CloudMonitor

User Guide

User Guide

The **Overview** page provides an overview of cloud services in terms of resource usage and alarms. It keeps you informed about the resource usages and alarms related to each cloud service in real time.

Cloud service overview

The cloud service overview provides a resource usage and alarm overview for cloud services such as ECS, RDS, OSS, CDN, ApsaraDB for MongoDB, ApsaraDB for Memcache, Container Service, and Log Service.

The cloud service overview keeps you informed about the resource quantity, resource usage, and alarm status under your account.

Clicking the cloud service resource quantity brings you to the monitoring page of the corresponding product. Click the status of an alarm rule to enter the relevant alarm rule page.

Note: To collect the ECS instance CPU, memory, and disk usage data, you must install the CloudMonitor agent. For the agent installation instructions, refer to ECS Monitoring Introduction.

Resource statistical methods: 95th percentile

Percentile is a term used in statistics. To find a percentile, data values are arranged in ascending order, and the corresponding cumulative percentile is calculated. Thus, the data value corresponding to a certain percentile is called the percentile.

The 95th percentile is the value of the 95th percentile. Assuming that the 95th percentile for the CPU usage for all ECS instances is 34%, for all ECS instances, 95% of the instance CPU usage values are less than 34%.

The 95th percentile statistics for various resources show the resource consumption level for the majority of cloud services.

Resource indicator descriptions

Product name	Indicator name	Statistical method	Statistical period	Statistical range
ECS	CPU usage	95th Percentile	Real-time	All instances
ECS	Memory usage	95th Percentile	Real-time	All instances
ECS	Disk usage	95th Percentile	Real-time	All instances

ECS	Outgoing Internet bandwidth	95th Percentile	Real-time	All instances
ApsaraDB for RDS	CPU usage	95th Percentile	Real-time	All instances
ApsaraDB for RDS	IOPS usage	95th Percentile	Real-time	All instances
ApsaraDB for RDS	Connection usage	95th Percentile	Real-time	All instances
ApsaraDB for RDS	Disk usage	95th Percentile	Real-time	All instances
OSS	Total outgoing Internet traffic for current month	Sum	The cumulative value from 00:00 on the first day of the month to the current time	All buckets
OSS	Total PUT requests for current month	Sum	The cumulative value from 00:00 on the first day of the month to the current time	All buckets
OSS	Total GET requests for current month	Sum	The cumulative value from 00:00 on the first day of the month to the current time	All buckets
OSS	Total traffic for current month	Sum	The cumulative value from 00:00 on the first day of the month to the current time	All buckets
CDN	Total traffic for current month	Sum	The cumulative value from 00:00 on the first day of the month to the current time	All domain names
CDN	Peak network bandwidth	95th Percentile	Real-time	All instances
CDN	Cache hit rate	95th Percentile	Real-time	All instances
ApsaraDB for MongoDB	CPU usage	95th Percentile	Real-time	All instances
ApsaraDB for MongoDB	Memory usage	95th Percentile	Real-time	All instances

ApsaraDB for MongoDB	IOPS usage	95th Percentile	Real-time	All instances
ApsaraDB for MongoDB	Connection usage	95th Percentile	Real-time	All instances
ApsaraDB for MongoDB	Disk usage	95th Percentile	Real-time	All instances
ApsaraDB for Memcache	Cache hit rate	95th Percentile	Real-time	All instances
ApsaraDB for Memcache	Cache used	95th Percentile	Real-time	All instances
Container Service	CPU usage	95th Percentile	Real-time	All instances
Container Service	Memory usage	95th Percentile	Real-time	All instances
Container Service	Outgoing Internet traffic	95th Percentile	Real-time	All instances
Log Service	Total incoming network traffic for current month	Sum	The cumulative value from 00:00 on the first day of the month to the current time	All projects
Log Service	Total outgoing network traffic for current month	Sum	The cumulative value from 00:00 on the first day of the month to the current time	All projects
Log Service	Total requests for current month	Sum	The cumulative value from 00:00 on the first day of the month to the current time	All projects

Site monitoring

Statistics are collected on the number of all sites created under your account and the current alarm status for all sites.

Click the number of monitored sites to go to the **Site Monitoring** page. Click the corresponding number of alarm rules to go to the **Alarm Rules** page.

Customized monitoring

Statistics are collected on the number of all custom metrics created under your account and the

current alarm status for all metrics.

Click the number of metrics to go to the **Customized monitoring** page. Click the corresponding number of alarm rules to go to the **Alarm Rules** page.

Dashboard

Overview

With the launch of the dashboard function in CloudMonitor, Alibaba Cloud provides you a one-stop metric visualization solution. It not only allows you to view detailed metrics for troubleshooting, but also gives you the big picture for a glimpse into all services.

Application scenarios

The dashboard function supports customized multi-dimensional query and display of cloud product metric data. The following are some types of typical application scenarios.

Display the metric data trend of multiple instances

For example, if one of your applications is deployed on multiple ECS instances, you can add metric data of these ECS instances to the same metric chart to view the change trend of the metric data of multiple machines.

For example, the CPU usage of multiple ECS instances can be displayed in the time sequence in one chart.

Display the data comparison of multiple metric items

For example, a metric chart can display multiple metrics of an ECS instance, including CPU usage, memory usage, and disk usage.

Display the ordering of machine resource consumption

For example, if you have 20 machines, you can view the CPU usage of them in descending order in a table. This allows you to quickly know about resource consumption, use resources more rationally, and avoid unnecessary cost.

Display the real-time metric data distribution of multiple instances

For example, the CPU usage distribution of an ECS instance group can be displayed in a heat map, so that you can compare the CPU usage of each machine. You can click a color block to view the metric data trend of the corresponding machine in a specified period of time.

Display the aggregated data of a specified metric item of multiple instances

For example, you can view the average aggregation value of the CPU usage of multiple ECS instances in one chart, so as to know about the overall CPU usage and check whether the resource usage of each instance is balanced.

Full screen display

The dashboard supports full screen display and automatic refresh of data. You can add various product metrics to a dashboard to display them in the dashboard in full screen mode.

You can create, modify, delete dashboards, and view charts on them.

View dashboards

Application scenario

The dashboard function of CloudMonitor supports custom display of metric data. You can view metric data on a monitoring dashboard across products and instances, and display instances of different products in a centralized manner.

Note:

CloudMonitor initializes ECS monitoring dashboards for you and displays ECS metric data.

Data of one hour, three hours, and six hours can be automatically refreshed. Data of more than six hours cannot be automatically refreshed.

Dashboard parameter description

Select the time range: You can click the timeframe selection button at the top of the **monitoring dashboard** page to quickly select the timeframe for displaying metric data on the dashboard. The selected time range applies to all charts of the monitoring dashboard.

Automatic refresh: When you click the **Automatic refresh** button, the automatic refresh function is enabled, then you can select the time range of “one hour”, “three hours”, or “six hours” to refresh data every minute.

The unit of metric items is displayed in a bracket in the chart name.

Metric values of all charts at the same time point are displayed as your move your mouse cursor.

Operation procedure

Log on to the CloudMonitor console.

Click the **Dashboard** option in the left menu to access the **Dashboard** page.

By default, **ECS global dashboard** initialized by CloudMonitor is displayed.

Click the monitoring dashboard name and select another monitoring dashboard from the drop-down list.

Click **Full screen** in the top-right corner of the page to view the monitoring dashboard in full screen.

Create a dashboard

Application scenario

If your business is complicated, and the default ECS monitoring dashboards cannot satisfy your monitoring visualization requirements, you can create a new monitoring dashboard and customize the charts to be displayed.

Operation procedure

Log on to the CloudMonitor console.

Click the **Dashboard** option in the left menu to access the **Dashboard** page.

In the top-right corner of the page, click **Add View Group**.

Enter the name of the monitoring dashboard, and click **Create** to complete the creation.

The page is automatically redirected to the new monitoring dashboard page where you can add various metric charts as you like.

Switch dashboards

Application scenario

If you create multiple monitoring dashboards, you can view the monitoring charts of different dashboards by switching monitoring dashboards.

Operation procedure

Log on to the CloudMonitor console.

Click the **Dashboard** option in the left menu to access the **Dashboard** page.

Click the name of a monitoring dashboard in the top-left corner of the page.

All monitoring dashboards created by you are displayed in a drop-down list. You can switch to another dashboard by selecting the name of that dashboard.

Delete a dashboard

Application scenario

You can delete a monitoring dashboard if you do not need it as your business changes.

Note: When you delete a monitoring dashboard, all metric charts added to the dashboard will all be deleted.

Operation procedure

Log on to the CloudMonitor console.

Click the **Dashboard** option in the left menu to access the **Dashboard** page.

In the top-right corner of the page, click the **Delete View Group** button to delete the

dashboard.

Modify a dashboard

Application scenario

You can modify a monitoring dashboard if you need to change the name of it as the content of the monitoring dashboard changes.

Operation procedure

Log on to the CloudMonitor console.

Click the **Dashboard** option in the left menu to access the **Dashboard** page.

Hover your mouse over the name of monitoring dashboard, and the **Change name** option is displayed on the right side. Click **Change name** to make it editable so that you can modify the name of the monitoring dashboard.

Application scenario

CloudMonitor initializes the ECS global dashboard of the user dimension. You can use the **Add cloud product metrics** function to view ECS data of other dimensions or other cloud product metric data.

Note:

By default, CloudMonitor initializes the ECS monitoring dashboard for you. Seven metric charts are displayed, showing the CPU usage, inbound network speed, outbound network speed, system disk BPS, system disk IOPS, network inflow, and network outflow respectively.

Limit of line chart view: A line chart can display 10 lines at most.

Limit of area chart view: An area chart can display 10 areas at most.

Table data limit: The ordered results can be displayed for a maximum of 1,000 data entries.

Limit on heat map view: One heat map can display a maximum of 1,000 color blocks.

Parameter description

Product selection: Choose to view metric data of a specified cloud product.

Metric item: Name of a metric that you need to view, such as outbound network traffic and CPU usage.

Statistical method: Common statistical methods for metric items including maximum value, minimum value, and average value. That is, how metric data is aggregated within the statistical period.

Filter: It is similar to the SQL Where statements and is used to filter metric data source that meets the criteria.

Group By: It is similar to SQL Group By and is used to group metric data that have been filtered by defined dimensions.

> **User dimension**: Group and aggregate metric data on the user account level. For example, if you want to view the average value of the overall memory usage of ECS instances A, B, and C, select **Memory usage** and **Average value** from metric items, select metric items A, B, and C as filter criteria, and set **Group By** to **User dimension**. User dimension is used to view the overall resource usage of multiple instances.
>
> **Instance dimension**: Group and aggregate metric data on the instance level. For example, if you want to view the average value of the memory usage of an ECS instance, select **Memory usage** and **Average value** from metric items, select this instance as filter criteria, and set **Group By** to **Instance dimension**. Instance dimension is used to view the resource usage of a single instance. If you need to view the monitoring status of multiple instances simultaneously, select multiple instances as filter criteria, and set **Group By** to **Instance dimension**.

Chart views: A view can be displayed in line chart, area chart, heat map, pie chart and table.

> **Line chart**: This chart displays metric data by time sequence. Multiple metric items can be added.
>
> **Area chart**: It displays metric data by time sequence. Multiple metric items can be added.

Heat map: It displays the real-time data of metric items. It is used to display distribution and comparison of real-time metric data of a specific metric item of multiple instances. For example, a heat map can display the distribution of the CPU usage of multiple instances. Only one metric item can be added.

Pie chart: This chart displays the real-time metric data, and is usually used for data comparison. Multiple metric items can be added.

Table: It displays metric item value in descending order. For example, a table can display the CPU usage of all machines in an ECS group in descending order. Only one metric item can be added.

Operation procedure

Log on to the CloudMonitor console.

Click the **Dashboard** option in the left menu to access the **Dashboard** page.

Click the **Add cloud product metrics** button in the top-right corner of a monitoring dashboard to access the **Add** page.

Select the cloud product to view and the region of the instance.

Select the product instance.

Select the region of the instance.

Define the chart name and chart type.

Define the chart name. The default chart name generated is “product name + region”.

Select the chart type.

Select the type of metric data to view and the mode of viewing metric data.

Select the metric item to view.

Select the way metric data is aggregated, for example, by maximum value, minimum value or average value.

Select filter criteria.

Select the dimension for **Group By**.

Click the **Add** button and repeat Step 6 if you need to add more metric items.

Click **Publish** to generate a chart in dashboard.

Drag the right border, bottom border or bottom right corner of a chart to resize its height and width (if needed).

Application scenario

By upgrading from custom monitoring to business metric monitoring, you can use the **Add Business Metric Monitoring** function on the data submitted through APIs or SDKs to CloudMonitor for data processing and display in dashboard.

With **Business Metric Monitoring**, metric data can be aggregated by time or space dimension. The time dimension can support the granularity of data aggregation down to a minimum of 1 minute. The space dimension controls the aggregation views with the **Group By** parameter.

Note:

When a chart is added, the data submitted in the last 60 minutes will be read. Therefore, if your data is submitted less frequently than every other 60 minutes, no data will be shown during a preview.

Limit on line chart view: 1 line chart can display up to 15 lines.

Limit on area chart view: 1 area chart may display up to 15 areas.

Table data limit: the ordered results can be displayed for a maximum of 1,000 data entries.

Limit on heat map view: 1 heat map can display a maximum of 1,000 color blocks.

By default, metric data is aggregated at a 1-minute granularity. If your data is submitted

once within less than 1 minute, when performing a query, you will only be able to get data submitted at a minimum granularity of 1 minute.

Parameter description

Chart title: the title of metric chart, displaying the name of metric item by default.

Metric name (required): you can customize name according to the meaning of a metric. It is a parameter for follow-up data query via APIs.

Metric item (required): the name of metric item for which data is submitted via APIs/SDKs.

Unit: the unit that is chosen according to the meaning of your metric.

Filter (optional): equivalent to the Where statement in SQL. If the filtering criteria is left blank, it means to process all the data.

Group By: equivalent to the Group By statement in SQL. The function can aggregate and group metric data by the space or other specified dimension. If no dimension is chosen for Group By, all the metric data will be aggregated using the aggregation methods.

Aggregation: aggregates the metric data within the aggregation period using the specific method. There are three aggregation methods available, including maximum, minimum and average values.

Chart views: a view can be displayed in line chart, area chart, heat map, pie chart and table.

Line chart: this chart displays metric data by time sequence.

Area chart: this chart displays metric data by time sequence.

Heat map: this map displays the real-time metric data, and is usually used to display distribution and comparison of metric data that is grouped by dimension and aggregated.

Pie chart: this chart displays the real-time metric data, and is usually used for data comparison.

Table: this table displays the real-time metric data.

Operation procedure

Log on to the CloudMonitor console.

Click the **Dashboard** option in the left menu to access the **Dashboard** page.

Click the **Add business metrics monitoring** button in the upper right corner of Monitor Dashboard.

Define the **Chart name**, **Metric name** and **Chart type**.

Choose the metric data you want to view and then define the processing method.

Select metric item and unit.

If you only want to view part of the data, select a filtering field.

If you want to aggregate the data grouped by dimension, choose the corresponding field in **Group By**.

Choose an aggregation method.

Click **Publish** to generate a chart in dashboard.

Drag the right border, bottom border or bottom right corner of a chart to resize its height and width (if needed).

Host monitoring

The host monitoring service of CloudMonitor allows you to install an agent on your servers for server system monitoring. Currently, host monitoring supports Linux and Windows operating systems.

Application scenarios

Host monitoring is applicable to Alibaba Cloud ECS servers as well as the servers or physical machines of other vendors. Host monitoring collects statistics using a diverse range of OS-related metrics, allowing you to query server resource usage and obtain metric data for troubleshooting.

Hybrid cloud monitoring solution

CloudMonitor uses an agent to collect server metric data. You can install the agent on a non-ECS server for basic monitoring on and off the cloud.

Enterprise-level monitoring solution

Host monitoring provides the application grouping feature, allowing you to allocate servers in different regions of Alibaba Cloud to the same group for server management from the business perspective. Host monitoring supports group-based alarm management. You need to configure only one alarm rule for the entire group, greatly improving O&M efficiency and management experience.

Note:

Host monitoring supports Linux and Windows, but does not support Unix.

Server resource consumption of the agent: The CloudMonitor agent installation package is 75 MB. After being installed, the agent is 200 MB, with 64-MB memory usage and smaller than 1% CPU usage.

Agent installation requires the root permission.

The TCP status statistics function is similar to the Linux netstat -anp command. When many TCP connections exist, a large amount of CPU time is consumed. Therefore, this function is disabled by default.

To enable this function in Linux, set “netstat.tcp.disable” in the “cloudmonitor/config/conf.properties” configuration file to “false”. Restart the agent after you modify the configuration.

To enable this function in Windows, set “netstat.tcp.disable” in the “C:\Program Files\Alibaba\cloudmonitor\config” configuration file to “false”. Restart the agent after you modify the configuration.

Monitoring capability

CloudMonitor provides more than 30 metric items covering CPU, memory, disk, and network, meeting the basic monitoring and O&M requirements of servers. For a full list of metrics, click here.

Alarm capability

CloudMonitor provides the alarm service for all metric items, allowing you to set alarm rules for individual servers, application groups, and all resources. You can use the alarm service from different business perspectives.

You can use the alarm service directly in the host monitoring list, or use it in your application group after you add servers to the group.

Process monitoring, by default, allows you to collect such information as CPU usage, memory usage and the number of files opened by an active process over a period of time. If you include a process keyword, the number of processes containing the keyword will be returned.

View the consumption status of an active process

Every minute, the agent singles out Top 5 processes with maximum CPU consumption within the last minute, showing their information like CPU usage, memory usage and the number of files they opened.

For the CPU and memory usage for a process, refer to the Linux top command. Here, CPUs means multi-core CPUs.

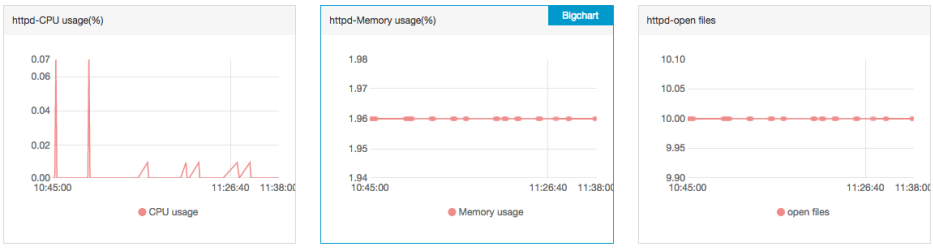
For the number of files opened by an active process, refer to the Linux lsof command.

Note:

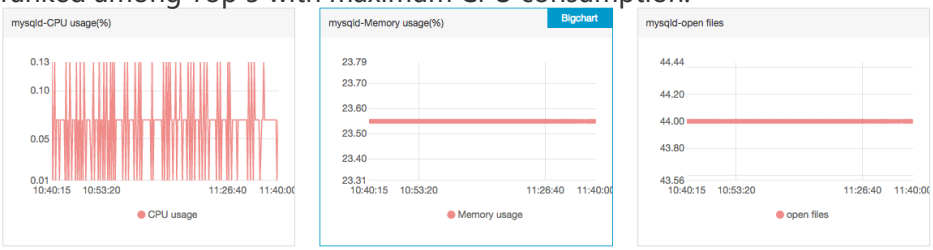
If the Top 5 processes are changing over the time span specified for your query, the process list will show all processes that have ever ranked among Top 5 over the specified time span. The times in the list indicate when any of the processes last ranked among Top 5.

Only on the Top 5 processes, information like CPU usage, memory usage and the number of files they opened will be collected. Therefore, if any of the processes has not ranked among Top 5 continually over the time span specified for the query, its data points will appear discontiguous in the metrics charts. The density of the data points for the process shows its degree of activity on the server.

As shown in the charts below for the http process, which has not continuously ranked among Top 5 processes with maximum server CPU consumption, the data points in the metric charts are sparse and discontiguous. Those data points indicate that the process has ranked among Top 5 at exactly the points of time for the data points.



As shown in the charts below for the mysql process, the data points in the metric charts are very dense and contiguous, indicating that the process has continuously ranked among Top 5 with maximum CPU consumption.



Manage the number of specified processes

You can get the number of key processes and their viability statuses through the process count metric item.

Add a specified process to monitor

Note: When adding a process, you can provide its absolute path or simply its keyword. Refer to the Linux ps aux|grep 'keyword' command.

For example, the server is currently running the following processes.

/usr/bin/java -Xmx2300m -Xms2300m

org.apache.catalina.startup.Bootstrap

/usr/bin/ruby

nginx -c /ect/nginx/nginx.conf

Assume that the user configures 6 keywords, the results are returned as follows respectively:

Keyword: ruby, number of processes returned: 1, hitting a process name.

Keyword: nginx, number of processes returned: 1, hitting a process name and a parameter.

Keyword: /usr/bin, number of processes returned: 2, hitting 2 paths (two processes under the paths respectively).

Keyword: apache.catalina, number of processes returned: 1, hitting part of a parameter.

Keyword: nginx.conf, number of processes returned: 1, hitting part of a parameter.

Keyword: -c, number of processes returned: 1, hitting part of a parameter.

Operation Procedure

Log on to the CloudMonitor console.

Select **Host Monitoring** in the left-side menu to go to the **Host Monitoring** page.

Click the name of the instance you want to monitor. Or click **Metric Chart** in the **Actions** column to access the instance monitoring details page.

Click **Process Monitoring** on top of the page to access the **Process Monitoring** page.

When hovering over the process count metric chart, click **Add Process to Monitor** button to add the process you want to monitor.

Delete a monitored process

Log on to the CloudMonitor console.

Select **Host Monitoring** in the left-side menu to go to the **Host Monitoring** page.

Click the name of the instance you want to monitor. Or click **Metric Chart** in the **Actions** column to access the instance monitoring details page.

Click **Process Monitoring** on top of the page to access the process monitoring page.

When hovering over the process count metric chart, click **Add Process to Monitor** button to access the list of processes added page.

Click **Delete** in the list to delete the process you want to delete.

Host monitoring metrics are divided into agent-collected metrics and ECS native metrics. Agent-collected metrics are collected every 15 seconds, and ECS basic metrics are collected every minute.

Note:

The ECS basic metric data may be inconsistent with the operating system (OS) metric data mainly because:

Different statistical frequencies

Metric chart data are the average values collected during measurement periods. The statistical frequency of basic monitoring is 1 minute, whereas that of OS monitoring is 15 seconds. In the case of large metric data fluctuations, basic metric data is smaller than OS metric data because the former data is de-peaked.

Different statistical perspectives

The network traffic billing data in basic monitoring does not include the unbilled network traffic between ECS and Server Load Balancer, whereas the network traffic statistics in OS monitoring records the actual network traffic of each network adapter. Therefore, the network data in OS monitoring is greater than that in basic monitoring (that is, the agent-collected data is greater than the actually purchased bandwidth or traffic quota).

Agent-collected metrics

CPU metrics

The following lists the metrics of CPU usage. You can refer to the Linux top command to understand the meaning of the metric items.

Metric item	Meaning	Unit	Description
Host.cpu.idle	Percentage of CPU in the idle state	%	
Host.cpu.system	CPU usage of the current kernel space	%	This metric item measures the consumption

			resulting from system context switchover. A great value indicates that many processes or threads are running on the server.
Host.cpu.user	CPU usage of the current user space	%	This metric item measures the CPU consumption of user processes.
Host.cpu.iowait	Percentage of CPU waiting for I/O operation	%	A great value indicates frequent I/O operations.
Host.cpu.other	CPU usage of other items	%	Other types of CPU consumption are calculated using the formula “Nice + SoftIrq + Irq + Stolen”.
Host.cpu.total	Current total CPU usage	%	This metric item measures the aggregate CPU usage of the preceding items and is typically used by the alarm service.

Memory metrics

The following lists the metric items of memory usage. You can refer to the Linux free command to understand the meanings of the metric items.

Metric item	Meaning	Unit	Description
Host.mem.total	Total memory	Bytes	Total memory of the server
Host.mem.used	Memory in use	Bytes	The calculation formula is as follows: “Memory occupied by user programs + Buffers + Cached”. Buffers indicates the memory occupied by the buffer, and cached indicates the memory occupied by the cache.
Host.mem.actualused	Memory occupied by the user	Bytes	The calculation formula is as follows: “Used – Buffers – Cached”.

Host.mem.free	Available memory	Bytes	The calculation formula is as follows: “Total memory – Used memory”.
Host.mem.freeutilization	Percentage of available memory	%	The calculation formula is as follows: “Available memory/Total memory x 100%”.
Host.mem.usedutilization	Memory usage	%	The calculation formula is as follows: “(Actual used/total*100%)”.

Metrics of average system load

The following lists the metrics of average system load. You can refer to the Linux top command to understand the meanings of the metrics. The greater the value, the busier the system.

Metric item	Meaning	Unit
Host.load1	Average system load during the past minute. This metric is not available in Windows.	None
Host.load5	Average system load during the past 5 minutes. This metric is not available in Windows.	None
Host.load15	Average system load during the past 15 minutes. This metric is not available in Windows.	None

Disk metrics

- For disk usage and inode usage metrics, refer to the Linux df command.
- For disk read/write metrics, refer to the Linux iostat command.

Metric item	Meaning	Unit
Host.diskusage.used	Disk space in use	Bytes
Host.disk.utilization	Disk usage	%
Host.diskusage.free	Available disk space	Bytes/s
Host.diskussage.total	Total disk storage	Bytes
Host.disk.readbytes	Number of bytes read from the disk per second	Bytes/s
Host.disk.writebytes	Number of bytes written to	Bytes/s

	the disk per second	
Host.disk.readiops	Number of read requests sent to the disk per second	Requests/s
Host.disk.writeiops	Number of write requests sent to the disk per second	Requests/s

File system metrics

Metric item	Meaning	Unit	Description
Host.fs.inode	Inode usage. Unix and Linux use inode numbers to identify files. When the disk is not full but all inode numbers have been allocated, no more files can be created in the disk. This metric is not available in Windows.	%	Inode numbers indicate the number of files in the file system. Inode usage is high when there are many small files.

Network metrics

The following lists the network metrics. You can refer to the Linux iftop command to understand the meanings of the metrics. For details about TCP connection statistics, refer to the Linux ss command.

By default, statistics are collected on the number of TCP connections by TCP_TOTAL (total connections), ESTABLISHED (normally established connections), and NON_ESTABLISHED (connections not in the established state). If you want to obtain the number of connections in each state, perform the following operations:

> **Linux:** Set netstat.TCP.disable in the "cloudmonitor/config/CONF.properties" configuration file to false to enable data collection. Restart the Agent after you modify the configuration.
>
> **Windows**: Set netstat.tcp.disable in the "C:\Program Files\Alibaba\cloudmonitor\config" configuration file to false to enable data collection. Restart the Agent after you modify the configuration.

Metric item	Meaning	Unit
Host.netin.rate	Number of bits received by the network adapter per second, that is, the uplink bandwidth of the network adapter	bits/s

Host.netout.rate	Number of bits sent by the network adapter per second, that is, the downlink bandwidth of the network adapter	bits/s
Host.netin.packages	Number of packets received by the network adapter per second	packets/s
Host.netout.packages	Number of packets sent by the network adapter per second	packets/s
Host.netin.errorpackage	Number of incoming error packets detected by the drive	packets/s
Host.netout.errorpackages	Number of outgoing error packets detected by the drive	packets/s
Host.tcpconnection	Number of TCP connections in various states, including LISTEN, SYN_SENT, ESTABLISHED, SYN_RECV, FIN_WAIT1, CLOSE_WAIT, FIN_WAIT2, LAST_ACK, TIME_WAIT, CLOSING, and CLOSED	Connections

Process metrics

For details about process-specific CPU usage and memory usage, refer to the Linux top command. CPU usage indicates the CPU consumption of multiple kernels.

For details about Host.process.openfile, refer to the Linux lsof command.

For details about Host.process.number, refer to the Linux ps aux |grep 'keyword' command.

Metric item	Meaning	Unit
Host.process.cpu	CPU usage of a process	%
Host.process.memory	Memory usage of a process	%
Host.process.openfile	Number of files opened by a process	Files
Host.process.number	Number of processes that match the specified keyword	Processes

ECS metrics

If your host is an ECS server, the following metric items are provided without agent installation after you bought an ECS instance. The collection granularity is 1 minute.

Metric item	Meaning	Unit
ECS.CPUUtilization	CPU usage	%
ECS.InternetInRate	Average rate of Internet inbound traffic	bits/s
ECS.IntranetInRate	Average rate of intranet inbound traffic	bits/s
ECS.InternetOutRate	Average rate of Internet outbound traffic	bits/s
ECS.IntranetOutRate	Average rate of intranet outbound traffic	bits/s
ECS.SystemDiskReadbps	Number of bytes read from the system disk per second	bytes/s
ECS.SystemDiskWritebps	Number of bytes written to the system disk per second	bytes/s
ECS.SystemDiskReadOps	Number of times data is read from the system disk per second	times/s
ECS.SystemDiskWriteOps	Number of times data is written to the system disk per second	times/s
ECS.InternetIn	Internet inbound traffic	Bytes
ECS.InternetOut	Internet outbound traffic	Bytes
ECS.IntranetIn	Intranet inbound traffic	Bytes
ECS.IntranetOut	Intranet outbound traffic	Bytes

Host monitoring provides the alarm service. You can set an alarm rule for a single server in host monitoring, or set an alarm rule in application group granularity after you add servers to the specified application group. For details, refer to Set alarm rules for application groups.

Create an alarm rule

Go to the Host Monitoring page of CloudMonitor.

Go to the **Alarm Rules** page.

Click **Create Alarm Rule** in the upper-right corner.

Set alarm parameters on the **New Alarm Rule** page. For details about the parameters, refer to Alarm parameters.

Save the rule settings to create an alarm rule.

Delete an alarm rule

Go to the Host Monitoring page of CloudMonitor.

Go to the **Alarm Rules** page.

Click **Delete** next to an alarm rule to delete this rule. To delete multiple alarm rules at the same time, select the rules and click **Delete** below the list.

Modify an alarm rule

Go to the Host Monitoring page of CloudMonitor.

Go to the **Alarm Rules** page.

Click **Modify** next to an alarm rule to modify this rule.

View alarm rules

Go to the Host Monitoring page of CloudMonitor.

Click **Alarm Rules** in the instance list to view the alarm rule of a single server.

Go to the **Alarm Rules** page to view all alarm rules.

Application group

Application group allows you to manage cloud product resources by group across products and regions and centrally manage service-related resources such as servers, databases, load balancing, and storage from the service perspective. You can manage alarm rules and view metric data from the service perspective, quickly improving O&M efficiency.

Application scenarios

If you have bought multiple Alibaba Cloud products, you can use the application group feature to add resources (such as servers, databases, object storage, and cache) related to the same service to the same application group. You can manage alarm rules and view metric data in the group dimension, greatly reducing management complexity and improving cloud monitoring efficiency.

Note:

A cloud account can create up to 100 application groups.

Up to 1,000 resource instances can be added to one application group.

Application scenarios

If you have bought multiple Alibaba Cloud products, you can use the application group feature to add resources (such as servers, databases, object storage, and cache) related to the same service to the same application group. You can manage alarm rules and view metric data in the group dimension, greatly reducing management complexity and improving O&M efficiency.

Note:

Up to 1,000 resource instances can be added to each application group.

If you select **Initialize Alarm Rule** when you create an application group, CloudMonitor checks whether the average value in 5 minutes exceeds the threshold value based on the resource types of the group. Alarm notifications are sent by email and TradeManager. The notification object is the alarm contact group you selected when creating the group.

The alarm rules for the metric items of the following products are initialized:

Product name	Metric item	Initialized alarm threshold value
Host	CPU usage	80%

Host	Disk usage	90%
Host	Memory usage	80%
Host	Inode usage	90%
ApsaraDB for RDS	CPU usage	80%
ApsaraDB for RDS	Disk usage	80%
ApsaraDB for RDS	IOPS usage	80%
ApsaraDB for RDS	Connection usage	80%
Redis	Percentage of capacity in use	80%
Redis	Percentage of connections in use	80%
Redis	Percentage of QPS in use	80%
Redis	Network bandwidth usage by the write operation	80%
Redis	Network bandwidth usage by the read operation	80%
MongoDB	CPU usage	80%
MongoDB	Memory usage	80%
MongoDB	Disk usage	80%
MongoDB	IOPS usage	80%
MongoDB	Connection usage	80%
Analytic DB	Disk usage	90%
Container service	CPU usage	80%
Container service	Memory usage	80%

Procedure

Log on to the CloudMonitor Console.

Select **Application Group** in the left-side menu to go to the **Application Group** page.

Click **Create Application Group** in the upper-right corner to go to the edit page.

Fill in the group name.

Select the product to be added.

By default, ECS and RDS are initialized. You can click **Add Product** and **Delete Product** to set the product range of the group.

Select the instances to be added to the group from the instance list of the product.

Select alarm notification objects.

Select **Initialize Alarm Rule** based on your needs.

Click **OK** to save the application group settings.

You can create, view, modify, delete, enable, and disable alarm rules in application groups.

Note: When you query alarm rules in the application group dimension, the system displays only the alarm rules applied to the specified application group. The alarm rules applied to instances or all resources are not displayed.

Create an alarm rule

Log on to the CloudMonitor Console.

Select **Application Group** in the left-side menu to go to the **Application Group** page.

Select the application group for which you want to create an alarm rule, and click the group name or **Manage** to go to the application group details page.

Click **Create Alarm Rule** in the upper-right corner.

Complete settings on the alarm rule page.

Delete an alarm rule

Log on to the CloudMonitor Console.

Select **Application Group** in the left-side menu to go to the **Application Group** page.

Select the application group for which you want to create an alarm rule, and click the group name or **Manage** to go to the application group details page.

Click **Alarm Rules** in the upper part of the page to go to the group’ s alarm rules page.

Click **Delete** next to an alarm rule under the **Action** column to delete this rule. To delete multiple alarm rules at the same time, select the rules and click **Delete** below the list.

Modify an alarm rule

Log on to the CloudMonitor Console.

Select **Application Group** in the left-side menu to go to the **Application Group** page.

Select the application group for which you want to create an alarm rule, and click the group name or **Manage** to go to the application group details page.

Click **Alarm Rules** in the upper part of the page to go to the group’ s alarm rules page.

Click **Modify** next to an alarm rule under the **Action** column to modify this rule.

Disable or enable group alarm rules

When you need to stop a service or perform application maintenance and upgrade, you can disable all alarm rules of the application group involved to avoid reception of many useless alarm notifications due to manual changes. After the changes are completed, you can re-enable the alarm rules.

Disable all alarm rules of an application group

Log on to the CloudMonitor Console.

Select **Application Group** in the left-side menu to go to the **Application Group** page.

Select a group name and click **More** under the **Action** column.

Select **Disable All Alarm Rules** under **More** to disable all alarm rules of the selected group.

Enable all alarm rules of an application group

Log on to the CloudMonitor Console.

Select **Application Group** in the left-side menu to go to the **Application Group** page.

Select a group name and click **More** under the **Action** column.

Select **Enable All Alarm Rules** under **More** to enable all alarm rules of the selected group.

Disable partial alarm rules of an application group

Log on to the CloudMonitor Console.

Select **Application Group** in the left-side menu to go to the **Application Group** page.

Select the application group for which you want to create an alarm rule, and click the group name or **Manage** to go to the application group details page.

Click **Alarm Rules** in the upper part of the page to go to the group’ s alarm rules page.

Click **Disable** next to an alarm rule under the **Action** column to disable this rule. To disable multiple alarm rules at the same time, select the rules and click “Disable” below the list.

Enable partial alarm rules of an application group

Log on to the CloudMonitor Console.

Select **Application Group** in the left-side menu to go to the **Application Group** page.

Select the application group for which you want to create an alarm rule, and click the group name or **Manage** to go to the application group details page.

Click **Alarm Rules** in the upper part of the page to go to the group’ s alarm rules page.

Click **Enable** next to an alarm rule under the **Action** column to enable this rule. To enable multiple alarm rules at the same time, select the rules and click **Enable** below the list.

Overview

The group details page includes a fault list, alert history, alarm rules, group resources, events, and group resource metric data.

Application group list

The application group list shows all your application groups on CloudMonitor, as well as the resources and health condition of each group.

Parameters

Group name: the name of an application group.

Health condition: indicates whether group resources have active alarms. The group is healthy when no group resources have active alarms. The group is unhealthy if one resource has an active alarm.

VM count: total number of servers (ECS servers and other servers) in the group.

Resource count: total number of resource types in the group. For example, if the group has ECS, ApsaraDB for RDS, and Server Load Balancer instances, the total number of resource types is three.

Unhealthy Count: total number of instances with active alarms in the group. For example, if two ECS instances and one ApsaraDB for RDS instance have active alarms, the number of unhealthy instances is three.

Creation time: time when the application group is created.

Actions: Four types of operation are supported - copy this group and create a new group, enable all alarm rules, disable all alarm rules, and delete groups.

Fault list

The fault list shows the resources with active alarms in your application group, allowing you to quickly view all unhealthy instances and troubleshoot faults in a timely manner.

Note:

When multiple metrics of a resource have active alarms at the same time, the fault list displays the resource multiple times. Each row of the list shows one metric with an active alarm.

After you disable the rule hit by active alarms, the resources and metrics associated with the rule disappear from the fault list.

Parameters

Faulty resource: resource with an active alarm.

Start time: time when the first alarm is generated.

Status: displays a message indicating that a resource has an active alarm.

Duration: total time during which the faulty resource is in the alarm state.

Alarm Rule name: name of the alarm rule applied to the faulty resource.

Actions: Click **Expand** to show the trends of the faulty resource’ s metric with an active alarm over the past six hours, and compare the metric data with the alarm threshold value.

Alert history

Display the alarm history of all alarm rules applied to an application group.

Note: You can query historical alarms over a period of up to three consecutive days. If the interval between the query start time and end time exceeds three days, the system prompts you to reselect the time range.

Parameters

Faulty resource: resource with an active alarm.

Duration: total time during which the faulty resource is in the alarm state.

Occurrence time: time when the alarm is generated.

Alarm Rule name: name of the alarm rule applied to the faulty resource.

Notification method: method by which alarm notifications are sent. Three methods are supported: SMS, email, and TradeManager.

Product type: product to which the faulty resource belongs.

Status: status of the alarm rule, including the alarm state, cleared state, and channel silence state.

Notification object: a group of contacts who receive alarm notifications.

Alarm rule

Lists all alarm rules applied to an application group. You can select an alarm rule from the list and disable, enable, and modify the rule.

Note: The alarm rule list only displays the alarm rules applied to the specified application group. It does not display the alarm rules with **Resource Range** set to **All Resources** or **Instance**.

Parameters

Rule name: name of an alarm rule, which is specified when the rule is created.

Status: shows whether the resources associated with the alarm rule have active alarms.

Normal state: all resources associated with the alarm rule are normal.

Alarm state: at least one instance associated with the alarm rule has an active alarm.

Insufficient data: at least one instance associated with the alarm rule has insufficient data and no instance has an active alarm.

Enable: indicates whether the alarm rule is enabled.

Product name: name of the product to which group resources belong.

Rule description: a brief description of alarm rule settings.

Actions: The optional operations include **Modify**, **Disable**, **Enable**, **Delete**, and **Alarm History**.

Modify: click to modify the alarm rule.

Disable: click to disable the alarm rule. After the alarm rule is disabled, the alarm service does not check whether metric data exceeds the threshold value.

Enable: click to enable the alarm rule. After you enable a previously disabled alarm rule, the alarm service checks metric data and determines whether to trigger an alarm based on the alarm rule.

Delete: click to delete the alarm rule.

Alarm History: click to view the alarm history of the alarm rule.

Group resources

Display all resources of an application group and the resource health condition.

Parameters

Instance name: instance name or ID of a resource.

Health condition: the resource is healthy if no alarm is generated according to the corresponding alarm rule. The resource is unhealthy if it has an active alarm.

Event

Currently, the alarm service provides the alarm history and records alarm rule operation events (add, delete, and modify), allowing you to trace any operation performed on a specific alarm rule.

Note: You can query event information over the last 90 days.

Parameters

Occurrence time: time when an event occurs.

Event name: alarm generated, alarm cleared, create alarm rule, modify alarm rule, or delete alarm rule.

Event type: events are classified into system events and alarm events. System events include **create alarm rule**, **delete alarm rule**, and **modify alarm rule**. Alarm events include **alarm generated** and **alarm cleared**.

Event details: details of an event.

Metric chart

The lower section of the application group details page displays the monitoring details of group resources. By default, CloudMonitor initializes frequently used metric data. If you want to display more metric data or change the chart type, modify the charts and customize metric data and the chart type.

Note: To obtain the OS metrics of ECS, you need to install the CloudMonitor agent.

Initialized metric data

The following application group data is initialized by default. If you want to view more metric data, click **Add Metric Chart** to add more metrics.

Product category	Metric item	Chart type	Description
ECS	CPU usage and Internet outbound bandwidth	Line chart	Displays the aggregate data of all servers in the group.
ApsaraDB for RDS	CPU usage, disk usage, IOPS usage, connection usage	Line chart	Displays the data of a single database instance.
Server Load Balancer	Outbound bandwidth and inbound bandwidth	Line chart	Displays the data of a single Server Load Balancer instance.
OSS	Storage size and GET/PUT request count	Line chart	Displays the data of a single bucket.
CDN	Downstream bandwidth and hit rate	Line chart	Displays the data of a single domain name.
EIP	Internet outbound bandwidth	Line chart	Displays the data of a single instance.
ApsaraDB for Redis	Memory usage, connection usage, and QPS usage	Line chart	Displays the data of a single instance.

ApsaraDB for MongoDB	CPU usage, memory usage, IOPS usage, and connection usage	Line chart	Displays the data of a single instance.

Application scenarios

When your applications need more cloud products to meet the requirements of service resizing or technical architecture improvement, you need to modify the resources in your application group

When the application O&M and development personnel changes, you need to modify the alarm notification objects of your application group.

Note:

After resources are removed from an application group, the alarm rule configured in the group dimension is no longer applicable to the removed instances.

After an instance is added to an application group, the instance is automatically associated with the alarm rule configured in the group dimension. You do not need to create an alarm rule for the instance.

Procedure

Log on to the CloudMonitor Console.

Select **Application Group** in the left-side menu to go to the **Application Group** page.

Select an application group you want to edit from the group list and go to the group details page.

Click **Modify Group** in the upper-right corner of the page.

Edit the group content.

Click **OK** to save the modification.

Application scenarios

You can quickly create groups with the same alert policies and metric charts using the group copy function, which can simplify group configuration, and allow you to set same alert policies and metric charts for different groups.

Operation procedure

1. Log in to the CloudMonitor console.
2. Select “Group ” in the left-side menu to go to the “Group” page.
3. Select a group to be duplicated on the group list page, click “Operation” > “More” > “Copy Group”.
4. Add instances and notification objects for the new group on the pop-up page. Then, you can duplicate a group with the same alert policies and metric charts.

Cloud service monitoring

Cloud service monitoring is a service that Alibaba Cloud provides for users to monitor the indicators of various cloud products. After buying an instance of a related product, you have access to the relevant monitoring services.

Currently, the following products are supported by the CloudMonitor service. You can click the product to get more information.

ECS

RDS

Server Load Balancer

OSS

EIP

ApsaraDB for Memcache

ApsaraDB for Redis

CDN

Message Service

Log Service

Container Service

E-MapReduce

API Gateway

Auto Scaling

Express Connect

RDS monitoring

Overview

CloudMonitor displays the RDS operation status based on four metrics: **Disk usage**, **IOPS usage**, **Connection usage**, and **CPU usage**. After you buy RDS products, CloudMonitor will automatically start monitoring the above four metrics without any additional operations.

Note:

RDS only provides monitoring and alarm services for primary and read-only instances.

By default, CloudMonitor will create alarm rules for each primary instance and read-only instance. These rules set up the thresholds of **CPU usage**, **Connection usage**, **IOPS usage**, and **Disk usage** all to 80%. When metric data exceeds any of the above thresholds, a text message and email will be sent to the alarm contact account.

Monitoring service

Metric item descriptions

Metric item	Definition	Dimension	Units	Minimum monitoring granularity
Disk usage	The percentage of disk space used by the RDS instance	Instance	Percentage	5 minutes
IOPS usage	The percentage of IO requests per second used by the RDS instance	Instance	Percentage	5 minutes
Connection usage	The connection count is the number of connections that application programs can establish with the RDS instance. Connection usage is the percentage of these connections currently in use.	Instance	Percentage	5 minutes
CPU usage	The percentage of CPU capacity consumed by the RDS instance (CPU performance is determined by the database memory size.)	Instance	Percentage	5 minutes
Memory usage	The percentage of the RDS instance’ s memory in use. Currently, the memory usage metric is only supported by MySQL databases.	Instance	Percentage	5 minutes
Incoming	The instance’ s	Instance	Bps	5 minutes

network traffic	input traffic per second			
Outgoing network traffic	The instance’ s output traffic per second	Instance	Bps	5 minutes

Note: The incoming and outgoing network traffic metrics are only supported by MySQL and SQLServer databases.

View metric data

Log on to the CloudMonitor console.

Go to the **RDS** instance list under **Cloud Service Monitoring**.

Click an instance name in the product instance list or click **Metric Chart** in the **Actions** column to access the **Instance Monitoring Details** page.

(Optional) Click the **Chart Size** button to switch to large chart display mode.

Alarm service

Parameter description

Metric items: The monitoring indicators provided by RDS.

Statistical cycle: The alarm system checks whether your monitoring data has exceeded the alarm threshold value based on the statistical cycle. For example, if the statistical cycle of the alarm rule for memory usage is set to one minute, the system checks whether the memory usage has exceeded the threshold value every other minute.

Statistical method: This sets the method used to determine if the data exceeds the threshold. You can set **Average**, **Maximum**, **Minimum**, and **Sum** in **Statistical method**.

Average: The average value of metric data within a statistical cycle. The statistical result is the average of all metric data collected within 15 minutes. An average value of over 80% is deemed to exceed the threshold.

Maximum: The maximum value of metric data within a statistical cycle. When the

maximum value of the metric data collected within the statistical cycle is over 80%, the value exceeds the threshold.

Minimum: The minimum value of metric data within a statistical cycle. When the minimum value of the metric data collected within the statistical cycle is larger than 80%, the value exceeds the threshold.

Sum: The sum of metric data within the statistical cycle. When the sum of the metric data collected within the statistical cycle is over 80%, it exceeds the threshold. The above statistical methods are needed for traffic-based indicators.

Trigger Alarm After Threshold Value Is Exceeded Several Times: This refers to an alarm which is triggered when the value of the metric item continuously exceeds the threshold value in several consecutive statistical cycles.

For example, you may set the alarm to go off when the CPU usage rate exceeds 80% within a 5-minute statistical cycle after the threshold value is exceeded for three times. If the CPU usage rate is found to exceed 80% for the first time, no warning notification is sent. No alarm is reported if the CPU usage rate exceeds 80% only twice in a row. An alarm is reported only if the CPU usage rate exceeds 80% for a third time. That is, from the first time when the actual data exceeds the threshold to the time when the alarm rule is triggered, the minimum time consumed is **Statistical cycle***(the quantity of consecutive detection times-1) = 5*(3-1) = 10 minutes.

Set an alarm rule

Log on to the CloudMonitor console.

Go to the **RDS** instance list under **Cloud Service Monitoring**.

Click **Alarm Rules** in instance list **Actions** to access the instance' s **Alarm Rules** page.

Click **Create Alarm Rule** at the bottom of the **Alarm Rules** page to create an alarm rule based on the entered parameters.

Server Load Balancer monitoring

CloudMonitor displays the status of Server Load Balancer based on seven metric items, including

inbound traffic and outbound traffic. This helps you to monitor the operational status of instances and allows you to configure alarm rules for these metric items. After you create a Server Load Balancer instance, CloudMonitor will automatically collect data on the metric items listed above.

Monitoring service

Metrics descriptions

Metric	Definition	Dimension	Units	Minimum monitoring granularity
Inbound traffic	Traffic consumed by access to the Server Load Balancer from the Internet	Instance	Bps	1 minute
Outbound traffic	Traffic consumed by access to the Internet from the Server Load Balancer	Instance	Bps	1 minute
Incoming packet count	Number of request packets that the Server Load Balancer receives per second	Instance	Count per second	1 minute
Outgoing packet count	Number of request packets that the Server Load Balancer sends per second	Instance	Count per second	1 minute
New connection count	The number of first-time SYN_SENT statuses for TCP three-way handshakes in a statistical period	Instance	Count	1 minute
Active connection count	The number of connections in the ESTABLISHED status in the current statistical	Instance	Count	1 minute

	period			
Inactive connection count	The number of all TCP connections except connections in the ESTABLISHED status	Instance	Count	1 minute

Note: New connection count, active connection count, and inactive connection count all indicate the TCP connection requests from clients to the Server Load Balancer.

View metric data

Log on to the CloudMonitor console.

Go to the **Server Load Balancer** instance list under **Cloud Service Monitoring**.

Click an instance name in the product instance list or click **Metric Chart** in the **Actions** column to access the **Instance monitoring details** page.

(Optional) Click the **Chart Size** button to switch to large chart display.

Alarm service

Parameter description

Metric items: The monitoring indicators provided by Server Load Balancer.

Statistical cycle: The alarm system checks whether your monitoring data has exceeded the alarm threshold value based on the statistical cycle. For example, if the statistical cycle of the alarm rule for memory usage is set to one minute, the system checks whether the memory usage has exceeded the threshold value every other minute.

Statistical method: This sets the method used to determine if the data exceeds the threshold. You can set **Average**, **Maximum**, **Minimum**, and **Sum** in **Statistical method**.

Average: The average value of metric data within a statistical cycle. The statistical result is the average of all metric data collected within 15 minutes. An average value

of over 80% is deemed to exceed the threshold.

Maximum: The maximum value of metric data within a statistical cycle. When the maximum value of the metric data collected within the statistical cycle is over 80%, the value exceeds the threshold.

Minimum: The minimum value of metric data within a statistical cycle. When the minimum value of the metric data collected within the statistical cycle is larger than 80%, the value exceeds the threshold.

Sum: The sum of metric data within the statistical cycle. When the sum of the metric data collected within the statistical cycle is over 80%, it exceeds the threshold. The above statistical methods are needed for traffic-based indicators.

Trigger Alarm After Threshold Value Is Exceeded Several Times: This refers to an alarm which is triggered when the value of the metric item continuously exceeds the threshold value in several consecutive statistical cycles.

For example, you may set the alarm to go off when the CPU usage rate exceeds 80% within a 5-minute statistical cycle after the threshold value is exceeded for three times. If the CPU usage rate is found to exceed 80% for the first time, no warning notification is sent. No alarm is reported if the CPU usage rate exceeds 80% only twice in a row. An alarm is reported only if the CPU usage rate exceeds 80% for a third time. That is, from the first time when the actual data exceeds the threshold to the time when the alarm rule is triggered, the minimum time consumed is **Statistical cycle***(the quantity of consecutive detection times-1) = 5*(3-1) = 10 minutes.

Set an alarm rule

Log on to the CloudMonitor console.

Go to the **Server Load Balancer** instance list under **Cloud Service Monitoring**.

Click **Alarm Rules** in instance list **Actions** to access the instance’ s **Alarm Rules** page.

Click **Create Alarm Rule** at the bottom of the **Alarm Rules** page to create an alarm rule based on the entered parameters.

OSS monitoring

The OSS monitoring service provides you with metric data which describes basic system operation status, performance, and metering. It also provides a custom alarm service to help you track requests, analyze usage, collect statistics on business trends, and promptly discover and diagnose system problems.

Monitoring service

Metric item descriptions

OSS metric indicators are classified into groups including basic service indicators, performance indicators, and metering indicators. For details, refer to OSS Metric Indicator Reference Manual.

Note: In order to maintain consistency with billing policies, the collection and presentation of metering indicators have the following special features:

Metering indicator data are output by the hour. This means that resource metering information for each hour is combined into a single value that represents the overall metering condition for the hour.

Metering indicator data have an output delay of nearly 30 minutes.

The data time of metering indicator data refers to the start time of the relevant statistical period.

The cutoff time of metering data acquisition is the end time of the last metering data statistical period of the current month. If no metering data are produced in the current month, the metering data acquisition cutoff is 00:00 on the first day of the current month.

A maximum amount of metering indicator data is pushed for presentation. For precise metering data, refer to Consumption Records.

For example, assume that you only use PutObject requests to upload data and perform this operation at an average of 10 times per minute. Then, in the hour between 2016-05-10 08:00:00 and 2016-05-10 09:00:00, the metering data value for your PUT requests will be 600 times (10*60 minutes), the data time will be 2016-05-10 08:00:00, this piece of data will be output at around 2016-05-10 09:30:00. If this piece of data is the last one since 2016-05-01 00:00:00, the metering data acquisition cutoff for the current month is 2016-05-10 09:00:00. If in May 2016, you have not produced any metering data, the metering data acquisition cutoff will be 2016-05-01 00:00:00.

Alarm service

Note: OSS buckets must be globally unique. After deleting a bucket, if you create another bucket with the same name, the monitoring and alarms rules set for the deleted bucket will be applied to the new bucket with the same name.

Besides metering indicators and statistical indicators, alarms rules can be configured for other metric indicators and added to alarm monitoring. Also, multiple alarm rules may be configured for a single metric indicator.

User guide

For information about the alarm service, refer to Alarm Service Overview.

For instructions on how to use the OSS alarm service, refer to OSS Alarm Service User Guide.

CDN monitoring

Overview

CloudMonitor displays the usage of CDN based on nine metric items, including **Queries Per Second (QPS)**, **Bytes Per Second (BPS)**, and **bytes hit rate**. After you add a CDN domain, CloudMonitor automatically monitors the domain.

You can access the **CDN monitoring** page to view the metric data. You can configure alarm rules for metric items so that an alarm is generated when a data exception occurs.

Monitoring service

Metric item descriptions

Metric item	Definition	Dimension	Units	Minimum monitoring granularity
QPS	Total access requests in a specific time interval/Time interval	Instance	Quantity	5 minutes
Peak	The maximum	Instance	Bps	5 minutes

bandwidth BPS	network traffic per unit time			
Hit rate	The probability that request bytes hit the cache in a specific time interval (Bytes = Number of requests x Traffic). The bytes hit rate directly reflects the back-to-source traffic.	Instance	Percentage	5 minutes
Internet outbound traffic	CDN Internet outbound traffic	Instance	Bytes	5 minutes
HTTP Return Code 4xx percentage	Percentage of HTTP Return Code 4xx in a specific time interval	Instance	Percentage	5 minutes
HTTP Return Code 5xx percentage	Percentage of HTTP Return Code 5xx in a specific time interval	Instance	Percentage	5 minutes

View metric data

Log on to the CloudMonitor console.

Go to the **CDN** instance list under **Cloud Service Monitoring**.

Click an instance name in the product instance list or click **Metric Chart** in the **Actions** column to access the **Instance monitoring details** page.

(Optional) Click the **Chart Size** button to switch to large chart display mode.

Alarm service

Parameter description

Metric items: The monitoring indicators provided by CDN.

Statistical cycle: The alarm system checks whether your monitoring data has exceeded the alarm threshold value based on the statistical cycle. For example, if the statistical cycle of the alarm rule for memory usage is set to one minute, the system checks whether the memory usage has exceeded the threshold value every other minute.

Statistical method: This sets the method used to determine if the data exceeds the threshold. You can set **Average**, **Maximum**, **Minimum**, and **Sum** in **Statistical method**.

> **Average**: The average value of metric data within a statistical cycle. The statistical result is the average of all metric data collected within 15 minutes. An average value of over 80% is deemed to exceed the threshold.
>
> **Maximum**: The maximum value of metric data within a statistical cycle. When the maximum value of the metric data collected within the statistical cycle is over 80%, the value exceeds the threshold.
>
> **Minimum**: The minimum value of metric data within a statistical cycle. When the minimum value of the metric data collected within the statistical cycle is larger than 80%, the value exceeds the threshold.
>
> **Sum**: The sum of metric data within the statistical cycle. When the sum of the metric data collected within the statistical cycle is over 80%, it exceeds the threshold. The above statistical methods are needed for traffic-based indicators.

Trigger Alarm After Threshold Value Is Exceeded Several Times: This refers to an alarm which is triggered when the value of the metric item continuously exceeds the threshold value in several consecutive statistical cycles.

For example, you may set the alarm to go off when the CPU usage rate exceeds 80% within a 5-minute statistical cycle after the threshold value is exceeded for three times. If the CPU usage rate is found to exceed 80% for the first time, no warning notification is sent. No alarm is reported if the CPU usage rate exceeds 80% only twice in a row. An alarm is reported only if the CPU usage rate exceeds 80% for a third time. That is, from the first time when the actual data exceeds the threshold to the time when the alarm rule is triggered, the minimum time consumed is **Statistical cycle***(the quantity of consecutive detection times-1) = 5*(3-1) = 10 minutes.

Set an alarm rule

Log on to the CloudMonitor console.

Go to the **CDN** instance list under **Cloud Service Monitoring**.

Click **Alarm Rules** in instance list **Actions** to access the instance' s **Alarm Rules** page.

Click **Create Alarm Rule** at the bottom of the **Alarm Rules** page to create an alarm rule based on the entered parameters.

EIP monitoring

Overview

CloudMonitor provides four EIP metric items (**outbound traffic**, **inbound traffic**, **outgoing packet count**, and **incoming packet count**), to help you monitor the service status. You can set alarm rules for these metric items. After you buy the EIP service, CloudMonitor will automatically collect data on the four metric items listed above.

Monitoring service

Metric item descriptions

Metric item	Definition	Dimension	Units	Minimum monitoring granularity
Inbound traffic	The volume of traffic per minute that passes through the EIP to an ECS instance	Instance	Bytes	1 minute
Outbound traffic	The volume of traffic per minute that passes through the EIP from an ECS instance	Instance	Bytes	1 minute
Incoming packet count	The number of packets per minute that pass through the EIP to an	Instance	Count	1 minute

	ECS instance			
Outgoing packet count	The number of packets per minute that pass through the EIP from an ECS instance	Instance	Count	1 minute

View metric data

Log on to the CloudMonitor console.

Go to the **EIP** instance list under **Cloud Service Monitoring**.

Click an instance name in the product instance list or click **Metric Chart** in the **Actions** column to access the **instance monitoring details** page.

(Optional) Click the **Chart Size** button to switch to large chart display mode.

Alarm service

Parameter description

Metric items: The monitoring indicators provided by EIP.

Statistical cycle: The alarm system checks whether your monitoring data has exceeded the alarm threshold value based on the statistical cycle. For example, if the statistical cycle of the alarm rule for memory usage is set to one minute, the system checks whether the memory usage has exceeded the threshold value every other minute.

Statistical method: This sets the method used to determine if the data exceeds the threshold. You can set **Average**, **Maximum**, **Minimum**, and **Sum** in **Statistical method**.

Average: The average value of metric data within a statistical cycle. The statistical result is the average of all metric data collected within 15 minutes. An average value of over 80% is deemed to exceed the threshold.

Maximum: The maximum value of metric data within a statistical cycle. When the maximum value of the metric data collected within the statistical cycle is over 80%,

the value exceeds the threshold.

Minimum: The minimum value of metric data within a statistical cycle. When the minimum value of the metric data collected within the statistical cycle is larger than 80%, the value exceeds the threshold.

Sum: The sum of metric data within the statistical cycle. When the sum of the metric data collected within the statistical cycle is over 80%, it exceeds the threshold. The above statistical methods are needed for traffic-based indicators.

Trigger Alarm After Threshold Value Is Exceeded Several Times: This refers to an alarm which is triggered when the value of the metric item continuously exceeds the threshold value in several consecutive statistical cycles.

For example, you may set the alarm to go off when the CPU usage rate exceeds 80% within a 5-minute statistical cycle after the threshold value is exceeded for three times. If the CPU usage rate is found to exceed 80% for the first time, no warning notification is sent. No alarm is reported if the CPU usage rate exceeds 80% only twice in a row. An alarm is reported only if the CPU usage rate exceeds 80% for a third time. That is, from the first time when the actual data exceeds the threshold to the time when the alarm rule is triggered, the minimum time consumed is **Statistical cycle***(the quantity of consecutive detection times-1) = 5*(3-1) = 10 minutes.

Set an alarm rule

Log on to the CloudMonitor console.

Go to the **EIP** instance list under **Cloud Service Monitoring**.

Click **Alarm Rules** in instance list **Actions** to access the **Instance' s alarm rules** page.

Click **Create Alarm Rule** at the bottom of the **Alarm Rules** page to create an alarm rule based on the entered parameters.

ApsaraDB for Memcache monitoring

Overview

CloudMonitor provides seven ApsaraDB for Memcache metric items, including **used cache** and **read**

hit rate, to help you monitor the status of the service. You can set alarm rules for these metric items. After you buy the Memcache service, CloudMonitor will automatically collect data on the metric items listed above.

Monitoring service

Metric item descriptions

Metric item	Definition	Dimension	Units	Minimum monitoring granularity
Used cache	Amount of cache in use	Instance	Bytes	1 minute
Read hit rate	The probability that key values (KVs) are read successfully	Instance	Percentage	1 minute
QPS	Total times KVs are read per second	Instance	Count	1 minute
Record count	Total number of KVs in the current measurement period	Instance	Count	1 minute
Cache inbound bandwidth	Traffic generated during access to the cache	Instance	Bps	1 minute
Cache outbound bandwidth	Traffic generated during read operations on the cache	Instance	Bps	1 minute
Eviction	Number of KVs evicted per second	Instance	KVs per second	1 minute

Note:

Metric data are saved for up to 31 days.

You can view metric data for up to 14 consecutive days.

View metric data

Log on to the CloudMonitor console.

Go to the **ApsaraDB for Memcache Monitoring** instance list under **Cloud Service Monitoring**.

Click an instance name in the product instance list or click **Metric Chart** in the **Actions** column to access the **Instance monitoring details** page and view the various indicators.

Click a **Time Range** shortcut on the top of the page or use the specific selection function.

Click the **Zoom In** button in the top-right corner of the metric chart to enlarge the graph.

Alarm service

CloudMonitor provides alarm services for all Memcache metric items. After setting an alarm rule for an important metric item, you will receive an alarm notification if the metric data exceeds the set threshold value. This allows for rapid troubleshooting and reduces the probability of faults.

Parameter descriptions

Metric items: The monitoring indicators provided by ECS for Redis.

Statistical cycle: The alarm system checks whether your monitoring data has exceeded the alarm threshold value based on the statistical cycle. For example, if the statistical cycle of the alarm rule for memory usage is set to one minute, the system checks whether the memory usage has exceeded the threshold value every other minute.

Statistical method: This sets the method used to determine if the data exceeds the threshold. You can set **Average**, **Maximum**, **Minimum**, and **Sum** in **Statistical method**.

Average: The average value of metric data within a statistical cycle. The statistical result is the average of all metric data collected within 15 minutes. An average value of over 80% is deemed to exceed the threshold.

Maximum: The maximum value of metric data within a statistical cycle. When the maximum value of the metric data collected within the statistical cycle is over 80%,

the value exceeds the threshold.

Minimum: The minimum value of metric data within a statistical cycle. When the minimum value of the metric data collected within the statistical cycle is larger than 80%, the value exceeds the threshold.

Sum: The sum of metric data within a statistical cycle. When the sum of the metric data collected within the statistical cycle is over 80%, it exceeds the threshold. The above statistical methods are needed for traffic-based indicators.

Consecutive times: Refers to an alarm which is triggered when the value of the metric item continuously exceeds the threshold value in several consecutive statistical cycles.

For example, you may set the alarm to go off when the CPU usage rate exceeds 80% within a 5-minute statistical cycle after the threshold value is exceeded for three times. If the CPU usage rate is found to exceed 80% for the first time, no warning notification is sent. No alarm is reported if the CPU usage rate exceeds 80% only twice in a row. An alarm is reported only if the CPU usage rate exceeds 80% for a third time. That is, from the first time when the actual data exceeds the threshold to the time when the alarm rule is triggered, the minimum time consumed is **Statistical cycle***(the quantity of consecutive detection times-1) = 5*(3-1) = 10 minutes.

Set an individual alarm rule

Log on to the CloudMonitor console.

Go to the **ApsaraDB for Memcache Monitoring** instance list under **Cloud Service Monitoring** .

Click an instance name in the product instance list or click **Metric Chart** in the **Actions** column to access the **Instance monitoring details** page.

Click the **Bell** button in the top-right corner of the metric chart to set an alarm for the corresponding metric item for this instance.

Batch set alarm rules

Log on to the CloudMonitor console.

Go to the **ApsaraDB for Memcache Monitoring** instance list under **Cloud Service Monitoring** .

Select the appropriate instance on the instance list page. Then, click **Set Alarm Rules** at the bottom of the page to add multiple alarm rules.

ApsaraDB for Redis monitoring

Overview

Cloud Monitor displays the status and usage of ApsaraDB for Redis based on various metric items, including **capacity usage** and **connection usage**. After you create a Redis instance, Cloud Monitor automatically starts monitoring the instance. You can access the **Cloud Monitor Redis** page to view the metric data. You can configure alarm rules for metric items so that an alarm is generated when a data exception occurs.

Monitoring service

Metric item descriptions

Metric item	Definition	Dimension	Units	Minimum monitoring granularity
Capacity used	The current Redis capacity used	Instance	Bytes	1 minute
Used connection count	The total number of client connections	Instance	Count	1 minute
Write speed	Network traffic generated per second during write operations on ApsaraDB for Redis	Instance	Bps	1 minute
Read speed	The network traffic generated per second during read operations on ApsaraDB	Instance	Bps	1 minute

	for Redis			
Failed operation count	Number of failed operations on ApsaraDB for Redis	Instance	Count	1 minute
Capacity usage	Percentage of ApsaraDB for Redis capacity in use	Instance	Percentage	1 minute
Connection usage	Established connections as a percentage of total connections	Instance	Percentage	1 minute
Write bandwidth usage	Percentage of bandwidth consumed by write operations	Instance	Percentage	1 minute
Read bandwidth usage	Percentage of bandwidth consumed by read operations	Instance	Percentage	1 minute

View metric data

Log on to Cloud Monitor console.

Go to the **ApsaraDB for Redis** instance list under **Cloud Service Monitoring**.

Click an instance name in the product instance list or click **Metric Chart** in the **Operation** column to access the **Instance monitoring details** page.

Click the **Chart Size** button to switch to large chart display (optional).

Alarm service

Parameter description

Metric items: The monitoring indicators provided by ECS for Redis.

Statistical cycle: The alarm system checks whether your monitoring data has exceeded the alarm threshold value based on the statistical cycle. For example, if the statistical cycle of the alarm rule for memory usage is set to one minute, the system checks whether the memory usage has exceeded the threshold value every other minute.

Statistical method: This sets the method used to determine if the data exceeds the threshold. You can set **Average**, **Maximum**, **Minimum**, and **Sum** in **Statistical method**.

Average: The average value of metric data within a statistical cycle. The statistical result is the average of all metric data collected within 15 minutes. An average value of over 80% is deemed to exceed the threshold.

Maximum: The maximum value of metric data within a statistical cycle. When the maximum value of the metric data collected within the statistical cycle is over 80%, the value exceeds the threshold.

Minimum: The minimum value of metric data within a statistical cycle. When the minimum value of the metric data collected within the statistical cycle is larger than 80%, the value exceeds the threshold.

Sum: The sum of metric data within a statistical cycle. When the sum of the metric data collected within the statistical cycle is over 80%, it exceeds the threshold. The above statistical methods are needed for traffic-based indicators.

Trigger Alarm After Threshold Value Is Exceeded Several Times: This refers to an alarm which is triggered when the value of the metric item continuously exceeds the threshold value in several consecutive statistical cycles.

For example, you may set the alarm to go off when the CPU usage rate exceeds 80% within a 5-minute statistical cycle after the threshold value is exceeded for three times. If the CPU usage rate is found to exceed 80% for the first time, no warning notification is sent. No alarm is reported if the CPU usage rate exceeds 80% only twice in a row. An alarm is reported only if the CPU usage rate exceeds 80% for a third time. That is, from the first time when the actual data exceeds the threshold to the time when the alarm rule is triggered, the minimum time consumed is **Statistical cycle***(the quantity of consecutive detection times-1) = 5*(3-1) = 10 minutes.

Set an alarm rule

Log on to Cloud Monitor console.

Go to the **ApsaraDB for Redis** instance list under **Cloud Service Monitoring**.

Click **Alarm Rules** in instance list **Operations** to access the instance' s **Alarm Rules** page.

Click C**reate Alarm Rule** at the bottom of the **Alarm Rules** page to create an alarm rule based on the entered parameters.

Message Service monitoring

Overview

CloudMonitor displays the usage of Message Service queues based on the following three metric items: **DelayMessage**, **InactiveMessages**, and **ActiveMessages**. After you create a message queue for Message Service, CloudMonitor automatically starts monitoring the queue. You can access the CloudMonitor **Message Service** page to view the metric data. You can configure alarm rules for metric items so that an alarm is generated when a data exception occurs.

Metric item descriptions

Metric item	Definition	Dimension	Units	Minimum monitoring granularity
ActiveMessages	Total number of active messages in the queue	userId, region, bid, and queue	Count	5 minutes
InactiveMessages	Total number of inactive messages in the queue	userId, region, bid, and queue	Count	5 minutes
DelayMessage	Total number of delayed messages in the queue	userId, region, bid, and queue	Count	5 minutes

View metric data

Log on to the CloudMonitor console.

Go to the **Message Service** instance list under **Cloud Service Monitoring**.

Click an instance name in the product instance list or click **Metric Chart** in the **Actions** column to access the **Instance monitoring details** page.

(Optional) Click the **Chart Size** button to switch to large chart display.

Alarm service

Parameter description

Metric items: The monitoring indicators provided by the Message Service.

Statistical cycle: The alarm system checks whether your monitoring data has exceeded the alarm threshold value based on the statistical cycle. For example, if the statistical cycle of the alarm rule for memory usage is set to one minute, the system checks whether the memory usage has exceeded the threshold value every other minute.

Statistical method: This sets the method used to determine if the data exceeds the threshold. You can set **Average**, **Maximum**, **Minimum**, and **Sum** in **Statistical method**.

Average: The average value of metric data within a statistical cycle. The statistical result is the average of all metric data collected within 15 minutes. An average value of over 80% is deemed to exceed the threshold.

Maximum: The maximum value of metric data within a statistical cycle. When the maximum value of the metric data collected within the statistical cycle is over 80%, the value exceeds the threshold.

Minimum: The minimum value of metric data within a statistical cycle. When the minimum value of the metric data collected within the statistical cycle is larger than 80%, the value exceeds the threshold.

Sum: The sum of metric data within the statistical cycle. When the sum of the metric data collected within the statistical cycle is over 80%, it exceeds the threshold. The above statistical methods are needed for traffic-based indicators.

Trigger Alarm After Threshold Value Is Exceeded Several Times: This refers to an alarm which is triggered when the value of the metric item continuously exceeds the threshold value in

several consecutive statistical cycles.

For example, you may set the alarm to go off when the CPU usage rate exceeds 80% within a 5-minute statistical cycle after the threshold value is exceeded for three times. If the CPU usage rate is found to exceed 80% for the first time, no warning notification is sent. No alarm is reported if the CPU usage rate exceeds 80% only twice in a row. An alarm is reported only if the CPU usage rate exceeds 80% for a third time. That is, from the first time when the actual data exceeds the threshold to the time when the alarm rule is triggered, the minimum time consumed is **Statistical cycle***(the quantity of consecutive detection times-1) = 5*(3-1) = 10 minutes.

Set an alarm rule

Log on to the CloudMonitor console.

Go to the **Message Service** instance list under **Cloud Service Monitoring**.

Click **Alarm Rules** in instance list **Actions** to access the instance' s **Alarm Rules** page.

Click **Create Alarm Rule** at the bottom of the alarm rules page to create an alarm rule based on the entered parameters.

Container Service monitoring

Overview

By monitoring seven indicators including Container Service CPU usage and memory usage, CloudMonitor informs you about Container Service usage. After you create a Container Service instance, CloudMonitor automatically starts monitoring the service. You can access the CloudMonitor **Container Service** page to view the metric data. You can configure alarm rules for metric items so that an alarm notification is generated in case of a data exception.

Monitoring service

Metric item descriptions

Metric item	Definition	Dimension	Units	Minimum monitoring

				granularity
containerCpuUtilization	The container CPU usage	User and container	Percentage	30 seconds
containerMemoryUtilization	The container memory usage	User and container	Percentage	30 seconds
containerMemoryAmount	The container memory usage amount	User and container	Bytes	30 seconds
containerInternetIn	The container’s incoming traffic	User and container	Bytes	30 seconds
containerInternetOut	The container’s outgoing traffic	User and container	Bytes	30 seconds
containerIORead	The container IO read speed	User and container	Bytes	30 seconds
containerIOWrite	The container IO write speed	User and container	Bytes	30 seconds

Note:

Metric data are saved for up to 31 days.

You can view metric data for up to 14 consecutive days.

View metric data

Log on to the CloudMonitor console.

Go to the **Container Service** instance list under **Cloud Service Monitoring**.

Click an instance name in the product instance list or click **Metric Chart** in the **Actions** column to access the **Instance monitoring details** page and view the various indicators.

Click a **Time Range** shortcut on the top of the page or use the specific selection function. Up to 14 consecutive days of metric data can be viewed.

Click the **Zoom In** button in the top-right corner of the **Container Service Monitoring** page to enlarge the graph.

Alarm service

Set individual alarm rules

Click the **Bell** button in the top-right corner of the metric chart to set an alarm for the corresponding metric item for this instance.

Batch set alarm rules

Select the appropriate instance on the **Instance List** page. Then click **Set Alarm Rules** at the bottom of the page to add multiple alarm rules.

Overview

CloudMonitor provides inbound traffic, outbound traffic, response time, and other metric data at the API gateway, and helps you obtain the API gateway service’ s use information. After you activate your API Gateway instance, CloudMonitor automatically starts monitoring the instance. You can access the CloudMonitor API Gateway page to view the metric data. You can configure alert policies for metrics so that an alert is reported when a data exception occurs.

Monitoring service

Metric descriptions

Metric	Meaning	Dimension	Units	Minimum monitoring granularity
Error distribution	Number of times of 2XX, 4XX, and 5XX status codes returned for an API in one monitoring period	User and API	Count	1 minute
Inbound traffic	The sum of traffic of requests from an API in one monitoring period	User and API	Bytes	1 minute
Outbound traffic	The sum of traffic of	User and API	Bytes	1 minute

	requests from an API in one monitoring period			
Response time	The difference between the time when the gateway calls a backend service through an API and the time when the backend service receives the return result in one monitoring period	User and API	seconds	1 minute
Total number of requests	The sum of requests received by an API in one monitoring period	User and API	per time	1 minute

View metric data

Log in to the CloudMonitor console.

Go to the **API gateway** instance list under **Cloud Service Monitoring**.

Click an instance name in the product instance list or click **Metric Chart** in the **Operation** column to access the instance monitoring details page.

Click the Chart Size button to switch to large chart display (optional).

Alert service

Parameter description

Metrics: The monitoring indicators provided by the API Gateway.

Period: The alert system checks whether your monitoring data has exceeded the alert

threshold value based on the period. For example, if the period of the alert policy for memory usage is set to one minute, the system checks whether the memory usage has exceeded the threshold value every other minute.

Statistic method: This sets the method used to determine if the data exceed the threshold. Average, maximum, minimum, and sum can be set in the statistic method.

Average: The average value of metric data within a statistical period. The statistic result is the average of all metric data collected within 15 minutes. An average value of over 80% is deemed to exceed the threshold.

Maximum: The maximum value of metric data within a statistical period. When the maximum value of the metric data collected within the period is over 80%, the value exceeds the threshold.

Minimum: The minimum value of metric data within a statistical period. When the minimum value of the metric data collected within the period is larger than 80%, the value exceeds the threshold.

Sum: The sum of metric data within a statistical period. When the sum of the metric data collected within the period is over 80%, it exceeds the threshold. The above statistic methods are needed for traffic-based indicators.

Trigger Alert After Threshold Value Is Exceeded Several Times: This refers to an alert which is triggered when the value of the metric continuously exceeds the threshold value in several consecutive periods.

For example, you may set the alert to go off when the CPU usage rate exceeds 80% within a 5-minute period after the threshold value is exceeded for three times. If the CPU usage rate is found to exceed 80% for the first time, no alert is triggered. If the CPU usage rate exceeds 80% again in the following 5-minute period, it still does not trigger an alert. An alert is reported only if the CPU usage rate exceeds 80% for a third time in the third period. That is, from the first time when the actual data exceeds the threshold to the time when the alert policy is triggered, the minimum time consumed is: the period*(the quantity of consecutive detection times-1) = 5*(3-1) = 10 minutes.

Set an alert policy

Log in to the CloudMonitor console.

Go to the **API gateway** instance list under **Cloud Service Monitoring**.

Click **Alert Policies** in instance list **Operations** to access the instance' s alert policies page.

Click **Create Alert Policy** at the upper right of the alert policies page to create an alert policy based on the entered parameters.

Auto scaling

Overview

CloudMonitor monitors multiple metrics, such as the minimum and maximum numbers of instances in an auto scaling group. It helps you monitor the status of instances in an auto scaling group and set alert policies for metrics. After you buy the auto scaling service, CloudMonitor will automatically collect data on the metrics listed above.

Monitoring service

Metrics

Metrics provided by CloudMonitor are listed in the table below.

Metric	Dimension	Units	Minimum monitoring granularity
Minimum number of instances	User and auto scaling group	Count	5 minutes
Maximum number of instances	User and auto scaling group	Count	5 minutes
Total number of instances	User and auto scaling group	Count	5 minutes
Number of running instances	User and auto scaling group	Count	5 minutes
Number of instances being added	User and auto scaling group	Count	5 minutes
Number of instances being removed	User and auto scaling group	Count	5 minutes

NOTE

Metric data are saved for up to 31 days.

You can view metric data for up to 14 consecutive days.

View metric data

Log in to the CloudMonitor console.

Go to the auto scaling group list in **Auto scaling** under **Cloud Service Monitoring**.

Click an instance name in the product instance list or click **Metric Chart** in the **Operation** column to access the instance monitoring details page and view the various indicators.

Click a **Time Range** shortcut on the top of the page or use the specific selection function. Up to 14 consecutive days of metric data can be viewed.

Click the **Zoom In** button in the top-right corner of the metric chart to enlarge the graph.

Alert service

Parameter description

Metrics: Monitoring indicators provided by the auto scaling service.

Period: The alert system checks whether your monitoring data has exceeded the alert threshold value based on the period. For example, if the period of the alert policy for memory usage is set to one minute, the system checks whether the memory usage has exceeded the threshold value every other minute.

Statistic method: This sets the method used to determine if the data exceed the threshold. Average, maximum, minimum, and sum can be set in the statistic method.

Average: The average value of metric data within a statistical period. For example, when the statistic result is the average of all metric data collected within 15 minutes. An average value of over 80% is deemed to exceed the threshold.

Maximum: The maximum value of metric data within a statistical period. For example, when the statistic result is the maximum value of all metric data collected within 15 minutes. An average value of over 80% is deemed to exceed the threshold.

Minimum: The minimum value of metric data within a statistical period. For example, when the statistic result is the minimum value of all metric data collected within 15 minutes. An average value of over 80% is deemed to exceed the threshold.

Sum: The sum of metric data within a statistical period. For example, when the statistic result is the sum of all metric data collected within 15 minutes. An average value of over 80% is deemed to exceed the threshold. The above statistic methods are needed for traffic-based indicators.

Consecutive times: Alert that is triggered when the value of the metric continuously exceeds the threshold value in several consecutive periods.

For example, you may set the alert to go off when the CPU usage rate exceeds 80% within a 5-minute period after the threshold value is exceeded for three times. If the CPU usage rate is found to exceed 80% for the first time, no alert is triggered. If the CPU usage rate exceeds 80% again in the following 5-minute period, it still does not trigger an alert. An alert is reported only if the CPU usage rate exceeds 80% for a third time in the third period. That is, from the first time when the actual data exceeds the threshold to the time when the alert policy is triggered, the minimum time consumed is: the period*(the quantity of consecutive detection times-1) = 5*(3-1) = 10 minutes.

Set an individual alert policy

Log in to the CloudMonitor console.

Go to the auto scaling group list in **Auto scaling** under **Cloud Service Monitoring**.

Click an instance name in the product instance list or click **Metric Chart** in the **Operation** column to access the instance monitoring details page.

Click the **Bell** button or **Create Alert Policy** in the top-right corner of the metric chart to set an alert policy for the corresponding metric for this instance.

Batch set alert policies

Log in to the CloudMonitor console.

Go to the **Auto scaling monitoring** instance list under **Cloud Service Monitoring**.

Select the appropriate instance on the instance list page. Then, click **Set Alert Policies** at the bottom of the page to add multiple alert policies.

Express Connect

Overview

CloudMonitor monitors multiple metrics, such as inbound network traffic and outbound network traffic of the Express Connect instance. It helps you monitor the network use of the instance and allows you to set alert policies for the metrics. After you buy the Express Connect service, CloudMonitor will automatically collect data on the metrics listed above.

Monitoring service

Metrics

Metrics provided by CloudMonitor are listed in the table below.

Metric	Dimension	Units	Minimum monitoring granularity
Inbound network traffic	User and instance	Bytes	1 minute
Outbound network traffic	User and instance	Bytes	1 minute
Inbound network bandwidth	User and instance	Bits/s	1 minute
Outbound network bandwidth	User and instance	Bits/s	1 minute

NOTE

Metric data are saved for up to 31 days.

You can view metric data for up to 14 consecutive days.

View metric data

Log in to the CloudMonitor console.

Go to the **Express Connect** instance list under **Cloud Service Monitoring**.

Click an instance name in the product instance list or click **Metric Chart** in the **Operation** column to access the instance monitoring details page and view the various indicators.

Click a **Time Range** shortcut on the top of the page or use the specific selection function. Up to 14 consecutive days of metric data can be viewed.

5. Click the **Zoom In** button in the top-right corner of the metric chart to enlarge the graph.

Alert service

Parameter description

Metrics: the monitoring indicators provided by Express Connect.

Period: The alert system checks whether your monitoring data has exceeded the alert threshold value based on the period. For example, if the period of the alert policy for memory usage is set to one minute, the system checks whether the memory usage has exceeded the threshold value every other minute.

Statistic method: This sets the method used to determine if the data exceed the threshold. Average, maximum, minimum, and sum can be set in the statistic method.

Average: The average value of metric data within a statistical period. For example, when the statistic result is the average of all metric data collected within 15 minutes. An average value of over 80% is deemed to exceed the threshold.

Maximum: The maximum value of metric data within a statistical period. For example, when the statistic result is the maximum value of all metric data collected within 15 minutes. An average value of over 80% is deemed to exceed the threshold.

Minimum: The minimum value of metric data within a statistical period. For example, when the statistic result is the minimum value of all metric data collected within 15 minutes. An average value of over 80% is deemed to exceed the threshold.

Sum: The sum of metric data within a statistical period. For example, when the statistic result is the sum of all metric data collected within 15 minutes. An average value of over 80% is deemed to exceed the threshold. The above statistic methods are needed for traffic-based indicators.

Consecutive times: Alert that is triggered when the value of the metric continuously exceeds the threshold value in several consecutive periods.

For example, you may set the alert to go off when the CPU usage rate exceeds 80% within a 5-minute period after the threshold value is exceeded for three times. If the CPU usage rate is found to exceed 80% for the first time, no alert is triggered. If the CPU usage rate exceeds 80% again in the following 5-minute period, it still does not trigger an alert. An alert is reported only if the CPU usage rate exceeds 80% for a third time in the third period. That is, from the first time when the actual data exceeds the threshold to the time when the alert policy is triggered, the minimum time consumed is: the period*(the quantity of consecutive detection times-1) = 5*(3-1) = 10 minutes.

Set an individual alert policy

Log in to the CloudMonitor console.

Go to the **Express Connect** instance list under **Cloud Service Monitoring**.

Click an instance name in the product instance list or click **Metric Chart** in the **Operation** column to access the instance monitoring details page.

Click the **Bell** button or **Create Alert Policy** in the top-right corner of the metric chart to set an alert policy for the corresponding metric for this instance.

Batch set alert policies

Log in to the CloudMonitor console.

Go to the **Express Connect** instance list under **Cloud Service Monitoring**.

Select the appropriate instance on the instance list page. Then, click **Set Alert Policies** at the bottom of the page to add multiple alert policies.

Alarm service

Overview

You can set alarm rules for probe points in site monitoring, instances in cloud service monitoring, and metric items in customized monitoring.

When you use the alarm function for the first time, you need to create an alarm contact, create an alarm contact group, and then set alarm rules for relevant services.

Site monitoring alarm rules

You can create alarm rules for probe points in site monitoring. The statistical cycle of alarm rules in site monitoring is the same as the detection cycle of probe points. That is, when the detection cycle of a probe point is five minutes, the statistical cycle of its alarm rule is also five minutes. The system monitors the data returned from the probe point every five minutes to check whether the actual value exceeds the threshold value.

Cloud service monitoring alarm rules

You can set alarm rules for instances in cloud service monitoring. Alarm rules can be set for metric items of each product.

Note: The SMS quota for a new user is 1,000 by default. You can submit a ticket or contact Alibaba Cloud through TradeManager to apply for additional free alarm SMS quota.

Manage alarm rules

The alarm service provides the monitoring alarm capability, allowing you to obtain up-to-date metric data for troubleshooting any cloud product abnormality in a timely manner.

Parameter description

Product: Host monitoring, RDS, OSS, and so on.

Resource range: Indicates the range in which an alarm rule takes effect. Three range types are provided: All resources, Application group, and Instance.

> **All resources**: The alarm rule takes effect for all instances of a product under a username. For example, if you set an alarm rule for MongoDB CPU usage greater than 80% and select the all resources range, the alarm is triggered when the CPU usage of a MongoDB instance under your username is greater than 80%.
>
> **Application group**: The alarm rule takes effect for all instances in an application group. For example, if you set an alarm rule for host CPU usage greater than 80% and select the application group range, the alarm is triggered when the CPU usage of a host in the specified application group is greater than 80%.
>
> **Instance**: The alarm rule takes effect only for a specific instance. For example, if you set an alarm rule for host CPU usage greater than 80% and select the instance range, the alarm is triggered when the CPU usage of the specified instance is greater than 80%.

Rule Name: Name of an alarm rule.

Rule Description: Subject of an alarm rule, which describes the conditions that metric data must meet to trigger the alarm.

For example, if you configure rule description as “1-minute average CPU usage >=90%”, the alarm service checks every minute whether the average value of metric data collected during 1 minute is greater than or equal to 90%.

In host monitoring, a single server metric reports one data point every 15 seconds, so 20 data points are reported in 5 minutes.

> “5-minute average CPU usage > 90%” indicates that the average value of the 20 data points about CPU usage in 5 minutes is greater than 90%.
>
> “5-minute CPU usage always > 90%” indicates that the values of the 20 data points about CPU usage in 5 minutes are all greater than 90%.

"5-minute CPU usage once > 90%" indicates that the value of at least one of the 20 data points about CPU usage in 5 minutes is greater than 90%.

"Total 5-minute Internet outbound traffic > 50 MB" indicates that the sum of the values of the 20 data points about Internet outbound traffic in 5 minutes is greater than 5 MB.

Alarm after the threshold value is exceeded multiple times consecutively: An alarm notification is sent in the case that the value has been detected multiple times to meet the alarm rule.

Effective Time: Time when an alarm rule takes effect. The alarm service checks metric data and determines whether to generate an alarm only when the alarm rule is effective.

Notification Object: A group of contacts who receive alarm notifications.

Notification Method: Method by which alarm notifications are sent. Two methods are available: E-mail+TradeManager and mobile phone+E-mail+TradeManager.

Email Remarks: Supplementary information customized for an alarm email. The remarks are sent together with the alarm notification email.

Alarm rule management

CloudMonitor provides three alarm rule management portals: **Application Group** page, monitoring list page for various metric items, and **Alarm Rule List** page of the alarm service.

Manage alarm rules in Application Group monitoring.

Manage alarm rules in Host monitoring.

Manage alarm rules in Cloud Service monitoring.

Use alarm rules in Site monitoring.

Manage alarm contact

The contact and contact group information is a prerequisite for the alarm notification service. You need to create a contact and contact group and select a contact group for the alarm rule to receive the alarm notifications.

Alarm contact management

You can manage the alarm contacts by creating, deleting or modifying the contact telephone, email, or other notification methods.

Create a contact

Log on to the CloudMonitor console.

Go to the **Alarm Contact** page.

Click the **New Contact** button on the right-top corner of the page, and fill in the telephone, email, and other information.

A short message or email is sent to the mobile phone number or email address you fill in for verification. This prevents that you cannot receive the alarm notification in time due to incorrect information.

Edit a contact

Log on to the CloudMonitor console.

Go to the **Alarm Contact** page.

Click **Edit** in the **Actions** column in the contact list to edit the contact information.

Delete a contact

Log on to the CloudMonitor console.

Go to the **Alarm Contact** page.

Click **Delete** in the **Actions** column in the contact list to delete the contact information.

After you delete a contact, no CloudMonitor alarm notification will be sent to the contact.

Alarm contact group management

An alarm group is a group of alarm contacts and may contain one or more alarm contacts. The same alarm contact can be added into multiple alarm contact groups. During the alarm rule setup, the alarm notifications can be sent through alarm contact group.

Create a contact group

Log on to the CloudMonitor console.

Go to the **Alarm Contact** page.

Click the **Alarm Contact Group** menu on the top of the page to switch to the alarm contact group list.

Click **Create a Contact Group** on the right-top corner to access the **Create a Contact Group** page.

Fill in the group name and add desired contacts into the group.

Edit a contact group

Log on to the CloudMonitor console.

Go to the **Alarm Contact** page.

Click the **Alarm Contact Group** menu on the top of the page to switch to the alarm contact group list.

Click **Edit** in the **Actions** column in the contact group list to modify contacts in the contact group.

Delete a contact group

Log on to the CloudMonitor console.

Go to the **Alarm Contact** page.

Click the **Alarm Contact Group** menu on the top of the page to switch to the alarm contact group list.

Click **Delete** in the **Actions** column in the contact group list to delete the contact group.

Batch add contacts to a contact group

Log on to the CloudMonitor console.

Go to the **Alarm Contact** page.

Tick contacts to be added in the alarm contact list.

Click **Add to the alarm contact group** on the page bottom.

Select the contact group on the page prompted and click **OK**.

Event subscription service

CloudMonitor terms of service

Change history

CloudMonitor RAM

Overview

CloudMonitor supports RAM. This allows you to control permissions for Cloud Service Monitoring metric data, alarm rule management, and contact and contact group management through subaccounts.

Note: At present, metric data queries are supported for the following cloud products.

ECS

RDS

Server Load Balancer

OSS

CDN

ApsaraDB for Memcache

EIP

ApsaraDB for Redis

Message Service

Log Service

Permission description

Considerations

In RAM system permissions, the Read-only CloudMonitor access permission only authorizes subaccounts to view metric data. If you want to authorize subaccounts to apply alarm rules, refer to the **Alarm management** section below to learn how to modify or create new authorizations.

Authentication type

Besides basic subaccount permission control, RAM currently supports time, MFA, and IP authentication.

Resource description

At present, RAM does not support fine-grained resource descriptions. Only the “*” wildcard is used for resource authorization.

Operation description

Metric data

Data query actions are divided into two groups: product instance list display and CloudMonitor metric data queries. When authorizing a subaccount to log on to the CloudMonitor portal and view metric data, you must also grant the subaccount permissions for the corresponding product’s instance list and metric data query.

For metric data authorization, simply access the RAM product’s system authorization policy and select **Read-only CloudMonitor access permission**.

Metric data query action: Query*.

Product instance list display actions are as follows.

Product name	Action
ECS	DescribeInstances
RDS	DescribeDBInstances
SLB	DescribeLoadBalancer*
OSS	ListBuckets
OCS	DescribeInstances
EIP	DescribeEipAddresses
ApsaraDB for Redis	DescribeInstances
MNS	ListQueue
CDN	DescribeUserDomains

Alarm management

At present, alarm management does not support fine-grained operations. After being granted the following permissions, a subaccount can add, delete, query, and modify alarm rules, contacts, and contact groups.

If you need to allow a subaccount to use alarm functions, add the following permissions.

```
{
"Version": "1",
"Statement": [
{
"Action": [
"cms:*"
],
"Resource": "*",
"Effect": "Allow"
}
]
}
```