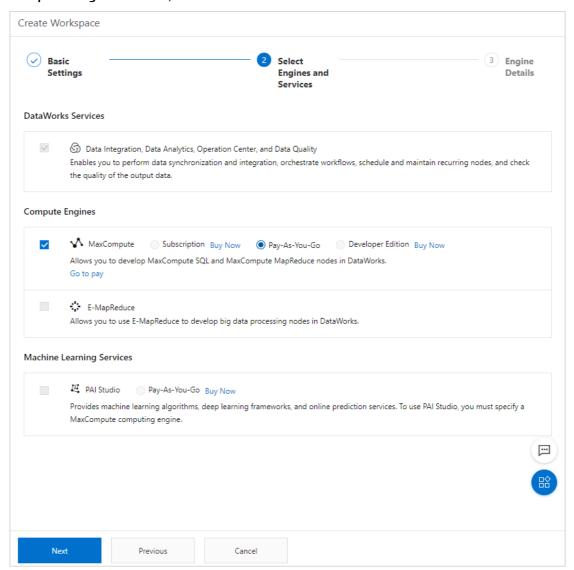


? Note

- The workspace name must be globally unique. We recommend that you use a distinct name.
- For ease of use, mode of the DataWorks workspace is set to Basic Mode (Production Environment Only) in this topic. In basic mode, a DataWorks workspace is associated with only one MaxCompute project. For more information, see Basic mode and standard mode.

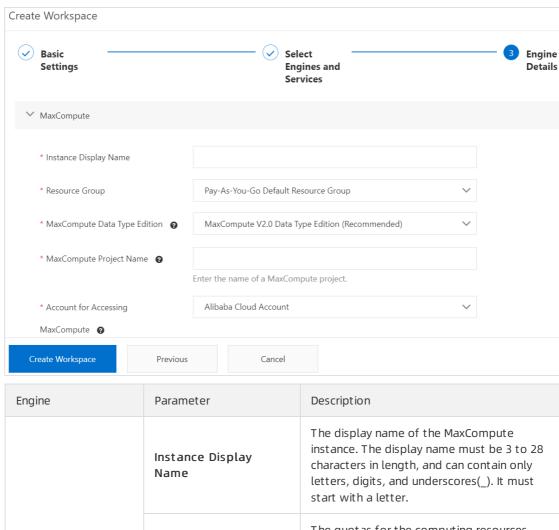
10

iii. In the Select Engines and Services step, select MaxCompute and Pay-As-You-Go in the Compute Engines section, and click Next.



iv. In the Engine Details step, set the required parameters and click Create Workspace.

After the workspace is created, you can view the information of the workspace on the **Workspaces** page.



The quotas for the computing resources Resource Group and disk space of the instance. MaxCompute Data The data type edition of the MaxCompute Type Edition project. MaxCompute By default, the name is the same as that of MaxCompute Project Name the DataWorks workspace. Account for accessing MaxCompute includes Account for Accessing Alibaba Cloud account and Node MaxCompute Owner. Specifies whether to encrypt the MaxCompute instance configuration based Whether to encrypt on your business requirements.

5. Activate Quick Bl. For more information, see Purchases, upgrades, downgrades, renewals, and overdue payments.

1.3. Build data models and analyze data

1.3.1. Create tables

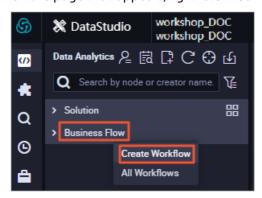
This topic describes how to create tables in MaxCompute to store raw data and processed data.

Prerequisites

- MaxCompute is activated. A DataWorks workspace in basic mode is created. For more information, see Activate services.
- You are granted the permissions to access data in Tablestore. If you use the same Alibaba Cloud
 account to activate MaxCompute and Tablestore, you can obtain the permissions with one click on
 the Cloud Resource Access Authorization page. If you use different Alibaba Cloud accounts to
 activate MaxCompute and Tablestore, you can obtain the permissions by using a custom policy. For
 more information, see Access Tablestore data.

Procedure

- 1. Go to the DataStudio page in the DataWorks console.
 - i. Log onto the DataWorks console. In the left-side navigation pane, click Workspaces. On the page that appears, select China (Shanghai) in the top navigation bar.
 - ii. Find the workspace that you created and click Data Development in the Actions column.
- 2. Create a workflow.
 - i. On the page that appears, right-click Business Flow and select Create Workflow.



- ii. In the Create Workflow dialog box, specify values for **Workflow Name** and **Description**, and click **Create**. In this example, Workflow Name is set to Workshop.
- 3. Create tables.
 - Create an external table named ots_user_trace_log.
 - Notice You cannot create external tables by using DDL statements in wizard mode. To create an external table, you can manually add fields to the table in wizard mode. You can also create an ODPS SQL node in MaxCompute and execute a DDL statement on the ODPS SQL node to create an external table. The following example shows how to execute a DDL statement on an ODPS SQL node to create an external table.

- a. On the Data Analytics tab of the DataStudio page, unfold the Workshop workflow, rightclick MaxCompute, and then choose Create > ODPS SQL. In the Create Node dialog box, set Node Name to workshop, and click Commit.
- b. Double-click the ODPS SQL node that you created. On the tab that appears, enter the following DDL statement in the code editor to create an external table named ots_user_trace_log.

Sample DDL statement:

```
CREATE EXTERNAL TABLE ots user trace log (
   md5 string COMMENT 'The first eight characters in the MD5 value of the user
ID (UID)',
   uid string COMMENT 'The UID',
   ts bigint COMMENT 'The timestamp when the user performs the operation',
   ip string COMMENT 'The IP address',
    status bigint COMMENT 'The status code returned by the server',
   bytes bigint COMMENT 'The number of bytes returned to the client',
   device string COMMENT 'The model of the terminal on which the client runs',
   system string COMMENT 'The version of the operating system in which the cli
ent runs. The version must be in the ios xxx or android xxx format',
   customize event string COMMENT 'The custom event. You can specify a custom
event that is reported for logon, logoff, purchase, registration, click, backgr
ound operations, account switching, browsing, or commenting',
    use time bigint COMMENT 'The duration for which the app is used at a time.
This field is required if you configure a custom event that is reported for log
off, background operations, or account switching',
    customize event content string COMMENT 'The content of the custom event. Th
is field is required if you configure a custom event that is reported for brows
ing or commenting'
STORED BY 'com.aliyun.odps.TableStoreStorageHandler'
WITH SERDEPROPERTIES (
    'tablestore.columns.mapping'=':md5,:uid,:ts, ip,status,bytes,device,system,
customize event, use time, customize event content',
    'tablestore.table.name'='user trace log'
LOCATION 'tablestore://workshop-bj-001.cn-beijing.ots.aliyuncs.com/';
```

■ STORED BY: required. Set the value to com.aliyun.odps.TableStoreStorageHandler .

This value specifies the storage handler that is built in MaxCompute to process the data stored in Tablestore. This parameter defines the logic of interactions between MaxCompute and Tablestore.

- SERDEPROPERITES: required. The WITH SERDEPROPERITES clause allows you to specify options. If you set the STORED BY parameter to com.aliyun.odps.TableStoreStorageHandler, you must specify the following options in the WITH SERDEPROPERITES clause:
 - tablestore.columns.mapping: the columns of the Tablestore table that you want to access from MaxCompute. The columns include primary key columns and attribute columns.

? Note

- Primary key columns are columns that begin with a colon (:). In this example, indicated, and its are primary key columns. The remaining columns are attribute columns.
- When you configure column mappings, you must specify all primary key columns in the Tablestore table. For the attribute columns, you need only to specify the columns that you want to access from MaxCompute. The Tablestore table must contain the specified attribute columns. Otherwise, errors are returned when you query data by using the external table that you created.
- tablestore.table.name: the name of the Tablestore table that you want to access from MaxCompute. If you specify an invalid table name or the table name that you specify does not exist, an error is returned. MaxCompute does not create a Tablestore table.
- LOCATION: the endpoint of Tablestore. For more information about how to specify a value for this parameter, see Activate services.
 - **Note** If you set this parameter to the public endpoint 'tablestore://workshop-bj-001.cn-beijing.ots.aliyuncs.com/', an error that indicates a network failure may be returned. In this case, you can set this parameter to the classic network endpoint, such as 'tablestore://workshop-bj-001.cn-beijing.ots-internal.aliyuncs.com/'.
- c. Click the Save icon. Then click the Run icon.

If the operational log indicates that the preceding statement is successfully executed, the external table is created.

Onte The first time you click the Run icon, you must select a resource group for scheduling. For more information about the scenarios in which resource groups for scheduling are used and how to use resource groups for scheduling, see Overview.

ii. Create a table named ods_user_trace_log.

Use the same method to create the table. The ods_user_trace_log table is a table at the operational data store (ODS) layer.

```
CREATE TABLE IF NOT EXISTS ods user trace log (
   md5 STRING COMMENT 'The first eight characters in the MD5 value of the UID',
   uid STRING COMMENT 'The UID',
   ts BIGINT COMMENT 'The timestamp when the user performs the operation',
   ip STRING COMMENT 'The IP address',
   status BIGINT COMMENT 'The status code returned by the server',
   bytes BIGINT COMMENT 'The number of bytes returned to the client',
   device STRING COMMENT 'The model of the terminal on which the client runs',
   system STRING COMMENT 'The version of the operating system in which the client
runs. The version must be in the ios xxx or android xxx format',
   customize event STRING COMMENT 'The custom event. You can specify a custom even
t that is reported for logon, logoff, purchase, registration, click, background ope
rations, account switching, browsing, or commenting',
    use time BIGINT COMMENT 'The duration for which the app is used at a time. This
field is required if you configure a custom event that is reported for logoff, back
ground operations, or account switching',
    customize event content STRING COMMENT 'The content of the custom event. This f
ield is required if you configure a custom event that is reported for browsing or c
ommenting'
PARTITIONED BY (
   dt STRING
);
```

iii. Create a table named dw_user_trace_log.

Use the same method to create the table. The dw_user_trace_log table is a table at the data warehouse (DW) layer.

```
CREATE TABLE IF NOT EXISTS dw user trace log (
   uid STRING COMMENT 'The UID',
   region STRING COMMENT 'The region in which the user resides. The region is obta
ined based on the IP address of the user',
   device brand string comment 'The brand of the terminal on which the client runs
   device STRING COMMENT 'The model of the terminal on which the client runs',
   system type STRING COMMENT 'The type of the operating system in which the clien
t runs. Valid values: Android, IOS, ipad, and Windows_phone',
    customize event STRING COMMENT 'The custom event. You can specify a custom even
t that is reported for logon, logoff, purchase, registration, click, background ope
rations, account switching, browsing, or commenting',
    use time BIGINT COMMENT 'The duration for which the app is used at a time. This
field is required if you configure a custom event that is reported for logoff, back
ground operations, or account switching',
   customize event content STRING COMMENT 'The content of the custom event. This f
ield is required if you configure a custom event that is reported for browsing or c
ommenting'
)
PARTITIONED BY (
   dt STRING
);
```

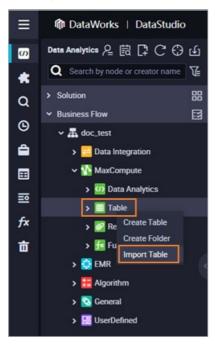
iv. Create a table named rpt_user_trace_log.

Use the same method to create the table. The rpt_user_trace_log table is a table at the application data store (ADS) layer.

```
CREATE TABLE IF NOT EXISTS rpt user trace log (
   country STRING COMMENT 'The country in which the user resides',
   province STRING COMMENT 'The province in which the user resides',
   city STRING COMMENT 'The city in which the user resides',
   device brand string comment 'The brand of the terminal on which the client runs
   device STRING COMMENT 'The model of the terminal on which the client runs',
   system type STRING COMMENT 'The type of the operating system in which the clien
t runs. Valid values: Android, IOS, ipad, and Windows phone',
    customize event STRING COMMENT 'The custom event. You can specify a custom even
t that is reported for logon, logoff, purchase, registration, click, background ope
rations, account switching, browsing, or commenting',
    use time BIGINT COMMENT 'The duration for which the app is used at a time. This
field is required if you configure a custom event that is reported for logoff, back
ground operations, or account switching',
   customize event content STRING COMMENT 'The content of the custom event. This f
ield is required if you configure a custom event that is reported for browsing or c
ommenting',
   pv bigint comment 'The number of page views (PVs)',
   uv bigint comment 'The number of unique visitors (UVs)'
PARTITIONED BY (
   dt STRING
);
```

v. Optional. If your workspace is in standard mode, the external table that you created is committed only to the development environment. To commit the external table to the production environment, perform the following steps:

a. On the Data Analytics tab of the DataStudio page, unfold your workflow below Business Flow, right-clickTable, and then select Import Table.

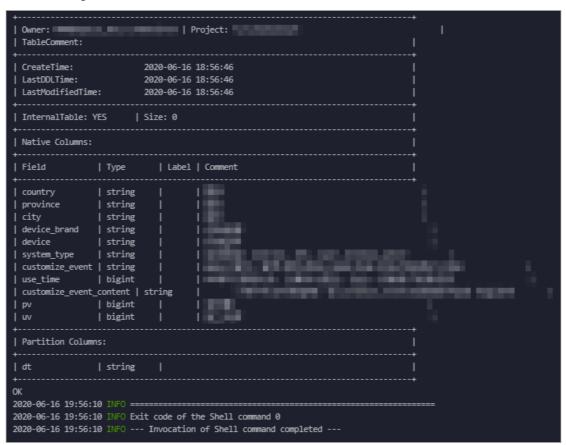


- b. In the Import Table dialog box, select the external table that is created in the development environment from the Table drop-down list, and click **OK**.
 - **? Note** You can move the pointer over the table name to check whether the table is created in the production environment or in the development environment. If you select a table that is created in the production environment, the system displays a message indicating that the table does not exist.
- c. Double-click the external table that you imported. On the tab that appears, click **Commit to Production Environment** in the upper part of the tab.
- 4. Verify the tables that you created.
 - i. After you create the tables, unfold the Workshop workflow below Business Flow, and choose MaxCompute > Table to view all tables that you created.
 - ii. Unfold the Workshop workflow below Business Flow, right-click Data Analytics below MaxCompute, and then choose Create > ODPS SQL.
 - iii. In the **Create Node** dialog box, specify a value for **Node Name**, and click **Commit** to create an ODPS SQL node.

iv. Enter the following SQL statements in the code editor of the ODPS SQL node, and click the or icon

```
DESCRIBE ots_user_trace_log;
DESCRIBE ods_user_trace_log;
DESCRIBE dw_user_trace_log;
DESCRIBE rpt_user_trace_log;
```

The following information about the table is returned:



1.3.2. Design workflows

You can design a workflow to arrange nodes to be run during data analytics. This topic describes how to design a workflow in which each node corresponds to a table at a data warehouse layer.

Procedure

- 1. Go to the DataStudio page in the DataWorks console.
 - i. Log on to the DataWorks console. In the left-side navigation pane, click Workspaces. On the page that appears, select **China (Shanghai)** in the top navigation bar.
 - ii. Find the workspace that you created and click Data Development in the Actions column.
- 2. Go to the DataStudio page. On the Data Analytics tab, double-click the Workshop workflow. The canvas for editing the workflow appears on the right.
- 3. Drag Zero-Load Node to the canvas. In the Create Node dialog box that appears, set Node Name to start and click Commit.

4. Drag ODPS SQL to the canvas. In the Create Node dialog box that appears, set Node Name to ods_user_trace_log and click Commit. Use the same method to create another two nodes and name them dw_user_trace_log and rpt_user_trace_log, respectively.

Note The ods_user_trace_log, dw_user_trace_log, and rpt_user_trace_log nodes represent the operational data store (ODS), common data model (CDM), and application data store (ADS) layers of a data warehouse, respectively. For more information, see <u>Divide a data</u> warehouse into layers.

1.3.3. Configure nodes

After you design a workflow, you need to configure each data analytics node in the workflow.

Prerequisites

A user-defined function (UDF) is registered. For more information, see Register a UDF.

Register a UDF

- 1. Add resources.
 - Download the GetAddr.jar package and ip.dat file for querying locations based on IP addresses.
 - ii. Go to the DataStudio page. On the Data Analytics tab, click the Workshop workflow and right-click MaxCompute. Choose Create > Resource > File to add a file resource, and choose Create > Resource > JAR to add a JAR resource.
 - When you add a file resource, upload the ip.dat file.
 - a. In the Create Resource dialog box, set Resource Name, select Large File (over 500 KB) and Upload to MaxCompute, and then click Upload. Select the ip.dat file and click OK.
 - b. On the tab that appears, click the **Submit** icon in the upper-left corner to commit the resource.
 - When you add a JAR resource, upload the **GetAddr.jar** package.
 - a. In the Create Resource dialog box, set Resource Name, select Upload to
 MaxCompute, and then click Upload. Select the GetAddr.jar package and click OK.
 - b. On the tab that appears, click the **Submit** icon in the upper-left corner to commit the resource.

Note When you commit the resource, if a message appears and indicates inconsistent data lineage, ignore it.

2. Register a UDF.

 Right-click MaxCompute under the Workshop workflow and choose Create > Function. In the Create Function dialog box that appears, set Function Name to getregion and click Commit. ii. On the Register Function tab that appears, set Class Name to odps.test.GetAddr, Resources to getaddr.jar,ip.dat, and Expression Syntax to getregion(ip string). Click the Save icon in the upper-left corner to save the settings and then click the iii icon to commit the UDF for registration.

Configure nodes

- 1. Configure the zero-load node named start.
 - i. On the Data Analytics tab, double-click the Workshop workflow. In the directed acyclic graph (DAG) of the workflow, double-click the start node. The node configuration tab appears.
 - ii. Click the **Properties** tab. On the Properties tab that appears, click **Use Root Node** in the **Dependencies** section.
 - iii. In the Schedule section, set Rerun to Allow Regardless of Running Status.
 - iv. Click the node.
- 2. Configure the ODPS SQL node named ods_user_trace_log.
 - i. In the DAG of the workflow, double-click the ods_user_trace_log node. On the node configuration tab that appears, enter the following SQL statements in the SQL editor:

```
insert overwrite table ods_user_trace_log partition (dt=${bdp.system.bizdate})
select
   md5,
   uid ,
   ts,
   ip,
   status,
   bytes,
   device,
   system,
   customize_event,
   use_time,
   customize_event_content
   from ots_user_trace_log
   where to_char(FROM_UNIXTIME(ts),'yyyymmdd')=${bdp.system.bizdate};
```

- Note For more information about the \${bdp.system.bizdate} parameter, see
 Overview of scheduling parameters.
- ii. Click the **Properties** tab. On the Properties tab that appears, set **Auto Parse** to No in the Dependencies section.
- iii. Remove incorrect parent nodes.
- iv. Search for each correct parent node based on the workflow design and click the **Add** icon to add the parent node. In this example, the parent node is the start node.
- v. In the Schedule section, set Rerun to Allow Regardless of Running Status.
- vi. Click the Submit icon.
- 3. Configure the ODPS SQL node named dw_user_trace_log.
 - You can configure the dw_user_trace_log node in the same way as configuring the

ods user trace log node. Enter the following SQL statements in the SQL editor:

```
INSERT OVERWRITE TABLE dw_user_trace_log PARTITION (dt=${bdp.system.bizdate})
SELECT uid, getregion(ip) AS region
    , CASE
       WHEN TOLOWER (device) RLIKE 'xiaomi' THEN 'xiaomi'
       WHEN TOLOWER (device) RLIKE 'meizu' THEN 'meizu'
       WHEN TOLOWER (device) RLIKE 'huawei' THEN 'huawei'
       WHEN TOLOWER (device) RLIKE 'iphone' THEN 'iphone'
       WHEN TOLOWER (device) RLIKE 'vivo' THEN 'vivo'
       WHEN TOLOWER (device) RLIKE 'honor' THEN 'honor'
       WHEN TOLOWER (device) RLIKE 'samsung' THEN 'samsung'
       WHEN TOLOWER (device) RLIKE 'leeco' THEN 'leeco'
       WHEN TOLOWER (device) RLIKE 'ipad' THEN 'ipad'
       ELSE 'unknown'
   END AS device brand, device
    . CASE
        WHEN TOLOWER(system) RLIKE 'android' THEN 'android'
       WHEN TOLOWER(system) RLIKE 'ios' THEN 'ios'
       ELSE 'unknown'
   END AS system type, customize event, use time, customize event content
FROM ods user trace log
WHERE dt = ${bdp.system.bizdate};
```

4. Configure the ODPS SQL node named rpt_user_trace_log.

You can configure the rpt_user_trace_log node in the same way as configuring the ods_user_trace_log node. Enter the following SQL statements in the SQL editor:

```
INSERT OVERWRITE TABLE rpt_user_trace_log PARTITION (dt=${bdp.system.bizdate})
SELECT split_part(split_part(region, ',', 1),'[',2) AS country
    , trim(split_part(region, ',', 2)) AS province
    , trim(split_part(region, ',', 3)) AS city
    , MAX(device_brand), MAX(device)
    , MAX(system_type), MAX(customize_event)
    , FLOOR(AVG(use_time / 60))
    , MAX(customize_event_content), COUNT(uid) AS pv
    , COUNT(DISTINCT uid) AS uv
FROM dw_user_trace_log
WHERE dt = ${bdp.system.bizdate}
GROUP BY uid,
    region;
```

5. Verify the configuration.

On the Data Analytics tab, double-click the Workshop workflow. The workflow editing canvas appears on the right. Click the \odot icon in the upper-left corner. If each node is marked with a green check sign $(\sqrt{})$, the nodes are run successfully.

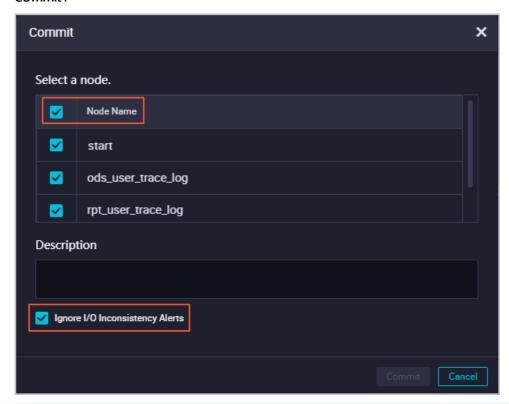
If a node fails to be run, right-click the node, select **View Log**, and then fix the issue based on logs.

1.3.4. Commit and test nodes

This topic describes how to commit nodes to Operation Center for testing after you configure the nodes.

Procedure

- 1. Optional. Commit a workflow. If you have committed nodes and have not modified the nodes after you configure them, skip this step.
 - i. Double-click the Workshop workflow. Then, click the 🖪 icon.
 - ii. Select all nodes that can be committed and **Ignore I/O Inconsistency Alerts**. Then, click **Commit**.

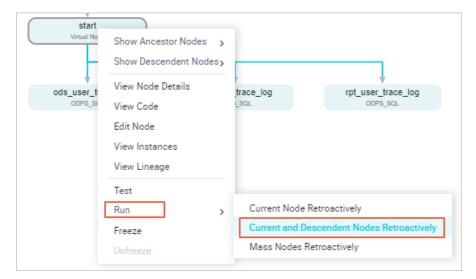


Note In a workspace in standard mode, after you commit the nodes, you must click Deploy and deploy the nodes to the production environment.

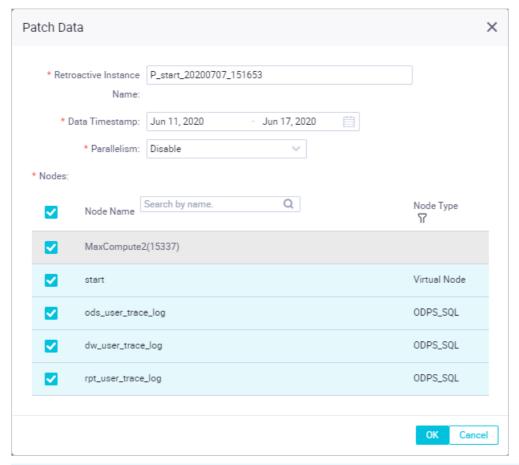
2. Click Operation Center in the upper-right corner.



- 3. In the left-side navigation pane, choose Cycle Task Maintenance > Cycle Task. Double-click the start node in the node list.
- 4. In the directed acyclic graph (DAG) that appears on the right, right-click the start node and choose Run > Current and Descendent Nodes Retroactively.



5. In the **Patch Data** dialog box, select the nodes for which retroactive data needs to be generated, set the data timestamp to the last week, and then click **OK**.

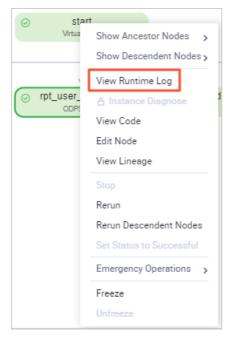


Note For more information about retroactive instances, see Perform retroactive data generation and view retroactive data generation instances.

6. In the left-side navigation pane, click **Patch Data**. View the running status of retroactive instances. You can click **Refresh** to refresh the page and view the running status of the instances in real time.



If a retroactive instance is abnormal, right-click the corresponding node in the DAG, select **View Runtime Log**, and then fix the issue.

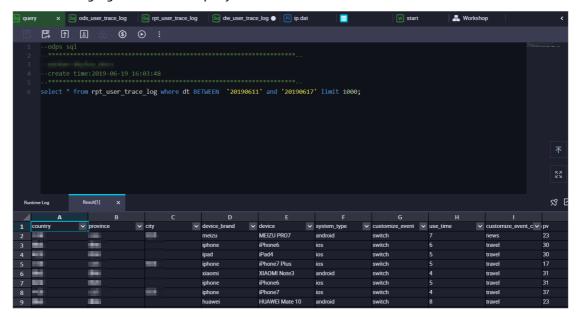


- 7. After the retroactive instances are run, verify the results.
 - i. In the left-side navigation pane, click the Workshop workflow and then MaxCompute. Rightclick Data Analytics and choose Create > ODPS SQL. Create an ODPS SQL node named query.
 - ii. To check whether data is written to the rpt_user_trace_log table, enter the following SQL statement to query data from June 11, 2019 to June 17, 2019 in the rpt_user_trace_log table:

```
select * from rpt_user_trace_log where dt BETWEEN '20190611' and '20190617' limit 1
000;
```

iii. Click the o icon.

The following figure shows the query results.



1.4. Display data on dashboards

After you process the rpt_user_trace_log table, you can create a dashboard in the Quick BI console to display the profiles of website users in the table.

Prerequisites

Before you start, make sure that you complete the steps in the Activate services and Data modeling and analytics topics.

Context

The rpt_user_trace_log table contains fields such as country, province, city, device_brand, use_time, and pv. You can perform the following steps to create a dashboard to display core metrics of users, periodic changes of user behavior data, and regional distribution and records of users.

- Step 1: Connect to a data source.
- Step 2: Create a dataset.
- Step 3: Display data analytics results on dashboards.

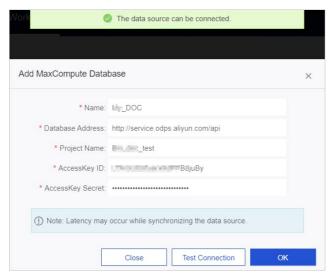
Step 1: Connect to a data source

- 1. Log on to the Quick BI console by using your Alibaba Cloud account.
- 2. On the homepage of the Quick BI console, select the workspace that you want to connect to a data source.
 - In this example, the default workspace is used.
- 3. On the page that appears, perform the steps that are shown in the following figure to connect the workspace to a data source.



4. In the **Add MaxCompute Database** dialog box, configure the parameters that are used to connect to the data source and test the network connectivity.

After the data source is added, you can view it in the data source list.



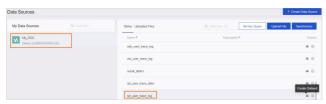
Parameter	Description
Name	The name of the data source. The name is displayed in the data source list. The name must be 1 to 50 characters in length, and can contain letters, digits, underscores (_), forward slashes (/), backslashes (\), vertical bars (), parentheses (), and brackets [].
	A default address is displayed. In most cases, you do not need to change the address.
Database Address	Note The address of the data source varies based on the region where the data source is deployed. For more information, see Endpoints.
Project Name	The name of the MaxCompute project. You can log on to the MaxCompute console and view the project name on the Project management tab.

Parameter	Description
AccessKey ID	The AccessKey ID of your Alibaba Cloud account or a RAM user of your Alibaba Cloud account. You can obtain the AccessKey ID on the AccessKey Pair page in the Resource Access Management (RAM) console.
AccessKey Secret	The AccessKey secret that corresponds to the AccessKey ID. You can obtain the AccessKey secret on the AccessKey Pair page in the RAM console.

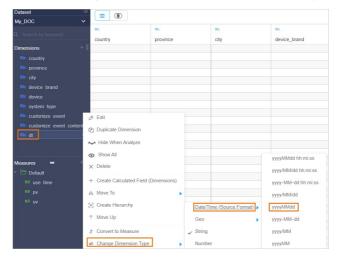
Step 2: Create a dataset

Datasets are the basis to display data analytics results in Quick BI. You can create a dataset for the data tables that you want to analyze. For more information about how to create a dataset, see Create a dataset.

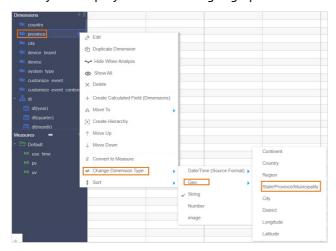
1. In the My Data Sources section of the **Data Sources** page, click the MaxCompute data source to which the workspace is connected. Find the **rpt_user_trace_log** table on the right side of the page, and click the **Create Dataset** icon in the Actions column.



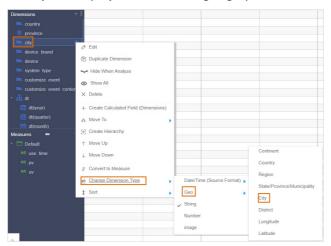
- 2. On the **Datasets** page, modify the display formats of fields.
 - i. Modify the display format of the date dimension type of the dt field.



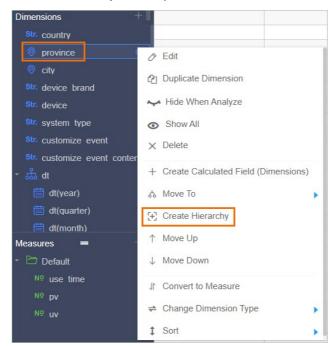
ii. Modify the display format of the geographic dimension type of the province field.



iii. Modify the display format of the geographic dimension type of the city field.



iv. Create a hierarchy for the province field.



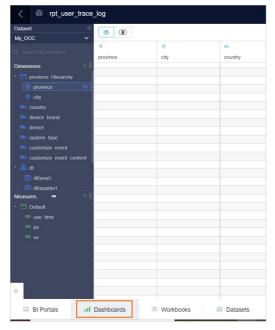
v. Drag the city field to the province hierarchy. The drilling effect is shown in the following figure.



Step 3: Display data analytics results on dashboards

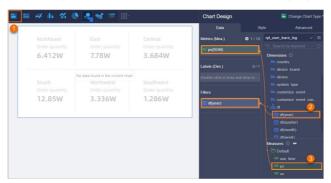
Quick BI provides you with charts to display data analytics results. For more information, see Create a dashboard.

1. On the **Datasets** page, click **Create Dashboards**.

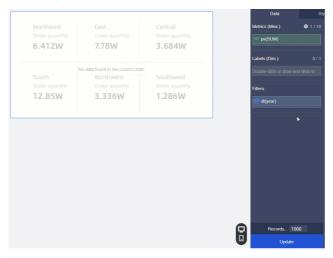


2. Create a dashboard.

i. Create a pv dashboard as shown in the following figure.

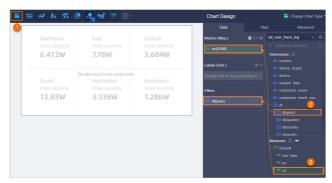


Set the pv dashboard filter as shown in the following figure.

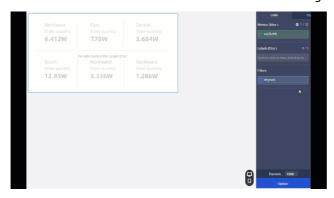


? Note The rpt_user_trace_log table is a partitioned table. Therefore, you must set a time range in the **filter**. The **time range** of the filter in the preceding figure is from 2021 to 2021.

ii. Create a uv dashboard as shown in the following figure.



Set the uv dashboard filter as shown in the following figure.



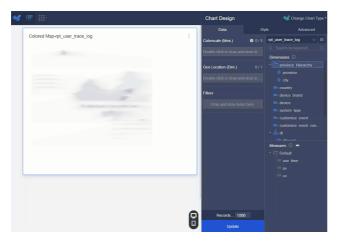
3. Create a line chart as shown in the following figure.



Set the line chart filter as shown in the following figure.



4. Create a filled map.



- 5. After you complete the configuration, click **Save** on the Dashboards page.
- 6. On the Dashboards page, click **Preview** to view the display effect.