ALIBABA CLOUD

# Alibaba Cloud

## DataWorks

## Task Operation

Document Version: 20211230

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.

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 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 |
| --- | --- | --- |
| Danger | 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. |
| Warning | 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 list --instanceid` *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}` |

# Table of Contents

# 1.Operation Center

## 1.1. Overview

Operation Center provides the following modules: Overview, RealTime Task Maintenance, Cycle Task Maintenance, Manual Task Maintenance, Alarm, Intelligent Diagnosis, Resource, and Engine Maintenance. You can use these modules to perform O&M operations on nodes, engines, and resources.

### Modules of Operation Center

In general, after you commit and deploy a node that is configured in DataStudio, you can perform O&M operations on the node in Operation Center. Operation Center provides multiple modules such as Overview and RealTime Task Maintenance. The following table describes the modules of Operation Center.

| Module | Description | Supported environment |
| --- | --- | --- |
| **Overview** | This module displays the statuses of nodes in charts. For more information, see View the dashboard. | This module is available only for the production environment of a DataWorks workspace in standard mode. |
| **RealTime Task Maintenance** | This module allows you to run, stop, and disable real-time sync nodes, and configure alert rules for monitoring real-time sync nodes. For more information, see Manage real-time synchronization nodes. | - |
| **Cycle Task Maintenance** | This module displays the auto triggered nodes that are committed to the scheduling system and the auto triggered node instances that are generated after the nodes are run by the scheduling system. On the Cycle Task page, you can view all existing auto triggered nodes and perform operations such as changing the resource groups and owners of auto triggered nodes.<br>DataWorks generates instances for auto triggered nodes each evening based on the point in time at which the auto triggered nodes are committed. You can perform operations on auto triggered nodes to generate retroactive instances and test instances for the nodes. For more information, see View auto triggered nodes. | - |

| Module | Description | Supported environment |
| --- | --- | --- |
| **Manual Task Maintenance** | This module displays the manually triggered nodes or workflows that are committed to the scheduling system and the manually triggered instances that are generated after the nodes or workflows are manually triggered. On the **Manual Task** page, you can view all existing manually triggered nodes or workflows and perform operations such as changing the resource groups and owners of manually triggered nodes or workflows.<br>In the upper part of the Manual Task page, you can set the Type parameter to Manually Triggered Workflow to view manually triggered workflows. You can trigger the manually triggered workflows to run and generate instances for the manually triggered workflows. Then, you can view the execution details of the manually triggered workflows on the Manual Instance page. For more information, see Manually triggered nodes. | - |
| **Alarm** | This module allows you to configure alert rules for monitoring auto triggered nodes. This module monitors the statuses of auto triggered node instances and the usage of exclusive resource groups in automatic scheduling scenarios.<br>This module also allows you to configure alert rules for monitoring a specific object such as a node, workflow, workspace, baseline, real-time compute node, exclusive resource group for scheduling, or exclusive resource group for Data Integration. In addition, this module allows you to configure alert rules for monitoring the baselines of the global line-of-business and sends alert notifications based on the notification method you specify. The notification methods include SMS messages, emails, and DingTalk chatbots. For more information, see Overview.<br>**Note** Only Singapore, Malaysia(Kuala Limpur), and Germany(Frankfurt) support the SMS reminding method. To use the SMS reminding method in other regions, submit a ticket to contact DataWorks technical support. | This module is available only for the production environment of a DataWorks workspace in standard mode. |

| Module | Description | Supported environment |
| --- | --- | --- |
| **Resource** | This module monitors the usage of exclusive resource groups for scheduling and exclusive resource groups for Data Integration and automatically performs O&M operations. You can view the usage of a resource group and the list of nodes that are using or waiting to use the resource group. For more information, see Resource O&M. | - |
| **Engine Maintenance** | This module is available only for E-MapReduce (EMR) compute engines. A DataWorks node instance that runs on an EMR compute engine contains multiple EMR jobs. These EMR jobs are run based on a specific sequence. You can use the Engine Maintenance module of DataWorks to view the details of each EMR job and identify and remove the jobs that fail to be run. This prevents failed jobs from affecting the execution of both the DataWorks node instance to which the jobs belong and the descendant nodes of the node that generates the node instance. For more information, see Use the engine O&M feature. | This module is available only for the production environment of a DataWorks workspace in standard mode. |
| **Intelligent Diagnosis** | This module helps you track the execution of nodes and identify problems. For more information, see Instance diagnosis. | This module is available only for the production environment of a DataWorks workspace in standard mode. |

When you use Operation Center of DataWorks, take note of the following items:

- In a DataWorks workspace in standard mode, you can click the switch icon next to Operation Center in the top navigation bar to switch between the development and production environments.
- Only Operation Center of the production environment supports automatic scheduling. You can view auto triggered node instances only on the Cycle Instance page in Operation Center of the production environment.

## Logic for running nodes

Before you run a node that is committed and deployed to Operation Center, an instance must be generated for the node. Then, DataWorks runs the node based on the trigger mechanism that is configured.

For example, an auto triggered node is run based on the following logic:

- After the auto triggered node is committed to Operation Center, DataWorks generates an instance each night for running the node the following day. Then, the generated instance is automatically triggered to run based on your scheduling configuration when the trigger conditions are met. The scheduling configuration includes node dependencies, the point in time when the node is automatically triggered to run, and the resources for running the node.
- You can manually generate retroactive data for the auto triggered node to generate a retroactive instance for the node. The retroactive instance can be run to obtain the retroactive data of the auto triggered node in the specified time range in the past. Alternatively, you can manually test an auto triggered node to generate a test instance for the node.

> **Note** To sum up, auto triggered node instances are automatically generated for auto triggered nodes and automatically run based on the scheduling configuration when the trigger conditions are met. Test instances or retroactive instances are generated after you manually trigger DataWorks to generate retroactive data for or test auto triggered nodes. Therefore, test instances and retroactive instances are generated based on the latest node configurations.

- On the **Cycle Instance**, **Patch Data**, or **Test Instance** page under **Cycle Task Maintenance**, you can view the statuses of auto triggered node instances, retroactive instances, or test instances. You can determine whether the data output of an instance is normal based on the status of the instance. For more information about instance statuses, see Related information: Instance status.

| Instance type | Scenario | How to generate the instance based on an auto triggered node | How to trigger the instance to run | Prerequisite for running the corresponding node |
| --- | --- | --- | --- | --- |
| Auto triggered node instance | You want to perform periodic extract, transform, load (ETL) operations. | DataWorks automatically generates an auto triggered node instance based on the snapshot information of the auto triggered node at a specific point in time.<br>**Note** If a DataWorks workspace in standard mode is used, an auto triggered node instance can be automatically generated and run only in the production environment. | DataWorks automatically triggers an auto triggered node instance to run. | |

| Instance type | Scenario | How to generate the instance based on an auto triggered node | How to trigger the instance to run | Prerequisite for running the corresponding node |
| --- | --- | --- | --- | --- |
| Retroactive instance | • You want to generate retroactive data for the current node and its descendant nodes for a specific time range in the past. In other words, you want to perform ETL operations on historical data.<br>• You want to generate retroactive data for the current node and its descendant nodes for a specific time range in the future. In other words, you want to perform ETL operations on future data in advance. | You need to manually trigger DataWorks to generate retroactive data for the auto triggered node and generate a retroactive instance for the node. | After you manually trigger DataWorks to generate retroactive data for the auto triggered node, a retroactive instance is generated and triggered to run. | The following prerequisites must be met before an auto triggered node starts to run:<br>• The parent node of this node is successfully run.<br>• The scheduled point in time for running the node is reached.<br>• Sufficient scheduling resources are provided for running the node.<br>• This node is not frozen. |

| Instance type | Scenario | How to generate the instance based on an auto triggered node | How to trigger the instance to run | Prerequisite for running the corresponding node |
| --- | --- | --- | --- | --- |
| Test instance | You want to test the current auto triggered node to check whether the node can be run as expected.<br>**Note** When you run a test instance, the code logic of the test instance is implemented. | You need to manually trigger DataWorks to test the auto triggered node and generate a test instance for the node. | After you manually trigger DataWorks to test the auto triggered node, a test instance is generated and triggered to run. | |

After an auto triggered node is deployed in the production environment, you can view the node on the Cycle Task page. However, when an instance is generated for the auto triggered node depends on the method that you use to generate the instance. For more information, see Configure time properties.

## Node O&M: Intelligent monitoring

You can configure alert rules for monitoring auto triggered node instances and exclusive resource groups.

- Custom alert rules

  You can configure a custom alert rule for monitoring the specified object.

  For example, you can specify that an alert notification is sent if a node of the specified object such as a baseline, workspace, or workflow is in one of the following states: Completed, Uncompleted, Error, Uncompleted in Cycle, Overtime, and The error persists after the node automatically reruns. You can specify that an alert notification is sent if an error occurs when a real-time compute node runs. You can also configure a custom alert rule for monitoring an exclusive resource group. For example, you can specify that an alert notification is sent if the resource usage or the number of node instances that are waiting for resources exceeds a specific threshold. In addition, you can use the automated O&M feature of intelligent monitoring to send notifications to the specified alert contact such as the node owner, specified responsible person, or on-duty engineer that you specify for the shift schedule. Notifications can be sent by using SMS messages, emails, or DingTalk chatbots. For more information, see Manage custom alert rules.

- Built-in global alert rules

You can monitor special events or implement global monitoring by using built-in global alert rules, including alert rules for node isolation, node loops, global events, and global baselines.

- Alert rule for node isolation

  Isolated nodes are nodes that do not have upstream dependencies. You can view the dependencies of an auto triggered node on the Cycle Task or Cycle Instance page. When DataWorks schedules auto triggered nodes to run, isolated nodes are excluded from the scheduling. Therefore, isolated nodes cannot be automatically triggered to run. If an isolated node has a large number of descendant nodes, the descendant nodes may fail to be run. After an isolated node is generated, an alert notification is automatically sent. Handle the issue at the earliest opportunity if you receive an alert notification that indicates an isolated node.

- Alert rule for node loops

  If a node serves as both the ancestor node and descendant node of another node, a node loop is formed. As a result, a dependency loop is formed. The nodes in a node loop cannot be automatically scheduled to run. After a node loop is formed, an alert notification is automatically sent. Handle the issue at the earliest opportunity if you receive an alert notification that indicates a node loop.

- Alert rule for global events

  You can configure alert rules for monitoring events that may affect the execution of nodes in the baselines of critical concern. You can specify the following information for a specific baseline: the maximum number of alerts to be handled, minimum intervals at which alert notifications are to be sent, notification methods, and alert contacts.

- Alert rule for global baselines

  You can specify the following information for a baseline of critical concern: the maximum number of alerts to be handled, minimum intervals at which alert notifications are to be sent, notification methods, and alert contacts.

- Baseline management

  You can use the baseline management feature to dynamically monitor lines-of-business and baselines.

  You can add the nodes of critical concern to a baseline. After the baseline is enabled, the nodes in the baseline are monitored. If the data output of the nodes in the baseline is affected, an alert is generated. For example, if an error is reported or the speed for running the nodes slows down, the data output is affected. Then, DataWorks calculates based on the historical duration required for running these nodes to learn whether the data output of the current day can be generated and generates alerts in advance. For more information, see Manage baselines.

- Automated O&M

  You can use the automated O&M feature to manage exclusive resource groups based on the configured custom alert rules for the resource groups when the specified conditions are met. For example, you can terminate the execution of node instances. For more information, see Automated O&M.

## Data O&M: Data quality

Data Quality of DataWorks can monitor the table data generated by auto triggered node instances, retroactive instances, and test instances.

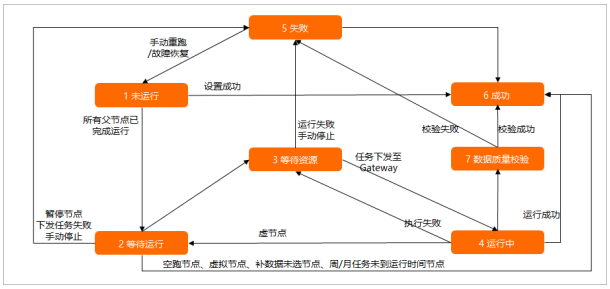
You can configure a data quality rule for a table generated by a node. Then, Data Quality matches the partitions in the table by using the partition expression configured for the table. If the node that is associated with the data quality rule is run, the data quality rule is triggered. You can specify the strength of the rule to determine whether a node exits when an error occurs. This limits the flow of dirty data. You can also subscribe to rules to receive corresponding alert notifications at the earliest opportunity. For more information, see Overview.

## Related information: Instance status

As mentioned in the Logic for running nodes section in this topic, the following conditions must be met before a node starts to run:

- The parent node of this node is successfully run.
- The scheduled point in time for running the node is reached.
- Sufficient scheduling resources are provided for running the node.
- This node is not frozen.
-

Two of the preceding conditions are related to node statuses. A node instance status model defines the six states of a node throughout the time to live (TTL). The following figure shows the logic of conversion between the states.

| No. | State | Icon | Conversion logic |
| --- | --- | --- | --- |
| 1 | Succeeded | |  |
| 2 | Not running | | |
| 3 | Failed | | |
| 4 | Running | | |
| 5 | Waiting | | |
| 6 | Suspended or frozen | | |



# 1.2. View the dashboard

The dashboard displays the overall operations and maintenance (O&M) information, including the metrics that require your special attention, overall running information about nodes, and trends on scheduling resources. It also displays the information about data integration, including the running status distribution and data synchronization progress of batch sync nodes and real-time sync nodes. The dashboard helps improve O&M efficiency.

## View the dashboard

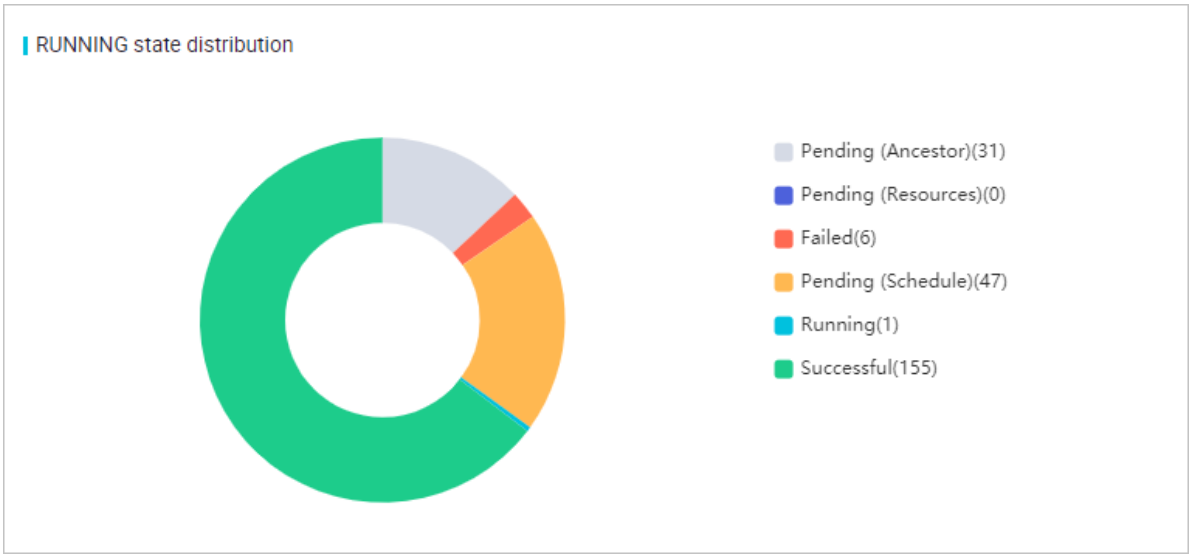
1. Log on to the DataWorks console.
2. In the left-side navigation pane, click **Workspaces**.
3. After you select the region in which the workspace that you want to manage resides, find the workspace and click **Data Analytics** in the Actions column.
4. Click the icon in the upper-left corner and choose **All Products > Task Operation > Operation Center**. The **Workbench Overview** tab of the **Overview** page appears.

## View the overall O&M information

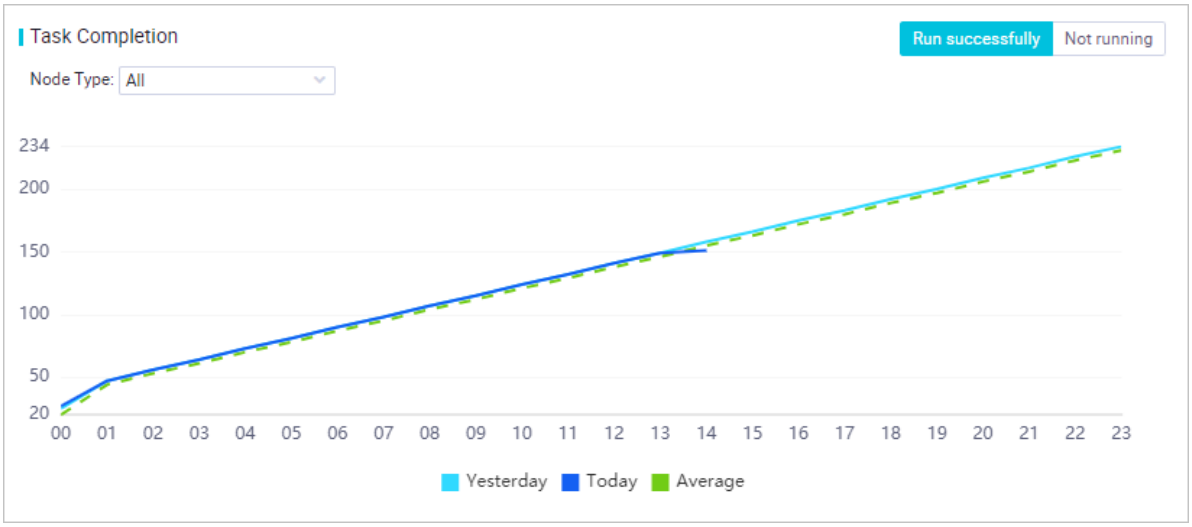
The Workbench Overview tab displays statistics on auto triggered nodes and auto triggered node instances. Other types of nodes or node instances are not included. You can view the following information on the **Overview** page:

- The **Focus on** section displays the following information:
  - The numbers of auto triggered node instances that require your special attention, including failed node instances, slow node instances, and node instances pending for resources. The statistics are collected on node instances whose data timestamp is the day before the current date. A node instance is considered a slow node instance when it meets the following conditions:
    - The node instance is running.
    - The running time of the node instance exceeds 30 minutes.
    - The running time of the node instance is at least 15 minutes longer than the average running time of the past 10 days.
  - The numbers of isolated nodes, paused nodes, and expired nodes.
    - An isolated node is a node that does not have an ancestor node. In this case, the node cannot be run. For example, if you change the output name for the parent node of a node, the dependency becomes invalid.
    - After a node is paused, the node no longer generates node instances and cannot be scheduled.
    - If a node is not triggered at the specified time, the node becomes expired.

  The statistics in this section are updated when you load the page. You can click a type of node or node instance to go to the details page and view the specific nodes or node instances. We recommend that you fix these nodes and node instances at the earliest opportunity to avoid impacts on your business.
- The **RUNNING state distribution** section displays the distribution of auto triggered node instances in different states. The data timestamp of these node instances is the day before the current date. The statistics in this section are updated when you load the page. You can click a sector in the pie chart to view the node instances in the specific state.



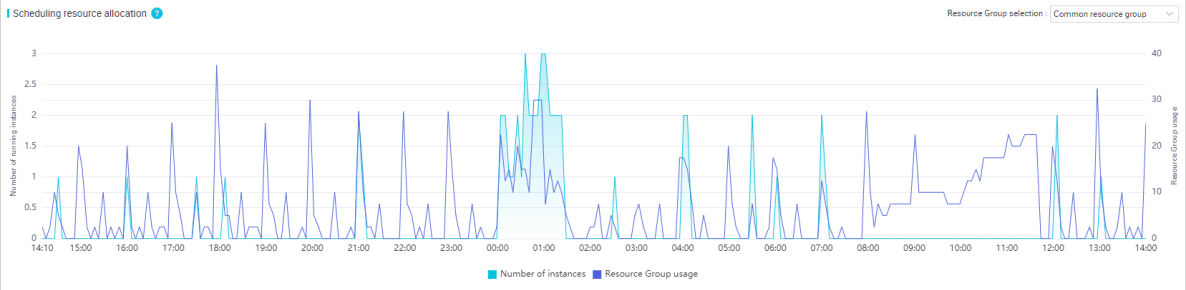

- The **Task Completion** section displays the completion status of node instances between 00:00 and 23:00 of the current date. You can view the number of the node instances that are completed or not run today and yesterday, and the historical average. You can also select a node type to view the status of specific node instances.




The line chart displays the numbers of auto triggered node instances that are completed today and yesterday, and the historical average. If the deviations among the three numbers are large, an exception occurred during a specific period of time. Further check and analysis are required.

> **Note** The statistical aggregation method used by the Operation Center service has been changed. Only node instances in the production environment are counted. Therefore, the line that represents the number of node instances completed today shows obvious fluctuations.

- The **Scheduling resource allocation** section displays the usage of a specific resource group and the number of node instances that were running at different time points in the last 24 hours. You can select a resource group from the **Resource Group selection** drop-down list in the upper-right corner.

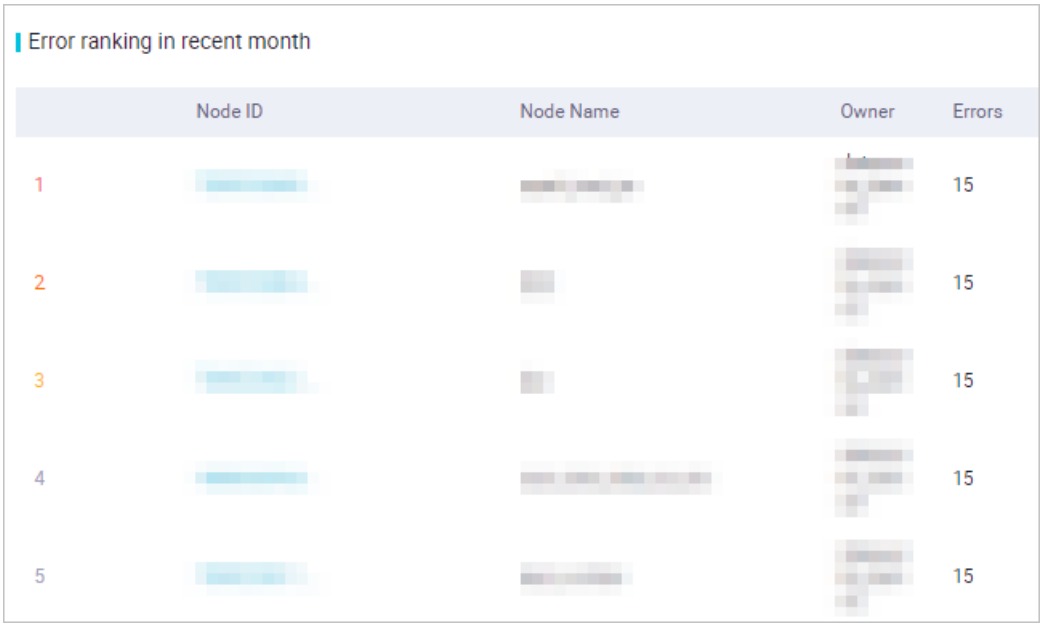



The **Number of instances** line shows the number of only the node instances in the current workspace. The **Resource Group usage** line shows the resource group usage occupied by node instances in all workspaces under the current Alibaba Cloud account.

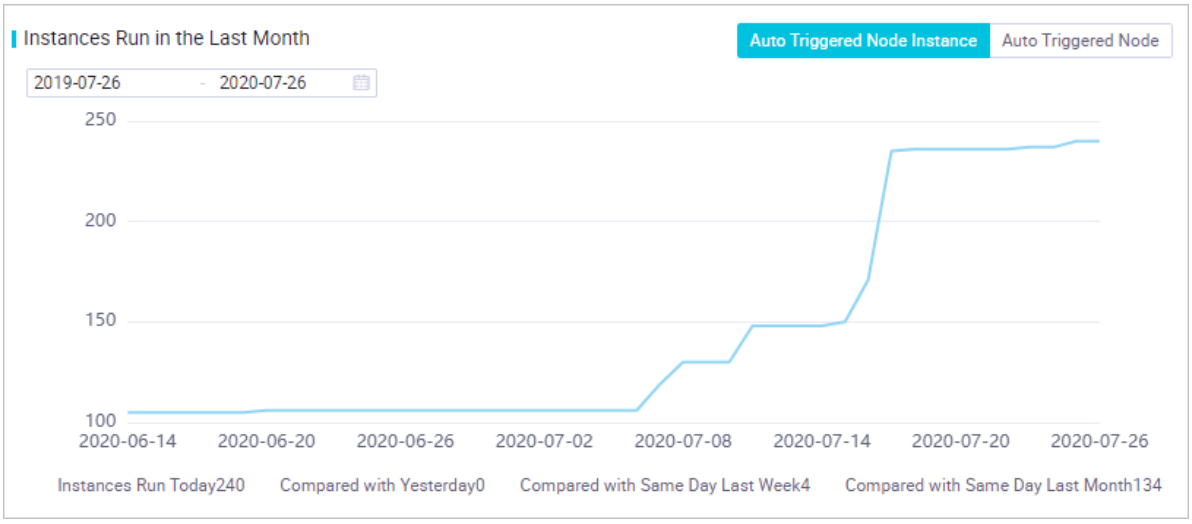
> **Note** The resource group usage occupied by data integration nodes is not counted.

- The **Runtime ranking** section ranks nodes based on their running time, time pending for resources, or excess running time. The statistics in this section are updated every day. Nodes that were completed on the day before the current date are ranked in this section.

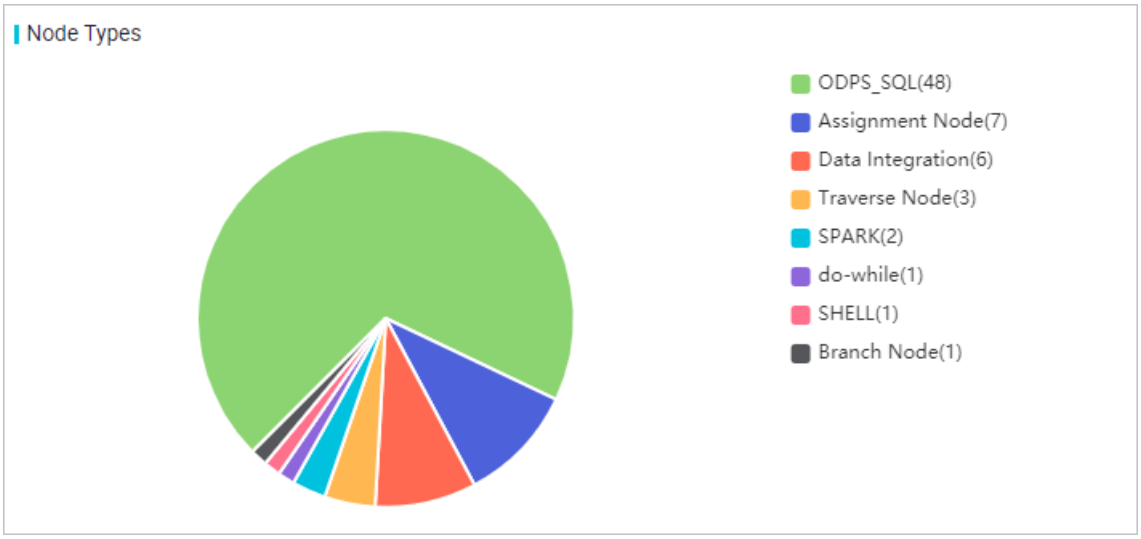

- The **Error ranking in recent month** section ranks nodes with the most errors in the last month and displays the top 10 nodes. The statistics in this section are updated every day. You can view the name, ID, number of occurred errors of each node.




- The **Instances Run in the Last Month** section displays the trends on the numbers of nodes and node instances that are scheduled in the specified time range in the production environment. The statistics in this section are updated every day. You can view the trend in the number of nodes and node instances that are scheduled in a time range as wide as one year.

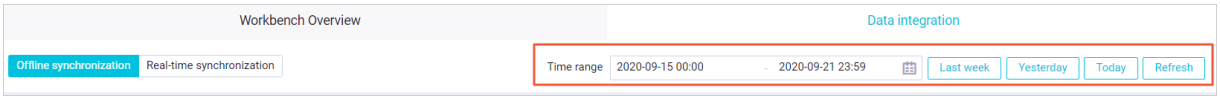



- The **Node Types** section displays the distribution of nodes in a pie chart. The statistics in this section are updated when you load the page. The pie chart displays a maximum of eight node types. If you have created more than eight types of nodes, specific node types are merged for display.



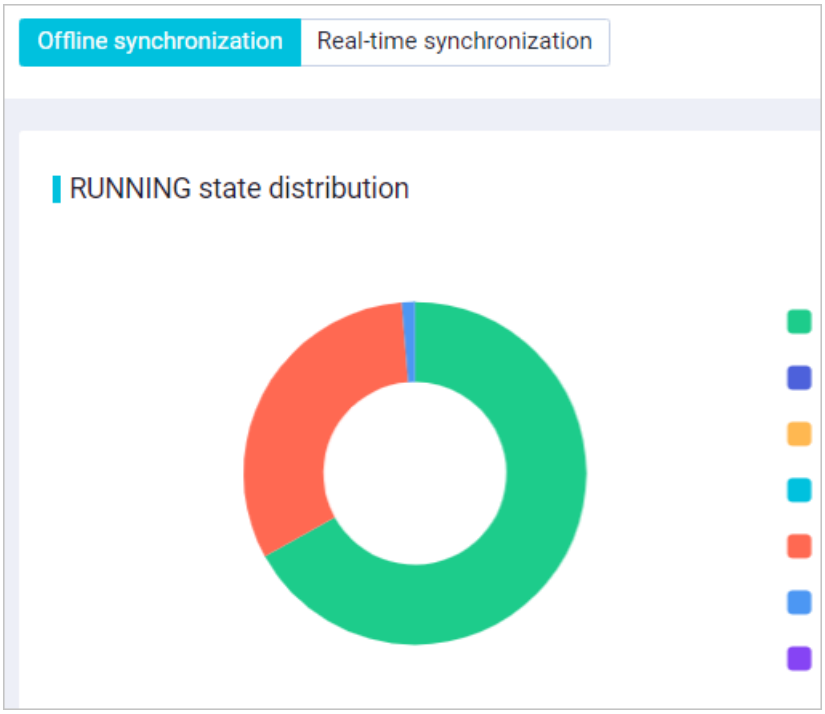

## View the O&M information about offline synchronization

On the **Overview** page, click the **Data integration** tab. The information about offline synchronization within a specific time range is displayed. You can specify the time range of the statistics in the upper-right corner.



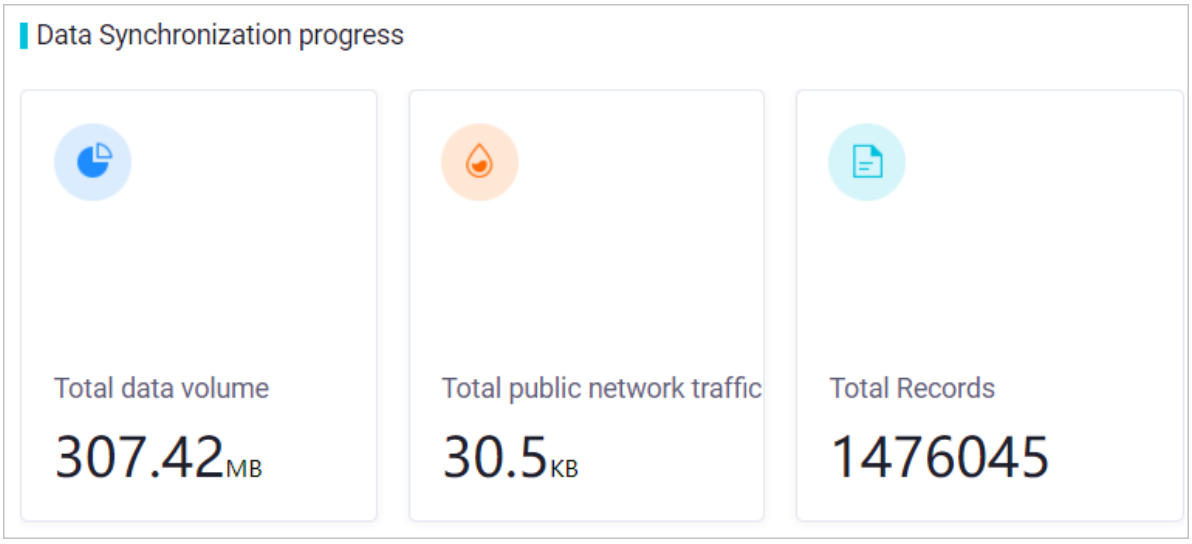

You can view the following information about batch sync nodes:

- The **RUNNING state distribution** section displays the distribution of auto triggered node instances in different states. The data timestamp of these node instances is in the specified time range. The statistics in this section are updated when you load the page. You can click a sector in the pie chart to view the node instances in the specific state.

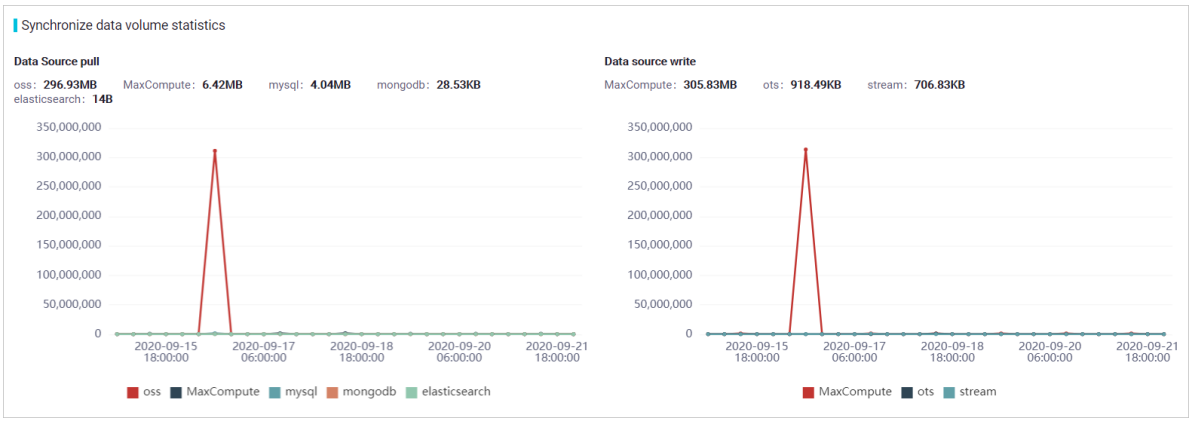



- The **Data Synchronization progress** section displays information about the data that is involved in

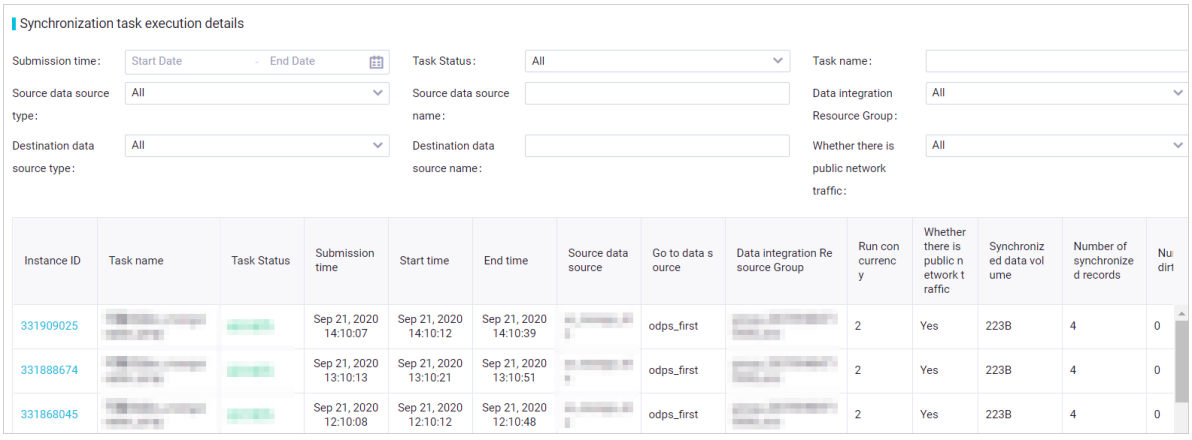
offline synchronization within the specified time range. The information includes the total amount of data, total amount of Internet traffic, and total number of records.




- The **Synchronize data volume statistics** section displays the curves of the data that is pulled from or written to different data stores within the specified time range.



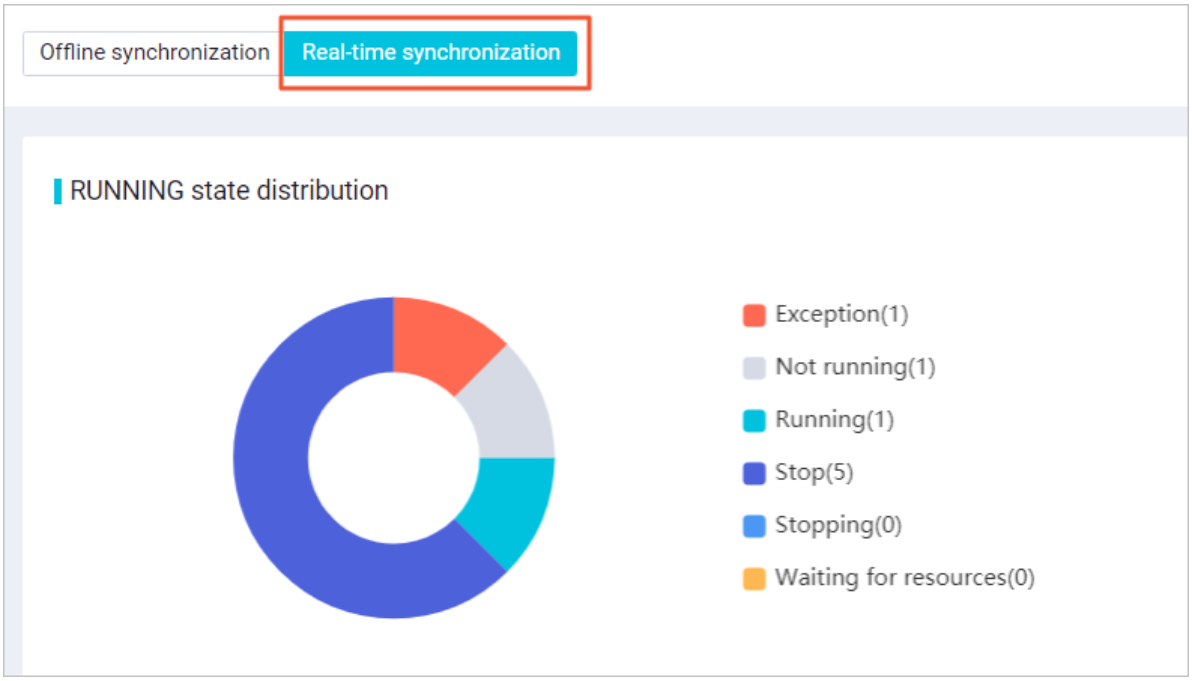

- The **Latest list Top10** section displays the latest 10 node instances that failed and the latest 10 node instances that are successful. The statistics allows you to have an overview of the latest node status.
- The **Synchronization task execution details** section allows you to search for a node instance by conditions such as the submission time, node status, and node name. You can also click the ID of a node instance to view the running details of the node instance.



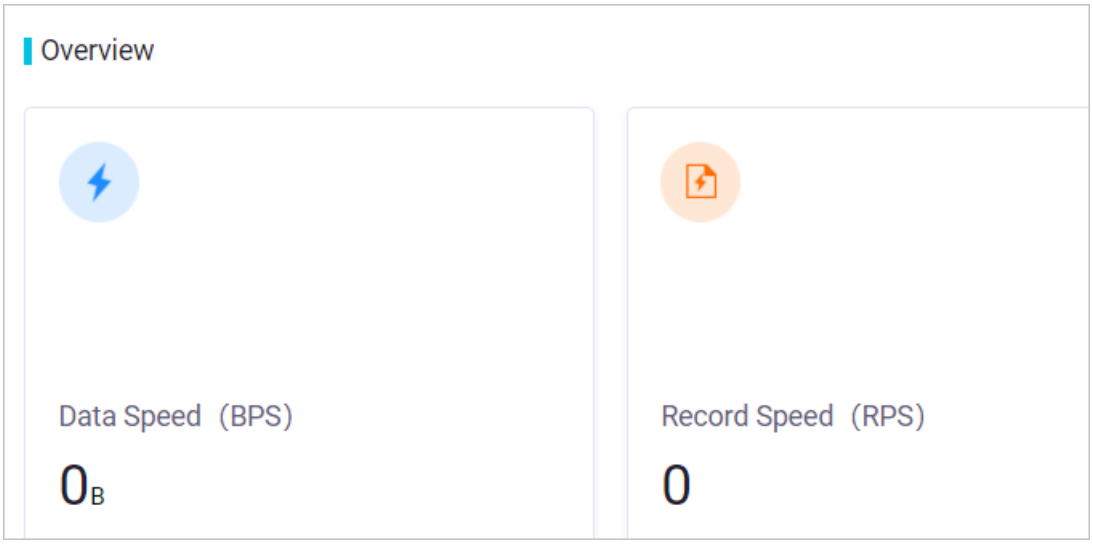
## View the O&M information about real-time synchronization

On the **Data integration** tab of the **Overview** page, click **Real-time synchronization**. You can view the following information about real-time sync nodes:

- The **RUNNING state distribution** section displays the distribution of real-time sync node instances. The statistics in this section are updated when you load the page. You can click a sector in the pie chart to view the node instances in the specific state.

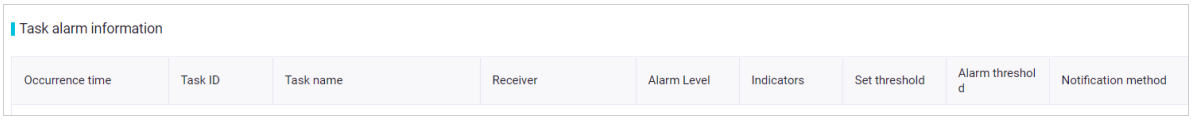



- The **Overview** section displays the total data transmission speed and total record transmission speed of all real-time sync nodes in the current workspace.

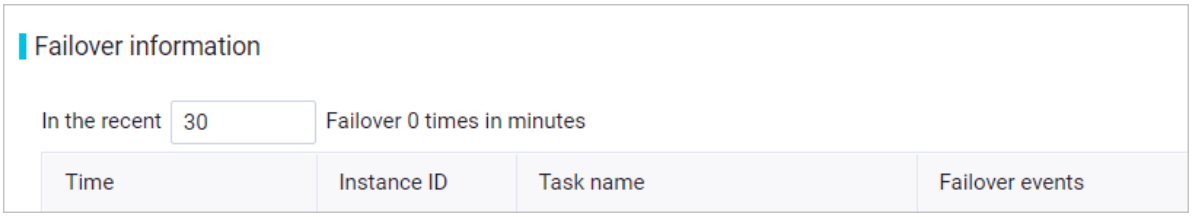



- The **Task delay Top10** section displays the top 10 nodes with the highest latency. This section allows you to find the nodes with high latency.

- The **Task alarm information** section displays information about the latest alerts. This section allows you to know exceptions with efficiency.




- The **Failover information** section displays information about failovers within a specified period. This section allows you to have an overview of node failovers.




# 1.3. Perform O&M on real-time nodes

## 1.3.1. Manage real-time computing nodes

The Stream Task page of the DataWorks console displays all real-time computing nodes and the basic information of the nodes. You can view the details about nodes on this page. You can also configure alert rules for nodes. This helps you quickly locate and fix errors if errors occur on a node.

### Limits

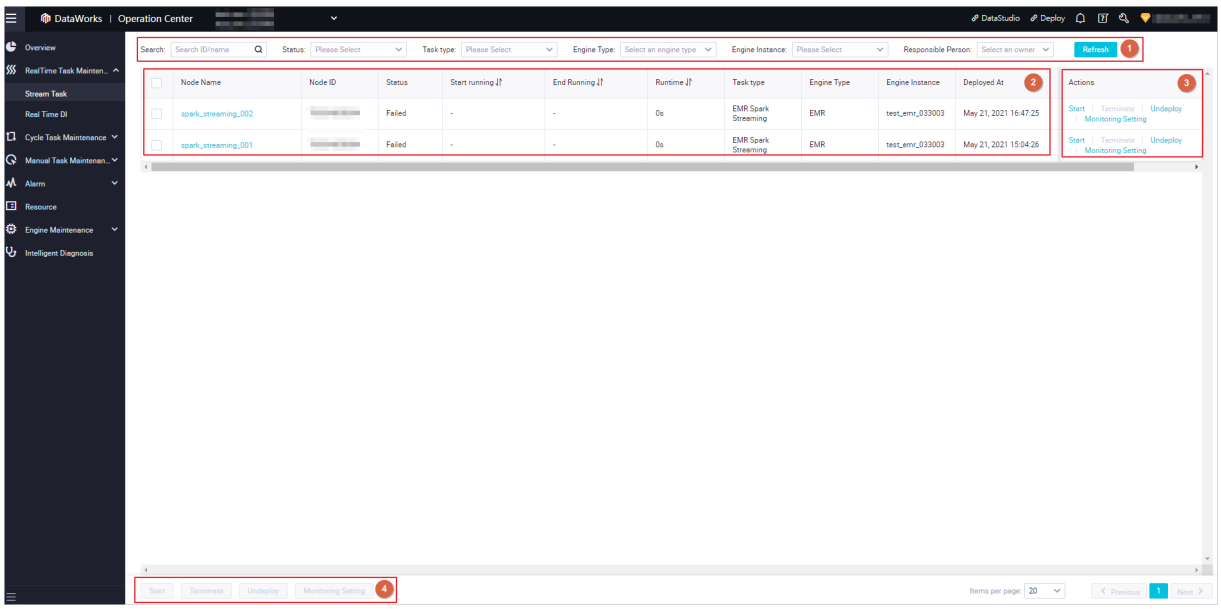
DataWorks supports O&M only for EMR Spark Streaming and EMR Streaming SQL nodes.

### Go to the Stream Task page

1. Log on to the DataWorks console.
2. In the left-side navigation pane, click **Workspaces**.
3. After you select the region in which the workspace that you want to manage resides, find the workspace and click **Data Analytics** in the Actions column.
4. On the **DataStudio** page, click the icon in the upper-left corner and choose **All Products > Task Operation > Operation Center**.
5. On the Operation Center page, choose **RealTime Task Maintenance > Stream Task** to go to the Stream Task page.

### View the real-time computing nodes in the node list

The Stream Task page displays real-time computing nodes in the production environment. You can view the basic information of, start, stop, or undeploy a real-time computing node. You can also configure alert rules for a real-time computing node.



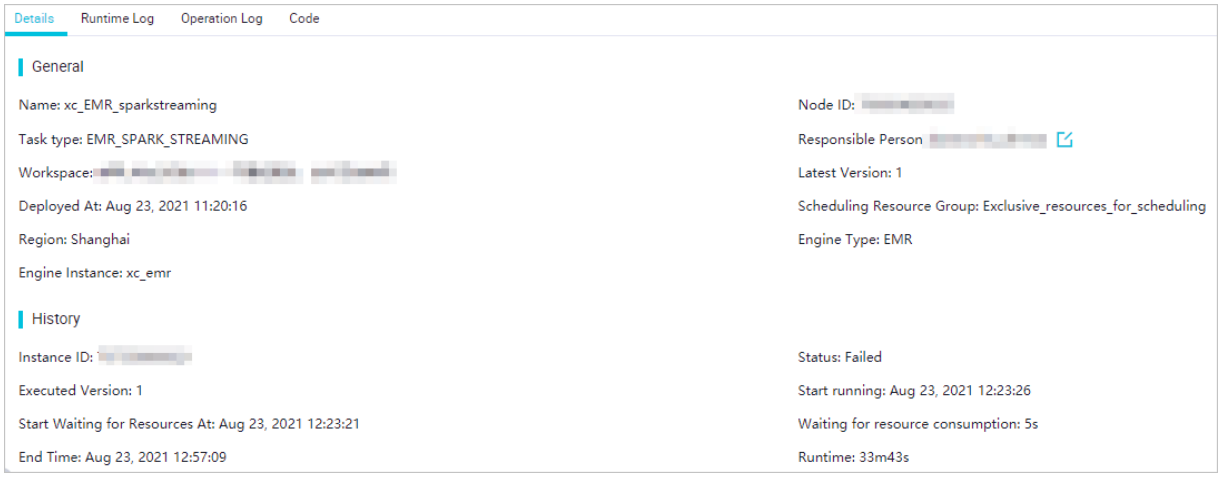

| Section | Parameter and description |
| --- | --- |
| 1 | In this section, you can search for real-time computing nodes by node ID or node name. You can also specify **Status**, **Task type**, **Engine Type**, **Engine Instance**, and **Responsible Person** to perform the operation.<br>**Note**<br>• If you search for nodes by node name or node ID, the search result is affected by other filter conditions that you specified. Only nodes that meet all the filter conditions that you specify are displayed.<br>• If you search for nodes by node name, fuzzy match is supported. After you enter a keyword, all real-time computing nodes whose names contain the keyword are displayed. |

| Section | Parameter and description |
| --- | --- |
| 2 | In this section, you can view the basic information and details of a real-time computing node.<br>• **Node Name**: the name of the real-time computing node. You can click the node name to go to the details page of the node.<br>• **Node ID**: the ID of the real-time computing node.<br>• **Status**: the status of the real-time computing node. The node can be in one of the following states: **Not Running**, **Generating**, **Pending (Resources)**, **Starting**, **Running**, **Stopped**, **Failed**, **Restarting**, and **Undeploying**.<br>• **Start running**: the time when the real-time computing node started to run.<br>• **End Running**: the time when the running of the node was complete.<br>• **Runtime**: the running duration of the real-time computing node. Unit: seconds.<br>• **Task type**: the type of the real-time computing node.<br>Note<br>DataWorks supports O&M only for EMR Spark Streaming and EMR Streaming SQL nodes.<br>• **Engine Type**: the type of the engine used to run the real-time computing node. DataWorks supports O&M only for real-time computing nodes that are run by using the EMR engine.<br>• **Engine Instance**: the name of the EMR compute engine instance that is associated with the workspace to which the real-time computing node belongs.<br>• **Deployed At**: the time when the real-time computing node was deployed. The time is in the yyyy-MM-dd HH:mm:ss format.<br>• **Responsible Person**: the owner of the workspace to which the real-time computing node belongs.<br>Note You can rank all real-time computing nodes in ascending or descending order of **Starting running**, **End Running**, or **Runtime**. |
| 3 | In this section, you can perform the following operations on a real-time computing node:<br>• **Start**: Start the node.<br>• **Terminate**: Terminate the node.<br>• **Undeploy**: Undeploy the node.<br>• **Monitoring Setting**: Configure alert rules for the node. If the node fails to run, the system sends an alert notification to the specified alert contact by **text message**, **email**, **mobile phone**, **DingTalk chatbot**, or **webhook URL**.<br>For more information about how to configure an email address and a mobile phone number for the alert contact, see What can I do if I fail to receive SMS or email alert notifications after I configure alert notifications in Operation Center?. For more information about how to configure a DingTalk chatbot and obtain a webhook URL, see Send alert notifications to a DingTalk group. |

| Section | Parameter and description |
| --- | --- |
| 4 | In this section, you can perform an operation on multiple real-time computing nodes at a time. You can select multiple nodes and click **Start**, **Terminate**, **Undeploy**, or **Monitoring Setting** to perform the operation on these nodes. |

## View the details of a real-time computing node

Find the real-time computing node that you want to query in the node list and click the node name to go to the details panel of the node. In the node details panel, you can view the details, operational logs, operation logs, and code of the node on the **Details**, **Runtime Log**, **Operations Log**, and **Code** tabs. This allows you to quickly locate operations performed on the node and operation records, obtain details about the errors reported for the node, and fix the errors.




# 1.3.2. Manage real-time synchronization nodes

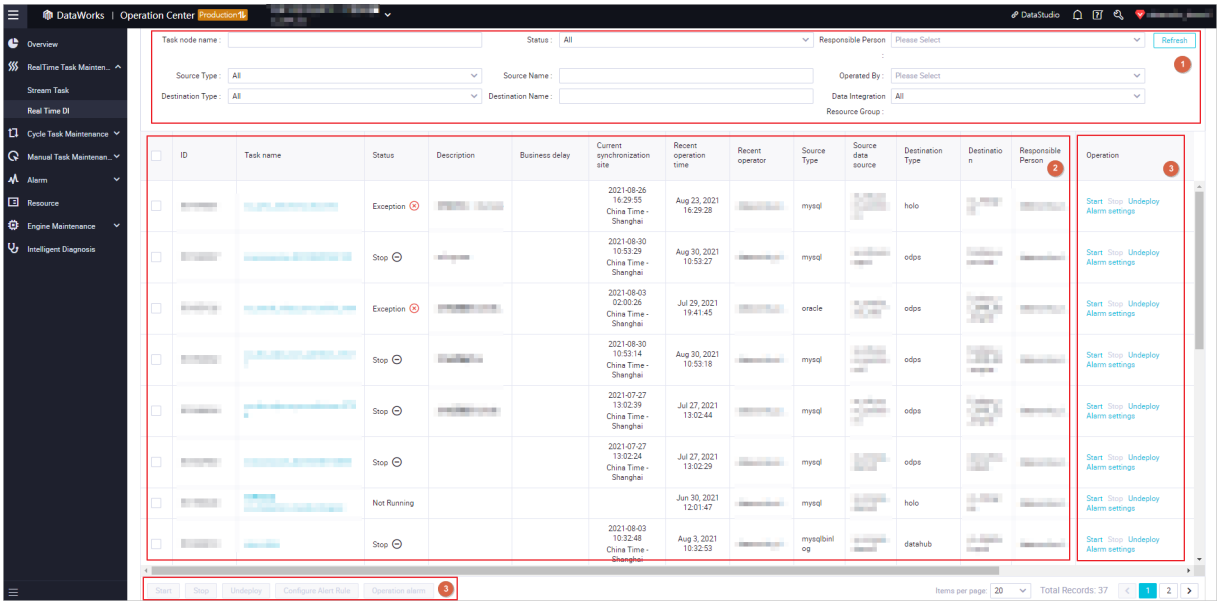
The Real Time DI page of the DataWorks console displays all real-time synchronization nodes that are committed to and run by the scheduling system.

## Go to the Real Time DI page

1. Log on to the DataWorks console.
2. In the left-side navigation pane, click **Workspaces**.
3. After you select the region in which the workspace that you want to manage resides, find the workspace and click **Data Analytics** in the Actions column.
4. On the **DataStudio** page, click the icon in the upper-left corner and choose **All Products > Task Operation > Operation Center**.
5. On the Operation Center page, click the icon next to **RealTime Task Maintenance** in the left-side navigation pane.
6. Select **Real Time DI** from the drop-down list. The Real Time DI page appears.

## View real-time synchronization nodes in the node list

The node list displays real-time synchronization nodes that are run by the scheduling system. You can manage these nodes and perform O&M operations on these nodes. For example, you can view the basic information, properties, and running information of a real-time synchronization node. You can also start or undeploy a real-time synchronization node.

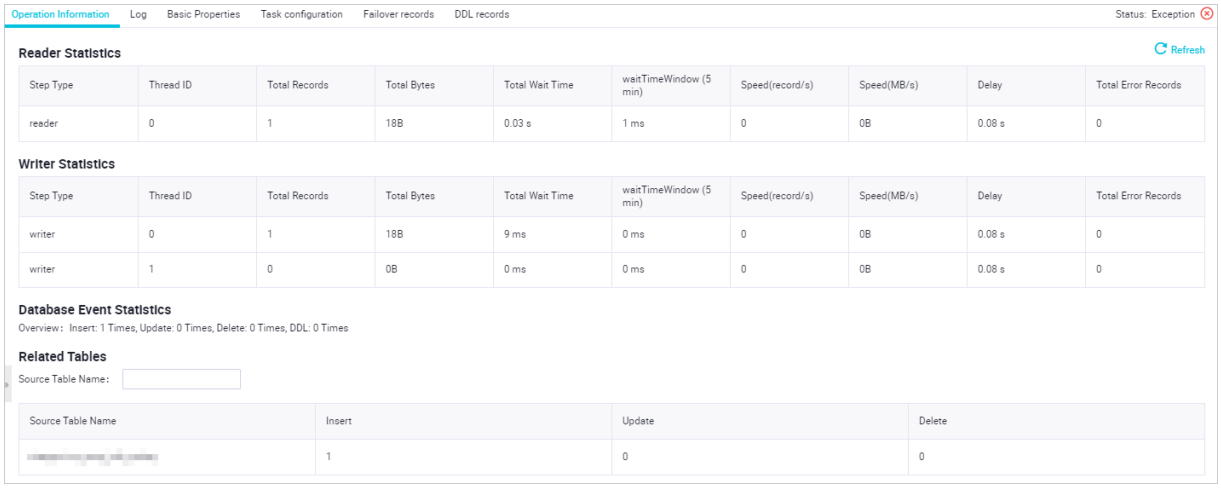



| Section | Description |
| --- | --- |
| 1 | In this section, you can specify filter conditions to search for real-time synchronization nodes.<br>You can specify the following filter conditions to search for nodes: **Task node name**, **Status**, **Responsible Person**, **Source Type**, **Source Name**, **Operated By**, **Destination Type**, **Destination Name**, and **Data Integration Resource Group**.<br>**Note** When you search for nodes by node name, the search result is affected by other filter conditions that you specified. Only the nodes that meet all filter conditions are displayed. |

| Section | Description |
| --- | --- |
| 2 | In this section, you can view the following information about the real-time synchronization node that you want to query:<br>• **ID**: the ID of the node.<br>• **Task name**: the name of the node. You can click the node name to go to the details page of the node.<br>• **Status**: the status of the node. The node can be in one of the following states: Running, Not Running, Waiting for Resources, Exception, and Stop or Stopping.<br>• **Description**: the description of the node.<br>• **Business delay**: the period of time between the current time and the offset from which incremental data starts to be synchronized.<br>• **Current synchronization site**: the offset at which incremental data is being synchronized.<br>• **Recent operation time**: the time when the node was last modified.<br>• **Recent operator**: the user who last modified the node.<br>• **Source Type**: the type of the source.<br>• **Source data source**: the name of the source.<br>• **Destination Type**: the type of the destination.<br>• **Destination**: the name of the destination.<br>• **Responsible Person**: the owner of the workspace to which the node belongs.<br>• **Current Start Site**: the offset from which incremental data starts to be synchronized.<br>• **Data Read Speed (Bytes per Second)**: the speed at which the data is read.<br>• **Recording Speed**: the speed at which the read data is written to logs.<br>• **Data Integration Resource Group**: the resource group that is used to run the node. |
| 3 | In this section, you can perform the following operations on multiple nodes at a time:<br>• **Start**: Start the nodes.<br>• **Stop**: Stop the running nodes.<br>• **Undeploy**: Undeploy the nodes that are not running, abnormal, or stopped.<br>• **Configure Alert Rule**: Configure alert rules for the nodes.<br>• **Operation alarm**: Delete, enable, or disable alert rules and modify alert rules based on alert metric types. |

## View the details about a real-time synchronization node

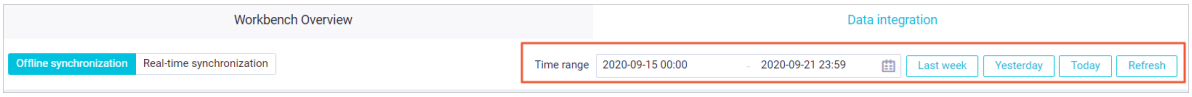
Find the real-time synchronization node that you want to query and click the node name to go to the details panel of the node. In the details panel of the node, you can view the details about the node on the **Operation Information**, **Log**, **Basic properties**, **Task configuration**, **Failover records**, and **DDL records** tabs.




## Perform operations on a real-time synchronization node

- Start a real-time synchronization node

  Find the node that you want to start and click **Start** in the **Operation** column. Then, the Start dialog box appears. After you configure the parameters in the dialog box, click **Confirm**. Then, the system starts to run the node.

  If you select **Reset site**, you must configure the **Start time point** and **Time zone** parameters.

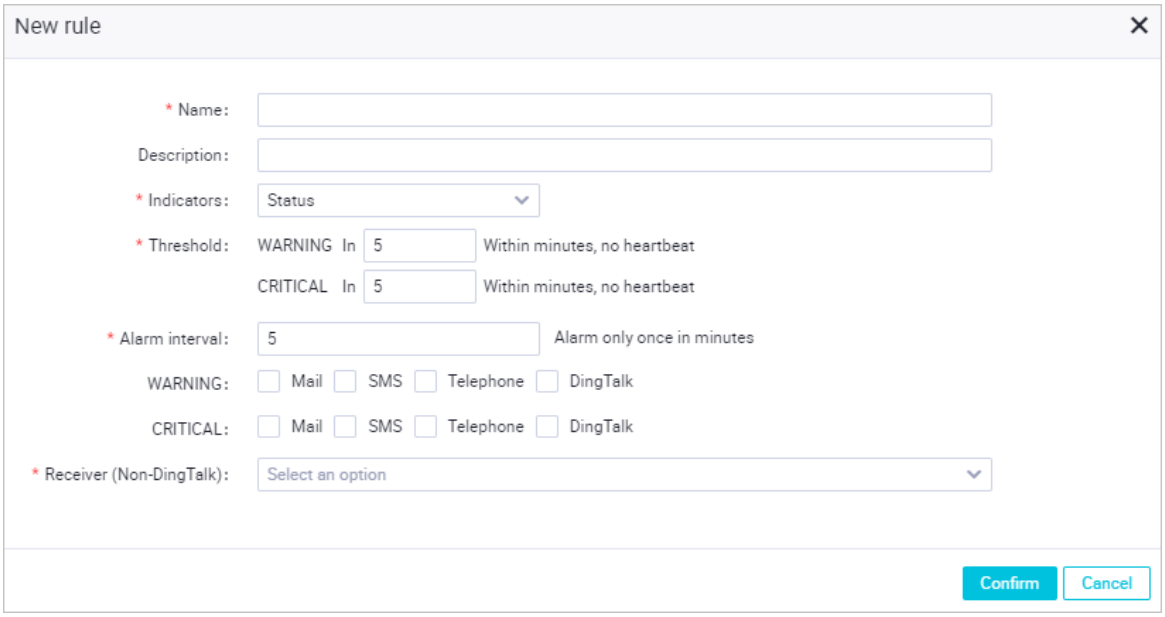
- Stop a real-time synchronization node

  Find the real-time synchronization node that you want to stop and click **Stop** in the Operation column. Then, click **Stop** in the message that appears to stop the node.

- Undeploy a real-time synchronization node

  Find the node that you want to undeploy and click **Undeploy** in the Operation column. Then, click **Undeploy** in the message that appears to undeploy the node.

- View the alert rule settings of a real-time synchronization node
  - Find the node whose alert rule settings you want to view and click Alarm settings in the Operation column. Then, the Alarm event tab appears. On this tab, you can specify **Occurrence time**, **Alarm Level**, and **Rules** to view the alarm events of the node.
  - Click **Alarm rules** to go to the Alarm rules tab. On this tab, you can view all the alert rules that are created for the node. You can view the metrics and status of an alert rule. You can also modify, pause, delete or test an alert rule. If you want to create an alert rule, click **New rule** and configure the following parameters.




| Parameter | Description | Required |
| --- | --- | --- |
| **Name** | The name of the alert rule. | Yes |
| **Description** | The description of the alert rule. | No |
| **Indicators** | The type of metric that triggers an alert. The Indicators parameter and the Threshold parameter must be used in pairs. The value of the Threshold parameter varies based on that of the Indicators parameter. Valid values of Indicators: Status, Business delay, Failover, Dirty Data, and DDL statements are not supported. | Yes |

| Parameter | Description | Required |
| --- | --- | --- |
| **Threshold** | The thresholds for the metrics that are specified by the Indicators parameter.<br>■ If you set the Indicators parameter to **Status**, you must specify the interval at which alerts are triggered.<br>■ If you set the Indicators parameter to **Business delay**, you must specify a duration for business delay and another duration during which an alert can last after the duration for business delay elapses.<br>■ If you set the Indicators parameter to **Failover**, you must specify a duration and the maximum number of failovers that can be performed within the duration before an alert is triggered.<br>■ If you set the Indicators parameter to **Dirty Data**, you must specify a duration and the maximum number of dirty data records that are allowed within the duration before an alert is triggered.<br>■ If you set the Indicators parameter to **DDL statements are not supported**, you need only to select an alert level. | Yes |
| **Alarm interval** | The minimum interval at which alerts are reported. Default value: 5. Minimum value: 1. Unit: minutes. | Yes |
| | The method used to receive a notification for a WARNING-level alert. Valid values: **Mail**, **SMS**, **Telephone**, and **DingTalk**. | |

| Parameter | Description | Required |
| --- | --- | --- |
| WARNING | Note<br>■ Mail: If you want to use a RAM user to receive the notification, you must use an Alibaba Cloud account to add the email address of the RAM user in user information.<br>■ SMS: If you want to use a RAM user to receive the notification, you must use an Alibaba Cloud account to add the mobile phone number of the RAM user in user information.<br>■ Telephone: If you want to use a RAM user to receive the notification, you must use an Alibaba Cloud account to add the mobile phone number of the RAM user in user information.<br>■ DingTalk: If you want to receive the notification by using a DingTalk chatbot, you must configure a DingTalk chatbot for your DingTalk group and add the keyword *DataWorks* for the DingTalk chatbot. | No |

| Parameter | Description | Required |
| --- | --- | --- |
| **CRITICAL** | The method used to receive a notification for a CRITICAL-level alert. Valid values: **Mail**, **SMS**, **Telephone**, and **DingTalk**. | No |
| **Receiver (Non-DingTalk)** | The alert contact to which alert notifications are sent. | Yes |

# 1.4. Auto triggered node O&M

## 1.4.1. View auto triggered nodes

Auto triggered nodes are automatically run as scheduled after they are committed to the scheduling system. You can view the details of an auto triggered node from the DAG of the node or the auto triggered node list.

### Context

By default, the **Cycle Task** page displays the nodes in all the workflows that are created by using the current Alibaba Cloud account.

> **Notice** Do not perform operations on the **projectname_root** node, which is the root node of the workspace. All the instances of auto triggered nodes depend on this node. If this node is frozen, the instances of auto triggered nodes cannot be run.

Instances can be generated in the following ways:

- **Next Day**:
  - If a node is committed and deployed before 23:30, instances are generated from the next day.
  - If a node is committed and deployed after 23:30, instances are generated from the day after the next day.
- **Immediately After Deployment**: Instances are generated immediately after a node is committed and deployed. However, if the node is committed and deployed after 23:30, instances cannot be immediately generated. For more information, see Configure immediate instance generation for a node.

If you cannot find the instances generated by your node, perform operations by following the instructions in What do I do if I cannot find the instance of a node?.

The **Cycle Task** page displays the auto triggered nodes that are committed to the scheduling system in a list or in directed acyclic graphs (DAGs).

- Manage auto triggered nodes

  You can specify filter conditions to search for your desired node. Then, you can view the basic information of the node and perform operations such as testing the node and generating retroactive data for the node. For more information, see Manage auto triggered nodes.

- DAG

A DAG presents the ancestor and descendant nodes of a node. You can view the details and lineage of a node and test it from the DAG of the node. A DAG provides the node aggregation, upstream analysis, and downstream analysis features. The node aggregation feature allows you to aggregate nodes in a DAG from the dimensions such as workspace, owner, and priority, and view the total number of nodes from your required dimension. The remaining two features allow you to analyze the ancestor and descendant nodes of a specific node. This way, you can quickly locate the ancestor node that blocks the running of the node and view the number of the descendant nodes of the node based on the analysis results. These features help you better understand the running status of all nodes. For more information, see Manage auto triggered nodes in a DAG.

## Limits

DAGs have the following limits:

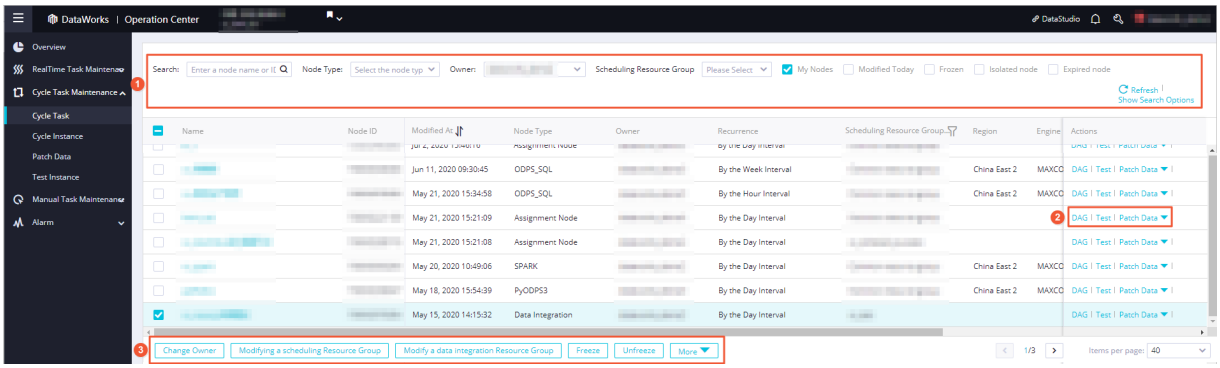
- Only users of the DataWorks Standard Edition or a more advanced edition can use the node aggregation, upstream analysis, and downstream analysis features provided by DAGs.
- Users of the DataWorks Basic Edition or Standard Edition can use the node aggregation, upstream analysis, and downstream analysis features provided by DAGs on a trial basis free of charge as of May 31, 2021. From June 1, 2021, users must update their DataWorks to the Professional Edition before they can use these features. For more information about DataWorks editions, see DataWorks advanced editions.
- You can use the node aggregation, upstream analysis, and downstream analysis features of the DAGs of auto triggered nodes that are deployed only in the China (Shenzhen) region.

## Go to the Cycle Task page

1. Go to the **DataStudio** page.
   i. Log on to the DataWorks console.
   ii. In the left-side navigation pane, click **Workspaces**.
   iii. In the top navigation bar, select the region where the required workspace resides, find the workspace, and then click **Data Analytics**.
2. On the **DataStudio** page, click the icon in the upper-left corner and choose **All Products > Operation Center**.
3. In the left-side navigation pane, choose **Cycle Task Maintenance > Cycle Task**. On the **Cycle Task** page, view the auto triggered nodes or the DAGs of the auto-triggered nodes.

## Manage auto triggered nodes

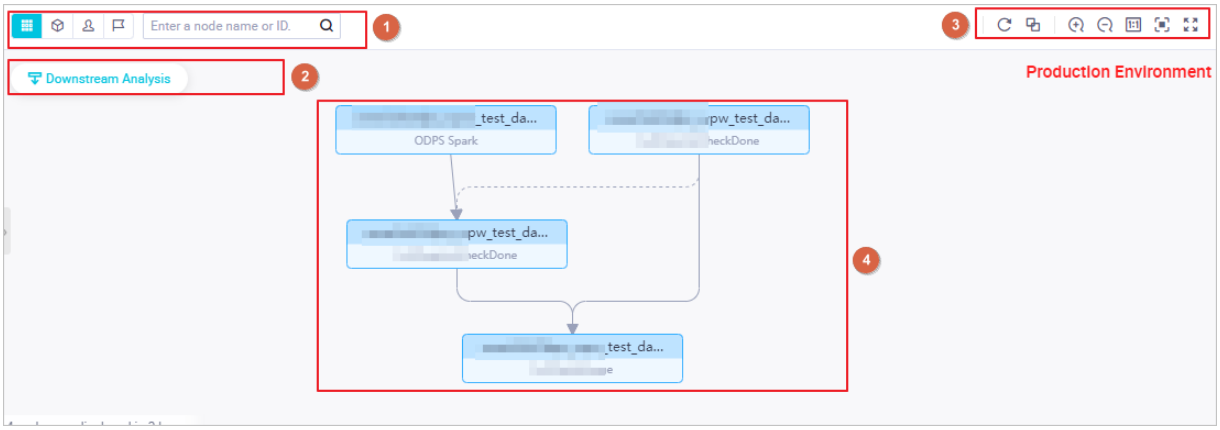
In the middle of the page, click the icon to show the node list.



| Operation | Description |
| --- | --- |
| **Filter** | Specify filter conditions to search for your desired node in the section marked with 1 in the preceding figure.<br>You can search for nodes by node name or node ID. You can also specify the following conditions to search for your desired node: **Solution**, **Workflow**, **Node Type**, **Owner**, **Region**, **Scheduling Resource Group**, **Engine Type**, **Engine Instance**, **Baseline**, **My Nodes**, **Modified Today**, **Frozen Nodes**, **Isolated Nodes**, and **Expired Nodes**.<br>**Note** If you search for nodes by node name, the search results are affected by other filter conditions you specified. Only nodes that meet both the specified search condition and other filter conditions are returned in the search result. |
| **DAG** | Click **DAG** in the Actions column that corresponds to a node to view the DAG of the node. You can view the node information, such as properties, operation logs, and code, in the DAG. |
| **Test** | Click **Test** in the Actions column that corresponds to a node to test the node. For more information, see Manage test instances. |
| **Patch Data** | Click **Patch Data** in the Actions column that corresponds to a node to generate retroactive data for the node. For more information, see Generate retroactive data and manage retroactive instances. |
| **More** | Click **More** in the Actions column that corresponds to a node to perform the following operations on the node:<br>• Select **Freeze** to freeze the node. After the node is frozen, the system generates instances for the node but does not run the instances of the node and its dependent descendant instances.<br>• Select **Unfreeze** to unfreeze the node. After the node is unfrozen, the system normally runs the instances of the node and its descendant instances.<br>• Select **View Instances** to view the instances of the node.<br>• Select **Configure Alert Rule** to configure alert rules for the node.<br>• Select **Change Owner** to change the owner of the node.<br>• Select **Add to Baseline** to add the node to a baseline.<br>• Select **Modify Scheduling Resource Group** to change the resource group that is used to run the node. You can perform this operation if multiple resource groups exist in the workspace. In the **Modify Scheduling Resource Group** dialog box, select a resource group from the **New Resource Group** drop-down list and click **OK**.<br>• Select **Configure Data Quality Rules** to configure rules for monitoring the data quality of the node.<br>• Select **View Lineage** to view the lineage of the node.<br>• Select **View Node Details** to go to the **Node Information** page. On this page, you can view the node information on the **Ancestor Nodes** and **Descendent Nodes** tabs. |
| **Batch operations** | In the section marked with 3 in the preceding figure, you can perform the following operations on the nodes: **Change Owner**, **Modify Scheduling Resource Group**, **Modify Data Integration Resource Group**, **Freeze**, **Unfreeze**, **Configure Alert Rule**, **Add to Baseline**, and **Undeploy**. |

## Manage auto triggered nodes in a DAG

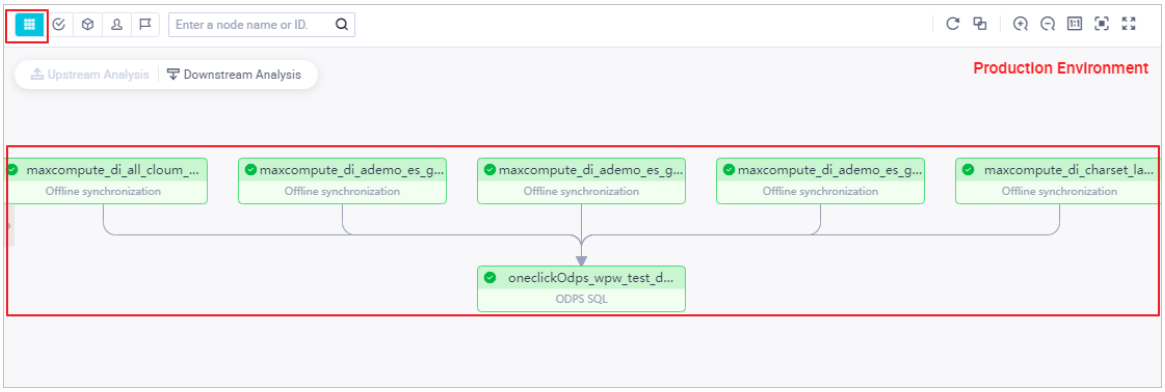
Click **DAG** in the Actions column that corresponds to a node to view the DAG of the node. You can perform the following operations in a DAG:



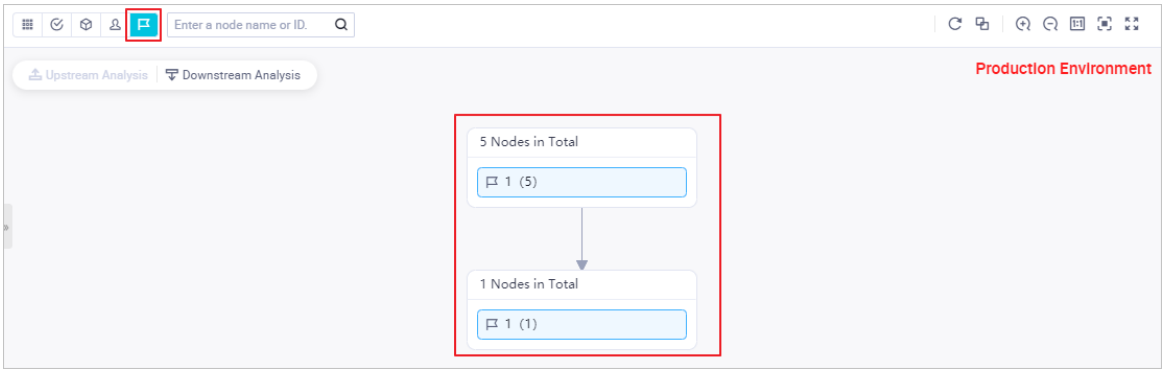

- Aggregate nodes

  If an auto triggered node has multiple ancestor and descendant nodes or the ancestor and descendant nodes are distributed at multiple levels, you can aggregate the nodes. The nodes can be aggregated from dimensions such as node status, workspace, owner, and priority. Then, you can view the number of nodes from your required dimension. This allows you to understand the number of nodes from different dimensions and helps the system run the nodes. The following figures show the node distribution when the ancestor and descendant nodes of an auto triggered node are **not aggregated** or are **aggregated by priority**.

  - The following figure shows the node distribution when the ancestor and descendant nodes of an auto triggered node are not aggregated.



- The following figure shows the node distribution when the ancestor and descendant nodes of an auto triggered node are aggregated by priority. From the figure, you can quickly understand that the current auto triggered node has *six* descendant nodes whose priorities are *1*.

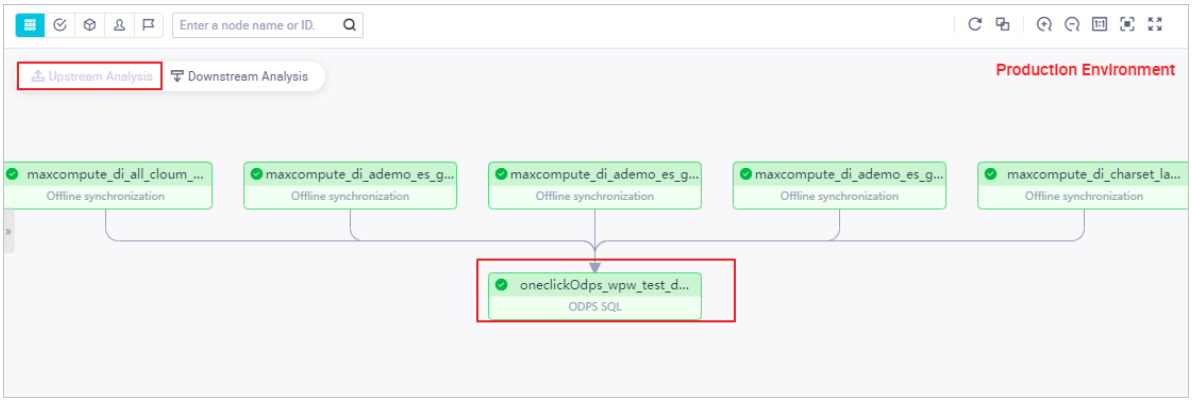



- Analyze ancestor nodes

  In most cases, an auto triggered node has both upstream and downstream relationships. If an auto triggered node is not run for a long time, you can analyze the ancestor nodes of the node. You can view the ancestor node that blocks the running of the node in the DAG of the node, and quickly locate and troubleshoot the issue. This improves the running efficiency of the node.

  > **Note** You can **analyze the ancestor nodes** of only the auto triggered nodes that are not run.

  The following figure shows how to analyze the ancestor nodes of an auto triggered node. For example, the **2_** node is not run for a long time. In this case, you can select the node and click **Upstream Analysis** in the upper-left corner to analyze the ancestor nodes of the node.




  The analysis results show that the ancestor nodes that block the running of the 2_ node are the **table data synchronization** and **metric statistics** nodes, as shown in the following figure. Then, you can quickly troubleshoot the issue based on the analysis result.

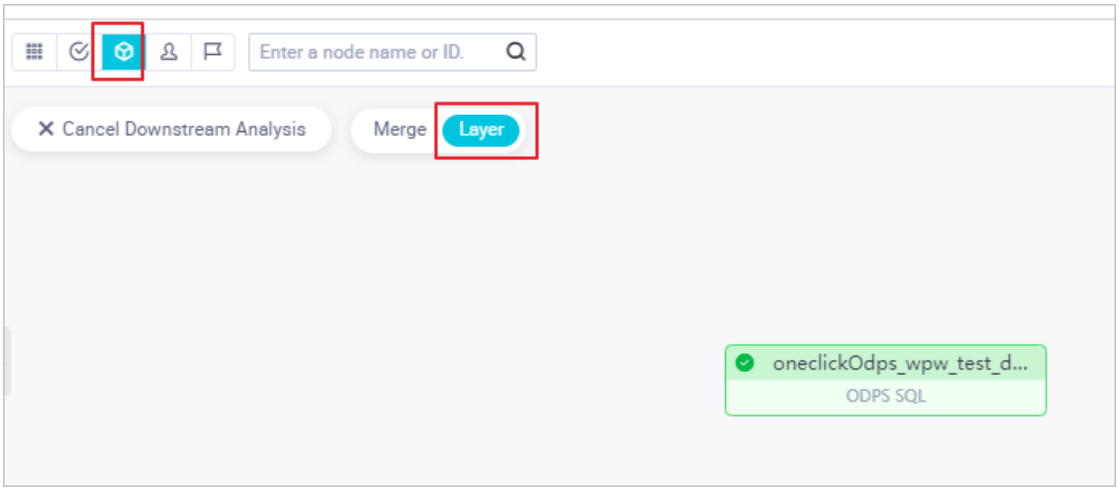
- Analyze descendant nodes

  If an auto triggered node has multiple descendant nodes or the descendant nodes of an auto triggered node are distributed at multiple levels, you can analyze the descendant nodes of the auto triggered node. You can aggregate the descendant nodes by workspace, owner, or priority. Then, you can view the number of nodes at different levels or the total number of nodes at all levels from your required dimension.

> **Note**
>
> - By default, the descendant nodes of an auto triggered node are aggregated by owner. The system calculates the total number of nodes at all levels from the owner dimension.
> - If you **analyze the descendant nodes** of an auto triggered node, the analysis results are displayed by **level**, and a maximum of six levels of nodes can be displayed. If you want to view more levels of nodes, click **Continue Analysis** in the upper-left corner.

In the following example, the descendant nodes of the **tag** node are analyzed. The following figures show the analysis results that are displayed by using different methods.

- The descendant nodes of the tag node are **aggregated based on the workspaces to which the descendant nodes belong**, and the analysis results are presented by **level**. This way, the number of the descendant nodes in different workspaces is displayed at different levels.



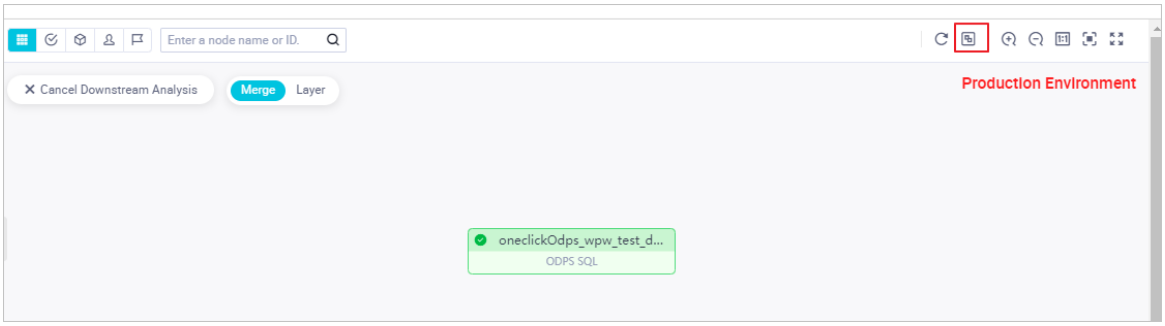

- The descendant nodes of the tag node are **aggregated based on the workspaces to which the descendant nodes belong**, and the analysis results are presented by using the **merging** method. This way, all the descendant nodes are placed at the same level, and the number of the descendant nodes that belong to different workspaces is displayed.


- Select a display pattern for a DAG

You can click the icons in the upper-right corner of a DAG panel to adjust the display pattern of the DAG based on your business requirements. For example, you can click **Toggle Full Screen View** or **Fit Screen** to perform the operation.

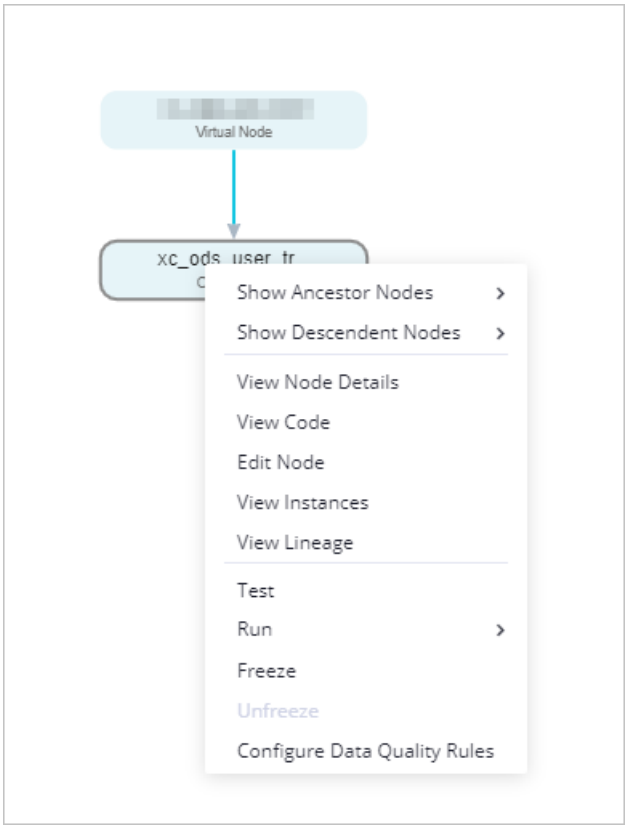
In the following examples, the DAG of the **0_2** node is displayed after the descendant nodes of the 0_2 node are **ungrouped** or **grouped**:

- The following figure shows the DAG of the 0_2 node when the descendant nodes of the 0_2 node are **ungrouped**. In this pattern, you can clearly view the upstream and downstream relationships of all the nodes.




- The following figure shows the DAG of the 0_2 node when the descendant nodes of the 0_2 node are **grouped**. In this pattern, every five descendant nodes of the 0_2 node are placed at the same level. This way, the descendant nodes are displayed in an orderly manner, and you can quickly obtain the total number of the descendant nodes.

- Right-click your desired node in a DAG and perform operations on the node.




| Operation | Description |
| --- | --- |
| **Show Ancestor Nodes** | If a workflow contains three or more nodes, specific nodes are automatically hidden in the DAG in Operation Center. You can select the number of levels to view all nodes at one or more levels. Select the largest number to view the most node dependencies. |
| **Show Descendant Nodes** | |
| **View Node Details** | Allows you to go to the **Node Information** page to view the node information, including the input table, output table, ancestor nodes, and descendant nodes. |
| **View Code** | Allows you to view the code of the node. |
| **Edit Node** | Allows you to go to the DataStudio page and modify the current node. |
| **View Instances** | Allows you to view the instances of the node. |
| **View Lineage** | Allows you to view the lineage of the node. |
| **Test** | Allows you to test the node. After you select Test, the **Test** dialog box appears. You must specify **Test Name** and **Data Timestamp** and click **OK**. Then, the **Test Instance** page appears. |

| Operation | Description |
| --- | --- |
| **Run** | Allows you to generate retroactive data for the node. The following modes for generating retroactive data are supported: **Current Node Retroactively**, **Current and Descendant Nodes Retroactively**, and **Mass Nodes Retroactively**. |
| **Freeze** | Allows you to freeze the current node and pause the scheduling of the node. |
| **Unfreeze** | Allows you to resume the scheduling of the frozen node. |
| **Configure Data Quality Rules** | Allows you to configure rules for monitoring the data quality of the node. |

# 1.4.2. View auto triggered node instances

An auto triggered node instance is a snapshot that is taken for an auto triggered node at the time the node is scheduled to run. You can view the details of and perform the required operations on an auto triggered node instance in the list of auto triggered node instances. You can also view the details of and perform the required operations on an auto triggered node instance in the directed acyclic graph (DAG) of the instance.

An instance is generated each time an auto triggered node is run as scheduled. You can manage auto triggered node instances that have been scheduled. For example, you can view the running status of instances, and stop, rerun, and unfreeze instances.

> **Note**
> - Instances are generated as scheduled for auto triggered nodes. Each generated instance runs the latest code. If you modify and recommit the node code after instances are generated for the node, the instances that have not been run will run the latest code.
> - If no alert is generated for a failed instance, you must check whether your mobile number and email address are configured on the Personal Account page.

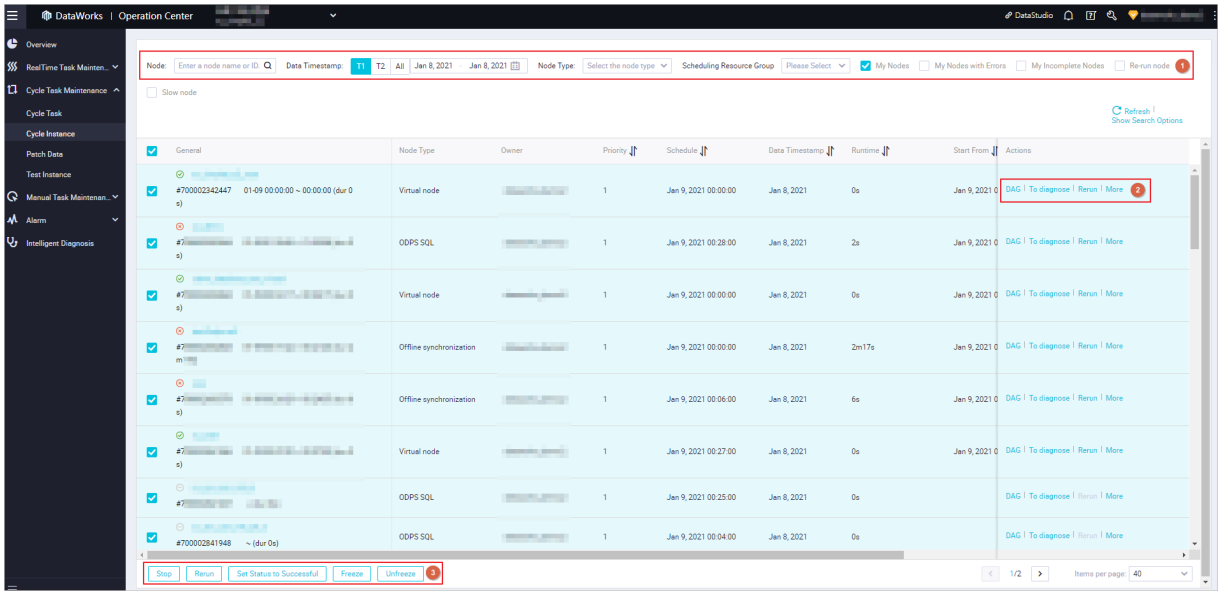
## Limits

DAGs have the following limits:

- Only users of the DataWorks Standard Edition or a more advanced edition can use the node aggregation, upstream analysis, and downstream analysis features provided by DAGs.
- Users of the DataWorks Basic Edition or Standard Edition can use the node aggregation, upstream analysis, and downstream analysis features provided by DAGs on a trial basis free of charge as of May 31, 2021. From June 1, 2021, users must update their DataWorks to the Professional Edition before they can use these features. For more information about DataWorks editions, see DataWorks advanced editions.
- You can use the node aggregation, upstream analysis, and downstream analysis features of the DAGs of auto triggered nodes that are deployed only in the China (Shenzhen) region.

## Manage auto triggered node instances

You can manage auto triggered node instances in the instance list. For example, you can check operational logs, rerun instances, and stop running instances.

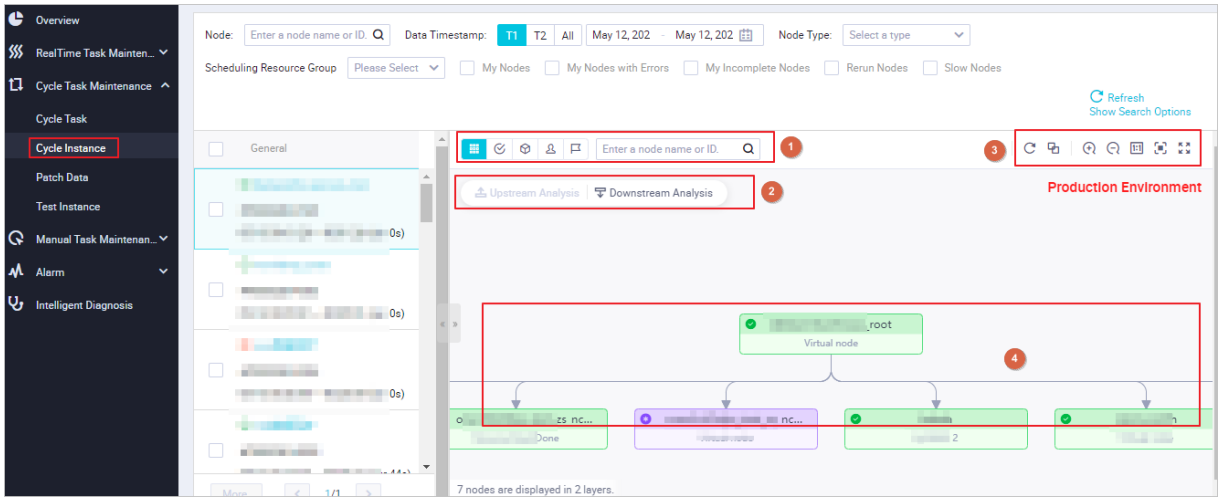



| Operation | Description |
| --- | --- |
| Filter | You can specify filter conditions to search for your desired instance in the section marked with 1 in the preceding figure.<br>You can search for instances by node name or node ID. You can also specify the following conditions to search for your desired instance: **Data Timestamp**, **Node Type**, **Run At**, **Solution**, **Workflow**, **Region**, **Engine Type**, **Engine Instance**, **Baseline**, **Owner**, **Scheduling Type**, **Status**, **My Nodes**, **My Nodes with Errors**, and **My Incomplete Nodes**.<br>**Note** By default, the data timestamp is set to the previous day of the current day. |
| Stop | Allows you to stop the instance. You can stop an instance only when it is in the **Pending (Schedule)**, **Pending (Resources)**, or **Running** state. After this operation is performed, the instance enters the **Failed** state. |
| Rerun | Allows you to rerun the instance. After the instance is rerun, its pending descendant instances are run as scheduled. Perform this operation if an instance fails to be run or it is not run as scheduled.<br>**Note** Only instances in the **Successful** or **Failed** state can be rerun. |

| Operation | Description |
| --- | --- |
| **Rerun Descendent Nodes** | Allows you to rerun the instance and its descendant instances. You must specify the instances that you want to rerun. After they are run, their pending descendant instances are run as scheduled. Perform this operation to recover data.<br>**Note** Only instances in the **Pending (Ancestor)**, **Successful**, or **Failed** state can be selected. The value No appears in the **Meet Rerun Condition** column of instances in other states, and you cannot select the instances. |
| **Set Status to Successful** | Allows you to set the status of the instance to **Successful** and run its pending descendant instances. Perform this operation if an instance fails to be run.<br>**Note** Only the status of a failed instance can be set to **Successful**. This operation does not apply to workflows. |
| **Freeze** | Allows you to freeze the instance that is in the Running state. |
| **Unfreeze** | Allows you to unfreeze the instance that is frozen.<br>• If the instance is not run, it is automatically run after the running of its ancestor instances is complete.<br>• If all the ancestor instances are run, the status of the instance is directly set to Failed. You must manually rerun the instance. |
| **Batch operations** | Allows you to perform batch operations. You can click the following buttons to perform batch operations on instances: **Stop**, **Rerun**, **Set Status to Successful**, **Freeze**, and **Unfreeze**. |

## Manage auto triggered node instances in a DAG

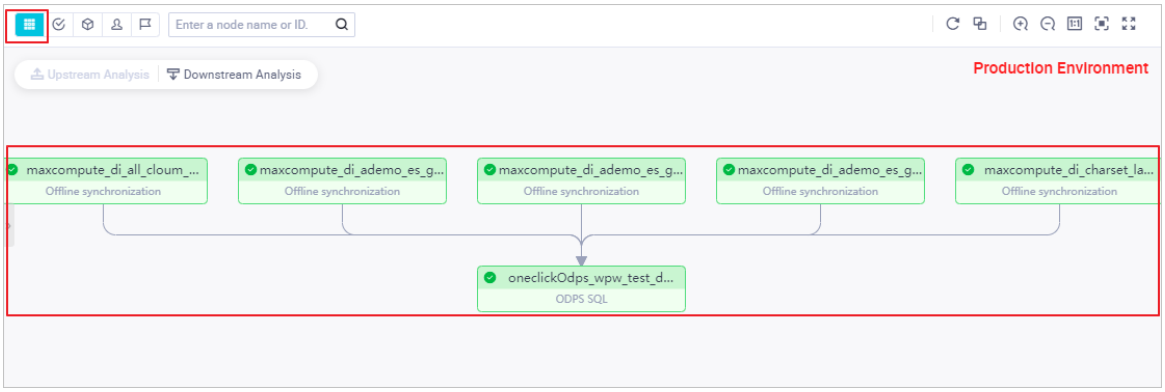
Click **DAG** in the Actions column that corresponds to a node to view the DAG of the node. You can perform the following operations in a DAG:



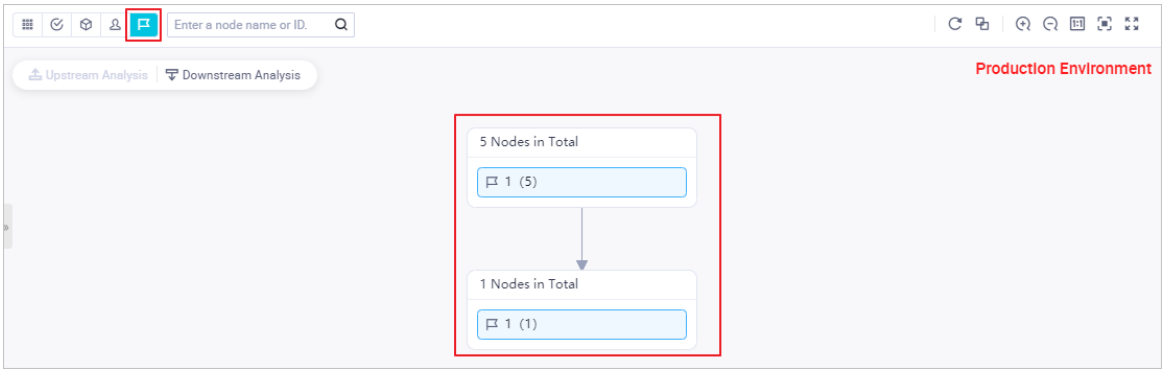

- Aggregate nodes

If an auto triggered node has multiple ancestor and descendant nodes or the ancestor and descendant nodes are distributed at multiple levels, you can aggregate the nodes. The nodes can be aggregated from dimensions such as node status, workspace, owner, and priority. Then, you can view the number of nodes from your required dimension. This allows you to understand the number of nodes from different dimensions and helps the system run the nodes. The following figures show the node distribution when the ancestor and descendant nodes of an auto triggered node are **not aggregated** or are **aggregated by priority**.

- The following figure shows the node distribution when the ancestor and descendant nodes of an auto triggered node are not aggregated.




- The following figure shows the node distribution when the ancestor and descendant nodes of an auto triggered node are aggregated by priority. From the figure, you can quickly understand that the current auto triggered node has *six* descendant nodes whose priorities are *1*.

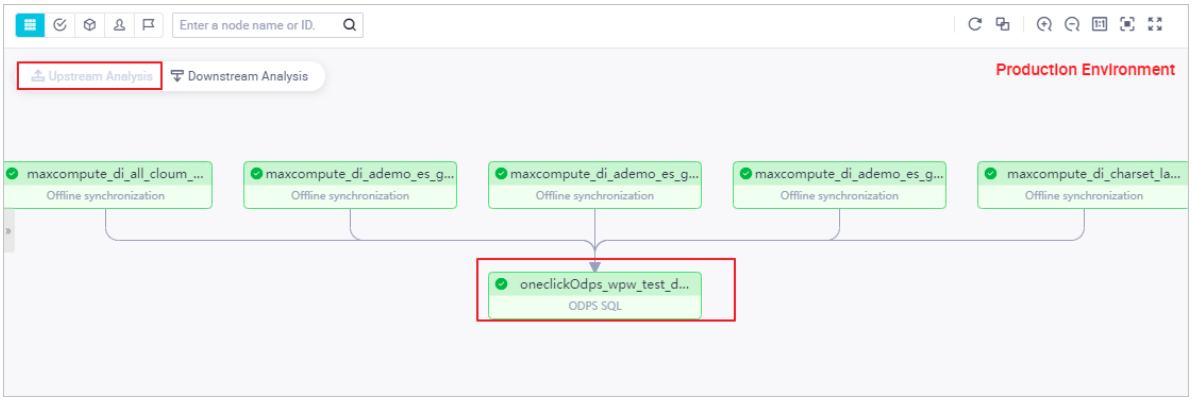



- Analyze ancestor nodes

  In most cases, an auto triggered node has both upstream and downstream relationships. If an auto triggered node is not run for a long time, you can analyze the ancestor nodes of the node. You can view the ancestor node that blocks the running of the node in the DAG of the node, and quickly locate and troubleshoot the issue. This improves the running efficiency of the node.

  > **Note** You can **analyze the ancestor nodes** of only the auto triggered nodes that are not run.

  The following figure shows how to analyze the ancestor nodes of an auto triggered node. For example, the **2_** node is not run for a long time. In this case, you can click the node and click **Upstream Analysis** in the upper-left corner to analyze the ancestor nodes of the node.




The analysis results show that the ancestor nodes that block the running of the 2_ node are the **table data synchronization** and **metric statistics** nodes. Then, you can quickly troubleshoot the issue based on the analysis results.

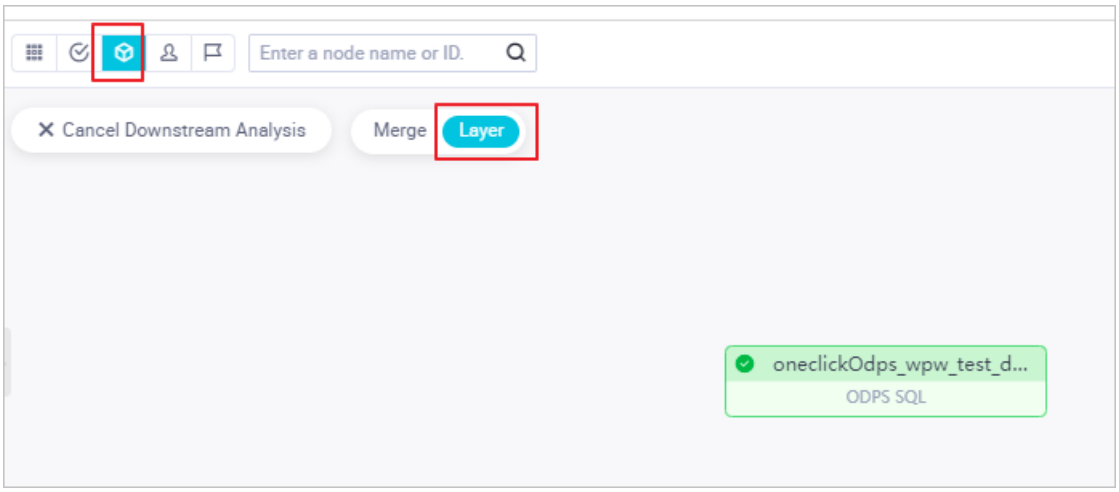
- Analyze descendant nodes

  If an auto triggered node has multiple descendant nodes or the descendant nodes of an auto triggered node are distributed at multiple levels, you can analyze the descendant nodes of the auto triggered node. You can aggregate the descendant nodes by node status, workspace, owner, or priority. Then, you can view the number of nodes at different levels from your required dimension or the total number of nodes at all levels from your required dimension.

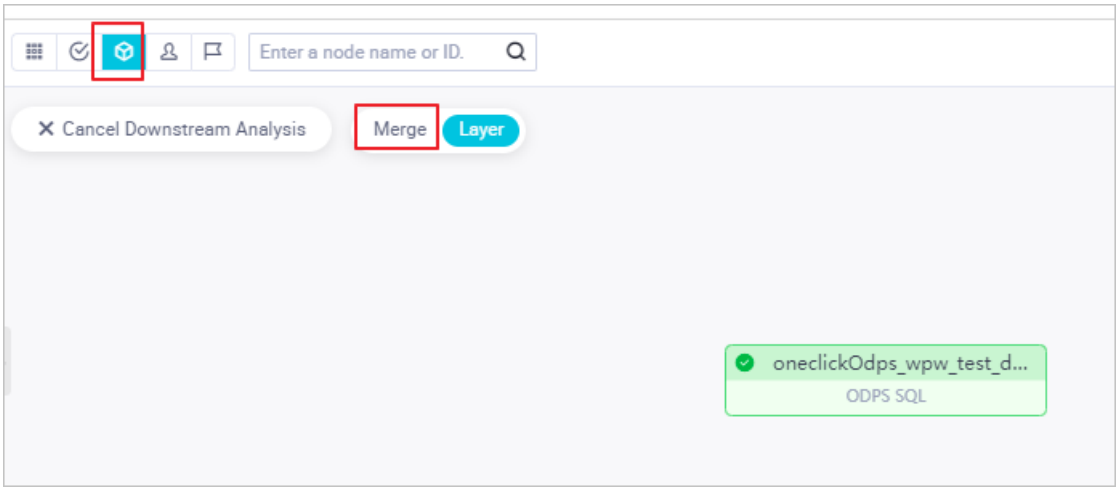
  > **Note**
  > - By default, the descendant nodes of an auto triggered node are aggregated by owner. The system calculates the total number of nodes at all levels from the owner dimension.
  > - If you **analyze the descendant nodes** of an auto triggered node, the analysis results are displayed by **level**, and a maximum of six levels of nodes can be displayed. If you want to view more levels of nodes, click **Continue Analysis** in the upper-left corner.

  In the following example, the **descendant nodes** of the **tag** node are analyzed. The following figures show the analysis results that are displayed by using different methods.

  - The descendant nodes of the tag node are **aggregated based on the workspaces to which the descendant nodes belong**, and the analysis results are presented by **level**. This way, the number of the descendant nodes in different workspaces is displayed at different levels.

- The descendant nodes of the tag node are **aggregated based on the workspaces to which the descendant nodes belong**, and the analysis results are presented by using the **merging** method. This way, all the descendant nodes are placed at the same level, and the number of the descendant nodes that belong to different workspaces is displayed.

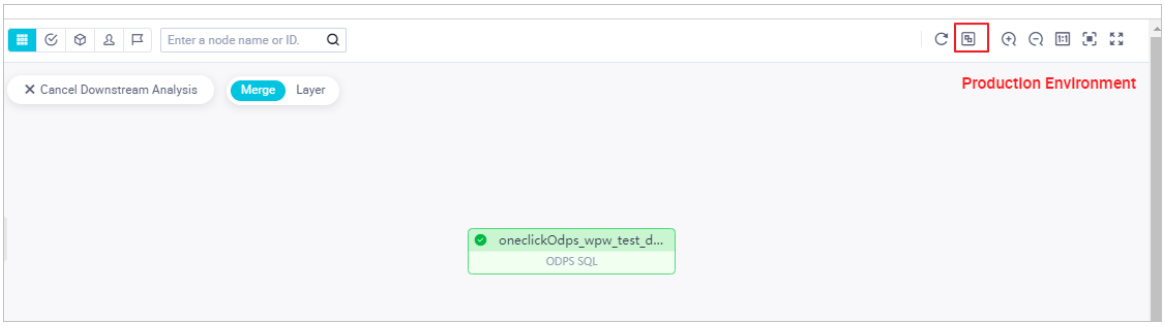



- Select a display pattern for a DAG

  You can click the icons in the upper-right corner of a DAG panel to adjust the display pattern of the DAG based on your business requirements. For example, you can click **Toggle Full Screen View** or **Fit Screen** to perform the operation.

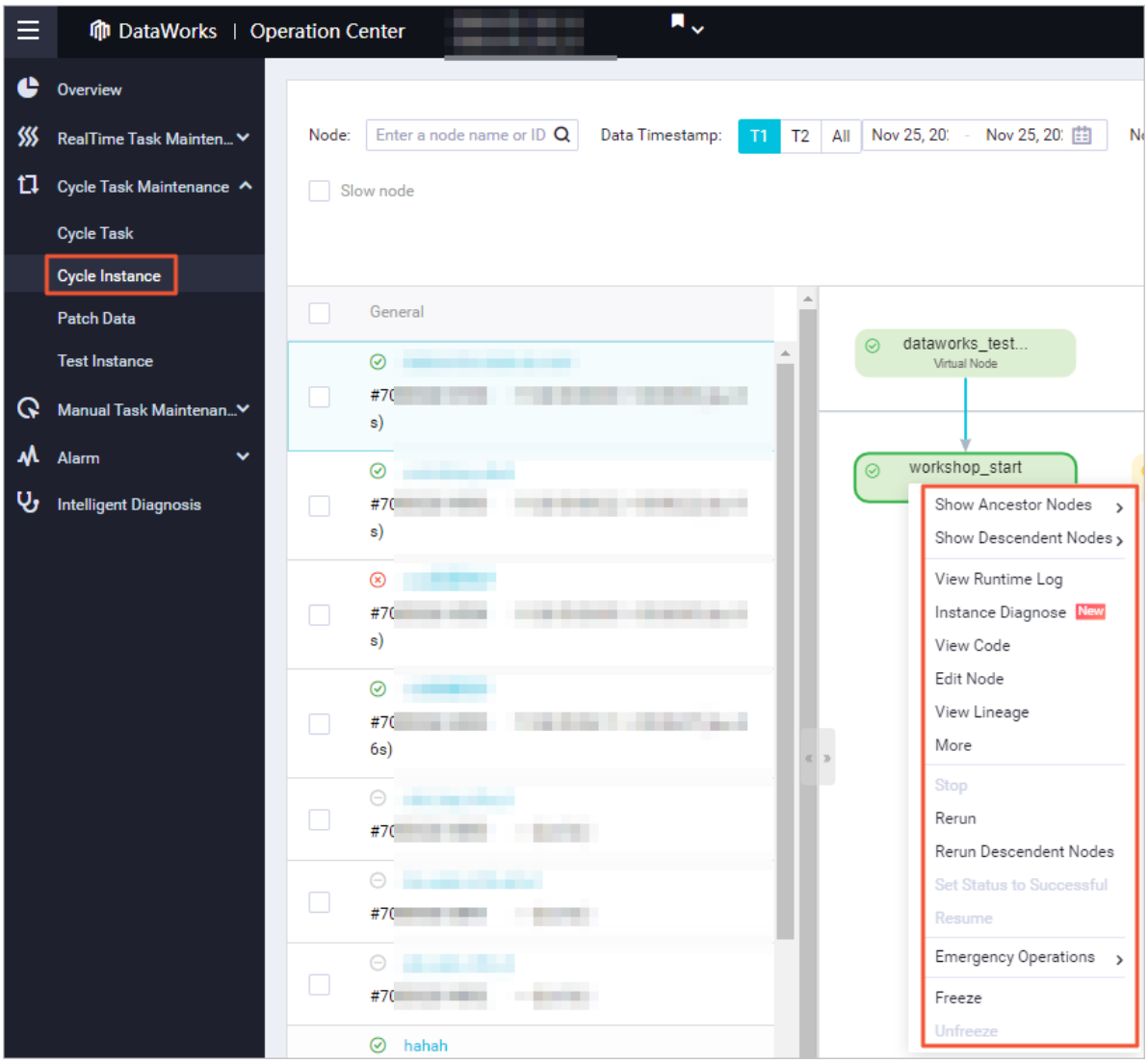
  In the following examples, the DAG of the **0_2** node is displayed after the descendant nodes of the 0_2 node are **ungrouped** or **grouped**:

  - The following figure shows the DAG of the 0_2 node when the descendant nodes of the 0_2 node are **ungrouped**. In this pattern, you can clearly view the upstream and downstream relationships of all the nodes.



- The following figure shows the DAG of the 0_2 node when the descendant nodes of the 0_2 node are **grouped**. In this pattern, every five descendant nodes of the 0_2 node are placed at the same level. This way, these descendant nodes are displayed in an orderly manner, and you can quickly obtain the total number of the descendant nodes.


- Right-click your desired node in a DAG and perform operations on the node.



| Operation | Description |
| --- | --- |
| **Show Ancestor Nodes or Show Descendent Nodes** | If a workflow contains three or more nodes, specific nodes are automatically hidden in the DAG in Operation Center. You can select the number of levels to view all nodes at one or more levels. |
| **View Runtime Log** | Allows you to view the operational logs of the current instance that is in the states such as Running, Successful, or Failed. |
| **Instance Diagnose** | DataWorks Operation Center provides the instance diagnosis feature to help you track the running of nodes and identify problems. For more information, see Instance diagnosis. |
| **View Code** | Allows you to view the code of the current instance. |
| **Edit Node** | Allows you to go to the DataStudio page and modify the current node. |
| **View Lineage** | Allows you to view the lineage of the current instance. |
| **More** | Allows you to view more instance information on the **General**, **Context**, **Runtime Log**, **Operation Log**, and **Code** tabs. |
| **Stop** | Allows you to stop the instance. Only instances in the Pending or Running state can be stopped. After this operation is performed, the instance enters the Failed state. |
| **Rerun** | Allows you to rerun the instance. After the instance is rerun, its pending descendant instances are run as scheduled. Perform this operation if an instance fails to be run or it is not run as scheduled.<br>**Note** Only instances in the **Pending (Ancestor)**, **Successful**, or **Failed** state can be rerun. |
| **Rerun Descendent Nodes** | Allows you to rerun the instance and its descendant instances. You must specify the instances that you want to rerun. After they are run, their pending descendant instances are run as scheduled. Perform this operation to recover data.<br>**Note** Only instances in the **Pending (Ancestor)**, **Successful**, or **Failed** state can be selected. The value No appears in the **Meet Rerun Condition** column of instances in other states, and you cannot select these instances. |
| **Set Status to Successful** | Allows you to set the status of the instance to Successful and run its pending descendant instances. Perform this operation if an instance fails to be run.<br>**Note** Only the status of a failed instance can be set to **Successful**. This operation does not apply to workflows. |
| **Resume** | Allows you to continue to run the instance after it fails. |

| Operation | Description |
| --- | --- |
| **Emergency Operations** | Allows you to perform emergency operations in emergencies only. Emergency operations take effect only once on the current node.<br>Select **Delete Dependencies** to delete the dependencies of the current node. You can perform this operation to start the current node if the ancestor instances fail and the current instance does not depend on the data of the ancestor instances. |
| **Freeze** | Allows you to freeze the instance that is in the Running state. |
| **Unfreeze** | Allows you to unfreeze the instance that is frozen.<br>◦ If the instance is not run, it is automatically run after the running of its ancestor instances is complete.<br>◦ If all the ancestor instances are run, the status of the instance is directly set to Failed. You must manually rerun the instance. |

# 1.4.3. Generate retroactive data and manage retroactive instances

You can generate retroactive data for an auto triggered node to run the node in a specified date range in DataWorks. You can stop, rerun and unfreeze the generated retroactive instances on the Patch Data page. This topic describes how to generate retroactive data and manage retroactive instances for an auto triggered node.

## Context

After an auto triggered node is developed, committed, and deployed to the scheduling system, the scheduling system runs the node as scheduled. If you want to run the auto triggered node in a specified date range, you can generate retroactive data for the node. For more information, see Generate retroactive data. You can select the following modes to generate retroactive data:

- **Current Node Retroactively**: This mode is used to generate retroactive data for the current node.
- **Current and Descendent Nodes Retroactively**: This mode is used to generate retroactive data for the current and descendant nodes at a time. We recommend that you use this mode when the number of the descendant nodes is small. In this mode, you can generate retroactive data for some of the descendant nodes.
- **Mass Nodes Retroactively**: This mode is used to generate retroactive data for the current and descendant nodes at a time. We recommend that you use this mode when the number of the descendant nodes is large. In this mode, you can filter descendant nodes by workspace. You can set a whitelist to generate retroactive instances for the nodes that are not in the selected workspaces. You can also set a blacklist to prevent the generation of retroactive instances for the nodes that are included in the selected workspaces.
- **Advanced Mode**: This mode is used to generate retroactive data for multiple nodes at a time. You can select nodes that may not have dependencies with each other. You can select nodes for which you want to generate retroactive data in the Directed acyclic graph (DAG) or in the node list on the Cycle Task page.
  - In the DAG, you can use the node aggregation feature to group nodes by workspace, owner, or priority. This way, you can generate retroactive data for the node group.

- You can also select nodes in the node list on the Cycle Task page. You can filter nodes based on specific conditions and select the nodes for which you want to generate retroactive data.

## Limits

- You can use the advanced mode only in workspaces in the China (Shenzhen) and UAE (Dubai) regions.
- You can stop multiple retroactive instances at a time, but you cannot delete them at a time. A retroactive instance is automatically deleted about 30 days after it expires.

## Considerations

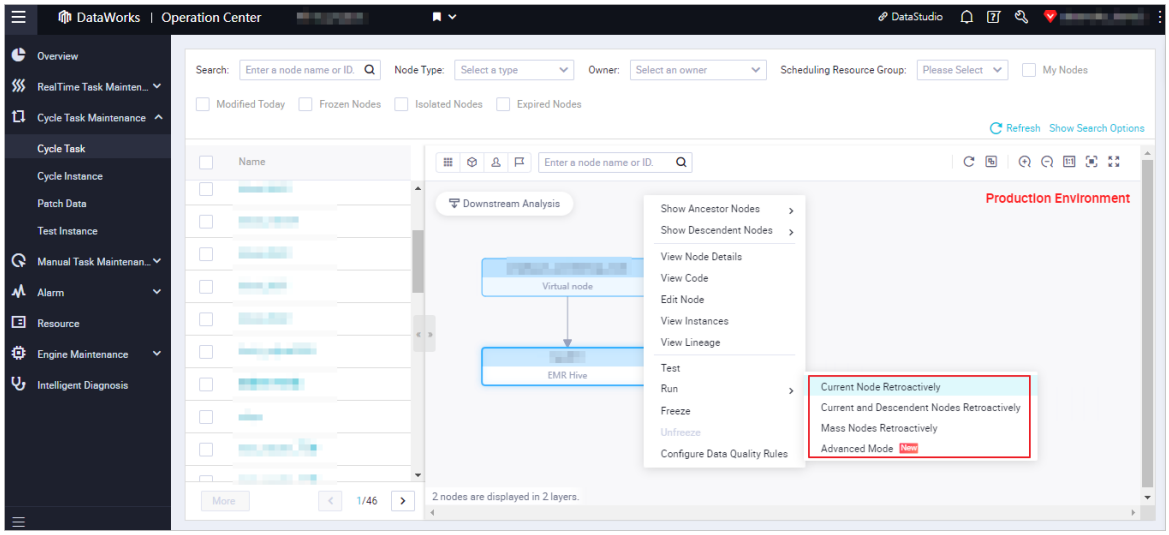
- When DataWorks generates retroactive data for a node for a specific time range, if one instance of the node fails on a day within the time range, the retroactive instance for that day is also set to failed. DataWorks will not run the instances of this node for the next day. To sum up, DataWorks runs the instances of a node on a day only when all its instances on the previous day are successful.
- For a self-dependent auto triggered node, if the first instance for which retroactive data needs to be generated has a last-cycle instance that is not run on the previous day, the retroactive instance cannot be run. If the first instance for which retroactive data needs to be generated does not have a last-cycle instance on the previous day, the retroactive instance is directly run.
- If both an auto triggered node instance and a retroactive instance are running for a node, you must stop the retroactive instance to ensure that the auto triggered node instance can be run as expected.
- A large number of retroactive instances or concurrent instances may lead to insufficient resources for the recurring schedule. Make sure that the number of instances is appropriate based on your business requirements.

## Generate retroactive data

1. Go to the **DataStudio** page.
   i. Log on to the DataWorks console.
   ii. In the left-side navigation pane, click **Workspaces**.
   iii. In the top navigation bar, select the region where the required workspace resides, find the workspace, and then click **Data Analytics**.
2. On the **DataStudio** page, click the icon in the upper-left corner and choose **All Products > Operation Center**.
3. In the left-side navigation pane, choose **Cycle Task Maintenance > Cycle Task**.
4. Generate retroactive data for nodes that you want.
   i. Click the name of an auto triggered node in the node list to open the DAG.

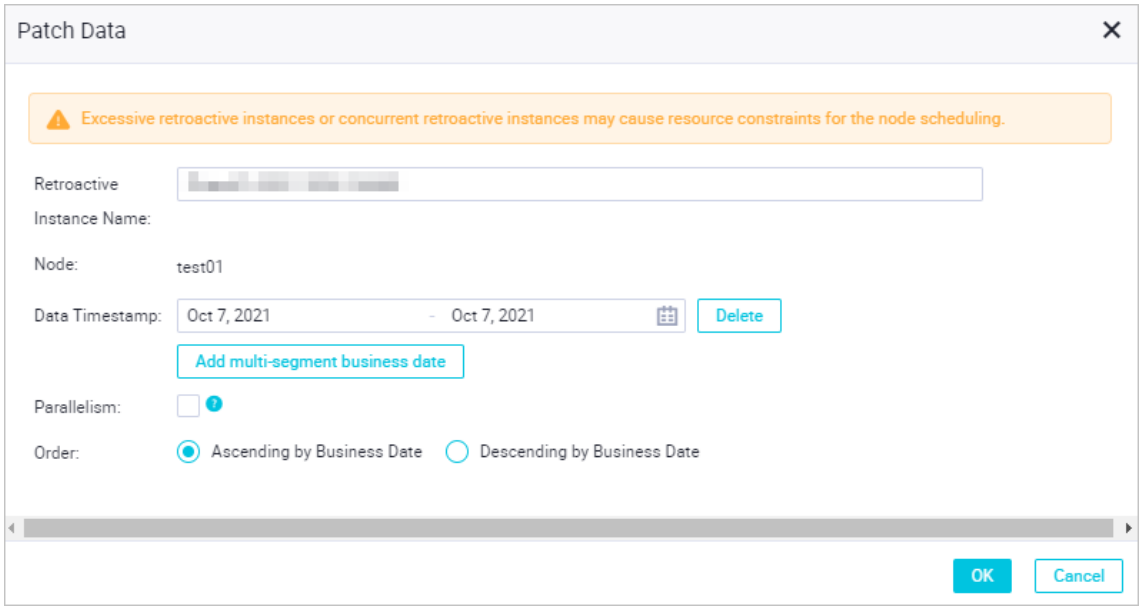
      You can also click the » icon to show the node list. Find the required node and click **DAG** in the **Actions** column to open the DAG.
   ii. Right-click the node in the DAG. In the shortcut menu that appears, move the pointer over **Run** and select a mode for generating retroactive data. In the dialog box that appears, the parameters.

> **Note** You can also perform this step on the **Cycle Task** page. Click the » icon to show the auto triggered node list. Find the required auto triggered node, click **Patch Data** in the **Actions** column, and then select a mode for generating retroactive data.




The following tables describe the parameters that you must set when you select different modes for generating retroactive data:

- Generate retroactive data in **Current Node Retroactively** mode.



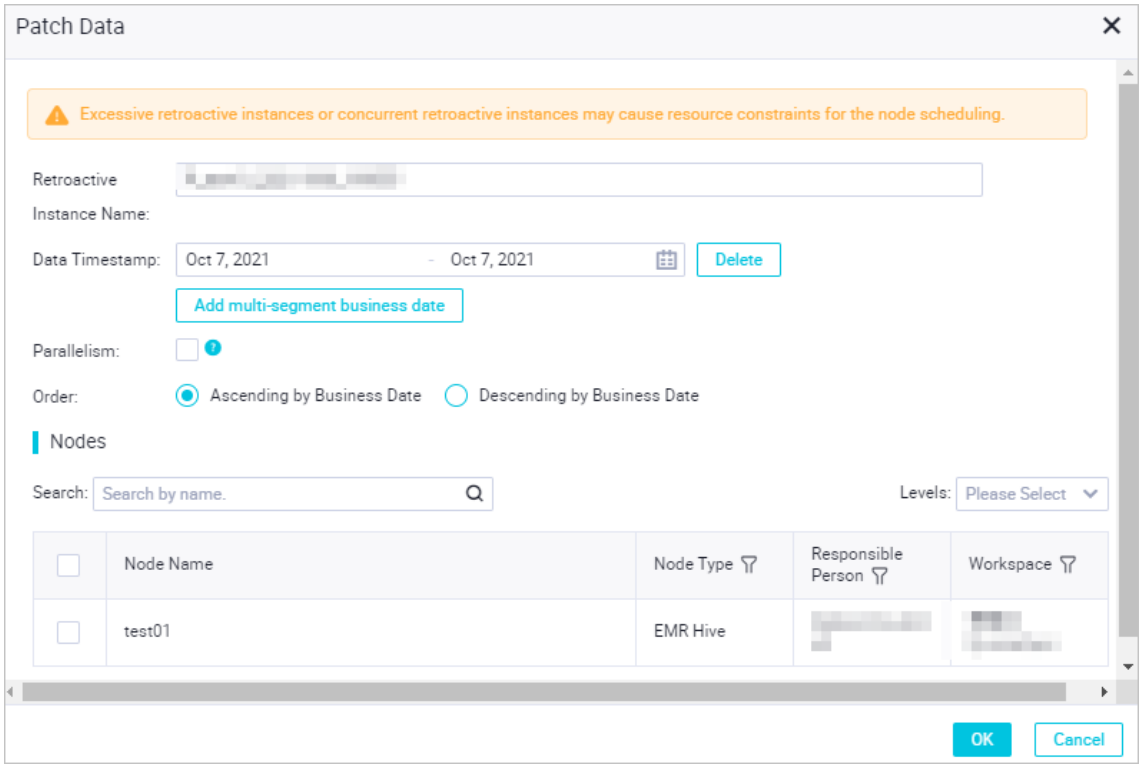

The following table describes the parameters.

| Parameter | Description |
| --- | --- |
| **Retroactive Instance Name** | DataWorks automatically generates a retroactive instance name. You can change the name as needed. |

| Parameter | Description |
| --- | --- |
| **Node** | The name of the node for which you want to generate retroactive data. |
| **Data Timestamp** | The data timestamp range of the retroactive instances. A data timestamp is a date-based timestamp.<br>■ If you want to generate retroactive data for the node in multiple non-consecutive date ranges, click **Add multi-segment business date**.<br>■ If the start date of the timestamp is later than the current date, you can select **Run Retroactive Instances Scheduled to Run after the Current Time**. When the start date passes, the system automatically runs the retroactive instance.<br>For example, if the current date is `August 24, 2021` and the start date of the timestamp is `September 17, 2021`, the system runs the retroactive instance on `September 18, 2021`.<br>**Note** We recommend that you do not set this parameter to a long range. Otherwise, the retroactive instance may be delayed due to insufficient resources. |
| **Parallelism** | Specifies whether to run multiple retroactive instances in parallel.<br>■ If you do not select Parallelism, the retroactive instances are run in sequence based on the data timestamps.<br>■ If you select Parallelism, a specific number of retroactive instances are generated based on the data timestamps and run in parallel. The number of retroactive instances is specified by the **Number of Concurrent Nodes** parameter. Instances with different data timestamps can be run at the same time. |

| Parameter | Description |
| --- | --- |
| **Number of Concurrent Nodes** | The number of retroactive instances that are created and run in parallel during the generation of retroactive data.<br><br>**Note** This parameter is required if **Parallelism** is selected.<br><br>You can set the Number of Concurrent Nodes parameter to an integer from *2* to *10*. The following rules are applied when multiple retroactive instances are run in parallel:<br>■ If the number of data timestamps is smaller than the number of parallel instances, the retroactive instances are run in parallel. For example, the data timestamps are from January 11 to January 13, and you set the **Number of Concurrent Nodes** parameter to 4. In this case, a retroactive instance is generated for each of the three data timestamps. These three retroactive instances are run in parallel.<br>■ If the number of data timestamps is larger than the number of parallel instances, the system runs specific instances in sequence and specific instances in parallel based on the data timestamps. For example, the data timestamps are from January 11 to January 13, and you set the **Number of Concurrent Nodes** parameter to 2. In this case, two retroactive instances are generated. They are run in parallel for once, and one of them must be run for the second time. |
| **Order** | Valid values: **Ascending by Business Date** and **Descending by Business Date**. You can generate retroactive data in an ascending or descending order of data timestamps. |

- Select **Current and Descendent Nodes Retroactively**.

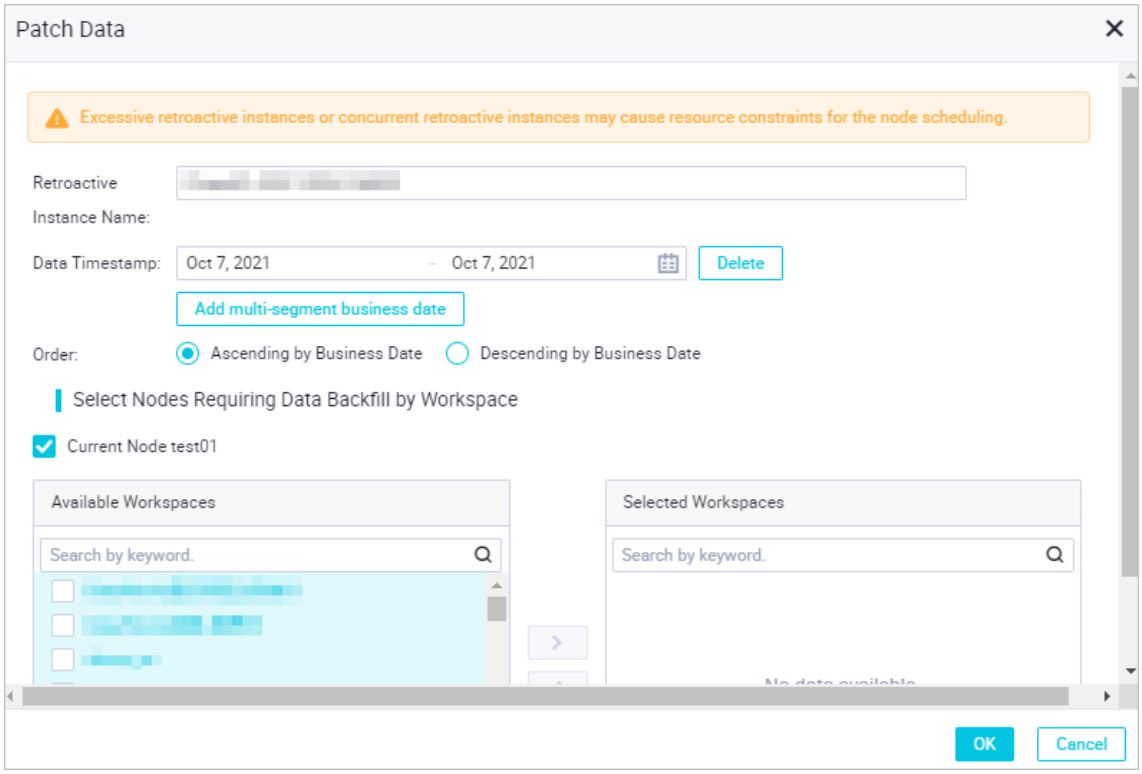



The following table describes the parameters.

| Parameter | Description |
| --- | --- |
| **Retroactive Instance Name** | DataWorks automatically generates a retroactive instance name. You can change the name as needed. |
| **Data Timestamp** | The data timestamp range of the retroactive instances. A data timestamp is a date-based timestamp.<br>■ If you want to generate retroactive data for the node in multiple non-consecutive date ranges, click **Add multi-segment business date**.<br>■ If the start date of the timestamp is later than the current date, you can select **Run Retroactive Instances Scheduled to Run after the Current Time**. When the start date passes, the system automatically runs the retroactive instance.<br>For example, if the current date is `August 24, 2021` and the start date of the timestamp is `September 17, 2021`, the system runs the retroactive instance on `September 18, 2021`.<br>**Note** We recommend that you do not set this parameter to a long range. Otherwise, the retroactive instance may be delayed due to insufficient resources. |

| Parameter | Description |
| --- | --- |
| **Parallelism** | Specifies whether to run multiple retroactive instances in parallel.<br>■ If you do not select Parallelism, the retroactive instances are run in sequence based on the data timestamps.<br>■ If you select Parallelism, a specific number of retroactive instances are generated based on the data timestamps and run in parallel. The number of retroactive instances is specified by the **Number of Concurrent Nodes** parameter. Instances with different data timestamps can be run at the same time. |
| **Number of Concurrent Nodes** | You can set the Number of Concurrent Nodes parameter to an integer from *2* to *10*. The following rules are applied when multiple retroactive instances are run in parallel:<br>■ If the number of data timestamps is smaller than the number of parallel instances, the retroactive instances are run in parallel. For example, the data timestamps are from January 11 to January 13, and you set the **Number of Concurrent Nodes** parameter to 4. In this case, a retroactive instance is generated for each of the three data timestamps. These three retroactive instances are run in parallel.<br>■ If the number of data timestamps is larger than the number of parallel instances, the system runs specific instances in sequence and specific instances in parallel based on the data timestamps. For example, the data timestamps are from January 11 to January 13, and you set the **Number of Concurrent Nodes** parameter to 2. In this case, two retroactive instances are generated. They are run in parallel for once, and one of them must be run for the second time. |
| **Order** | Valid values: **Ascending by Business Date** and **Descending by Business Date**. You can generate retroactive data in an ascending or descending order of data timestamps. |
| **Nodes** | You can filter nodes by name and level and select the nodes for which you want to generate retroactive data.<br>**Note**<br>■ You can search nodes by name in fuzzy match mode. When you enter a keyword, all the nodes whose name contains the keyword appear in the table below the search box.<br>■ The search scope includes the current node and its descendant nodes of all levels. You can select the current node and some or all of its descendant nodes. |

- Select **Mass Nodes Retroactively**.



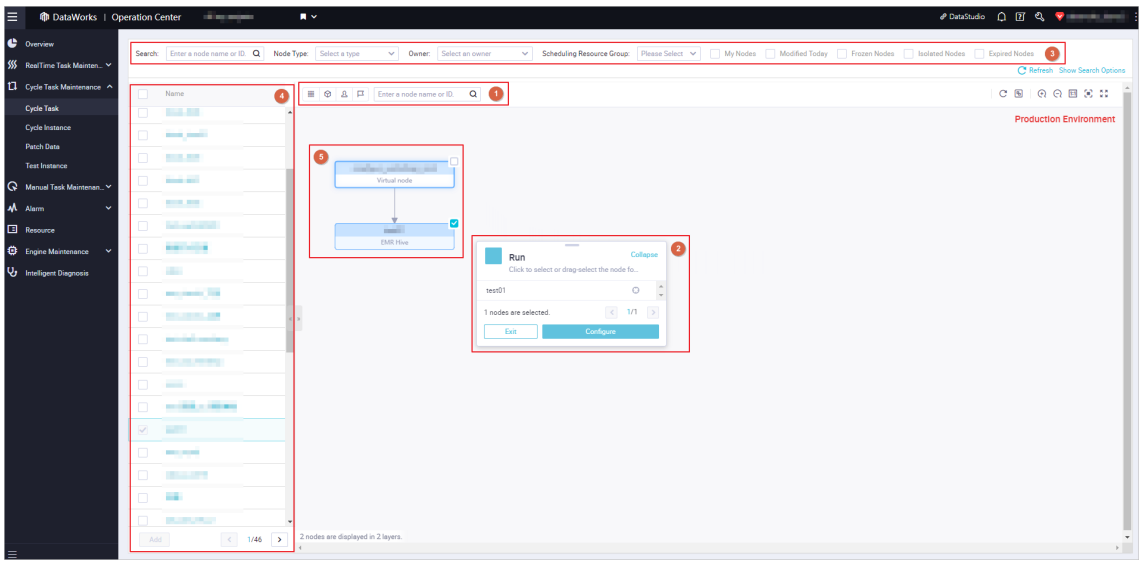

The following table describes the parameters.

| Parameter | Description |
| --- | --- |
| **Retroactive Instance Name** | DataWorks automatically generates a retroactive instance name. You can change the name as needed. |
| **Data Timestamp** | The data timestamp range of the retroactive instances. A data timestamp is a date-based timestamp.<br>■ If you want to generate retroactive data for the node in multiple non-consecutive date ranges, click **Add multi-segment business date**.<br>■ If the start date of the timestamp is later than the current date, you can select **Run Retroactive Instances Scheduled to Run after the Current Time**. When the start date passes, the system automatically runs the retroactive instance.<br>For example, if the current date is `August 24, 2021` and the start date of the timestamp is `September 17, 2021`, the system runs the retroactive instance on `September 18, 2021`.<br>**Note** We recommend that you do not set this parameter to a long range. Otherwise, the retroactive instance may be delayed due to insufficient resources. |
| **Order** | Valid values: **Ascending by Business Date** and **Descending by Business Date**. You can generate retroactive data in an ascending or descending order of data timestamps. |

| Parameter | Description |
| --- | --- |
| **Select Nodes Requiring Data Backfill by Workspace** | You can select workspaces in the **Available Workspaces** section and add them to the **Selected Workspaces** section. This way, you can generate retroactive data for all the nodes in the workspaces that you select.<br>**Note**<br>■ You can search workspaces by name in fuzzy match mode. When you enter a keyword, the workspaces whose name contains the keyword appear in the sections.<br>■ You can select workspaces only in the current region.<br>■ You can set a whitelist to generate retroactive data for the nodes that are not in the selected workspaces. You can also set a blacklist to prevent the generation of retroactive data for the nodes that are included in the selected workspaces.<br>■ You can specify whether to generate retroactive data for the current node.<br>■ If you select **Current Node**, retroactive instances are generated for the current node and its descendant nodes.<br>■ If you clear **Current Node**, a dry-run instance is generated for the current node and retroactive instances are generated for its descendant nodes. |
| **Node Whitelist** | You can select the nodes that are not in the selected workspaces and generate retroactive data for the nodes.<br>**Note** You can search for nodes only by node ID. |
| **Node Blacklist** | You can select the nodes in the selected workspaces for which you do not want to generate retroactive data.<br>**Note** You can search for nodes only by node ID. |

- Select **Advanced Mode**.

In advanced mode, you can use the node aggregation feature of the DAG to group nodes by different conditions such as the node type and owner. You can generate retroactive data for nodes that have no dependencies with each other.




The following steps describe how to generate retroactive data in this mode:

a. Select the nodes for which you want to generate retroactive data.

- In the DAG of the auto triggered node, you can use the node aggregation feature by clicking the **Not Aggregate**, **Aggregate By Workspace**, **Aggregate By Owner** or **Aggregate By Priority** icons in Section 1. This way, you group nodes by workspace, owner, or priority. You can select the check box in the upper-right corner of a group to select all the nodes in the group in Section 2.
- You can also select nodes in the node list on the Cycle Task page. You can filter nodes based on different conditions such as the node name, node type, owner, and scheduling resource group dimension in Section 3. You can select the auto triggered node for which you want to generate retroactive data in Section 4. Click **Add** at the bottom of the node list.

> **Note** This way, the system generates retroactive data for the current node and all its descendant nodes of the auto triggered node. If you want to generate retroactive data for only some of the descendant nodes of the auto triggered node, click the name of the auto triggered node to enter the DAG and select the descendant nodes for which you want to generate retroactive data.

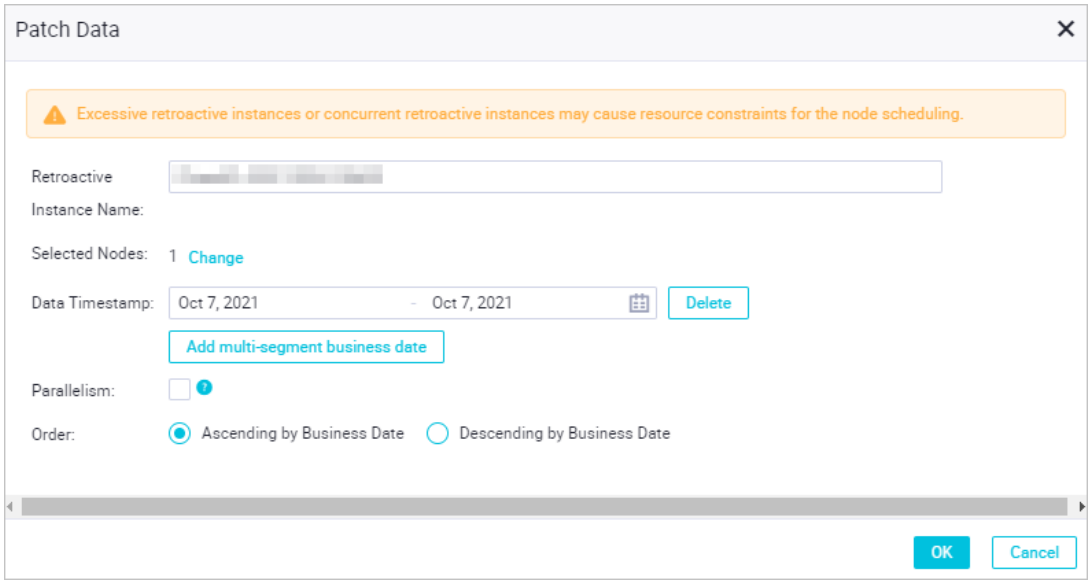
b. View the selected nodes.

After the nodes for which you want to generate retroactive data are selected, you can view the selected nodes in the **Run** panel in Section 5. You can also perform the following operations:

- Click the icon after the name of a node to open the DAG of the node. You can re-select the nodes for which you want to generate retroactive data.
- Click the icon after the name of a node to remove the node.

c. In the **Run** panel in Section 5, click **Configure** to set the parameters for the generation of

retroactive data.




The following table describes the parameters.

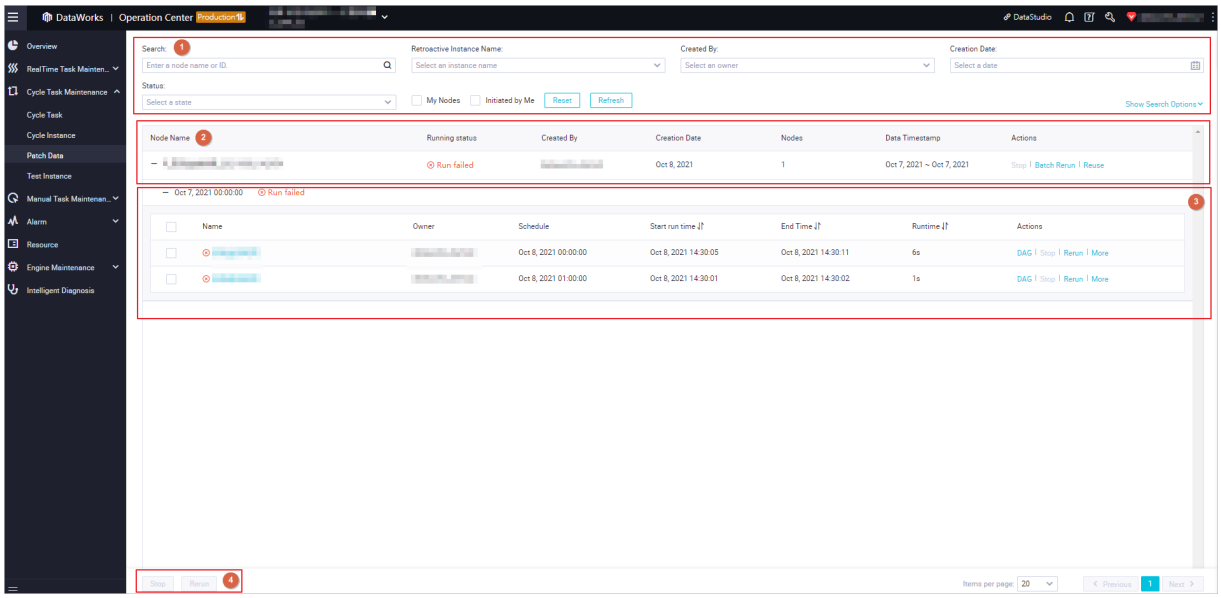
| Parameter | Description |
| --- | --- |
| **Retroactive Instance Name** | DataWorks automatically generates a retroactive instance name. You can change the name as needed. |
| **Selected Nodes** | The number of nodes for which you want to generate retroactive data. You can click **Change** to change the nodes for which you want to generate retroactive data. |
| **Data Timestamp** | The data timestamp range of the retroactive instances. A data timestamp is a date-based timestamp.<br>■ If you want to generate retroactive data for the node in multiple non-consecutive date ranges, click **Add multi-segment business date**.<br>■ If the start date of the timestamp is later than the current date, you can select **Run Retroactive Instances Scheduled to Run after the Current Time**. When the start date passes, the system automatically runs the retroactive instance.<br>For example, if the current date is `August 24, 2021` and the start date of the timestamp is `September 17, 2021`, the system runs the retroactive instance on `September 18, 2021`.<br>**Note** We recommend that you do not set this parameter to a long range. Otherwise, the retroactive instance may be delayed due to insufficient resources. |

| Parameter | Description |
| --- | --- |
| **Parallelism** | Specifies whether to run multiple retroactive instances in parallel.<br>■ If you do not select Parallelism, the retroactive instances are run in sequence based on the data timestamps.<br>■ If you select Parallelism, a specific number of retroactive instances are generated based on the data timestamps and run in parallel. The number of retroactive instances is specified by the **Number of Concurrent Nodes** parameter. Instances with different data timestamps can be run at the same time. |
| **Number of Concurrent Nodes** | You can set the Number of Concurrent Nodes parameter to an integer from *2* to *10*. The following rules are applied when multiple retroactive instances are run in parallel:<br>■ If the number of data timestamps is smaller than the number of parallel instances, the retroactive instances are run in parallel. For example, the data timestamps are from January 11 to January 13, and you set the **Number of Concurrent Nodes** parameter to 4. In this case, a retroactive instance is generated for each of the three data timestamps. These three retroactive instances are run in parallel.<br>■ If the number of data timestamps is larger than the number of parallel instances, the system runs specific instances in sequence and specific instances in parallel based on the data timestamps. For example, the data timestamps are from January 11 to January 13, and you set the **Number of Concurrent Nodes** parameter to 2. In this case, two retroactive instances are generated. They are run in parallel for once, and one of them must be run for the second time. |
| **Order** | Valid values: **Ascending by Business Date** and **Descending by Business Date**. You can generate retroactive data in an ascending or descending order of data timestamps. |

5. Click **OK** to start to generate retroactive data.

## Manage retroactive instances

In the left-side navigation pane of the **Operation Center** page, choose **Cycle Task Maintenance > Patch Data**. You can view the details and status of retroactive instances, and stop or rerun the instances. For more information about how to go to **Operation Center**, see the steps described in the "Generate retroactive data" section. The following table describes the operations that you can perform in different sections in the following figure.




| Section | Description |
| --- | --- |
| 1 | In this section, you can specify filter conditions to search for specific retroactive instances.<br>You can set filter conditions such as the node name, node ID, retroactive instance name, creator, creation time, status, and data timestamp. You can also filter retroactive instances based on the nodes that you own or initiate.<br>**Note**<br>• You must click **Show Search Options** if you want to set more filter conditions such as **Node Type**, **Scheduling Resource Group**, and **Engine Instance**.<br>• You can search for nodes by node name in fuzzy match mode. When you enter a keyword, all the nodes whose name contains the keyword appear in the table below. |

| Section | Description |
| --- | --- |
| 2 | In this section, you can view details of retroactive instances, including:<br>• **Node Name**: the name of the retroactive instance. Click the + icon before the name of the retroactive instance to check information in Section 3, such as the date when the retroactive instance is run and details of the nodes for which the instance is generated.<br>• **Check Status**: the check status of the retroactive instance.<br>• **Running status**: the status of the retroactive instance. The retroactive instance can be in the status of running, not running, waiting for resources, exception, and stopping.<br>• **Created by**: the Alibaba Cloud account that is used to create the retroactive instance.<br>• **Creation Date**: the date when the retroactive instance is created.<br>• **Nodes**: the number of nodes for which the instance is generated.<br>• **Data Timestamp**: the date when the instance is run.<br>In this section, you can also perform the following operations on retroactive instances:<br>• **Stop**: You can stop multiple retroactive instances that are running or waiting for resources at a time. This way, the instances are set to failed.<br>Note<br>◦ You can stop multiple retroactive instances at a time, but you cannot delete them at a time. A retroactive instance is automatically deleted about 30 days after it expires.<br>◦ You cannot stop instances that are not running, succeed or fail to be run.<br>• **Batch Rerun**: You can rerun multiple retroactive instances at a time.<br>Note Only instances that are set to failed can be rerun at a time.<br>• **Reuse**: You can reuse a group of nodes for which the retroactive instance is generated. This facilitates your selection of nodes for which you want to generate a retroactive instance. |

| Section | Description |
| --- | --- |
| 3 | In this section, you can view the details of the nodes for which the retroactive instance is generated. The details include:<br>• **Name**: the name of the node. Click the name. Then, you can view the DAG and details of the node.<br>• **Owner**: the owner of the workspace to which the node belongs.<br>• **Schedule**: the time scheduled to run the node.<br>• **Start run time**: the time when the node starts to run.<br>• **End time**: the time when the node stops running.<br>• **Runtime**: the amount of time consumed to run the node.<br>In this section, you can also perform the following operations on nodes:<br>• **Stop**: You can stop the nodes that are running or waiting for resources. This way, the node is set to failed.<br>**Note** You cannot stop the nodes that are not running, succeeded, or failed.<br>• **Rerun**: You can rerun nodes.<br>**Note** You can rerun only nodes that are successful or failed.<br>• **More > Rerun Descendant Nodes**: Rerun the descendant nodes of a node.<br>• **More > Set Status to Successful**: Set the status of a node to succeeded.<br>• **More > Freeze**: Freeze a node and pause the scheduling of the node.<br>• **More > Unfreeze**: Resume the scheduling of a frozen node.<br>• **More > View Lineage**: View the lineage of a node. |
| 4 | You can select multiple nodes in Section 3 and click **Stop** or **Rerun** in Section 4. This way, you can stop or rerun multiple nodes at a time. |

## Instance states

| No. | State | Icon |
| --- | --- | --- |
| 1 | Succeeded | |
| 2 | Not running | |
| 3 | Failed | |
| 4 | Running | |

| No. | State | Icon |
| --- | --- | --- |
| 5 | Waiting for resources |  |
| 6 | Frozen |  |

# 1.4.4. Manage related auto triggered nodes in the DAG of a retroactive instance

This topic describes how to manage related auto triggered nodes in the directed acyclic graph (DAG) of a retroactive instance.

## Context

After an auto triggered node is developed, committed, and deployed to the scheduling system, the scheduling system runs the node as scheduled. If you want to synchronize the retroactive data of the auto triggered node in a historical date range, you can run the node in a specific retroactive data generation mode. After the auto triggered node is run, a retroactive instance is generated. The generated retroactive instance is run based on the specified data timestamp range. For more information, see Generate retroactive data and manage retroactive instances. You can select the following modes to generate retroactive data:

- **Current Node Retroactively**: This mode is used to generate retroactive data for the current node.
- **Current and Descendent Nodes Retroactively**: This mode is used to generate retroactive data for the current and descendant nodes at a time. In this mode, you can generate retroactive data for specific descendant nodes.
- **Mass Nodes Retroactively**: This mode is used to generate retroactive data for all the nodes in the selected workspaces. In this mode, you can filter nodes by workspace. You can set a whitelist to generate retroactive data for the nodes that are not in the selected workspaces. You can also set a blacklist to prevent the generation of retroactive data for the nodes that are included in the selected workspaces.
- **Advanced Mode**: This mode is used to generate retroactive data for multiple nodes at a time. You can select nodes that may not have dependencies with each other. You can select nodes for which you want to generate retroactive data in the DAG or in the node list on the Cycle Task page.
  - In the DAG, you can use the node aggregation feature to group nodes by workspace, owner, or priority. This way, you can generate retroactive data for the node group.
  - You can also select nodes in the node list on the Cycle Task page. You can filter nodes based on specific conditions and select the nodes for which you want to generate retroactive data.

## Limits

DAGs have the following limits:

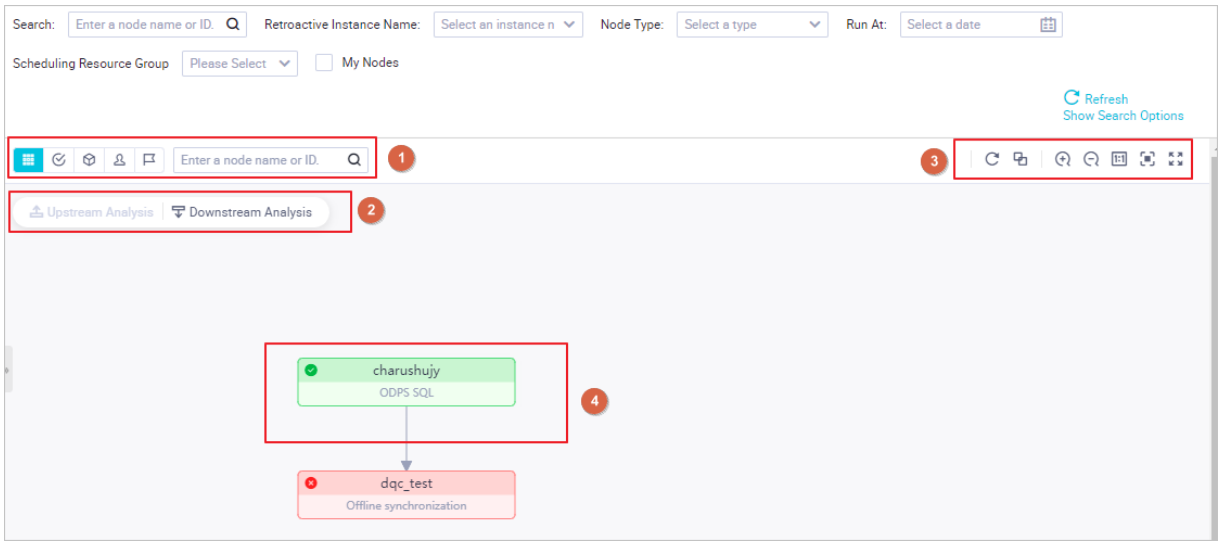
- Only users of DataWorks Standard Edition or a more advanced edition can use the node aggregation, upstream analysis, and downstream analysis features provided by DAGs.
- Users of DataWorks Basic Edition or Standard Edition can use the node aggregation, upstream analysis, and downstream analysis features provided by DAGs on a trial basis free of charge as of May 31, 2021. From June 1, 2021, users must update their DataWorks edition to Professional Edition before

they can use these features. For more information about DataWorks editions, see DataWorks advanced editions.

- You can use the node aggregation, upstream analysis, and downstream analysis features of DAGs only in the China (Shenzhen) region.

## Manage auto triggered nodes in a DAG

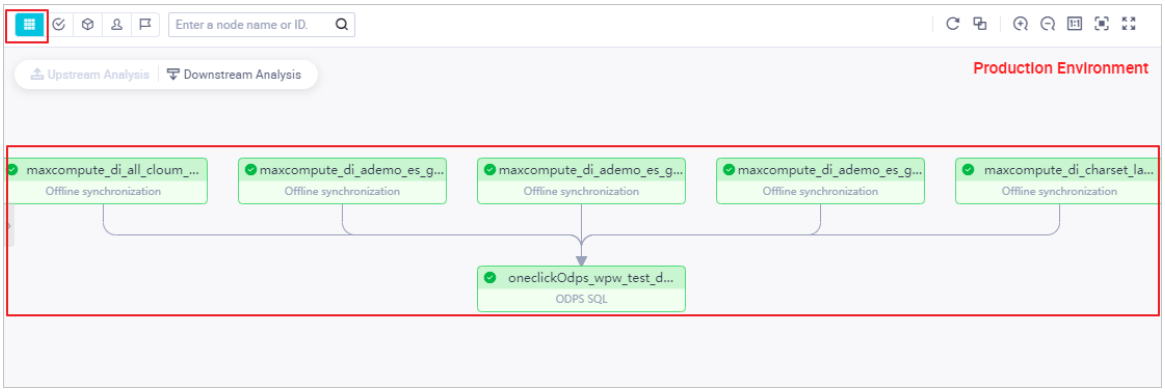
Find the retroactive instance based on which you want to manage auto triggered nodes and click **DAG** in the Actions column. You can perform the following operations in a DAG:



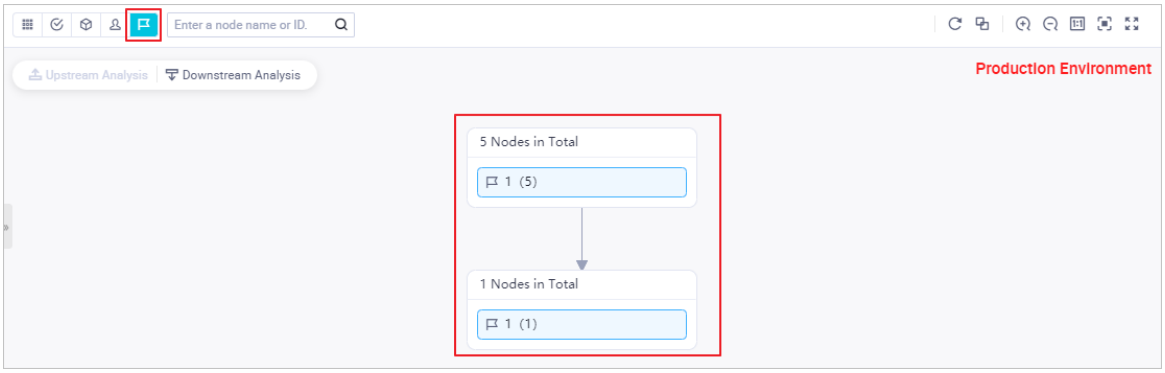

- Aggregate nodes

  If an auto triggered node has multiple ancestor and descendant nodes or the ancestor and descendant nodes are distributed at multiple levels, you can aggregate the nodes. The nodes can be aggregated from dimensions such as node status, workspace, owner, and priority. Then, you can view the number of nodes from your required dimension. This allows you to understand the numbers of nodes from different dimensions and helps you plan node running. The following figures show the node distribution when the ancestor and descendant nodes of an auto triggered node are not aggregated or are aggregated by priority.

  - The following figure shows the node distribution when the ancestor and descendant nodes of an auto triggered node are not aggregated.



- The following figure shows the node distribution when the ancestor and descendant nodes of an auto triggered node are aggregated by priority. The current auto triggered node has six descendant nodes whose priorities are 1, as shown in the following figure.

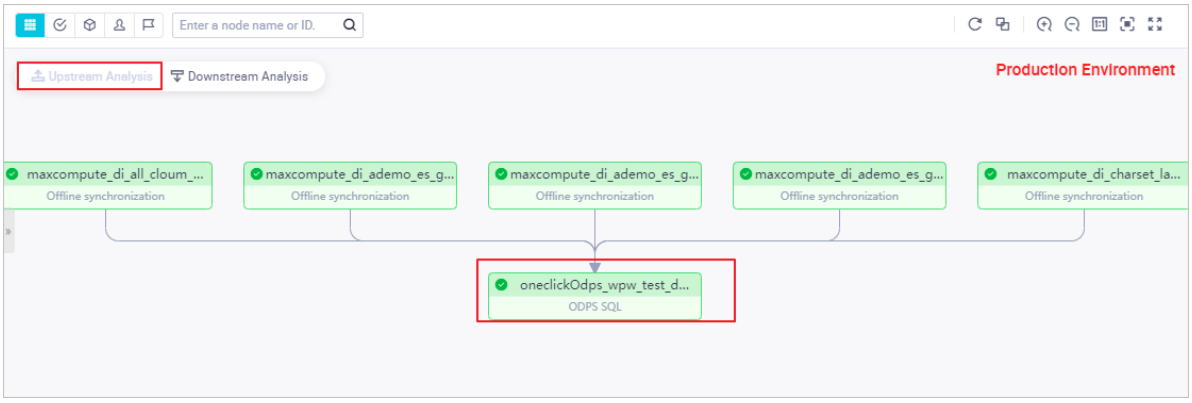



- Analyze ancestor nodes

  In most cases, an auto triggered node has both upstream and downstream nodes. If an auto triggered node is not run for a long time, you can analyze the ancestor nodes of the node. You can view the ancestor node that blocks the running of the node in the DAG and troubleshoot the issue. This improves the running efficiency of the node.

  > **Note** You can analyze the ancestor nodes of only the auto triggered nodes that are not run.

  The following figure shows how to analyze the ancestor nodes of an auto triggered node. For example, the **2_** node is not run for a long time. In this case, you can click the node and click **Upstream Analysis** in the upper-left corner to analyze the ancestor nodes of the node.




  The analysis results show that the ancestor nodes that block the running of the 2_ node are the **table data synchronization** and **metric statistics** nodes. Then, you can troubleshoot the issue based on the analysis results.

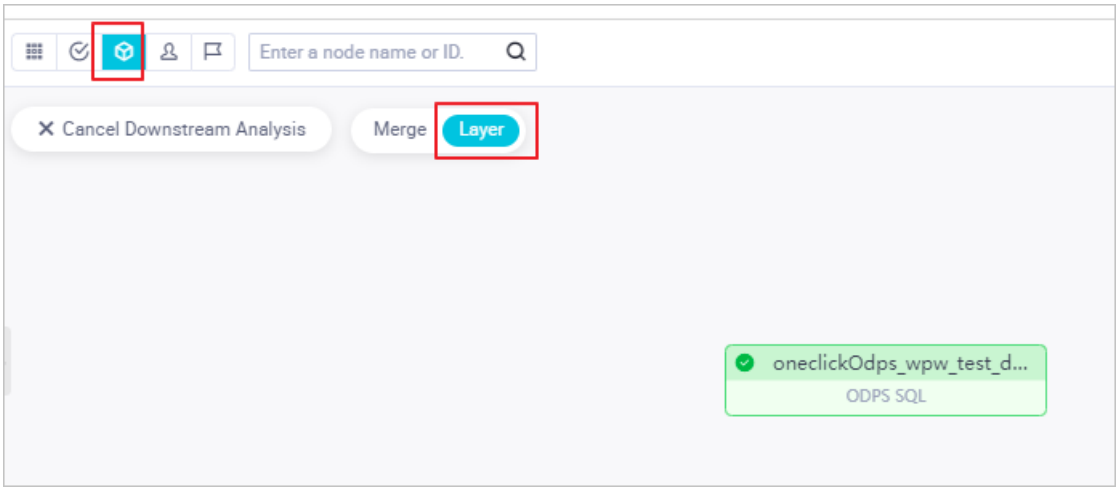
- Analyze descendant nodes

  If an auto triggered node has multiple descendant nodes or the descendant nodes of an auto triggered node are distributed at multiple levels, you can analyze the descendant nodes of the auto triggered node. You can aggregate the descendant nodes by node status, workspace, owner, or priority. Then, you can view the numbers of nodes at different levels from your required dimension or the total number of nodes at all levels from your required dimension.

> **Note**
>
> - By default, the descendant nodes of an auto triggered node are aggregated by owner. The system calculates the total number of nodes at all levels from the owner dimension.
> - If you analyze the descendant nodes of an auto triggered node, the analysis results are displayed by level. A maximum of six levels of nodes can be displayed. If you want to view more levels of nodes, click **Continue Analysis** in the upper-left corner.

In the following example, the descendant nodes of the **tag** node are analyzed. The following figures show the analysis results that are displayed by using different methods.

- The descendant nodes of the tag node are aggregated based on the workspaces to which the descendant nodes belong. The analysis results are displayed by level. This way, the numbers of the descendant nodes in different workspaces are displayed at different levels.



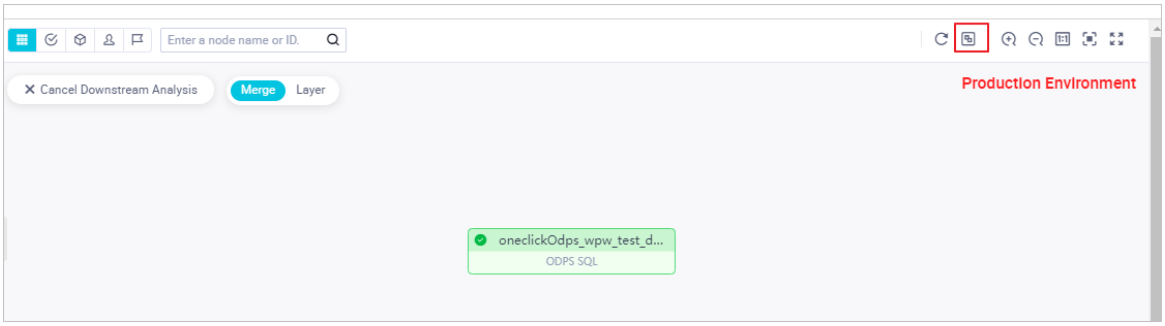

- The descendant nodes of the tag node are aggregated based on the workspaces to which the descendant nodes belong. The analysis results are displayed by using the merging method. This way, all the descendant nodes are placed at the same level, and the numbers of the descendant nodes that belong to different workspaces are displayed.


- Select a display pattern for a DAG

You can click the icons in the upper-right corner of a DAG panel to adjust the display pattern of the DAG based on your business requirements. For example, you can click **Toggle Full Screen View** or **Fit Screen** to perform the operation.

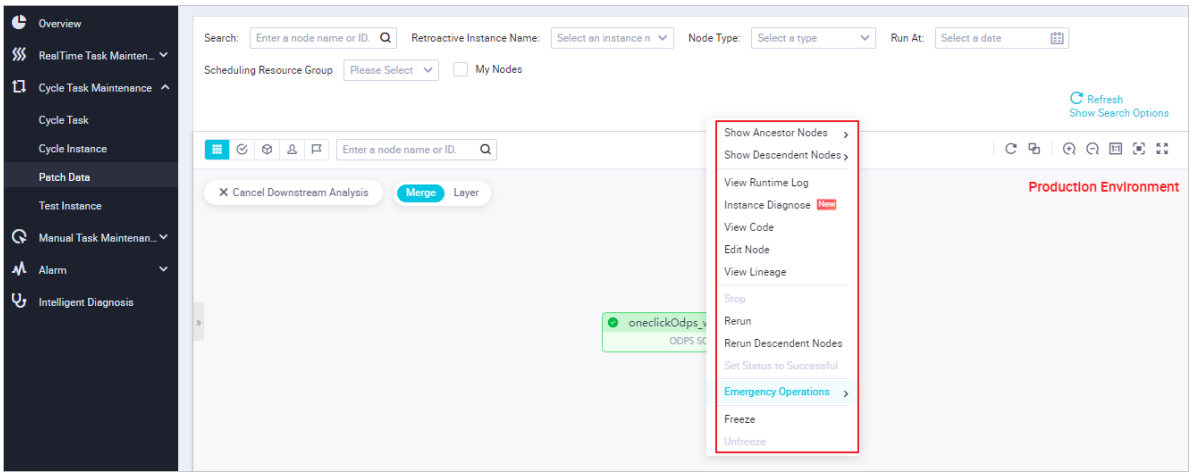
In the following examples, the DAG of the **0_2** node is displayed after the descendant nodes of the 0_2 node are ungrouped or grouped:

- The following figure shows the DAG of the 0_2 node when the descendant nodes of the 0_2 node are ungrouped. In this pattern, you can clearly view the upstream and downstream dependencies of all the nodes.




- The following figure shows the DAG of the 0_2 node when the descendant nodes of the 0_2 node are grouped. In this pattern, every five descendant nodes of the 0_2 node are placed at the same level. This way, these descendant nodes are displayed in an orderly manner. You can obtain the total number of the descendant nodes.


- Right-click the node that you want to manage in a DAG and perform operations on the node.




> **Note** After you click the **Update Node Status** icon in the upper-right corner, the DAG of the retroactive instance is refreshed, but the operational logs of the instance are not.

| Action | Description |
| --- | --- |
| **Show Ancestor Nodes or Show Descendant Nodes** | Show ancestor nodes or descendant nodes. If a workflow contains three or more nodes, specific nodes are automatically hidden in the DAG in Operation Center. You can select the number of levels to view all nodes at one or more levels. |
| **View Runtime Log** | View the operational logs of the node. A node may be in a state such as Running, Successful, or Failed. |
| **View Code** | View the code of the node. |
| **Edit Node** | Go to the **DataStudio** page to modify the configurations of the node. |
| **View Lineage** | View the dependencies of the node. |
| **Stop** | Stop the node if the node is in the **Pending** or **Running** state. After you perform this operation, the node enters the **Failed** state. |
| **Rerun** | Rerun the node if the node is in the Failed or an abnormal state. |
| **Rerun Descendent Nodes** | Rerun all the descendant nodes of the node. If the node has multiple descendant nodes, all these nodes are rerun. |
| **Set Status to Successful** | Set the status of the node to Successful and run descendant nodes that are not run. We recommend that you perform this operation if a node fails.<br>**Note** Only the status of a failed node can be set to successful. This operation does not apply to workflows. |
| **Emergency Operations** | Perform an emergency operation. An emergency operation takes effect only once on the current node.<br>Select **Delete Dependencies** to delete the dependencies of the node. You can perform this operation to start the node if the ancestor nodes of the node fail and the node does not depend on the data of the ancestor nodes. |
| **Freeze** | Freeze the node and pause the scheduling of the node. |
| **Unfreeze** | Resume the scheduling of the node if the node is frozen. |

# 1.4.5. Manage test instances

Test instances are generated when you test auto triggered nodes. You can manage test instances.

## Limits

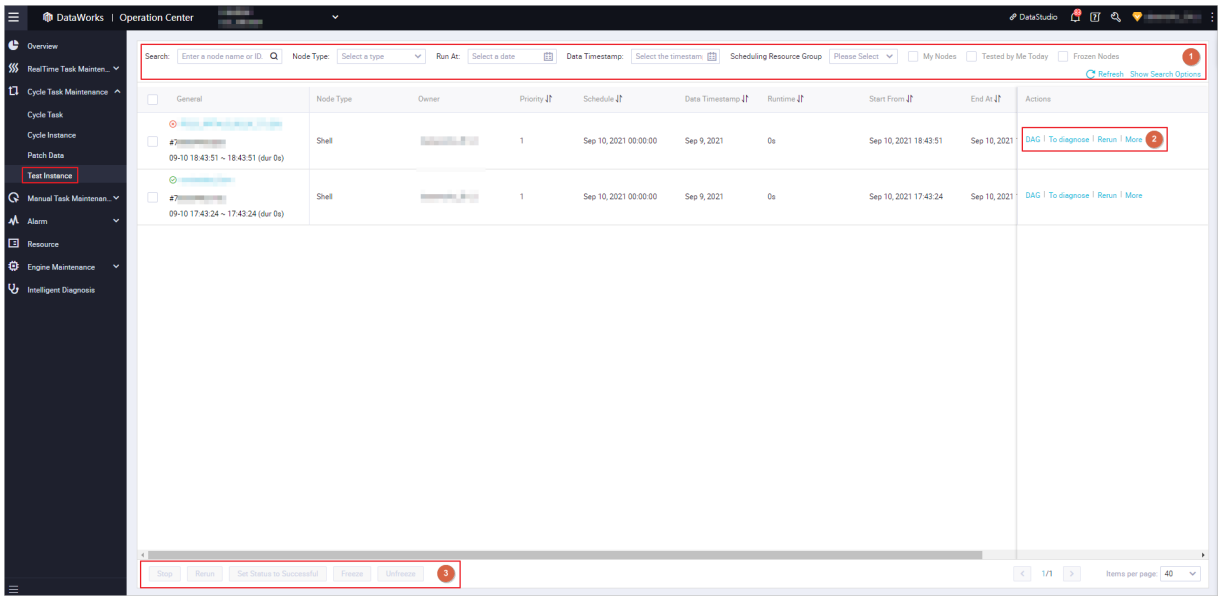
Test instances of a node can be run only after the running of an auto triggered node instance is complete.

## Go to the Test Instance page

1. Log on to the DataWorks console.
2. In the left-side navigation pane, click **Workspaces**.
3. Find your workspace and click **Data Analytics** in the Actions column.
4. On the DataStudio page, click the icon in the upper-left corner and choose **All Products > Operation Center**.
5. In the left-side navigation pane of the Operation Center page, choose **Cycle Task Maintenance > Test Instance**. Then, you can view test instances and the directed acyclic graphs (DAGs) of the test instances.

## Manage test instances in the instance list

You can manage test instances in the instance list. For example, you can rerun, freeze, or unfreeze test instances, set the status of test instances to successful, view the lineage of test instances, and check operational logs.

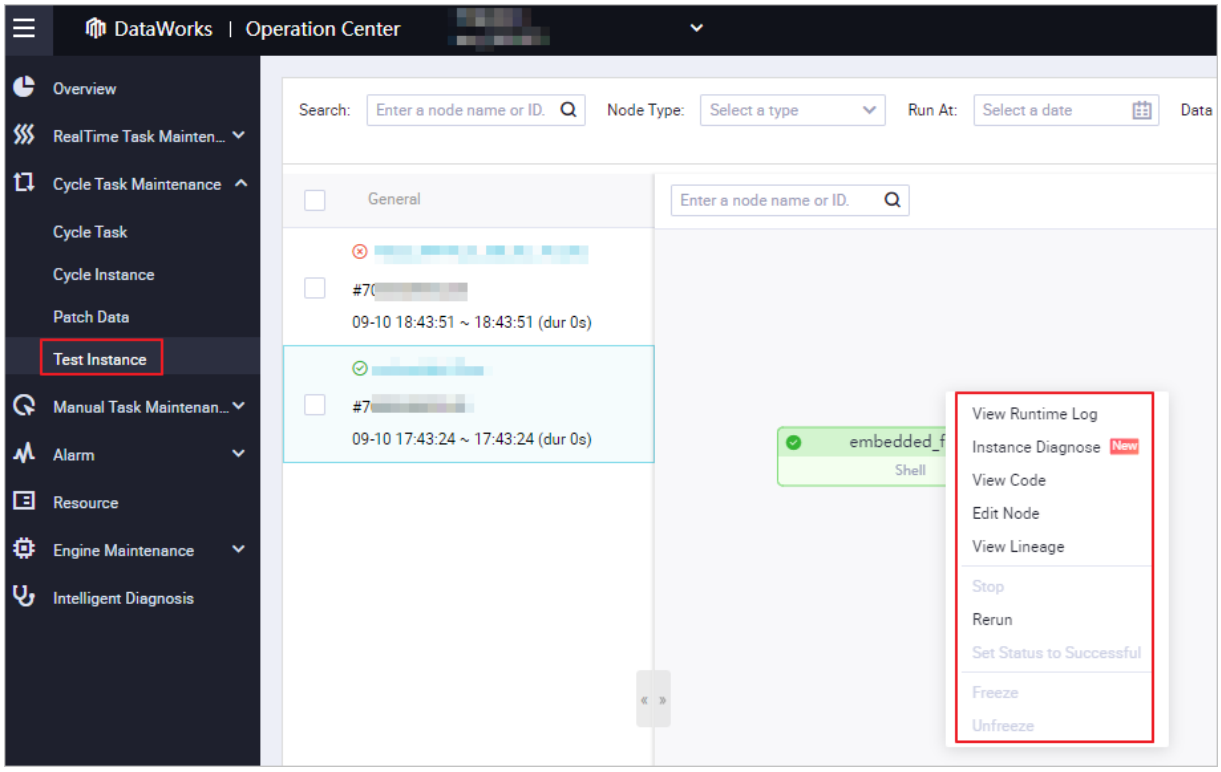



| Operation | Description |
| --- | --- |
| **Filter** | Allows you to specify filter conditions to search for your desired test instance in the section marked with 1 in the preceding figure.<br>You can search for a test instance by node name or node ID. You can also specify filter conditions, such as **Owner**, **Run At**, **Data Timestamp**, **Status**, **Region**, **Engine Type**, **Engine Instance**, **Baseline**, **My Nodes**, **Tested by Me Today**, and **Frozen Nodes**, to perform the operation. |
| **Stop** | Allows you to stop the test instance. You can stop a test instance only in the **Pending** or **Running** state. After you perform this operation, the test instance enters the **Failed** state. |

| Operation | Description |
| --- | --- |
| **Rerun** | Allows you to rerun the test instance. After the test instance is rerun, its pending descendant test instances are run as scheduled. Perform this operation if a test instance fails to be run or a test instance is not run as scheduled.<br>**Note** Only the test instances in the **Successful** or **Failed** state can be rerun. |
| **More** | Allows you to perform other operations on the test instance. You can click **More** in the Actions column of the test instance and select an operation. The operations include **Set Status to Successful**, **Freeze**, **Unfreeze**, **View Lineage**, and **View Runtime Log**. |
| **Batch operations** | Allows you to perform an operation on multiple test instances at a time. You can click one of the following buttons in the section marked with 3 in the preceding figure to perform an operation: **Stop**, **Rerun**, **Set Status to Successful**, **Freeze**, and **Unfreeze**. |

## Manage test instances in a DAG

Click the name of a test instance or **DAG** in the Actions column to view the DAG of the test instance. In the DAG, you can right-click the test instance to perform the related operations.




| Operation | Description |
| --- | --- |
| **View Runtime Log** | Allows you to view the operational logs of the test instance in a state such as Running, Successful, or Failed. |

| Operation | Description |
| --- | --- |
| **View Code** | Allows you to view the code of the test instance. |
| **Edit Node** | Allows you to go to the **DataStudio** page to modify the node to which the test instance belongs. |
| **View Lineage** | Allows you to view the lineage of the test instance. |
| **Stop** | Allows you to stop the test instance. You can stop a test instance only in the **Pending** or **Running** state. After you perform this operation, the test instance enters the **Failed** state. |
| **Rerun** | Allows you to rerun the test instance. After the test instance is rerun, its pending descendant test instances are run as scheduled. Perform this operation if a test instance fails to be run or a test instance is not run as scheduled.<br>**Note** Only the test instances in the **Successful** or **Failed** state can be rerun. |
| **Set Status to Successful** | Allows you to set the status of the test instance to **Successful** and run its pending descendant test instances. Perform this operation if a test instance fails to be run.<br>**Note** This operation applies only to **failed** test instances. |
| **Freeze** | Allows you to freeze the test instance and pause the scheduling of the test instance. |
| **Unfreeze** | Allows you to unfreeze the frozen test instance.<br>• If the test instance is not run, the system automatically runs this test instance after its ancestor test instances are successfully run.<br>• If all the ancestor test instances of the test instance are successfully run, the status of the test instance is directly set to Failed. You must manually rerun the test instance. |

# 1.4.6. Instance diagnosis

DataWorks Operation Center provides the instance diagnosis feature to help you track the running of nodes and identify problems.

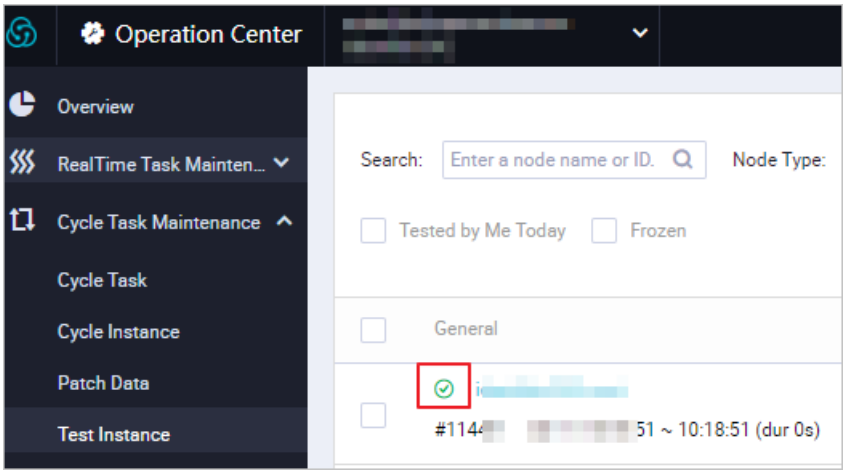
## Go to the Intelligent Diagnosis page

**Notice** The instance diagnosis feature is available only in DataWorks Professional Edition or more advanced editions.

1. Log on to the DataWorks console.
2. In the left-side navigation pane, click **Workspaces**.
3. After you select the region in which the workspace that you want to manage resides, find the workspace and click **Data Analytics** in the Actions column.

4. Click the icon in the upper-left corner and choose **All Products > Task Operation > Operation Center**.
5. In the left-side navigation pane, choose **Cycle Task Maintenance > Cycle Instance**.

   You can also click **Patch Data** or **Test Instance** or choose **Manual Task Maintenance > Manual Instance** to go to the **Intelligent Diagnosis** page of an instance in the same way.
6. Click the status icon next to an instance to go to the **Intelligent Diagnosis** page.




You can also go to the **Intelligent Diagnosis** page of an instance in one of the following ways:

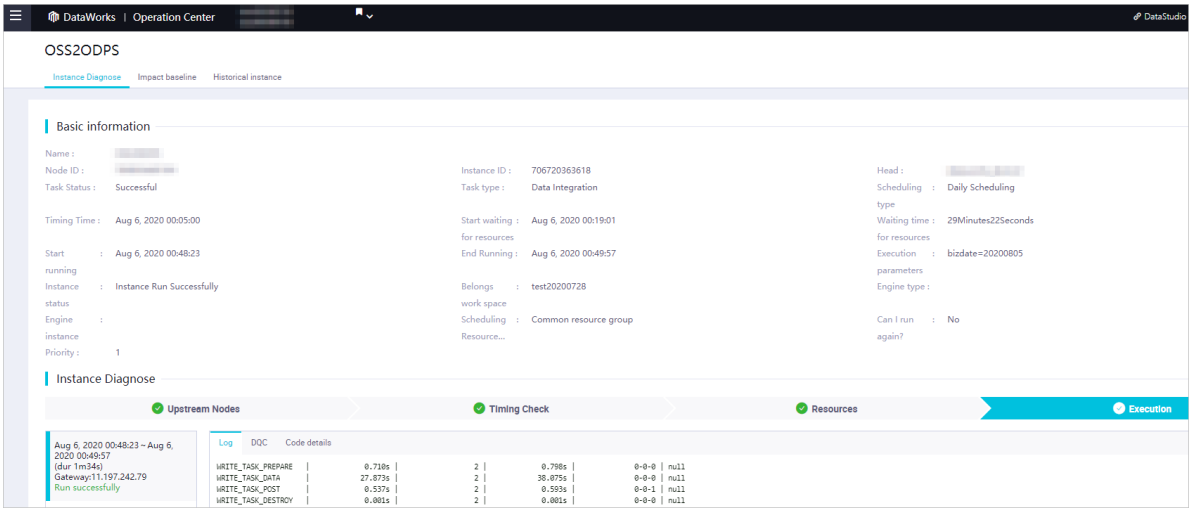
- Right-click the instance in the directed acyclic graph (DAG) and select **View Runtime Log**. On the log details page, click **Use Instance Diagnose now**.
- Right-click the instance in the DAG and select **Instance Diagnose**.
- Show the instance list. Find an instance and choose **More > Instance Diagnose**.

## Instance Diagnose

The instance diagnostic information of an instance is displayed in the **Upstream Nodes**, **Resources**, and **Execution** steps.

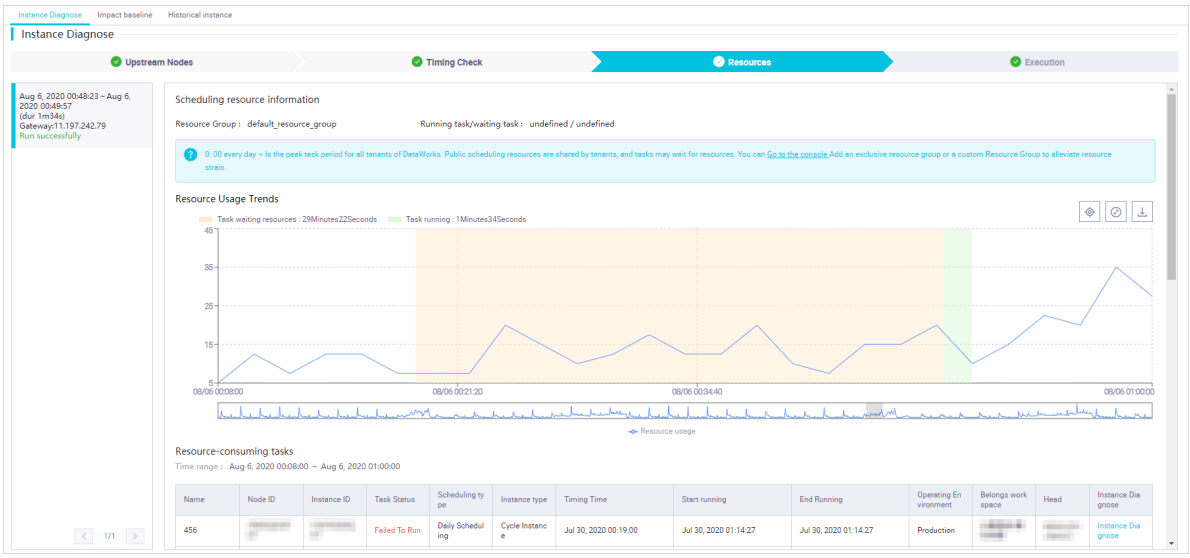
- **Upstream Nodes**

The **Upstream Nodes** step on the Instance Diagnose tab displays the running status of ancestor nodes of the current node. This tab helps you find the ancestor node that blocks the running of the current node.



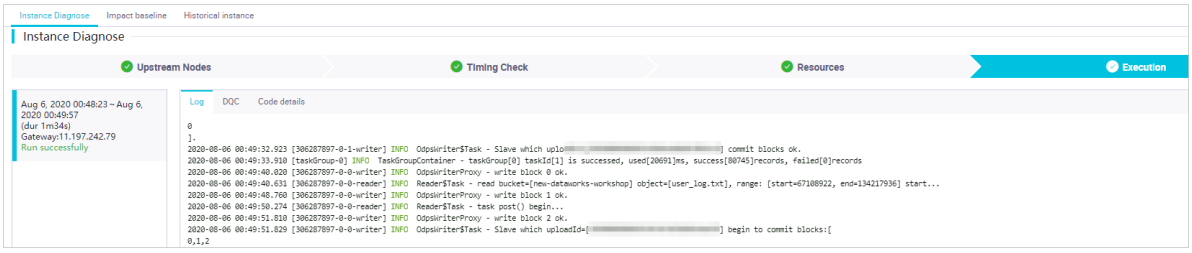

- Resources

  The **Resources** step on the Instance Diagnose tab displays the resource usage. Based on the information on this tab, you can arrange the scheduled time of the current node to avoid peak hours.




- Execution

  If the current node fails to run, the **Execution** step provides you with intelligent diagnostic suggestions based on log information. You can also click **Instance List** to conduct further analysis.



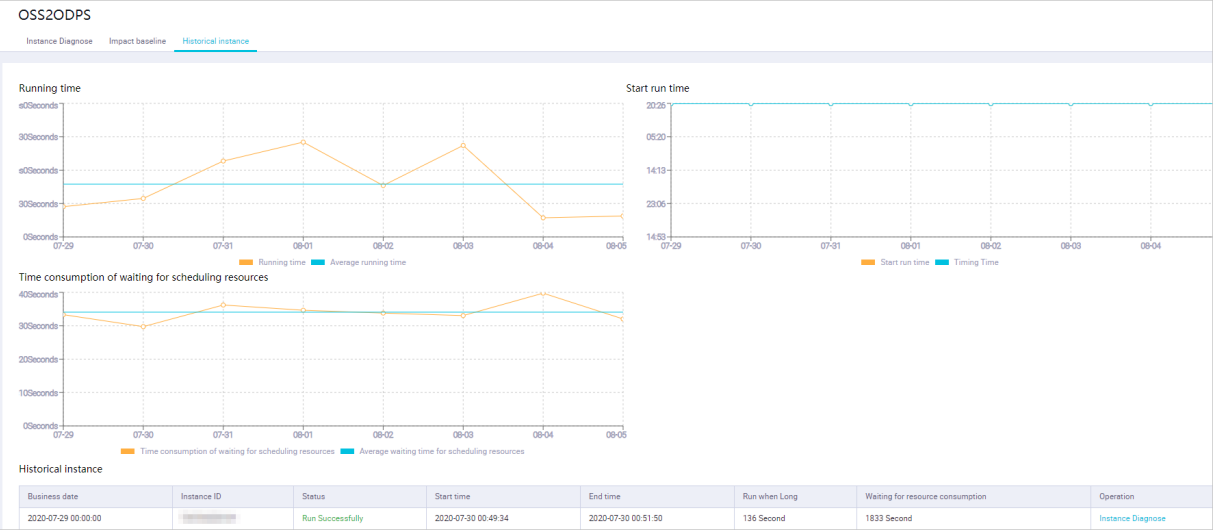
> **Notice** The **Instance List** tab is available only for MaxCompute nodes.

### Impact baseline

In the top navigation bar, click **Impact baseline** to view the baseline of the current node.

### Historical instance

In the top navigation bar, click **Historical instance** to view the information about the last 15 instances of the current node. The information includes **Running time**, **Start run time**, **Time consumption of waiting for scheduling resources**, and **Historical instance**.




# 1.5. Manually triggered node O&M

## 1.5.1. Manually triggered nodes

Manually triggered nodes are nodes whose scheduling type is set to manual before these nodes are committed to the scheduling system.

> **Note**
>
> - After a manually triggered node is committed to the scheduling system, the node is run only after it is manually triggered.
> - Manually triggered nodes created in DataWorks V1.0 are listed as the **Manually Triggered Node** type, and those created in DataWorks V2.0 or later are listed as the **Manually Triggered Workflow** type.

### Manage manually triggered workflows in the workflow list

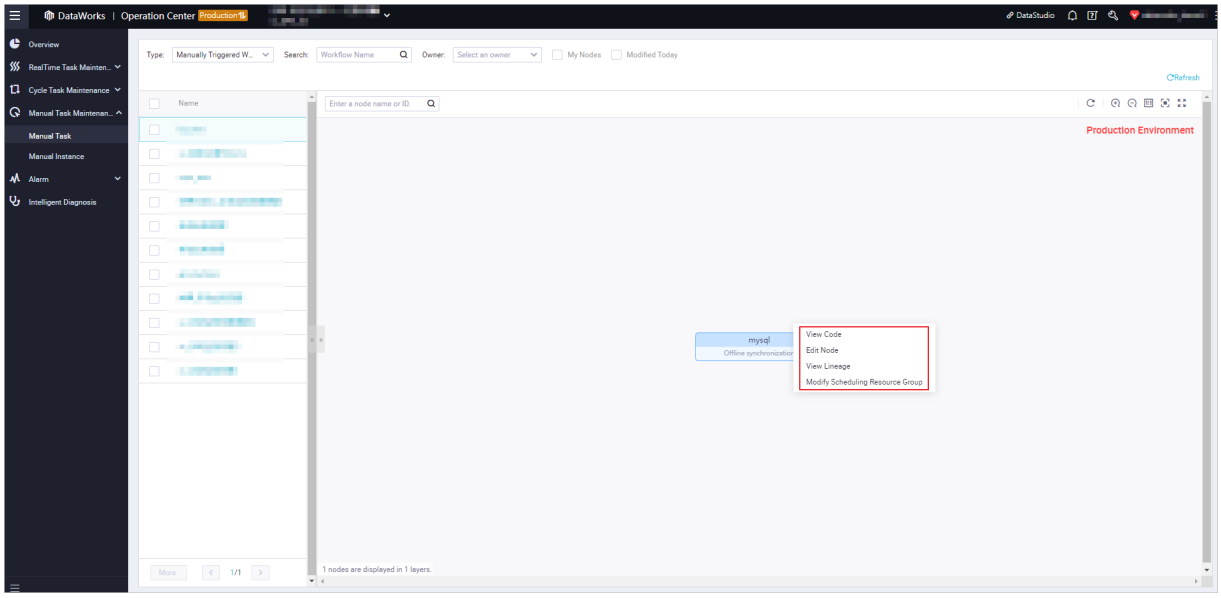
The manually triggered workflow list displays manually triggered workflows that are committed.

| Action | Description |
| --- | --- |

| Action | Description |
| --- | --- |
| **Filter** | Allows you to find the required workflows by configuring parameters in the red box that is marked with 1 in the preceding figure.<br>You can search for a workflow based on the **workflow name** and configurations of **Type**, **Owner**, **My Nodes**, and **Modified Today**.<br>**Note** Only the workflows that meet the workflow name filter condition and other filter conditions are returned. |
| **DAG** | Allows you to view the directed acyclic graph (DAG) of the workflow. You can view the workflow information, such as the code and lineage, in the DAG. |
| **Run** | Allows you to run the workflow to generate manually triggered node instances. |
| **View Instances** | Allows you to go to the **View Instances** page to view the manually triggered node instances that are generated for the workflow. |
| **More** | Allows you to change the owner of the workflow. You can choose **More > Change Owner** to perform the operation. |
| **Change Owner, Modify Scheduling Resource Group, Undeploy Manually Triggered Workflow** | To change the owners of multiple workflows at a time, select multiple workflows and click **Change Owner** as shown in the red box that is marked with 3 in the preceding figure. |

## DAG of a manually triggered workflow

To view the DAG of a workflow, click the name of the workflow or **DAG** in the Actions column. In the DAG, you can right-click the workflow to perform related operations.




| Action | Description |
| --- | --- |
| **View Code** | Allows you to view the code of the workflow. |
| **Edit Node** | Allows you to go to the **DataStudio** page to modify the workflow. |
| **View Lineage** | Allows you to view the lineage of the workflow. |
| **Modify Scheduling Resource Group** | Allows you to change the resource group that the workflow uses to run. |

# 1.5.2. Manually triggered node instances

Manually triggered node instances are generated when manually triggered nodes are run. Manually triggered nodes do not have node dependencies. They must be manually run.

> **Notice**
>
> - Alerts can be generated only when auto triggered node instances fail. Alerts are not generated for manually triggered node instances, retroactive instances, or test instances.
> - You must purchase DataWorks Professional Edition or a more advanced edition to use the instance diagnostics feature for manually triggered node instances.

## Go to the Manual Instance page

1. Log on to the DataWorks console.
2. In the left-side navigation pane, click **Workspaces**.
3. Find the required workspace and click **Data Analytics**.
4. Click the icon in the upper-left corner and choose **All Products > Operation Center**.

5. In the left-side navigation pane, choose **Manual Task Maintenance > Manual Instance**. You can view the list and directed acyclic graphs (DAGs) of manually triggered workflows or manually triggered node instances.

### Manually triggered workflow list

| Operation | Description |
| --- | --- |
| **Filter** | Find the required workflows by setting the filter conditions in the section marked with 1 in the preceding figure.<br>You can search for workflows by workflow name and set filter conditions such as **Type**, **Owner**, **Data Timestamp**, and **Run At**. |
| **DAG** | Click **DAG** in the Actions column to open the DAG of the workflow. You can view the running result of the instances. |
| **Stop** | Click **Stop** to stop the instance if it is in the Running state. |
| **Rerun** | Reschedule the instance. |
| **Batch operations** | Select multiple instances and click **Stop** in the section marked with 3 in the preceding figure to stop the selected instances at a time. |

### Manually triggered workflow DAG

Click the name of a workflow or **DAG** in the Actions column to open the DAG of the workflow. In the DAG, you can right-click the workflow to perform related operations.

> **Note** A manually triggered workflow does not have dependencies. Therefore, only the current instance appears in the DAG.

| Operation | Description |
| --- | --- |
| **View Runtime Log** | Allow you to view the operational logs of the current instance if it is in the Running, Successful, or Failed state. |
| **View Code** | Allow you to view the code of the node. |
| **Edit Node** | Click this menu item to go to the **DataStudio** page to modify the workflow. |
| **View Lineage** | Allow you to view the lineage of the current instance. |
| **Stop** | Stop the instance, which takes effects only on the current instance. |
| **Rerun** | Rerun the instance if it is in the Failed state or an abnormal state. |

# 1.6. Monitor

## 1.6.1. Overview

The intelligent monitoring system monitors the status of nodes. If the global rules, custom alert rules, and intelligent baselines that you configured are met, the intelligent monitoring system sends alert notifications based on the notification method that you specified, such as emails, text messages, phone calls, DingTalk group messages, and webhook addresses. This way, you can identify and handle exceptions at the earliest opportunity.

## Background information

In addition to the basic monitoring operations that are supported by a traditional monitoring system, DataWorks provides the following capabilities:

- Accurately identifies the nodes that you want to monitor based on your business requirements.

  DataWorks runs a large number of nodes, and the dependencies between the nodes are complex. This makes it difficult to find and monitor all the ancestor nodes of a node, even if you know which node is the most important. If you monitor all nodes, a large number of alerts are triggered. This causes you to miss the alerts of the nodes that you want to monitor, which leads to poor monitoring performance.

- Provides different alerting methods for monitored nodes.

  For example, some monitoring tasks require the monitored nodes to run for more than 1 hour before alerts are triggered. Other monitoring tasks require the monitored nodes to run for more than 2 hours before alerts are triggered. A long period of time is required to configure alert rules for each node, and it is difficult to determine the alert threshold for each node.

- Provides a switch to disable alerting.

  The intelligent monitoring system provides a switch to disable alerting.

The **intelligent monitoring** system provides comprehensive monitoring and alerting logic. You need only to specify the names of important nodes in your business. Then, the intelligent monitoring system automatically monitors the entire process of your nodes and generates standard alert rules for them. To create custom alert rules, you can configure the basic alert rule settings.

The full-path monitoring feature of the intelligent monitoring system ensures the overall data output of the important business within Alibaba Group. The intelligent monitoring system allows you to analyze ancestor and descendant node paths to identify risks and provide O&M information for business departments. The features that are provided by the intelligent monitoring system ensure the high stability of business within Alibaba Group.

## Limits

The following features are supported only in DataWorks Standard Edition and more advanced editions: **Baseline Instance**, **Baseline Management**, and **Event Management**.

## Considerations

If you want to receive alert notifications by text messages or phone calls as a RAM user, you must log on to the RAM console by using your Alibaba Cloud account. Then, enter the required information about this RAM user such as the phone number and email address. For more information about how to enter the personal information of a RAM user, see Modify the basic information about a RAM user.

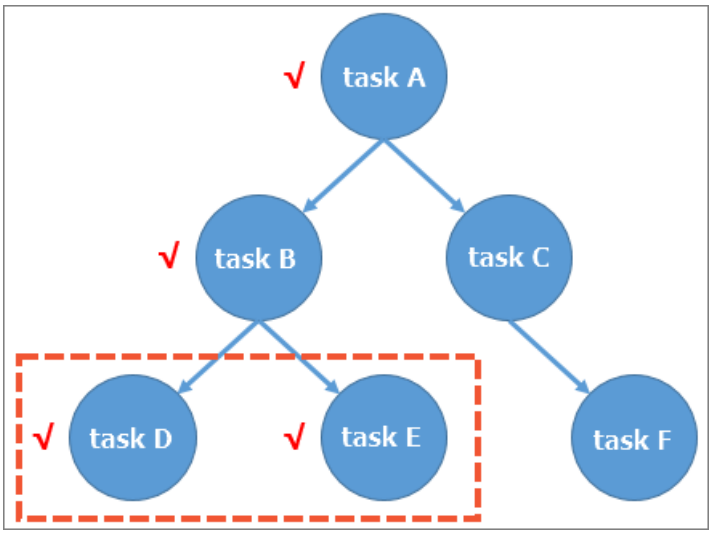
## Monitoring methods

The intelligent monitoring system provides the following monitoring methods: intelligent baseline and custom alert rules. This section describes the monitoring principles and configuration rules of the two monitoring methods.

- Intelligent baseline
  - Monitoring scope

    A baseline is a group of nodes. You can monitor nodes by baseline.

    After you configure a baseline as a monitored object, all nodes in the baseline and the ancestor nodes of the nodes are monitored. After you configure a monitored baseline, the intelligent monitoring system does not monitor all nodes. A descendant node of a monitored node is monitored only if the descendant node is added to the monitored baseline. If the descendant node is not added to the monitored baseline, the intelligent monitoring system does not report an alert even if the descendant node fails.

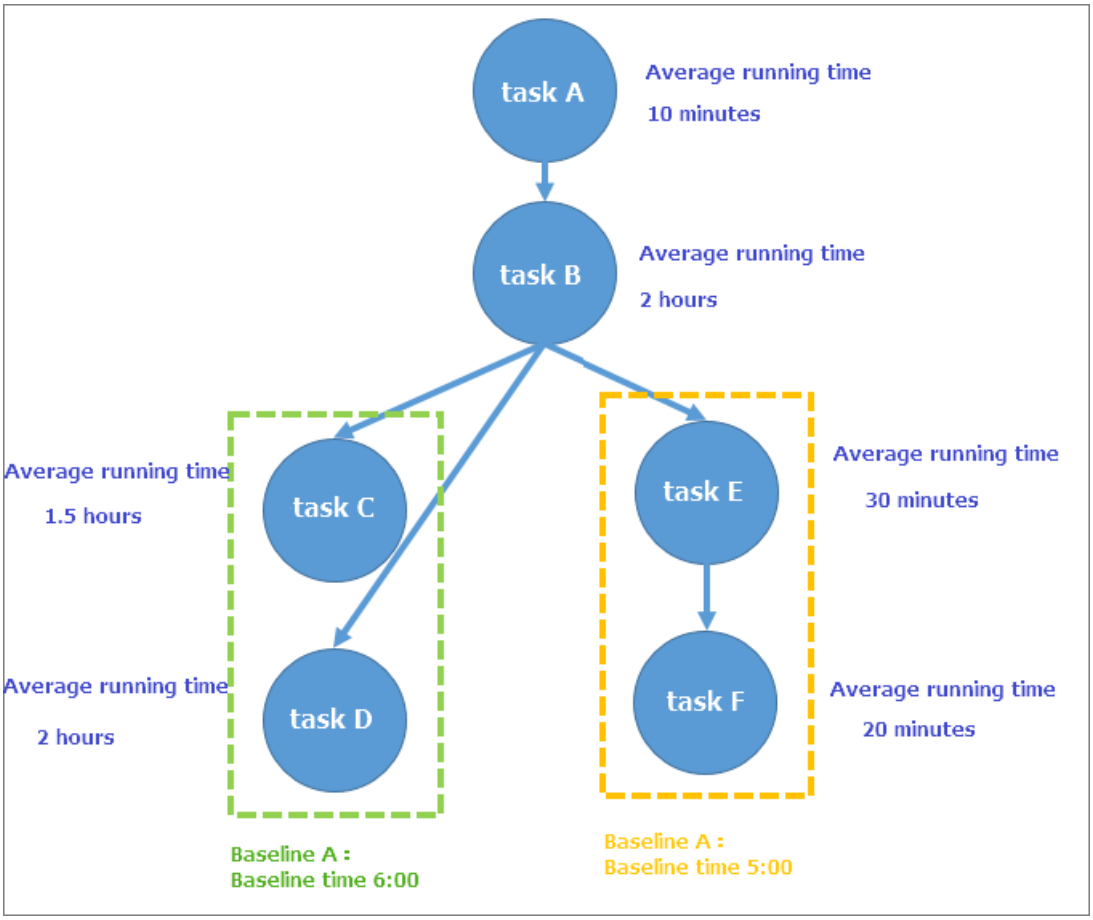
    In the preceding figure, DataWorks has six nodes, but only Nodes D and E are added to the monitored baseline. The intelligent monitoring system monitors Nodes D and E and all the ancestor nodes of these two nodes. If an exception such as an error or slowdown occurs on Node A, B, D, or E, the intelligent monitoring system detects the exception. However, the intelligent monitoring system does not monitor Node C or F.
  - Node capturing

    If an exception occurs on a node that is within the specified monitoring scope, the intelligent monitoring system generates an alert event and reports an alert based on the analysis of the alert event. Two types of node exceptions are monitored. You can go to the Event Management page to view the node exceptions. For more information, see Manage events.

    - **Error**: indicates that a node fails to run.
    - **Slow**: indicates that the running time of a node is significantly longer than the average running time of the node in the previous periods.

    > **Note** If a node times out and then encounters an error, two events are generated.
  - Alerting time judgment

**Margin** is the maximum period of time that the intelligent monitoring system allows you to wait before you start to run a node. `The latest start time of a node is calculated by using the following formula: Baseline time - Average running time` .




In the preceding figure, to make sure that the baseline time of Baseline A is 05:00, you must set the latest start time of Node E to 04:10. The time 04:10 is obtained by subtracting the average running time of Node F and the average running time of Node E from the baseline time 05:00. In this example, the average running time of Node F is 20 minutes, and the average running time of Node E is 30 minutes. This time is the latest completion time of Node B in Baseline A.

In the preceding figure, to make sure that the baseline time of Baseline B is 06:00, you must configure the latest completion time of Node B to 04:00. The time 04:00, which is earlier than 04:10, is obtained by subtracting the average running time of Node D from the baseline time 06:00. In this example, the average running time of Node D is 2 hours. To meet the baseline time of both Baseline A and Baseline B, you must set the latest completion time of Node B to 04:00.

The latest completion time of Node A is 02:00, which is the average running time of Node B subtracted from 04:00. In this example, the average running time of Node B is 2 hours. The latest start time of Node A is 01:50, which is the average running time of Node A subtracted from 02:00. In this example, the average running time of Node A is 10 minutes. If Node A fails to run before 01:50, Baseline A may enter the Overtime state. If Node A fails to run at 01:00, the margin of Node A is 50 minutes, which is the difference between 01:00 and 01:50. As shown in this example, **margin** indicates the degree of caution for a node exception.

- Baseline alerting

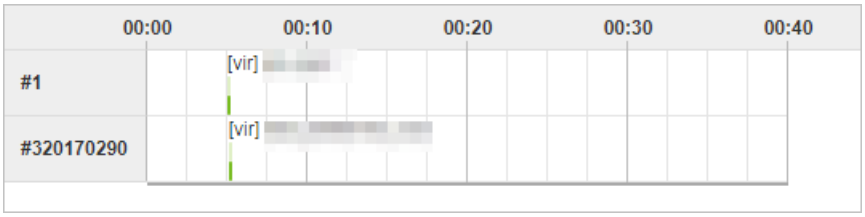
  The baseline alerting feature provides alerts for baselines for which baseline monitoring is enabled. You must set the **Margin Threshold** and **Committed Time** parameters for each baseline. When baseline alerting is enabled, the specified alert recipient is notified three times at intervals of 30 minutes when the baseline completion time that is estimated by the intelligent monitoring system exceeds the alert margin.

- Notification method

  By default, baseline alerts are sent to the baseline owner. You can change the notification method based on your business requirements. To perform the change, go to the **Rule Management** page, find **Global Baseline Alert Rule**, and then click View Details in the Actions column. For more information, see Configure alert details.

- Gantt chart

  The Gantt chart feature specifies the key path of a node. The Gantt chart feature is provided by the **baseline instances** of the intelligent monitoring system.

- Custom alert rules

  To create a trigger for a custom alert, configure the following parameters based on your business requirements:

  - Object: You can specify nodes, baselines, workspaces, and workflows as objects.
  - Trigger Condition: Valid values are Completed, Uncompleted, Error, Uncompleted in Cycle, Overtime, and The error persists after the node automatically reruns.
  - Notification Method: Valid values are SMS, Email, and Phone.
  - Maximum Alerts: the maximum number of times that alerts can be reported. If the maximum number is exceeded, no alerts are reported.
  - Minimum Alert Interval: the time interval at which DataWorks reports alerts.
  - Quiet Hours: the period of time during which no alerts are reported.
  - Recipient: You can set this parameter to the node owner or another recipient.

  The following information shows the trigger conditions for custom alert rules:

- **Completed**

  You can configure an alert rule for nodes, baselines, workspaces, and workflows. If all nodes of the specified object are completed within the specified time, an alert is reported.

  > **Note**
  > - If you configure an alert rule for a baseline, an alert is reported when all nodes of the baseline are completed.
  > - If you configure an alert rule for a workflow, an alert is reported when all nodes of the workflow are completed.
  > - If you configure an alert rule for multiple nodes, an alert is reported when all the configured nodes are completed.
  > - You cannot configure such an alert for a workspace.

- **Uncompleted**

  You can configure an alert rule for nodes, baselines, workspaces, and workflows. If one node of the specified object is not completed within the specified time, an alert is reported. The intelligent monitoring system starts to monitor the object that you specify in the Create Custom Rule dialog box from the time at which the object starts to run until the time specified by the **Alert At** parameter in the same dialog box.

  > **Note** If you configure an alert rule for a node that is scheduled by hour or minute, an alert is reported if one of the instances of the node is not completed within the time specified by the Alert At parameter on the current day. Example:
  > - You set Trigger Condition to Uncompleted and Alert At to 10:00 for a baseline. If a node in the baseline is not completed at 10:00, an alert that contains a list of the uncompleted nodes is reported.
  > - If a node scheduled by hour is configured to run on an hourly basis from 00:00 to 23:59 every day and you set the Alert At parameter to 12:00, an alert is triggered every day. For a node that is scheduled by hour or minute, we recommend that you set Trigger Condition to Uncompleted in Cycle in the Create Custom Rule dialog box. This way, an alert is reported if one of the instances of the node scheduled by hour or minute is not completed within the specified cycle.

- **Error**

  You can configure an alert rule for nodes, baselines, workspaces, and workflows. If an error is reported for a node, an alert that contains detailed information about the error is sent to the recipient.

  > **Note**
  > - If you configure an alert rule for a baseline, workspace, or workflow, an alert is reported when a node in the baseline, workspace, or workflow fails to run.
  > - After you set Trigger Condition to Error, an alert is reported every time the node fails. If the node fails to rerun three times, the alert is triggered three times.

- **Uncompleted in Cycle**

  If node instances continue to run at the end of the specified cycle, an alert is reported. In most cases, you can configure this alert rule for node instances that are scheduled by hour or minute. For example, Node A is scheduled to run every 2 hours, and the time to complete each run is 25 minutes. If Node A starts to run at 00:00 every day, Node A runs 12 times within 24 hours. The first cycle starts at 00:00, the second cycle starts at 02:00, and so on. The twelfth cycle starts at 22:00. If Node A runs as expected, the running of Node A is complete at the specified point in time in each cycle, such as 00:25 or 02:25. If the node continues to run at the end of a cycle, an alert is reported.

  > **Note** By default, frozen instances are in the Completed state.

- **Overtime**

  You can configure alerts for nodes, baselines, workspaces, and workflows. An alert is reported if the execution duration of a node of the monitored object exceeds the specified execution duration and the node status is unsuccessful.

  > **Note** The execution duration is specified by the **Timer** parameter in the Create Custom Rule dialog box.

- **The error persists after the node automatically reruns**

  You can configure an alert rule for nodes, baselines, workspaces, and workflows. If an error persists after a node of the monitored object automatically reruns, an alert is reported.

## Features

To reduce your configuration costs and highlight the alerts of important nodes that you want to monitor, the intelligent monitoring system provides the following features:

- **Baseline Instance**: allows you to view the information about a baseline. For more information, see Manage baseline instances.
- **Baseline Management**: allows you to create and define a baseline. For more information, see Manage baselines.
- **Event Management**: allows you to view all alert events that are related to slowdown or errors. For more information, see Manage events.
- **Rule Management**: allows you to create custom alert rules, add a DingTalk chatbot, and obtain a webhook URL. The intelligent monitoring system monitors the status of your nodes based on your custom alert rules. This way, you can identify and handle exceptions at the earliest opportunity. For more information, see Manage custom alert rules.
- **Alert Management**: allows you to view details about all alerts. This way, you can identify and handle exceptions at the earliest opportunity. For more information, see Manage alerts.
- **Schedule**: allows you to use the shift schedule feature of DataWorks to specify a shift schedule for alert recipients during the O&M of nodes. DataWorks sends an alert notification to a recipient of a custom alert rule based on the shift schedule that you specify. This way, the alert recipient can identify and handle exceptions at the earliest opportunity. For more information, see Create and manage a shift schedule.

For more information on frequently asked questions about the intelligent monitoring system, see FAQ.

# 1.6.2. Intelligent Baseline

## 1.6.2.1. Manage baselines

On the Baseline Management page, you can create and manage baselines.

> **Note** To use this feature, you must purchase DataWorks Standard Edition or higher.

### Create a baseline

1. Log on to the DataWorks console. In the left-side navigation pane, click Workspaces. On the Workspaces page, find the target workspace and click **Data Analytics** in the Actions column.
2. On the DataStudio page that appears, click the icon in the upper-left corner and choose **All Products > Operation Center**.
3. On the Operation Center page that appears, choose **Alarm > Baseline Management**.
4. On the **Baseline Management** page, click **Create Baseline** in the upper-right corner.

   > **Note** Currently, only the workspace administrator can create baselines.

5. In the **Create Baseline** dialog box that appears, set the parameters and click **OK**.

| Parameter | Description |
| --- | --- |
| **Baseline Name** | The name of the baseline. |
| **Workspace** | The workspace of the node associated with the baseline. |
| **Owner** | The name or ID of the owner. |
| **Recurrence** | Specifies whether the baseline detects nodes by day or hour.<br>○ **By the Day Interval**: Select this option for daily scheduled nodes.<br>○ **By the Hour Interval**: Select this option for hourly scheduled nodes. |
| **Node** | ○ **Node**: the node to be associated with the baseline. Enter the name or ID of a node and click the icon on the right to add the node. You can add multiple nodes.<br>○ **Workflow**: the workflow to be associated with the baseline. Enter the name or ID of a workflow and click the icon on the right to add the workflow.<br>**Note** We recommend that you add only the node at the end of a workflow. We recommend that you do not add all nodes of a workflow by adding the workflow. |

| Parameter | Description |
| --- | --- |
| **Priority** | The priority of the baseline. A baseline with a higher priority is scheduled preferentially. Default value: 1.<br>**Note**<br>○ You can only change the priority of nodes in a workspace whose billing method is subscription. Valid values: 1, 3, 5, 7, and 8.<br>○ The larger the value, the higher the priority. |
| **Estimated Completion Time** | The completion time of the node estimated based on the average running time of the node during previous scheduling. If no historical data is available, the message **The completion time cannot be estimated due to a lack of historical data** appears. |
| **Committed Time** | The time point when the node should be completed. An alert is triggered if the node is not completed until the time point obtained by subtracting the alert margin threshold from the committed completion time. |
| **Margin Threshold** | The interval before an alert is triggered. For example, if you set Committed Time to 3:30 and Margin Threshold to 10 minutes, an alert is triggered if the node is not completed at 3:20. Assume that the average running time of the node is 30 minutes. If the node is not started at 2:50, an alert is triggered.<br>**Note** The average running time of a node can be calculated based on the data of the last 15 days. |

6. After a baseline is created, click **Enable** in the Actions column to enable the baseline.

   You can click **View Details**, **Change**, **Enable**, Disable, or **Delete** in the Actions column to perform the corresponding operation on a baseline.

   - **View Details**: Click **View Details** to view the basic information about the baseline.
   - **Change**: Click **Change** to modify the baseline.
   - **Enable** or Disable: Click Enable or Disable to enable or disable the baseline. A baseline instance can be generated only when the corresponding baseline is enabled.
   - **Delete**: Click **Delete** to delete the baseline.

## Add a node to a baseline

By default, all nodes in the production environment are in the default baseline of each workspace. When you create a baseline, you actually move nodes from the default baseline to the baseline that you create.

> **Note** A node must belong to a baseline, so you cannot directly remove nodes from the default baseline. Instead, you can create a baseline to move nodes from the default baseline to the new baseline. When you delete a baseline that you create, you actually move the nodes in the baseline to the default baseline.

To change the baseline of a node, perform one of the following operations:

- On the **Baseline Management** page, click **Create Baseline** in the upper-right corner. Then, create a baseline by following the instructions in the "Create a baseline" section.
- In the left-side navigation pane, choose Cycle Task Maintenance > **Cycle Task**. On the page that appears, find the node and choose **More > Add to Baseline** in the Actions column.

# 1.6.2.2. Manage baseline instances

You can manage baseline instances on the Baseline Instance page.

> **Note**
>
> - To use this feature, you must purchase DataWorks Standard Edition or a more advanced edition.
> - After a baseline is created, you must enable the baseline. This way, instances can be generated for the baseline.

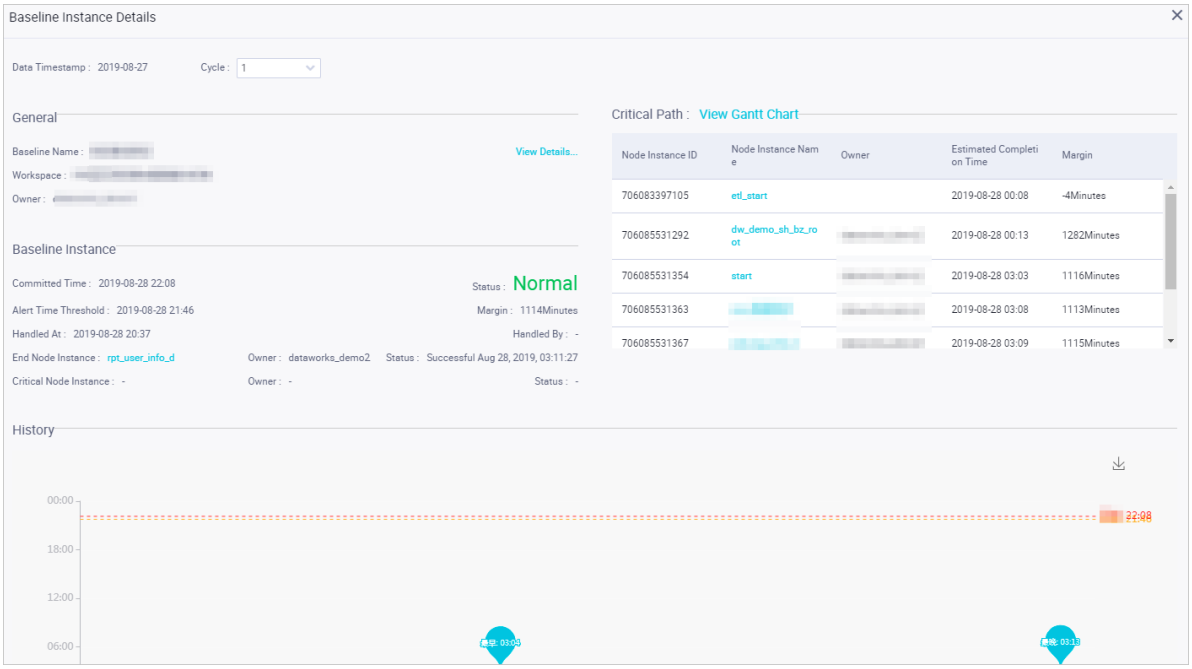
On the **Baseline Instance** page, you can search for baseline instances based on the configurations of the **Data Timestamp**, **Owner**, **Event ID**, **Workspace**, **Baseline Name**, **Recurrence**, **Priority**, and **Baseline Status** parameters. You can click **View Details**, **Handle**, or **View Gantt Chart** in the Actions column that corresponds to a specific baseline instance to perform the related operation.

You can filter baseline instances based on the status by select one or more of the following options:

- **Normal**: indicates that all the nodes associated with the baseline finish execution before the alerting time.
- **Alerting**: indicates that one or more nodes associated with the baseline do not finish execution after the alerting time but before the committed time.
- **Overtime**: indicates that one or more nodes associated with the baseline do not finish execution after the committed time.
- **Others**: indicates that all nodes in the baseline are paused or the baseline is not associated with any node.

You can click **View Details**, **Handle**, or **View Gantt Chart** in the Actions column as required.

- **View Details**: Click **View Details** to view the details about the baseline instance in the **Baseline Instance Details** dialog box.




In the Baseline Instance Details dialog box, you can view the information in the following sections: **General**, **Critical Path**, **Baseline Instance**, **History**, and **Events**.

> **Note**
> - The data timestamp is one day earlier than the system date.
> - You can set the Cycle parameter only for hourly baselines.

- **Handle**: Click Handle to handle the alert that is generated for the baseline instance. During this period, alerting is paused.
- **View Gantt Chart**: Click **View Gantt Chart** to view the critical paths of the nodes that are associated with the baseline instance.

FAQ: Why is a baseline in the Empty Baseline state?

In the following scenarios, a baseline instance may be in the Empty Baseline state:

- Scenario 1: A node can belong to only a baseline. When you add a node to another baseline, you actually move the node from the existing baseline to another one. If all nodes are removed from the existing baseline, the status of the existing baseline changes to Empty Baseline.
- Scenario 2: On the day when a baseline is created, the baseline is in the Empty Baseline state. After you enable the baseline, a baseline instance is generated on the next day.
- Scenario 3: You specify an invalid point in time for an auto triggered node instance in the hourly baseline.

> **Note** For example, you specify that the node is run at 6:00 and 18:00 every day. However, you specify that the baseline detects the node at 6:00 and 18:00.

# 1.6.2.3. Manage events

On the Event Management page, you can view all events related to slowdown or errors.

> **Note** To use this feature, you must purchase DataWorks Standard Edition or higher.

Go to the **Operation Center** page. In the left-side navigation pane, choose **Alarm > Event Management**. On the page that appears, you can search for events by node name, node ID, instance name, or instance ID and set **Owner**, **Detection Time**, **Event Status**, and **Event Type** to filter events.

In the search result, each event occupies a row and is associated with a node that encounters errors. The worst baseline is the baseline with the minimum margin among the baselines affected by the event.

- Click **View Details** in the Actions column of an event. You can view the event occurrence time, alerting time, clearance time, historical operational logs of the node, and detailed node logs.

  You can assign an alert recipient. After you click **View Alerts**, the alert details page corresponding to the event appears. Affected baselines are all descendant baselines affected by the node associated with the event. You can observe the descendant baselines and the impact on these baselines and analyze node logs to identify the cause of the event.
- If you click **Handle**, DataWorks records the event handling operation and pauses alerting for the event when the event is being handled.
- If you click **Ignore**, DataWorks keeps the event ignorance record and permanently stops alerting for the event.

# 1.6.3. Custom Rules

## 1.6.3.1. Manage custom alert rules

Custom alert rules allow you to monitor the status of specified node instances based on your business requirements. This helps you identify and handle exceptions at the earliest opportunity. This topic describes how to create a custom alert rule on the Rule Management page. This topic also describes how to add a DingTalk chatbot and obtain the webhook URL of the chatbot.

### Limits

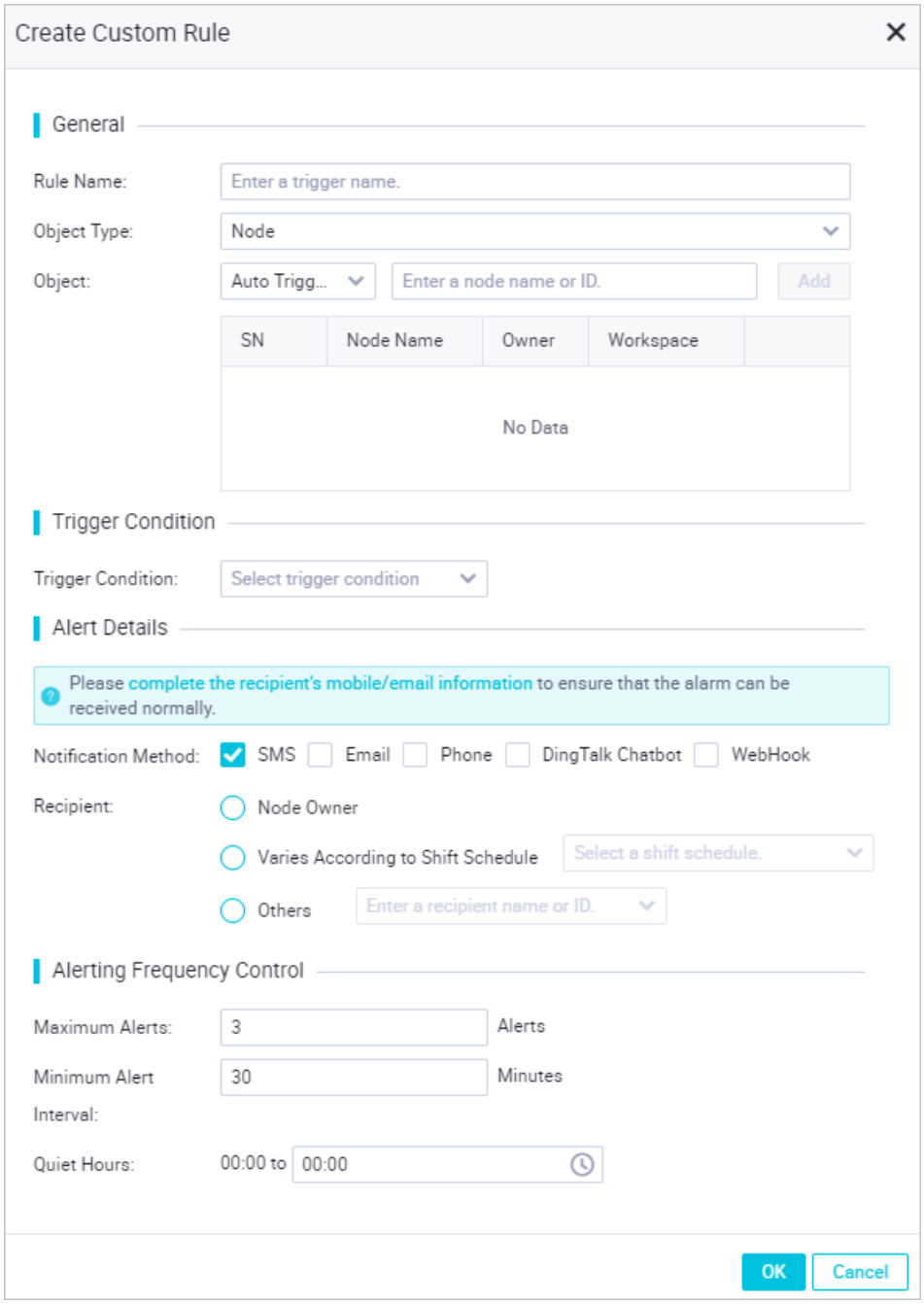
- Custom alert rules take effect only on auto triggered node instances.
- Only Singapore, Malaysia(Kuala Limpur), and Germany(Frankfurt) support the SMS reminding method. To use the SMS reminding method in other regions, submit a ticket to contact DataWorks technical support.
- The webhook alerting feature has the following limits:
  - You can use the webhook alerting feature only in DataWorks Enterprise Edition or Ultimate Edition.
  - Custom alert rules and baselines support the webhook alerting feature only in the Germany (Frankfurt) and Singapore (Singapore) regions.
  - DataWorks allows you to use the webhook alerting feature to send alert notifications only to Enterprise WeChat and Feishu.

### Create a custom alert rule

1. Go to the **DataStudio** page.
   i. Log on to the DataWorks console.
   ii. In the left-side navigation pane, click **Workspaces**.

iii. In the top navigation bar, select the region where the required workspace resides, find the workspace, and then click **Data Analytics**.

2. Click the icon in the upper-left corner and choose **All Products > Task Operation > Operation Center**. The **Operation Center** page appears.
3. In the left-side navigation pane, choose **Alarm > Rule Management**.
4. On the **Rule Management** page, click **Create Custom Rule** in the upper-right corner.
5. In the **Create Custom Rule** dialog box, set the parameters that are described in the following table.



| Section | Parameter | Description |
| --- | --- | --- |
| General | **Rule Name** | The name of the custom alert rule. |
| | **Object Type** | The type of object that you want to monitor. Valid values: **Node**, **Baseline**, **Workspace**, **Workflow**, **Exclusive Resource Groups for Scheduling**, and **Exclusive Resource Groups for Data Integration**. |
| | **Object** | If you set the **Object Type** parameter to Node, Baseline, Workspace, or Workflow, you must specify one or more objects. After you specify the name or ID of an object that you want to monitor, select the object from the drop-down list. Then, click the  icon. |
| | **Resource Group Name** | If you set the **Object Type** parameter to Exclusive Resource Groups for Scheduling or Exclusive Resource Groups for Data Integration, you must select a resource group name. |
| | | If you set the **Object Type** parameter to Node, Baseline, Workspace, or Workflow, the valid values of the Trigger Condition parameter are:<br>◦ **Completed**<br>Node instances are monitored from the time when they start to run. When the node instances are successfully run, an alert is reported.<br>◦ **Uncompleted**<br>Node instances are monitored from the time when they start to run. If the node instances are still running at the specified point in time, an alert is reported. For example, a node instance is scheduled to run at 01:00, and you set the alert time to 02:00. If the node instance is still running at 02:00, an alert is reported.<br>◦ **Error**<br>Node instances are monitored from the time when they start to run. If an error occurs when the node instances are running, an alert is reported.<br>If an error occurs for a node instance, the  icon is displayed in the General column on the **Cycle Instance** page under **Cycle Task Maintenance** in **Operation Center**.<br>◦ **Uncompleted in Cycle**<br>If node instances are still running at the end of the specified cycle, an alert is reported. In most cases, you can configure this trigger condition for node instances that are scheduled by hour. |

| Section | Parameter | Description |
| --- | --- | --- |
| Trigger Condition | Trigger Condition | For example, Node A is scheduled to run every 2 hours, and each run takes 25 minutes. If Node A starts to run at 00:00 every day, the node runs 12 times within 24 hours. The first cycle starts at 00:00, the second cycle starts at 02:00, and this goes on until the twelfth cycle. The twelfth cycle starts at 22:00. If the node runs as expected, the node instance in each cycle stops running at the specified point in time, such as 00:25 or 02:25. If the node instance is still running at the specified point in time in a cycle, an alert is reported.<br><br>**Note** You can configure the **Uncompleted in Cycle** trigger condition to monitor nodes in workflows.<br><br>If the Trigger Condition parameter is set to **Uncompleted in Cycle** for workflows, the system monitors nodes that are scheduled by day, hour, or minute in the workflows based on the cycle number (N) that you specified. If the number of node instances for a node is less than the value of N, the system ignores the alerts reported for the node.<br><br>For example, you set the cycle number to 3, and two nodes are configured in a workflow. The following examples show detailed alerting and monitoring information:<br><br>■ Node A is scheduled by hour: Node A is scheduled to run every 2 hours, and each run takes 25 minutes. If Node A starts to run at 00:00 every day, the node runs 12 times within 24 hours. The first cycle starts at 00:00, and the third cycle starts at 04:00. If the node runs as expected, the node instance in the third cycle stops running at 04:25. If the node instance in the third cycle is still running at 04:25, an alert is reported.<br><br>■ Node B is scheduled by minute: Node B is scheduled to run every 10 minutes, and each run takes 2 minutes. If Node B starts to run at 00:00 every day, the node runs six times within 1 hour. The first cycle starts at 00:00, and the third cycle starts at 00:20. If the node runs as expected, the node instance in the third cycle stops running at 00:22. If the node instance in the third cycle is still running at 00:22, an alert is reported.<br><br>○ **Overtime**<br><br>Node instances are monitored from the time when they start to run. If the node instances are still running after the specified period ends, an alert is reported. In most cases, you can configure this trigger condition to monitor the duration of node instances. |

| Section | Parameter | Description |
| --- | --- | --- |
| | | ○ **The error persists after the node automatically reruns**<br>Node instances are monitored from the time when they start to run. If an error still occurs after the node instances are rerun, an alert is reported. |
| | **Trigger Condition** | If you set the **Object Type** parameter to Exclusive Resource Groups for Scheduling or Exclusive Resource Groups for Data Integration, the valid values of the Trigger Condition parameter are:<br>○ **Resource Group Usage**: If the resource usage is greater than a specific percentage for a specific period of time, an alert is reported.<br>For example, if the resource usage is greater than 50% for 15 minutes, an alert is reported.<br>○ **Number of Instances Waiting for Resources in Resource Group**: If the number of node instances that are waiting for resources is greater than a specific number for a specific period of time, an alert is reported.<br>For example, if the number of node instances that are waiting for resources is greater than 10 for 15 minutes, an alert is reported. |

| Section | Parameter | Description |
| --- | --- | --- |
| Alert Details | Notification Method | The methods used to send alert notifications. Valid values: **Email**, **SMS**, **DingTalk Chatbot**, and **WebHook**. You can add a DingTalk chatbot to receive alert notifications. For more information about how to send alert notifications to a DingTalk group, see the following section. If you want to send alert notifications to multiple DingTalk groups, add multiple webhook URLs.<br>**Notice** Only the webhook URLs of DingTalk chatbots are supported. |
| | Recipient | The user who receives alert notifications. Valid values: **Node Owner**, **Varies According to Shift Schedule**, and **Others**. |
| Alerting Frequency Control | Maximum Alerts | The maximum number of times an alert is reported. If the number of times an alert is reported exceeds the specified threshold, the alert is no longer reported. |
| | Minimum Alert Interval | The minimum interval at which an alert is reported. |
| | Quiet Hours | The specified period during which no alerts are reported. |

6. Click **OK**. An alert rule is created.

   On the **Rule Management** page, you can find the created alert rule and click **View Details** in the Actions column to view the details of the alert rule.

## Send alert notifications to a DingTalk group

1. Go to the DingTalk group to which you want to send alert notifications and click the **Group Settings** icon in the upper-right corner.
2. In the Group Settings panel, click **Group Assistant**.
3. In the **Group Assistant** panel, click **Add Robot**.
4. In the **ChatBot** dialog box, click the + icon.
5. In the **Please choose which robot to add** section, click **Custom**.
6. In the **Robot details** message, click **Add**.
7. In the **Add Robot** dialog box, set the parameters that are described in the following table.

| Parameter | Description |
| --- | --- |
| Chatbot name | The name of the custom chatbot. |
| Add to Group | The DingTalk group to which the chatbot is added. This group cannot be changed. |

| Parameter | Description |
| --- | --- |
| Custom Keywords | After you specify custom keywords, messages are sent only if these messages contain the specified keywords. We recommend that you specify the keyword DataWorks.<br>**Note** You can specify a maximum of 10 keywords. A message can be sent only if it contains at least one of the specified keywords. |

8. Read the terms of service, select **I have read and accepted <<DingTalk Custom Robot Service Terms of Service>>**, and then click **Finished**.
9. After you complete the security settings, copy the webhook URL of the chatbot and click **Finished**.

**Notice**

- Save the copied webhook URL and paste it in the **Webhook Address** field when you create an alert rule.
- Keep the webhook URL confidential. If the webhook URL is leaked, your business is at risk.

# 1.6.3.2. Create and manage a shift schedule

DataWorks provides the shift schedule feature, which allows you to create shift schedules. If you set the Recipient parameter to Varies According to Shift Schedule and select a shift schedule when you create a custom alert rule, DataWorks can send alert notifications to the on-duty engineers that you specify for the shift schedule. After the engineers receive the alert notifications, they can identify and handle exceptions at the earliest opportunity. This topic describes how to create and manage a shift schedule.

## Limits

You can use the shift schedule feature only in DataWorks of the Professional Edition or a more advanced edition. If you are using DataWorks of the Basic or Standard Edition, upgrade your DataWorks service to the Professional Edition or a more advanced edition before you use this feature. For more information about DataWorks editions, see DataWorks advanced editions.

Only Singapore, Malaysia(Kuala Limpur), and Germany(Frankfurt) support the SMS reminding method. To use the SMS reminding method in other regions, submit a ticket to contact DataWorks technical support.

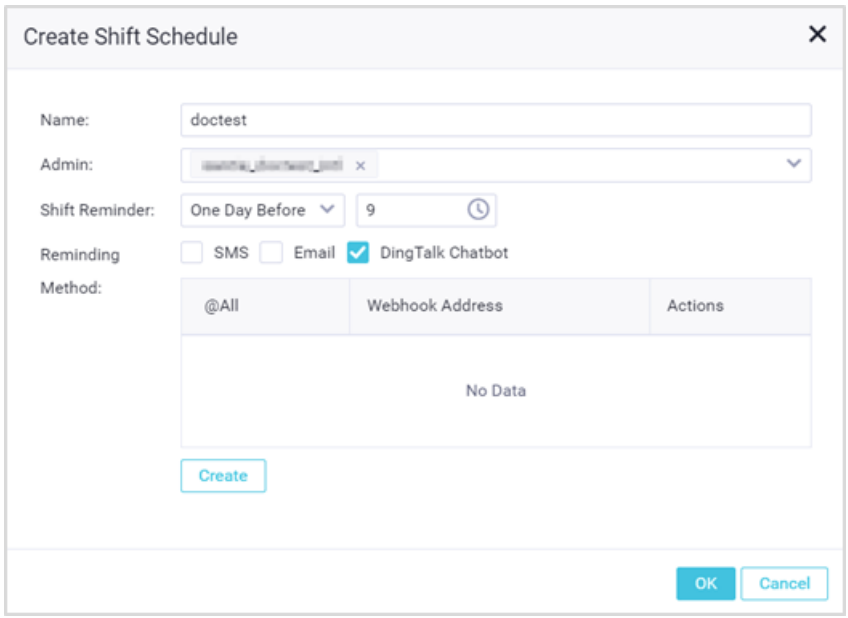
## Precautions

- After your DataWorks edition expires, you cannot access the Schedule page. In addition, you cannot create, view, modify, or delete a shift schedule and the shift plans in the shift schedule.
- If you do not renew the subscription of your DataWorks edition in a timely manner or the edition is even downgraded after it expires, DataWorks retains the shift schedules you created. After you upgrade your DataWorks service to the Professional Edition or a more advanced edition again, you can still use the shift schedules.

- To use a shift schedule, you must set the Recipient parameter to Varies According to Shift Schedule and select the shift schedule when you create a custom alert rule. This way, the on-duty engineers that you specify for the shift schedule can receive alert notifications when an alert is reported. For more information about how to create a custom alert rule, see Manage custom alert rules.

## Create a shift schedule and a shift plan

1. Go to the Operation Center page.
   i. Log on to the DataWorks console.
   ii. In the left-side navigation pane, click **Workspaces**.
   iii. In the top navigation bar, select the region where your workspace resides. Find your workspace and click **Operation Center** in the Actions column.
2. In the left-side navigation pane of the **Operation Center** page, choose **Alarm > Schedule**.
3. In the **Shift Schedule** pane, click the + icon. In the **Create Shift Schedule** dialog box, configure the parameters.

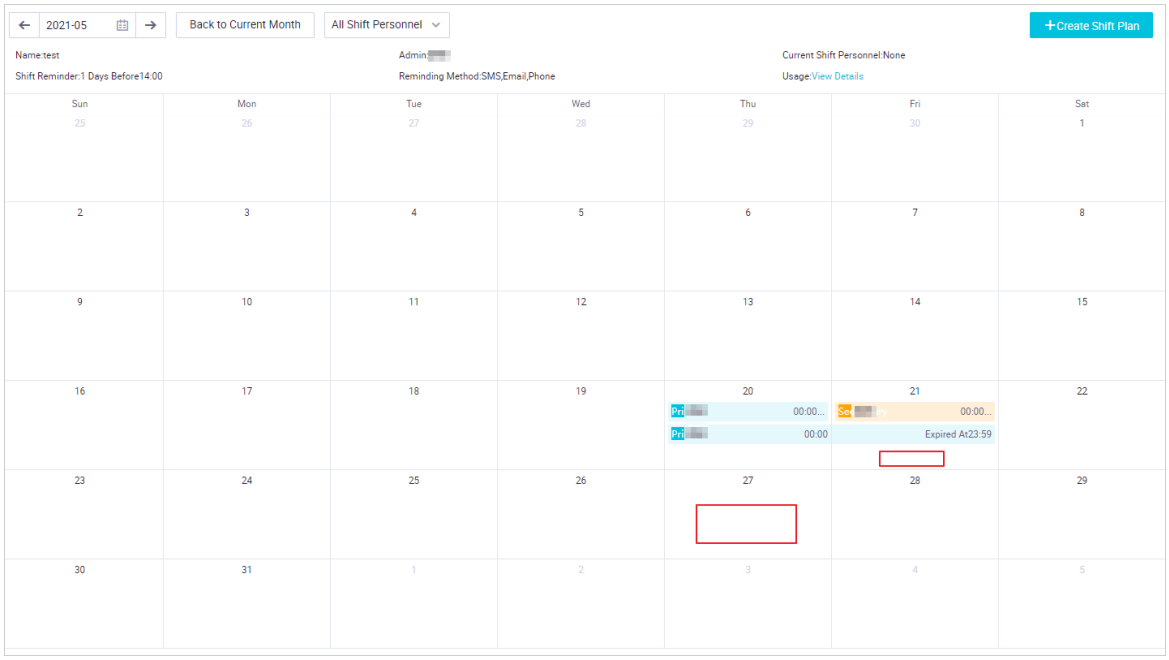



| Parameter | Description |
| --- | --- |
| **Name** | The name of the shift schedule. |
| **Admin** | The administrator of the shift schedule. The default value is the account that is used to log on to the Alibaba Cloud Management Console. You can set this parameter to the Alibaba Cloud account or the RAM user that is added to your DataWorks workspace. |

| Parameter | Description |
| --- | --- |
| **Shift Reminder** | You can specify this parameter to notify on-duty engineers of the time of the next shift. Default value: **No Reminder**. You can set this parameter to **One Day Before**, **Two Days Before**, or **One Week Before** based on your business requirements.<br>If you set this parameter to **One Day Before**, **Two Days Before**, or **One Week Before**, you can also configure the specific notification time and notification method. |
| **Reminding Method** | The notification method. Valid values: **SMS**, **Email**, and **DingTalk Chatbot**. For more information about how to add a DingTalk chatbot and obtain the webhook URL of the chatbot, see Send alert notifications to a DingTalk group.<br>This parameter is required only when you set the **Shift Reminder** parameter to **One Day Before**, **Two Days Before**, or **One Week Before**. |

4. Click **OK**.
5. In the upper-right corner of the pane that appears, click **Create Shift Plan**.

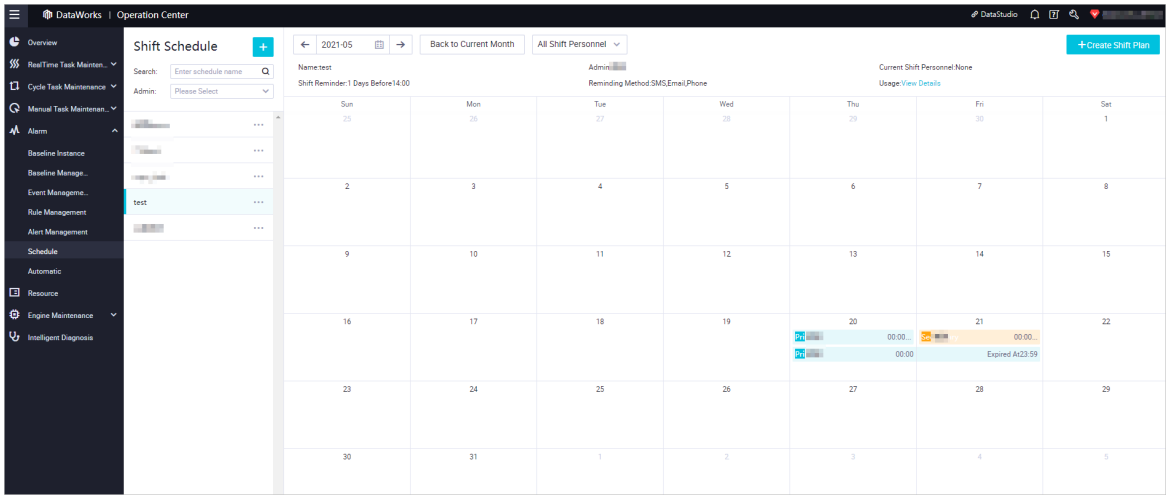
   You can also double-click the white block of a specific date to create a shift plan.




6. In the **Create Shift Plan** dialog box, configure the parameters.

| Parameter | Description |
| --- | --- |
| **Shift Personnel** | The on-duty engineer of the current shift schedule. You can set this parameter to the Alibaba Cloud account or the RAM user that is added to the DataWorks workspace. |
| **Type** | ○ **Chief**: The on-duty engineer serves as the main engineer in charge.<br>○ **Vice**: The on-duty engineer serves as the secondary engineer in charge. |
| **All Day** | ○ **Yes**: The shift plan is effective for 24 hours a day. The values of the **Start Time** and **End Time** parameters are both a specific date.<br>○ **No**: You can configure the **Start Time** and **End Time** parameters based on your business requirements. |
| **Start Time** | The time when the shift plan starts. You can customize a point in time only when the **All Day** parameter is set to **No**. |
| **End Time** | The time when the shift plan ends. You can customize a point in time only when the **All Day** parameter is set to **No**. |
| **Repeat** | Default value: **No**. You can also set this parameter to **Every Day**, **Every Week**, **Every Monday to Every Friday**, **Every Month**, **Every Year**, or **Custom Time**.<br>After you configure a repetition cycle, the information about the created shift plan is added to the white blocks of the specified dates based on the cycle. |
| **Repetition Interval** | You can set the **Repetition Interval** parameter to any number of days before or after the current shift period.<br>This parameter is required only when you set the **Repeat** parameter to **Custom Time**. |
| **Ended At** | Specifies the date on which the repetition cycle ends. Default value: Never.<br>This parameter is required only when you set the **Repeat** parameter to **Every Day**, **Every Week**, **Every Monday to Every Friday**, **Every Month**, **Every Year**, or **Custom Time**. |

The following figure shows a sample shift plan.



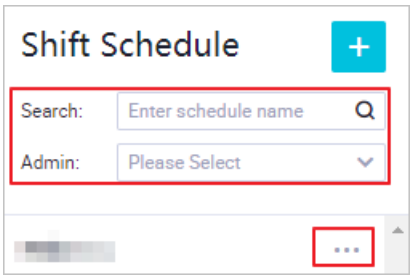

> **Note**
>
> - The information marked in blue is the shift information about the main engineer in charge. The information marked in yellow is the shift information about the secondary engineer in charge.
> - You can create a shift plan only for the current date or a later date.
> - DataWorks allows you to specify multiple main engineers or secondary engineers in charge for the same shift plan.

## Manage a shift schedule

- In the **Shift Schedule** pane, you can perform the following operations:
  - Search for a shift schedule by specifying the **Search** or **Admin** parameter.
  - Move the pointer over the ... icon on the right side of the name of a shift schedule and select **Change** or **Delete** to modify or delete the shift schedule.

    > **Note** Only the administrator of a shift schedule can modify or delete the shift schedule or a shift plan created in the shift schedule.

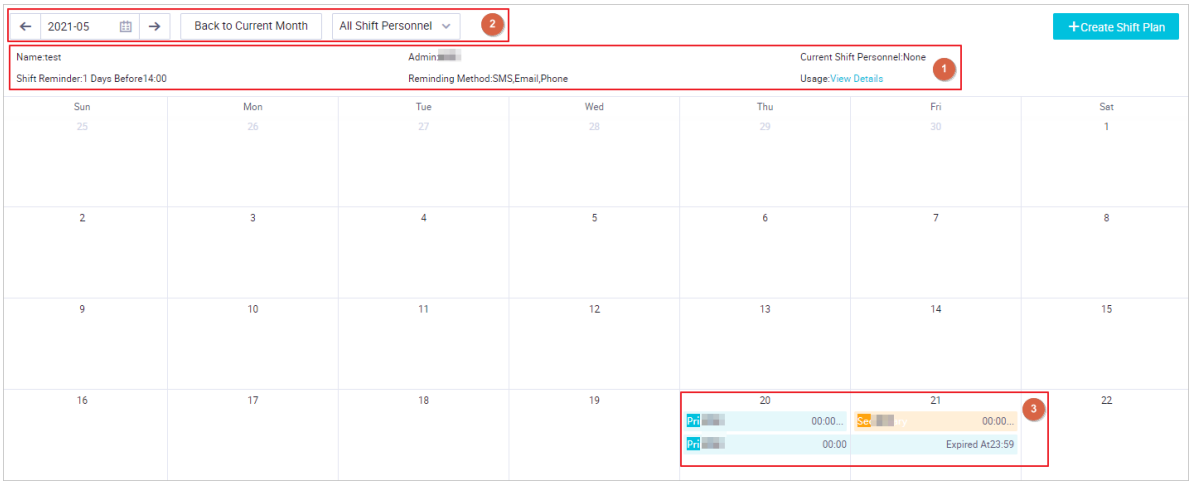
  - Double-click a shift schedule to go to the details pane and view the detailed information about the shift schedule.




- In the details pane of a shift schedule, you can perform the following operations:

- In Area 1, you can view the basic information about the shift schedule and view the current engineer in charge. If the custom alert rule that you configured uses this shift schedule, related alert notifications are sent based on the shift plan created in the shift schedule. You can click **View Details** next to **Usage** to view detailed information. For more information about how to manage custom alert rules, see Manage custom alert rules.
- In Area 2, you can select the type of engineers in charge to view a specific shift plan. You can also specify a month to view a shift plan. By default, the shift plans of all engineers in charge within the current month are displayed.

  After you view a shift plan that is not created in the current month, you can click **Back to Current Month** to go back to the details pane of a shift schedule in the current month.
- In the details pane of a shift schedule, you can view the detailed information about a created shift plan, including engineers in charge, the shift dates, and the types of engineers in charge. You can also click a created shift plan to modify or delete it.




# 1.6.4. Automated O&M

The Operation Center service of DataWorks provides the automated O&M feature for node instances that are running on exclusive resource groups. You can customize O&M rules for the node instances based on your business requirements. This topic describes how to manage an automated O&M rule.

## Context

The automated O&M feature enables DataWorks to automatically perform O&M operations on node instances that are running on exclusive resource groups. DataWorks performs O&M operations based on the O&M rules that you created and the monitoring rules that are associated with the O&M rules. You can customize metrics and create an O&M rule for the node instances based on your business requirements. Then, you can associate the O&M rule with a monitoring rule. The monitoring rule can be an existing monitoring rule or a new monitoring rule. If the conditions specified in the O&M rule are met, the system automatically performs the O&M operation defined in the rule.

## Limits

- Only workspace administrators can create, modify, or delete automated O&M rules.
- You can create automated O&M rules only for node instances that are running on exclusive resource groups.
- You can view the execution records that are generated only within 30 days for automated O&M rules.

- DataWorks supports only the automated O&M operation that stops running node instances when the resource usage of the exclusive resource groups used by the node instances is excessively high.
- You can associate one automated O&M rule with only one monitoring rule. However, you can associate multiple automated O&M rules with the same monitoring rule.

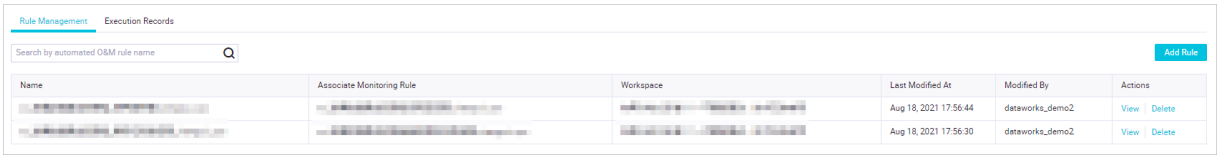
## Go the Automatic page

1. Go to the **DataStudio** page.
   i. Log on to the DataWorks console.
   ii. In the left-side navigation pane, click **Workspaces**.
   iii. In the top navigation bar, select the region where the required workspace resides, find the workspace, and then click **Data Analytics**.
2. On the **DataStudio** page, click the icon in the upper-left corner and choose **All Products > Operation Center**.
3. On the Operation Center page, choose **Alarm** > Automatic. Then, the Rule Management tab of the Automatic page appears.

## Manage automated O&M rules

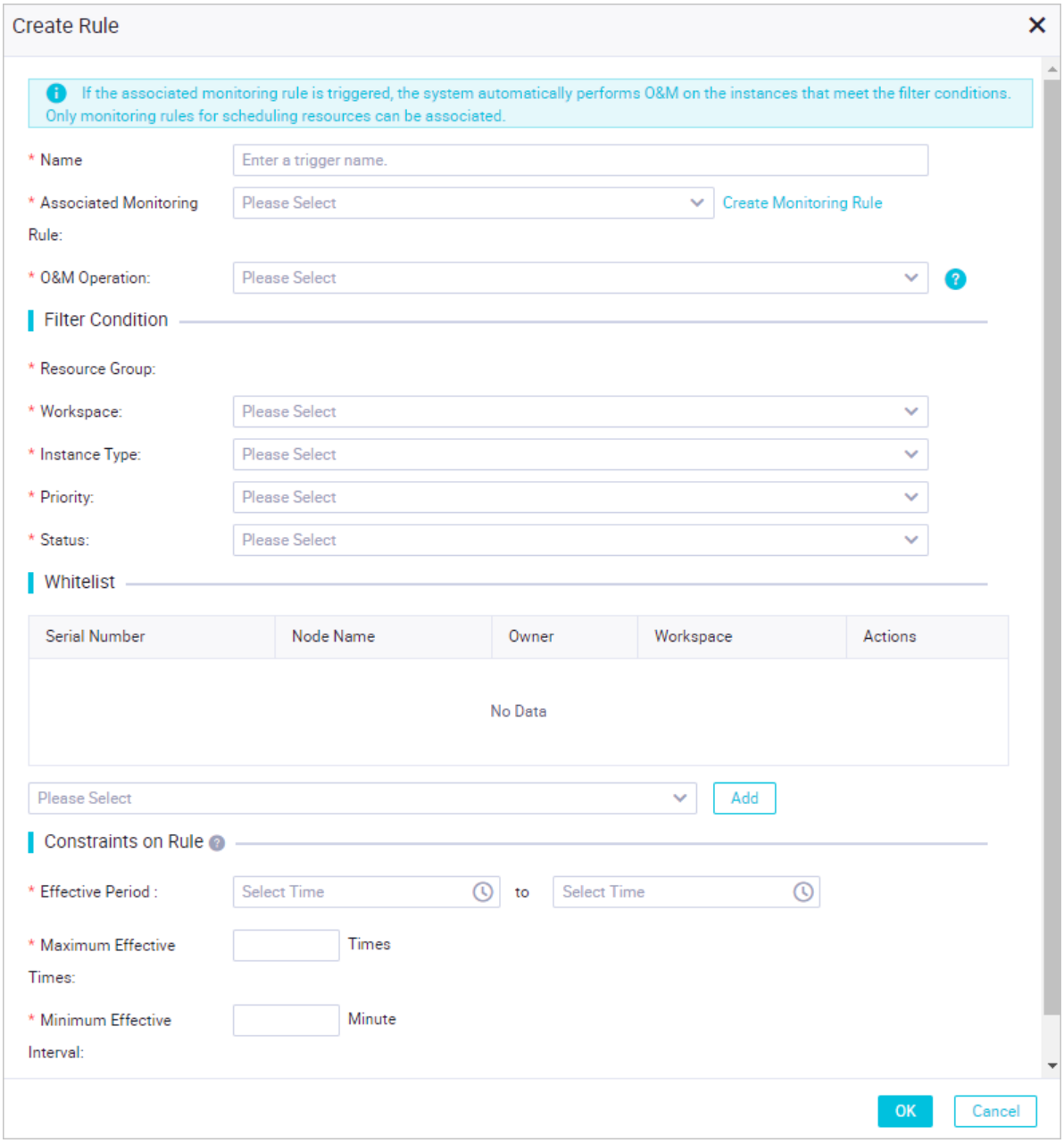
The Automatic page displays all the automated O&M rules that are created and the execution records of the rules. You can perform the following operations on this page.

> **Note** Only workspace administrators can create or modify automated O&M rules. For more information about how to obtain the permissions of the workspace administrator role, see Add workspace members.




- Add an automated O&M rule

  On the **Rule Management** tab, click **Add Rule** in the upper-right corner. In the Created Rule dialog box, configure the following parameters.




| Section | Parameter | Description |
| --- | --- | --- |
| Basic information | **Name** | The name of the automated O&M rule. |
| | **Associated Monitoring Rule** | The monitoring rule with which you want to associate the automated O&M rule. For more information about how to create a monitoring rule, see Manage custom alert rules.<br>**Note** You can associate only monitoring rules for exclusive resource groups for scheduling with automated O&M rules. |

| Section | Parameter | Description |
| --- | --- | --- |
| | **O&M Operation** | The automated O&M operation that you want to perform if the automated O&M rule is triggered.<br>**Note** Only the operation **Terminate Running Instance** is supported. |
| Filter Condition | **Resource Group** | The name of the resource group that you specified when you create the monitoring rule.<br>**Note** Only the names of exclusive resource groups for Data Integration and exclusive resource groups for scheduling are displayed. |
| | **Workspace** | The name of the workspace to which the automated O&M rule is applied. |
| | **Instance Type** | The type of the node instance to which the automated O&M rule is applied. Valid values: **Auto Triggered Node Instances**, **Data Backfill Instance**, **Test Instances**, and **Manually Triggered Workflows**. |
| | **Scheduling Cycle** | The scheduling cycle of the node instance. Valid values: Minute, Hours, Day, Week, and Month. |
| | **Priority** | The priority of the automated O&M rule. A larger value indicates a higher priority. Valid values:<br>○ 1<br>○ 3<br>○ 5<br>○ 7<br>○ 8 |
| | **Status** | The status of the node instance. Valid values: **Waiting for Resources** and **Running**. |

| Section | Parameter | Description |
| --- | --- | --- |
| Whitelist | **Whitelist** | The list of the node instances for which the automated O&M rule does not take effect. If you want to add another node instance to the list, you can select the name of the node instance from the drop-down list below the node instance list and click **Add**.<br>**Note** Automated O&M rules do not take effect for the node instances that are contained in the node instance list. |
| Constraints on Rule | **Effective Period** | The time range within which the automated O&M rule is effective. |
| | **Maximum Effective Times** | The maximum number of times that the automated O&M rule can be triggered. |
| | **Minimum Effective Interval** | The minimum interval at which the automated O&M rule can be triggered. |

- Search for an automated O&M rule

  In the search box of the **Rule Management** tab, you can enter the name of an automated O&M rule to search for the rule.

- View, modify, and delete an automated O&M rule
  - If you want to view the information about an automated O&M rule, find the desired rule in the automated O&M rule list of the **Rule Management** tab and click **View** in the Actions column.
  - If you want to modify an automated O&M rule, find the desired rule in the automated O&M rule list of the Rule Management tab and click View in the Actions column. In the **View Rule** dialog box, click **Change** in the lower-right corner.
  - If you want to delete an automated O&M rule, find the desired rule in the automated O&M rule list of the Rule Management tab and click **Delete** in the Actions column. In the Delete Rule message, click **OK**.

- View the execution records of an automated O&M rule

  The **Execution Records** tab displays the execution information about automated O&M rules, including the time when the rules are executed, rule owners, and the number of node instances to which the rules are applied. If you want to view the detailed execution information about a rule, click **View Details** in the Actions column of the rule. In the Record Details dialog box, you can view the line chart that displays the resource usage of the desired resource group within the last 24 hours and the number of node instances that are waiting for resources since the rule took effect.

  - **Instances Waiting for Resources/Resource Usage**: This section provides a chart that displays the number of node instances that are waiting for resources and the resource usage of the desired resource group. You can move the pointer over a point in the chart to view the number of node instances that are waiting for resources and the resource usage of the desired resource group at the related point in time.

- **Terminated Node Instances**: This section displays all the node instances whose running is stopped.

### Summary

After you create an automated O&M rule, the system automatically monitors the resource usage of the resource group defined in the automated O&M rule. After the automated O&M rule is triggered, the system performs the O&M operation on the node instances that are running on the resource group. For more information about the O&M of resources in resource groups, see Resource O&M.

# 1.6.5. Manage alerts

On the Alert Management page, you can view all alerts reported by the Monitor module.

Go to the **Operation Center** page. In the left-side navigation pane, choose **Alarm > Alert Management**. On the page that appears, you can set **Alert Trigger ID/Name**, **Recipient**, **Alert Time**, **Notification Method**, and **Trigger Type** to filter alerts.

You can view alert information such as **Notification Method** and **Status**. Click **View Details** in the Actions column to view more information about each alert.

# 1.6.6. FAQ

This topic provides answers to some frequently asked questions about the intelligent monitoring system in Operation Center.

## What do I do if I did not receive an email notification about a node run failure?

- Problem description: I configured the required parameters in the intelligent monitoring system in Operation Center. However, I did not receive an alert notification by email after a node failed to run.
- Cause:

  The email address of the alert recipient that you configured is invalid.

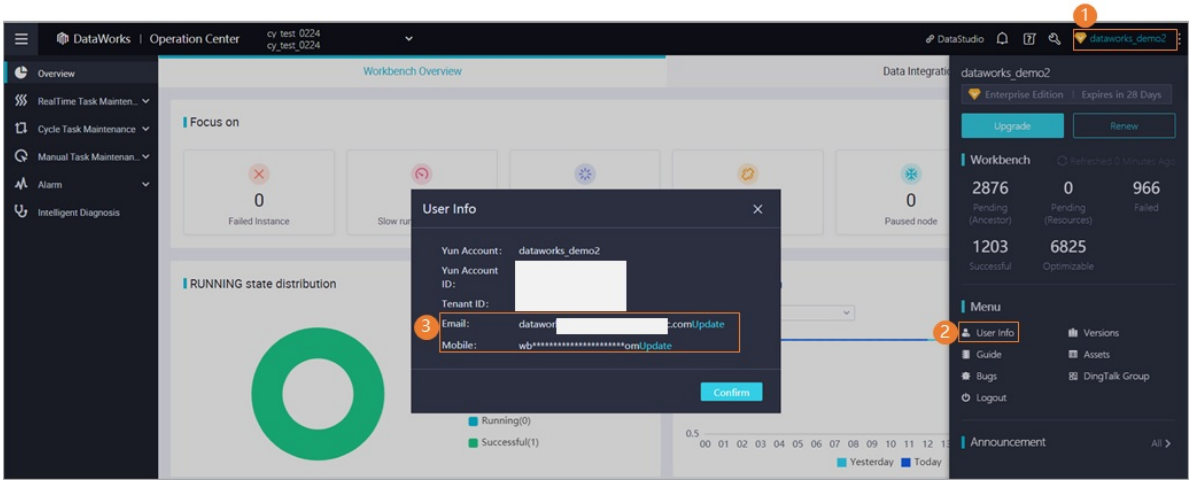
  If a DataWorks node fails to run, the system sends an alert notification to the email address of the alert recipient that you configured in the intelligent monitoring system in Operation Center. If the email address is invalid, the system sends the alert notification to the recipient that you configured on the Manage Contacts page, which appears after you click Manage Contacts in the upper-right corner of the Common Settings page of your Alibaba Cloud account. The recipient is also the person who receives messages, such as product overdue payment, suspension, and imminent release notifications. In this case, the alert recipient that you configured in the intelligent monitoring system of Operation Center will not receive the alert notification.

- Solution:

  Make sure that the email address of the alert recipient that you configured in the intelligent monitoring system of Operation Center is validso that the system can send the alert notification to the recipient.

  In the upper-right corner of the Operation Center page, click the account name. In the panel that appears, click User Info in the Menu section. In the User Info dialog box, view or change the email address and phone number.

> **Note** If a RAM user is used, only the Alibaba Cloud account to which the RAM user belongs can change the email address and phone number of the RAM user.




# 1.7. Resource O&M

This topic describes how to use the resource O&M feature of DataWorks to monitor the usage of resource groups that are used to run a node.

## Context

You can customize metrics and rules for automated O&M for a node that runs on a resource group based on your business requirements. You can view the usage of resource groups and status of node execution on the Resource page. The resource O&M feature supports intelligent monitoring and automated O&M of resource groups and nodes. This reduces the number of complex manual operations and improves O&M management efficiency.

## Limits

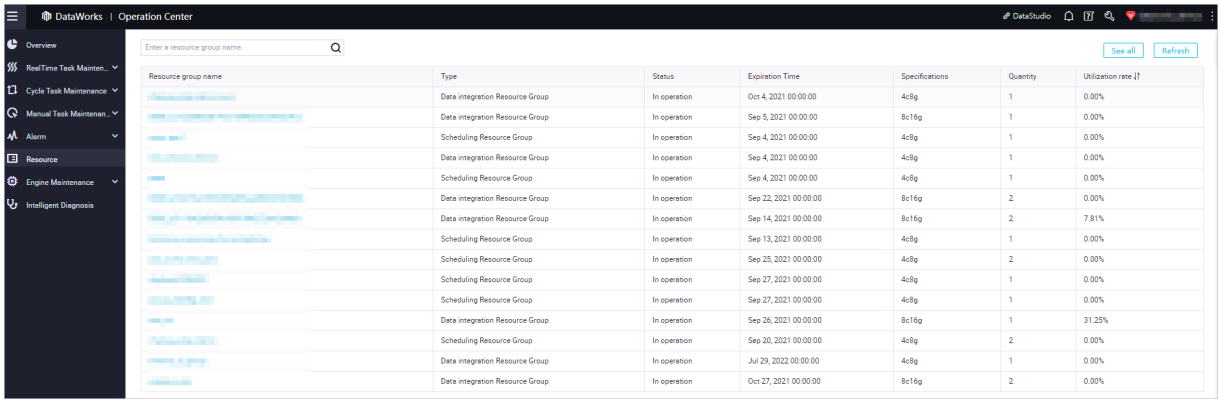
- The resource O&M feature of DataWorks is unavailable in the China (Shanghai) and China (Hangzhou) regions.
- The resource O&M feature of DataWorks can be used only for exclusive resource groups for scheduling and exclusive resource groups for Data Integration.

## Go to the Resource page

1. Log on to the DataWorks console.
2. In the left-side navigation pane, click **Workspaces**.
3. After you select the region in which the workspace that you want to manage resides, find the workspace and click **Data Analytics** in the Actions column.
4. On the **DataStudio** page, click the icon in the upper-left corner and choose **All Products > Operation Center**.
5. In the left-side navigation pane, click **Resource**. A list of exclusive resource groups that are used by DataWorks nodes appears.

## View the list of resource groups

The Resource page displays the resource groups that are used by DataWorks nodes. The status of the resource groups is Running. The resource O&M feature supports only exclusive resource groups for scheduling and exclusive resource groups for Data Integration. You can perform the following operations based on your business requirements.




- Search for the resource group whose usage you want to view.

  Enter the name of the resource group whose usage you want to view in the search box in the upper-right corner and click the icon to search for the resource group in fuzzy match mode.

- View other resource groups.

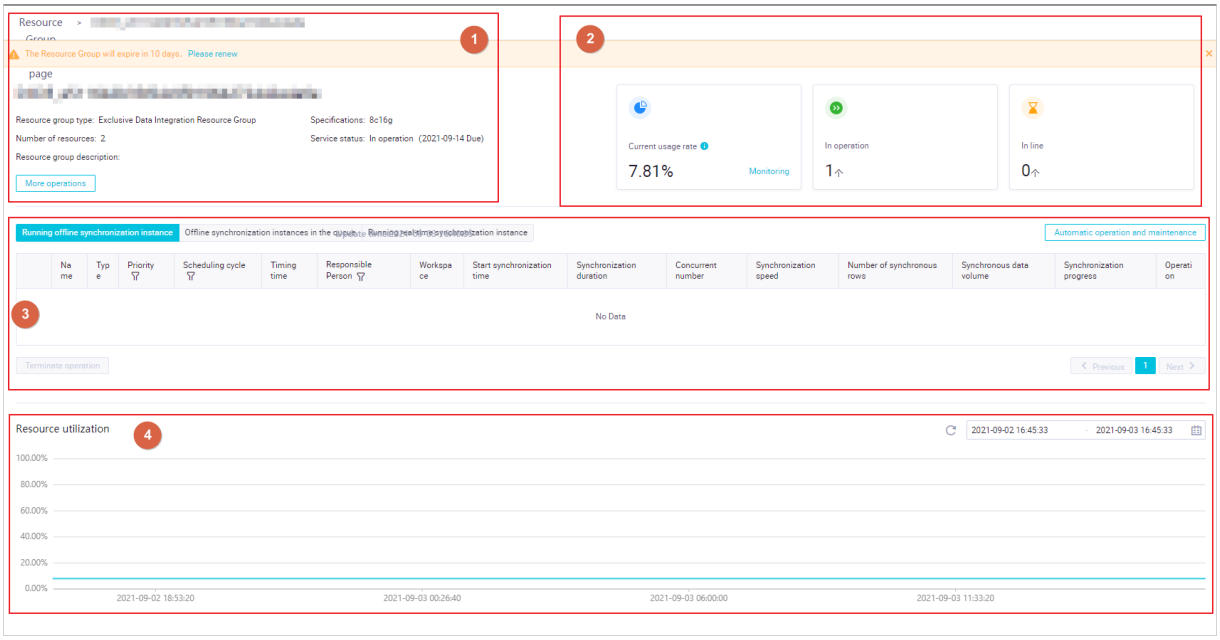
  In the upper-right corner of the Resource page, click See all. The **Resource Groups** page appears.

| Parameter | Description |
| --- | --- |
| **Resource group name** | The name of the resource group that you created. After you click the resource group name, the details page of the resource group appears. For more information, see View the details about a resource group. |
| **Type** | The resource group type that you selected when you created the resource group. The resource O&M feature supports only exclusive resource groups for Data Integration and exclusive resource groups for scheduling.<br>**Note** The resource O&M feature supports only exclusive resource groups. |
| **Status** | The status of the resource group. Valid values:<br>• **Running**<br>• **Expired**<br>• **Released**<br>• **Starting**<br>• **Failed to perform the create operation**<br>• **Updating**<br>• **Failed to perform the update operation**<br>• **Released**<br>• **Timed Out** |

| Parameter | Description |
| --- | --- |
| **Expiration Time** | The time when the resource group expires. |
| **Specifications** | The specifications of the resource group that you defined when you created the resource group. |
| **Quantity** | The number of resources that you configured when you created the resource group. |
| **Utilization rate** | The usage of the resource group. Click the ⇵ icon to the right of **Utilization rate** to sort all the resource groups in the list in ascending or descending order of the resource usage. |

## View the details about a resource group

On the details page of a resource group, you can view the basic information, usage curve, and nodes of the resource group. You can customize monitoring rules for intelligent monitoring and automated O&M of resources.




| No. | Parameter | Description |
| --- | --- | --- |
| 1 | Basic information about the resource group | If you want to perform other operations on the resource group, such as scaling, click **More operations**. The **Resource Groups** page appears. |

| No. | Parameter | Description |
| --- | --- | --- |
| 2 | Usage of the resource group. You can view the current usage of the resource group, the number of nodes that are in the Running state and are using the resource group, and the number of nodes that are in the In state in the queue and are waiting for resources. | If you want to view the monitoring rules that are configured for the resource group, click **Monitoring**. The Rule Management page appears. You can also choose **Operation Center** > **Alarm** > **Rule Management** to go to the Rule Management page. |
| 3 | All nodes that run on the current resource group. The nodes are classified into the following types based on the status of the nodes and the synchronization mode:<br>• **Running offline synchronization instance**<br>• **Offline synchronization instances in the queue**<br>• **Running real-time synchronization instance** | Click **Automatic operation and maintenance**. The Automatic page appears. You can also choose **Operation Center** > **Alarm** > **Automatic** to go to the Automatic page. |
| 4 | Usage curve of the resource group | To view the usage of the resource group in a specified time period, you can use the date and time picker to select the start time and end time.<br>You can also move the pointer over a point in the curve to view the usage of the resource group at a specific point in time. |

# 1.8. Use the engine O&M feature

A DataWorks node instance that runs on the E-MapReduce (EMR) compute engine contains multiple EMR jobs. These EMR jobs are run based on a specific sequence. You can use the engine operations and maintenance (O&M) feature of DataWorks to view the details of each EMR job and find and remove jobs that fail to be run. This prevents failed jobs from affecting the running of both the DataWorks node instances to which the jobs belong and the nodes that generate the node instances.

## Limits

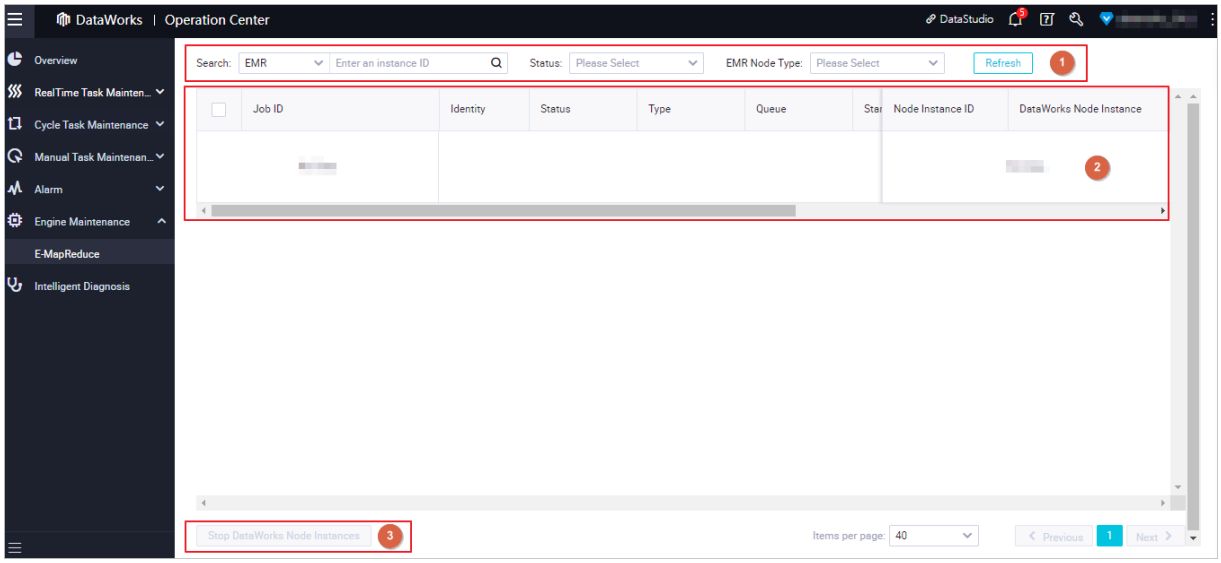
- DataWorks allows you to perform O&M only for the EMR compute engine. You must upgrade your EMR execution package to obtain O&M data. If you want to use the engine O&M feature, submit a ticket to upgrade your EMR execution package.
- **Engine Maintenance** is displayed in the left-side navigation pane of the **Operation Center** page only after you associate an EMR compute engine instance with your workspace.
- The engine O&M feature is available only for workspaces that reside in the China (Shenzhen) region.

## Go to the EMR engine O&M page

1. Go to the **DataStudio** page.
   i. Log on to the DataWorks console.
   ii. In the left-side navigation pane, click **Workspaces**.
   iii. In the top navigation bar, select the region where the required workspace resides, find the workspace, and then click **Data Analytics**.
2. On the **DataStudio** page, click the icon in the upper-left corner and choose **All Products > Operation Center**.
3. In the left-side navigation pane, choose **Engine Maintenance > E-MapReduce** to go to the EMR engine O&M page.

## View EMR jobs

On the EMR engine O&M page, you can view the details of the EMR jobs that are running for the EMR compute engine instance associated with your DataWorks workspace in the current region. You can perform the following operations based on your business requirements:




- Search for an EMR job.

  You can specify different conditions in the upper part of the **EMR engine O&M** page to search for an EMR job.

| Condition | Description |
| --- | --- |

| Condition | Description |
| --- | --- |
| **instance ID** | You can search for a specific EMR job by **job ID**. Alternatively, you can specify an **instance ID** to search for all the EMR jobs that belong to the related DataWorks node instance.<br>**Note**<br>◦ On the **EMR engine O&M** page, the data of an EMR compute engine instance over the last three days is displayed.<br>◦ If you search for a specific EMR job by **job ID** or **instance ID**, the data of the EMR compute engine instance only over the last 30 days is displayed. |

| Condition | Description |
| --- | --- |
| Status | ○ **NEW**: indicates that an EMR job is newly created.<br>○ **NEW_SAVING**: indicates that the EMR job is being saved.<br>○ **SUBMITTED**: indicates that a running application has been submitted for the EMR job.<br>○ **ACCEPTED**: indicates that the scheduling system approves the running application.<br>○ **RUNNING**: indicates that the EMR job is being run.<br>**Note** If an EMR job is in the RUNNING state for a long time, you can manually stop the running of the DataWorks node instance to which the EMR job belongs. This prevents the EMR job from occupying resources and affecting the running of the DataWorks node that generates the node instance. For more information, see *Stop a DataWorks node instance*.<br>○ **FINISHED**: indicates that the running of the EMR job is complete.<br>○ **SUCCESSED**: indicates that the EMR job is successfully run.<br>○ **FAILED**: indicates that the EMR job fails to be run. If an EMR job is in the FAILED state, you must locate and troubleshoot the issue at the earliest opportunity. This can prevent the EMR job from affecting both the running of the descendant jobs and the DataWorks node generating the node instance to which the EMR job belongs.<br>○ **KILLED**: indicates that the EMR job is stopped by the user who runs the job or the administrator. |
| EMR Node Type | Valid values: **MAPREDUCE** and **SPARK**. |

- View the details of an EMR job.
  - You can view the basic information of an EMR job, such as the ID, status, start time, and end time of the job. You can also view the DataWorks node instance to which an EMR job belongs and the ID of the DataWorks node instance.

    **Note** Different EMR jobs may belong to DataWorks node instances with the same name. However, if these EMR jobs start to run at different points in time, they actually belong to different DataWorks node instances. To determine whether EMR jobs belong to the same **DataWorks node instance**, you can view the ID in the Node Instance ID column that corresponds to each EMR job.

- You can find an EMR job in the **FAILED** state, and click the ID of the job in the **Job ID** column or click the name in the **DataWorks Node Instance** column. On the details page of the job, view and troubleshoot issues.
- You can click **Start time** or **End At** to arrange the EMR jobs in chronological or reverse chronological order. This way, you can clearly understand the running sequence of EMR jobs and the status of your desired EMR job.

- Stop a DataWorks node instance.

  If an EMR job is in the **RUNNING** state for a long time, you can select the EMR job and click **Stop DataWorks Node Instances** in the lower-left corner of the **EMR engine O&M** page. This way, you can stop the running of the DataWorks node instance to which the EMR job belongs. An EMR job may be in the RUNNING state for a long time because of an internal error and the job cannot be automatically stopped. To prevent the job from occupying resources and affecting the running of other jobs, you must manually stop the job and troubleshoot issues at your earliest opportunity.

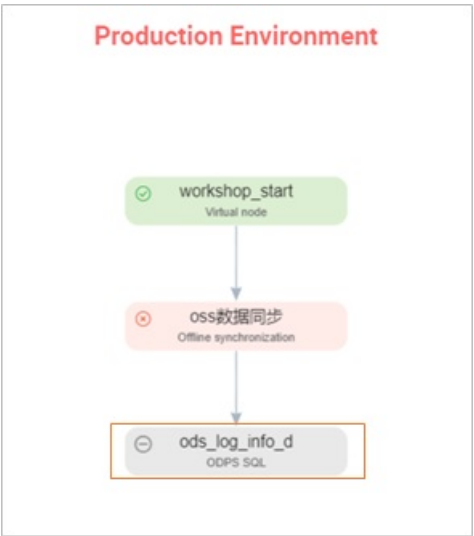
  > Notice
  > - If multiple EMR jobs belong to the same DataWorks node instance and you stop one of these EMR jobs, the DataWorks node instance enters the FAILED state.
  > - You can stop EMR jobs only in the RUNNING state in a DataWorks node instance.
  > - After you stop a running EMR job in a DataWorks node instance, the DataWorks node instance enters the FAILED state. In this case, the DataWorks node that generates the DataWorks node instance is blocked. Exercise caution when you stop a running EMR job.

# 1.9. FAQ

## 1.9.1. What do I do if a node fails to be run or does not generate data?

### Problem description

When I view data in Data Map, I find that a node does not generate data on the day. When I view the node in Operation Center, I find that the node is in the Not Running state, as shown in the following figure.




## Possible cause

The ancestor node of the node fails to be run. The possible cause is that the instance generated by the ancestor node is frozen or waiting to be run. The node can be run only after the running of the ancestor node is successful.

## Solution

You must open the directed acyclic graph (DAG) of the node and right-click the node name to view the ancestor node of the node. Then, view the instance generated by the ancestor node. The instance may be in the states indicated by the following colors:

- Purple

  The color purple indicates that the instance generated by the ancestor node is frozen, and both the ancestor node and the node fail to be run. In this case, you can view the operation record of the frozen instance on the instance information page of the instance.

- Yellow

  The color yellow indicates that the ancestor node is waiting to be run.

  - The ancestor node is scheduled to be run, but the scheduled time does not arrive. In this case, you can select the ancestor node and click **Show Details** in the lower-right corner to view the scheduled time for running the ancestor node. When the scheduled time arrives, the system starts to run the ancestor node. After the running of the ancestor node is complete, the system starts to run the node.
  - The ancestor node may be waiting for scheduling resources. You can view the operational logs of the ancestor node to check whether the ancestor node is waiting for scheduling resources. If the ancestor node is waiting for scheduling resources, the scheduling resources used by the running nodes in the current workspace have reached the upper limit. You can adjust the scheduled time for running the ancestor node or purchase an exclusive resource group for scheduling to run the ancestor node. For more information about exclusive resource groups for scheduling, see Exclusive resource groups for scheduling.

- Gray

The color gray indicates that the ancestor node fails to be run. In this case, you must view the status of the ancestor node of the current ancestor node.

# 1.9.2. What do I do if I cannot find the instance of a node?

## Problem description

After I commit a node, I cannot find the instance of the node in Operation Center.

## Possible cause

The instance is not generated. The system generates an instance for your node based on the values of **Start Instantiation** and parameters in the **Dependencies** section. The instance may fail to be generated due to the following reasons:

- The node is an isolated node because no **dependent ancestor node** is configured for the node. In this case, the system cannot generate an instance for the node.

  In most cases, this issue occurs if you delete the **dependent ancestor node** configured for the node before you commit the node. As a result, the node becomes an isolated node.

- If you set **Start Instantiation** to **Next Day**, the system generates an instance at the following time points:
  - If you commit your node before 23:30, the system immediately generates an instance for the node. You can view the instance in the instance list of Operation Center on the following day.
  - If you commit your node after 23:30, the system generates an instance at 23:30 of the following day. You can view the instance in the instance list of Operation Center two days after you commit the node.

  You can view the instance of the node in the instance list of Operation Center only after the instance is generated.

- If you set **Start Instantiation** to **Immediately After Deployment**, the system generates an instance at the following time points:

  > **Note** You cannot set **Start Instantiation** to **Immediately After Deployment** for a node that is created in a workflow of DataWorks of an earlier edition.

  - If you commit your node after 23:30, the system does not generate an instance for the node. As a result, you cannot find the instance of the node in the instance list.
  - If you commit your node before 23:30, the system immediately generates an instance for the node. In this case, you can view the instance in the instance list.
    - If the scheduled time for running your node is 10 or more minutes later than the time you commit your node, the system generates an instance for the node and runs the node. Then, you can view the instance in the instance list. For example, you commit your node at 18:00, but your node is scheduled to run at 18:30. In this case, the preceding situation occurs.
    - If the scheduled time for running your node is less than 10 minutes from the time you commit the code, the system generates an instance whose running is complete for the node. The instance is an expired instance that is generated in real time. For example, you commit your node at 18:00, but your node is scheduled to run at 18:05. In this case, the preceding situation occurs.

If you commit your node before 23:30, you can find the instance of the node in the instance list. However, you must check whether the instance is an expired instance.

## Solution

1. Check the parameter configurations in the **Dependencies** section to ensure that the node is not an isolated node.

   You must check whether **dependent ancestor nodes** are configured for the node in the **Dependencies** section of the DataStudio page.

2. Check whether the node is committed before 23:30.

   You must check the time when the node is committed on the **Deploy Tasks** page. If the node is committed after 23:30, you can commit the node again before 23:30 on the following day.

   > **Note** If you set **Start Instantiation** to **Immediately After Deployment**, you can check whether the scheduled time for running the node is 10 or more minutes later than the time you commit the node. If the interval is within 10 minutes, you must commit the node again and ensure that the interval is 10 or more minutes. Otherwise, an expired instance is generated for the node, and the node fails to be run.

# 2.Publish nodes

## 2.1. Delete a node

In some scenarios, you need to delete a node from the development or production environment.

### Delete a node from the development environment

1. Log on to the DataWorks console. In the left-side navigation pane, click Workspaces. On the Workspaces page, find the target workspace and click **Data Analytics** in the Actions column.
2. Search by node type and keyword for the node to be deleted.
3. Right-click the node and select **Delete**. The node is deleted from the development environment.

### Delete a node from the production environment

To delete a node from the production environment, you need to delete the node from the development environment and create a deployment task to deploy the node.

> **Note** Before deleting a node from the production environment, you need to remove its dependencies with child nodes. Otherwise, a message appears, indicating that the node to be deleted has child nodes and cannot be deleted.
>
> To remove the dependencies, follow these steps:
>
> 1. Find each child node of the target node. You can view the dependencies of the target node in the workflow.
> 2. On the configuration tab of a child node, click the Properties tab in the right-side navigation pane, and change the parent node. Alternatively, delete the child node.
>
> If a message appears indicating that the child node has next-level child nodes, repeat these steps to remove the dependencies.

1. Delete the node from the development environment.

   For more information, see the preceding section.

2. Create a deployment task for the deleted node.

   > **Note** Only administrators and administration experts have the permission to create a deployment task. If you are not an administrator or administration expert, we recommend that you seek help from an administration expert.

   i. After deleting the node from the production environment, click **Deploy** in the upper-right corner.

   ii. On the **Create Deploy Task** page, select the node.

   iii. Click **Deploy Selected**.

   You can also click **Add to List** to add the node to **To-Be-Deployed Node List** and deploy the node in the Nodes to Deploy pane.

3. Deploy the node.

In the **Create Deploy Task** dialog box that appears, click **Deploy** to deploy the node.

# 3.Cross-workspace cloning

## 3.1. Overview

For workspaces in basic mode under the same Alibaba Cloud account, you can use the cross-workspace cloning feature to isolate the development environment from the production environment. You can also use this feature to clone and migrate nodes, such as computing or sync nodes, across workspaces. This topic describes how to process the dependencies between nodes during cross-workspace cloning.

If you clone nodes across workspaces by using the **cross-workspace cloning** feature, DataWorks automatically modifies the output names in the destination workspace to distinguish nodes in different workspaces under the same Alibaba Cloud account. This allows you to smoothly clone or maintain node dependencies.

> Note
> - **Cross-workspace cloning** cannot be used across regions.
> - You cannot clone nodes in a workflow of earlier versions to a workspace in the latest version. You must clone the nodes to a folder under **Business Flow** and then clone the entire workflow to the destination workspace.

You can set the owner of cloned nodes to **Default** or **Clone Task Creator**.

- If the owner of the nodes to be cloned is the workspace administrator, you can set the owner of the cloned nodes to **Default** or **Clone Task Creator**.

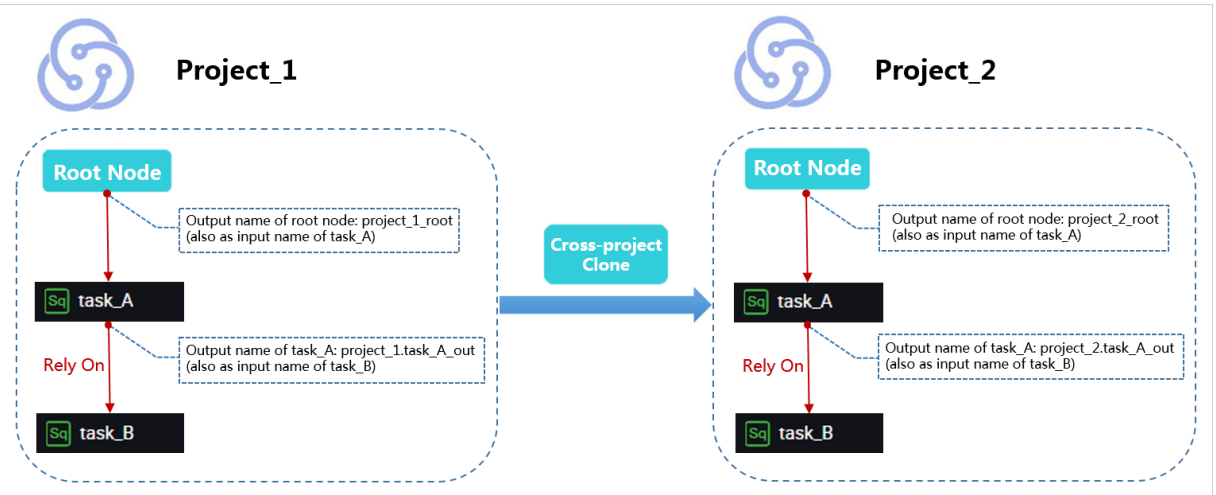
  After the nodes are cloned, their owner is preferentially set to the original owner. If the original owner is not added to the destination workspace, the clone task creator becomes the owner.

- If the owner of the nodes to be cloned is the clone task creator, you can set the owner of the cloned nodes to **Default** or **Clone Task Creator**.

  After the nodes are cloned, their owner is preferentially set to the original owner. If the original owner is not added to the destination workspace, the system asks whether to change the owner. If you agree to change the owner, the clone task creator becomes the owner of the cloned nodes and the clone task succeeds. If you do not agree to change the owner, the clone task is canceled.

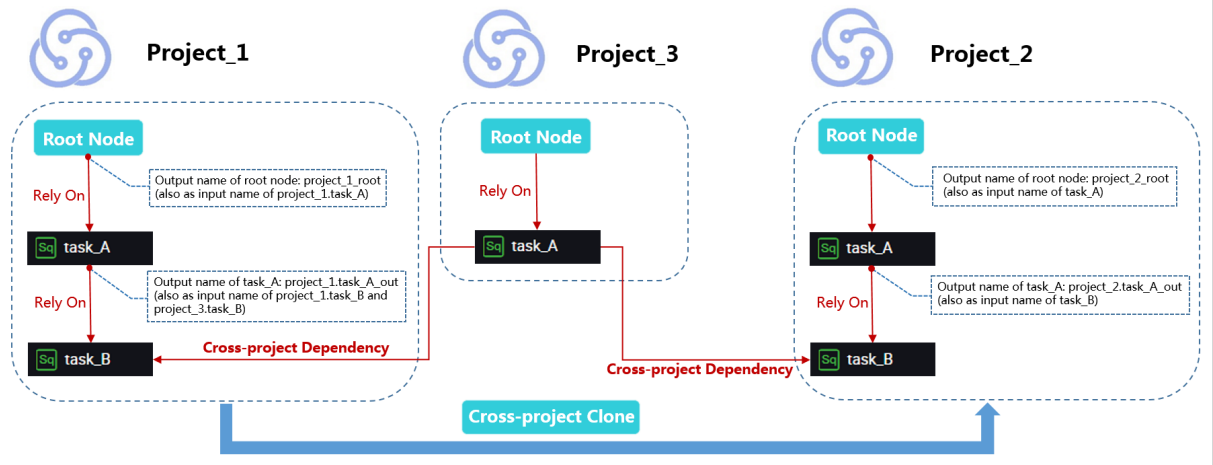
### Complete workflow cloning

Assume that the output name of the task_A node in the project_1 workspace is project_1.task_A_out. If you clone the task_A node to the project_2 workspace, the node output name changes to project_2.task_A_out.




## Cross-workspace dependency cloning

Assume that the task_ B node in the project_1 workspace depends on the task_A node in the project_3 workspace. If you clone the task_B node in the project_1 workspace to the project_2 workspace, the dependency between the task_A and task_B nodes is also cloned. The task_B node in the project_2 workspace also depends on the task_A node in the project_3 workspace.




# 3.2. Clone nodes across workspaces

This topic describes how to clone nodes across workspaces.

## Scenarios

You can clone nodes across workspaces in the following scenarios:

- Clone nodes from a workspace in basic mode to another workspace in basic mode.
- Clone nodes from a workspace in basic mode to another workspace in standard mode.

> Note
>
> - After you clone a node, the folder and workflow to which the node belongs are also cloned to the destination workspace.
> - Before you clone a node, make sure that its ancestor nodes are cloned to the destination workspace.

## Procedure

1. Go to the **DataStudio** page and double-click the target workflow on the Data Analytics tab.
2. Click **Cross-Project Cloning** in the upper-right corner. The Create Clone Task page appears.
3. Set **Target Workspace** and select the nodes to be cloned.
4. Click **Set Compute Engine Mapping**. In the dialog box that appears, set the mapping between the compute engines of the current workspace and the destination workspace.

   If the destination workspace has multiple compute engines, you must set the mapping between compute engines of the current workspace and the destination workspace before cloning nodes. If no mapping is set, the nodes are cloned to the default compute engine of the destination workspace.

   > Note
   >
   > - If the engine type of some nodes to be cloned does not exist in the destination workspace, a message appears in the **Compute Engine Mapping** dialog box. You can select **Skip Nodes Without Target Engine Instance** to ignore these nodes. Otherwise, an error will be reported during cloning.
   > - The **Set Compute Engine Mapping** button only appears when an engine type in the source or destination workspace has more than two engine instances.

5. Click **Add to List**. The selected nodes are added to **To-Be-Cloned Node List**.
6. Click **To-Be-Cloned Node List** in the upper-right corner. On the Nodes to Deploy pane that appears, click **Clone All**.
7. In the **Compute Engine Mapping** dialog box that appears, confirm the engine mapping and click **OK**.
8. After the nodes are cloned, view the cloned nodes in the destination workspace. Generally, the overall folder structure of the workflow is cloned.