cloudera®

Cloudera Operation

Important Notice

Cloudera, Inc.
395 Page Mill Road
Palo Alto, CA 94306
info@cloudera.com
US: 1-888-789-1488
Intl: 1-650-362-0488
www.cloudera.com

Release Information

Version: Cloudera Enterprise 5.14.x
Date: May 21, 2020

Table of Contents

About Cloudera Operation

This guide shows how to monitor the health of a Cloudera deployment and diagnose issues. You can obtain metrics and usage information and view processing activities. This guide also describes how to examine logs and reports to troubleshoot issues with cluster configuration and operation as well as monitor compliance.

Monitoring and Diagnostics

This section is for system administrators who want to use Cloudera Manager to monitor and diagnose their CDH installation. You can use the Cloudera Manager Admin Console to monitor cluster health, metrics, and usage, view processing activities, and view events, logs, and reports to troubleshoot problems and monitor compliance.

Introduction to Cloudera Manager Monitoring

Cloudera Manager provides many features for monitoring the health and performance of the components of your clusters (hosts, service daemons) as well as the performance and resource demands of the jobs running on your clusters. This guide has information on the following monitoring features:

- **Monitoring Services on page 34** - describes how to view the results of health tests at both the service and role instance level. Various types of metrics are displayed in charts that help with problem diagnosis. Health tests include advice about actions you can take if the health of a component becomes concerning or bad. You can also view the history of actions performed on a service or role, and can view an audit log of configuration changes.
- **Monitoring Hosts on page 46** - describes how to view information pertaining to all the hosts on your cluster: which hosts are up or down, current resident and virtual memory consumption for a host, what role instances are running on a host, which hosts are assigned to different racks, and so on. You can look at a summary view for all hosts in your cluster or drill down for extensive details about an individual host, including charts that provide a visual overview of key metrics on your host.
- **Monitoring Activities on page 51** - describes how to view the activities running on the cluster, both at the current time and through dashboards that show historical activity, and provides many statistics, both in tabular displays and charts, about the resources used by individual jobs. You can compare the performance of similar jobs and view the performance of individual task attempts across a job to help diagnose behavior or performance problems.
- **Events on page 89** - describes how to view events and make them available for alerting and for searching, giving you a view into the history of all relevant events that occur cluster-wide. You can filter events by time range, service, host, keyword, and so on.
- **Alerts** - describes how to configure Cloudera Manager to generate alerts from certain events. You can configure thresholds for certain types of events, enable and disable them, and configure alert notifications by email or using SNMP trap for critical events. You can also suppress alerts temporarily for individual roles, services, hosts, or even the entire cluster to allow system maintenance/troubleshooting without generating excessive alert traffic.
- **Lifecycle and Security Auditing on page 97** - describes how to view service, role, and host lifecycle events such as creating a role or service, making configuration revisions for a role or service, decommissioning and recommissioning hosts, and running commands recorded by Cloudera Manager management services. You can filter audit event entries by time range, service, host, keyword, and so on.
- **Charting Time-Series Data on page 99** - describes how to search metric data, create charts of the data, group (facet) the data, and save those charts to user-defined dashboards.
- **Logs on page 123 -** describes how to access logs in a variety of ways that take into account the current context you are viewing. For example, when monitoring a service, you can easily click a single link to view the log entries related to that specific service, through the same user interface. When viewing information about a user's activity, you can easily view the relevant log entries that occurred on the hosts used by the job while the job was running.
- **Reports on page 126** - describes how to view historical information about disk utilization by user, user group, and by directory and view cluster job activity user, group, or job ID. These reports are aggregated over selected time periods (hourly, daily, weekly, and so on) and can be exported as XLS or CSV files. You can also manage HDFS directories as well, including searching and setting quotas.
- **Troubleshooting Cluster Configuration and Operation on page 132** - contains solutions to some common problems that prevent you from using Cloudera Manager and describes how to use Cloudera Manager log and notification management tools to diagnose problems.

Time Line

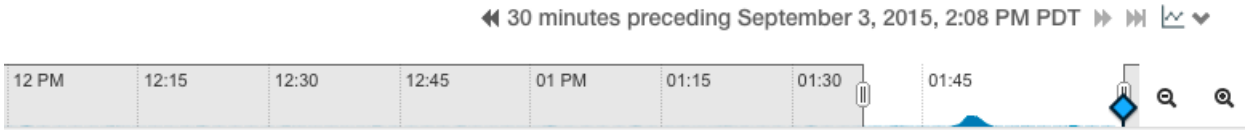
The Time Line appears on many pages in Cloudera Manager. When you view the top level service and Hosts tabs, the Time Line shows status and health only for a specific point in time. When you are viewing the Logs and Events tabs, and when you are viewing the Status, Commands, Audits, Jobs, Applications, and Queries pages of individual services, roles, and hosts, the Time Line appears as a Time Range Selector, which lets you highlight a range of time over which to view historical data.

Click the (📈) icon at the far right to turn on and turn off the display of the Time Line.

Cloudera Manager displays timestamped data using the time zone of the host where Cloudera Manager server is running. The time zone information can be found under the **Support** > **About** menu.

The background chart in the Time Line shows the percentage of CPU utilization on all hosts in the cluster, updated at approximately one-minute intervals, depending on the total visible time range. You can use this graph to identify periods of activity that may be of interest.

In the pages that support a time range selection, the area between the handles shows the selected time range.



There are a variety of ways to change the time range in this mode.

The Reports screen (**Clusters** > **Reports**) does not support the Time Range Selector: the historical reports accessed from the Reports screen have their own time range selection mechanism.

Zooming the Time Line In or Out

Use the Zoom In and Zoom Out buttons (🔍 and 🔍) to zoom the time line graph in or out.

- **Zoom In** shows a shorter time period with more detailed interval segments. Zooming does not change your selected time range. However, the ability to zoom the Time Line can make it easier to use the selector to highlight a time range.
- **Zoom Out** lets you show a longer time period on the time range graph (with correspondingly less granular segmentation).

Selecting a Point In Time or a Time Range

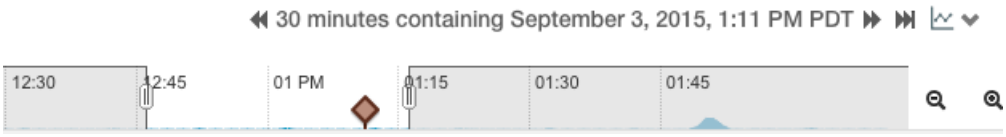
Depending on what page the Time Line appears, you can select a point in time or a time range. There are two ways to look at information about your cluster—its current status and health, or its status and health at some point (or during some interval) in the past. When you are looking at a point in the past, some functions may not be available. For example, on a Service Status page, the **Actions** menu (where you can take actions like stopping, starting, or restarting services or roles) is accessible only when you are looking at **Current** status.

Selecting a Point in Time

Status information on pages such as the service **Status** pages, reflects the state at a single point in time (a snapshot of the health and status). When displayed data is from a single point in time (a snapshot), the panel or column displays a small version of the Time Marker icon (◆) in the panel. This indicates that the data corresponds to the time at the location of the Time Marker on the Time Line.

By default, the status is shown at the current time. If you specify an earlier point on the time range graph, you see the status as it was at the selected point in the past.

- When the Time Marker is set to the current time, it is blue (◆).
- When the Time Marker is set to a time in the past, it is orange (◆).



You can select the point in time in one of the following ways:

- By moving the Time Marker ( )
- When the Time Marker is set to a past time, you can quickly switch back to view the current time using the Now button ( ).
- By clicking the date, choosing the date and time, and clicking **Apply**.

Selecting a Time Range

Pages such as the Logs, Events, and Activities show data over a time range rather than at a single point. These default to showing the past 30 minutes of data (ending at the current time). The charts that appear on the individual Service Status and Host Status pages also show data over a time range. For this type of display, there are several ways to select a time range of interest:

- Drag one (or both) edges of the time range handles to expand or contract the range.
- Choose a duration by clicking a duration link

  30m 1h 2h 6h 12h 1d 7d 30d

  and then do one of the following:
  - Click the next or previous buttons to select the next or previous duration.
  - Click somewhere in the dark portion of the time range to choose the selected duration.
- Click the date range

  30 minutes containing September 3, 2015, 1:11 PM PDT

  to open the time selection widget. Enter a start and end time and click **Apply** to put your choice into effect.
- When you are under the **Clusters** tab with an individual activity selected, a **Zoom to Duration** button is available. This lets you zoom the time selection to include just the time range that corresponds to the duration of your selected activity.

Health Tests

Cloudera Manager monitors the health of the services, roles, and hosts that are running in your clusters using ***health tests***. The Cloudera Management Service also provides health tests for its roles. Role-based health tests are enabled by default. For example, a simple health test is whether there's enough disk space in every NameNode data directory. A more complicated health test may evaluate when the last checkpoint for HDFS was compared to a threshold or whether a DataNode is connected to a NameNode. Some of these health tests also aggregate other health tests: in a distributed system like HDFS, it's normal to have a few DataNodes down (assuming you've got dozens of hosts), so we allow for setting thresholds on what percentage of hosts should color the entire service down.

Health tests can return one of three values: **Good**, **Concerning**, and **Bad**. A test returns **Concerning** health if the test falls below a warning threshold. A test returns **Bad** if the test falls below a critical threshold. The overall health of a service or role instance is a roll-up of its health tests. If any health test is **Concerning** (but none are **Bad**) the role's or service's health is **Concerning**; if any health test is **Bad**, the service's or role's health is **Bad**.

In the Cloudera Manager Admin Console, health tests results are indicated with colors: **Good** , **Concerning** , and **Bad** .

There are two types of health tests:

- **Pass-fail tests** - there are two types:
  - Compare a property to a yes-no value. For example, whether a service or role started as expected, a DataNode is connected to its NameNode, or a TaskTracker is (or is not) blacklisted.

  - Exercise a service lightly to confirm it is working and responsive. HDFS (NameNode role), HBase, and ZooKeeper services perform these tests, which are referred to as "canary" tests.

  Both types of pass-fail tests result in the health reported as being either **Good** or **Bad**.
- **Metric tests** - compare a property to a numeric value. For example, the number of file descriptors in use, the amount of disk space used or free, how much time spent in garbage collection, or how many pages were swapped to disk in the previous 15 minutes. In these tests the property is compared to a threshold that determine whether everything is **Good**, (for example, plenty of disk space available), whether it is **Concerning** (disk space getting low), or is **Bad** (a critically low amount of disk space).

By default most health tests are enabled and (if appropriate) configured with reasonable thresholds. You can modify threshold values by editing the monitoring properties under the entity's **Configuration** tab. You can also enable or disable individual or summary health tests, and in some cases specify what should be included in the calculation of overall health for the service, role instance, or host. See Configuring Monitoring Settings on page 15 for more information.

Viewing Health Test Results

Health test results are available in the following locations:

- **Home** > **Status** tab where various health results determine an overall health assessment of the service or role. The overall health of a role or service is a roll-up of its health tests; if any health test is **Bad**, the service's or role's health will be **Bad**. If any health test is **Concerning** (but none are **Bad**) the role's or service's health will be **Concerning**.
- **Hosts** tab, which shows summary result for the hosts.
- **Status** tab - which shows metrics for services, role instances, and hosts. These are reflected in the results shown in the **Health Tests** panel when you have selected a service, role instance, or host.
- The **All Health Issues** tab of the **Home** page displays all health issues. You can sort the display by entity or by Health Test.

For some health test results, you can chart the associated metrics over a time range. See Viewing Service Status on page 35, Viewing Role Instance Status on page 40, and Host Details on page 47 for details.

Suppressing Health Test Results

Cloudera Manager displays warnings when health tests indicate a problem in the cluster. Sometimes these warnings are expected or do not indicate a real problem in your deployment. You can suppress display of these warnings in Cloudera Manager.

You can suppress health test warnings as they appear or before any tests run. Suppressed health tests are hidden in Cloudera Manager and their status does not affect the roll-up of health tests that display for a service, host, or role instance. (If your cluster is monitored by a Multi Cloudera Manager Dashboard, the roll-up of health tests displayed there is also not affected by suppressed warnings.) Suppressed health test warnings remain available in Cloudera Manager, and the tests continue to run but the results are hidden. You can unsuppress a suppressed health test at any time.

> **Note:** *Suppressing* a health test is different than *disabling* a health test. A disabled health test never runs, whereas a suppressed health test runs but its results are hidden.

Suppressing a Health Test

1. Go to the health test you want to suppress. (See Viewing Health Test Results on page 9.)
2. Click the **Suppress...** link to the right of the health test description.

   A dialog box opens where you can enter a comment about the suppression action.
3. Click **Confirm**.

   The display changes to **Suppressing...** while the change is propagated.

Managing Suppressed Health Tests

On pages where you have suppressed validations, you will see a link that says **Show # Suppressed Test**. On this screen, you can:

- Click the **Show # Suppressed Test** link to view all suppressed health tests for the page.
- Click the **Unsuppress...** link to unsuppress the health test.
- Click **Hide Suppressed Tests** to re-hide the suppressed tests.

Configuring Suppression of Health Tests Before Tests Run

1. Go to the service or host with the health test you want to suppress.
2. Click the **Configuration** tab.
3. In the filters on the left, select **Category** > **Suppressions**.

   A list of suppression properties displays. The names of the properties begin with **Suppress Health Test**.
4. Select a health test suppression property to suppress the test.
5. Click **Save Changes** to commit the changes.

Viewing a List of Suppressed Health Tests

1. From the **Home** page or the **Status** page of a cluster, select **Configuration** > **Suppressed Health and Configuration Issues**.
2. Select **Status** > **Non-default**.

   A list of suppressed health tests and configuration issues displays.
3. To limit the list to health tests, enter “health test” in the **Search** box.

Unsuppressing Health Tests

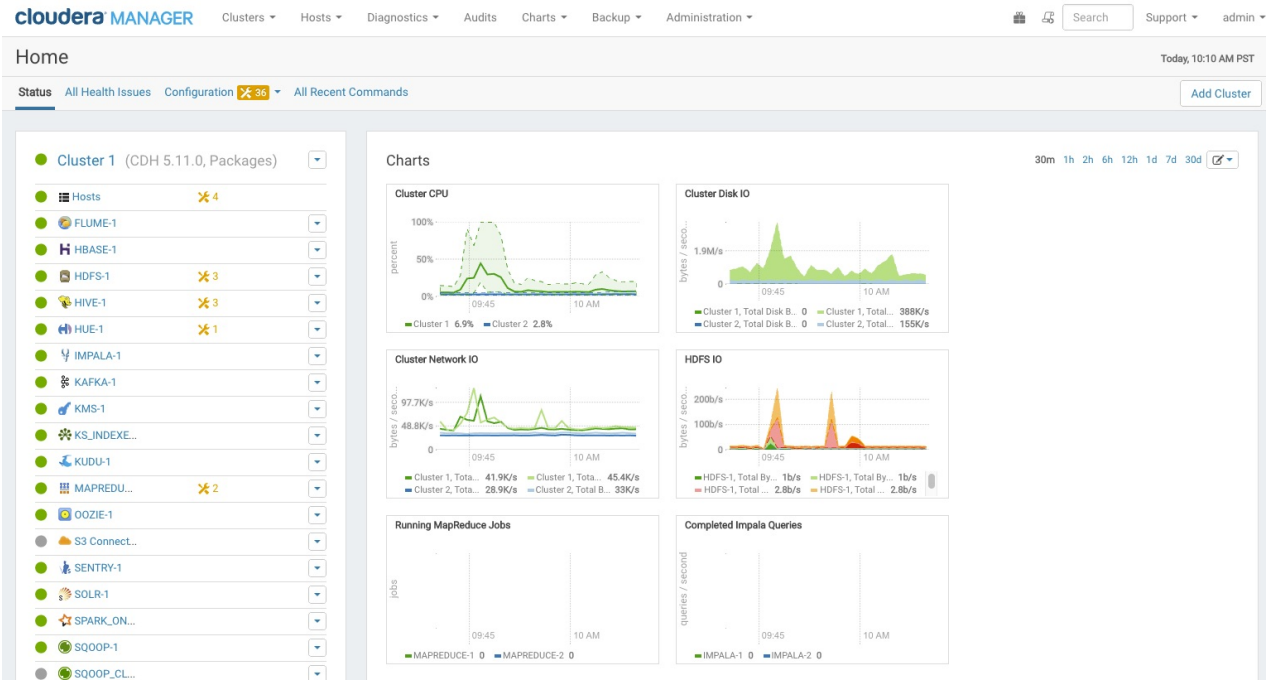
You can unsuppress a health test by doing one of the following:

- To unsuppress a single health test where it displays, click the **Unsuppress...** link next to a suppressed test. (You may need to click the **Show # Suppressed Test** link first.)
- To unsuppress one or more health tests from the configuration screen:
  1. Go to the service or host with the health test you want to unsuppress.
  2. Select **Status** > **Non-default**.

     A list of suppressed health tests and configuration issues displays.
  3. (Optional) Type the name of the health test in the **Search** box to locate it.
  4. Clear the suppression property for the health test.
  5. Click **Save Changes** to commit the changes.

Cloudera Manager Admin Console Home Page

When you start the Cloudera Manager Admin Console, the **Home** > **Status** tab displays.



You can also go to the **Home** > **Status** tab by clicking the Cloudera Manager logo in the top navigation bar.

Status

The Status tab contains:

- **Clusters** - The clusters being managed by Cloudera Manager. Each cluster is displayed either in summary form or in full form depending on the configuration of the **Administration** > **Settings** > **Other** > **Maximum Cluster Count Shown In Full** property. When the number of clusters exceeds the value of the property, only cluster summary information displays.
  - **Summary Form** - A list of links to cluster status pages. Click **Customize** to jump to the **Administration** > **Settings** > **Other** > **Maximum Cluster Count Shown In Full** property.
  - **Full Form** - A separate section for each cluster containing a link to the cluster status page and a table containing links to the Hosts page and the status pages of the services running in the cluster.

    Each service row in the table has a menu of actions that you select by clicking

    and can contain one or more of the following indicators:

Indicator	Meaning	Description
2	Health issue	Indicates that the service has at least one health issue. The indicator shows the number of health issues at the highest severity level. If there are Bad health test results, the indicator is red. If there are no Bad health test results, but Concerning test results exist, then the indicator is yellow. No indicator is shown if there are no Bad or Concerning health test results. **Important:** If there is one Bad health test result and two Concerning health results, there will be three health issues, but the number will be one. Click the indicator to display the **Health Issues** pop-up dialog box. By default only Bad health test results are shown in the dialog box. To display Concerning health test results, click the **Also show *n* concerning issue(s)**

Indicator	Meaning	Description
		link.Click the link to display the Status page containing with details about the health test result.
✕4	Configuration issue	Indicates that the service has at least one configuration issue. The indicator shows the number of configuration issues at the highest severity level. If there are configuration errors, the indicator is red. If there are no errors but configuration warnings exist, then the indicator is yellow. No indicator is shown if there are no configuration notifications. **Important:** If there is one configuration error and two configuration warnings, there will be three configuration issues, but the number will be one. Click the indicator to display the **Configuration Issues** pop-up dialog box. By default only notifications at the Error severity level are listed, grouped by service name are shown in the dialog box. To display Warning notifications, click the **Also show *n* warning(s)** link.Click the message associated with an error or warning to be taken to the configuration property for which the notification has been issued where you can address the issue.See Managing Services.
Restart Needed Refresh Needed	Configuration modified	Indicates that at least one of a service's roles is running with a configuration that does not match the current configuration settings in Cloudera Manager. Click the indicator to display the Stale Configurations page.To bring the cluster up-to-date, click the Refresh or Restart button on the Stale Configurations page or follow the instructions in Refreshing a Cluster, Restarting a Cluster, or Restarting Services and Instances after Configuration Changes.
	Client configuration redeployment required	Indicates that the client configuration for a service should be redeployed. Click the indicator to display the Stale Configurations page.To bring the cluster up-to-date, click the Deploy Client Configuration button on the Stale Configurations page or follow the instructions in Manually Redeploying Client Configuration Files.

- **Cloudera Management Service** - A table containing a link to the Cloudera Manager Service. The Cloudera Manager Service has a menu of actions that you select by clicking
  
  .
- **Charts** - A set of charts (dashboard) that summarize resource utilization (IO, CPU usage) and processing metrics.

  Click a line, stack area, scatter, or bar chart to expand it into a full-page view with a legend for the individual charted entities as well more fine-grained axes divisions.

  By default the time scale of a dashboard is 30 minutes. To change the time scale, click a duration link
  
  at the top-right of the dashboard.

  To set the dashboard type, click  and select one of the following:

  - **Custom** - displays a custom dashboard.
  - **Default** - displays a default dashboard.
  - **Reset** - resets the custom dashboard to the predefined set of charts, discarding any customizations.

All Health Issues

Displays all health issues by cluster. The number badge has the same semantics as the per service health issues reported on the Status tab.

- By default only Bad health test results are shown in the dialog box. To display Concerning health test results, click the **Also show *n* concerning issue(s)** link.
- To group the health test results by entity or health test, click the buttons on the **Organize by Entity/Organize by Health Test** switch.
- Click the link to display the Status page containing with details about the health test result.

All Configuration Issues

Displays all configuration issues by cluster. The number badge has the same semantics as the per service configuration issues reported on the Status tab. By default only notifications at the Error severity level are listed, grouped by service name are shown in the dialog box. To display Warning notifications, click the **Also show *n* warning(s)** link. Click the message associated with an error or warning to be taken to the configuration property for which the notification has been issued where you can address the issue.

All Recent Commands

Displays all commands run recently across the clusters. A badge  indicates how many recent commands are still running. Click the command link to display details about the command and child commands. See also Viewing Running and Recent Commands on page 44.

Starting and Logging into the Cloudera Manager Admin Console

1. In a web browser, enter `http://Server host:7180`, where *Server host* is the FQDN or IP address of the host where the Cloudera Manager Server is running.

   The login screen for Cloudera Manager Admin Console displays.

2. Log into Cloudera Manager Admin Console using the credentials assigned by your administrator. User accounts are assigned roles that constrain the features available to you.

> **Note:** You can configure the Cloudera Manager Admin Console to automatically log out a user after a configurable period of time. See Automatic Logout.

Displaying the Cloudera Manager Server Version and Server Time

To display the version, build number, and time for the Cloudera Manager Server:

1. Open the Cloudera Manager Admin Console.
2. Select **Support** > **About**.

Viewing Charts for Cluster, Service, Role, and Host Instances

For cluster, service, role, and host instances you can see dashboards of charts of various metrics relevant to the entity you are viewing. While the metrics displayed are different for each entity, the basic functionality works in the same way.

The **Home** > **Status** tab for clusters and the Status tab for a service, role, or host display dashboards containing a limited set of charts.

The Status page Charts Library tab displays a dashboard containing a much larger set of charts, organized by categories such as process charts, host charts, CPU charts, and so on, depending on the entity (service, role, or host) that you are viewing.

A custom dashboard is displayed by default when you view the Status tab for an entity. You can switch between custom and default dashboards by using the edit button

to the upper right of the chart.

Displaying Information from Charts

There are various ways to display information from charts.

- Click the ⚙ icon at the top right to see a menu for opening the chart in the Chart Builder or exporting its data.
- Change the size of a chart on a dashboard by dragging the lower-right corner of the chart.
- Hovering with the mouse over a stream on a chart (for example, a line on a line chart) opens a small pop-up window that displays information about that stream. Move the mouse horizontally to see the data values change in the small pop-up window, based on the time represented at the mouse's position along the chart's horizontal axis. Click any stream within the chart to display a larger pop-up window that includes additional information for the stream at the point in time where the mouse was clicked. At the bottom of the large pop-up window is a button for viewing the Cloudera Manager page for the entity (service, host, role, query, or application) associated with the chart, if applicable (**View Service**, **View Host**, and so on). Click the button **View Entity Chart** to display a chart for the stream on its own page. If the chart displays more than one stream, the new chart displays only the stream that was selected when the button was clicked.
- The chart page includes an editable text field containing a default title based on the select statement that was used to create the chart. This title will be used if you save the chart as a dashboard. Type a new title for the chart into this field, if desired.

Exporting Data from Charts

The menu displayed by clicking the ⚙ icon at the top right includes the selections **Export JSON**, and **Export CSV**.

- Click **Export JSON** to display the chart data in JSON format in a new browser window.
- Click **Export CSV** to open a **Save** dialog box enabling you to save the data as a CSV file, choose a program to open the CSV, or open the file with your system's default program for editing and displaying CSV files.

> **Note:** Time values that appear in Cloudera Manager charts reflect the time zone setting on the Cloudera Manager client machine, but time values returned by the Cloudera Manager API (including those that appear in JSON and CSV files exported from charts) reflect Coordinated Universal Time (UTC). For more information on the timestamp format, see the Cloudera Manager API documentation, for example, `ApiTimeSeriesData.java`.

Adding and Removing Charts from a Dashboard

Minimum Required Role: Configurator (also provided by **Cluster Administrator, Full Administrator**)

- With a custom dashboard, the menu displayed by clicking the ⚙ icon at the top right includes the selection **Remove** for users with the required roles. The **Remove** button does not appear in the menu when the default dashboard is used because the default dashboard does not allow removing the original charts. Use the edit button  to the upper right of the chart to switch between custom and default dashboards.
- Charts can also be added to a custom dashboard. Click the ⚙ icon at the top right and click **Add to Dashboard**. You can add the chart to an existing dashboard by selecting **Add chart to an existing custom or system dashboard** and selecting the dashboard name. Add the chart to a new dashboard by clicking **Add chart to a new custom dashboard** and enter a new name in the **Dashboard Name** field.

Creating Triggers from Charts

Minimum Required Role: Full Administrator

- For many charts, the menu opened with the ⚙ icon will also include **Create Trigger**. Triggers allow you to define actions to be taken when a specified condition is met. For information on creating triggers, see Triggers on page 91.

Configuring Monitoring Settings

Minimum Required Role: Configurator (also provided by **Cluster Administrator, Full Administrator**)

There are several types of monitoring settings you can configure in Cloudera Manager:

- Health tests - For a service or role for which monitoring is provided, you can enable and disable selected health tests and events, configure how those health tests factor into the overall health of the service, and modify thresholds for the status of certain health tests. For hosts you can disable or enable selected health tests, modify thresholds, and enable or disable health alerts.
- Free space - For hosts, you can set threshold-based monitoring of free space in the various directories on the hosts Cloudera Manager monitors.
- Activities - For MapReduce, YARN, and Impala services, you can configure aspects of how Cloudera Manager monitors activities, applications, and queries.
- Alerts - For all roles you can configure health alerts and configuration change alerts. You can also configure some service specific alerts and how alerts are delivered.
- Log events - For all roles you can configure logging thresholds, log directories, log event capture, when log messages become events, and when to generate log alerts.
- Monitoring roles - For the Cloudera Management Service you can configure monitoring settings for the monitoring roles themselves—enable and disable health tests on the monitoring processes as well as configuring some general settings related to events and alerts (specifically with the Event Server and Alert Publisher). Each of the Cloudera Management Service roles has its own parameters that can be modified to specify how much data is retained by that service. For some monitoring functions, the amount of retained data can grow very large, so it may become necessary to adjust the limits.

For general information about modifying configuration settings, see Modifying Configuration Properties Using Cloudera Manager.

Configuring Health Monitoring

The initial health monitoring configuration is handled during the installation and configuration of your cluster, and most monitoring parameters have default settings. However, you can set or modify these at any time.

Depending on the service or role you select, and the configuration category, you can enable or disable health tests, determine when health tests cause alerts, or determine whether specific health tests are used in computing the overall health of a role or service. In most cases you can disable these "roll-up" health tests separately from the individual health tests.

As a rule, a health test whose result is considered "Concerning" or "Bad" is forwarded as an event to the Event Server. That includes health tests whose results are based on configured Warning or Critical thresholds, as well pass-fail type health tests. An event is also published when the health test result returns to normal.

You can control when an individual health test is forwarded as an event or as an alert by modifying the threshold values for the relevant health test.

Configuring Service Monitoring

1. Select **Clusters** > ***cluster_name*** > ***service_name***.
2. Click the **Configuration** tab.
3. Select **Scope** > ***service_name*** **(Service-Wide)**.
4. Select **Category** > **Monitoring**.
5. Locate the property to change or search for it by typing its name in the Search box.
6. Configure the property.
7. Click **Save Changes** to commit the changes.
8. Return to the Home page by clicking the Cloudera Manager logo.
9. Click the icon next to any stale services to invoke the cluster restart wizard.

Configuring Host Monitoring

1. Click the **Hosts** tab.
2. Select a host.
3. Click the **Configuration** tab.
4. Select **Scope** > **All**.
5. Click the **Monitoring** category.
6. Configure the property.
7. Click **Save Changes** to commit the changes.
8. Return to the Home page by clicking the Cloudera Manager logo.
9. Click the icon next to any stale services to invoke the cluster restart wizard.

Configuring Directory Monitoring

Cloudera Manager can perform threshold-based monitoring of free space in the various directories on the hosts it monitors—such as log directories or checkpoint directories (for the Secondary NameNode).

These thresholds can be set in one of two ways—as absolute thresholds (in terms of MiB and GiB, and so on) or as percentages of space. As with other threshold properties, you can set values that trigger events at both the Warning and Critical levels.

If you set both thresholds, the Absolute Threshold setting is used.

Configuring Activity Monitoring

The Activity Monitor monitors the MapReduce MRv1 jobs running on your cluster. This also includes the higher-level activities, such as Pig, Hive, and Oozie workflows that run as MapReduce tasks.

You can monitor for slow-running jobs or jobs that fail, and alert on these events. To detect jobs that are running too slowly, you must configure a set of activity duration rules that specify what jobs to monitor, and what the limits on duration are for those jobs. A "slow activity" event occurs when a job exceeds the duration limit configured for it in an activity duration rule. Activity duration rules are not defined by default; you must configure these rules if you want to see events for jobs that exceed the duration defined by these rules.

To configure Activity Monitor settings:

1. Go to the MapReduce service.
2. Click the **Configuration** tab.
3. Select **Scope** > ***MapReduce service_name*** **(Service-Wide)**.
4. Click the **Monitoring** category.
5. Specify one or more activity duration rules.
6. Click **Save Changes** to commit the changes.
7. Return to the Home page by clicking the Cloudera Manager logo.
8. Click the icon next to any stale services to invoke the cluster restart wizard.

Activity Duration Rules

An ***activity duration rule*** is a regular expression (used to match an activity name (that is, a Job ID)) combined with a run time limit which the job should not exceed. You can add as many rules as you like, one per line, in the **Activity Duration Rules** property.

The format of each rule is *regex*=*number* where the *regex* is a regular expression to match against the activity name, and *number* is the job duration limit, in minutes. When a new activity starts, each *regex* expression is tested against the name of the activity for a match.

The list of rules is tested in order, and the first match found is used. For example, if the rule set is:

```
foo=10
bar=20
```

any activity named "foo" would be marked slow if it ran for more than 10 minutes. Any activity named "bar" would be marked slow if it ran for more than 20 minutes.

Since Java regular expressions can be used, if the rule set is:

```
foo.*=10
bar=20
```

any activity with a name that starts with "foo" (for example, fool, food, foot) matches the first rule.

If there is no match for an activity, then that activity is not monitored for job duration. However, you can add a "catch-all" as the last rule that always matches any name:

```
foo.*=10
bar=20
baz=30
.*=60
```

In this case, any job that runs longer than 60 minutes is marked slow and generates an event.

Configuring YARN Application Monitoring

You can configure the visibility of the YARN application monitoring results.

Configuring Application Visibility

To configure whether admin and non-admin users can view all applications, only that user's applications, or no applications:

1. Go to the YARN service.
2. Click the **Configuration** tab.
3. Select **Scope** > ***YARN service_name*** **(Service-Wide)**.
4. Click the **Monitoring** category.
5. Set the **Applications List Visibility Settings** properties for admin and non-admin users.
6. Click **Save Changes** to commit the changes.
7. Return to the Home page by clicking the Cloudera Manager logo.
8. Click the icon next to any stale services to invoke the cluster restart wizard.

Configuring Impala Query Monitoring

You can configure the visibility of the Impala query results and the size of the storage allocated to Impala query results.

Configuring Query Visibility

To configure whether admin and non-admin users can view all queries, only that user's queries, or no queries:

1. Go to the Impala service.
2. Click the **Configuration** tab.
3. Select **Scope** > ***Impala service_name*** **(Service-Wide)**.
4. Click the **Monitoring** category.
5. Set the **Visibility Settings** properties for admin and non-admin users.
6. Click **Save Changes** to commit the changes.
7. Return to the Home page by clicking the Cloudera Manager logo.
8. Click the icon next to any stale services to invoke the cluster restart wizard.

Configuring Impala Query Data Store Maximum Size

The query store stores enough information to make the query searchable through the filter language.

1. Do one of the following:

- Select **Clusters** > **Cloudera Management Service**.
- On the **Home** > **Status** tab, in **Cloudera Management Service** table, click the **Cloudera Management Service** link.

2. Click the **Configuration** tab.
3. Select **Scope** > **Service Monitor**.
4. Click the **Main** category.
5. In the **Impala Storage** section, set the **firehose_impala_storage_bytes** property. The default is 1 GiB.
6. Click **Save Changes** to commit the changes.
7. Return to the Home page by clicking the Cloudera Manager logo.
8. Click the icon next to any stale services to invoke the cluster restart wizard.

The `firehose_impala_storage_bytes` property determines the approximate amount of disk space dedicated to storing Impala query data. Once the store reaches its maximum size, older data is deleted to make room for newer queries. The disk usage is approximate because data deletion begins only when the limit has been reached.

Configuring Alerts

Enabling Activity Monitor Alerts

You can enable alerts when an activity runs too slowly or fails.

1. Go to the MapReduce service.
2. Click the **Configuration** tab.
3. Select **Scope** > ***MapReduce service_name*** **(Service-Wide)**.
4. Click the **Monitoring** category.
5. Check the **Alert on Slow Activities** or **Alert on Activity Failure** checkboxes.
6. Click **Save Changes** to commit the changes.
7. Return to the Home page by clicking the Cloudera Manager logo.
8. Click the icon next to any stale services to invoke the cluster restart wizard.

Enabling Configuration Change Alerts

Configuration change alerts can be set service wide, or on specific roles for the service.

1. Click a service, role, or host.
2. Click the **Configuration** tab.
3. Select **Scope** > **All**.
4. Click the **Monitoring** category.
5. Check the **Enable Configuration Change Alerts** checkbox.
6. Click **Save Changes** to commit the changes.
7. Return to the Home page by clicking the Cloudera Manager logo.
8. Click the icon next to any stale services to invoke the cluster restart wizard.

Enabling HBase Alerts

1. Go to the HBase service.
2. Click the **Configuration** tab.
3. Select **Scope** > ***HBase service_name*** **(Service-Wide)**.
4. Click the **Monitoring** category.
5. Set one of the region or Hbck alerts:

   - Hbck Region Error Count
   - Hbck Error Count
   - Hbck Alert Error Codes
   - Hbck Slow Run
   - Region Health Canary Slow Run
   - Canary Unhealthy Region Count

- Canary Unhealthy Region Percentage

6. Click **Save Changes** to commit the changes.
7. Return to the Home page by clicking the Cloudera Manager logo.
8. Click the icon next to any stale services to invoke the cluster restart wizard.

Configuring Health Alerts

Enabling Health Alerts

You can enable alerts when the health of a role or service crosses a threshold.

1. Select **Clusters** > ***cluster_name*** > ***service_name*** or open the page for a role.
2. Click the **Configuration** tab.
3. Select **Scope** > ***role_name*** **or** ***service_name*** **(Service-Wide)**.
4. Click the **Monitoring** category.
5. Check the **Enable Health Alerts for this Role** or **Enable Service Level Health Alerts** checkbox, depending on whether you are configuring a role or a service.
6. Click **Save Changes** to commit the changes.
7. Return to the Home page by clicking the Cloudera Manager logo.
8. Click the icon next to any stale services to invoke the cluster restart wizard.

Modifying the Health Threshold

You can configure the threshold when a health alert is raised.

1. Select **Administration** > **Alerts**.
2. Click to the right of **Health Alert Threshold**.
3. Select **Scope** > **Event Server**.
4. Click the **Main** category.
5. Select the **Bad** or **Concerning** option.
6. Click **Save Changes** to commit the changes.
7. Return to the Home page by clicking the Cloudera Manager logo.
8. Click the icon next to any stale services to invoke the cluster restart wizard.

Configuring Alerts Transitioning Out of Alerting Health Threshold

You can configure an alert when a service or role instance transitions from an alerting to a non-alerting health threshold.

1. Select **Administration** > **Alerts**.
2. Click to the right of **Alert on Transitions out of Alerting Health**.
3. Select **Scope** > ***role_name*** **or** ***service_name*** **(Service-Wide)**.
4. In the category **Event Server Default Group**, check the **Alert on Transitions out of Alerting Health** checkbox.
5. Click **Save Changes** to commit the changes.
6. Return to the Home page by clicking the Cloudera Manager logo.
7. Click the icon next to any stale services to invoke the cluster restart wizard.

Configuring Log Alerts

You can configure an alert when a daemon emits a log message that matches a specified regular expression. See Configuring Log Alerts on page 22.

Configuring Alert Delivery

You can configure alerts to be delivered by email or sent as SNMP traps. If you choose email delivery, you can add to or modify the list of alert recipient email addresses. You can also send a test alert email. See Managing Alerts.

Note: If alerting is enabled for events, you can search for and view alerts in the Events tab, even if you do not have email notification configured.

Configuring Log Events

You can enable or disable the forwarding of selected log events to the Event Server. This is enabled by default, and is a service-wide setting (**Enable Log Event Capture**) for each service for which monitoring is provided. You can enable and disable event capture for CDH services or for the Cloudera Management Service.

Important: We do not recommend logging to a network-mounted file system. If a role is writing its logs across the network, a network failure or the failure of a remote file system can cause that role to freeze up until the network recovers.

Configuring Logs

1. Go to a service.
2. Click the **Configuration** tab.
3. Select ***role_name*** **(Service-Wide)** > **Logs** .
4. Edit a log property.
5. Click **Save Changes** to commit the changes.
6. Return to the Home page by clicking the Cloudera Manager logo.
7. Click the icon next to any stale services to invoke the cluster restart wizard.

Configuring Logging Thresholds

A logging threshold determines what level of log message is reported. The available levels are:

- `TRACE` - Informational events finer-grained than DEBUG.
- `DEBUG` - Informational events useful to debug an application.
- `INFO` - Informational events that highlight progress at coarse-grained level.
- `WARN` - Events that indicate a potential problem which is handled by the application.
- `ERROR` - Error events that allows the application to continue running.
- `FATAL` - Very severe error events that typically lead the application to abort.

The number of messages is greater and severity is least for `TRACE`. The default setting is `INFO`.

1. Go to a service.
2. Click the **Configuration** tab.
3. Enter `Logging Threshold` in the **Search** text field.
4. For the desired role group, select a logging threshold level.
5. Click **Save Changes** to commit the changes.
6. Return to the Home page by clicking the Cloudera Manager logo.
7. Click the icon next to any stale services to invoke the cluster restart wizard.

Configuring Log Directories

1. Do one of the following:
   - Cluster:
     1. On the **Home** > **Status** tab, click a cluster name.
     2. Select **Configuration** > **Log Directories**.
     3. Edit a ***role_name*** **Log Directory** property.
   - Service:

1. Go to a service.
2. Click the **Configuration** tab.
3. Select ***role_name*** **(Service-Wide)** > **Logs**.
4. Edit the **Log Directory** property.

2. Click **Save Changes** to commit the changes.
3. Return to the Home page by clicking the Cloudera Manager logo.
4. Click the icon next to any stale services to invoke the cluster restart wizard.

Enabling and Disabling Log Event Capture

1. Select **Clusters** > ***cluster_name*** > ***service_name***.
2. Click the **Configuration** tab.
3. Select **Scope** > ***service_name*** **(Service-Wide)**.
4. Click the **Monitoring** category.
5. Modify the **Enable Log Event Capture** setting.
6. Click **Save Changes** to commit the changes.
7. Return to the Home page by clicking the Cloudera Manager logo.
8. Click the icon next to any stale services to invoke the cluster restart wizard.

You can also modify the rules that determine how log messages are turned into events. Editing these rules is not recommended.

For each role, there are rules that govern how its log messages are turned into events by the custom log4j appender for the role. These are defined in the **Rules to Extract Events from Log Files** property.

Configuring Which Log Messages Become Events

1. Select **Clusters** > ***cluster_name*** > ***service_name***.
2. Click the **Configuration** tab.
3. Enter `Rules to Extract Events from Log Files` in the **Search** text field.
4. Click the **Monitoring** category.
5. Select the role group for the role for which you want to configure log events, or search for "Rules to Extract Events from Log Files". Note that for some roles there may be more than one role group, and you may need to modify all of them. The easiest way to ensure that you have found all occurrences of the property you need to modify is to search for the property by name. Cloudera Manager shows all copies of the property that matches the search filter.
6. In the Content field, edit the rules as needed. Rules can be written as regular expressions.
7. Click **Save Changes** to commit the changes.
8. Return to the Home page by clicking the Cloudera Manager logo.
9. Click the icon next to any stale services to invoke the cluster restart wizard.

Cloudera defines a number of rules by default. For example:

- The line `{"rate": 10, "threshold":"FATAL"},` means log entries with severity FATAL should be forwarded as events, up to 10 a minute.
- The line `{"rate": 0, "exceptiontype": "java.io.EOFException"},` means log entries with the exception `java.io.EOFException` should always be forwarded as an event.

The syntax for these rules is defined in the **Description** field for this property: the syntax lets you create rules that identify log messages based on log4j severity, message content matching, or the exception type. These rules must result in valid JSON.

Note: Editing these rules is not recommended. Cloudera Manager provides a default set of rules that should be sufficient for most users.

Configuring Log Alerts

You specify that a log event should generate an alert (by setting `"alert":true` in the rule). If you specify a content match, the entire content must match — if you want to match on a partial string, you must provide wildcards as appropriate to allow matching the entire string.

Monitoring Clusters

There are several ways to monitor clusters.

The Clusters tab in the top navigation bar displays each cluster's services in its own section, with the Cloudera Management Service separately below. You can select the following cluster-specific pages: hosts, reports, activities, and resource management.

The **Home** > **Status** tab displays the clusters being managed by Cloudera Manager. Each cluster is displayed either in summary form or in full form depending on the configuration of the **Administration** > **Settings** > **Other** > **Maximum Cluster Count Shown In Full** property. When the number of clusters exceeds the value of the property, only cluster summary information displays.

To display a cluster Status page, click the cluster name on the **Home** > **Status** tab Status tab. The cluster Status page displays a table containing links to the Hosts page and the status pages of the services running in the cluster.

Each service row in the table has a menu of actions that you select by clicking

and can contain one or more of the following indicators:

Indicator	Meaning	Description
	Health issue	Indicates that the service has at least one health issue. The indicator shows the number of health issues at the highest severity level. If there are Bad health test results, the indicator is red. If there are no Bad health test results, but Concerning test results exist, then the indicator is yellow. No indicator is shown if there are no Bad or Concerning health test results. **Important:** If there is one Bad health test result and two Concerning health results, there will be three health issues, but the number will be one. Click the indicator to display the **Health Issues** pop-up dialog box. By default only Bad health test results are shown in the dialog box. To display Concerning health test results, click the **Also show *n* concerning issue(s)** link.Click the link to display the Status page containing with details about the health test result.
	Configuration issue	Indicates that the service has at least one configuration issue. The indicator shows the number of configuration issues at the highest severity level. If there are configuration errors, the indicator is red. If there are no errors but configuration warnings exist, then the indicator is yellow. No indicator is shown if there are no configuration notifications. **Important:** If there is one configuration error and two configuration warnings, there will be three configuration issues, but the number will be one. Click the indicator to display the **Configuration Issues** pop-up dialog box.

Indicator	Meaning	Description
		By default only notifications at the Error severity level are listed, grouped by service name are shown in the dialog box. To display Warning notifications, click the **Also show *n* warning(s)** link.Click the message associated with an error or warning to be taken to the configuration property for which the notification has been issued where you can address the issue.See Managing Services.
Restart Needed Refresh Needed	Configuration modified	Indicates that at least one of a service's roles is running with a configuration that does not match the current configuration settings in Cloudera Manager. Click the indicator to display the Stale Configurations page.To bring the cluster up-to-date, click the Refresh or Restart button on the Stale Configurations page or follow the instructions in Refreshing a Cluster, Restarting a Cluster, or Restarting Services and Instances after Configuration Changes.
	Client configuration redeployment required	Indicates that the client configuration for a service should be redeployed. Click the indicator to display the Stale Configurations page.To bring the cluster up-to-date, click the Deploy Client Configuration button on the Stale Configurations page or follow the instructions in Manually Redeploying Client Configuration Files.

The right side of the status page displays charts (dashboard) that summarize resource utilization (IO, CPU usage) and processing metrics. .

Note: If you delete a cluster, the deleted cluster still displays in some charts. This is because the charts also show historical data. Over time, data from the deleted cluster will drop off as older data is replaced by more current data. You can work around this by:

- Waiting for the data from the deleted cluster to drop off.
- Editing the `where` clause of query for the chart to include only the cluster(s) you are interested in. (For example: `clusterDisplayName=Cluster_1`). You can revert to the original query at a later date, after the data for the deleted cluster has dropped off. See Charting Time-Series Data on page 99.
- Deleting all data in the Host Monitor and Service Monitor storage directories and starting from scratch. You will, however, loose all historical data from both current and deleted clusters. See Configuring Host Monitor Data Storage and Configuring Service Monitor Data Storage to learn where the storage directories are located.

Monitoring Multiple CDH Deployments Using the Multi Cloudera Manager Dashboard

Note:

This item is deprecated and will be removed in a future release. Cloudera supports items that are deprecated until they are removed. For more information about deprecated and removed items, see Deprecated Items.

The **Multi Cloudera Manager Dashboard** is a special mode of Cloudera Manager that enables you to view monitoring data aggregated from multiple Cloudera Manager instances that manage multiple CDH clusters. In a single-page view, you can use the **Multi Cloudera Manager Dashboard** to:

- Display health information.
- Display metrics showing capacity and utilization levels.
- Go to the Cloudera Manager Admin console for Cloudera Manager instances on the dashboard.

Using the **Multi Cloudera Manager Dashboard**, you define the Cloudera Manager instances that display on the dashboard. You can also define a **Profile** that defines the information displayed on the dashboard. The **Multi Cloudera Manager Dashboard** functionality is included with every installation of Cloudera Manager, and you enable it by setting a property. Cloudera Manager instances monitored by the Multi Cloudera Manager Dashboard must be version 5.5.0 or higher.

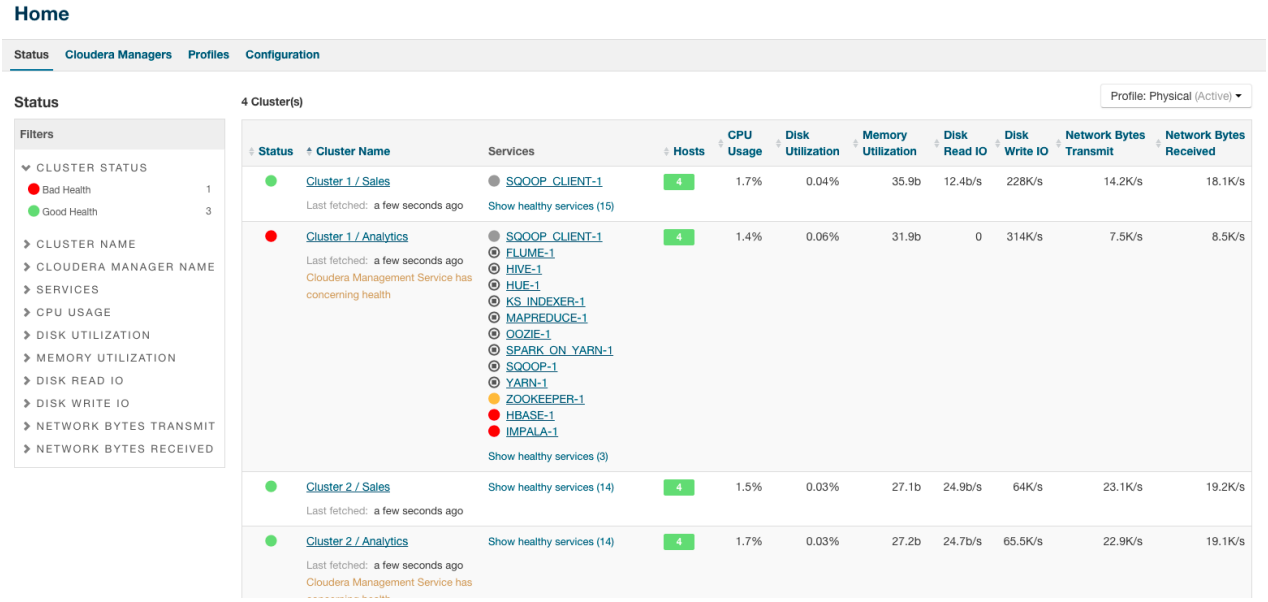


Figure 1: Multi Cloudera Manager Status Dashboard

Multi Cloudera Manager Dashboard Modes

The **Multi Cloudera Manager Dashboard** has two modes of operation:

- **Hosted mode** - You can use an instance of Cloudera Manager that manages CDH clusters to also display the **Multi Cloudera Manager Dashboard**. As shown in Figure 2: Multi Cloudera Manager Dashboard Configured in Hosted Mode on page 25, the **Multi Cloudera Manager Dashboard**, hosted on the Cloudera Manager instance 1, fetches status updates from the clusters managed by Cloudera Manager instances 1, 2, and 3 and displays them on the **Multi Cloudera Manager Dashboard**.

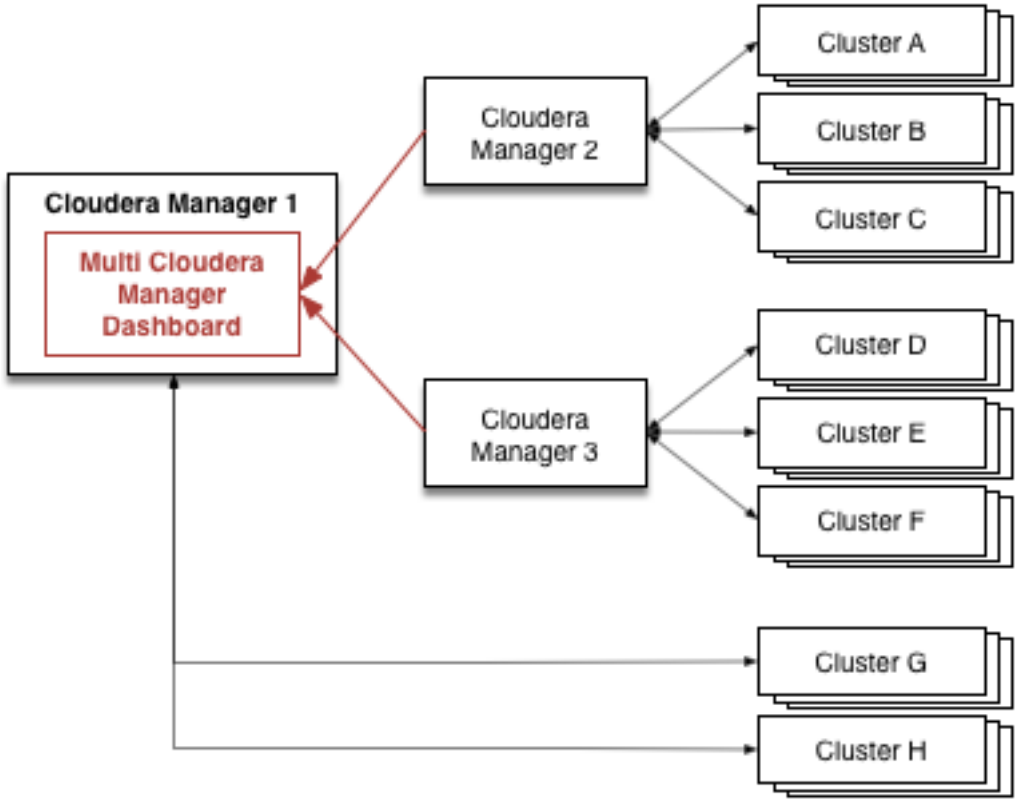


Figure 2: Multi Cloudera Manager Dashboard Configured in *Hosted* Mode

- **Dedicated Mode** - You can configure a standalone instance of Cloudera Manager for which the only function is to display and manage the **Multi Cloudera Manager Dashboard**. As shown in Figure 3: Multi Cloudera Manager Dashboard Configured in Dedicated Mode on page 26, Cloudera Manager 1 fetches status updates from the clusters managed by Cloudera Manager instances 2, 3, and 4 and displays them on the **Multi Cloudera Manager Dashboard**.

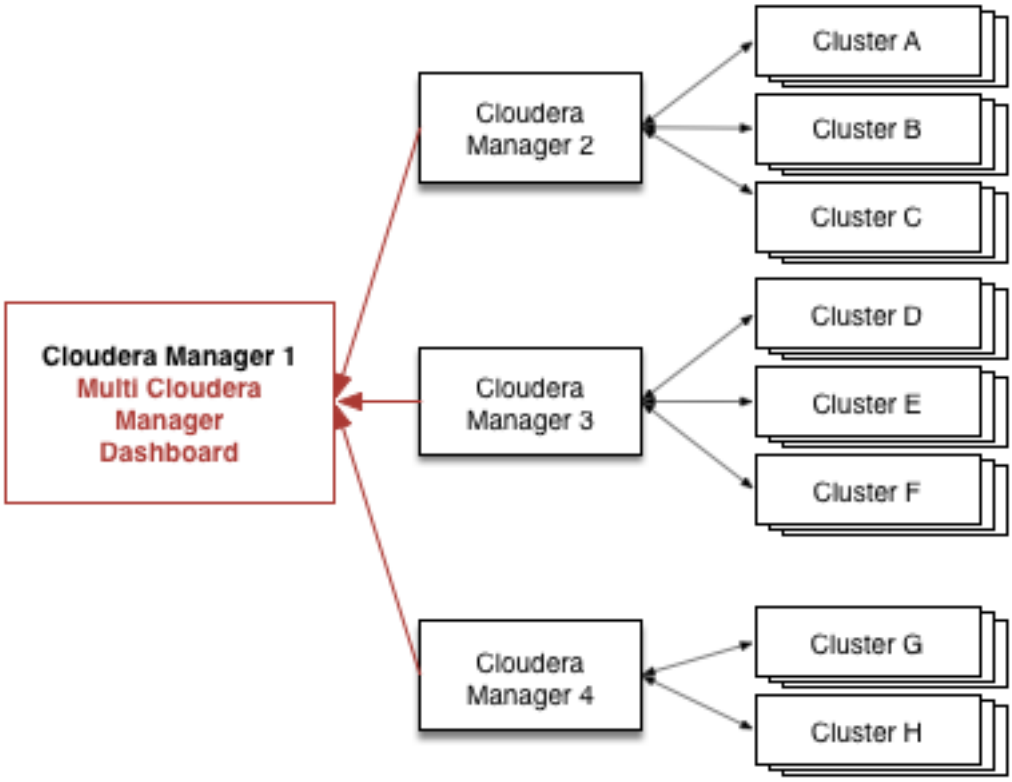


Figure 3: Multi Cloudera Manager Dashboard Configured in *Dedicated* Mode

Installing and Managing the Multi Cloudera Manager Dashboard

To use the **Multi Cloudera Manager Dashboard** to monitor multiple CDH clusters, follow the steps in the following sections.

Installing and Enabling the Multi Cloudera Manager Status Dashboard

Minimum Required Role: **Cluster Administrator** (also provided by **Full Administrator**)

> **Note:** Your installation must have a Cloudera Enterprise license installed to use this feature. See Managing Licenses.

You can install the **Multi Cloudera Manager Dashboard** in either *Hosted* or *Dedicated* mode. (See Multi Cloudera Manager Dashboard Modes on page 24.

Hosted Mode Installation

Hosted mode uses an existing installation of Cloudera Manager that manages one or more clusters to also display the **Multi Cloudera Manager Dashboard**.

1. Log in to the Cloudera Manager instance as a user with Administrator privileges.
2. Select **Administration** > **Settings**.
3. Under **Category**, select **Multi Cloudera Manager Dashboard**.
4. Select the **Enable Multi Cloudera Manager Status Dashboard** property.
5. Click **Save Changes** to commit the changes.

You can now access the **Multi Cloudera Manager Status Dashboard** at the following URL:

```
http:/Cloudera_Manager_Host:Cloudera_Manager_port/cmf/aggregator/status
```

Dedicated Mode Installation

To install **Multi Cloudera Manager Dashboard** in Dedicated mode, you install Cloudera Manager on a single sever using the Installation Wizard, but you do not install CDH and other services.

1. Install Cloudera Manager using one of the procedures described in Installing Cloudera Manager and CDH.
2. Use a web browser to open the Installation Wizard by logging in to the Cloudera Manager instance at `http://<Cloudera Manager Server Host>:7180`. Use the default credentials: **Username:** `admin` **Password:** `admin`.
3. Select a Cloudera edition.

   > **Note:** Your installation must have a Cloudera Enterprise license installed to use this feature. See Managing Licenses.

4. Click **Continue**.

   The **Thank You for choosing Cloudera Manager and CDH** notice appears.

5. Click the **ClouderaManager** logo to open the **Home** page.
6. Select **Administration** > **Settings**.
7. Under **Category**, select **Multi Cloudera Manager Dashboard**.
8. Select the **Enable Multi Cloudera Manager Status Dashboard** property.
9. Click **Save Changes** to commit the changes.
10. Log in as `root` to the host where the Cloudera Manager Server is running.
11. Open `/etc/default/cloudera-scm-server` in a text editor and append the following line to the end of the file:

```
export CMF_JAVA_OPTS="$CMF_JAVA_OPTS -DCM_AGGREGATOR_UI=true"
```

12. Restart Cloudera Manager on the host where the Cloudera Manager Server is running, using the following command:

```
service cloudera-scm-server restart
```

> **Note:** If you restart Cloudera Manager without having installed a Cloudera Enterprise license, the restart will fail. To recover, disable Dedicated mode, install the license, reenable Dedicated mode, and restart Cloudera Manager Server.

You can now access the **Multi Cloudera Manager Status Dashboard** at the following URL:

```
http:/Cloudera_Manager_Host:Cloudera_Manager_port
```

Adding, Editing, or Deleting a Cloudera Manager Instance to Display on the Dashboard

Minimum Required Role: Cluster Administrator (also provided by **Full Administrator**)

To select the Cloudera Manager instances that manage cluster(s) you want to monitor using the **Multi Cloudera Manager Dashboard**, specify a name, URL, and credentials for each instance, using the steps in this section.Cloudera Manager instances monitored by the Multi Cloudera Manager Dashboard must be version 5.5.0 or higher.

Adding a Cloudera Manager to the Dashboard

1. Open the **Multi Cloudera Manager Dashboard**.
2. Select the **Add Cloudera Managers** tab.
3. Click **Add Cloudera Manager**.

   The **Add Cloudera Manager** screen displays.

4. Enter a descriptive name for the Cloudera Manager in the **Cloudera Manager Name** field.
5. Enter the URL of the Cloudera Manager in the **Cloudera Manager URL** field.
6. Select one of the following options for **Credentials**:
   - **Create a New Account** - Creates a new account on the Cloudera Manager instance. If you select this option, enter administrative credentials for the Cloudera Manager instance. The Cloudera Manager instance uses these credentials to create a new, internal account, and the **Multi Cloudera Manager Dashboard** uses this new account to access monitoring data from Cloudera Manager. The username for this account is not available and is not required. You can use this option to keep access to from the **Multi Cloudera Manager Dashboard** to the Cloudera Manager instance isolated from changes in the administrative user's credentials.
   - **Use an Existing Account** - Uses the credentials you provide to access monitoring data from the Cloudera Manager instance. If you select this option, enter read-only credentials for the Cloudera Manager instance.
7. Click **Add Cloudera Manager**.

   The **Cloudera Managers** tab displays a line containing the Cloudera Manager instance you just added.
8. Click **Test Connectivity** to make sure you can connect to the Cloudera Manager instance using the URL and credentials you provided. If the test fails, click **Edit** to modify the connection parameters. If the connection is correct, you can click the URL to open the Cloudera Manager Admin Console.
9. Click **Add Cloudera Manager** again to add additional Cloudera Manager instances.

Deleting a Cloudera Manager Instance from the Dashboard

1. Open the **Multi Cloudera Manager Dashboard**.
2. Select the **Cloudera Managers** tab.
3. Click **Delete** in the row containing the Cloudera Manager instance you want to delete.

Editing a Cloudera Manager on the Dashboard

1. Open the **Multi Cloudera Manager Dashboard**.
2. Select the **Cloudera Managers** tab.
3. Click **Edit** in the row containing the Cloudera Manager instance you want to edit.
4. Make the desired changes.
5. Click **Update Cloudera Manager**.

Managing Multi Cloudera Manager Dashboard Users

Minimum Required Role: User Administrator (also provided by **Full Administrator**)

Adding an Internal User Account

1. Select **Administration** > **Users**.
2. Click the **Add User** button.
3. Enter a username and password.
4. In the Role drop-down menu, select a role for the new user.
5. Click **Add**.

Assigning User Roles

1. Select **Administration** > **Users**.
2. Check the checkbox next to one or more usernames.
3. Select **Actions for Selected** > **Assign User Roles**.
4. In the drop-down menu, select the role.
5. Click the **Assign Role** button.

User Roles for the Multi Cloudera Manager Status Dashboard

The valid User Roles for the **Multi Cloudera Manager Dashboard** are:

- **Read-Only**
  - View configuration and monitoring information in Cloudera Manager.
  - View service and monitoring information.
  - View events and logs.
  - View replication jobs and snapshot policies.
  - View YARN applications and Impala queries.

  The Read-Only role does not allow the user to:

  - Add services or take any actions that affect the state of the cluster.
  - Use the HDFS file browser.
  - Use the HBase table browser.
  - Use the Solr Collection Statistics browser.
- **User Administrator**
  - View configuration and monitoring information in Cloudera Manager.
  - View service and monitoring information.
  - Manage user accounts and configuration of external authentication.
  - Use the HDFS file browser, the HBase table browser, and the Solr Collection browser.
  - Perform the same tasks as the Read-Only role.
- **Full Administrator** - Full Administrators have permissions to use all of the functionality available in Cloudera Manager and perform all actions on all clusters. Additionally, the Full Administrator can view the data related to Cloudera Manager, such as file metadata, snapshots, quotas, and file size. The Full Administrator cannot see things like the content of files stored by HDFS or other components.

Viewing User Sessions

1. Select **Administration** > **Users**.
2. Click the tab **User Sessions**.

Managing Multi Cloudera Manager Dashboard Profiles

On the **Status** tab, each row displays aggregated monitoring information for each Cloudera Manager instance you have added and for each cluster managed by those instances. You can define custom *Profiles* to define the information that displays on each row on the **Status** tab. A separate set of custom profiles are maintained for each Cloudera Manager user. Custom profiles are not visible to other users.

The following profiles are available to all users by default:

- Physical
- HDFS
- HBASE
- IMPALA
- YARN

Adding a Multi Cloudera Manager Status Dashboard Profile

To add a profile:

1. Select the **Profiles** tab.
2. Click **Add Profile**. (You can also duplicate an existing profile by clicking the  icon next to the profile name and selecting **Duplicate**.)

   The **Add Profile** dialog box displays.
3. Enter a name for the profile.

   The new profile is added to the list of profiles, and the configuration for the new profile displays.

4. Select the columns and metrics you want to display for this profile by selecting any of the following:
   - **Show Column** - The column displays in the **Status** table.
   - **Show Filter** - The column is included in the **Filters** section.
   - **Expand Filter** - For columns included in the **Filters** section, you can select whether to automatically show the filter expanded when you go to the **Status** tab.

   For more information about these columns, see Multi Cloudera Manager Status Dashboard Columns on page 31.

   For more information about metrics, see Multi Cloudera Manager Status Dashboard Metrics on page 33. You can also add your own queries to create metrics that display information you want.

5. To use the new profile on the dashboard, click **Activate** next to the Profile name.

   You can also select the active profile on the **Status** tab. Select a profile from the drop-down **Profile** button located on the right of the **Status** tab, above the list of Cloudera Manager instances.

Editing a Multi Cloudera Manager Status Dashboard Profile

To edit a profile:

1. Select the **Profiles** tab.
2. Click the name of the profile you want to edit.
3. Change the profile as needed.

To see your changes, activate the profile by clicking the drop-down arrow next to its name and selecting **Activate**.

Deleting a Multi Cloudera Manager Status Dashboard Profile

To delete a profile:

1. Select the **Profiles** tab.
2. Click the  icon next to the Profile name.
3. Select **Delete**.

Configuring Metrics to Display on the Multi Cloudera Manager Status Dashboard

Minimum Required Role: Cluster Administrator (also provided by **Full Administrator**)

The **Multi Cloudera Manager Dashboard** has a number of default queries you can use to display metrics. See Multi Cloudera Manager Status Dashboard Metrics on page 33.

You can also create custom queries by using the tsquery Language to write queries on data from the Cloudera Manager time-series datastore. Queries must return a single stream for display in the **Multi Cloudera Manager Dashboard**.

Adding a Custom Metric

To add a metric that displays on the **Status** tab:

1. Select the **Profiles** tab.
2. Click **Configure Metrics**.
3. In the **Aggregate Metrics** area, click a ✚ icon to add a new metric beneath the current metric.

   A dialog box to define a new metric displays.

4. Enter a name for the metric in the **Name** field.
5. Enter a tsquery statement in the **Query** field.
6. Click **Save Changes** to commit the changes.
7. Click the **Profiles** tab.
8. Select a profile.
9. Locate the new metric in the list of metrics.
10. Select **Show Column**, **Show Filter**, or **Expand Filter** as desired.

Editing a Metric

1. Select the **Profiles** tab.
2. Click **Configure Metrics**.
3. Expand the query you want to edit by clicking the ❯ icon.
4. Change the query statement or the name of the metric.
5. Click **Save Changes** to commit the changes.

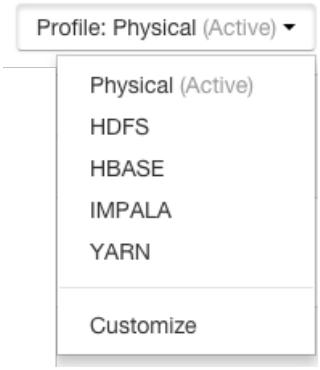
Deleting a Metric

1. Select the **Profiles** tab.
2. Click **Configure Metrics**.
3. Click the ━ icon in the row containing the query you want to delete.
4. Click **Save Changes** to commit the changes.

Using the Multi Cloudera Manager Status Dashboard

Use the Multi Cloudera Manager Status Dashboard **Status** page to monitor the CDH clusters managed by the Cloudera Manager instancess you have added to the **Multi Cloudera Manager Dashboard**. The **Status** tab displays a table with rows that contain information about each cluster. Within each row, you can click links to display additional information and go to the relevant Cloudera Manager, cluster, or service.

You configure which columns display by creating Profiles. The name of the active profile displays in a button on the right of the screen above the table of Cloudera Managers. You can choose which profile is active by clicking ▾ next to the profile name and selecting a new profile.



In a profile, you can also select **Metrics**, which are queries that display metrics about the functionality of each cluster. A set of metrics are provided, and you can create custom metrics using the tsquery Language on page 111.

Using the **Filters** section, you can limit the rows displayed by selecting various values of the displayed columns. You can define which columns are used as filters in the profile.

Multi Cloudera Manager Status Dashboard Columns

The Multi Cloudera Manager Status Dashboard displays a table with a row for each cluster managed by a Cloudera Manager. Click the arrows next to a column name to sort the table by the values in the column. Some columns contain links that you can click to display additional details or to go to other displays. You can select which columns to display by editing a Profile. The following table describes the columns.

Table 1: Multi Cloudera Manager Status Dashboard Columns

Column	Description
Status	Displays an icon representing the health of the cluster. See Health Status Icons on page 33.

Column	Description
Cluster Name	Displays the cluster name and the Cloudera Manager that manages it; for example: **Cluster 1 / Sales** The cluster name is a link that you can click to display a pop-up box containing additional information about the cluster. The pop-up includes the following links: • **Visit Cluster** - Opens the **Status** page in the Cloudera Manager Admin Console for the cluster. • **Visit Cloudera Manager** - Opens the **Home** page of the Cloudera Manager Admin console for the Cloudera Manager instance. • **Cloudera Management Service** - Opens the **Cloudera Management Service** > **Status** page for the Cloudera Manager instance. The Cluster Name column also displays the latest time that data was fetched from the Cloudera Manager and displays text indicating the health status of the Cloudera Management Service for the cluster. **Note:** When managing a Profile, you cannot hide this column.
Services	Displays icons indicating the health status of each service defined in the cluster. See Health Status Icons on page 33. Click the service name to open a pop-up box that displays details about the service's status. Click **Visit Service** to go to the service page in the Cloudera Manager instance that manages that service. Healthy services are hidden by default. Click the **Show healthy services(#)** link to show the list of services. Click **Hide healthy services** to hide them.
Hosts	Displays colored badges displaying the number of hosts in the cluster by their health status. For example: 6 Click the badge to go to the **Hosts** page of the Cloudera Manager instance that manages the cluster.
CM Version	Displays the version of the Cloudera Manager instance.
Cloudera Management Service Status	Displays icons indicating the health status of the Cloudera Management Service. See Health Status Icons on page 33.
Host Count	Number of hosts in the cluster.

Filtering the Status Display

Click the › icon next to an item in the **Filters** section to see the values you can select for filtering. By selecting one or more items, you can limit the rows displayed to those that match the values you select. You can select multiple values for multiple items.

Click the **Clear** link to remove the filter.

For example, Figure 4: Selecting Filter Values on page 33 shows the Services filter expanded to include only Flume and HDFS:

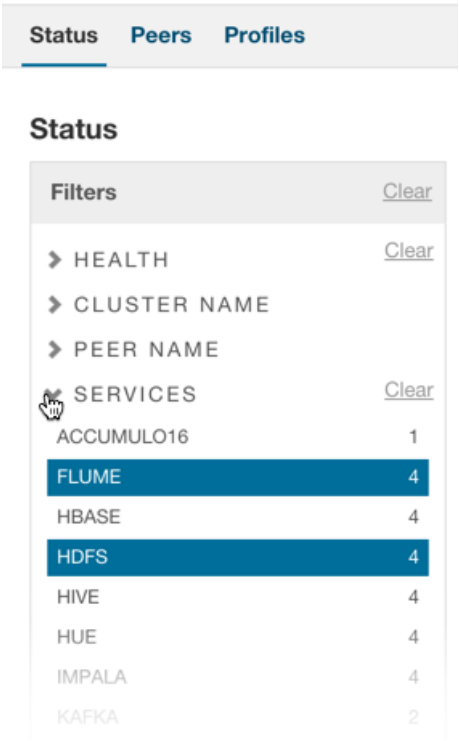


Figure 4: Selecting Filter Values

Multi Cloudera Manager Status Dashboard Metrics

The **Multi Cloudera Manager Dashboard** has a number of default queries you can use to display metrics. You select which metrics display by defining a Profile. See Installing and Managing the Multi Cloudera Manager Dashboard on page 26. You can also create custom queries by using the tsquery Language on page 111 to write queries on data from the Cloudera Manager time-series datastore. See Adding a Custom Metric on page 30.

Health Status Icons

The following icons indicate the health of hosts and services:

Table 2: Status

Indicator	Status	Description
	Started with outdated configuration	For a service, this indicates the service is running, but at least one of its roles is running with a configuration that does not match the current configuration settings in Cloudera Manager. For a role, this indicates a configuration change has been made that requires a restart, and that restart has not yet occurred. Click the indicator to display the Stale Configurations page.
	Starting	The entity is starting up but is not yet running.
	Stopping	The entity is stopping but has not stopped yet.
	Stopped	The entity is stopped, as expected.
	Down	The entity is not running, but it is expected to be running.

Indicator	Status	Description
	History not available	Cloudera Manager is in historical mode, and the entity does not have historical monitoring support. This is the case for services other than HDFS, MapReduce and HBase such as ZooKeeper, Oozie, and Hue.
	None	The entity does not have a status. For example, it is not something that can be running and it cannot have health. Examples are the HDFS Balancer (which runs from the HDFS Rebalance action) or Gateway roles. The Start and Stop commands are not applicable to these instances.
	Good health	The entity is running with good health. For a specific health test, the returned result is normal or within the acceptable range. For a role or service, this means all health tests for that role or service are **Good**.
	Concerning health	The entity is running with concerning health. For a specific health test, the returned result indicates a potential problem. Typically this means the test result has gone above (or below) a configured Warning threshold. For a role or service, this means that at least one health test is **Concerning**.
	Bad health	The entity is running with bad health. For a specific health test, the test failed, or the returned result indicates a serious problem. Typically this means the test result has gone above (or below) a configured Critical threshold. For a role or service, this means that at least one health test is **Bad**.
	Disabled health	The entity is running, but all of its health tests are disabled.
	Unknown health	The status of a service or role instance is unknown. This can occur for a number of reasons, such as the Service Monitor is not running, or connectivity to the Agent doing the health monitoring has been lost.

Monitoring Services

Cloudera Manager's Service Monitoring feature monitors dozens of service health and performance metrics about the services and role instances running on your cluster:

- Presents health and performance data in a variety of formats including interactive charts
- Monitors metrics against configurable thresholds
- Generates events related to system and service health and critical log entries and makes them available for searching and alerting
- Maintains a complete record of service-related actions and configuration changes

Monitoring Service Status

From a service page, you can:

- Monitor the status of the services running on your clusters.
- Manage the services and roles in your clusters.
- Add new services.
- Access the client configuration files generated by Cloudera Manager that enable Hadoop client users to work with the HDFS, MapReduce, HBase, and YARN services you added. (These configuration files are normally deployed automatically when you install a cluster or add a service).
- View the maintenance mode status of a cluster.

You can also pull down a menu from an individual service name to go directly to one of the tabs for that service to its Status, Instances, Commands, Configuration, Audits, or Charts Library tabs.

Viewing the URLs of the Client Configuration Files

To allow Hadoop client users to work with the services you created, Cloudera Manager generates client configuration files that contain the relevant configuration files with the settings from your services. These files are deployed automatically by Cloudera Manager based on the services you have installed, when you add a service, or when you add a Gateway role on a host.

You can manually download and distribute these client configuration files to the users of a service, if necessary.

The **Actions** > **Client Configuration URLs** command opens a pop-up that displays links to the client configuration zip files created for the services installed in your cluster. You can download these zip files by clicking the link.

The Actions button is not enabled if you are viewing status for a point of time in the past.

See Client Configuration Files for more information on this topic.

Viewing the Status of a Service Instance

Do one of the following:

- In **Home** > **Status** tab, select ***ClusterName*** > ***ServiceName***.
- Select **Clusters** > ***ClusterName*** > ***ServiceName***.

This opens the **Status** page where you can view a variety of information about a service and its performance. See Viewing Service Status on page 35 for details.

Viewing the Health and Status of a Role Instance

Click the role instance under the **Role Counts** column.

If there is just one instance of this role, this opens the **Status** tab for the role instance.

If there are multiple instances of a role, clicking the role link under **Role Counts** will open the **Instances** tab for the service, showing instances of the role type you have selected. See Viewing Role Instance Status on page 40 for details.

If you are viewing a point in time in the past, the Role Count links will be greyed out, but still functional. Their behavior will depend on whether historical data is available for the role instance.

Viewing the Maintenance Mode Status of a Cluster

Select **Actions** > **View Maintenance Mode Status...** button to view the status of your cluster in terms of which components (service, roles, or hosts) are in maintenance mode. This pops up a dialog box that shows the components in your cluster that are in maintenance mode, and indicates which are in effective maintenance mode as well as those that have been explicitly placed into maintenance mode. (See Maintenance Mode for an explanation of explicit maintenance mode and effective maintenance mode.)

From this dialog box you can select any of the components shown there and remove them from maintenance mode.

If individual services are in maintenance mode, you will see the maintenance mode icon next to the **Actions** button for that service.

The Actions button is not enabled if you are viewing status for a point of time in the past.

Viewing Service Status

To view service status, do one of the following:

- In the **Home** > **Status** tab, if the cluster is displayed in full form, click *ServiceName* in a *ClusterName* table.
- In the **Home** > **Status** tab, click *ClusterName* and then click *ServiceName*.
- Select **Clusters** > ***ClusterName*** > ***ServiceName***.

For all service types there is a **Status Summary** that shows, for each configured role, the overall status and health of the role instance(s).

Note: Not all service types provide complete monitoring and health information. Hive, Hue, Oozie, Solr, and YARN (CDH 4 only) only provide the basic Status Summary on page 36.

Each service that supports monitoring provides a set of monitoring properties where you can enable or disable health tests and events, and set thresholds for tests and modify thresholds for the status of certain health tests. For more information see Configuring Monitoring Settings on page 15.

The HDFS, MapReduce, HBase, ZooKeeper, and Flume services also provide additional information: a snapshot of service-specific metrics, health test results, health history, and a set of charts that provide a historical view of metrics of interest.

Viewing Past Status

The health and status information on the Status page represents the state of the service or role instance at a given *point in time*. The charts (and the Logs and Events under Diagnostics) represent the time range selected on the Time Range Selector (which defaults to the past 30 minutes). You can view health, status, and chart historical data by expanding the Time Range Selector (click the mini line chart under "admin") and moving the time marker (◆) to a point in the past.

When you move the time marker to a point in the past (for services and roles that support health history), the entire Status page updates to the time selected. A Now button (⏭) allows you to quickly return to the current state of the service. The Actions menu is disabled while viewing a past status to ensure that you cannot accidentally act on outdated status information.

See Time Line on page 7 for more details.

Status Summary

The Status Summary shows the status of each service instance being managed by Cloudera Manager. Even services such as Hue, Oozie, or YARN (which are not monitored by Cloudera Manager) show a status summary. The overall status for a service is a roll-up of the health test results for the service and all its role instances. The **Status** can be:

Table 3: Status

Indicator	Status	Description
⏻	Started with outdated configuration	For a service, this indicates the service is running, but at least one of its roles is running with a configuration that does not match the current configuration settings in Cloudera Manager. For a role, this indicates a configuration change has been made that requires a restart, and that restart has not yet occurred. Click the indicator to display the Stale Configurations page.
	Starting	The entity is starting up but is not yet running.
	Stopping	The entity is stopping but has not stopped yet.
	Stopped	The entity is stopped, as expected.
⛔	Down	The entity is not running, but it is expected to be running.
	History not available	Cloudera Manager is in historical mode, and the entity does not have historical monitoring support. This is the case for services other than HDFS, MapReduce and HBase such as ZooKeeper, Oozie, and Hue.
	None	The entity does not have a status. For example, it is not something that can be running and it cannot have health. Examples are the HDFS Balancer (which runs from the HDFS Rebalance action) or Gateway roles. The Start and Stop commands are not applicable to these instances.

Indicator	Status	Description
	Good health	The entity is running with good health. For a specific health test, the returned result is normal or within the acceptable range. For a role or service, this means all health tests for that role or service are **Good**.
	Concerning health	The entity is running with concerning health. For a specific health test, the returned result indicates a potential problem. Typically this means the test result has gone above (or below) a configured Warning threshold. For a role or service, this means that at least one health test is **Concerning**.
	Bad health	The entity is running with bad health. For a specific health test, the test failed, or the returned result indicates a serious problem. Typically this means the test result has gone above (or below) a configured Critical threshold. For a role or service, this means that at least one health test is **Bad**.
	Disabled health	The entity is running, but all of its health tests are disabled.
	Unknown health	The status of a service or role instance is unknown. This can occur for a number of reasons, such as the Service Monitor is not running, or connectivity to the Agent doing the health monitoring has been lost.

To see the status of one or more role instances, click the role type link under **Status Summary**. If there is a single instance of the role type, the link directs you to the Status page of the role instance.

If there are multiple role instances (such as for DataNodes, TaskTrackers, and RegionServers), the role type link directs you to the Role Instances page for that role type. Click on each instance, under Role Type, to be taken to the corresponding Status page.

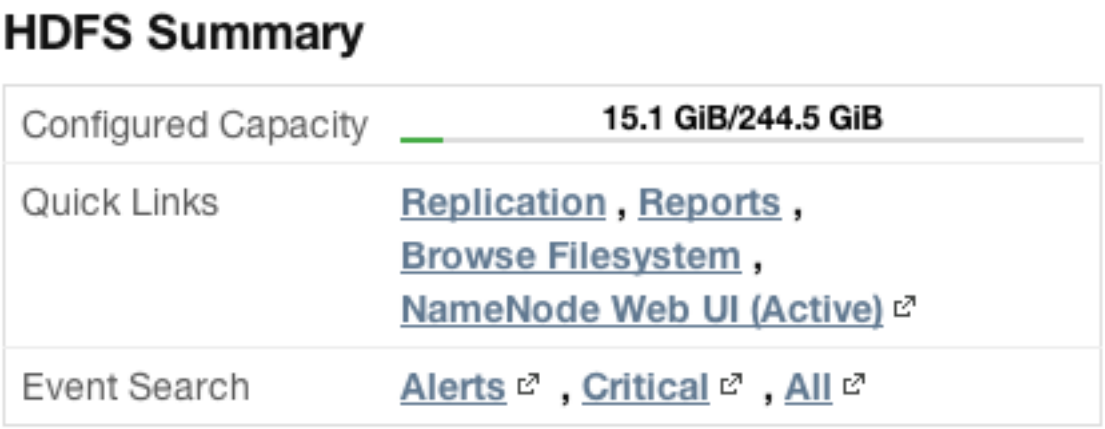
To display the results for each health test that applies to this role type, expand the **Health Tests** filter on the left and expand **Good Health**, **Warnings**, **Bad Health**, or **Disabled Health**. Health test results that have been filtered out by your role type selection appear as unavailable.

Service Summary

Some services (specifically HDFS, MapReduce, HBase, Flume, and ZooKeeper) provide additional statistics about their operation and performance. These are shown in a Summary panel at the left side of the page. The contents of this panel depend on the service:

- The HDFS Summary shows disk space usage.
- The MapReduce Summary shows statistics on slot usage, jobs and so on.
- The Flume Summary provides a link to a page of Flume metric details. See Flume Metric Details on page 38.
- The ZooKeeper Summary provides links to the ZooKeeper role instances (nodes) as well as Zxid information if you have a ZooKeeper Quorum (multiple ZooKeeper servers).

For example:



Other services such as Hue, Oozie, Impala, and Cloudera Manager itself, do not provide a Service Summary.

Health Tests and Health History

The Health Tests and Health History panels appear for HDFS, MapReduce, HBase, Flume, Impala, ZooKeeper, and the Cloudera Manager Service. Other services such as Hue, Oozie, and YARN do not provide a Health Test panel.

The Health Tests panel shows health test results in an expandable and collapsible list, typically with the specific metrics that the test returned. (You can Expand All or Collapse All from the links at the upper right of the Health Tests panel).

- The color of the text (and the background color of the field) for a Health Test result indicates the status of the results. The tests are sorted by their health status – Good, Concerning, Bad, or Disabled. The entries are collapsed by default. Click the arrow to the left of an entry to expand the entry and display further information.
- Clicking the **Details** link for a health test displays further information about the test, such as the meaning of the test and its possible results, suggestions for actions you can take or how to make configuration changes related to the test. The help text may include a link to the relevant monitoring configuration section for the service. See Configuring Monitoring Settings on page 15 for more information.
- In the Health Tests panel:
  - Clicking ❯ displays the lists of health tests that contributed to the health test.
  - Clicking the **Details** link displays further information about the health test.
- In the Health History panel:
  - Clicking ❯ displays the lists of health tests that contributed to the health history.
  - Clicking the **Show** link moves the time range to the historical time period.

Charts

HDFS, MapReduce, HBase, ZooKeeper, Flume, and Cloudera Management Service all display charts of some of the critical metrics related to their performance and health. Other services such as Hive, Hue, Oozie, and Solr do not provide charts.

See Viewing Charts for Cluster, Service, Role, and Host Instances on page 13 for detailed information on the charts that are presented, and the ability to search and display metrics of your choice.

Flume Metric Details

From the Flume Service Status page, click the **Flume Metric Details** link in the **Flume Summary** panel to display details of the Flume agent roles.

On this page you can view a variety of metrics about the Channels, Sources and Sinks you have configured for your various Flume agents. You can view both current and historical metrics on this page.

The **Channels** section shows the metrics for all the channel components in the Flume service. These include metrics related to the channel capacity and throughput.

The **Sinks** section shows metrics for all the sink components in the Flume service. These include event drain statistics as well as connection failure metrics.

The **Sources** section shows metrics for all the source components in the Flume service.

This page maintains the same navigation bar as the Flume service status page, so you can go directly to any of the other tabs (Instances, Commands, Configuration, or Audits).

Viewing Service Instance Details

1. Do one of the following:
   - In the **Home** > **Status** tab, if the cluster is displayed in full form, click *ServiceName* in a *ClusterName* table.
   - In the **Home** > **Status** tab, click *ClusterName* and then click *ServiceName*.
   - Select **Clusters** > ***ClusterName*** > ***ServiceName***.
2. Click the **Instances** tab on the service's navigation bar. This shows all instances of all role types configured for the selected service.

You can also go directly to the Instances page to view instances of a specific role type by clicking one of the links under the **Role Counts** column. This will show only instances of the role type you selected.

The Instances page displays the results of the configuration validation checks it performs for all the role instances for this service.

> **Note:** The information on this page is always the **Current** information for the selected service and roles. This page does not support a historical view: thus, the Time Range Selector is not available.

The information on this page shows:

- The name of the role instance. Click the name to view the role status for that role.
- The host on which it is running. Click the hostname to view the host status details for the host.
- The rack assignment.
- The status. A single value summarizing the state and health of the role instance.
- Whether the role is currently in maintenance mode. If the role has been set into maintenance mode explicitly, you will see the following icon (). If it is in effective maintenance mode due to the service or its host having been set into maintenance mode, the icon will be this ().
- Whether the role is currently decommissioned.

You can sort or filter the Instances list by criteria in any of the displayed columns:

- **Sort**
  1. Click the column header by which you want to sort. A small arrow indicates whether the sort is in ascending or descending order.
  2. Click the column header again to reverse the sort order.
- **Filter** - Type a property value in the Search box or select the value from the facets at the left of the page.

Role Instance Reference

The following tables contain reference information on the status, role state, and health columns for role instances.

Table 4: Status

Indicator	Status	Description
	Started with outdated configuration	For a service, this indicates the service is running, but at least one of its roles is running with a configuration that does not match the current configuration settings in Cloudera Manager. For a role, this indicates a configuration change has been made that requires a restart, and that restart has not yet occurred. Click the indicator to display the Stale Configurations page.
	Starting	The entity is starting up but is not yet running.
	Stopping	The entity is stopping but has not stopped yet.
	Stopped	The entity is stopped, as expected.
	Down	The entity is not running, but it is expected to be running.
	History not available	Cloudera Manager is in historical mode, and the entity does not have historical monitoring support. This is the case for services other than HDFS, MapReduce and HBase such as ZooKeeper, Oozie, and Hue.
	None	The entity does not have a status. For example, it is not something that can be running and it cannot have health. Examples are the HDFS Balancer (which runs from the HDFS Rebalance action) or Gateway roles. The Start and Stop commands are not applicable to these instances.
	Good health	The entity is running with good health. For a specific health test, the returned result is normal or within the acceptable range. For a role or service, this means all health tests for that role or service are **Good**.
	Concerning health	The entity is running with concerning health. For a specific health test, the returned result indicates a potential problem. Typically this means the test result has gone above (or below) a configured Warning threshold. For a role or service, this means that at least one health test is **Concerning**.
	Bad health	The entity is running with bad health. For a specific health test, the test failed, or the returned result indicates a serious problem. Typically this means the test result has gone above (or below) a configured Critical threshold. For a role or service, this means that at least one health test is **Bad**.
	Disabled health	The entity is running, but all of its health tests are disabled.
	Unknown health	The status of a service or role instance is unknown. This can occur for a number of reasons, such as the Service Monitor is not running, or connectivity to the Agent doing the health monitoring has been lost.

Viewing Role Instance Status

To view status for a role instance:

1. Select a service instance to display the **Status** page for that service.
2. Click the **Instances** tab.
3. From the list of roles, select one to display that role instance's **Status** page.

The Actions Menu

Minimum Required Role: Operator (also provided by **Configurator**, **Cluster Administrator**, **Full Administrator**)

The **Actions** menu provides a list of commands relevant to the role type you are viewing. These commands typically include Stopping, Starting, or Restarting the role instance, accessing the Web UI for the role, and may include many other commands, depending on the role you are viewing.

The **Actions** menu is available from the Role Status page only when you are viewing **Current** time status. The menu is disabled if you are viewing a point of time in the past.

Viewing Past Status

The status and health information shown on this page represents the state of the service or role instance at a given point in time. The exceptions are the charts tabs, which show information for the time range currently selected on the Time Range Selector (which defaults to the past 30 minutes). By default, the information shown on this page is for the current time. You can view status for a past point in time simply by moving the time marker (◆) to a point in the past.

When you move the time marker to a point in the past (for Services/Roles that support health history), the Health Status clearly indicates that it is referring to a past time. A Now button (⏭) enables you to quickly switch to view the current state of the service. In addition, the Actions menu is disabled while you are viewing status in the past – to ensure that you cannot accidentally take an action based on outdated status information. See Time Line on page 7 for more details.

You can also view past status by clicking the **Show** link in the Health Tests and Health History on page 41 panel.

Summary

The Summary panel provides basic information about the role instance, where it resides, and the health of its host.

All role types provide the **Summary** panel. Some role instances related to HDFS, MapReduce, and HBase also provide a Health Tests panel and associated charts.

Health Tests and Health History

The Health Tests and Health History panels are shown for roles that are related to HDFS, MapReduce, or HBase. Roles related to other services such as Hue, ZooKeeper, Oozie, and Cloudera Manager itself, do not provide a Health Tests panel. The Health Tests panel shows health test results in an expandable/collapsible list, typically with the specific metrics that the test returned. (You can Expand All or Collapse All from the links at the upper right of the Health Tests panel).

- The color of the text (and the background color of the field) for a Health Test result indicates the status of the results. The tests are sorted by their health status – Good, Concerning, Bad, or Disabled. The entries are collapsed by default. Click the arrow to the left of an entry to expand the entry and display further information.
- Clicking the **Details** link for a health test displays further information about the test, such as the meaning of the test and its possible results, suggestions for actions you can take or how to make configuration changes related to the test. The help text may include a link to the relevant monitoring configuration section for the service. See Configuring Monitoring Settings on page 15 for more information.
- In the Health Tests panel:
  - Clicking › displays the lists of health tests that contributed to the health test.
  - Clicking the **Details** link displays further information about the health test.
- In the Health History panel:
  - Clicking › displays the lists of health tests that contributed to the health history.
  - Clicking the **Show** link moves the time range to the historical time period.

Status Summary

The Status Summary panel reports a roll-up of the status of all the roles.

Charts

Charts are shown for roles that are related to HDFS, MapReduce, HBase, ZooKeeper, Flume, and Cloudera Management Service. Roles related to other services such as Hue, Hive, Oozie, and YARN, do not provide charts.

See Viewing Charts for Cluster, Service, Role, and Host Instances on page 13 for detailed information on the charts that are presented, and the ability to search and display metrics of your choice.

The Processes Tab

To view the processes running for a role instance:

1. Select a service instance to display the Status page for that service.
2. Click the **Instances** tab.
3. From the list of roles, select one to display that role instance's Status page.
4. Click the **Processes** tab.

The Processes page shows the processes that run as part of this service role, with a variety of metrics about those processes.

- To see the location of a process' configuration files, and to view the Environment variable settings, click the **Show** link under **Configuration Files/Environment**.
- If the process provides a Web UI (as is the case for the NameNode, for example) click the link to open the Web UI for that process
- To see the most recent log entries, click the **Show Recent Logs** link.
- To see the full log, stderr, or stdout log files, click the appropriate links.

Running Diagnostic Commands for Roles

Minimum Required Role: **Cluster Administrator** (also provided by **Full Administrator**)

Cloudera Manager allows administrators to run the following diagnostic utility tools against most Java-based role processes:

- List Open Files (`lsof`) - Lists the open files of the process.
- Collect Stack Traces (`jstack`) - Captures Java thread stack traces for the process.
- Heap Dump (`jmap`) - Captures a heap dump for the process.
- Heap Histogram (`jmap -histo`) - Produces a histogram of the heap for the process.

These commands are found on the **Actions** menu of the Cloudera Manager page for the instance of the role. For example, to run diagnostics commands for the HDFS active NameNode, perform these steps:

1. Click the HDFS service on the **Home** > **Status** tab or select it on the Clusters menu.
2. Click **Instances** > **NameNode (Active)**.
3. Click the **Actions** menu.
4. Choose one of the diagnostics commands listed in the lower section of the menu.
5. Click the button in the confirmation dialog box to confirm your choice.
6. When the command is executed, click **Download Result Data** and save the file to view the command output.

Periodic Stacks Collection

Periodic stacks collection allows you to enable and configure the periodic collection of thread stack traces in Cloudera Manager. When stacks collection is enabled for a role, call stacks are output to a log file at regular intervals. The logs can help with diagnosis of performance issues such as deadlock, slow processing, or excessive numbers of threads.

Stacks collection may impact performance for the processes being collected as well as other processes on the host, and is turned off by default. For troubleshooting performance issues, you may be asked by Cloudera Support to enable stacks collection and send the resulting logs to Cloudera for analysis.

Stacks collection is available for the majority of roles in Cloudera Manager. For the HDFS service, for example, you can enable stacks collection for the DataNode, NameNode, Failover Controller, HttpFS, JournalNode, and NFS Gateway. If the **Stacks Collection** category does not appear in the role's configuration settings, the feature is not available for that role.

Configuring Periodic Stacks Collection

To enable and configure periodic stacks collection, open the Cloudera Manager page for a specific service or role. Access the configuration settings in one of the following ways:

- From the service page in Cloudera Manager:
  - Click the **Configuration** tab.
  - Select **Scope** > **NameNode**.
  - Select **Category** > **Stacks Collection**.
- From the service page in Cloudera Manager:
  - Click the **Instances** tab.
  - Click the **Configuration** tab.
  - Select **Scope** > ***role type***.
  - Select **Category** > **Stacks Collection**.

The configuration settings are as follows:

- **Stacks Collection Enabled** - Whether or not periodic stacks collection is enabled.
- **Stacks Collection Directory** - The directory in which stack logs will be placed. If not set, stacks will be logged into a `stacks` subdirectory of the role's log directory.
- **Stacks Collection Frequency** - The frequency with which stacks will be collected.
- **Stacks Collection Data Retention** - The amount of stacks data that will be retained. When the retention limit is reached, the oldest data will be deleted.
- **Stacks Collection Method** - The method that will be used to collect stacks. The jstack option involves periodically running the jstack command against the role's daemon process. The servlet method is available for those roles with an HTTP server endpoint that exposes the current stacks traces of all threads. When the servlet method is selected, that HTTP endpoint is periodically scraped.

As an example, to configure stacks collection for an HDFS NameNode, perform the following steps:

1. Go to the HDFS service page.
2. Click the **Configuration** tab.
3. Select **Scope** > **NameNode**.
4. Select **Category** > **Stacks Collection**.
5. Locate the property or search for it by typing its name in the Search box.
6. Modify the configuration settings if desired.
7. Click **Save Changes**.

Stacks collection configuration settings are stored in a per-role configuration file called `cloudera-stacks-monitor.properties`. Cloudera Manager reads the configuration file and coordinates stack collection. Changes to the configuration settings take effect after a short delay. It is not necessary to restart the role.

Viewing and Downloading Stacks Logs

Stacks are collected and logged to a compressed, rotated log file. A certain amount of the log data is in an uncompressed file. When that file reaches a limit, the file is rotated and bzip2 compressed. Once the total number of files exceeds the configured retention limit, the oldest files are deleted.

Collected stacks data is available for download through the Cloudera Manager UI and API. To view or download stacks logs through the UI, perform the following steps:

1. On the service page, click the **Instances** tab.
2. Click the role in the **Role Type** column.
3. In the Summary section of the role page, click **Stacks Logs**.
4. Click **Stacks Log File** to view the most recent stacks file. Click **Download Stacks Logs** to download a zipped bundle of the stacks logs.

Viewing Running and Recent Commands

Viewing Running and Recent Commands For a Cluster

The indicator  positioned just to the left of the Search field on the right side of the Admin Console main navigation bar displays the number of commands currently running for all services or roles. To display the running commands, click the indicator.

To display all commands that have run and finished recently, do one of the following:

- Click the **All Recent Commands** button in the window that pops up when you click the indicator. This command displays information on all running and recent commands in the same form, as described below.
- Click the Cloudera Manager logo in the main navigation bar and click the **All Recent Commands** tab.

Select a value from the pager 

to control how many commands are listed, or click the arrows to view pages.

The command indicator shows the number of commands running on all clusters you are managing. Likewise, **All Recent Commands** shows all commands that were run and finished within the search time range you specified, across all your managed clusters.

Viewing Running and Recent Commands for a Service or Role

For a selected service or role instance, the **Commands** tab shows which commands are running or have been run for that instance, and what the status, progress, and results are. For example, if you go to the HDFS service shortly after you have installed your cluster and look at the **Commands** tab, you will see recent commands that created the directories, started the HDFS role instances (the NameNode, Secondary NameNode, and DataNode instances), and the command that initially formatted HDFS on the NameNode. This information is useful if a service or role seems to be taking a long time to start up or shut down, or if services or roles are not running or do not appear to have been started correctly. You can view both the status and progress of currently running commands, as well as the status and results of commands run in the past.

1. Click the **Clusters** tab on the top navigation bar.
2. Click the service name to go to the Status tab for that service.
3. For a role instance, click the **Instances** tab and select the role instance name to go to its Status tab.
4. Click the **Commands** tab.

Command Details

The details available for a command depend on whether the command is running or recently completed.

Running Commands

The Running Commands area shows commands that are in progress.

While the status of the command is **In Progress**, an **Abort** button displays so that you can abort the command if necessary.

The Commands status information is updated automatically while the command is running.

After the command has finished running (all its subcommands have finished), the status is updated, the **Abort** buttons disappear, and the information for **Recent Commands** appears as described below.

Recent Commands

The Recent Commands area shows commands that were run and finished within the search time range you specified.

If no commands were run during the selected time range, you can double the time range selection by clicking the **Try expanding the time range selection** link. If you are in the "current time" mode, the beginning time will move; if you

are looking at a time range in the past, both the beginning and ending times of the range are changed. You can also change the time range using the options described in Time Line on page 7.

Select a value from the pager 

to control how many commands are listed, or click the arrows to view pages.

Commands are shown with the most recent ones at the top.

The icon associated with the status (which typically includes the time that the command finished) plus the result message tells you whether the command succeeded ✔ or failed ⊗. If the command failed, it indicates if it was one of the subcommands that actually failed. In many cases, multiple subcommands result from the top level command.

The **First Run** command runs during the initial startup of your cluster. Click this link to view the command history of the cluster startup.

Command Details

In the Running Commands dialog box or Recent Commands page, click a command in the **Command** column to display its details and any subcommands. The page title is the name of the command.

The **Summary** section at the top shows information about the command:

- The current status
- The context, which can be a cluster, service, host, or role
- The time the command started
- The duration of the command
- A message about the command completion
- If the context is a role, links to role instance logs

The **Details** section shows how many steps, if any, the selected command has and lists any subcommands.

Expand a command to view subcommands. In the Running Commands dialog box, each subcommand also has an **Abort** button that is present as long as the subcommand is in progress.

You can perform the following actions:

- Select the option to display all the subcommands or only failed or running commands.
- Click the link in the **Context** column to go to the **Status** page for the component (host, service, or role instance) to which this command is related.
- Click a **Role Log** tab to display the log for that role, and `stdout` and `stderr` if available for the role.

Monitoring Resource Management

With Cloudera Manager 5, statically allocating resources using cgroups is configurable through a single *static service pool* wizard. You allocate services as a percentage of total resources, and the wizard configures the cgroups.

Monitoring Static Service Pools

Static service pools isolate the services in your cluster from one another, so that load on one service has a bounded impact on other services. Services are allocated a static percentage of total resources—CPU, memory, and I/O weight—which are not shared with other services. When you configure static service pools, Cloudera Manager computes recommended memory, CPU, and I/O configurations for the worker roles of the services that correspond to the percentage assigned to each service. Static service pools are implemented per role group within a cluster, using Linux control groups (cgroups) and cooperative memory limits (for example, Java maximum heap sizes). Static service pools can be used to control access to resources by HBase, HDFS, Impala, MapReduce, Solr, Spark, YARN, and add-on services. Static service pools are not enabled by default.

Viewing Static Service Pools

Select **Clusters** > ***Cluster name*** > **Static Service Pools**.If the cluster has a YARN service, the Static Service Pools Status tab displays and shows whether resource management is enabled for the cluster, and the currently configured service pools.

Static Service Pool Status

The Status tab of the Static Service Pools page contains a list of current services that can or have been allocated resources and a set of resource usage charts for the cluster.

Click **Historical Data** to display detailed resource usage charts for each service.

Click a duration link 30m 1h 2h 6h 12h 1d 7d 30d at the top right of the charts to change the time period for which the resource usage displays.

Monitoring Dynamic Resource Pools

A ***dynamic resource pool*** is a named configuration of resources and a policy for scheduling the resources among YARN applications and Impala queries running in the pool. Dynamic resource pools allow you to schedule and allocate resources to YARN applications and Impala queries based on a user's access to specific pools and the resources available to those pools. If a pool's allocation is not in use, it can be preempted and distributed to other pools. Otherwise, a pool receives a share of resources according to the pool's weight. Access control lists (ACLs) restrict who can submit work to dynamic resource pools and administer them.

Viewing Dynamic Resource Pools

1. Go to the YARN service.
2. Click the **Resource Pools** tab.

Click a duration link 30m 1h 2h 6h 12h 1d 7d 30d at the top right of the charts to change the time period for which the resource usage displays.

- **Status** - a summary of the virtual CPU cores and memory that can be allocated by the YARN scheduler.
- **Resource Pools Usage** - a list of pools that have been explicitly configured and pools created by YARN, and properties of the pools. The **Configuration** link takes you to the Dynamic Resource Pool Configuration page.
  - **Allocated Memory** - The memory assigned to the pool that is currently allocated to applications and queries.
  - **Allocated VCores** - The number of virtual CPU cores assigned to the pool that are currently allocated to applications and queries.
  - **Allocated Containers** - The number of YARN containers assigned to the pool whose resources have been allocated.
  - **Pending Containers** - The number of YARN containers assigned to the pool whose resources are pending.

Monitoring Hosts

Cloudera Manager's Host Monitoring features let you manage and monitor the status of the hosts in your clusters.

Viewing All Hosts

To display summary information about all the hosts managed by Cloudera Manager, click **Hosts** in the main navigation bar. The All Hosts page displays with a list of all the hosts managed by Cloudera Manager.

The list of hosts shows the overall status of the Cloudera Manager-managed hosts in your cluster.

- The information provided varies depending on which columns are selected. To change the columns, click the **Columns: *n* Selected** drop-down and select the checkboxes next to the columns to display.
- Click › to the left of the number of roles to list all the role instances running on that host.

- Filter the hosts list by entering search terms (hostname, IP address, or role) in the search box separated by commas or spaces. Use quotes for exact matches (for example, strings that contain spaces, such as a role name) and brackets to search for ranges. Hosts that match any of the search terms are displayed. For example:

```
hostname[1-3], hostname8 hostname9, "hostname.example.com"
hostname.example.com “HDFS DataNode”
```

- You can also search for hosts by selecting a value from the facets in the **Filters** section at the left of the page.
- If the Configuring Agent Heartbeat and Health Status Options are configured as follows:
  - Send Agent heartbeat every *x*
  - Set health status to Concerning if the Agent heartbeats fail *y*
  - Set health status to Bad if the Agent heartbeats fail *z*

  The value *v* for a host's Last Heartbeat facet is computed as follows:

  - *v* < *x* * *y* = Good
  - *v* >= *x* * *y* and <= *x* * *z* = Concerning
  - *v* >= *x* * *z* = Bad

Role Assignments

You can view the assignment of roles to hosts as follows:

1. Click the **Roles** tab.
2. Click a cluster name or **All Clusters**.

Disks Overview

Click the **Disks Overview** tab to display an overview of the status of all disks in the deployment. The statistics exposed match or build on those in `iostat`, and are shown in a series of histograms that by default cover every physical disk in the system.

Adjust the endpoints of the time line to see the statistics for different time periods. Specify a filter in the box to limit the displayed data. For example, to see the disks for a single rack rack1, set the filter to: `logicalPartition = false and rackId = "rack1"` and click **Filter**. Click a histogram to drill down and identify outliers. Mouse over the graph and click ⤢ to display additional information about the chart.

Viewing the Hosts in a Cluster

Do one of the following:

- Select **Clusters** > ***Cluster name*** > **Hosts**.
- In the Home screen, click Hosts in a full form cluster table.

The All Hosts page displays with a list of the hosts filtered by the cluster name.

Viewing Individual Hosts

You can view detailed information about an individual host—resources (CPU/memory/storage) used and available, which processes it is running, details about the host agent, and much more—by clicking a host link on the All Hosts page. See Host Details on page 47.

Host Details

You can view detailed information about each host, including:

- Name, IP address, rack ID
- Health status of the host and last time the Cloudera Manager Agent sent a heartbeat to the Cloudera Manager Server
- Number of cores

- System load averages for the past 1, 5, and 15 minutes
- Memory usage
- File system disks, their mount points, and usage
- Health test results for the host
- Charts showing a variety of metrics and health test results over time.
- Role instances running on the host and their health
- CPU, memory, and disk resources used for each role instance

To view detailed host information:

1. Click the **Hosts** tab.
2. Click the name of one of the hosts. The Status page is displayed for the host you selected.
3. Click tabs to access specific categories of information. Each tab provides various categories of information about the host, its services, components, and configuration.

From the status page you can view details about several categories of information.

Status

The Status page is displayed when a host is initially selected and provides summary information about the status of the selected host. Use this page to gain a general understanding of work being done by the system, the configuration, and health status.

If this host has been decommissioned or is in maintenance mode, you will see the following icon(s) ( , ) in the top bar of the page next to the status message.

Details

This panel provides basic system configuration such as the host's IP address, rack, health status summary, and disk and CPU resources. This information summarizes much of the detailed information provided in other panes on this tab. To view details about the Host agent, click the Host Agent link in the Details section.

Health Tests

Cloudera Manager monitors a variety of metrics that are used to indicate whether a host is functioning as expected. The Health Tests panel shows health test results in an expandable/collapsible list, typically with the specific metrics that the test returned. (You can Expand All or Collapse All from the links at the upper right of the Health Tests panel).

- The color of the text (and the background color of the field) for a health test result indicates the status of the results. The tests are sorted by their health status – Good, Concerning, Bad, or Disabled. The list of entries for good and disabled health tests are collapsed by default; however, Bad or Concerning results are shown expanded.
- The text of a health test also acts as a link to further information about the test. Clicking the text will pop up a window with further information, such as the meaning of the test and its possible results, suggestions for actions you can take or how to make configuration changes related to the test. The help text for a health test also provides a link to the relevant monitoring configuration section for the service. See Configuring Monitoring Settings on page 15 for more information.

Health History

The Health History provides a record of state transitions of the health tests for the host.

- Click the arrow symbol at the left to view the description of the health test state change.
- Click the **View** link to open a new page that shows the state of the host at the time of the transition. In this view some of the status settings are greyed out, as they reflect a time in the past, not the current status.

File Systems

The File systems panel provides information about disks, their mount points and usage. Use this information to determine if additional disk space is required.

Roles

Use the Roles panel to see the role instances running on the selected host, as well as each instance's status and health. Hosts are configured with one or more role instances, each of which corresponds to a service. The role indicates which daemon runs on the host. Some examples of roles include the NameNode, Secondary NameNode, Balancer, JobTrackers, DataNodes, RegionServers and so on. Typically a host will run multiple roles in support of the various services running in the cluster.

Clicking the role name takes you to the role instance's status page.

You can delete a role from the host from the Instances tab of the Service page for the parent service of the role. You can add a role to a host in the same way. See Role Instances.

Charts

Charts are shown for each host instance in your cluster.

See Viewing Charts for Cluster, Service, Role, and Host Instances on page 13 for detailed information on the charts that are presented, and the ability to search and display metrics of your choice.

Processes

The Processes page provides information about each of the processes that are currently running on this host. Use this page to access management web UIs, check process status, and access log information.

Note: The Processes page may display exited startup processes. Such processes are cleaned up within a day.

The Processes tab includes a variety of categories of information.

- **Service** - The name of the service. Clicking the service name takes you to the service status page. Using the triangle to the right of the service name, you can directly access the tabs on the role page (such as the Instances, Commands, Configuration, Audits, or Charts Library tabs).
- **Instance** - The role instance on this host that is associated with the service. Clicking the role name takes you to the role instance's status page. Using the triangle to the right of the role name, you can directly access the tabs on the role page (such as the Processes, Commands, Configuration, Audits, or Charts Library tabs) as well as the status page for the parent service of the role.
- **Name** - The process name.
- **Links** - Links to management interfaces for this role instance on this system. These is not available in all cases.
- **Status** - The current status for the process. Statuses include stopped, starting, running, and paused.
- **PID** - The unique process identifier.
- **Uptime** - The length of time this process has been running.
- **Full log file** - A link to the full log (a file external to Cloudera Manager) for this host log entries for this host.
- **Stderr** - A link to the stderr log (a file external to Cloudera Manager) for this host.
- **Stdout** - A link to the stdout log (a file external to Cloudera Manager) for this host.

Resources

The Resources page provides information about the resources (CPU, memory, disk, and ports) used by every service and role instance running on the selected host.

Each entry on this page lists:

- The service name
- The name of the particular instance of this service
- A brief description of the resource
- The amount of the resource being consumed or the settings for the resource

The resource information provided depends on the type of resource:

- **CPU** - An approximate percentage of the CPU resource consumed.
- **Memory** - The number of bytes consumed.
- **Disk** - The disk location where this service stores information.
- **Ports** - The port number being used by the service to establish network connections.

Commands

The Commands page shows you running or recent commands for the host you are viewing. See Viewing Running and Recent Commands on page 44 for more information.

Configuration

Minimum Required Role: Full Administrator

The Configuration page for a host lets you set properties for the selected host. You can set properties in the following categories:

- **Advanced** - Advanced configuration properties. These include the Java Home Directory, which explicitly sets the value of `JAVA_HOME` for all processes. This overrides the auto-detection logic that is normally used.
- **Monitoring** - Monitoring properties for this host. The monitoring settings you make on this page will override the global host monitoring settings you make on the Configuration tab of the Hosts page. You can configure monitoring properties for:
  - health check thresholds
  - the amount of free space on the filesystem containing the Cloudera Manager Agent's log and process directories
  - a variety of conditions related to memory usage and other properties
  - alerts for health check events

  For some monitoring properties, you can set thresholds as either a percentage or an absolute value (in bytes).
- **Other** - Other configuration properties.
- **Parcels** - Configuration properties related to parcels. Includes the **Parcel Director** property, the directory that parcels will be installed into on this host. If the `parcel_dir` variable is set in the Agent's `config.ini` file, it will override this value.
- **Resource Management** - Enables resource management using control groups (cgroups).

For more information, see the description for each or property or see Modifying Configuration Properties Using Cloudera Manager.

Components

The Components page lists every component installed on this host. This may include components that have been installed but have not been added as a service (such as YARN, Flume, or Impala).

This includes the following information:

- **Component** - The name of the component.
- **Version** - The version of CDH from which each component came.
- **Component Version** - The detailed version number for each component.

Audits

The Audits page lets you filter for audit events related to this host. See Lifecycle and Security Auditing on page 97 for more information.

Charts Library

The Charts Library page for a host instance provides charts for all metrics kept for that host instance, organized by category. Each category is collapsible/expandable. See Viewing Charts for Cluster, Service, Role, and Host Instances on page 13 for more information.

Host Inspector

You can use the host inspector to gather information about hosts that Cloudera Manager is currently managing. You can review this information to better understand system status and troubleshoot any existing issues. For example, you might use this information to investigate potential DNS misconfiguration.

The inspector runs tests to gather information for functional areas including:

- Networking
- System time
- User and group configuration
- HDFS settings
- Component versions

Common cases in which this information is useful include:

- Installing components
- Upgrading components
- Adding hosts to a cluster
- Removing hosts from a cluster

Running the Host Inspector

1. Click the **Hosts** tab and select **All Hosts**.
2. Click the **Inspect All Hosts** button. Cloudera Manager begins several tasks to inspect the managed hosts.
3. After the inspection completes, click **Download Result Data** or **Show Inspector Results** to review the results.

The results of the inspection displays a list of all the validations and their results, and a summary of all the components installed on your managed hosts.

If the validation process finds problems, the **Validations** section will indicate the problem. In some cases the message may indicate actions you can take to resolve the problem. If an issue exists on multiple hosts, you may be able to view the list of occurrences by clicking a small triangle that appears at the end of the message.

The **Version Summary** section shows all the components that are available from Cloudera, their versions (if known) and the CDH distribution to which they belong.

Viewing Past Host Inspector Results

You can view the results of a past host inspection by looking for the Host Inspector command using the **Recent Commands** feature.

1. Click the Running Commands indicator () just to the left of the Search box at the right hand side of the navigation bar.
2. Click the **Recent Commands** button.
3. If the command is too far in the past, you can use the Time Range Selector to move the time range back to cover the time period you want.
4. When you find the Host Inspector command, click its name to display its subcommands.
5. Click the **Show Inspector Results** button to view the report.

See Viewing Running and Recent Commands on page 44 for more information about viewing past command activity.

Monitoring Activities

Cloudera Manager's activity monitoring capability monitors the MapReduce, Pig, Hive, Oozie, and streaming jobs, Impala queries, and YARN applications running or that have run on your cluster. When the individual jobs are part of larger workflows (using Oozie, Hive, or Pig), these jobs are aggregated into MapReduce jobs that can be monitored as a whole, as well as by the component jobs.

If you are running multiple clusters, there will be a separate link in the Clusters tab for each cluster's MapReduce activities, Impala queries, and YARN applications.

The following sections describe how to view and monitor activities that run on your cluster.

Monitoring MapReduce Jobs

A MapReduce job is a unit of processing (query or transformation) on the data stored within a Hadoop cluster. You can view information about the different jobs that have run in your cluster during a selected time span.

- The list of jobs provides specific metrics about the jobs that were submitted, were running, or finished within the time frame you select.
- You can select charts that show a variety of metrics of interest, either for the cluster as a whole or for individual jobs.

You can use the Time Range Selector or a duration link (30m 1h 2h 6h 12h 1d 7d 30d) to set the time range. (See Time Line on page 7 for details).

Note: Activity Monitor treats the original job start time as immutable. If a job is resubmitted due to failover it will retain its original start time.

You can select an activity and drill down to look at the jobs and tasks spawned by that job:

- View the children (MapReduce jobs) of a Pig or Hive activity.
- View the task attempts generated by a MapReduce job.
- View the children (MapReduce, Pig, or Hive activities) of an Oozie job.
- View the activity or job statistics in a detail report format.
- Compare the selected activity to a set of other similar activities, to determine if the selected activity showed anomalous behavior. For example, if a standard job suddenly runs much longer than usual, this may indicate issues with your cluster.
- Display the distribution of task attempts that made up a job, by different metrics compared to task duration. You can use this, for example, to determine if tasks running on a certain host are performing slower than average.
- Kill a running job, if necessary.

Note: Some activity data is sampled at one-minute intervals. This means that if you run a very short job that both starts and ends within the sampling interval, it may not be detected by the Activity Monitor, and thus will not appear in the Activities list or charts.

Viewing and Filtering MapReduce Activities

This section describes the various actions you can perform in the MapReduce Activities page:

Viewing MapReduce Activities

1. Select **Clusters** > ***Cluster name*** > ***MapReduce service name*** **Jobs**. The *MapReduce service name* page displays a list of activities. The columns in the Activities list show statistics about the performance of and resources used by each activity (and you can modify the default display by adding or removing columns).

   - The leftmost column holds a shortcut menu button (▼). Click this button to display a menu of commands relevant to the job shown in that row. The possible commands are:

Children	For a Pig, Hive or Oozie activity, takes you to the **Children** tab of the individual activity page. You can also go to this page by clicking the activity ID in the activity list. This command only appears for Pig, Hive or Oozie activities.
Tasks	For a MapReduce job, takes you to the **Tasks** tab of the individual job page. You can also go to this page by clicking the job ID in the activity or activity children list. This command only appears for a MapReduce job.
Details	Takes you to the **Details** tab where you can view the activity or job statistics in report form.
Compare	Takes you to the **Compare** tab where you can see how the selected activity compares to other similar activities in terms of a wide variety of metrics.
Task Distribution	Takes you to the **Task Distribution** tab where you can view the distribution of task attempts that made up this job, by amount of data and task duration. This command is available for MapReduce and Streaming jobs.
Kill Job	A pop-up asks for confirmation that you want to kill the job. This command is available only for MapReduce and Streaming jobs.

- The second column shows a chart icon (). Select this to chart statistics for the job. If there are charts showing similar statistics for the cluster or for other jobs, the statistics for the job are added to the chart. See Activity Charts on page 55 for more details.
- The third column shows the status of the job, if the activity is a MapReduce job:

	The job has been submitted.
	The job has been started.
	The job is assumed to have succeeded.
	The job has finished successfully.
	The job's final state is unknown.
	The job has been suspended.
	The job has failed.
	The job has been killed.

- The fourth column shows the type of activity:

	MapReduce job
	Pig job
	Hive job
	Oozie job
	Streaming job

Selecting Columns to Show in the Activities List

In the Activities list, you can display or hide any of the statistics that Cloudera Manager collects. By default only a subset of the possible statistics are displayed.

1. Click the **Select Columns to Display** icon (). A pop-up panel lets you turn on or off a variety of metrics that may be of interest.
2. Check or uncheck the columns you want to include or remove from the display. As you check or uncheck an item, its column immediately appears or disappears from the display.
3. Click the × in the upper right corner to close the panel.

> **Note:** You cannot hide the shortcut menu or chart icon columns. Also, column selections are retained only for the current session.

Sorting the Activities List

You can sort the Activities list by the contents of any column:

1. Click the column header to initiate a sort. The small arrow that appears next to the column header indicates the sort direction.
2. Click the column header to reverse the sort direction.

Filtering the Activities List

You can filter the list of activities based on values of any of the metrics that are available. You can also easily filter for certain common queries from the drop-down menu next to the Search button at the top of the Activities list. By default, it is set to show **All Activities**.

To use one of the predefined filters:

- Click the
  
  to the right of the **Search** button and select the filter you want to run. There are predefined filters to search by job type (for example Pig activities, MapReduce jobs, and so on) or for running, failed, or long-running activities.

To create a filter:

1. Click the
   
   to the right of the **Search** button and select **Custom**.
2. Select a metric from the drop-down list in the first field; you can create a filter based on any of the available metrics.
3. Once you select a metric, fill in the rest of the fields; your choices depend on the type of metric you have selected. Use the percent character `%` as a wildcard in a string; for example, `Id matches job%0001` will look for any MapReduce job ID with suffix `0001`.
4. To create a compound filter, click the plus icon at the end of the filter row to add another row. If you combine filter criteria, all criteria must be true for an activity to match.
5. To remove a filter criteria from a compound filter, click the minus icon at the end of the filter row. Removing the last row removes the filter.
6. To include any children of a Pig, Hive, or Oozie activity in your search results, check the **Include Child Activities** checkbox. Otherwise, only the top-level activity will be included, even if one or more child activities matched the filter criteria.
7. Click the **Search** button (which appears when you start creating the filter) to run the filter.

> **Note:** Filters are remembered across user sessions — that is, if you log out the filter will be preserved and will still be active when you log back in. Newly-submitted activities will appear in the Activity List only if they match the filter criteria.

Activity Charts

By default the charts show aggregated statistics about the performance of the cluster: Tasks Running, CPU Usage, and Memory Usage. There are additional charts you can enable from a pop-up panel. You can also superimpose individual job statistics on any of the displayed charts.

Most charts display multiple metrics within the same chart. For example, the **Tasks Running** chart shows two metrics: **Cluster, Running Maps** and **Cluster, Running Reduces** in the same chart. Each metric appears in a different color.

- To see the exact values at a given point in time, move the cursor over the chart – a movable vertical line pinpoints a specific time, and a tooltip shows you the values at that point.
- You can use the time range selector at the top of the page to zoom in – the chart display will follow. In order to zoom out, you can use the Time Range Selector at the top of the page or click the link below the chart.

To select additional charts:

1. Click  at the top right of the chart panel to open the Customize dialog box.
2. Check or uncheck the boxes next to the charts you want to show or hide.

To show or hide cluster-wide statistics:

- Check or uncheck the **Cluster** checkbox at the top of the Charts panel.

To chart statistics for an individual job:

- Click the chart icon () in the row next to the job you want to show on the charts. The job ID will appear in the top bar next to the Cluster checkbox, and the statistics will appear on the appropriate chart.
- To remove a job's statistics from the chart, click the ✖ next to the job ID in the top bar of the chart.

Note: Chart selections are retained only for the current session.

To expand, contract, or hide the charts

- Move the cursor over the divider between the Activities list and the charts, grab it and drag to expand or contract the chart area compared to the Activities list.
- Drag the divider all the way to the right to hide the charts, or all the way to the left to hide the Activities list.

Viewing the Jobs in a Pig, Oozie, or Hive Activity

The Activity **Children** tab shows the same information as does the Activities tab, except that it shows only jobs that are children of a selected Pig, Hive or Oozie activity. In addition, from this tab you can view the details of the Pig, Hive or Oozie activity as a whole, and compare it to similar activities.

1. Click the **Activities** tab.
2. Click the Pig, Hive or Oozie activity you want to inspect. This presents a list of the jobs that make up the Pig, Hive or Oozie activity.

The functions under the **Children** tab are the same as those seen under the **Activities** tab. You can filter the job list, show and hide columns in the job list, show and hide charts and plot job statistics on those charts.

- Click an individual job to view Task information and other information for that child. See Viewing and Filtering MapReduce Activities on page 52 for details of how the functions on this page work.

In addition, viewing a Pig, Hive or Oozie activity provides the following tabs:

- The **Details** tab shows Activity details in a report form. See Viewing Activity Details in a Report Format for more information.
- The **Compare** tab compares this activity to other similar activity. The main difference between this and a comparison for a single MapReduce activity is that the comparison is done looking at other activities of the same type (Pig,

Hive or Oozie) but does include the child jobs of the activity. See Comparing Similar Activities for an explanation of that tab.

Task Attempts

The Tasks tab contains a list of the Map and Reduce task attempts that make up a job.

Viewing a Job's Task Attempts

1. From the **Clusters** tab, in the section marked **Other**, select the activity you want to inspect.
   - If the activity is a MapReduce job, the **Tasks** tab opens.
   - If the activity is a Pig, Hive, or Oozie activity, select the job you want to inspect from the activity's **Children** tab to open the **Tasks** tab.

The columns shown under the **Tasks** tab display statistics about the performance of and resources used by the task attempts spawned by the selected job. By default only a subset of the possible metrics are displayed — you can modify the columns that are displayed to add or remove the columns in the display.

- The status of an attempt is shown in the Attempt Status column:

	The attempt is running.
	The attempt has succeeded.
	The attempt has failed.
	The attempt has been unassigned.
	The attempt has been killed.
	The attempt's final state is unknown.

- Click the task ID to view details of the individual task.

You can use the **Zoom to Duration** button to zoom the Time Range Selector to the exact time range spanned by the activity whose tasks you are viewing.

Selecting Columns to Show in the Tasks List

In the Tasks list, you can display or hide any of the metrics the Cloudera Manager collects for task attempts. By default a subset of the possible metrics are displayed.

1. Click the **Select Columns to Display** icon (). A pop-up panel lets you turn on or off a variety of metrics that may be of interest.
2. Check or uncheck the columns you want to include or remove from the display. As you check or uncheck an item, its column immediately appears or disappears from the display.
3. Click the in the upper right corner to close the panel.

Sorting the Tasks List

You can sort the tasks list by any of the information displayed in the list:

1. Click the column header to initiate a sort. The small arrow that appears next to the column header indicates the sort direction.
2. Click the column header to reverse the sort direction.

Filtering the Tasks List

You can filter the list of tasks based on values of any of the metrics that are available.

To use one of the predefined filters:

- Click the
  
  to the right of the **Search** button and select the filter you want to run. There are predefined filters to search by job type (for example Pig activities, MapReduce jobs, and so on) or for running, failed, or long-running activities.

To create a filter:

1. Click the
   
   to the right of the **Search** button and select **Custom**.
2. Select a metric from the drop-down list in the first field; you can create a filter based on any of the available metrics.
3. Once you select a metric, fill in the rest of the fields; your choices depend on the type of metric you have selected. Use the percent character `%` as a wildcard in a string; for example, `Id matches job%0001` will look for any MapReduce job ID with suffix `0001`.
4. To create a compound filter, click the plus icon at the end of the filter row to add another row. If you combine filter criteria, all criteria must be true for an activity to match.
5. To remove a filter criteria from a compound filter, click the minus icon at the end of the filter row. Removing the last row removes the filter.
6. To include any children of a Pig, Hive, or Oozie activity in your search results, check the **Include Child Activities** checkbox. Otherwise, only the top-level activity will be included, even if one or more child activities matched the filter criteria.
7. Click the **Search** button (which appears when you start creating the filter) to run the filter.

Note: The filter persists only for this user session — when you log out, tasks list filter is removed.

Viewing Activity Details in a Report Format

The Details tab for an activity shows the job or activity statistics in a report format.

To view activity details for an individual MapReduce job:

1. Select a MapReduce job from the Clusters tab *or* Select a Pig, Hive or Oozie activity, then select a MapReduce job from the **Children** tab.
2. Select the **Details** tab after the job page is displayed.

This display information about the individual MapReduce job in a report format.

From this page you can also access the **Job Details** and **Job Configuration** pages on the JobTracker web UI.

- Click the **Job Details** link at the top of the report to be taken to the job details web page on the JobTracker host.
- Click the **Job Configuration** link to be taken to the job configuration web page on the JobTracker host.

To view activity details for a Pig, Hive, or Oozie activity:

1. Select a Pig, Hive or Oozie activity.
2. Select the **Details** tab after the list of child jobs is displayed.

This displays information about the Pig, Oozie, or Hive job as a whole.

Note that this the same data you would see for the activity if you displayed all possible columns in the Activities list.

Comparing Similar Activities

It can be useful to compare the performance of similar activities if, for example, you suspect that a job is performing differently than other similar jobs that have run in the past.

The **Compare** tab shows you the performance of the selected job compared with the performance of other similar jobs. Cloudera Manager identifies jobs that are similar to each other (jobs that are basically running the same code – the same Map and Reduce classes, for example).

To compare an activity to other similar activities:

1. Select the job or activity from the Activities list.
2. Click the **Compare** tab.

The activity comparison feature compares performance and resource statistics of the selected job to the mean value of those statistics across a set of the most recent similar jobs. The table provides visual indicators of how the selected job deviates from the mean calculated for the sample set of jobs, as well as providing the actual statistics for the selected job and the set of the similar jobs used to calculate the mean.

- **The first row** in the comparison table displays a set of visual indicators of how the selected job deviates from the mean of all the similar jobs (the combined Average values). This is displayed for each statistic for which a comparison makes sense. The diagram in the ID column shows the elements of the indicator, as follows:
  - The line at the midpoint of the bar represents the mean value of all similar jobs. The colored portion of the bar indicates the degree of deviation of your selected job from the mean. The top and bottom of the bar represent two standard deviations (plus or minus) from the mean.
  - For a given metric, if the value for your selected job is within two standard deviations of the mean, the colored portion of the bar is blue.
  - If a metric for your selected job is more than two standard deviations from the mean, the colored portion of the bar is red.
- **The following rows** show the actual values for other similar jobs. These are the sets of values that were used to calculate the mean values shown in the Combined Averages row. The most recent ten similar jobs are used to calculate the average job statistics, and these are the jobs that are shown in the table.

Viewing the Distribution of Task Attempts

The Task Distribution tab provides a graphical view of the performance of the Map and Reduce tasks that make up a job.

To display the task distribution metrics for a job:

1. Do one of the following:
   - Select a MapReduce job from the **Activities** list.
   - Select a job from the **Children** tab of a Pig, Hive, or Oozie activity.
2. Click the **Task Distribution** tab.

The chart that appears initially shows the distribution of Map Input Records by Duration; you can change the Y-axis to chart a number of different metrics.

You can use the **Zoom to Duration** button to zoom the Time Range Selector to the exact time range spanned by the activity whose tasks you are viewing.

The Task Distribution Chart

The Task Distribution chart shows the distribution of attempts according to their duration on the X-axis and a number of different metrics on the Y-axis. Each cell represents the number of tasks whose performance statistics fall within the parameters of the cell.

The Task Distribution chart is useful for detecting tasks that are outliers in your job, either because of skew, or because of faulty TaskTrackers. The chart can clearly show if some tasks deviate significantly from the majority of task attempts.

Normally, the distribution of tasks will be fairly concentrated. If, for example, some Reducers receive much more data than others, that will be represented by having two discrete sections of density on the graph. That suggests that there may be a problem with the user code, or that there's skew in the underlying data. Alternately, if the input sizes of various Map or Reduce tasks are the same, but the time it takes to process them varies widely, it might mean that certain TaskTrackers are performing more poorly than others.

You can click in a cell and see a list of the TaskTrackers that correspond to the tasks whose performance falls within the cell.

The X-axis show the task duration is seconds. From the drop-down you can chose different metrics for the Y-axis: Input or Output records or bytes for Map tasks, or the number of CPU seconds for the user who ran the job:

- Map Input Records vs. Duration
- Map Output Records vs. Duration
- Map Input Bytes vs. Duration
- Map Output Bytes vs. Duration
- Map Total User CPU seconds vs. Duration
- Reduce Input Records vs. Duration
- Reduce Output Records vs. Duration
- Reduce Total User CPU seconds vs. Duration

TaskTracker Hosts

To the right of the chart is a table that shows the TaskTracker hosts that processed the tasks in the selected cell, along with the number of task attempts each host executed.

You can select a cell in the table to view the TaskTracker hosts that correspond to the tasks in the cell.

- The area above the TaskTracker table shows the type of task and range of data volume (or User CPUs) and duration times for the task attempts that fall within the cell.
- The table itself shows the TaskTracker hosts that executed the tasks that are represented within the cell, and the number of task attempts run on that host.

Clicking a TaskTracker hostname takes you to the Role Status page for that TaskTracker instance.

Monitoring Impala Queries

The Impala Queries page displays information about Impala queries that are running and have run in your cluster. You can filter the queries by time period and by specifying simple filtering expressions.

> **Note:** The Impala query monitoring feature requires Impala 1.0.1 and higher.

Viewing Queries

1. Do one of the following:
   - Select **Clusters** > ***Cluster name*** > ***Impala service name*** **Queries**.
   - On the **Home** > **Status** tab, select ***Impala service name*** and click the **Queries** tab.

   The Impala queries run during the selected time range display in the Results Tab on page 60.

You can also perform the following actions on this page:

Table 5: Viewing Queries Actions

Action	Description
Filter the displayed queries	Create filter expressions manually, select preconfigured filters, or use the **Workload Summary** section to build a query interactively. See Filtering Queries on page 61.
Select additional attributes for display.	Click **Select Attributes**. Selected attributes also display as available filters in the **Workload Summary** section. To display information about attributes, hover over a field label. See Filter Attributes on page 63.

Action	Description
	Only attributes that support filtering appear in the **Workload Summary** section. See the Table 6: Attributes on page 63 table.
View a histogram of the attribute values.	Click the icon to the right of each attribute displayed in the **Workload Summary** section.
Display charts based on the filter expression and selected attributes.	Cick the **Charts** tab.
View charts that help identify whether Impala best practices are being followed.	Click the **Best Practices** link.
Export a JSON file with the query results that you can use for further analysis.	Click **Export**.

Configuring Impala Query Monitoring

You can configure the visibility of the Impala query results and the size of the storage allocated to Impala query results.

For information on how to configure whether admin and non-admin users can view all queries, only that user's queries, or no queries, see Configuring Query Visibility on page 17.

Query information is stored in-memory in a ring buffer. If you restart Service Monitor, all queries are lost, and older queries eventually are dropped. For information on how to configure the query store, see Configuring Impala Query Data Store Maximum Size on page 17.

Impala Best Practices

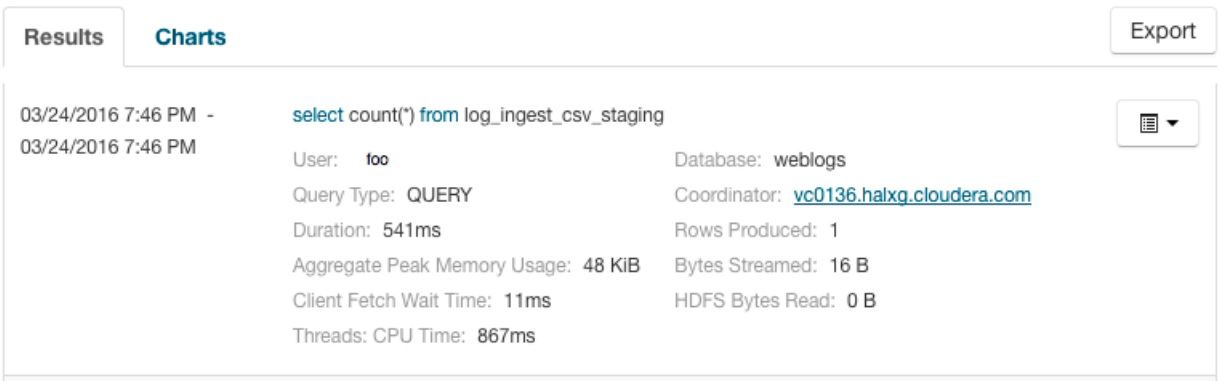
To open the Impala Best Practices page, click the **Best Practices** tab on the Impala service page. The page contains charts that include description of each best practice and how to determine if it is being followed. See the Impala documentation for more detail on each best practice and for additional best practices.

Adjust the time range to see data on queries run at different times. Click the charts to get more detail on individual queries. Use the filter box at the top right of the Best Practices page to adjust which data is shown on the page. For example, to see just the queries that took more than ten seconds, make the filter `query_duration > 10s`.

Create a trigger based on any best practice by choosing **Create Trigger** from the individual chart drop-down menu.

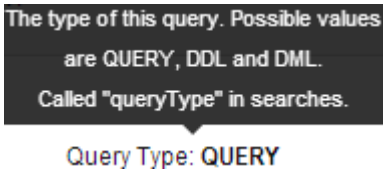
Results Tab

Queries appear on the **Results** tab, with the most recent at the top. Each query has summary and detail information. A query summary includes the following default attributes: start and end timestamps, statement, duration, rows produced, user, coordinator, database, and query type. For example:

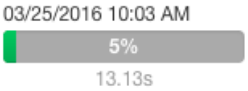


You can add additional attributes to the summary by clicking the Attribute Selector. In each query summary, the query statement is truncated if it is too long to display. To display the entire statement, click ❯. The query entry expands to

display the entire query string. To collapse the query display, click ❯. To display information about query attributes and possible values, hover over a field in a query. For example:



A running job displays a progress bar under the starting timestamp:



If an error occurred while processing the query, Error displays under the complete timestamp.

Use the Actions drop-down menu to the right of each query listing to do the following. (Not all options display, depending on the type of job.)

- Query Details – Opens a details page for the job. See query details.
- User's Impala Queries – Displays a list of queries run by the user for the current job.
- Cancel (running queries only) – Cancel a running query (administrators only). Canceling a running query creates an audit event. When you cancel a query, Canceled replaces the progress bar.
- Queries in the same YARN pool – Displays queries that use the same resource pool.

Filtering Queries

You filter queries by selecting a time range and specifying a filter expression in the search box.

You can use the Time Range Selector or a duration link (30m 1h 2h 6h 12h 1d 7d 30d) to set the time range. (See Time Line on page 7 for details).

Filter Expressions

Filter expressions specify which entries should display when you run the filter. The simplest expression consists of three components:

- **Attribute** - Query language name of the attribute.
- **Operator** - Type of comparison between the attribute and the attribute value. Cloudera Manager supports the standard comparator operators =, `!=`, >, <, >=, <=, and `RLIKE`. (RLIKE performs regular expression matching as specified in the Java Pattern class documentation.) Numeric values can be compared with all operators. String values can be compared with =, `!=`, and `RLIKE`. Boolean values can be compared with = and `!=`.
- **Value** - The value of the attribute. The value depends on the type of the attribute. For a Boolean value, specify either `true` or `false`. When specifying a string value, enclose the value in double quotes.

You create compound filter expressions using the AND and OR operators. When more than one operator is used in an expression, AND is evaluated first, then OR. To change the order of evaluation, enclose subexpressions in parentheses.

Compound Expressions

To find all the queries issued by the root user that produced over 100 rows, use the expression:

```
user = "root" AND rowsProduced > 100
```

To find all the executing queries issued by users Jack or Jill, use the expression:

```
executing = true AND (user = "Jack" OR user = "Jill")
```

Choosing and Running a Filter

1. Do one of the following:

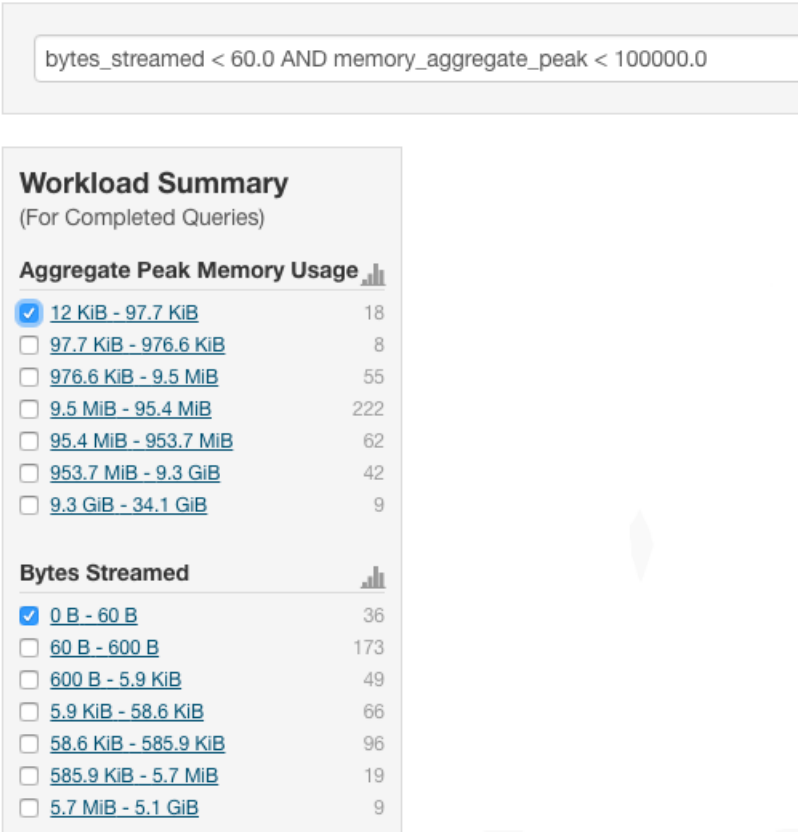
   - **Select a Suggested or Recently Run Filter**

     Click the
     
     to the right of the **Search** button to display a list of sample and recently run filters, and select a filter. The filter text displays in the text box.

   - **Construct a Filter from the Workload Summary Attributes**

     Optionally, click **Select Attributes** to display a dialog box where you can chose which attributes to display in the **Workload Summary** section. Select the checkbox next to one or more attributes, and click **Close**.

     The attributes display in the **Workload Summary** section along with values or ranges of values that you can filter on. The values and ranges display as links with checkboxes. Select one or more checkboxes to add the range or value to the query. Click a link to run a query on that value or range. For example:

   - **Type a Filter**

     1. Start typing or press **Spacebar** in the text box. As you type, filter attributes matching the typed letter display. If you press **Spacebar**, standard filter attributes display. These suggestions are part of typeahead, which helps build valid queries. For information about the attribute name and supported values for each field, hover over the field in an existing query.
     2. Select an attribute and press **Enter**.
     3. Press **Spacebar** to display a drop-down list of operators.
     4. Select an operator and press **Enter**.
     5. Specify an attribute value in one of the following ways:

        - For attribute values that support typeahead, press **Spacebar** to display a drop-down list of values and press **Enter**.

- Type a value.

2. Click in the text box and press **Enter** or click **Search**. The list displays the results that match the specified filter. The Workload Summary section refreshes to show only the values for the selected filter. The filter is added to the Recently Run list.

Filter Attributes

The following table includes available filter attributes and their names in Cloudera Manager, types, and descriptions.

Note: Only attributes for which the **Supports Filtering?** column value is TRUE appear in the **Workload Summary** section.

Table 6: Attributes

Display Name (Attribute Name)	Type	Supports Filtering?	Description
Admission Result (admission_result)	STRING	TRUE	The result of admission, whether immediately, queued, rejected, or timed out. Called 'admission_result' in searches.
Admission Wait Time (admission_wait)	MILLISECONDS	TRUE	The time from submission for admission to completion of admission. Called 'admission_wait' in searches.
Aggregate Peak Memory Usage (memory_aggregate_peak)	BYTES	TRUE	The highest amount of memory allocated by this query at a particular time across all nodes. Called 'memory_aggregate_peak' in searches.
Bytes Streamed (bytes_streamed)	BYTES	TRUE	The total number of bytes sent between Impala Daemons while processing this query. Called 'bytes_streamed' in searches.
Client Fetch Wait Time (client_fetch_wait_time)	MILLISECONDS	TRUE	The total amount of time the query spent waiting for the client to fetch row data. Called 'client_fetch_wait_time' in searches.
Client Fetch Wait Time Percentage (client_fetch_wait_time_percentage)	NUMBER	TRUE	The total amount of time the query spent waiting for the client to fetch row data divided by the query duration. Called 'client_fetch_wait_time_percentage' in searches.
Connected User (connected_user)	STRING	TRUE	The user who created the Impala session that issued this query. This is distinct from 'user' only if delegation is in use. Called 'connected_user' in searches.
Coordinator (coordinator_host_id)	STRING	TRUE	The host coordinating this query. Called 'coordinator_host_id' in searches.
Database (database)	STRING	TRUE	The database on which the query was run. Called 'database' in searches.
DDL Type (ddl_type)	STRING	TRUE	The type of DDL query. Called 'ddl_type' in searches.

Display Name (Attribute Name)	Type	Supports Filtering?	Description
Delegated User (delegated_user)	STRING	TRUE	The effective user for the query. This is set only if delegation is in use. Called 'delegated_user' in searches.
Duration (query_duration)	MILLISECONDS	TRUE	The duration of the query in milliseconds. Called 'query_duration' in searches.
Estimated per Node Peak Memory (estimated_per_node_peak_memory)	BYTES	TRUE	The planning process's estimate of per-node peak memory usage for the query. Called 'estimated_per_node_peak_memory' in searches.
Executing (executing)	BOOLEAN	FALSE	Whether the query is currently executing. Called 'executing' in searches.
File Formats (file_formats)	STRING	FALSE	An alphabetically sorted list of all the file formats used in the query. Called 'file_formats' in searches.
HBase Bytes Read (hbase_bytes_read)	BYTES	TRUE	The total number of bytes read from HBase by this query. Called 'hbase_bytes_read' in searches.
HBase Scanner Average Read Throughput (hbase_scanner_average_bytes_read_per_second)	BYTES_PER_SECOND	TRUE	The average HBase scanner read throughput for this query. This is computed by dividing the total bytes read from HBase by the total time spent reading by all HBase scanners. Called 'hbase_scanner_average_bytes_read_per_second' in searches.
HDFS Average Scan Range (hdfs_average_scan_range)	BYTES	TRUE	The average HDFS scan range size for this query. HDFS scan nodes that contained only a single scan range are not included in this computation. Low numbers for a query might indicate reading many small files which negatively impacts performance. Called 'hdfs_average_scan_range' in searches.
HDFS Bytes Read (hdfs_bytes_read)	BYTES	TRUE	The total number of bytes read from HDFS by this query. Called 'hdfs_bytes_read' in searches.
HDFS Bytes Read From Cache (hdfs_bytes_read_from_cache)	BYTES	TRUE	The total number of bytes read from HDFS that were read from the HDFS cache. This is only for completed queries. Called 'hdfs_bytes_read_from_cache' in searches.
HDFS Bytes Read From Cache Percentage (hdfs_bytes_read_from_cache_percentage)	NUMBER	TRUE	The percentage of all bytes read by this query that were read from the HDFS cache. This is only for completed queries. Called 'hdfs_bytes_read_from_cache_percentage' in searches.
HDFS Bytes Skipped (hdfs_bytes_skipped)	BYTES	TRUE	The total number of bytes that had to be skipped by this query while reading from HDFS. Any number above zero may indicate a problem. Called 'hdfs_bytes_skipped' in searches.

Display Name (Attribute Name)	Type	Supports Filtering?	Description
HDFS Bytes Written (hdfs_bytes_written)	BYTES	TRUE	The total number of bytes written to HDFS by this query. Called 'hdfs_bytes_written' in searches.
HDFS Local Bytes Read (hdfs_bytes_read_local)	BYTES	TRUE	The total number of local bytes read from HDFS by this query. This is only for completed queries. Called 'hdfs_bytes_read_local' in searches.
HDFS Local Bytes Read Percentage (hdfs_bytes_read_local_percentage)	NUMBER	TRUE	The percentage of all bytes read from HDFS by this query that were local. This is only for completed queries. Called 'hdfs_bytes_read_local_percentage' in searches.
HDFS Remote Bytes Read (hdfs_bytes_read_remote)	BYTES	TRUE	The total number of remote bytes read from HDFS by this query. This is only for completed queries. Called 'hdfs_bytes_read_remote' in searches.
HDFS Remote Bytes Read Percentage (hdfs_bytes_read_remote_percentage)	NUMBER	TRUE	The percentage of all bytes read from HDFS by this query that were remote. This is only for completed queries. Called 'hdfs_bytes_read_remote_percentage' in searches.
HDFS Scanner Average Read Throughput (hdfs_scanner_average_bytes_read_per_second)	BYTES_PER_SECOND	TRUE	The average HDFS scanner read throughput for this query. This is computed by dividing the total bytes read from HDFS by the total time spent reading by all HDFS scanners. Called 'hdfs_scanner_average_bytes_read_per_second' in searches.
HDFS Short Circuit Bytes Read (hdfs_bytes_read_short_circuit)	BYTES	TRUE	The total number of bytes read from HDFS by this query that used short-circuit reads. This is only for completed queries. Called 'hdfs_bytes_read_short_circuit' in searches.
HDFS Short Circuit Bytes Read Percentage (hdfs_bytes_read_short_circuit_percentage)	NUMBER	TRUE	The percentage of all bytes read from HDFS by this query that used short-circuit reads. This is only for completed queries. Called 'hdfs_bytes_read_short_cicuit_percentage' in searches.
Impala Version (impala_version)	STRING	TRUE	The version of the Impala Daemon coordinating this query. Called 'impala_version' in searches.
Memory Accrual (memory_accrual)	BYTE_SECONDS	TRUE	The total accrued memory usage by the query. This is computed by multiplying the average aggregate memory usage of the query by the query's duration. Called 'memory_accrual' in searches.
Memory Spilled (memory_spilled)	BYTES	TRUE	Amount of memory spilled to disk. Called 'memory_spilled' in searches.
Network Address (network_address)	STRING	TRUE	The network address that issued this query. Called 'network_address' in searches.
Node with Peak Memory Usage	STRING	TRUE	The node with the highest peak memory usage for this query. See Per Node Peak Memory Usage for the

Display Name (Attribute Name)	Type	Supports Filtering?	Description
(memory_per_node_peak_node)			actual peak value. Called 'memory_per_node_peak_node' in searches.
Out of Memory (oom)	BOOLEAN	TRUE	Whether the query ran out of memory. Called 'oom' in searches.
Per Node Peak Memory Usage (memory_per_node_peak)	BYTES	TRUE	The highest amount of memory allocated by any single node that participated in this query. See Node with Peak Memory Usage for the name of the peak node. Called 'memory_per_node_peak' in searches.
Planning Wait Time (planning_wait_time)	MILLISECONDS	TRUE	The total amount of time the query spent waiting for planning to complete. Called 'planning_wait_time' in searches.
Planning Wait Time Percentage (planning_wait_time_percentage)	NUMBER	TRUE	The total amount of time the query spent waiting for planning to complete divided by the query duration. Called 'planning_wait_time_percentage' in searches.
Pool (pool)	STRING	TRUE	The name of the resource pool in which this query executed. Called 'pool' in searches. If YARN is in use, this corresponds to a YARN pool. Within YARN, a pool is referred to as a queue.
Query ID (query_id)	STRING	FALSE	The id of this query. Called 'query_id' in searches.
Query State (query_state)	STRING	TRUE	The current state of the query (running, finished, and so on). Called 'query_state' in searches.
Query Status (query_status)	STRING	TRUE	The status of the query. If the query hasn't failed the status will be 'OK', otherwise it will provide more information on the cause of the failure. Called 'query_status' in searches.
Query Type (query_type)	STRING	TRUE	The type of the query's SQL statement (DML, DDL, Query). Called 'query_type' in searches.
Resource Reservation Wait Time (resources_reserved_wait_time)	MILLISECONDS	TRUE	The total amount of time the query spent waiting for pool resources to become available . Called 'resources_reserved_wait_time' in searches.
Resource Reservation Wait Time Percentage (resources_reserved_wait_time_percentage)	NUMBER	TRUE	The total amount of time the query spent waiting for pool resources to become available divided by the query duration. Called 'resources_reserved_wait_time_percentage' in searches.
Rows Inserted (rows_inserted)	NUMBER	TRUE	The number of rows inserted by the query. Called 'rows_inserted' in searches.
Rows Produced (rows_produced)	NUMBER	TRUE	The number of rows produced by the query. Called 'rows_produced' in searches.

Display Name (Attribute Name)	Type	Supports Filtering?	Description
Service Name (service_name)	STRING	FALSE	The name of the Impala service. Called 'service_name' in searches.
Session ID (session_id)	STRING	TRUE	The ID of the session that issued this query. Called 'session_id' in searches.
Session Type (session_type)	STRING	TRUE	The type of the session that issued this query. Called 'session_type' in searches.
Statement (statement)	STRING	FALSE	The query's SQL statement. Called 'statement' in searches.
Statistics Missing (stats_missing)	BOOLEAN	TRUE	Whether the query was flagged with missing table or column statistics warning during the planning process. Called 'stats_missing' in searches.
Threads: CPU Time (thread_cpu_time)	MILLISECONDS	TRUE	The sum of the CPU time used by all threads of the query. Called 'thread_cpu_time' in searches.
Threads: CPU Time Percentage (thread_cpu_time_percentage)	NUMBER	TRUE	The sum of the CPU time used by all threads of the query divided by the total thread time. Called 'thread_cpu_time_percentage' in searches.
Threads: Network Receive Wait Time (thread_network_receive_wait_time)	MILLISECONDS	TRUE	The sum of the time spent waiting to receive data over the network by all threads of the query. A query will almost always have some threads waiting to receive data from other nodes in the query's execution tree. Unlike other wait times, network receive wait time does not usually indicate an opportunity for improving a query's performance. Called 'thread_network_receive_wait_time' in searches.
Threads: Network Receive Wait Time Percentage (thread_network_receive_wait_time_percentage)	NUMBER	TRUE	The sum of the time spent waiting to receive data over the network by all threads of the query divided by the total thread time. A query will almost always have some threads waiting to receive data from other nodes in the query's execution tree. Unlike other wait times, network receive wait time does not usually indicate an opportunity for improving a query's performance. Called 'thread_network_receive_wait_time_percentage' in searches.
Threads: Network Send Wait Time (thread_network_send_wait_time)	MILLISECONDS	TRUE	The sum of the time spent waiting to send data over the network by all threads of the query. Called 'thread_network_send_wait_time' in searches.
Threads: Network Send Wait Time Percentage (thread_network_send_wait_time_percentage)	NUMBER	TRUE	The sum of the time spent waiting to send data over the network by all threads of the query divided by the total thread time. Called 'thread_network_send_wait_time_percentage' in searches.

Display Name (Attribute Name)	Type	Supports Filtering?	Description
Threads: Storage Wait Time (thread_storage_wait_time)	MILLISECONDS	TRUE	The sum of the time spent waiting for storage by all threads of the query. Called 'thread_storage_wait_time' in searches.
Threads: Storage Wait Time Percentage (thread_storage_wait_time_percentage)	NUMBER	TRUE	The sum of the time spent waiting for storage by all threads of the query divided by the total thread time. Called 'thread_storage_wait_time_percentage' in searches.
Threads: Total Time (thread_total_time)	MILLISECONDS	TRUE	The sum of thread CPU, storage wait and network wait times used by all threads of the query. Called 'thread_total_time' in searches.
User (user)	STRING	TRUE	The effective user for the query. This is the delegated user if delegation is in use. Otherwise, this is the connected user. Called 'user' in searches.
Work CPU Time (cm_cpu_milliseconds)	MILLISECONDS	TRUE	Attribute measuring the sum of CPU time used by all threads of the query, in milliseconds. Called 'work_cpu_time' in searches. For Impala queries, CPU time is calculated based on the 'TotalCpuTime' metric. For YARN MapReduce applications, this is calculated from the 'cpu_milliseconds' metric.

Examples

Consider the following filter expressions: `user = "root"`, `rowsProduced > 0`, `fileFormats RLIKE ".TEXT.*"`, and `executing = true`. In the examples:

- The filter attributes are `user`, `rowsProduced`, `fileFormats`, and `executing`.
- The operators are `=`, `>`, and `RLIKE`.
- The filter values are `root`, `0`, `.TEXT.*`, and `true`.

Query Details

The **Query Details** page contains the low-level details of how a SQL query is processed through Impala. The initial information on the page can help you tune the performance of some kinds of queries, primarily those involving joins. The more detailed information on the page is primarily for troubleshooting with the assistance of Cloudera Support; you might be asked to attach the contents of the page to a trouble ticket.

The **Query Details** page displays the following information that is also available in Query Profile:

- Query Plan
- Query Info
- Query Timeline on page 69
- Planner Timeline on page 69
- Query Fragments

To download the contents of the query details, select one of the following:

- **Download Profile...** or **Download Profile...** > **Download Text Profile...** - to download a text version of the query detail.
- **Download Profile...** > **Download Thrift Encoded Profile...** - to download a binary version of the query detail.

Query Plan

The Query Plan section can help you diagnose and tune performance issues with queries. This information is especially useful to understand performance issues with join queries, such as inefficient order of tables in the SQL statement, lack of table and column statistics, and the need for query hints to specify a more efficient join mechanism. You can also learn valuable information about how queries are processed for partitioned tables.

The information in this section corresponds to the output of the `EXPLAIN` statement for the Impala query. Each fragment shown in the query plan corresponds to a processing step that is performed by the central coordinator host or distributed across the hosts in the cluster.

Query Timeline

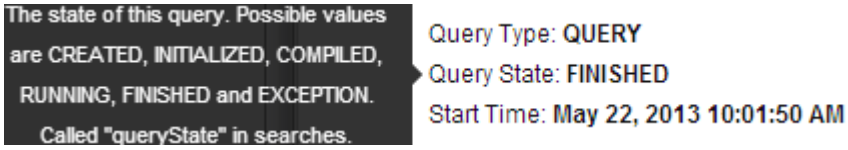
The Query Timeline section reports statistics about the execution time for phases of the query.

Planner Timeline

The Planner Timeline reports statistics about the execution time for phases of the query planner.

Query Info

The Query Info section reports the attributes of the query, start and end time, duration, and statistics about HDFS access. You can hover over an attribute for information about the attribute name and supported values (for enumerated values). For example:



Query Fragments

The Query Fragments section reports detailed low-level statistics for each query plan fragment, involving physical aspects such as CPU utilization, disk I/O, and network traffic. This is the primary information that Cloudera Support might use to help troubleshoot performance issues and diagnose bugs. The details for each fragment display on separate tabs.

Monitoring YARN Applications

The YARN Applications page displays information about the YARN jobs that are running and have run in your cluster. You can filter the jobs by time period and by specifying simple filtering expressions.

Viewing Jobs

1. Do one of the following:
   - Select **Clusters** > ***Cluster name*** > ***YARN service name*** **Applications**.
   - On the **Home** > **Status** tab, select ***YARN service name*** and click the **Applications** tab.

The YARN jobs run during the selected time range display in the Results Tab on page 70. The results displayed can be filtered by creating filter expressions.

You can also perform the following actions on this page:

Table 7: Viewing Jobs Actions

Action	Description
Filter jobs that display.	Create filter expressions manually, select preconfigured filters, or use the **Workload Summary** section to build a query interactively. See Filtering Jobs on page 71.
Select additional attributes for display.	Click **Select Attributes**. Selected attributes also display as available filters in the **Workload Summary** section. To display information about attributes, hover over a field label. See Filter Attributes on page 72 Only attributes that support filtering appear in the **Workload Summary** section. See the Table 8: Attributes on page 73 table.
View a histogram of the attribute values.	Click the icon to the right of each attribute displayed in the **Workload Summary** section.
Display charts based on the filter expression and selected attributes.	Click the **Charts** tab.
Send a YARN application diagnostic bundle to Cloudera support.	Click **Collect Diagnostics Data**. See Sending Diagnostic Data to Cloudera for YARN Applications on page 82.
Export a JSON file with the query results that you can use for further analysis.	Click **Export**.

Configuring YARN Application Monitoring

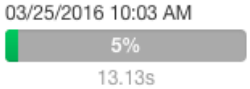
You can configure the visibility of the YARN application monitoring results.

For information on how to configure whether admin and non-admin users can view all applications, only that user's applications, or no applications, see Configuring Application Visibility on page 17.

Results Tab

Jobs are ordered with the most recent at the top. Each job has summary and detail information. A job summary includes start and end timestamps, query (if the job is part of a Hive query) name, pool, job type, job ID, and user. For example:

A running job displays a progress bar under the start timestamp:



Use the Actions drop-down menu  to the right of each job listing to do the following. (Not all options display, depending on the type of job.)

- Application Details – Open a details page for the job.
- Collect Diagnostic Data – Send a YARN application diagnostic bundle to Cloudera support.

- Similar MR2 Jobs – Display a list of similar MapReduce 2 jobs.
- User's YARN Applications – Display a list of all jobs run by the user of the current job.
- View on JobHistory Server – View the application in the YARN JobHistory Server.
- Kill (running jobs only) – Kill a job (administrators only). Killing a job creates an audit event. When you kill a job, 

replaces the progress bar.
- Applications in Hive Query (Hive jobs only)
- Applications in Oozie Workflow (Oozie jobs only)
- Applications in Pig Script (Pig jobs only)

Filtering Jobs

You filter jobs by selecting a time range and specifying a filter expression in the search box.

You can use the Time Range Selector or a duration link (30m 1h 2h 6h 12h 1d 7d 30d) to set the time range. (See Time Line on page 7 for details).

Filter Expressions

Filter expressions specify which entries should display when you run the filter. The simplest expression consists of three components:

- **Attribute** - Query language name of the attribute.
- **Operator** - Type of comparison between the attribute and the attribute value. Cloudera Manager supports the standard comparator operators =, `!=`, >, <, >=, <=, and `RLIKE`. (RLIKE performs regular expression matching as specified in the Java Pattern class documentation.) Numeric values can be compared with all operators. String values can be compared with =, `!=`, and `RLIKE`. Boolean values can be compared with = and `!=`.
- **Value** - The value of the attribute. The value depends on the type of the attribute. For a Boolean value, specify either `true` or `false`. When specifying a string value, enclose the value in double quotes.

You create compound filter expressions using the AND and OR operators. When more than one operator is used in an expression, AND is evaluated first, then OR. To change the order of evaluation, enclose subexpressions in parentheses.

Compound Expressions

To find all the jobs issued by the root user that ran for longer than ten seconds, use the expression:

```
user = "root" AND application_duration >= 100000.0
```

To find all the jobs that had more than 200 maps issued by users Jack or Jill, use the expression:

```
maps_completed >= 200.0 AND (user = "Jack" OR user = "Jill")
```

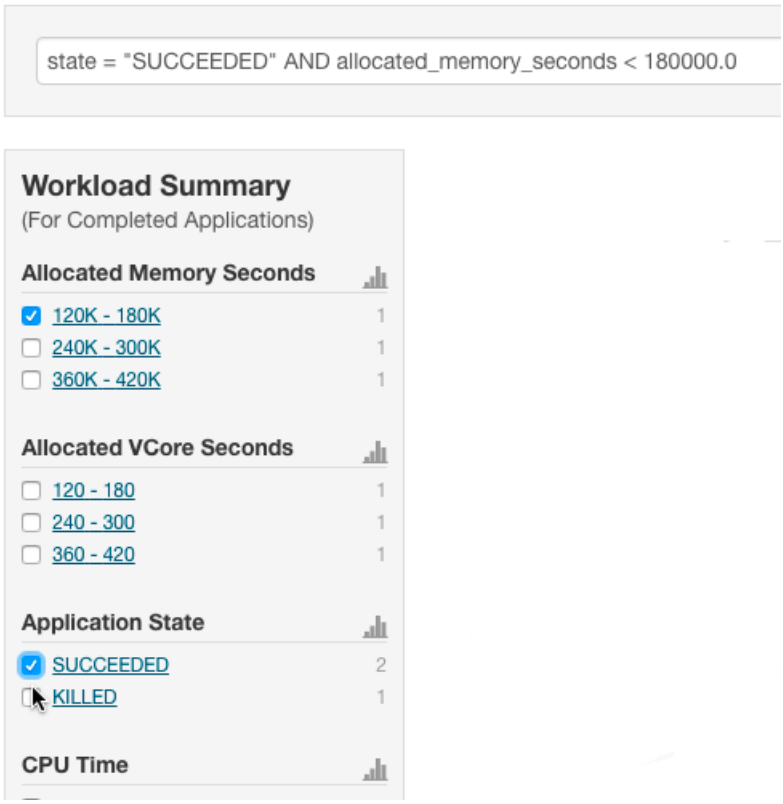
Choosing and Running a Filter

1. Do one of the following:
   - **Select a Suggested or Recently Run Filter**

     Click the 
     to the right of the **Search** button to display a list of sample and recently run filters, and select a filter. The filter text displays in the text box.
   - **Construct a Filter from the Workload Summary Attributes**

     Optionally, click **Select Attributes** to display a dialog box where you can chose attributes to display in the **Workload Summary** section. Select the checkbox next to one or more attributes and click **Close**. Only attributes that support filtering appear in the **Workload Summary** section. See the Table 8: Attributes on page 73 table.

The attributes display in the **Workload Summary** section along with values or ranges of values that you can filter on. The values and ranges display as links with checkboxes. Select one or more checkboxes to add the range or value to the query. Click a link to run a query on that value or range. For example:



- **Type a Filter**
  1. Start typing or press **Spacebar** in the text box. As you type, filter attributes matching the typed letter display. If you press **Spacebar**, standard filter attributes display. These suggestions are part of typeahead, which helps build valid queries. For information about the attribute name and supported values for each field, hover over the field in an existing query.
  2. Select an attribute and press **Enter**.
  3. Press **Spacebar** to display a drop-down list of operators.
  4. Select an operator and press **Enter**.
  5. Specify an attribute value in one of the following ways:
     - For attribute values that support typeahead, press **Spacebar** to display a drop-down list of values and press **Enter**.
     - Type a value.

2. Click in the text box and press **Enter** or click **Search**. The list displays the results that match the specified filter. If the histograms are showing, they are redrawn to show only the values for the selected filter. The filter is added to the Recently Run list.

Filter Attributes

Filter attributes, their names as they are displayed in Cloudera Manager, their types, and descriptions, are enumerated below.

Note: Only attributes where the **Supports Filtering?** column value is TRUE appear in the **Workload Summary** section.

Table 8: Attributes

Display Name (Attribute Name)	Type	Supports Filtering?	Description
Allocated Memory (allocated_mb)	NUMBER	FALSE	The sum of memory in MB allocated to the application's running containers. Called 'allocated_mb' in searches.
Allocated Memory Seconds (allocated_memory_seconds)	NUMBER	TRUE	The amount of memory the application has allocated (megabyte-seconds). Called 'allocated_memory_seconds' in searches.
Allocated VCores (allocated_vcores)	NUMBER	FALSE	The sum of virtual cores allocated to the application's running containers. Called 'allocated_vcores' in searches.
Allocated VCore Seconds (allocated_vcore_seconds)	NUMBER	TRUE	The amount of CPU resources the application has allocated (virtual core-seconds). Called 'allocated_vcore_seconds' in searches.
Application ID (application_id)	STRING	FALSE	The ID of the YARN application. Called 'application_id' in searches.
Application State (state)	STRING	TRUE	The state of this YARN application. This reflects the ResourceManager state while the application is running and the JobHistory Server state after the application has completed. Called 'state' in searches.
Application Tags (application_tags)	STRING	FALSE	A list of tags for the application. Called 'application_tags' in searches.
Application Type (application_type)	STRING	TRUE	The type of the YARN application. Called 'application_type' in searches.
Bytes Read (bytes_read)	BYTES	TRUE	Bytes read. Called 'bytes_read' in searches.
Bytes Written (bytes_written)	BYTES	TRUE	Bytes written. Called 'bytes_written' in searches.
Combine Input Records (combine_input_records)	NUMBER	TRUE	Combine input records. Called 'combine_input_records' in searches.
Combine Output Records (combine_output_records)	NUMBER	TRUE	Combine output records. Called 'combine_output_records' in searches.
Committed Heap (committed_heap_bytes)	BYTES	TRUE	Total committed heap usage. Called 'committed_heap_bytes' in searches.
Completed Maps and Reduces (tasks_completed)	NUMBER	TRUE	The number of completed map and reduce tasks in this MapReduce job. Called 'tasks_completed' in searches. Available only for running jobs.

Display Name (Attribute Name)	Type	Supports Filtering?	Description
CPU Allocation (vcores_millis)	NUMBER	TRUE	CPU allocation. This is the sum of 'vcores_millis_maps' and 'vcores_millis_reduces'. Called 'vcores_millis' in searches.
CPU Time (cpu_milliseconds)	MILLISECONDS	TRUE	CPU time. Called 'cpu_milliseconds' in searches.
Data Local Maps (data_local_maps)	NUMBER	TRUE	Data local maps. Called 'data_local_maps' in searches.
Data Local Maps Percentage (data_local_maps_percentage)	NUMBER	TRUE	The number of data local maps as a percentage of the total number of maps. Called 'data_local_maps_percentage' in searches.
Diagnostics (diagnostics)	STRING	FALSE	Diagnostic information on the YARN application. If the diagnostic information is long, this may only contain the beginning of the information. Called 'diagnostics' in searches.
Duration (application_duration)	MILLISECONDS	TRUE	How long YARN took to run this application. Called 'application_duration' in searches.
Executing (executing)	BOOLEAN	FALSE	Whether the YARN application is currently running. Called 'executing' in searches.
Failed Map and Reduce Attempts (failed_tasks_attempts)	NUMBER	TRUE	The number of failed map and reduce attempts for this MapReduce job. Called 'failed_tasks_attempts' in searches. Available only for failed jobs.
Failed Map Attempts (failed_map_attempts)	NUMBER	TRUE	The number of failed map attempts for this MapReduce job. Called 'failed_map_attempts' in searches. Available only for running jobs.
Failed Maps (num_failed_maps)	NUMBER	TRUE	Failed maps. Called 'num_failed_maps' in searches.
Failed Reduce Attempts (failed_reduce_attempts)	NUMBER	TRUE	The number of failed reduce attempts for this MapReduce job. Called 'failed_reduce_attempts' in searches. Available only for running jobs.
Failed Reduces (num_failed_reduces)	NUMBER	TRUE	Failed reduces. Called 'num_failed_reduces' in searches.
Failed Shuffles (failed_shuffle)	NUMBER	TRUE	Failed shuffles. Called 'failed_shuffle' in searches.
Failed Tasks (num_failed_tasks)	NUMBER	TRUE	The total number of failed tasks. This is the sum of 'num_failed_maps' and 'num_failed_reduces'. Called 'num_failed_tasks' in searches.
Fallow Map Slots Time (fallow_slots_millis_maps)	MILLISECONDS	TRUE	Fallow map slots time. Called 'fallow_slots_millis_maps' in searches.

Display Name (Attribute Name)	Type	Supports Filtering?	Description
Fallow Reduce Slots Time (fallow_slots_millis_reduces)	MILLISECONDS	TRUE	Fallow reduce slots time. Called 'fallow_slots_millis_reduces' in searches.
Fallow Slots Time (fallow_slots_millis)	MILLISECONDS	TRUE	Total fallow slots time. This is the sum of 'fallow_slots_millis_maps' and 'fallow_slots_millis_reduces'. Called 'fallow_slots_millis' in searches.
File Bytes Read (file_bytes_read)	BYTES	TRUE	File bytes read. Called 'file_bytes_read' in searches.
File Bytes Written (file_bytes_written)	BYTES	TRUE	File bytes written. Called 'file_bytes_written' in searches.
File Large Read Operations (file_large_read_ops)	NUMBER	TRUE	File large read operations. Called 'file_large_read_ops' in searches.
File Read Operations (file_read_ops)	NUMBER	TRUE	File read operations. Called 'file_read_ops' in searches.
File Write Operations (file_write_ops)	NUMBER	TRUE	File write operations. Called 'file_large_write_ops' in searches.
Garbage Collection Time (gc_time_millis)	MILLISECONDS	TRUE	Garbage collection time. Called 'gc_time_millis' in searches.
HDFS Bytes Read (hdfs_bytes_read)	BYTES	TRUE	HDFS bytes read. Called 'hdfs_bytes_read' in searches.
HDFS Bytes Written (hdfs_bytes_written)	BYTES	TRUE	HDFS bytes written. Called 'hdfs_bytes_written' in searches.
HDFS Large Read Operations (hdfs_large_read_ops)	NUMBER	TRUE	HDFS large read operations. Called 'hdfs_large_read_ops' in searches.
HDFS Read Operations (hdfs_read_ops)	NUMBER	TRUE	HDFS read operations. Called 'hdfs_read_ops' in searches.
HDFS Write Operations (hdfs_write_ops)	NUMBER	TRUE	HDFS write operations. Called 'hdfs_write_ops' in searches.
Hive Query ID (hive_query_id)	STRING	FALSE	If this MapReduce job ran as a part of a Hive query, this field contains the ID of the Hive query. Called 'hive_query_id' in searches.
Hive Query String (hive_query_string)	STRING	TRUE	If this MapReduce job ran as a part of a Hive query, this field contains the string of the query. Called 'hive_query_string' in searches.

Display Name (Attribute Name)	Type	Supports Filtering?	Description
Hive Sentry Subject Name (hive_sentry_subject_name)	STRING	TRUE	If this MapReduce job ran as a part of a Hive query on a secured cluster using impersonation, this field contains the name of the user that initiated the query. Called 'hive_sentry_subject_name' in searches.
Input Directory (input_dir)	STRING	TRUE	The input directory for this MapReduce job. Called 'input_dir' in searches.
Input Split Bytes (split_raw_bytes)	BYTES	TRUE	Input split bytes. Called 'split_raw_bytes' in searches.
Killed Map and Reduce Attempts (killed_tasks_attempts)	NUMBER	TRUE	The number of map and reduce attempts that were killed by user(s) for this MapReduce job. Called 'killed_tasks_attempts' in searches. Available only for killed jobs.
Killed Map Attempts (killed_map_attempts)	NUMBER	TRUE	The number of map attempts killed by user(s) for this MapReduce job. Called 'killed_map_attempts' in searches. Available only for running jobs.
Killed Reduce Attempts (killed_reduce_attempts)	NUMBER	TRUE	The number of reduce attempts killed by user(s) for this MapReduce job. Called 'killed_reduce_attempts' in searches. Available only for running jobs.
Launched Map Tasks (total_launched_maps)	NUMBER	TRUE	Launched map tasks. Called 'total_launched_maps' in searches.
Launched Reduce Tasks (total_launched_reduces)	NUMBER	TRUE	Launched reduce tasks. Called 'total_launched_reduces' in searches.
Map and Reduce Attempts in NEW State (new_tasks_attempts)	NUMBER	TRUE	The number of map and reduce attempts in NEW state for this MapReduce job. Called 'new_tasks_attempts' in searches. Available only for running jobs.
Map Attempts in NEW State (new_map_attempts)	NUMBER	TRUE	The number of map attempts in NEW state for this MapReduce job. Called 'new_map_attempts' in searches. Available only for running jobs.
Map Class (mapper_class)	STRING	TRUE	The class used by the map tasks in this MapReduce job. Called 'mapper_class' in searches. You can search for the mapper class using the class name alone, for example 'QuasiMonteCarlo$QmcMapper', or the fully qualified classname, for example, 'org.apache.hadoop.examples.QuasiMonteCarlo$QmcMapper'.
Map CPU Allocation (vcores_millis_maps)	NUMBER	TRUE	Map CPU allocation. Called 'vcores_millis_maps' in searches.
Map Input Records (map_input_records)	NUMBER	TRUE	Map input records. Called 'map_input_records' in searches.

Display Name (Attribute Name)	Type	Supports Filtering?	Description
Map Memory Allocation (mb_millis_maps)	NUMBER	TRUE	Map memory allocation. Called 'mb_millis_maps' in searches.
Map Output Bytes (map_output_bytes)	BYTES	TRUE	Map output bytes. Called 'map_output_bytes' in searches.
Map Output Materialized Bytes (map_output_materialized_bytes)	BYTES	TRUE	Map output materialized bytes. Called 'map_output_materialized_bytes' in searches.
Map Output Records (map_output_records)	NUMBER	TRUE	Map output records. Called 'map_output_records' in searches.
Map Progress (map_progress)	NUMBER	TRUE	The percentage of maps completed for this MapReduce job. Called 'map_progress' in searches. Available only for running jobs.
Maps Completed (maps_completed)	NUMBER	TRUE	The number of map tasks completed as a part of this MapReduce job. Called 'maps_completed' in searches.
Map Slots Time (slots_millis_maps)	MILLISECONDS	TRUE	Total time spent by all maps in occupied slots. Called 'slots_millis_maps' in searches.
Maps Pending (maps_pending)	NUMBER	TRUE	The number of maps waiting to be run for this MapReduce job. Called 'maps_pending' in searches. Available only for running jobs.
Maps Running (maps_running)	NUMBER	TRUE	The number of maps currently running for this MapReduce job. Called 'maps_running' in searches. Available only for running jobs.
Maps Total (maps_total)	NUMBER	TRUE	The number of Map tasks in this MapReduce job. Called 'maps_total' in searches.
Memory Allocation (mb_millis)	NUMBER	TRUE	Total memory allocation. This is the sum of 'mb_millis_maps' and 'mb_millis_reduces'. Called 'mb_millis' in searches.
Merged Map Outputs (merged_map_outputs)	NUMBER	TRUE	Merged map outputs. Called 'merged_map_outputs' in searches.
Name (name)	STRING	TRUE	Name of the YARN application. Called 'name' in searches.
Oozie Workflow ID (oozie_id)	STRING	FALSE	If this MapReduce job ran as a part of an Oozie workflow, this field contains the ID of the Oozie workflow. Called 'oozie_id' in searches.
Other Local Maps (other_local_maps)	NUMBER	TRUE	Other local maps. Called 'other_local_maps' in searches.

Display Name (Attribute Name)	Type	Supports Filtering?	Description
Other Local Maps Percentage (other_local_maps_percentage)	NUMBER	TRUE	The number of other local maps as a percentage of the total number of maps. Called 'other_local_maps_percentage' in searches.
Output Directory (output_dir)	STRING	TRUE	The output directory for this MapReduce job. Called 'output_dir' in searches.
Pending Maps and Reduces (tasks_pending)	NUMBER	TRUE	The number of maps and reduces waiting to be run for this MapReduce job. Called 'tasks_pending' in searches. Available only for running jobs.
Physical Memory (physical_memory_bytes)	BYTES	TRUE	Physical memory. Called 'physical_memory_bytes' in searches.
Pig Script ID (pig_id)	STRING	FALSE	If this MapReduce job ran as a part of a Pig script, this field contains the ID of the Pig script. Called 'pig_id' in searches.
Pool (pool)	STRING	TRUE	The name of the resource pool in which this application ran. Called 'pool' in searches. Within YARN, a pool is referred to as a queue.
Progress (progress)	NUMBER	TRUE	The progress reported by the application. Called 'progress' in searches.
Rack Local Maps (rack_local_maps)	NUMBER	TRUE	Rack local maps. Called 'rack_local_maps' in searches.
Rack Local Maps Percentage (rack_local_maps_percentage)	NUMBER	TRUE	The number of rack local maps as a percentage of the total number of maps. Called 'rack_local_maps_percentage' in searches.
Reduce Attempts in NEW State (new_reduce_attempts)	NUMBER	TRUE	The number of reduce attempts in NEW state for this MapReduce job. Called 'new_reduce_attempts' in searches. Available only for running jobs.
Reduce Class (reducer_class)	STRING	TRUE	The class used by the reduce tasks in this MapReduce job. Called 'reducer_class' in searches. You can search for the reducer class using the class name alone, for example 'QuasiMonteCarlo$QmcReducer', or fully qualified classname, for example, 'org.apache.hadoop.examples.QuasiMonteCarlo$QmcReducer'.
Reduce CPU Allocation (vcores_millis_reduces)	NUMBER	TRUE	Reduce CPU allocation. Called 'vcores_millis_reduces' in searches.
Reduce Input Groups (reduce_input_groups)	NUMBER	TRUE	Reduce input groups. Called 'reduce_input_groups' in searches.
Reduce Input Records (reduce_input_records)	NUMBER	TRUE	Reduce input records. Called 'reduce_input_records' in searches.

Display Name (Attribute Name)	Type	Supports Filtering?	Description
Reduce Memory Allocation (mb_millis_reduces)	NUMBER	TRUE	Reduce memory allocation. Called 'mb_millis_reduces' in searches.
Reduce Output Records (reduce_output_records)	NUMBER	TRUE	Reduce output records. Called 'reduce_output_records' in searches.
Reduce Progress (reduce_progress)	NUMBER	TRUE	The percentage of reduces completed for this MapReduce job. Called 'reduce_progress' in searches. Available only for running jobs.
Reduces Completed (reduces_completed)	NUMBER	TRUE	The number of reduce tasks completed as a part of this MapReduce job. Called 'reduces_completed' in searches.
Reduce Shuffle Bytes (reduce_shuffle_bytes)	BYTES	TRUE	Reduce shuffle bytes. Called 'reduce_shuffle_bytes' in searches.
Reduce Slots Time (slots_millis_reduces)	MILLISECONDS	TRUE	Total time spent by all reduces in occupied slots. Called 'slots_millis_reduces' in searches.
Reduces Pending (reduces_pending)	NUMBER	TRUE	The number of reduces waiting to be run for this MapReduce job. Called 'reduces_pending' in searches. Available only for running jobs.
Reduces Running (reduces_running)	NUMBER	TRUE	The number of reduces currently running for this MapReduce job. Called 'reduces_running' in searches. Available only for running jobs.
Reduces Total (reduces_total)	NUMBER	TRUE	The number of reduce tasks in this MapReduce job. Called 'reduces_total' in searches.
Running Containers (running_containers)	NUMBER	FALSE	The number of containers currently running for the application. Called 'running_containers' in searches.
Running Map and Reduce Attempts (running_tasks_attempts)	NUMBER	TRUE	The number of map and reduce attempts currently running for this MapReduce job. Called 'running_tasks_attempts' in searches. Available only for running jobs.
Running Map Attempts (running_map_attempts)	NUMBER	TRUE	The number of running map attempts for this MapReduce job. Called 'running_map_attempts' in searches. Available only for running jobs.
Running MapReduce Application Information Retrieval Duration. (running_application_info_retrieval_time)	NUMBER	TRUE	How long it took, in seconds, to retrieve information about the MapReduce application.
Running Maps and Reduces (tasks_running)	NUMBER	TRUE	The number of maps and reduces currently running for this MapReduce job. Called 'tasks_running' in searches. Available only for running jobs.

Display Name (Attribute Name)	Type	Supports Filtering?	Description
Running Reduce Attempts (running_reduce_attempts)	NUMBER	TRUE	The number of running reduce attempts for this MapReduce job. Called 'running_reduce_attempts' in searches. Available only for running jobs.
Service Name (service_name)	STRING	FALSE	The name of the YARN service. Called 'service_name' in searches.
Shuffle Bad ID Errors (shuffle_errors_bad_id)	NUMBER	TRUE	Shuffle bad ID errors. Called 'shuffle_errors_bad_id' in searches.
Shuffle Connection Errors (shuffle_errors_connection)	NUMBER	TRUE	Shuffle connection errors. Called 'shuffle_errors_connection' in searches.
Shuffled Maps (shuffled_maps)	NUMBER	TRUE	Shuffled maps. Called 'shuffled_maps' in searches.
Shuffle IO Errors (shuffle_errors_io)	NUMBER	TRUE	Shuffle IO errors. Called 'shuffle_errors_io' in searches.
Shuffle Wrong Length Errors (shuffle_errors_wrong_length)	NUMBER	TRUE	Shuffle wrong length errors. Called 'shuffle_errors_wrong_length' in searches.
Shuffle Wrong Map Errors (shuffle_errors_wrong_map)	NUMBER	TRUE	Shuffle wrong map errors. Called 'shuffle_errors_wrong_map' in searches.
Shuffle Wrong Reduce Errors (shuffle_errors_wrong_reduce)	NUMBER	TRUE	Shuffle wrong reduce errors. Called 'shuffle_errors_wrong_reduce' in searches.
Slots Time (slots_millis)	MILLISECONDS	TRUE	Total slots time. This is the sum of 'slots_millis_maps' and 'slots_millis_reduces'. Called 'slots_millis' in searches.
Spilled Records (spilled_records)	NUMBER	TRUE	Spilled Records. Called 'spilled_records' in searches.
Successful Map and Reduce Attempts (successful_tasks_attempts)	NUMBER	TRUE	The number of successful map and reduce attempts for this MapReduce job. Called 'successful_tasks_attempts' in searches. Available only for successful jobs.
Successful Map Attempts (successful_map_attempts)	NUMBER	TRUE	The number of successful map attempts for this MapReduce job. Called 'successful_map_attempts' in searches. Available only for running jobs.
Successful Reduce Attempts (successful_reduce_attempts)	NUMBER	TRUE	The number of successful reduce attempts for this MapReduce job. Called 'successful_reduce_attempts' in searches. Available only for running jobs.
Total Maps and Reduces Number (total_task_num)	NUMBER	TRUE	The number of map and reduce tasks in this MapReduce job. Called 'tasks_total' in searches. Available only for running jobs.

Display Name (Attribute Name)	Type	Supports Filtering?	Description
Total Tasks (total_launched_tasks)	NUMBER	TRUE	The total number of tasks. This is the sum of 'total_launched_maps' and 'total_launched_reduces'. Called 'total_launched_tasks' in searches.
Tracking Url (tracking_url)	STRING	FALSE	The MapReduce application tracking URL.
Uberized Job (uberized)	BOOLEAN	FALSE	Whether this MapReduce job is uberized - running completely in the ApplicationMaster. Called 'uberized' in searches. Available only for running jobs.
Unused Memory Seconds (unused_memory_seconds)	NUMBER	TRUE	The amount of memory the application has allocated but not used (megabyte-seconds). This metric is available only from CDH 5.7 onwards and is calculated hourly if container usage metric aggregation is enabled. Called 'unused_memory_seconds' in searches.
Unused VCore Seconds (unused_vcore_seconds)	NUMBER	TRUE	The amount of CPU resources the application has allocated but not used (virtual core-seconds). This metric is available only from CDH 5.7 onwards and is calculated hourly if container usage metric aggregation is enabled. Called 'unused_vcore_seconds' in searches.
Used Memory Max (used_memory_max)	NUMBER	TRUE	The maximum container memory usage for a YARN application. This metric is calculated hourly if container usage metric aggregation is enabled and a Cloudera Manager Container Usage Metrics Directory is specified. For information about how to enable metric aggregation and the Container Usage Metrics Directory, see Enabling the Cluster Utilization Report.
User (user)	STRING	TRUE	The user who ran the YARN application. Called 'user' in searches.
Virtual Memory (virtual_memory_bytes)	BYTES	TRUE	Virtual memory. Called 'virtual_memory_bytes' in searches.
Work CPU Time (cm_cpu_milliseconds)	MILLISECONDS	TRUE	Attribute measuring the sum of CPU time used by all threads of the query, in milliseconds. Called 'work_cpu_time' in searches. For Impala queries, CPU time is calculated based on the 'TotalCpuTime' metric. For YARN MapReduce applications, this is calculated from the 'cpu_milliseconds' metric.

Examples

Consider the following filter expressions: `user = "root"`, `rowsProduced > 0`, `fileFormats RLIKE ".TEXT.*"`, and `executing = true`. In the examples:

- The filter attributes are `user`, `rowsProduced`, `fileFormats`, and `executing`.
- The operators are =, >, and `RLIKE`.

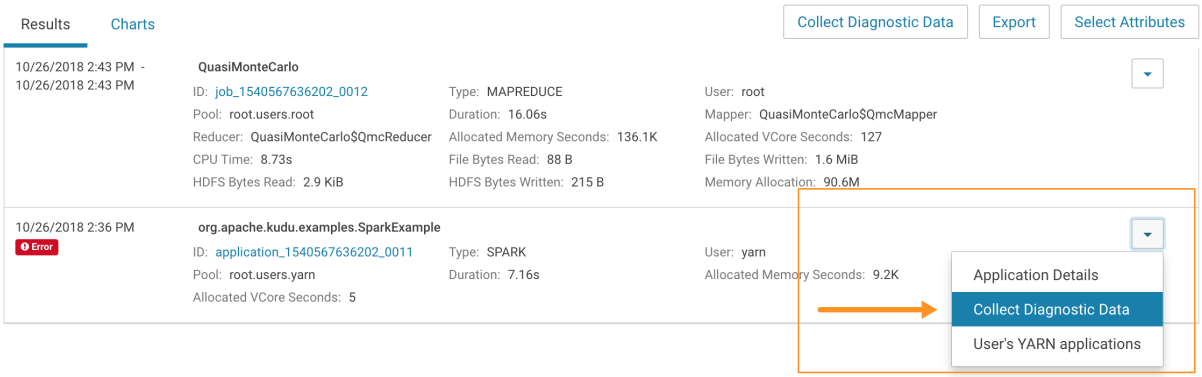
- The filter values are `root`, `0`, `.TEXT.*`, and `true`.

Sending Diagnostic Data to Cloudera for YARN Applications

Minimum Required Role: **Cluster Administrator** (also provided by **Full Administrator**)

You can send diagnostic data collected from YARN applications, including metadata, configurations, and log data, to Cloudera Support for analysis. Include a support ticket number if one exists to enable Cloudera Support to address the issue more quickly and efficiently. To send YARN application diagnostic data, perform the following steps:

1. From the YARN page in Cloudera Manager, click the **Applications** menu.
2. Collect diagnostics data. There are two ways to do this:
   - To collect data from all applications that are visible in the list, click the top **Collect Diagnostics Data** button on the upper right, above the list of YARN applications.
   - To collect data from only one specific application, click the down arrow on the right-hand end of the row of the application and select **Collect Diagnostics Data**.



3. In the **Send YARN Applications Diagnostic Data** dialog box, provide the following information:
   - If applicable, the Cloudera Support ticket number of the issue being experienced on the cluster.
   - Optionally, add a comment to help the support team understand the issue.
4. Click the checkbox **Send Diagnostic Data to Cloudera**.
5. Click the button **Collect and Send Diagnostic Data**.

> **Note:** Passwords from configuration will not be retrieved.

Monitoring Spark Applications

To obtain information about Spark application behavior you can consult cluster manager logs and the Spark web application UI. These two methods provide complementary information. Logs enable you to see fine grained events in the lifecycle of an application. The web UI provides both a broad overview of the various aspects of Spark application behavior and fine grained metrics. This section provides an overview of both methods.

For further information on Spark monitoring, see Monitoring and Instrumentation.

Viewing and Debugging Spark Applications Using Logs

To see overview information about all running Spark applications, depending on which cluster manager you are using, do one of the following:

- YARN - Go to the YARN applications page in the Cloudera Manager Admin Console.

  To debug Spark applications running on YARN, view the logs for the NodeManager role:

  1. Open the log event viewer.

2. Filter the event stream to choose a time window, log level, and display the NodeManager source.
3. For any event, click **View Log File** to view the entire log file.

- Spark Standalone - Go to the Spark Master UI, by default at `http://spark_master:18080`. The master and each worker show cluster and job statistics. In addition, detailed log output for each job is also written to the work directory of each worker.

Visualizing Spark Applications Using the Web Application UI

Every Spark application launches a web application UI that displays useful information about the application:

- An event timeline that displays the relative ordering and interleaving of application events. The timeline view is available on three levels: across all jobs, within one job, and within one stage. The timeline also shows executor allocation and deallocation.
- A list of stages and tasks.
- The execution directed acyclic graph (DAG) for each job.
- A summary of RDD sizes and memory usage.
- Environment - runtime information, property settings, library paths.
- Information about Spark SQL jobs.

The web UI is available in different ways depending on whether the application is running or has completed.

Accessing the Web UI of a Running Spark Application

To access the web application UI of a running Spark application, open `http://spark_driver_host:4040` in a web browser. If multiple applications are running on the same host, the web application binds to successive ports beginning with 4040 (4041, 4042, and so on). The web application is available only for the *duration of the application*.

Accessing the Web UI of a Completed Spark Application

To access the web application UI of a completed Spark application, do the following:

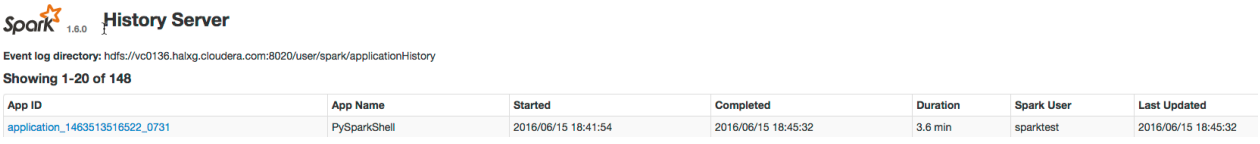
1. Open the Spark History Server UI in one of the following ways:
   - Open the URL `http://spark_history_server_host:18088`.
   - Open the UI in the Cloudera Manager Admin Console:
     1. Go to the Spark service.
     2. Click the **History Server Web UI** link.

   The History Server displays a list of completed applications.
2. In the list of applications, click an **App ID** link. The application UI displays.

> **Note:** In CDH 5.10 and higher, and in CDS 2 Powered by Apache Spark, the **Storage** tab of the Spark History Server is always blank. To see storage information while an application is running, use the web UI of the application as described in the previous section. After the application is finished, storage information is not available.

Example Spark Application Web Application

Consider a job consisting of a set of transformation to join data from an accounts dataset with a weblogs dataset in order to determine the total number of web hits for every account and then an action write the result to HDFS. In this example, the write is performed twice, resulting in two jobs. To view the application UI, in the History Server click the link in the App ID column:

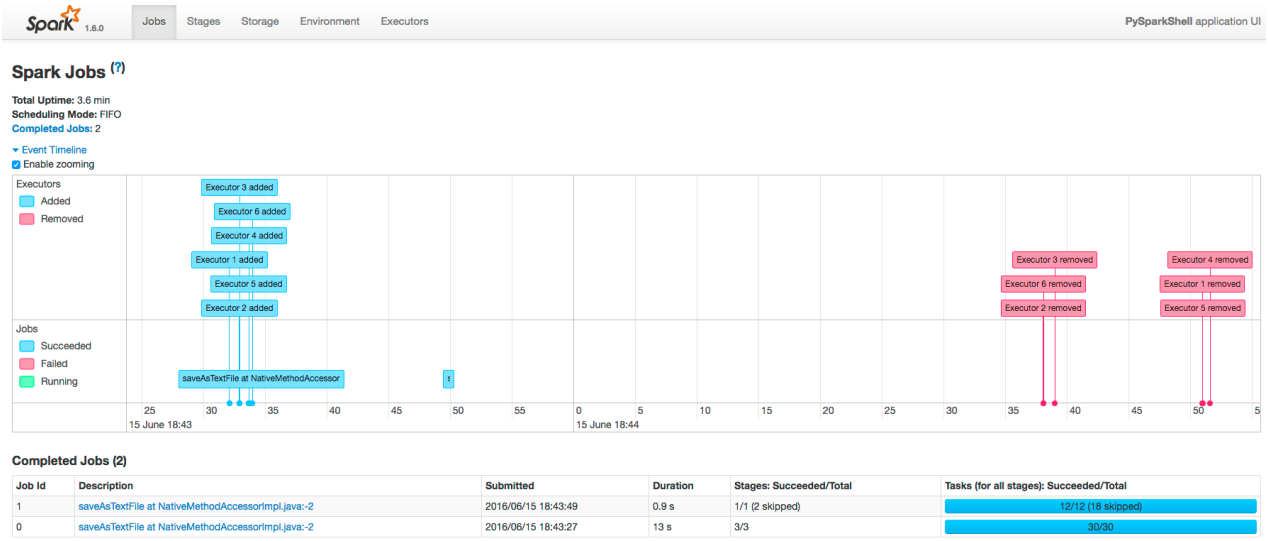

Spark 1.6.0 History Server

Event log directory: hdfs://vc0136.halxg.cloudera.com:8020/user/spark/applicationHistory

Showing 1-20 of 148

App ID	App Name	Started	Completed	Duration	Spark User	Last Updated
application_1463513516522_0731	PySparkShell	2016/06/15 18:41:54	2016/06/15 18:45:32	3.6 min	sparktest	2016/06/15 18:45:32

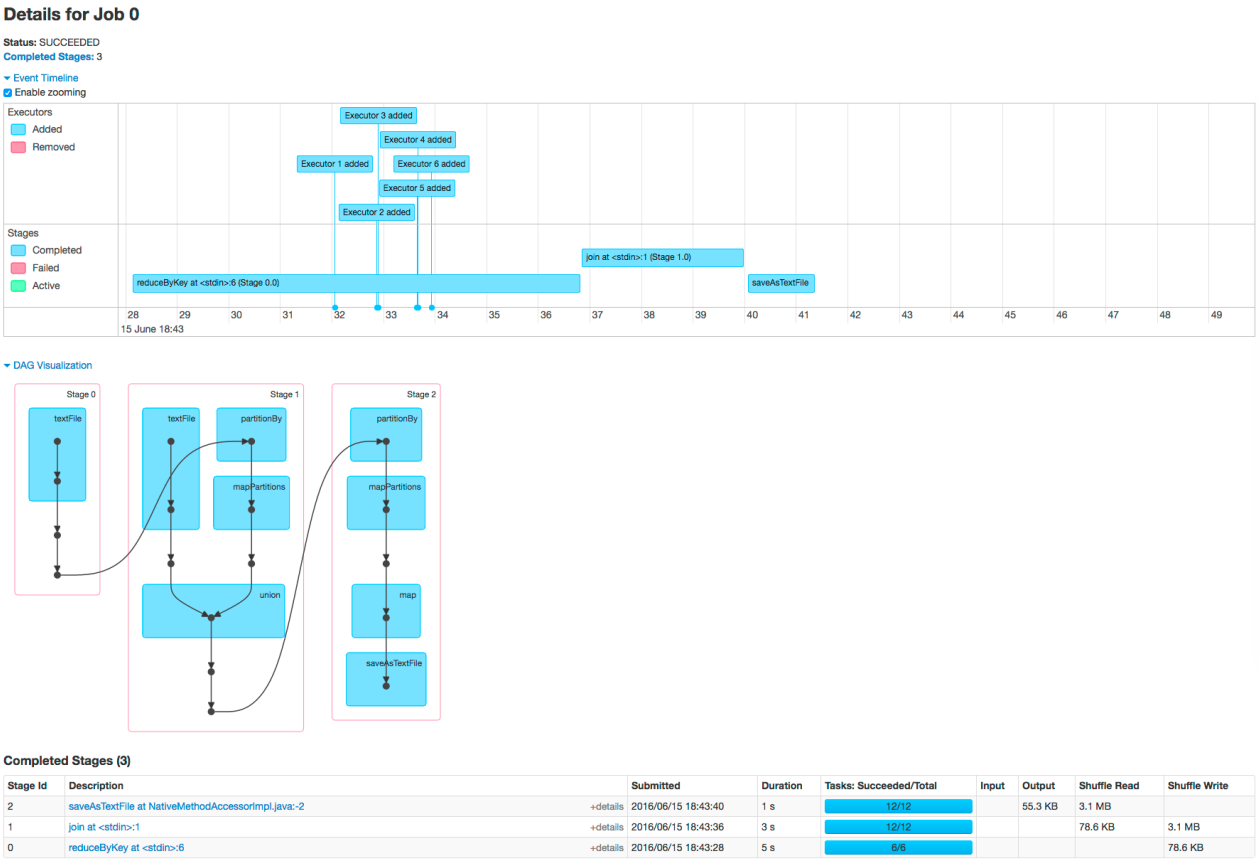
The following screenshot shows the timeline of the events in the application including the jobs that were run and the allocation and deallocation of executors. Each job shows the last action, `saveAsTextFile`, run for the job. The timeline shows that the application acquires executors over the course of running the first job. After the second job finishes, the executors become idle and are returned to the cluster.



You can manipulate the timeline as follows:

- Pan - Press and hold the left mouse button and swipe left and right.
- Zoom - Select the **Enable zooming** checkbox and scroll the mouse up and down.

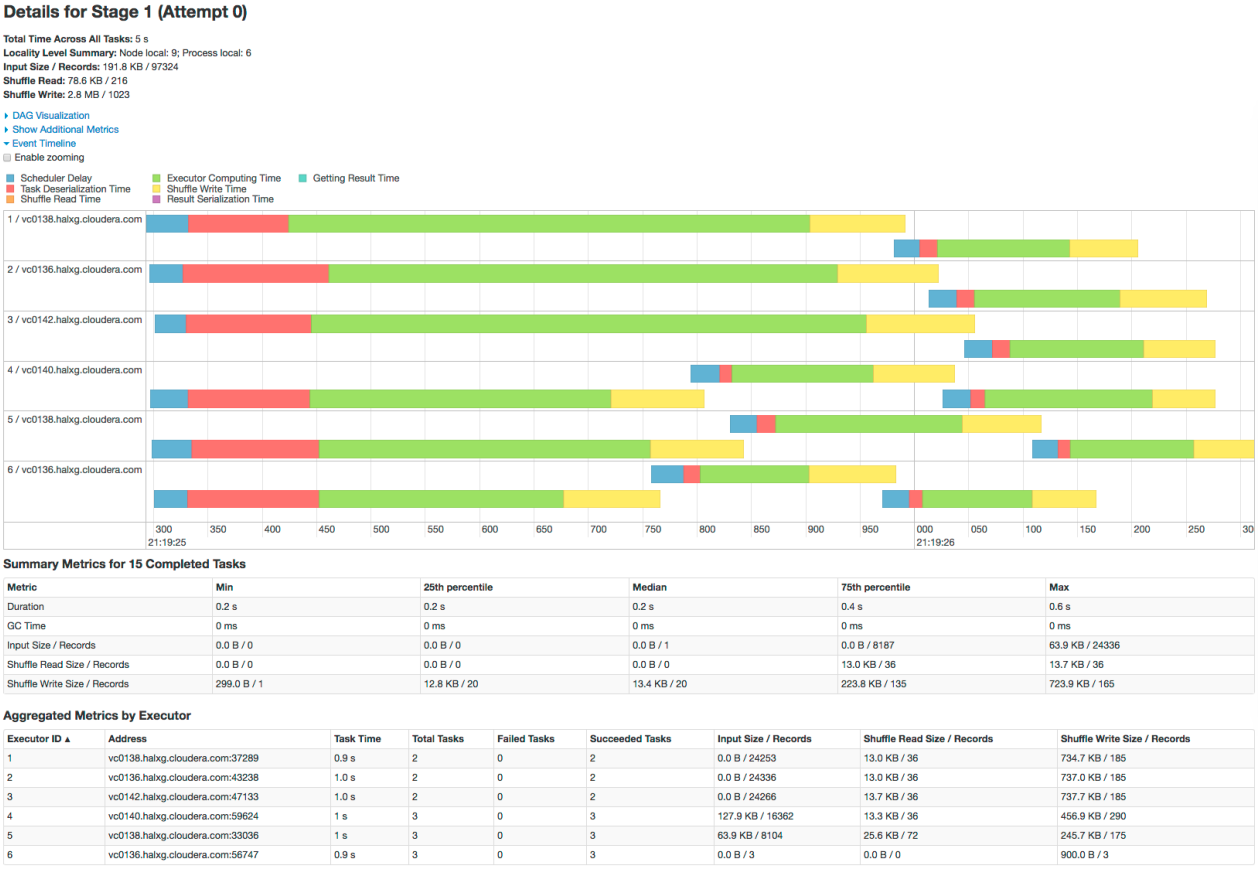
To view the details for Job 0, click the link in the Description column. The following screenshot shows details of each stage in Job 0 and the DAG visualization. Zooming in shows finer detail for the segment from 28 to 42 seconds:



Completed Stages (3)

Stage Id	Description	Submitted	Duration	Tasks: Succeeded/Total	Input	Output	Shuffle Read	Shuffle Write
2	saveAsTextFile at NativeMethodAccessorImpl.java:-2 +details	2016/06/15 18:43:40	1 s	12/12		55.3 KB	3.1 MB	
1	join at <stdin>:1 +details	2016/06/15 18:43:36	3 s	12/12			78.6 KB	3.1 MB
0	reduceByKey at <stdin>:6 +details	2016/06/15 18:43:28	5 s	6/6				78.6 KB

Clicking a stage shows further details and metrics:



Details for Stage 1 (Attempt 0)

Total Time Across All Tasks: 5 s
Locality Level Summary: Node local: 9; Process local: 6
Input Size / Records: 191.8 KB / 97324
Shuffle Read: 78.6 KB / 216
Shuffle Write: 2.8 MB / 1023

DAG Visualization
Show Additional Metrics
Event Timeline
Enable zooming

Scheduler Delay
Task Deserialization Time
Shuffle Read Time
Executor Computing Time
Shuffle Write Time
Result Serialization Time
Getting Result Time

1 / vc0138.halxg.cloudera.com
2 / vc0136.halxg.cloudera.com
3 / vc0142.halxg.cloudera.com
4 / vc0140.halxg.cloudera.com
5 / vc0138.halxg.cloudera.com
6 / vc0136.halxg.cloudera.com

300 350 400 450 500 550 600 650 700 750 800 850 900 950 000 050 100 150 200 250 30
21:19:25 21:19:26

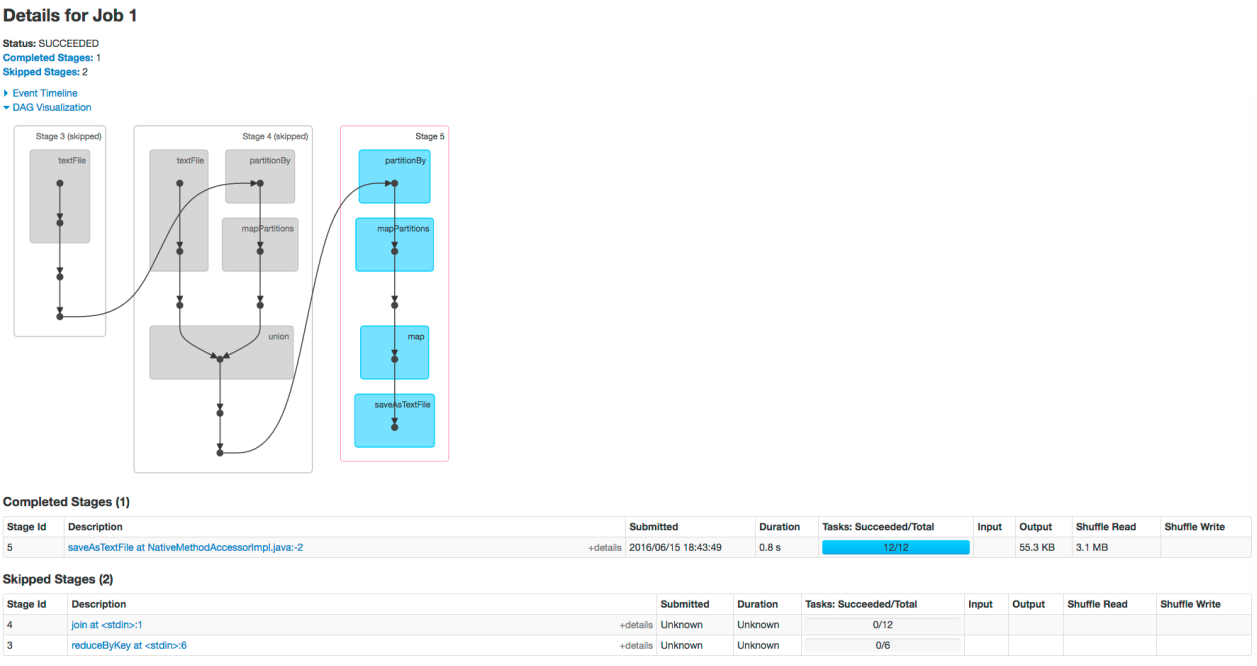
Summary Metrics for 15 Completed Tasks

Metric	Min	25th percentile	Median	75th percentile	Max
Duration	0.2 s	0.2 s	0.2 s	0.4 s	0.6 s
GC Time	0 ms	0 ms	0 ms	0 ms	0 ms
Input Size / Records	0.0 B / 0	0.0 B / 0	0.0 B / 1	0.0 B / 8187	63.9 KB / 24336
Shuffle Read Size / Records	0.0 B / 0	0.0 B / 0	0.0 B / 0	13.0 KB / 36	13.7 KB / 36
Shuffle Write Size / Records	299.0 B / 1	12.8 KB / 20	13.4 KB / 20	223.8 KB / 135	723.9 KB / 165

Aggregated Metrics by Executor

Executor ID ▲	Address	Task Time	Total Tasks	Failed Tasks	Succeeded Tasks	Input Size / Records	Shuffle Read Size / Records	Shuffle Write Size / Records
1	vc0138.halxg.cloudera.com:37289	0.9 s	2	0	2	0.0 B / 24253	13.0 KB / 36	734.7 KB / 185
2	vc0136.halxg.cloudera.com:43238	1.0 s	2	0	2	0.0 B / 24336	13.0 KB / 36	737.0 KB / 185
3	vc0142.halxg.cloudera.com:47133	1.0 s	2	0	2	0.0 B / 24266	13.7 KB / 36	737.7 KB / 185
4	vc0140.halxg.cloudera.com:59624	1 s	3	0	3	127.9 KB / 16362	13.3 KB / 36	456.9 KB / 290
5	vc0138.halxg.cloudera.com:33036	1 s	3	0	3	63.9 KB / 8104	25.6 KB / 72	245.7 KB / 175
6	vc0136.halxg.cloudera.com:56747	0.9 s	3	0	3	0.0 B / 3	0.0 B / 0	900.0 B / 3

The web page for Job 1 shows how preceding stages are skipped because Spark retains the results from those stages:



Details for Job 1

Status: SUCCEEDED
Completed Stages: 1
Skipped Stages: 2

Event Timeline
DAG Visualization

Stage 3 (skipped)
textFile
Stage 4 (skipped)
textFile
partitionBy
mapPartitions
union
Stage 5
partitionBy
mapPartitions
map
saveAsTextFile

Completed Stages (1)

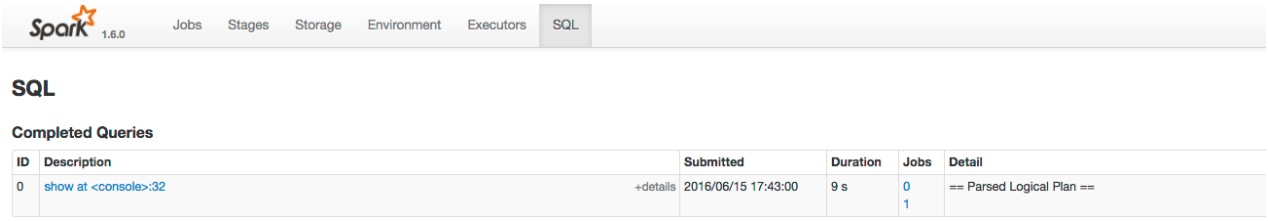
Stage Id	Description	Submitted	Duration	Tasks: Succeeded/Total	Input	Output	Shuffle Read	Shuffle Write
5	saveAsTextFile at NativeMethodAccessorImpl.java:-2 +details	2016/06/15 18:43:49	0.8 s	12/12		55.3 KB	3.1 MB	

Skipped Stages (2)

Stage Id	Description	Submitted	Duration	Tasks: Succeeded/Total	Input	Output	Shuffle Read	Shuffle Write
4	join at <stdin>:1 +details	Unknown	Unknown	0/12				
3	reduceByKey at <stdin>:6 +details	Unknown	Unknown	0/6				

Example Spark SQL Web Application

In addition to the screens described above, the web application UI of an application that uses the Spark SQL API also has an SQL tab. Consider an application that loads the contents of two tables into a pair of DataFrames, joins the tables, and then shows the result. After you click the application ID, the SQL tab displays the final action in the query:

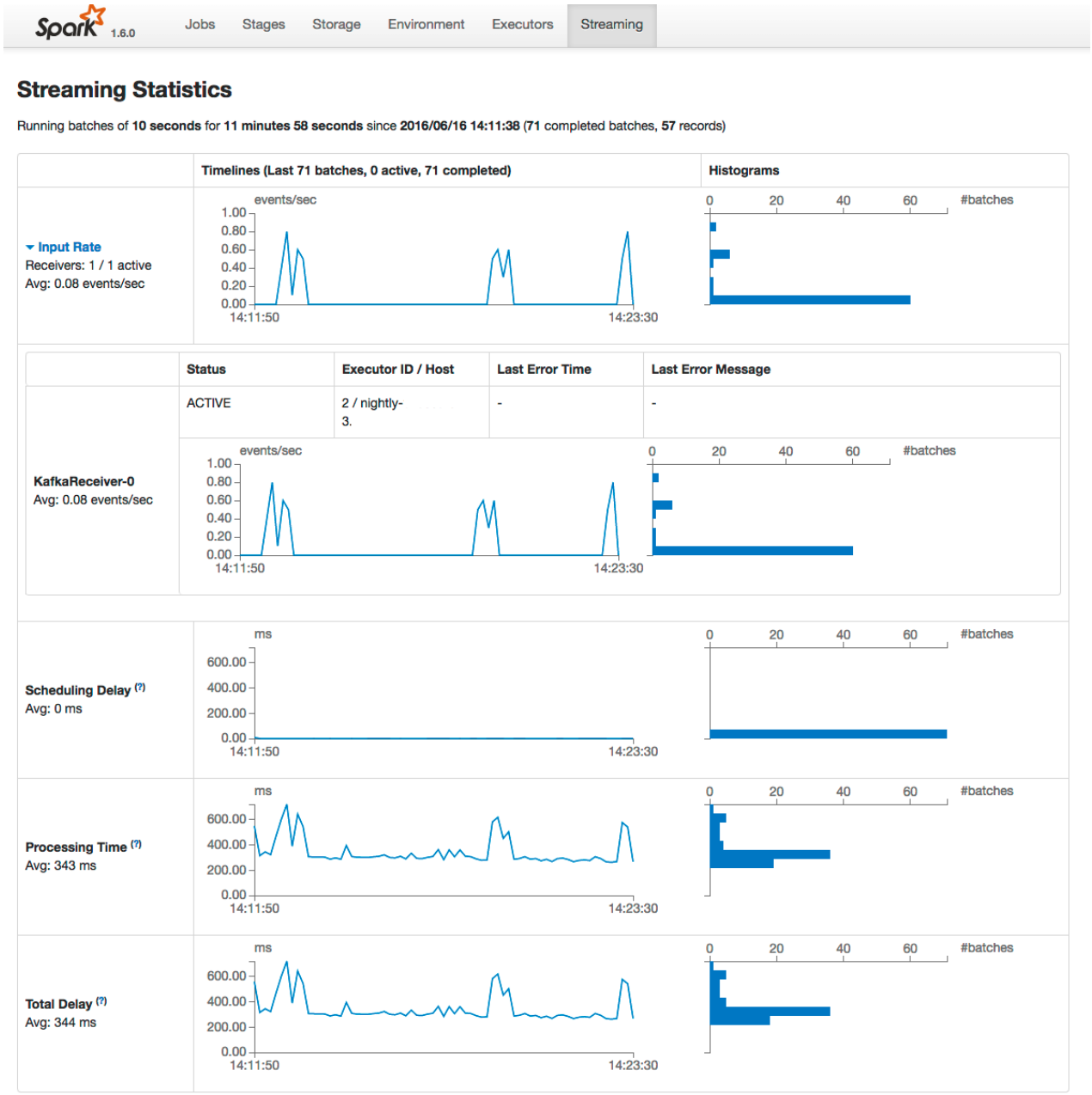


If you click the show link you see the DAG of the job. Clicking the **Details** link on this page displays the logical query plan:

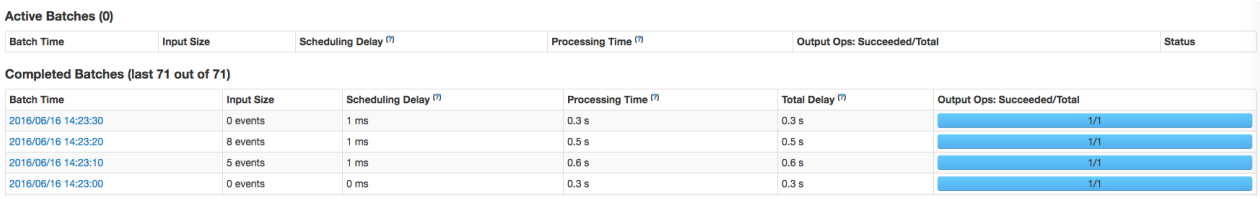
Example Spark Streaming Web Application

> **Note:** The following example demonstrates the Spark driver web UI. Streaming information is not captured in the Spark History Server.

The Spark driver web application UI also supports displaying the behavior of streaming applications in the **Streaming** tab. If you run the example described in Spark Streaming Example, and provide three bursts of data, the top of the tab displays a series of visualizations of the statistics summarizing the overall behavior of the streaming application:



The application has one receiver that processed 3 bursts of event batches, which can be observed in the events, processing time, and delay graphs. Further down the page you can view details of individual batches:



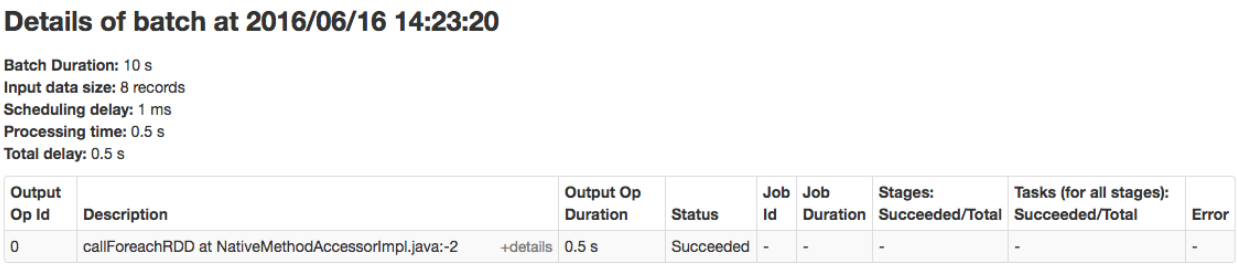
Active Batches (0)

Batch Time	Input Size	Scheduling Delay (?)	Processing Time (?)	Output Ops: Succeeded/Total	Status

Completed Batches (last 71 out of 71)

Batch Time	Input Size	Scheduling Delay (?)	Processing Time (?)	Total Delay (?)	Output Ops: Succeeded/Total
2016/06/16 14:23:30	0 events	1 ms	0.3 s	0.3 s	1/1
2016/06/16 14:23:20	8 events	1 ms	0.5 s	0.5 s	1/1
2016/06/16 14:23:10	5 events	1 ms	0.6 s	0.6 s	1/1
2016/06/16 14:23:00	0 events	0 ms	0.3 s	0.3 s	1/1

To view the details of a specific batch, click a link in the **Batch Time** column. Clicking the **2016/06/16 14:23:20** link with 8 events in the batch, provides the following details:



Details of batch at 2016/06/16 14:23:20

Batch Duration: 10 s
Input data size: 8 records
Scheduling delay: 1 ms
Processing time: 0.5 s
Total delay: 0.5 s

Output Op Id	Description	Output Op Duration	Status	Job Id	Job Duration	Stages: Succeeded/Total	Tasks (for all stages): Succeeded/Total	Error
0	callForeachRDD at NativeMethodAccessorImpl.java:-2 +details	0.5 s	Succeeded	-	-	-	-	-

Events

An ***event*** is a record that something of interest has occurred – a service's health has changed state, a log message (of the appropriate severity) has been logged, and so on. Many events are enabled and configured by default.

From the Events page you can filter for events for services or role instances, hosts, users, commands, and much more. You can also search against the content information returned by the event.

The Event Server aggregates relevant events and makes them available for alerting and for searching. This way, you have a view into the history of all relevant events that occur cluster-wide.

Cloudera Manager supports the following categories of events:

Category	Description
ACTIVITY_EVENT	Generated by the Activity Monitor; specifically, for jobs that fail, or that run slowly (as determined by comparison with duration limits). In order to monitor your workload for slow-running jobs, you must specify Activity Duration Rules on page 16.
AUDIT_EVENT	Generated by actions performed • In Cloudera Manager, such as creating, configuring, starting, stopping, and deleting services or roles • By services that are being audited by Cloudera Navigator.
HBASE	Generated by HBase with the exception of log messages, which have the LOG_MESSAGE category.
HEALTH_CHECK	Indicate that certain health test activities have occurred, or that health test results have met specific conditions (thresholds). Thresholds for various health tests can be set under the **Configuration** tabs for HBase, HDFS, Impala, and MapReduce service instances, at both the service and role level. See Configuring Health Monitoring on page 15 for more information.
LOG_MESSAGE	Generated for certain types of log messages from HDFS, MapReduce, and HBase services and roles. Log events are created when a log entry matches a set of rules for identifying messages of interest. The default set of rules is based on Cloudera experience supporting Hadoop clusters. You can configure additional log event rules if necessary.

Category	Description
SYSTEM	Generated by system events such as parcel availability.

Viewing Events

The **Events** page lets you display events and alerts that have occurred within a time range you select anywhere in your clusters. From the Events page you can filter for events for services or role instances, hosts, users, commands, and much more. You can also search against the content information returned by the event.

To view events, click the **Diagnostics** tab on the top navigation bar, then select **Events**.

Events List

Event entries are ordered (within the time range you've selected) with the most recent at the top. If the event generated an alert, that is indicated by a red alert icon (Alert) in the entry.

This page supports infinite scrolling: you can scroll to the end of the displayed results and the page will fetch more results and add them to the end of the list automatically.

To display event details, click **Expand** at the right side of the event entry.

Clicking the **View** link at the far right of the entry has different results depending on the category of the entry:

- **ACTIVITY_EVENT** - Displays the activity Details page.
- **AUDIT_EVENT** - If the event was a restart, displays the service's Commands page. If the event was a configuration change, the Revision Details dialog box displays.
- **HBASE** - Displays a health report or log details.
- **HEALTH_CHECK** - Displays the status page of the role instance.
- **LOG_MESSAGE** - Displays the event's log entry. You can also click **Expand** to display details of the entry, then click the **URL** link. When you perform one of these actions the time range in the Time Line is shifted to the time the event occurred.
- **SYSTEM** - Displays the Parcels page.

Filtering Events

You filter events by selecting a time range and adding filters.

You can use the Time Range Selector or a duration link (30m 1h 2h 6h 12h 1d 7d 30d) to set the time range. (See Time Line on page 7 for details). The time it takes to perform a search will typically increase for a longer time range, as the number of events to be searched will be larger.

Adding a Filter

To add a filter, do one of the following:

- Click the icon that displays next to a property when you hover in one of the event entries. A filter containing the property, operator, and its value is added to the list of filters at the left and Cloudera Manager redisplays all events that match the filter.
- Click the **Add a filter** link. A filter control is added to the list of filters.
  1. Choose a property in the drop-down list. You can search by properties such as Username, Service, Command, or Role. The properties vary depending on the service or role.
  2. If the property allows it, choose an operator in the operator drop-down list.
  3. Type a property value in the value text field. For some properties you can include multiple values in the value field. For example, you can create a filter like `Category = HEALTH_CHECK LOG_MESSAGE`. To drop individual values, click the × to the right of the value. For properties where the list of values is finite and known, you can start typing and then select from a drop-down list of potential matches.
  4. Click **Search**. The log displays all events that match the filter criteria.

5. Click ➕ to add more filters and repeat steps 1 through 4.

> **Note:** You can filter on a string by adding a filter, selecting the property CONTENT, operator =, and typing the string to search for in the value field.

Removing a Filter

1. Click the ➖ at the right of the filter. The filter is removed.
2. Click **Search**. The log displays all events that match the filter criteria.

Re-running a Search

To re-run a recently performed search, click

to the right of the Search button and select a search.

Triggers

A ***trigger*** is a statement that specifies an action to be taken when one or more specified conditions are met for a service, role, role configuration group, or host. The conditions are expressed as a tsquery statement, and the action to be taken is to change the health for the service, role, role configuration group, or host to either Concerning (yellow) or Bad (red).

Triggers can be created for services, roles, role configuration groups, or hosts. Create a trigger by doing one of the following:

- Directly editing the configuration for the service, role (or role configuration group), or host configuration.
- Clicking **Create Trigger** on the drop-down menu for most charts. Note that the Create Trigger command is not available on the drop-down menu for charts where no context (role, service, and so on) is defined, such as on the **Home** > **Status** tab.

  > **Important:** Because triggers are a new and evolving feature, backward compatibility between releases is not guaranteed at this time.

- Use the **Create Trigger** expression builder. See Creating a Trigger Using the Expression Editor on page 92.

The Structure of Triggers

A trigger is defined by a JSON formatted string that includes four parts:

- Name
- Expression
- Stream threshold
- Whether or not the trigger should be enabled

Each of the four parts of a trigger is described in the following sections.

Name (required)

A trigger's name must be unique in the context for which the trigger is defined. That is, there cannot be two triggers for the same service or role with the same name. Different services or different roles can have triggers with the same name.

Expression (required)

A trigger expression takes the form:

`IF (`*`CONDITIONS`*`) DO `*`HEALTH_ACTION`*

When the conditions of the trigger are met, the trigger is considered to be firing. A condition is any valid tsquery statement. In most cases conditions employ stream filters to filter out streams below or above a certain threshold. For example, the following tsquery can be used to retrieve the streams for DataNodes with more than 500 open file descriptors:

```
SELECT fd_open WHERE roleType=DataNode AND last(fd_open) > 500
```

The stream filter used here, `last(fd_open) > 50`, is composed of four parts:

- A scalar producing function "last" that takes a stream and returns its last data point
- A metric to operate on
- A comparator
- A scalar value

Other scalar producing functions are available, like `min` or `max`, and they can be combined to create arbitrarily complex expressions:

```
last(moving_avg(fd_open)) >= 500
```

See the tsquery documentation for more details.

Conditions can be combined using the logical operators AND and OR. For example, here is a trigger expression with two conditions:

```
IF ((SELECT fd_open WHERE roleType=DataNode AND last(fd_open) > 500) OR (SELECT fd_open
 WHERE roleType=NameNode AND last(fd_open) > 500)) DO health:bad
```

A condition is met if it returns more than the number of streams specified by the `streamThreshold` (see below). A trigger fires if the logical evaluation of all of its conditions results in a met condition. When a trigger fires, two actions can be taken: `health:concerning` or `health:bad`. These actions change the health of the entity on which the trigger is defined.

Stream Threshold (optional)

The stream threshold determines the number of streams that need to be returned by the tsquery before the condition is met. The default is 0; that is, if the tsquery returns any results the condition will be met. For example if the stream threshold is set to 10 and the condition is `SELECT fd_open WHERE roleType=DataNode AND last(fd_open) > 500` the condition will be considered *met* only if there are at least 10 DataNodes that have more than 500 file descriptor opened, so at least 10 streams were returned by the tsquery.

Enabled (optional)

Whether the trigger is enabled. The default is `true`, (enabled).

Trigger Example

The following is a JSON formatted trigger that fires if there are more than 10 DataNodes with more than 500 file descriptors opened:

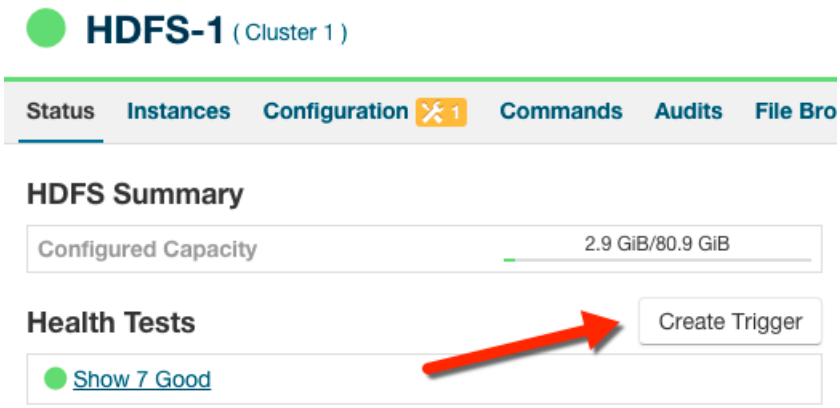
```
[{"triggerName": "sample-trigger", "triggerExpression": "IF (SELECT fd_open WHERE roleType
 = DataNode and last(fd_open) > 500) DO health:bad", "streamThreshold": 10, "enabled":
 "true"}]
```

Creating a Trigger Using the Expression Editor

The **Create New Trigger** screen allows you to use a graphical editor to build the JSON string that defines a trigger. You can use the expression editor section to build the tsquery statement, or you can edit the tsquery statement manually. Triggers use the tsquery Language on page 111 to create trigger expressions.

To create a trigger using the expression editor:

1. Go to a service, role, role configuration group, or host configuration page and click the **Create Trigger** button next to the **Health Test** section.



   The **Create New Trigger** screen displays.

   As you build the trigger, the actual query text displays to the right, along with a preview of a chart returned by the query.

2. Enter a name for the trigger in the **Name** field.
3. Build the Metric Expressions on page 112:
   a. Select the function to use in your expression, either **Last**, **Min**, or **Max**.
   b. Select the metric by typing its name in the **Metric** field. A list of available metrics displays as you type.
   c. Select the operator, either **>**, **>=**, **=**, **<**, or **<=**.
   d. Enter the value to use for the comparison in the **Value** field.
   e. (Optional) Click the **+** icon to add additional expressions. Additional expressions are added to the query using the logical operator **AND**.
4. (Optional) Create a predicate for the query. Under **Attribute Conditions**, click the **+** icon to add an attribute condition.

   A set of fields displays that you use to build an expression for the predicate.

   a. Type the attribute name in the **Attribute** field. A list of attributes displays as you type.
   b. Select the operator, either **=** or **RLIKE**.
   c. Enter the value for the comparison in the **Value** field.
   d. (Optional) Click the **+** icon to add additional expressions. Additional expressions are added to the predicate using the logical operator **AND**.
5. Select an **Action** from the drop-down menu to define the action taken when the trigger fires:
   - **Mark as bad** (red)
   - **Mark as concerning** (yellow)
6. Enter a value for the **Stream Threshold**. Leave the value set to `0` to include all streams; enter an integer to set the number of streams required to meet the condition. See Stream Threshold.
7. Select **Enabled** to enable the trigger. If you disable the trigger, it does not run.
8. (Optional) Select **Suppressed**. A suppressed trigger still runs but does not impact the health display of the owning entity.
9. Verify your expression:

   In the area to the right of the expression builder, in the **Preview** section, the expression you have built displays. A chart also displays the result of the query. Click **Show Filtered Streams** to see all streams. Click **Hide Filtered Streams** to hide streams that do not meet the Stream Threshold (optional) on page 92.

You can edit your trigger using the fields in the expression builder, or you can click the **Edit Manually** link to display a text box in which you can manually edit the trigger. Click **Use Editor** to return to the expression builder.

> **Important:** If you select **Edit Manually**, changes you make manually do not appear in the expression builder when you click **Use Editor**.

10. Click **Create Trigger** to save your trigger.

Editing, Deleting, Suppressing, or Deleting a Trigger

1. Go to the service, role, role configuration group, or host configuration page where the trigger was created. (For example: select **Clusters** > **HDFS**.)
2. In the **Health Tests** section, click the trigger name. (You may need to click a **Show ...** link to expand the list of triggers.)

   A page displays showing the query and chart for the trigger. Click **Show Filtered Streams** to see all streams. Click **Hide Filtered Streams** to hide streams that do not meet the Stream Threshold (optional) on page 92.
3. Click the **Actions** drop-down menu and select one of the following actions:
   - **Edit Trigger**

     A page opens where you can edit the query. Click **Save Trigger** to save your changes.
   - **Disable Trigger** or **Enable Trigger**
   - **Suppress Trigger** or **Unsuppress Trigger**
   - **Delete Trigger**

Cloudera Manager Trigger Use Cases

Cloudera Manager allows you to monitor cluster performance. Some indicators require timely attention to keep your data safe. Triggers let you track occurrence and severity of issues so that you can fix problems before they result in system failures.

The conditions you create for your trigger can be quite complex, but they do not need to be in order to be useful. This topic describes two triggers that alert you when you are approaching capacity limits for your cluster.

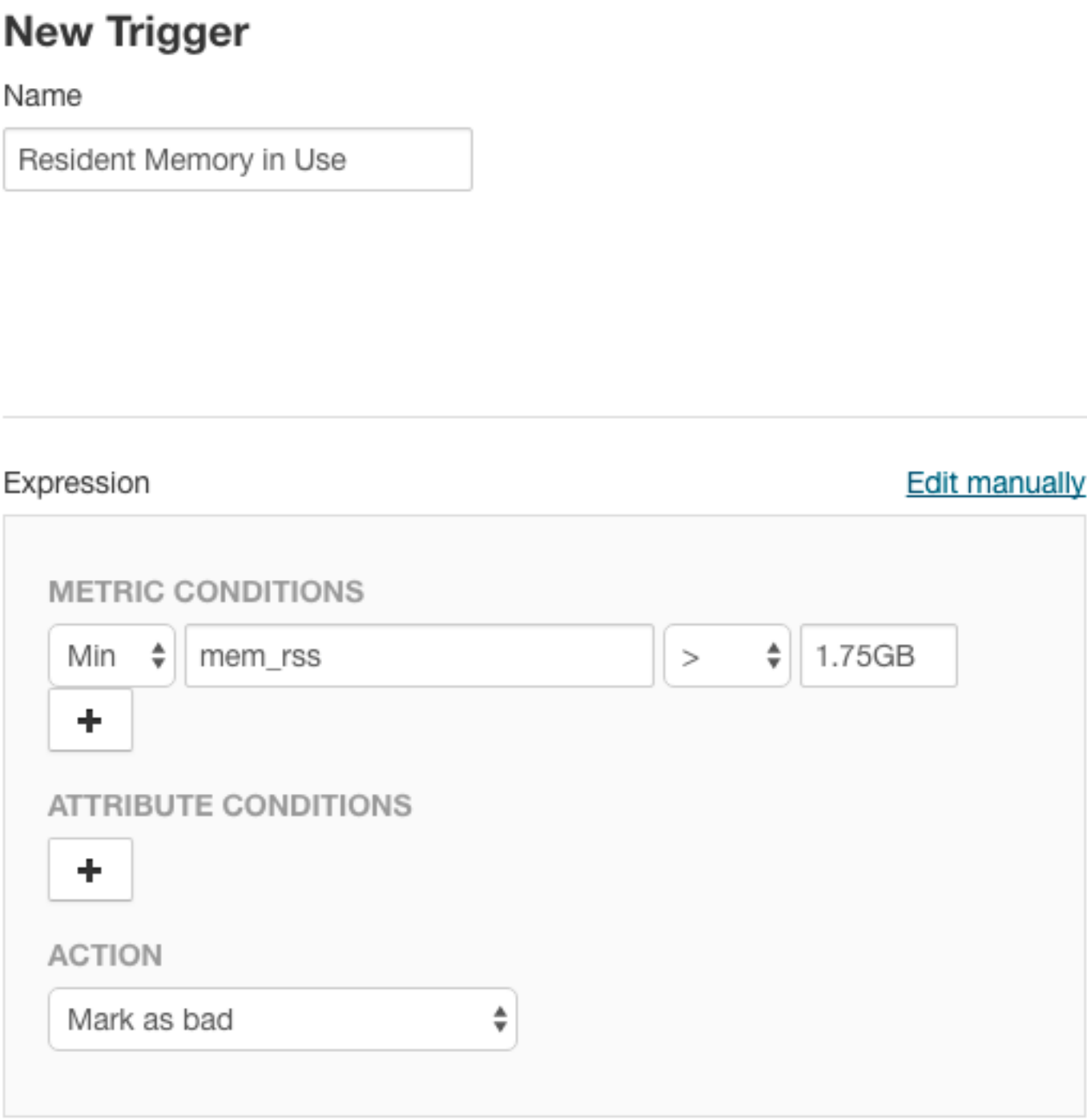
Creating a Trigger for Memory Capacity

A common use case is to monitor memory usage, and trigger a warning if your system is approaching its upper limit.

To create a memory usage trigger, do the following.

1. In Cloudera Manager, go to the **Hosts** page.
2. Click a link in the **Name** column to open a host status page.
3. In the **Health Tests** section, click **Create Trigger**.
4. On the **New Trigger** page, enter the name `Resident Memory In Use`.
5. In **Expression**, set these metric conditions.
   a. **Scalar Function**: `Min`.
   b. **Metric**: `mem_rss`.
   c. **Comparator**: **>** (greater than).
   d. **Scalar Value**: `1.75GB`. (This is a low value for demonstration purposes, so that it will trigger the action. In practice, use a value that more accurately reflects the memory limits of your cluster.)
6. Set **Action** to `Mark as bad`.

   As you work, the **Preview** shows the resulting chart and current status of your host.



7. Scroll down and choose whether to apply this trigger to **All hosts**.
8. Click **Create Trigger**.

Creating a Trigger for CPU Capacity

Another key indicator for performance is CPU capacity. Ideally, you will consistently use most of your allocated CPU resources on a regular basis without exceeding capacity.

To create a CPU capacity trigger, do the following.

1. In Cloudera Manager, go to the **Hosts** page.

2. Click a link in the **Name** column to open a host status page.
3. In the **Health Tests** section, click **Create Trigger**.
4. On the **New Trigger** page, enter the name `CPU Capacity`.
5. In **Expression**, set these metric conditions.
    a. **Scalar Function**: `Min`.
    b. **Metric**: `cpu_percent`.
    c. **Comparator**: **>** (greater than).
    d. **Scalar Value**: `90`.
6. Set **Action** to `Mark as concerning`.
7. Click **Create Trigger**.

This trigger fires whenever the CPU percentage exceeds 90%. However, just exceeding 90% of available CPU resources is not necessarily a bad thing. What would be of more concern is if CPU resources were to consistently exceed 90% over an extended period. You can modify the trigger to evaluate the average CPU usage over time.

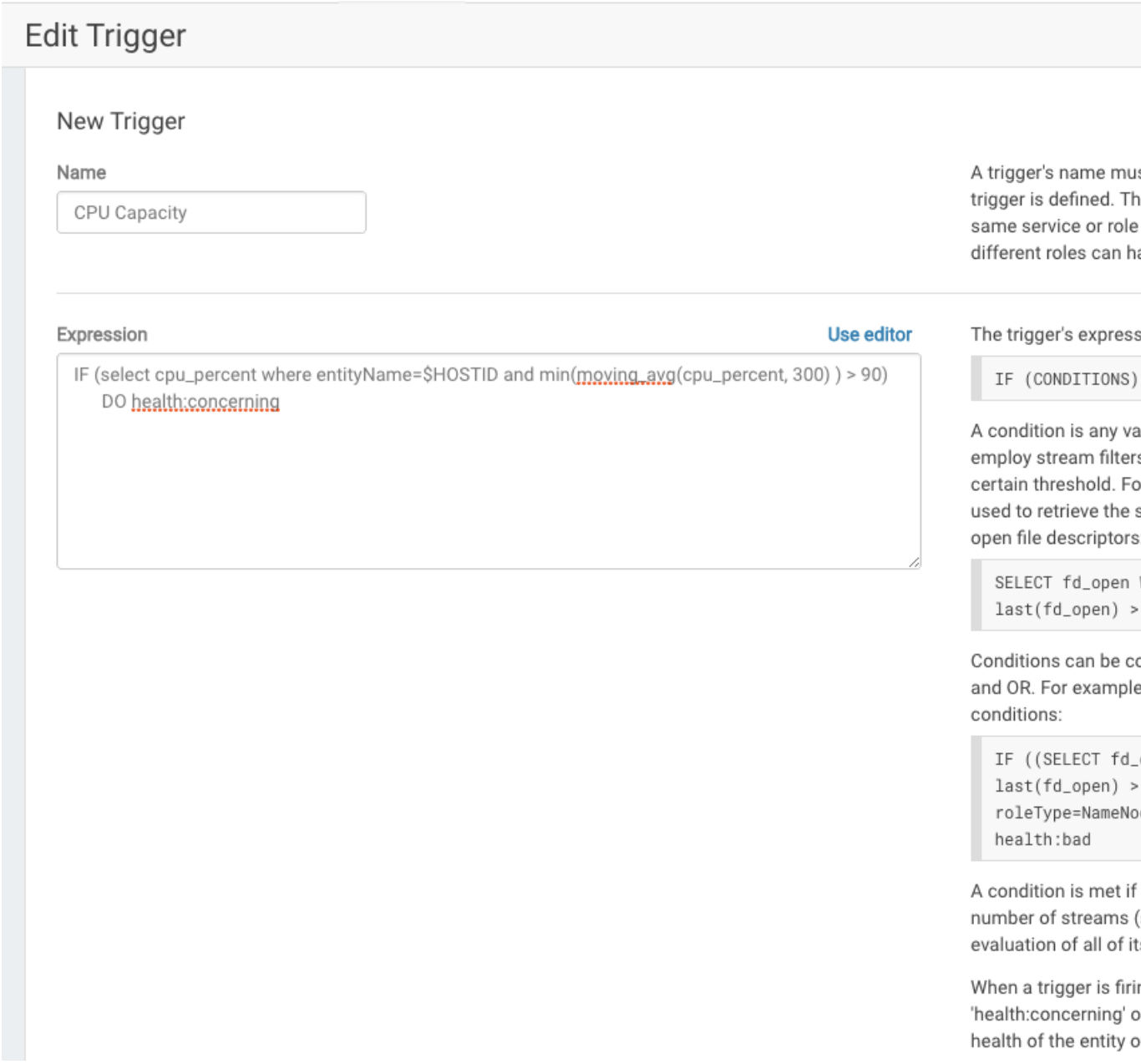
To modify the trigger to capture high CPU usage in a five minute window, do the following.

1. In Cloudera Manager, go to the **Hosts** page.
2. Click a link in the **Name** column to open the host status page.
3. In the **Health Tests** section, click the **Show *n* Good** link.
4. Click the link for **CPU Capacity**.
5. Choose **Actions** > **Edit Trigger**.

    The metric expression function for average over time, `moving_avg`, is not available from the pop-up menu in the editor. You can edit the expression directly using tsquery language.

6. Above the **Expression** editor, click **Edit manually**.
7. Revise the expression as follows.

```
IF (select cpu_percent where entityName=$HOSTID and min(moving_avg(cpu_percent, 300) )
 > 90)
      DO health:concerning
```



8. Click **Save Trigger**.

For more information on defining triggers, see Triggers on page 91.

For more information on writing custom queries, see tsquery Language on page 111.

For the complete list of available metrics, see Cloudera Manager Metrics.

Lifecycle and Security Auditing

Minimum Required Role: Auditor (also provided by **Full Administrator**)

An ***audit event*** is an event that describes an action that has been taken for a cluster, host, license, parcel, role, service or user.

Cloudera Manager records cluster, host, license, parcel, role, and service ***lifecycle events*** (activate, create, delete, deploy, download, install, start, stop, update, upgrade, and so on), user ***security-related events*** (add and delete user,

login failed and succeeded), and provides an audit UI and API to view, filter, and export such events. The Cloudera Manager audit log does not track the progress or results of commands (such as starting or stopping a service or creating a directory for a service), it just notes the command that was executed and the user who executed it. To view the progress or results of a command, follow the procedures in Viewing Running and Recent Commands on page 44.

The Cloudera Navigator Audit Server records ***service access events*** and the Cloudera Navigator Metadata Server provides an audit UI and API to view, filter, and export both service access events and the lifecycle and security events retrieved from Cloudera Manager. For information on Cloudera Navigator auditing features, see Exploring Audit Data.

Viewing Audit Events

You can view audit events for a cluster, service, role, or host.

Object	Procedure
Cluster	1. Click the **Audits** tab on the top navigation bar.
Service	1. Click the **Clusters** tab on the top navigation bar. 2. Select a service. 3. Click the **Audits** tab on the service navigation bar.
Role	1. Click the **Clusters** tab on the top navigation bar. 2. Select a service. 3. Click the **Instances** tab on the service navigation bar. 4. Select a role. 5. Click the **Audits** tab on the role navigation bar.
Host	1. Click the **Hosts** tab on the top navigation bar. 2. Select a host. 3. Click the **Audits** tab on the host navigation bar.

Audit event entries are ordered with the most recent at the top.

Audit Event Properties

The following properties can appear in an audit event entry:

- **Date** - Date and time the action was performed.
- **Command** - The action performed.
- **Source** - The object affected by the action.
- **User** - The name of the user that performed the action.
- **IP Address** - The IP address of the client that initiated the action.
- **Host IP Address** - The IP address of the host on which the action was performed.
- **Service** - The name of the service on which the action was performed.
- **Role** - The name of the role on which the action was performed.

Filtering Audit Events

You filter audit events by selecting a time range and adding filters.

You can use the Time Range Selector or a duration link (30m 1h 2h 6h 12h 1d 7d 30d) to set the time range. (See Time Line on page 7 for details). When you select the time range, the log displays all events in that range. The time it takes to perform a search will typically increase for a longer time range, as the number of events to be searched will be larger.

Adding a Filter

To add a filter, do one of the following:

- Click the icon that displays next to a property when you hover in one of the event entries. A filter containing the property, operator, and its value is added to the list of filters at the left and Cloudera Manager redisplays all events that match the filter.
- Click the **Add a filter** link. A filter control is added to the list of filters.

  1. Choose a property in the drop-down list. You can search by properties such as Username, Service, Command, or Role. The properties vary depending on the service or role.
  2. If the property allows it, choose an operator in the operator drop-down list.
  3. Type a property value in the value text field. To match a substring, use the `like` operator and specify `%` around the string. For example, to see all the audit events for files created in the folder `/user/joe/out` specify `Source like %/user/joe/out%`.
  4. Click **Search**. The log displays all events that match the filter criteria.
  5. Click + to add more filters and repeat steps 1 through 4.

Removing a Filter

1. Click the − at the right of the filter. The filter is removed.
2. Click **Search**. The log displays all events that match the filter criteria.

Downloading Audit Events

You can download audit events in CSV formats.

1. Specify desired filters and time range.
2. Click the **Download CSV** button. A file with the following fields is downloaded: `service`, `username`, `command`, `ipAddress`, `resource`, `allowed`, `timestamp`, `operationText`. The structure of the `resource` field depends on the type of the service:

   - HDFS - A file path
   - Hive, Hue, and Impala - *database*:*tablename*
   - HBase - *table family*:*qualifier*

   For Hive, Hue, and Impala query and load commands, `operationText` is the query string.

HDFS Service Audit Log

```
service,username,command,ipAddress,resource,allowed,timestamp
hdfs1,cloudera,setPermission,10.20.187.242,/user/hive,false,"2013-02-09T00:59:34.430Z"
hdfs1,cloudera,getfileinfo,10.20.187.242,/user/cloudera,true,"2013-02-09T00:59:22.667Z"
hdfs1,cloudera,getfileinfo,10.20.187.242,/,true,"2013-02-09T00:59:22.658Z"
```

In this example, the first event access was denied, and therefore the `allowed` field has the value `false`.

Charting Time-Series Data

Cloudera Manager enables you to enter a query for a time series, chart the time-series data, group (facet) individual time series if your query produced multiple time series, and save the results as a dashboard.

The following sections have more details on the terminology used, how to query for time-series data, displaying chart details, editing charts, and modifying chart properties.

Terminology

Entity
A Cloudera Manager component that has metrics associated with it, such as a service, role, or host.

Metric
A property that can be measured to quantify the state of an entity or activity, such as the number of open file descriptors or CPU utilization percentage.

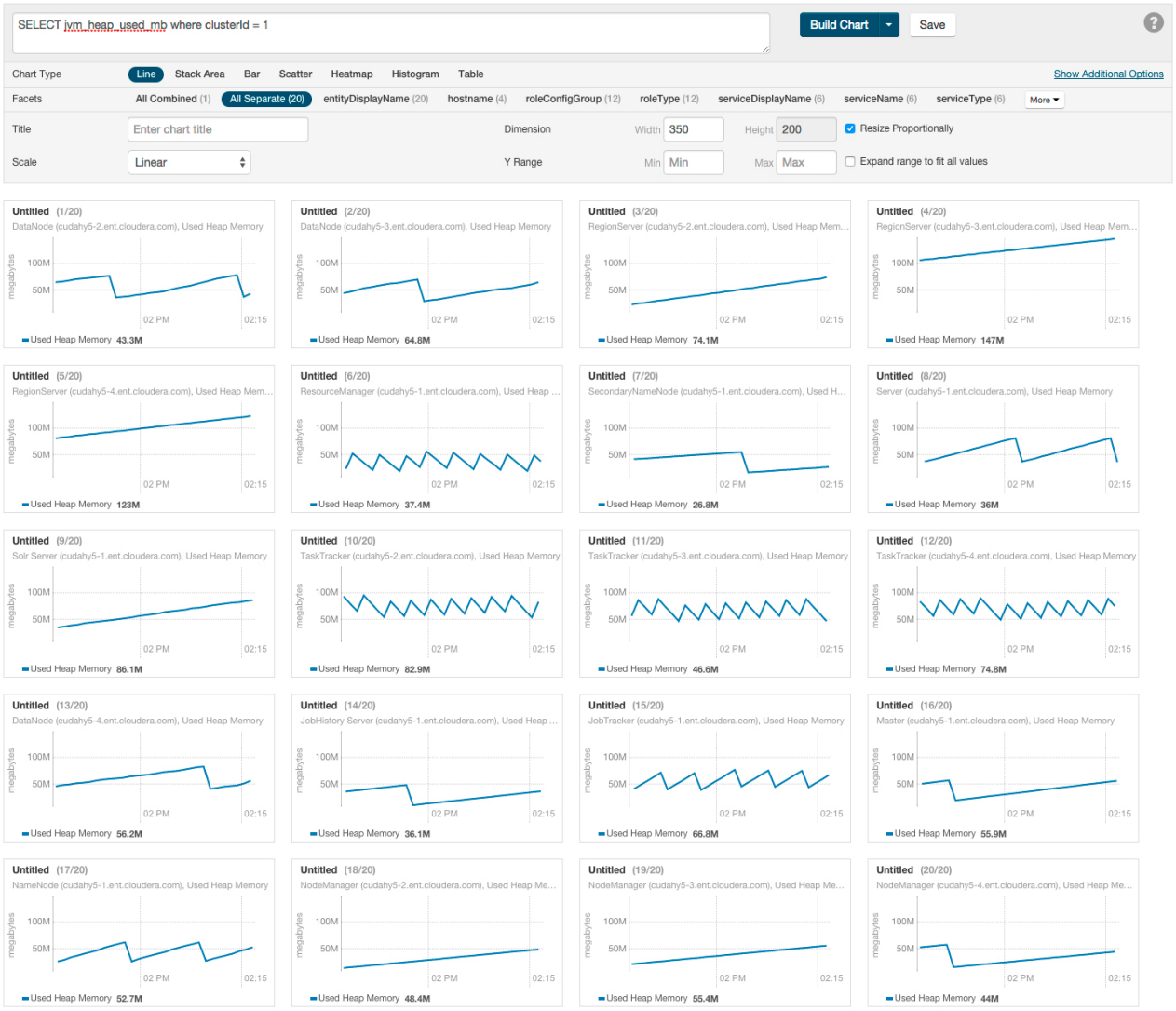
Time series
A list of (time, value) pairs that is associated with some (entity, metric) pair such as, `(datanode-1, fd_open)`, `(hostname, cpu_percent)`. In more complex cases, the time series can represent operations on other time series. For example, `(datanode-1 , cpu_user + cpu_system)`.

Facet
A display grouping of a set of time series. By default, when a query returns multiple time series, they are displayed in individual charts. Facets allow you to display the time series in separate charts, in a single chart, or grouped by various attributes of the set of time series.

Building a Chart with Time-Series Data

1. Select **Charts** > **Chart Builder**.
2. Display time series in one of the following ways:
   - **Select a recently used statement**
     1. Click the 
        to the right of the **Build Chart** button to display a list of recently run statements and select a statement. The statement text displays in the text box and the chart(s) that display that time series will display.
   - **Select from the list of Chart Examples**
     1. Click the question mark icon 
        to the right of the **Build Chart** button to display a list of examples with descriptions.
     2. Click **Try it** to create a chart based on the statement text in the example.
   - **Type a new statement**
     1. Press **Spacebar** in the text box. tsquery statement components display in a drop-down list. These suggestions are part of type ahead, which helps build valid queries. Scroll to the desired component and click **Enter**. Continue choosing query components by pressing **Spacebar** and **Enter** until the tsquery statement is complete.

For example, the query `SELECT jvm_heap_used_mb where clusterId = 1` could return a set of charts like the following:



Configuring Time-Series Query Results

Minimum Required Role: Configurator (also provided by **Cluster Administrator, Full Administrator**)

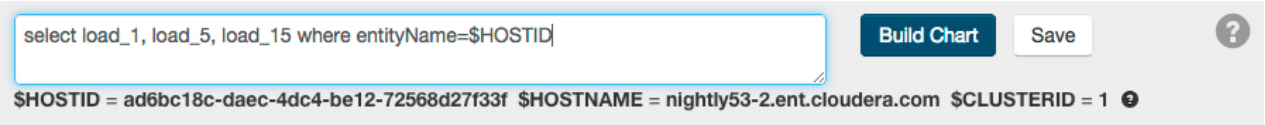
A time-series query returns one or more time series or scalar values. By default a maximum of 250 time series will be returned.

To change this value:

1. Select **Administration** > **Settings**.
2. In the **Advanced** category, set the **Maximum Number Of Time-Series Streams Returned Per Time-Series Query** or the **Maximum Number of Time-Series Streams Returned Per Heatmap** property.
3. Click **Save Changes**.

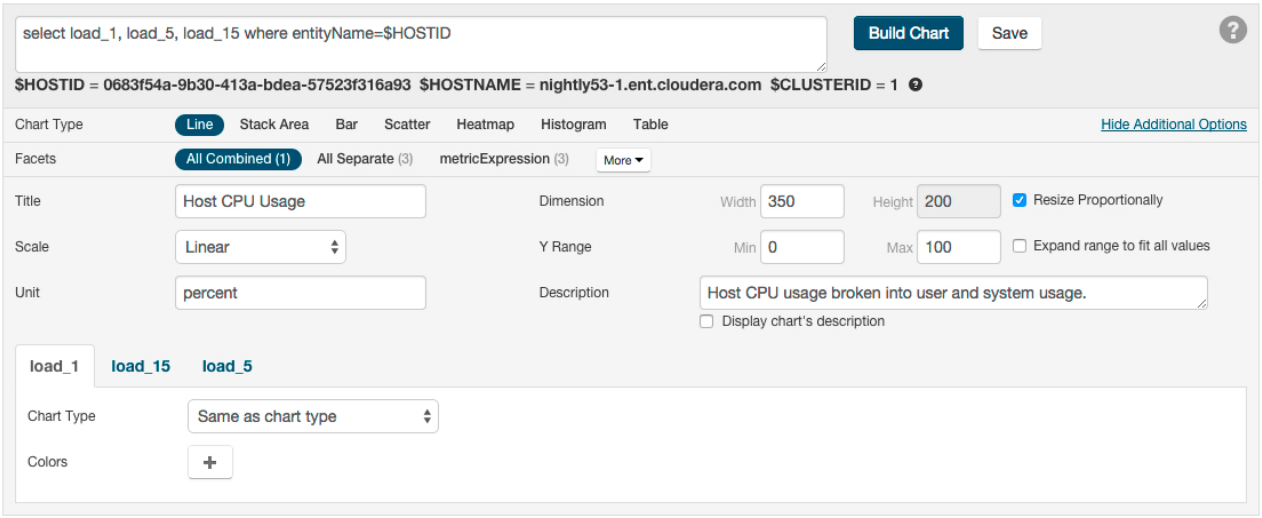
Using Context-Sensitive Variables in Charts

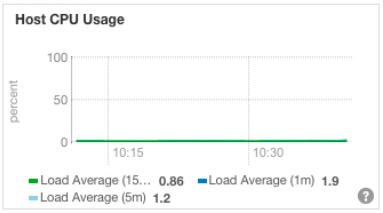
When editing charts from a service, role or host status or charts page, or when adding a chart to a status page, a set of context-sensitive variables (each beginning with '$') will be displayed below the query box on the Chart Builder page. For example, you might see variables similar to those in the query below:



Notice the `$HOSTNAME` portion of the query string. `$HOSTNAME` is a variable that will be resolved to a specific value based on the page before the query is actually issued. In this case, `$HOSTNAME` will become `nightly53-2.ent.cloudera.com`.

The chart below shows an example of the output of a similar query.





Context-sensitive variables are useful since they allow portable queries to be written. For example the query above may be on the host status page or any role status page to display the appropriate host's swap rate. Variables cannot be used in queries that are part of user-defined dashboards since those dashboards have no service, role or host context.

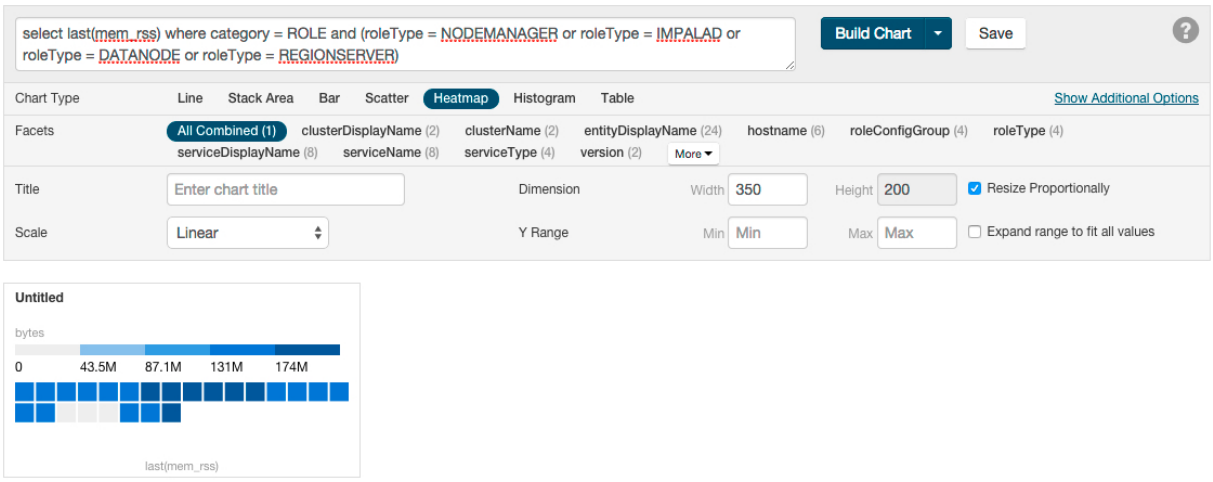
Chart Properties

By default, the time-series data retrieved by the tsquery is displayed on its own chart, using a **Line** style chart, a default size, and a default minimum and maximum for the Y-axis. You can change the chart type, facet the data, set the chart scale and size, and set X- and Y-axis ranges.

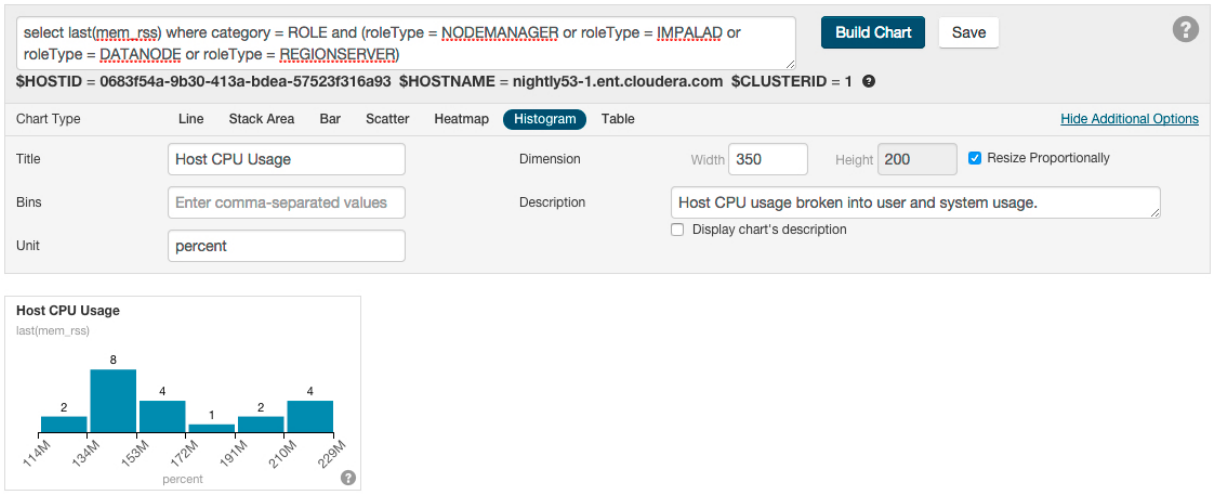
Changing the Chart Type

To change the chart type, click one of the chart types on the left:

- **Line** - Displays the points in the time series as continuous line.
- **Stack Area** - Displays the points in the time series as continuous line and the area under the line filled in.
- **Bar** - Displays each the value of the metric averaged over a second as a bar.
- **Scatter** - Displays the points in the time series as dots.
- **Heatmap** - Displays a metric thermometer and grid of colored squares. The thermometer displays buckets that represent a range of metric values and a color coding for the bucket. Each square represents an entity and the color of the square represents the value of a metric within a range. The following heatmap shows the last value of the resident memory for the NodeManager, ImpalaD, DataNode, and RegionServer roles.



- **Histogram** - Displays the time series values as a set of bars where each bar represents a range of metric values and the height of the bar represents the number of entities whose value falls within the range. The following histogram shows the number of roles in each range of the last value of the resident memory.



- **Table** - Displays the time series values as a table with each row containing the data for a single time value.

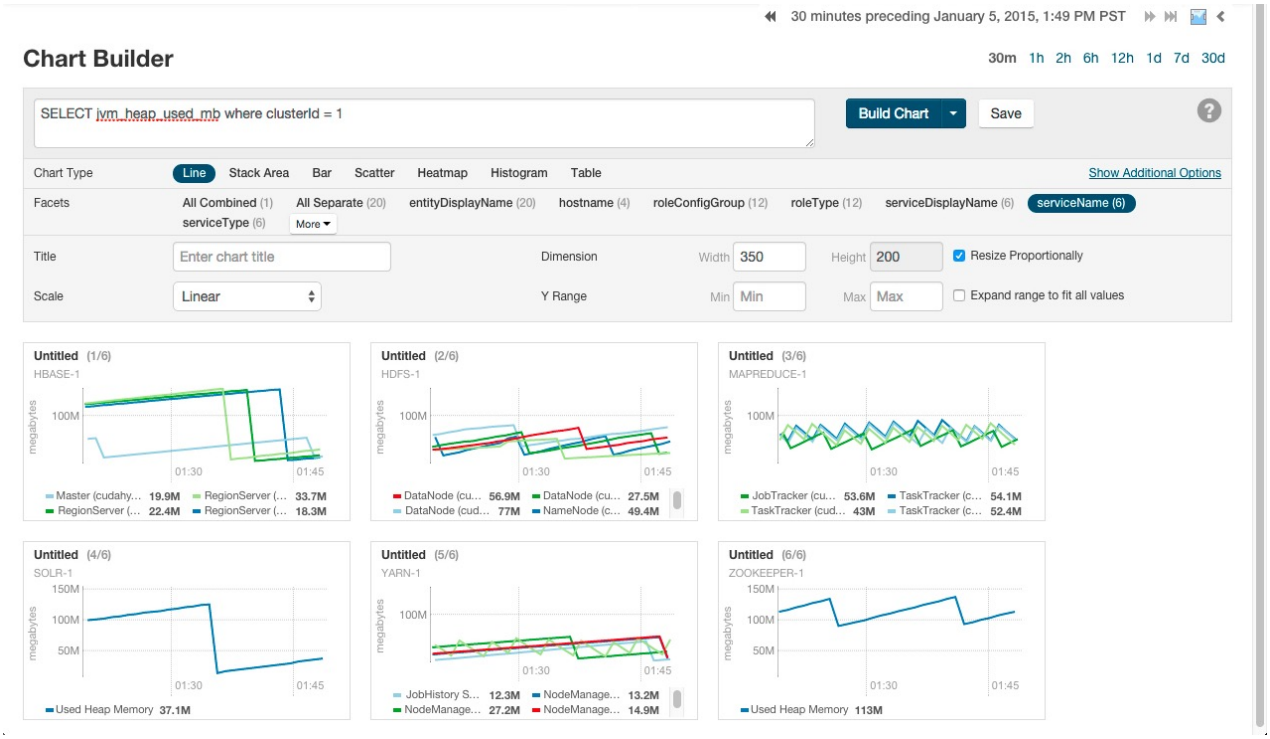
> **Note:** Heatmaps and histograms render charts for a single point as opposed to time series charts that render a series of points. For queries that return time series, Cloudera Manager will generate the heatmap or histogram based on the last recorded point in the series, and will issue the warning: "Query returned more than one value per stream. Only the last value was used." To eliminate this warning, use a scalar returning function to choose a point. For example, use `select last(cpu_percent)` to use the last point or `select max(cpu_percent)` to use the maximum value (in the selected time range).

Grouping (Faceting) Time Series

A time-series plot for a service, role, or host may actually be a composite of multiple individual time series. For example, the query `SELECT jvm_heap_used_mb where clusterId = 1` returns time-series data for the JVM heap used. Each time series has hostname, role type, metric, and entity name attributes. By default each attribute is displayed all on a single chart.

Using facets, you can combine time series based their attributes. To change the organization of the chart data, click one of the facets in the facet section in the upper part of the screen. The number in parentheses indicates how many charts will be displayed for that facet. As shown in the image below if the **serviceName** facet is selected for the JVM heap query, the time series is grouped into six charts, one chart each for each service name. The charts for service types with multiple roles contain multiple lines (for example, HBase, HDFS) while services that have only one role (for

example, ZooKeeper) contain just a single line. When a chart contains multiple lines, each entity is identified by a different color line.



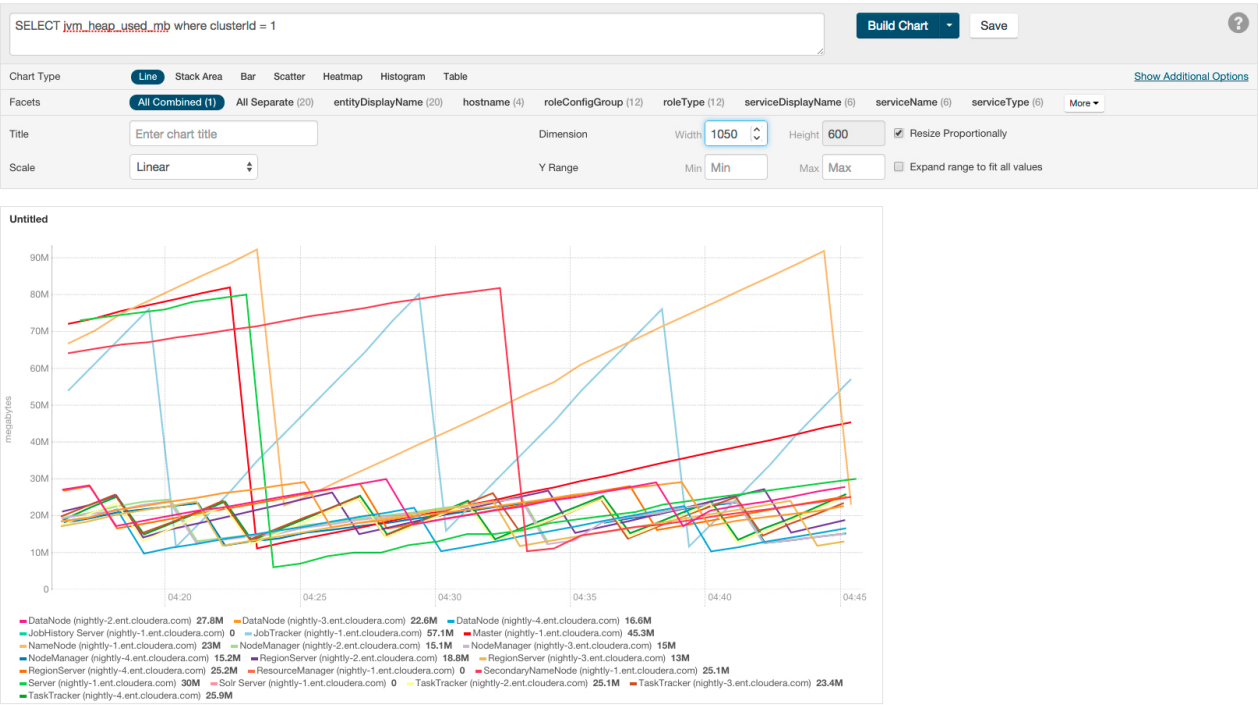
Changing Scale

You can set the scale of the chart to linear, logarithmic, and power.

Changing Dimensions

You can change the size of your charts by modifying the values in the **Dimension** fields. They change in 50-pixel increments when you click the up or down arrows, and you can type values in as long as they are multiples of 50. If you have multiple charts, depending on the dimensions you specify and the size of your browser window, your charts may appear in rows of multiple charts. If the **Resize Proportionally** checkbox is checked, you can modify one dimension and the other will be modified automatically to maintain the chart's width and height proportions.

The following chart shows the same query as the previous chart, but with **All Combined** selected (which shows all time series in a single chart) and with the Dimension values increased to expand the chart.



Changing Axes

You can change the Y-axis range using the **Y Range** minimum and maximum fields.

The X-axis is based on clock time, and by default shows the last hour of data. You can use the Time Range Selector or a duration link (30m 1h 2h 6h 12h 1d 7d 30d) to set the time range. (See Time Line on page 7 for details).

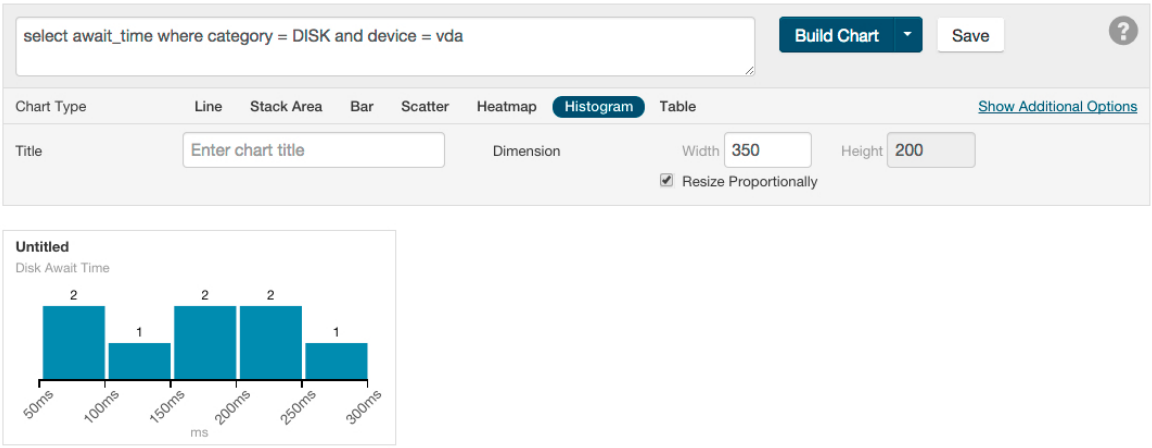
Displaying Chart Details

When you move your mouse over a chart, its background turns gray, indicating that you can act upon it.

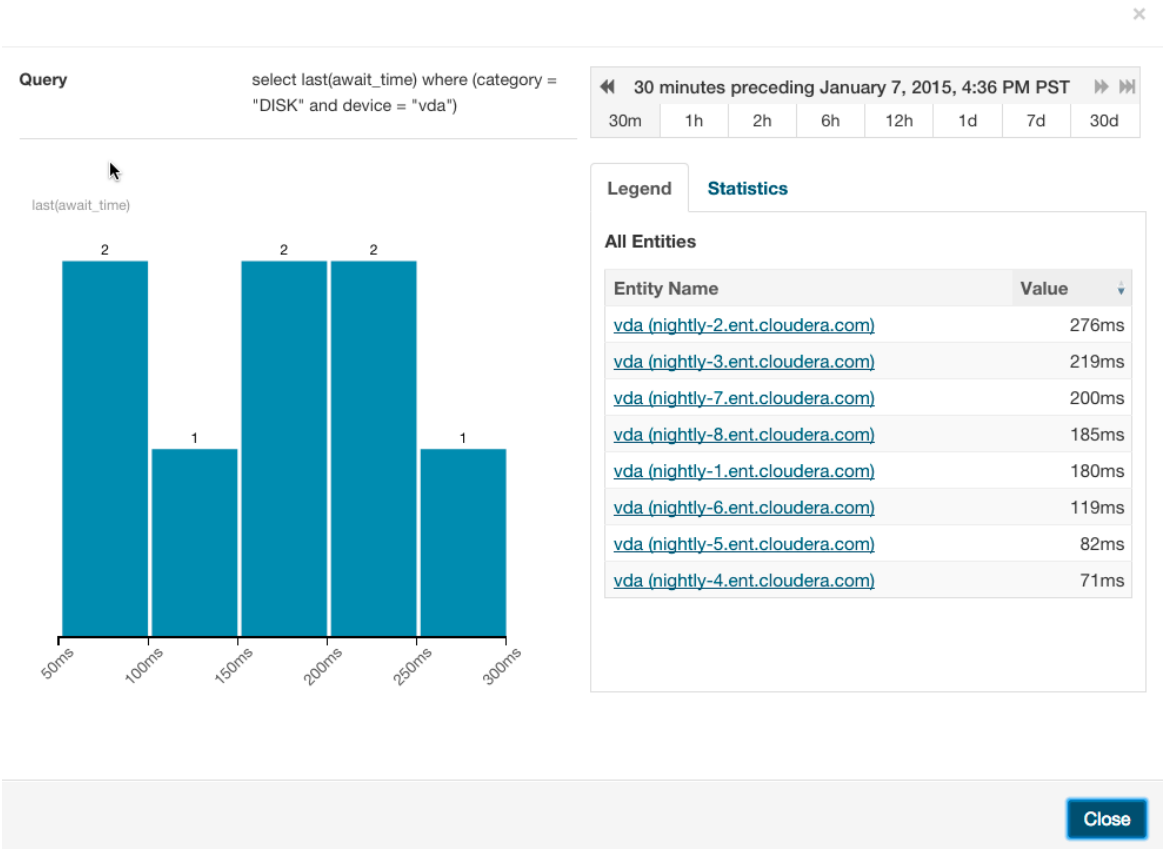
- Moving the mouse to a data point on a line, stack area, or bar chart shows the details about that data point in a pop-up tooltip.
- Click a line, stack area, scatter, or bar chart to expand it into a full-page view with a legend for the individual charted entities as well more fine-grained axes divisions.
  - If there are multiple entities in the chart, you can
    - Check and uncheck the legend item to hide or show the time series for the entities on the chart.

      ✓ vda (tcdn5-4.ent.cloudera.com), await_time View ✓ vda (tcdn5-3.ent.cloudera.com), await_time View

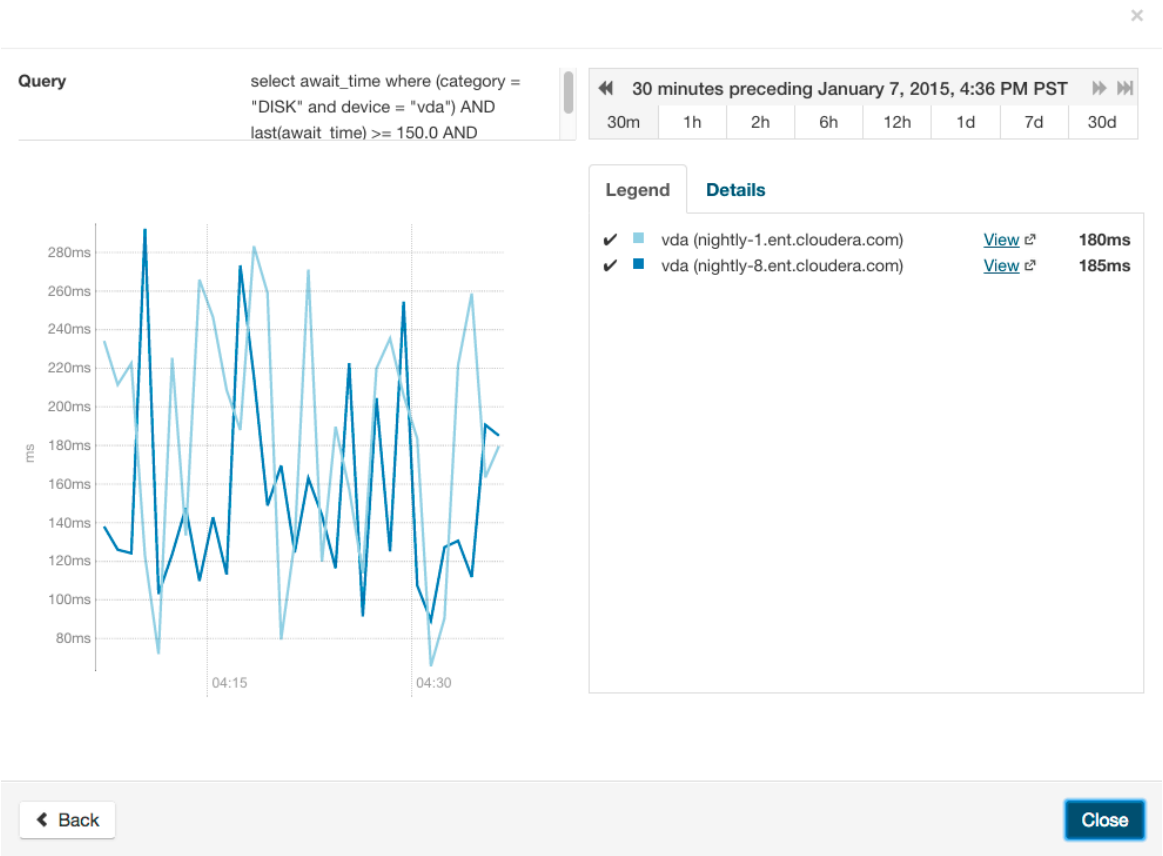
    - If there are service, role, or host instances in the chart, click the **View** link to display the instance's **Status** page.
  - Click the **Close** button to return to the regular chart view.
- **Heatmap** - Clicking a square in a heatmap displays a line chart of the time series for that entity.
- **Histogram** -
  - Mousing over the upper right corner of a histogram and clicking opens a pop-up containing the query that generated the chart, an expanded view of the chart, a list of entity names and links to the entities whose metrics are represented by the histogram bars, and the value of the metric for each entity. For example, clicking the following histogram



displays the following:



– Clicking a bar in the expanded histogram displays a line chart of the time series from which the histogram was generated:



Clicking the **< Back** link at the bottom left of the line chart returns to the expanded histogram.

Editing a Chart

You can edit a chart from the custom dashboard and save it back into the same or another existing dashboard, or to a new custom dashboard. Editing a chart only affects the copy of the chart in the current dashboard – if you have copied the chart into other dashboards, those charts are not affected by your edits.

1. Move the cursor over the chart, and click the gear icon
   ⚙
   at the top right.
2. Click **Open in Chart Builder**. This opens the **Chart Builder** page with the chart you selected already displayed.
3. Edit the chart's select statement and click **Build Chart**.

Saving a Chart

Minimum Required Role: Configurator (also provided by **Cluster Administrator, Full Administrator**)

After editing a chart you can save it to a new or existing custom dashboard.

1. Modify the chart's properties and click **Build Chart**.
2. Click **Save** to open the **Save Chart** dialog box, and select one of the following:
   a. Update chart in current dashboard: <name of current dashboard>.
   b. Add chart to another dashboard.
   c. Add chart to a new custom dashboard.
3. Click **Save Chart**.
4. Click **View Dashboard** to go to the dashboard where the chart has been saved.

See the following topics for more information:

- Saving Charts to a New Dashboard on page 109
- Saving Charts to an Existing Dashboard on page 110

Saving a chart only affects the copy of the chart in the dashboard where you save it – if you have previously copied the chart into other dashboards, those charts are not affected by your edits.

Users with Read-Only, Limited Operator, or Operator user roles can edit charts and view the results, but cannot save them to a dashboard.

Obtaining Time-Series Data Using the API

Time-series data can be obtained using the Cloudera Manager API. For details about using a tsquery statement to obtain time-series data, see the `/timeseries` API documentation at `http://cmServerHost:7180/static/apidocs/path__timeseries.html`. To see the API call that returns the time-series data for an existing chart, click the blue down-arrow at the upper-right corner of the chart and click **Export JSON**. A new web browser window opens, displaying the time-series data in JSON format. The query string of the URL for that window displays the API call that retrieved the time-series data.

Dashboards

A ***dashboard*** is a set of charts. This topic covers:

Dashboard Types

A ***default dashboard*** is a predefined set of charts that you cannot change. In a default dashboard you can:

- Display chart details.
- Edit a chart and then save back to a new or existing custom dashboard.

A ***custom dashboard*** contains a set of charts that you can change. In a custom dashboard you can:

- Display chart details.
- Edit a chart and then save back to a new or existing custom dashboard.
- Save a chart, make any modifications, and then save to a new or existing dashboard.
- Remove a chart.

When you first display a page containing charts it has a custom dashboard with the same charts as a default dashboard.

Creating a Dashboard

1. Do one of the following:
   - Select **Charts** > **New Dashboard**.
   - Select **Charts** > **Manage Dashboards** and click **Create Dashboard**.
   - Save a chart to a new dashboard.
2. Specify a name and optionally a duration.

   Conventional dashboard names follow the patterns given by one of the following Java regular expressions:

```
Pattern.compile("^(.+):(\\d):" + STATUS_VIEW_SUFFIX + "$");
```

```
Pattern.compile("^(.+):" + STATUS_VIEW_SUFFIX + "$");
```

   Examples of expected names are: `HOST:STATUS_VIEW`, `MGMT:STATUS_VIEW`, or `HDFS:5:STATUS_VIEW`. If the dashboard name does not match the expected pattern, a warning will be displayed in the server log.
3. Click **Create Dashboard**.

Managing Dashboards

To manage dashboards, select **Charts** > **Manage Dashboards**. You can create, clone, edit, export, import, and remove dashboards.

- **Create Dashboard** - create a new dashboard.
- **Clone** - clones an existing dashboard.
- **Edit** - edit an existing dashboard.
- **Export** - exports the specifications for the dashboard as a JSON file.
- **Import Dashboard** - reads an exported JSON file and recreates the dashboard.
- **Remove** - deletes the dashboard.

Configuring Dashboards

You can change the time scale of a dashboard, switch between default and custom dashboards, and reset a custom dashboard.

Setting the Time Scale of a Dashboard

By default the time scale of a dashboard is 30 minutes. To change the time scale, click a duration link

30m 1h 2h 6h 12h 1d 7d 30d

at the top-right of the dashboard.

Setting the Dashboard Type

To set the dashboard type, click  and select one of the following:

- **Custom** - displays a custom dashboard.
- **Default** - displays a default dashboard.
- **Reset** - resets the custom dashboard to the predefined set of charts, discarding any customizations.

Saving Charts to Dashboards

Minimum Required Role: **Configurator** (also provided by **Cluster Administrator, Full Administrator**)

You can save the charts and their configurations (type, dimension, and y-axis minimum and maximum) to a new dashboard or to an existing dashboard.

If your tsquery statement resulted in multiple charts, those charts are saved as a unit (either to a new or existing dashboard). You cannot edit the individual plots in that set of charts, but you can edit the set as a whole. A single edit button appears for the set that you saved — typically on the last chart in the set.

You can edit a copy of the individual charts in the set, but the edited copy does not change the original chart in the dashboard from which it was copied.

Saving Charts to a New Dashboard

1. Optionally modify the chart properties.
2. If the chart was created with the Chart Builder, optionally type a name for the chart in the Title field.
3. Do one of the following:
   - New chart - Click **Save**.
   - Existing chart - Move the cursor over the chart, and click the ⚙ icon at the top right.
4. Optionally edit the chart name.
5. Select the **Add chart to a new custom dashboard** option.
6. Enter a dashboard name.
7. Click **Save Chart**. The new dashboard appears on the menu under the top-level **Charts** tab.

Saving Charts to an Existing Dashboard

1. Optionally modify the chart properties.
2. If the chart was created with the Chart Builder, optionally type a name for the chart in the Title field.
3. Do one of the following:
   - New chart - Click **Save**.
   - Existing chart - Move the cursor over the chart, and click the ⚙ icon at the top right.
4. Optionally edit the chart name.
5. Select the **Add chart to an existing custom or system dashboard** option.
6. Select a dashboard from the **Dashboard Name** drop-down list.
7. Click **Save Chart**. The chart is added (appended) to the dashboard you select.

Adding a New Chart to the Home Page Custom Dashboard

You can add new charts to the custom dashboard on the **Home** > **Status** tab.

1. Click  and select **Add From Chart Builder** - displays the **Add Chart To** *Dashboard* page, with variables preset for the specific cluster where you want to add the dashboard.
   a. Click the question mark icon  to the right of the **Build Chart** button and select a metric from the **List of Metrics**, type a metric name or description into the **Basic** text field, or type a query into the **Advanced** field.
   b. Click **Build Chart**. The charts that result from your query are displayed, and you can modify their chart type, combine them using facets, change their size and so on.
2. Click **Add**.

Adding a New Chart to the Custom Dashboard

You can add new charts to the custom dashboard on the Status tab of a service, host, or role.

1. Click  and select one of the following:
   - **Add From Charts Library** - displays the charts page.
     1. Select one or more charts.
   - **Add From Chart Builder** - displays the **Add Chart To** *Dashboard* page, with variables preset for the specific service, role, or host where you want to add the dashboard.
     1. Click the question mark icon  to the right of the **Build Chart** button and select a metric from the **List of Metrics**, type a metric name or description into the **Basic** text field, or type a query into the **Advanced** field.
     2. Click **Build Chart**. The charts that result from your query are displayed, and you can modify their chart type, combine them using facets, change their size and so on.
2. Click **Add**.

Note: If the query you've chosen has resulted in multiple charts, all the charts are added to the dashboard as a set. Although the individual charts in this set can be copied, you can only edit the set as a whole.

Removing a Chart from a Custom Dashboard

Minimum Required Role: **Configurator** (also provided by **Cluster Administrator, Full Administrator**)

1. Move the cursor over the chart, and click the ⚙ icon at the top right.
2. Click **Remove**.

Moving and Resizing Charts on a Dashboard

You can move or resize the charts on a dashboard:

- Drag charts to a dashboard to change their relative positions.
- Change the size of a chart on a dashboard by dragging the lower-right corner of the chart.

tsquery Language

The tsquery language is used to specify statements for retrieving time-series data from the Cloudera Manager time-series datastore.

Before diving into the tsquery language specification, here's how you perform some common queries using the tsquery language:

1. Retrieve time series for all metrics for all DataNodes.

```
select * where roleType=DATANODE
```

2. Retrieve `cpu_user_rate` metric time series for all DataNodes.

```
select cpu_user_rate where roleType=DATANODE
```

3. Retrieve the `jvm_heap_used_mb` metric time series divided by 1024 and the `jvm_heap_committed` metric time series divided by 1024 for all roles running on the host named "my host".

```
select jvm_heap_used_mb/1024, jvm_heap_committed_mb/1024 where category=ROLE and
hostname="my host"
```

4. Retrieve the `jvm_total_threads` and `jvm_blocked_threads` metric time series for all entities for which Cloudera Manager collects these two metrics.

```
select jvm_total_threads,jvm_blocked_threads
```

tsquery Syntax

A tsquery statement has the following structure:

`SELECT [`*`metric expression`*`]WHERE [`*`predicate`*`]`

Note the following properties of tsquery statements:

- The statement `select *` is invalid.
- Tokens are case insensitive. For example, `Select`, `select`, and `SeLeCt` are all equivalent to `SELECT`.
- Multiple statements can be concatenated with semi-colons. Thus example 3 can be written as:

```
select jvm_heap_used_mb/1024 where category=ROLE and hostname=myhost; select
jvm_heap_commited_mb/1024 where category=ROLE and hostname=myhost
```

- The metric expression can be replaced with an asterisk (*), as shown in example 1. In that case, all metrics that are applicable for selected entities, such as `DATANODE` in example 1, are returned.
- The predicate can be omitted, as shown in example 4. In such cases, time series for all entities for which the metrics are appropriate are returned. For this query you would see the `jvm_new_threads` metric for NameNodes, DataNodes, TaskTrackers, and so on.

Metric Expressions

A ***metric expression*** generates the time series. It is a comma-delimited list of one or more metric expression statements. A ***metric expression statement*** is the name of a metric, a metric expression function, or a scalar value, joined by one or more metric expression operators.

See the FAQ on page 119 which answers questions concerning how to discover metrics and use cases for scalar values.

Metric expressions support the binary operators: `+`, `-`, `*`, `/`.

Here are some examples of metric expressions:

- `jvm_heap_used_mb, cpu_user, 5`
- `1000 * jvm_gc_time_ms / jvm_gc_count`
- `total_cpu_user + total_cpu_system`
- `max(total_cpu_user)`

Metric Expression Functions

Metric expressions support the functions listed in the following table. A function can return a time series or a scalar computed from a time series.

Functions that return scalars must be used for heatmap charts.

Function	Returns Scalar?	Description
`avg(`*`metric expression`*`)`	N	Computes a simple average for a time series.
`count_service_roles()`	Y	Returns the number of roles. There are three variants of this function: • `count_service_roles(roleType, roleState)` - Returns the number of roles of the specified `roleType` and `roleState`. For example, `count_service_roles(datanode, running)` returns the number of running DataNodes. • `count_service_roles(roleType)` - Returns the number of roles with the specified `roleType`. • `count_service_roles()` - Return the number of roles. For example, `select events_critical where count_service_roles() > 100` returns the `event_critical` metric when the number of roles is greater than 100.
`dt(`*`metric expression`*`)`	N	Derivative with negative values. The change of the underlying metric expression per second. For example: `dt(jvm_gc_count)`.
`dt0(`*`metric expression`*`)`	N	Derivative where negative values are skipped (useful for dealing with counter resets). The change of the underlying metric expression per second. For example: `dt0(jvm_gc_time_ms) / 10`.
`getClusterFact(string factName, double defaultValue)`	Y	Retrieves a fact about a cluster. Currently supports one fact: `numCores`. If the number of cores cannot be determined, `defaultValue` is returned.
`getHostFact(string factName, double defaultValue)`	Y	Retrieves a fact about a host. Currently supports one fact: `numCores`. If the number of cores cannot be determined, `defaultValue` is returned. For example, `select dt(total_cpu_user) / getHostFact(numCores, 2) where category=HOST` divides the results of `dt(total_cpu_user)` by the current number of cores for each host.

Function	Returns Scalar?	Description
		The following query computes the percentage of total user and system CPU usage each role is using on the host. It first computes the CPU seconds per second for the number of cores used by taking the derivative of the total user and system CPU times. It normalizes the result to the number of cores on the host by using the `getHostFact` function and multiplies the result by 100 to get the percentage. `select dt0(total_cpu_user)/getHostFact(numCores,1)*100, dt0(total_cpu_system)/getHostFact(numCores,1)*100 where category=ROLE and clusterId=1`
`greatest(`*`metric expression, scalar metric expression`*`)`	N	Compares two metric expressions, one of which one is a scalar metric expression. Returns a time series where each point is the result of evaluating `max(`*`point, scalar metric expression`*`)`.
`integral(`*`metric expression`*`)`	N	Computes the integral value for a stream and returns a time-series stream within which each data point is the integral value of the corresponding data point from the original stream. For example, `select integral(maps_failed_rate)` will return the count of the failed number of maps.
`counter_delta(`*`metric expression`*`)`	N	Computes the difference in counter value for a stream and returns a time-series stream within which each data point is the difference in counter value of the corresponding data point from the counter value of previous data point in the original stream. For example: `select counter_delta(maps_failed_rate)` returns the count of the failed number of maps. This method is more accurate than the `integral()` function. However there are a few caveats: • This function is only implemented for single time-series streams. For streams of cross-entity aggregates, continue to use the `integral()` function. • If you apply this method for time-series streams which was created using a version of Cloudera Manager older than 5.7, Cloudera Manager fills in the older data points using the `integral()` function.
`last(`*`metric expression`*`)`	Y	Returns the last point of a time series. For example, to use the last point of the `cpu_percent` metric time series, use the expression `select last(cpu_percent)`.
`least(`*`metric expression, scalar metric expression`*`)`	N	Compares two metric expressions, of which one is a scalar metric expression. Returns a time series where each point is the result of evaluating `min(`*`point, scalar metric expression`*`)`.
`max(`*`metric expression`*`)`	Y	Computes the maximum value of the time series. For example, `select max(cpu_percent)`.
`min(`*`metric expression`*`)`	Y	Computes the minimum value of the time series.
`moving_avg(`*`metric expression`*`, time_window_sec)`	N	Computes the moving average for a time series over a time window `time_window_sec` specified in seconds (2, 0.1, and so on)
`stats(`*`metric expression, stats name`*`)`	N	Some time-series streams have additional statistics for each data point. These include rollup time-series streams, cross-entity aggregates, and rate metrics. The following statistics are available for rollup and cross-entity aggregates: max, min, avg, std_dev, and sample. For rate metrics, the underlying counter value is available using the "counter" statistics. For

Function	Returns Scalar?	Description
		example, `stats(fd_open_across_datanodes, max)` or `stats(swap_out_rate, counter)`.
`sum(`*`metric expression`*`)`	Y	Computes the sum value of the time-series.

Predicates

A ***predicate*** limits the number of streams in the returned series and can take one of the following forms:

- *time_series_attribute operator value*, where
  - *time_series_attribute* is one of the supported attributes.
  - *operator* is one of = and `rlike`
  - *value* is an attribute value subject to the following constraints:
    - For attributes values that contain spaces or values of attributes of the form *xxx*`Name` such as `displayName`, use quoted strings.
    - The value for the `rlike` operator must be specified in quotes. For example: `hostname rlike "host[0-3]+.*"`.
    - *value* can be any regular expression as specified in regular expression constructs in the Java Pattern class documentation.

- *`scalar_producing_function`*`(`*`metric_expression`*`)` *`comparator number`*, where
  - *scalar_producing_function* is any function that takes a time series and produces a scalar. For example, `min` or `max`.
  - *metric_expression* is a valid metric expression. For example, `total_cpu_user + total_cpu_system`.
  - *comparator* is a comparison operator: <, <=, =, !=, >=, >.
  - *number* is any number expression or a number expression with units. For example, `5`, `5mb`, `5s` are all valid number expressions. The valid units are:
    - Time - `ms` (milliseconds), `s` (seconds), `m` (minutes), `h` (hours), and `d` (days).
    - Bytes -`b` (bytes), `kb` or `kib` (kilobytes), `mb` or `mib` (megabytes), `gb` or `gib` (gigabytes), `tb` or `tib` (terabytes), and `pb` or `pib` (petabytes)
    - Bytes per second - Bytes and Time: `bps`, `kbps`, `kibps`, `mbps`, `mibps`, and so on. For example, 5 kilobytes per second is `5 kbps`.
    - Bytes time - Bytes and Time combined: `bms`, `bs`, `bm`, `bh`, `bd`, `kms`, `ks`, and so on. For example, 5 kilobytes seconds is `5 ks` or `5 kis`.

You use the `AND` and `OR` operators to compose compound predicates.

Example Statements with Compound Predicates

1. Retrieve all time series for all metrics for DataNodes or TaskTrackers.

```
select * where roleType=DATANODE or roleType=TASKTRACKER
```

2. Retrieve all time series for all metrics for DataNodes or TaskTrackers that are running on host named "myhost".

```
select * where (roleType=DATANODE or roleType=TASKTRACKER) and hostname=myhost
```

3. Retrieve the `total_cpu_user` metric time series for all hosts with names that match the regular expression "host[0-3]+.*"

```
select total_cpu_user where category=role and hostname rlike "host[0-3]+.*"
```

Example Statements with Predicates with Scalar Producing Functions

1. Return the entities where the last count of Java VM garbage collections was greater than 10:

```
select jvm_gc_count where last(jvm_gc_count) > 10
```

2. Return the number of open file descriptors where processes have more than 500Mb of `mem_rss`:

```
select fd_open where min(mem_rss) > 500Mb
```

Filtering by Day of Week or Hour of Day

You can add an expression to the predicate of a tsquery statement that limits the stream to specified days of the week or to a range of hours in each day.

By Day – Limits the stream to selected days of the week.

The `day in ()` expression takes an argument with a comma-separated list of days of the week, enclosed in parentheses. The days of the week are numbered 1 through 7; 1 = Monday, 2 = Tuesday, and so on. Use the following syntax:

```
day in (#, #, ...)
```

For example, the following expression limits the stream to events that occurred only on weekdays:

```
day in (1,2,3,4,5)
```

By Hour – Limits the stream to a range of hours each day.

The `hour in` expression takes an argument with a range of hours separated by a colon and enclosed in square brackets. Valid values are integers 0–23:

```
hour in [#:#]
```

For example, the following expression limits the stream to events that occur only between 9:00 a.m. and 5:00 p.m.:

```
hour in [9:17]
```

Add the day or time range expression after the `WHERE` clause. Do not use the `AND` keyword. For example:

```
select fd_open where category = ROLE and roleType = SERVICEMONITOR day in (1,2,3,4,5)
```

You can also combine `day in` and `hour in` expressions. Always put the `day` expression before the `hour` expression. The following example limits the stream to weekdays between 9:00 a.m. and 5:00 p.m.:

```
select fd_open where category = ROLE and roleType = SERVICEMONITOR day in (1,2,3,4,5)
hour in [9:17]
```

Time Series Attributes

Attribute names and most attribute values are case insensitive. `displayName` and `serviceType` are two attributes whose values are *case sensitive*.

Name	Description
`active`	Indicates whether the entities to be retrieved must be active. A nonactive entity is an entity that has been removed or deleted from the cluster. The default is to retrieve only active entities (that is, `active=true`). To access time series for deleted or removed entities, specify `active=false` in the query. For example: `SELECT fd_open WHERE roleType=DATANODE and active=false`

Name	Description
agentName	A Flume agent name.
applicationName	One of the Cloudera Manager monitoring daemon names.
cacheId	The HDFS cache directive ID.

category	The category of the entities returned by the query: CLUSTER, DIRECTORY, DISK, FILESYSTEM, FLUME_SOURCE, FLUME_CHANNEL, FLUME_SINK, HOST, HTABLE, IMPALA_QUERY_STREAM, NETWORK_INTERFACE, ROLE, SERVICE, USER, YARN_APPLICATION_STREAM, YARN_QUEUE.
	Some metrics are collected for more than one type of entity. For example, total_cpu_user is collected for entities of category HOST and ROLE. To retrieve the data only for hosts use:
	`select total_cpu_user where category=HOST`
	The ROLE category applies to all role types (see roleType attribute). The SERVICE category applies to all service types (see serviceType attribute). For example, to retrieve the committed heap for all roles on host1 use:
	`select jvm_committed_heap_mb where category=ROLE and hostname="host1"`

clusterDisplayName	The user-defined display name of a cluster.
clusterName	The cluster ID. To specify the cluster by its display name, use the clusterDisplayName attribute.
componentName	A Flume component name. For example, channel1, sink1.
device	A disk device name. For example, sda.
displayName	The display name of an entity. **Note:** The displayName attribute was removed in Cloudera Manager 5. Older queries should be modified to use the cluster or service display name attributes, clusterDisplayName and serviceDisplayName.
entityName	A display name plus unique identifier. For example: HDFS-1-DATANODE-692d141f436ce70aac080aedbe83f887.
expired	A Boolean that indicates whether an HDFS cache directive expired.
groupName	A user group name.
hbaseNamespace	The name of the HBase namespace.
hostId	The canonical identifier for a host in Cloudera Manager. It is unique and immutable. For example: 3d645222-2f7e-4895-ae51-cd43b91f1e7a.
hostname	A hostname.
hregionName	The HBase region name. For example, 4cd887662e5c2f3cd5dd227bb03dd760.
hregionStartTimeMs	Milliseconds from UNIX epoch since Cloudera Manager monitoring started collecting metrics for the HBase region.
htableName	The name of an HBase table.
iface	A network interface name. For example, eth0.
logicalPartition	A Boolean indicating whether or not the disk is a logical partition. Applies to disk entity types.

Name	Description
mountpoint	A mount point name. For example, /var, /mnt/homes.
nameserviceName	The name of the HDFS nameservice.
ownerName	The owner username.
partition	A partition name. Applies to partition entity types.
path	A filesystem path associated with the time-series entity.
poolName	A pool name. For example, hdfs cache pool, yarn pools.
queueName	The name of a YARN queue.
rackId	A Rack ID. For example, /default.
roleConfigGroup	The role group that a role belongs to.
roleName	The role ID. For example, HBASE-1-REGIONSERVER-0b0ad09537621923e2b460e5495569e7.
roleState	The role state: BUSY, HISTORY_NOT_AVAILABLE, NA, RUNNING, STARTING, STOPPED, STOPPING, UNKNOWN
roleType	The role type: ACTIVITYMONITOR, AGENT, ALERTPUBLISHER, BEESWAX_SERVER, CATALOGSERVER, DATANODE, EVENTSERVER, FAILOVERCONTROLLER, HBASE_INDEXER, HBASERESTSERVER, HBASETHRIFTSERVER, HIVEMETASTORE, HIVESERVER2, HOSTMONITOR, HTTPFS, HUESERVER, IMPALAD, JOBHISTORY,JOBTRACKER, JOURNALNODE, KT_RENEWER, LLAMA, MASTER, NAVIGATOR, REGIONSERVER, SERVICEMONITOR, NAMENODE, NODEMANAGER, REPORTSMANAGER, SECONDARYNAMENODE, SERVER, SOLR_SERVER, SQOOP_SERVER, STATESTORE, TASKTRACKER.
rollup	The time-series store table rollup type.
schedulerType	The scheduler type associated with the pool service.
serviceDisplayName	The user-defined display name of a service entity.
serviceName	The service ID. To specify a service by its display name use the serviceDisplayName attribute.
serviceState	The service state: HISTORY_NOT_AVAILABLE, NA, RUNNING, STARTING, STOPPED, STOPPING, UNKNOWN
serviceType	The service type: ACCUMULO,FLUME, HDFS, HBASE, HIVE, HUE, IMPALA, KS_INDEXER, MAPREDUCE, MGMT, OOZIE,SOLR, SPARK,SQOOP,YARN, ZOOKEEPER.
solrCollectionName	The Solr collection name. For example, my_collection.
solrReplicaName	The Solr replica name. For example, my_collection_shard1_replica1.
solrShardName	The Solr shard name. For example, shard1.
systemTable	A boolean indicating whether the HBase table is a system table or not.
tableName	The name of a table.
userName	The name of the user.
version	The version of the cluster. The value can be any of the supported CDH major versions: 4 for CDH 4 and 5 for CDH 5.

Time Series Entities and their Attributes

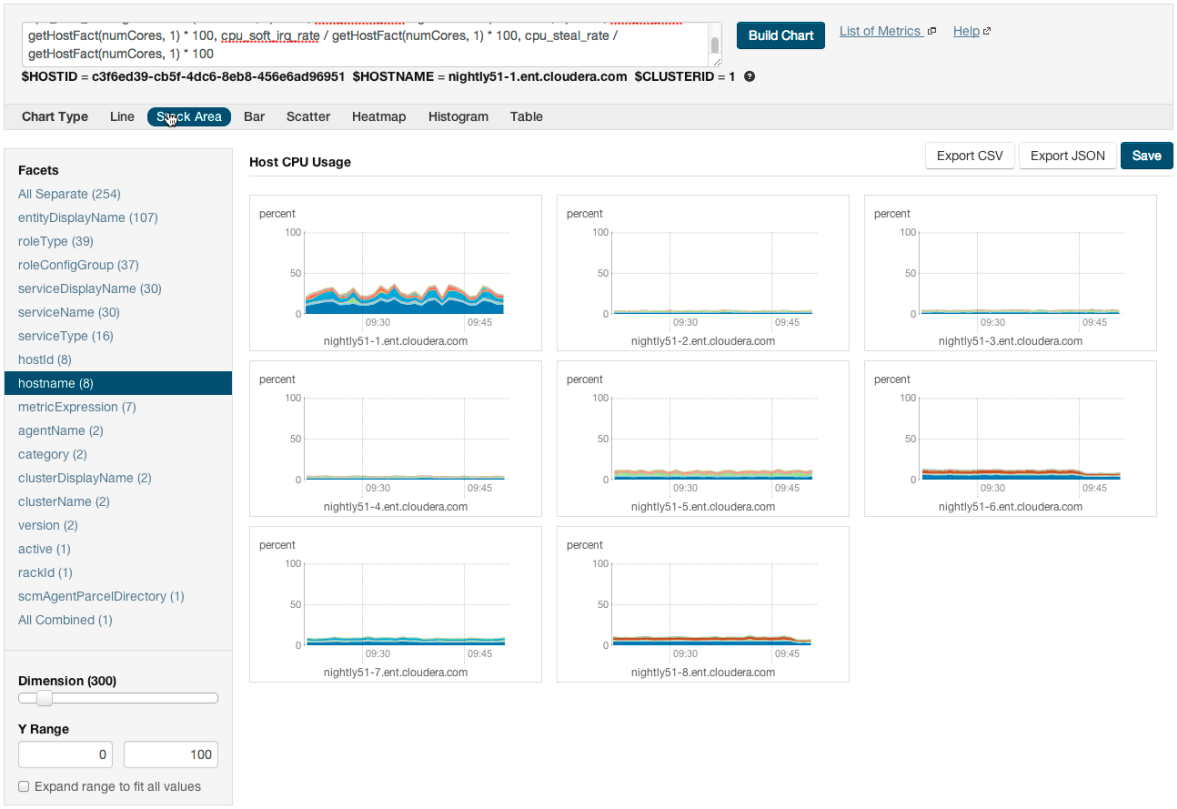
The following table shows the entities and associated attributes that can appear in the predicate ("where" clause) of a tsquery statement.

Entity	Attributes
All Roles	roleType, hostId, hostname, rackId, serviceType, serviceName
All Services	serviceName, serviceType, clusterId, version, serviceDisplayName, clusterDisplayName
Agent	roleType, hostId, hostname, rackId, serviceType, serviceName, clusterId, version, agentName, serviceDisplayName, clusterDisplayName
Cluster	clusterId, version, clusterDisplayName
Directory	roleName, hostId, path, roleType, hostname, rackId, serviceType, serviceName, clusterId, version, agentName, hostname, clusterDisplayName
Disk	device, logicalPartition, hostId, rackId, clusterId, version, hostname, clusterDisplayName
File System	hostId, mountpoint, rackId, clusterId, version, partition, hostname, clusterDisplayName
Flume Channel	serviceName, hostId, rackId, roleName, flumeComponent, roleType, serviceType, clusterId, version, agentName, serviceDisplayName, clusterDisplayName
Flume Sink	serviceName, hostId, rackId, roleName, flumeComponent, roleType, serviceType, clusterId, version, agentName, serviceDisplayName, clusterDisplayName
Flume Source	serviceName, hostId, rackId, roleName, flumeComponent, roleType, serviceType, clusterId, version, agentName, serviceDisplayName, clusterDisplayName
HDFS Cache Pool	serviceName, poolName, nameserviceName, serviceType, clusterId, version, groupName, ownerName, serviceDisplayName, clusterDisplayName
HNamespace	serviceName, namespaceName, serviceType, clusterId, version, serviceDisplayName, clusterDisplayName
Host	hostId, rackId, clusterId, version, hostname, clusterDisplayName
HRegion	htableName, hregionName, hregionStartTimeMs, namespaceName, serviceName, tableName, serviceType, clusterId, version, roleType, hostname, roleName, hostId, rackId , serviceDisplayName, clusterDisplayName
HTable	namespaceName, serviceName, tableName, serviceType, clusterId, version, serviceDisplayName, clusterDisplayName
Network Interface	hostId, networkInterface, rackId, clusterId, version, hostname, clusterDisplayName
Rack	rackId
Service	serviceName, serviceType, clusterId, serviceDisplayName
Solr Collection	serviceName, serviceType, clusterId, version, serviceDisplayName, clusterDisplayName
Solr Replica	serviceName, solrShardName, solrReplicaName, solrCollectionName, serviceType, clusterId, version, roleType, hostId, hostname, rackId, roleName, serviceDisplayName, clusterDisplayName
Solr Shard	serviceName, solrCollectionName, solrShardName, serviceType, clusterId, version, serviceDisplayName, clusterDisplayName
Time Series Table	tableName, roleName, roleType, applicationName, rollup, path
User	userName
YARN Pool	serviceName, queueName, schedulerType

FAQ

How do I compare information across hosts?

1. Click **Hosts** in the top navigation bar and click a host link.
2. In the Charts pane, choose a chart, for example **Host CPU Usage** and select ⚙ and then **Open in Chart Builder**.
3. In the text box, remove the `where entityName=$HOSTID` clause and click **Build Chart**.
4. In the Facets list, click **hostname** to compare the values across hosts.
5. Configure the time scale, minimums and maximums, and dimension. For example:



How do I compare all disk IO for all the DataNodes that belong to a specific HDFS service?

Use a query of the form:

```
select bytes_read, bytes_written where roleType=DATANODE and serviceName=hdfs1
```

replacing `hdfs1` with your HDFS service name. Then facet by **metricDisplayName** and compare all DataNode `byte_reads` and `byte_writes` metrics at once. See Grouping (Faceting) Time Series on page 103 for more details about faceting.

When would I use a derivative function?

Some metrics represent a counter, for example, `bytes_read`. For such metrics it is sometimes useful to see the rate of change instead of the absolute counter value. Use `dt` or `dt0` derivative functions.

When should I use the `dt0` function?

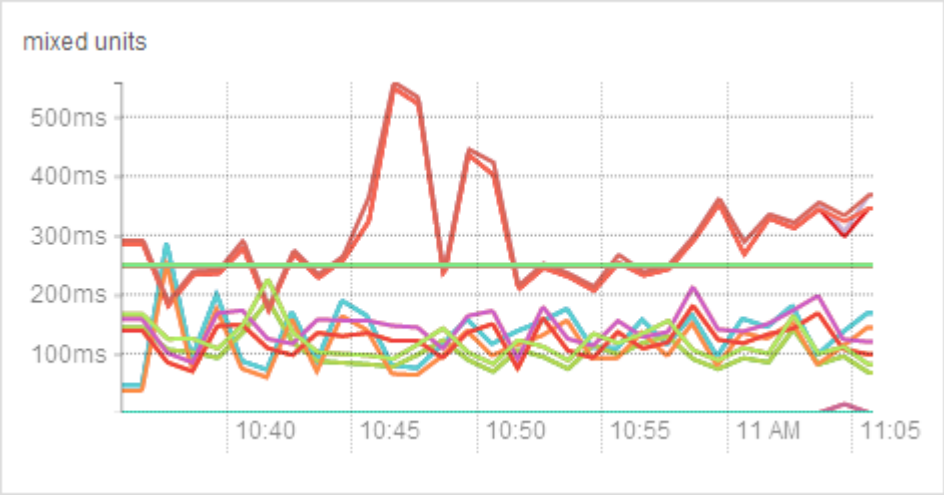
Some metrics, like `bytes_read` represent a counter that always grows. For such metrics a negative rate means that the counter has been reset (for example, process restarted, host restarted, and so on). Use `dt0` for these metrics.

How do I display a threshold on a chart?

Suppose that you want to retrieve the latencies for all disks on your hosts, compare them, and show a threshold on the chart to easily detect outliers. Use the following query to retrieve the metrics and the threshold:

```
select await_time, await_read_time, await_write_time, 250 where category=disk
```

Then choose **All Combined (1)** in the Facets list. The scalar threshold `250` will also be rendered on the chart:



See Grouping (Faceting) Time Series on page 103 for more details about faceting.

I get the warning "The query hit the maximum results limit". How do I work around the limit?

There is a limit on the number of results that can be returned by a query. When a query results in more time-series streams than the limit a warning for "partial results" is issued. To circumvent the problem, reduce the number of metrics you are trying to retrieve or see Configuring Time-Series Query Results on page 101.

You can use the `rlike` operator to limit the query to a subset of entities. For example, instead of

```
select await_time, await_read_time, await_write_time, 250 where category=DISK
```

you can use

```
select await_time, await_read_time, await_write_time, 250 where category=DISK and hostname
 rlike "host1[0-9]?.cloudera.com"
```

The latter query retrieves the disk metrics for ten hosts.

How do I discover which metrics are available for which entities?

- Type **Select** in the text box and then press **Space** or continue typing. Metrics matching the letters you type display in a drop-down list.
- Select **Charts** > **Chart Builder**, click the question mark icon
  
  to the right of the **Build Chart** button and click the **List of Metrics** link
- Retrieve all metrics for the type of entity:

```
select * where roleType=DATANODE
```

Metric Aggregation

In addition to collecting and storing raw metric values, the Cloudera Manager Service Monitor and Host Monitor produce a number of aggregate metrics from the raw metric data. Where a raw data point is a timestamp value pair, an aggregate metric point is a timestamp paired with a bundle of statistics including the minimum, maximum, average, and standard deviation of the data points considered by the aggregate.

Individual metric streams are aggregated across time to produce statistical summaries at different data granularities. For example, an individual metric stream of the number of open file descriptors on a host will be aggregated over time to the ten-minute, hourly, six-hourly, daily and weekly data granularities. A point in the hourly aggregate stream will include the maximum number of open file descriptors seen during that hour, the minimum, the average and so on. When servicing a time-series request, either for the Cloudera Manager UI or API, the Service Monitor and Host Monitor automatically choose the appropriate data granularity based on the time-range requested.

Cross-Time Aggregate Example

Consider the following fd_open raw metric values for a host:

```
9:00, 100 fds
9:01, 101 fds
9:02, 102 fds
. . .
9:09, 109 fds
```

The ten minutely cross-time aggregate point covering the ten-minute window from 9:00 - 9:10 would have the following statistics and metadata:

```
min: 100 fds
min timestamp: 9:00
max 109 fds
max timestamp 9:09
mean 104.5 fds
standard deviation: 3.02765 fds
count: 10 points
sample: 109 fds
sample timestamp: 9:09
```

The Service Monitor and Host Monitor also produce cross-entity aggregates for a number of entities in the system. Cross-entity aggregates are produced by considering the metric value of a particular metric across a number of entities of the same type at a particular time. For each stream considered, two metrics are produced. The first tracks statistics such as the minimum, maximum, average and standard deviation across all considered entities as well as the identities of the entities that had the minimum and maximum values. The second tracks the sum of the metric across all considered entities.

An example of the first type of cross-entity aggregate is the fd_open_across_datanodes metric. For an HDFS service this metric contains aggregate statistics on the fd_open metric value for all the DataNodes in the service. For a rack this metric contains statistics for all the DataNodes within that rack, and so on. An example of the second type of cross-entity aggregate is the total_fd_open_across_datanodes metric. For an HDFS service this metric contains the total number of file descriptors open by all the DataNodes in the service. For a rack this metric contains the total number of file descriptors open by all the DataNodes within the rack, and so on. Note that unlike the first type of cross-entity aggregate, this total type of cross-entity aggregate is a simple timestamp, value pair and not a bundle of statistics.

Cross-Entity Aggregate Example

Consider the following fd_open raw metric values for a set of ten DataNodes in an HDFS service at a given timestamp:

```
datanode-0, 200 fds
datanode-1, 201 fds
datanode-2, 202 fds
…
datanode-9, 209 fds
```

The cross-entity aggregate fd_open_across_datanodes point for that HDFS service at that time would have the following statistics and metadata:

```
min: 200 fds
min entity: datanode-0
max: 209 fds
max entity: datanode-9
mean: 204.5 fds
standard deviation: 3.02765 fds
count: 10 points
sample: 209 fds
sample entity: datanode-9
```

Just like every other metric, cross-entity aggregates are aggregated across time. For example, a point in the hourly aggregate of fd_open_across_datanodes for an HDFS service will include the maximum fd_open value of any

DataNode in that service over that hour, the average value over the hour, and so on. A point in the hourly aggregate of `total_fd_open_across_datanodes` for an HDFS service will contain statistics on the value of the `total_fd_open_across_datanodes` for that service over the hour.

Presentation of Aggregate Data

Aggregate data points returned from the Cloudera Manager API appear as shown in this section.

A cross-time aggregate:

```
{
        "timestamp" : "2014-02-24T00:00:00.000Z",
        "value" : 0.014541698027508003,
        "type" : "SAMPLE",
        "aggregateStatistics" : {
          "sampleTime" : "2014-02-23T23:59:35.000Z",
          "sampleValue" : 0.0,
          "count" : 360,
          "min" : 0.0,
          "minTime" : "2014-02-23T18:00:35.000Z",
          "max" : 2.9516129032258065,
          "maxTime" : "2014-02-23T19:37:36.000Z",
          "mean" : 0.014541698027508003,
          "stdDev" : 0.17041289765265377
        }
      }
```

A raw cross-entity aggregate:

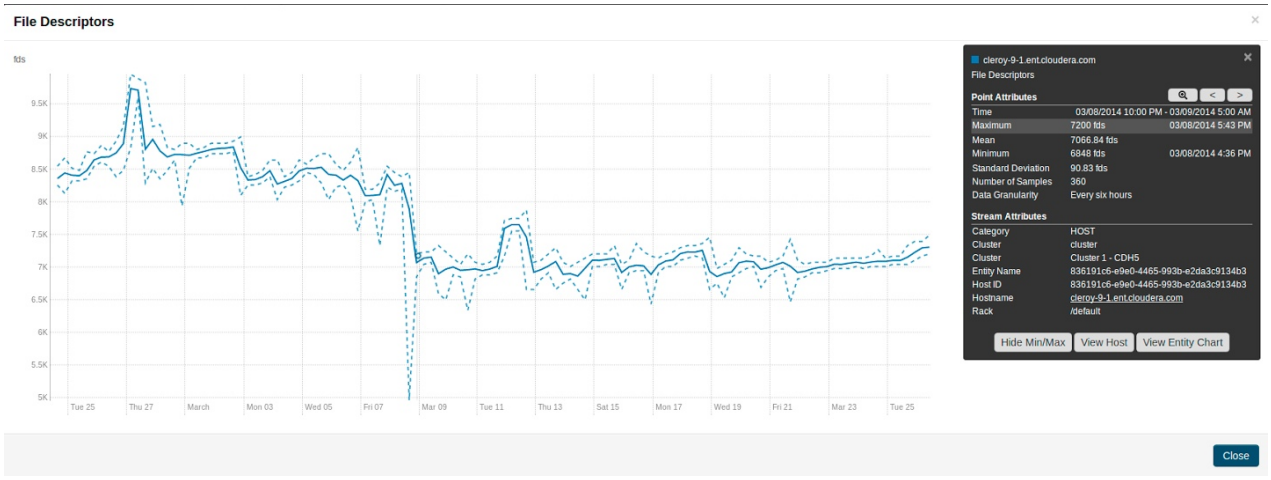
```
{
        "timestamp" : "2014-03-26T00:50:15.725Z",
        "value" : 3288.0,
        "type" : "SAMPLE",
        "aggregateStatistics" : {
          "sampleTime" : "2014-03-26T00:49:19.000Z",
          "sampleValue" : 7232.0,
          "count" : 4,
          "min" : 1600.0,
          "minTime" : "2014-03-26T00:49:42.000Z",
          "max" : 7232.0,
          "maxTime" : "2014-03-26T00:49:19.000Z",
          "mean" : 3288.0,
          "stdDev" : 2656.7549127961856,
          "crossEntityMetadata" : {
            "maxEntityDisplayName" : "cleroy-9-1.ent.cloudera.com",
            "minEntityDisplayName" : "cleroy-9-4.ent.cloudera.com",
            "numEntities" : 4.0
          }
        }
      }
```

A cross-time, cross-entity aggregate:

```
{
        "timestamp" : "2014-03-11T00:00:00.000Z",
        "value" : 3220.818863879957,
        "type" : "SAMPLE",
        "aggregateStatistics" : {
          "sampleTime" : "2014-03-10T22:28:48.000Z",
          "sampleValue" : 7200.0,
          "count" : 933,
          "min" : 1536.0,
          "minTime" : "2014-03-10T21:02:17.000Z",
          "max" : 7200.0,
          "maxTime" : "2014-03-10T22:28:48.000Z",
          "mean" : 3220.818863879957,
          "stdDev" : 2188.6143063503378,
          "crossEntityMetadata" : {
            "maxEntityDisplayName" : "cleroy-9-1.ent.cloudera.com",
            "minEntityDisplayName" : "cleroy-9-4.ent.cloudera.com",
```

```
            "numEntities" : 3.9787037037037036
          }
        }
      }
```

These differ from non-aggregate data points by having the aggregateStatistics structure. Note that the value field in the point structure will always be the same as the aggregteStatistics mean field. The Cloudera Manager UI presents aggregate statistics in a number of ways. First, aggregate statistics are made available in the hover detail and chart popover when dealing with aggregate data. Second, it is possible to turn on and turn off the display of minimum and maximum time-series streams in line charts of aggregate data. These streams are displayed using dotted lines and give a visual indication of the underlying metric values data range over the time considered, entities considered or both. These lines are displayed by default for single stream line charts of aggregate data. For all line charts this behavior can be turned on and turned off using the chart popover.



Accessing Aggregate Statistics Through tsquery

The stats function can be used to access aggregate statistics directly in tsquery. For example, `select stats(fd_open_across_datanodes, max) where category = service and serviceDisplayName = "my-hdfs-service"` will return a single time-series stream containing the just the maximum statistic values from the `fd_open_across_datanodes` stream. The following statistics are available through the stats function: `min`, `max`, `avg`, `std_dev`, and `sample`. See tsquery Language for more details on the stats function.

Logs

The Logs page presents log information for Hadoop services, filtered by service, role, host, or search phrase as well log level (severity).

To configure logs, see Configuring Log Events on page 20.

Viewing Logs

1. Select **Diagnostics** > **Logs** on the top navigation bar.
2. Click **Search**.

The logs for all roles display. If any of the hosts cannot be searched, an error message notifies you of the error and the host(s) on which it occurred.

Logs List

Log results are displayed in a list with the following columns:

- **Host** - The host where this log entry appeared. Clicking this link will take you to the Host Status page (see Host Details on page 47).

- **Log Level** - The log level (severity) associated with this log entry.
- **Time** - The date and time this log entry was created.
- **Source** - The class that generated the message.
- **Message** - The message portion of the log entry. Clicking **View Log File** displays the Log Details on page 124 page, which presents a display of the full log, showing the selected message (highlighted) and the 100 messages before and after it in the log.

If there are more results than can be shown on one page (per the Results per Page setting you selected), **Next** and **Prev** buttons let you view additional results.

Filtering Logs

You filter logs by selecting a time range and specifying filter parameters.

You can use the Time Range Selector or a duration link (30m 1h 2h 6h 12h 1d 7d 30d) to set the time range. (See Time Line on page 7 for details). However, logs are, by definition, historical, and are meaningful only in that context. So the Time Marker, used to pinpoint status at a specific point in time, is not available on this page. The Now button (⏭) is available.

1. Specify any of the log filter parameters:
   - **Search Phrase** - A string to match against the log message content. The search is case-insensitive, and the string can be a regular expression, such that wildcards and other regular expression primitives are supported.
   - **Select Sources** - A list of all the service instances and roles currently instantiated in your cluster. By default, all services and roles are selected to be included in your log search; the All Sources checkbox lets you select or clear all services and roles in one operation. You can expand each service and limit the search to specific roles by selecting or clearing individual roles.
   - **Hosts** - The hosts to be included in the search. As soon as you start typing a hostname, Cloudera Manager provides a list of hosts that match the partial name. You can add multiple names, separated by commas. The default is to search all hosts.
   - **Minimum Log Level** - The minimum severity level for messages to be included in the search results. Results include all log entries at the selected level or higher. This defaults to WARN (that is, a search will return log entries with severity of WARN, ERROR, or FATAL only.
   - **Additional Settings**
     - **Search Timeout** - A time (in seconds) after which the search will time out. The default is 20 seconds.
     - **Results per Page** - The number of results (log entries) to be displayed per page.
2. Click **Search**. The Logs list displays the log entries that match the specified filter.

Log Details

The Log Details page presents a portion of the full log, showing the selected message (highlighted), and messages before and after it in the log. The page shows you:

- The host
- The role
- The full path and name of the log file you are viewing.
- Messages before and after the one you selected.

The log displays the following information for each message:

- Time - the time the entry was logged
- Log Level - the severity of the entry
- Source - the source class that logged the entry
- Log Message

You can switch to display only messages or all columns using the  buttons.

In addition, from the Log Details page you can:

- View the log entries in either expanded or contracted form using the buttons to the left of the date range at the top of the log.
- Download the full log using the **Download Full Log** button at the top right of the page.
- View log details for a different host or for a different role on the current host, by clicking the **Change...** link next to the host or role at the top of the page. In either case this shows a pop-up where you can select the role or host you want to see.

Viewing the Cloudera Manager Server Log

To help you troubleshoot problems, you can view the Cloudera Manager Server log. You can view the logs in the Logs page or in specific pages for the log.

Viewing Cloudera Manager Server Logs in the Logs Page

1. Select **Diagnostics** > **Logs** on the top navigation bar.
2. Click **Select Sources** to display the log source list.
3. Uncheck the **All Sources** checkbox.
4. Click ▸ to the left of Cloudera Manager and select the **Server** checkbox.
5. Click **Search**.

For more information about the Logs page, see Logs on page 123.

Viewing the Cloudera Manager Server Log

1. Select **Diagnostics** > **Server Log** on the top navigation bar.

Note: You can also view the Cloudera Manager Server log at `/var/log/cloudera-scm-server/cloudera-scm-server.log` on the Server host.

Viewing the Cloudera Manager Agent Logs

To help you troubleshoot problems, you can view the Cloudera Manager Agent logs. You can view the logs in the Logs page or in specific pages for the logs.

Viewing Cloudera Manager Agent Logs in the Logs Page

1. Select **Diagnostics** > **Logs** on the top navigation bar.
2. Click **Select Sources** to display the log source list.
3. Uncheck the **All Sources** checkbox.
4. Click ▸ to the left of Cloudera Manager and select the **Agent** checkbox.
5. Click **Search**.

For more information about the Logs page, see Logs on page 123.

Viewing the Cloudera Manager Agent Log

1. Click the **Hosts** tab.
2. Click the link for the host where you want to see the Agent log.
3. In the **Details** panel, click the **Details** link in the **Host Agent** field.
4. Click the **Agent Log** link.

You can also view the Cloudera Manager Agent log at `/var/log/cloudera-scm-agent/cloudera-scm-agent.log` on the Agent hosts.

Managing Disk Space for Log Files

All CDH cluster hosts write out separate log files for each role instance assigned to the host. Cluster administrators can monitor and manage the disk space used by these roles and configure log rotation to prevent log files from consuming too much disk space.

Disk Space Requirements

For each role assigned to a host, you should generally provision 2GB of disk space for log files. This recommendation is based on the default values of configuration properties that set the maximum log file size (200MB) and the maximum number of files (10). To calculate the disk space required for each host, multiply the configured maximum size of the log file by the configured maximum number of logs. Perform this calculation for each role on a host and add them together. (Note that Gateway roles do not generate log files.)

To determine the roles assigned to each host, open the Cloudera Manager Admin Console and go to **Hosts** > **All Hosts** and expand the list of roles in the **Roles** column.

Managing Log Files

To manage log file configurations for all role instances of a service:

1. Go to ***Service Name*** > **Configuration**.
2. Select **Category** > **Logs**.
3. Edit the logging parameters.
4. Click **Save Changes**.

> **Note:** You can also manage these configurations using role groups, which you can use to configure similar hosts with the same configuration values. See Managing Roles.

There are the parameters you use to manage log files:

Table 9: Log File Properties

Property	Description	Default Value
Role Type **Max Log Size**	Maximum size for a log file before the log file rolls over into a new file.	200 MB
Role Type **Maximum Log File Backups**	Maximum number of rolled-over log files to retain.	10
Role Type **Log Directory**	The path to the directory where the log files are saved.	`/var/log/log_file_name`
Role Type **Logging Threshold** (not available for all roles)	Logging level to limit the number of entries saved in the log file.	Depends on the role.

Reports

> **Important:** This feature requires a Cloudera Enterprise license. It is not available in Cloudera Express. See Managing Licenses for more information.

The **Reports** page lets you create reports about the usage of HDFS in your cluster—data size and file count by user, group, or directory. It also lets you report on the MapReduce activity in your cluster, by user.

To display the **Reports** page, select **Clusters** > ***Cluster name*** > **Reports**.

For users with the Administrator role, the **Search Files and Manage Directories** button on the **Reports** page opens a file browser for searching files, managing directories, and setting quotas.

If you are managing multiple clusters, or have multiple nameservices configured (if high availability or federation is configured) there will be separate reports for each cluster and nameservice.

Directory Usage Report

Minimum Required Role: BDR Administrator (also provided by **Full Administrator** and **Cluster Administrator**)

The directory usage report allows you to browse the HDFS filesystem in a way that is similar to the HDFS File Browser. However, the **Directory Usage Report** also allows you to sort the listings and select multiple items and perform actions on them. You can also view the last access time, the last modified time of any file in a directory, and the total size of all files in the directory. This usage information is updated on an hourly basis.

You can customize the report by adding filters. A number of preconfigured filters are available, and you can create a custom filter.

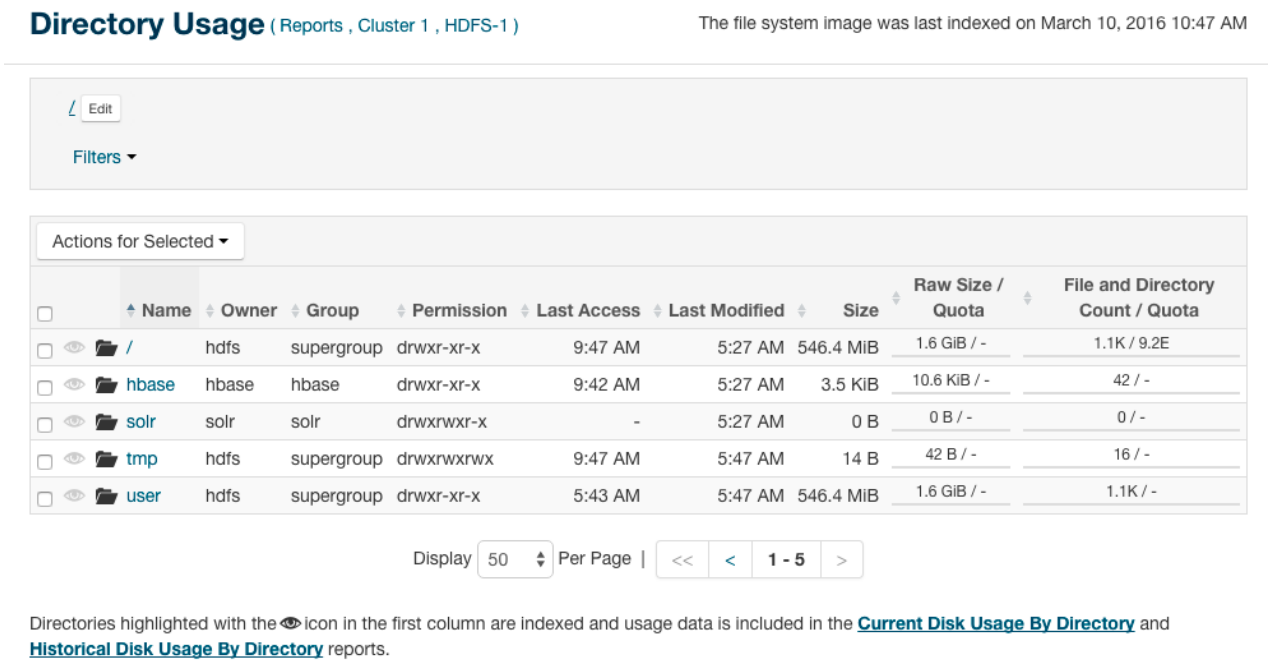
Accessing the Directory Usage Report

To view the directory usage report, click **Clusters** > ***Cluster Name*** > **Reports** > **Directory Usage**.

You can also access this report from the HDFS File Browser. (To access the File Browser, click **Clusters** > ***HDFS service*** > **File Browser**.) Click the **Directory Usage** link located in the lower-right portion of the File Browser.

Using the Directory Usage Report

When you first open the report, the top level of the HDFS filesystem displays:



Click the **Reports** link next to the **Directory Usage** title to go back to the **Reports** menu. You can also click links to go to the cluster and HDFS service home pages.

Click any column header to sort the display.

Click a directory name to view the files and subdirectories in the directory.

Select one or more rows by checking the boxes on the left and then choose an action to perform on the selection from the **Actions for Selected** drop-down menu. You can select the following actions:

- **Manage Quota** – A dialog box opens in which you can set a quota for the number of files or disk space. These values are displayed in columns in the file listing.
- **Include selected directories in disk usage reports** – The selected directories appear in the Disk Usage Reports.
- **Exclude selected directories from disk usage reports** – The selected directories do not appear in the Disk Usage Reports.

Filters

You can use filters to limit the display and to search for files. To apply filters to the directory usage report, click the **Filters** drop-down menu near the top of the page and select one of the following preconfigured filters:

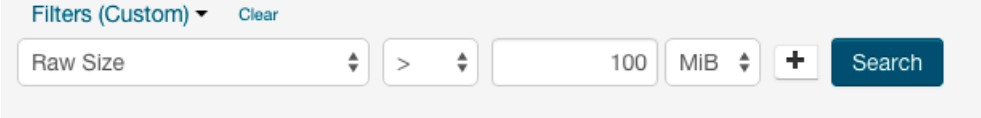
- Large Files
- Large Directories
- By Specific Owner
- By Specific Group
- Old Files
- Old Directories
- Files with Low Replication
- Overpopulated Directories
- Directories with Quotas
- Directories Watched

To modify any of these filters, click the **Customize** link and select new criteria. Click **Clear** to revert to the preconfigured criteria for the filter.

Click the **Search** button to display the report with the filters applied.

You can also select **Custom** from the **Filters** drop-down menu to create a report in which you define the criteria. To create a custom report:

1. Select any of the following criteria from the drop-down menu on the left:
   - Filename
   - Owner
   - Group
   - Path
   - Last Modified
   - Size
   - Diskspace Quota
   - Namespace Quota
   - Last Access
   - File and Directory Count
   - Replication
   - Parent
   - Raw Size
2. Select an operator from the drop-down menu.
3. Enter a value and units of measure for the comparison.
4. Select the units of measure for the comparison from the drop-down menu. (Some criteria do not require units of measure.)
5. Click the + icon to add additional criteria.
6. Click the **Search** button to display the directory usage report with the custom filter applied.



The report changes to display the result of applying the filter. A new column, **Parent** is added that contains the full path to each file or subdirectory.

Disk Usage Reports

There are two types of disk usage reports: **Current Disk Usage By Directory** and **Historical Disk Usage By Directory**. To use these reports, select one or more directories to watch by clicking the icon for the directory. You can also select multiple directories, and then click **Actions for Selected** > **Include selected directories in disk usage reports**.

For information on using and configuring the Disk Usage Report, see Disk Usage Reports on page 129.

Disk Usage Reports

The following reports show HDFS disk usage statistics, either current or historical, by user, group, or directory.

The **By Directory** reports display information about the directories in the Watched list, so if you are not watching any directories there will be no results found for these reports. You can also specify which directories to watch by selecting them from the Directory Usage Report on page 127.

Viewing Current Disk Usage by User, Group, or Directory

These reports show "current" disk usage in both chart and tabular form. The data for these reports comes from the `fsimage` kept on the NameNode, so the data in a report will be only as current as when the last checkpoint was performed. Typically the checkpoint interval is (by default) once per hour, but if checkpoints are not being performed as frequently, the disk usage report may not be up to date. The disk usage report displays the current usage and does not account for deleted files that only exist in snapshots. These files are included in the usage information when you run the du command.

To create a disk usage report:

- Click the report name (link) to produce the resulting report.

Each of these reports show:

Bytes	The logical number of bytes in the files, aggregated by user, group, or directory. This is based on the actual files sizes, not taking replication into account.
Raw Bytes	The physical number of bytes (total disk space in HDFS) used by the files aggregated by user, group, or directory. This does include replication, and so is actually Bytes times the number of replicas.
File and Directory Count	The number of files aggregated by user, group, or directory.

Bytes and Raw Bytes are shown in IEC binary prefix notation (1 GiB = $1 * 2^{30}$).

The directories shown in the **Current Disk Usage by Directory** report are the HDFS directories you have set as watched directories. You can add or remove directories to or from the watch list from this report; click the **Search Files and Manage Directories** button at the top right of the set of reports for the cluster or nameservice (see Designating Directories to Include in Disk Usage Reports on page 131).

The report data is also shown in chart format:

- Move the cursor over the graph to highlight a specific period on the graph and see the actual value (data size) for that period.
- You can also move the cursor over the user, group, or directory name (in the graph legend) to highlight the portion of the graph for that name.
- You can right-click within the chart area to save the whole chart display as a single image (a .PNG file) or as a PDF file. You can also print to the printer configured for your browser.

Viewing Historical Disk Usage by User, Group, or Directory

You can use these reports to view disk usage over a time range you define. You can have the usage statistics reported per hour, day, week, month, or year.

To create one of these reports:

- Click the report name (link) to produce the initial report. This generates a report that shows Raw Bytes for the past month, aggregated daily.

To change the report parameters:

- Select the **Start Date** and **End Date** to define the time range of the report.
- Select the **Graph Metric** you want to graph: bytes, raw bytes, or files and directories count.
- In the **Report Period** field, select the period over which you want the metrics aggregated. The default is Daily. This affects both the number of rows in the results table, and the granularity of the data points on the graph.
- Click **Generate Report** to produce a new report.

As with the current reports, the report data is also presented in chart format, and you can use the cursor to view the data shown on the charts, as well as save and print them.

For weekly or monthly reports, the Date indicates the date on which disk usage was measured.

The directories shown in the **Historical Disk Usage by Directory** report are the HDFS directories you have set as watched directories (see Designating Directories to Include in Disk Usage Reports on page 131).

Downloading Reports as CSV and XLS Files

Any report can be downloaded to your local system as an XLS file (Microsoft Excel 97-2003 worksheet) or CSV (comma-separated value) text file.

To download a report, do one of the following:

- From the main page of the Report tab, click CSV or XLS link next to in the column to the right of the report name
- From any report page, click the **Download CSV** or **Download XLS** buttons.

Either of these opens the Open file dialog box where you can open or save the file locally.

Activity, Application, and Query Reports

The Reports page contains links for displaying metrics on the following types of activities in your cluster:

- Disk usage
- MapReduce jobs
- YARN applications
- Impala queries
- HBase tables and namespaces

To view the Reports page, click **Clusters** > ***ClusterName*** > **Reports**. You can generate a report to view aggregate job activity per hour, day, week, month, or year, by user or for all users.

1. Click the **Start Date** and **End Date** fields and choose a date from the date control.
2. In the **Report Period** drop-down, select the period over which you want the metrics aggregated. Default is Daily.
3. Click **Generate Report**.

For weekly reports, the Date column indicates the year and week number (for example, 2013-01 through 2013-52). For monthly reports, the Date column indicates the year and month by number (2013-01 through 2013-12).

The File Browser

Minimum Required Role: BDR Administrator (also provided by **Full Administrator** and **Cluster Administrator**)

The **File Browser** tab on the HDFS service page lets you browse and search the HDFS namespace and manage your files and directories. The File Browser page initially displays the root directory of the HDFS file system in the gray panel

at the top and its immediate subdirectories below. Click any directory to drill down into the contents of that directory or to select that directory for available actions.

Searching Within the File System

To search the file system, click **Custom report** in the **Reports** section. The **Choose** drop down lets you select from custom search criteria such as filename, owner, file size, and so on. The file and directory listings are taken from the `fsimage` stored on the NameNode, so the listings will be only as current as the last checkpoint. Typically the checkpoint interval is (by default) once per hour, but if checkpoints are not being performed as frequently, the listings may not be up to date.

To search the file system:

1. From the HDFS service page, select the **File Browser** tab.
2. Click **Choose** and do one of the following:
   - Select a predefined query. Depending on what you select, you may be presented with different fields to fill in or different views of the file system. For example, selecting **Size** will provide a choice of arithmetic operators and fields where you provide the size to be used as the search criteria.
     1. Select a property in the **Choose...** drop-down.
     2. Select an operator.
     3. Specify a value.
     4. Click ✚ to add another criteria (all of which must be satisfied for a file to be considered a match) and repeat the preceding steps.
3. Click the **Generate Report** button to generate a custom report containing the search results.

If you search within a directory, only files within that directory will be found. For example, if you browse `/user` and do a search, you might find `/user/foo/file`, but you will not find `/bar/baz`.

Enabling Snapshots

To enable snapshots for an HDFS directory and its contents, see Managing HDFS Snapshots.

Setting Quotas

To set quotas for an HDFS directory and its contents, see Setting HDFS Quotas.

Designating Directories to Include in Disk Usage Reports

1. To add or remove directories from the directory-based Disk Usage reports, navigate through the file system to see the directory you want to add. You can include a directory at any level without including its parent.
2. Check the checkbox **Include this directory in Disk Usage reports**. As long as the checkbox is checked, the directory appears in the usage reports. To discontinue inclusion of the directory in Disk Usage reports, clear the checkbox.

Downloading HDFS Directory Access Permission Reports

The Directory Access By Group feature in the User Access category on the Reports page is a Cloudera data management feature.

Minimum Required Role: **Cluster Administrator** (also provided by **Full Administrator**)

For each HDFS service, you can download a report that details the HDFS directories a group has permission to access.

1. In the Cloudera Manager Admin Console, click **Clusters** > ***ClusterName*** > **Reports**.
2. In the Directory Access by Group row, click **CSV** or **XLS**. The Download User Access Report pop-up displays.
   a. In the pop-up, type a group and directory.
   b. Click **Download**. A report of the selected type will be generated containing the following information – path, owner, permissions, and size – for each directory contained in the specified directory that the specified group has access to.

Troubleshooting Cluster Configuration and Operation

This section contains solutions to some common problems that prevent you from using Cloudera Manager and describes how to use Cloudera Manager log and notification management tools to diagnose problems.

Solutions to Common Problems

Symptom	Reason	Solution
Cloudera Manager		
The Cloudera Manager service will not be running as it exited abnormally. Running `service cloudera-scm-server status` will print following message "cloudera-scm-server dead but pid file exists". The Cloudera Manager Server log file `/var/log/cloudera-scm-server/cloudera-scm-server.log` will have a stacktrace with "`java.lang.OutOfMemoryError`" logged.	Out of memory.	Examine the heap dump that the Cloudera Manager Server creates when it runs out of memory. The heap dump file is created in the `/tmp` directory, has file extension `.hprof` and file permission of 600. Its owner and group will be the owner and group of the Cloudera Manager server process, normally `cloudera-scm:cloudera-scm`.
You are unable to start service on the Cloudera Manager server, that is, `service cloudera-scm-server start` does not work and there are errors in the log file located at `/var/log/cloudera-scm-server/cloudera-scm-server.log`	The server has been disconnected from the database or the database has stopped responding or has shut down.	Go to `/etc/cloudera-scm-server/db.properties` and make sure the database you are trying to connect to is listed there and has been started.
Logs include `APPARENT DEADLOCK` entries for c3p0.	These deadlock messages are cause by the c3p0 process not making progress at the expected rate. This can indicate either that c3p0 is deadlocked or that its progress is slow enough to trigger these messages. In many cases, progress is occurring and these messages should not be seen as catastrophic.	There are a variety of ways to react to these log entries. • You may ignore these messages if system performance is not otherwise affected. Because these entries often occur during slow progress, they may be ignored in some cases. • You may modify the timer triggers. If c3p0 is making slow progress, increasing the period of time during which progress is evaluated stop the log entries from occurring. The default time between Timer triggers is 10 seconds and is configurable indirectly by configuring `maxAdministrativeTaskTime`. For more information, see maxAdministrativeTaskTime. • You may increase the number of threads in the c3p0 pool, thereby increasing the resources available to make progress on tasks. For more information, see numHelperThreads.
Starting Services		

Symptom	Reason	Solution
After you click the **Start** button to start a service, the **Finished** status does not display. This may not be merely a case of the status not getting displayed. It could be for a number of reasons such as network connectivity issues or subcommand failures.	The host is disconnected from the Server, as will be indicated by missing heartbeats on the **Hosts** tab.	• Look at the logs for the service for causes of the problem. • Restart the Agents on the hosts where the heartbeats are missing.
	Subcommands failed resulting in errors in the log file indicating that either the command timed out or the target port was already occupied	• Look at the log file at */var/log/cloudera-scm-server/cloudera-scm-server.log* for more details on the errors. For example, if the port is already occupied you should see an "Address in use" error. • Go to the **Hosts > Status** tab. Click the **Name** of the host you want to inspect. Now go to the **Processes** tab and check the **Stdout/Stderr** logs to diagnose the cause of the failure. For example, if any binaries are missing or if Java could not be found.
After you click **Start** to start a service, the **Finished** status displays but there are error messages. The subcommands to start service components (such as JobTracker and one or more TaskTrackers) do not start.	A port specified in the **Configuration** tab of the service is already being used in your cluster. For example, the JobTracker port is in use by another process.	Enter an available port number in the port property (such as **JobTracker port**) in the **Configuration** tab of the service.
	There are incorrect directories specified in the **Configuration** tab of the service (such as the log directory).	Enter correct directories in the **Configuration** tab of the service.
Job is Failing	No space left on device.	One approach is to use a system monitoring tool such as Nagios to alert on the disk space or quickly check disk space across all systems. If you do not have Nagios or equivalent you can do the following to determine the source of the space issue: In the JobTracker Web UI, drill down from the job, to the map or reduce, to the task attempt details to see which TaskTracker the task executed and failed on due to disk space. For example: *http://JTHost*:50030/taskdetails.jsp?tipid=*TaskID*. You can see on which host the task is failing in the Machine column. In the NameNode Web UI, inspect the % used column on the NameNode Live Nodes page: *http://namenode*:50070/dfsnodelist.jsp?whatNodes=LIVE.
Send Test Alert and Diagnose SMTP Errors		

Symptom	Reason	Solution
You have enabled sending alerts from the Cloudera Manager Admin Console, however, Cloudera Manager does not seem to be sending any alerts. Using the **Send Test Alert** link under **Administration** > **Alerts** shows success even though you do not receive an alert email.	There is possibly a mismatch of protocol or port numbers between your mail server and the Alert Publisher. For example, if the Alert Publisher is sending alerts to SMTPS on port 465 and your mail servers are not configured for SMTPS, you wouldn't receive any alerts.	Use the following steps to make changes to the Alert Publisher configuration: 1. In the Cloudera Manager Admin Console, click the Cloudera Management Service. 2. Click the **Configuration** tab. 3. Select **Scope** > **Alert Publisher**. 4. Click the **Main** category. 5. Change **Alerts: Mail Server Protocol** to `smtp` (or `smtps`). 6. Click the **Ports and Addresses** category and change **Alerts: Mail Server TCP Port** to 25 (or to 465 for SMTPS) 7. Click **Save Changes** to commit the changes. 8. Restart the Alert Publisher.

Logs and Events

For information about problems, check the logs and events:

- Logs on page 123 present log information for services, filtered by role, host, or keywords as well log level (severity).
- Viewing the Cloudera Manager Server Log on page 125 contains information on the server and host agents.
- The Events tab lets you search for and display events and alerts that have occurred within a selected time range filtered by service, hosts, or keywords.

Appendix: Apache License, Version 2.0

SPDX short identifier: Apache-2.0

Apache License
Version 2.0, January 2004
http://www.apache.org/licenses/

TERMS AND CONDITIONS FOR USE, REPRODUCTION, AND DISTRIBUTION

1. Definitions.

"License" shall mean the terms and conditions for use, reproduction, and distribution as defined by Sections 1 through 9 of this document.

"Licensor" shall mean the copyright owner or entity authorized by the copyright owner that is granting the License.

"Legal Entity" shall mean the union of the acting entity and all other entities that control, are controlled by, or are under common control with that entity. For the purposes of this definition, "control" means (i) the power, direct or indirect, to cause the direction or management of such entity, whether by contract or otherwise, or (ii) ownership of fifty percent (50%) or more of the outstanding shares, or (iii) beneficial ownership of such entity.

"You" (or "Your") shall mean an individual or Legal Entity exercising permissions granted by this License.

"Source" form shall mean the preferred form for making modifications, including but not limited to software source code, documentation source, and configuration files.

"Object" form shall mean any form resulting from mechanical transformation or translation of a Source form, including but not limited to compiled object code, generated documentation, and conversions to other media types.

"Work" shall mean the work of authorship, whether in Source or Object form, made available under the License, as indicated by a copyright notice that is included in or attached to the work (an example is provided in the Appendix below).

"Derivative Works" shall mean any work, whether in Source or Object form, that is based on (or derived from) the Work and for which the editorial revisions, annotations, elaborations, or other modifications represent, as a whole, an original work of authorship. For the purposes of this License, Derivative Works shall not include works that remain separable from, or merely link (or bind by name) to the interfaces of, the Work and Derivative Works thereof.

"Contribution" shall mean any work of authorship, including the original version of the Work and any modifications or additions to that Work or Derivative Works thereof, that is intentionally submitted to Licensor for inclusion in the Work by the copyright owner or by an individual or Legal Entity authorized to submit on behalf of the copyright owner. For the purposes of this definition, "submitted" means any form of electronic, verbal, or written communication sent to the Licensor or its representatives, including but not limited to communication on electronic mailing lists, source code control systems, and issue tracking systems that are managed by, or on behalf of, the Licensor for the purpose of discussing and improving the Work, but excluding communication that is conspicuously marked or otherwise designated in writing by the copyright owner as "Not a Contribution."

"Contributor" shall mean Licensor and any individual or Legal Entity on behalf of whom a Contribution has been received by Licensor and subsequently incorporated within the Work.

2. Grant of Copyright License.

Subject to the terms and conditions of this License, each Contributor hereby grants to You a perpetual, worldwide, non-exclusive, no-charge, royalty-free, irrevocable copyright license to reproduce, prepare Derivative Works of, publicly display, publicly perform, sublicense, and distribute the Work and such Derivative Works in Source or Object form.

3. Grant of Patent License.

Subject to the terms and conditions of this License, each Contributor hereby grants to You a perpetual, worldwide, non-exclusive, no-charge, royalty-free, irrevocable (except as stated in this section) patent license to make, have made, use, offer to sell, sell, import, and otherwise transfer the Work, where such license applies only to those patent claims

licensable by such Contributor that are necessarily infringed by their Contribution(s) alone or by combination of their Contribution(s) with the Work to which such Contribution(s) was submitted. If You institute patent litigation against any entity (including a cross-claim or counterclaim in a lawsuit) alleging that the Work or a Contribution incorporated within the Work constitutes direct or contributory patent infringement, then any patent licenses granted to You under this License for that Work shall terminate as of the date such litigation is filed.

4. Redistribution.

You may reproduce and distribute copies of the Work or Derivative Works thereof in any medium, with or without modifications, and in Source or Object form, provided that You meet the following conditions:

1. You must give any other recipients of the Work or Derivative Works a copy of this License; and
2. You must cause any modified files to carry prominent notices stating that You changed the files; and
3. You must retain, in the Source form of any Derivative Works that You distribute, all copyright, patent, trademark, and attribution notices from the Source form of the Work, excluding those notices that do not pertain to any part of the Derivative Works; and
4. If the Work includes a "NOTICE" text file as part of its distribution, then any Derivative Works that You distribute must include a readable copy of the attribution notices contained within such NOTICE file, excluding those notices that do not pertain to any part of the Derivative Works, in at least one of the following places: within a NOTICE text file distributed as part of the Derivative Works; within the Source form or documentation, if provided along with the Derivative Works; or, within a display generated by the Derivative Works, if and wherever such third-party notices normally appear. The contents of the NOTICE file are for informational purposes only and do not modify the License. You may add Your own attribution notices within Derivative Works that You distribute, alongside or as an addendum to the NOTICE text from the Work, provided that such additional attribution notices cannot be construed as modifying the License.

You may add Your own copyright statement to Your modifications and may provide additional or different license terms and conditions for use, reproduction, or distribution of Your modifications, or for any such Derivative Works as a whole, provided Your use, reproduction, and distribution of the Work otherwise complies with the conditions stated in this License.

5. Submission of Contributions.

Unless You explicitly state otherwise, any Contribution intentionally submitted for inclusion in the Work by You to the Licensor shall be under the terms and conditions of this License, without any additional terms or conditions. Notwithstanding the above, nothing herein shall supersede or modify the terms of any separate license agreement you may have executed with Licensor regarding such Contributions.

6. Trademarks.

This License does not grant permission to use the trade names, trademarks, service marks, or product names of the Licensor, except as required for reasonable and customary use in describing the origin of the Work and reproducing the content of the NOTICE file.

7. Disclaimer of Warranty.

Unless required by applicable law or agreed to in writing, Licensor provides the Work (and each Contributor provides its Contributions) on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied, including, without limitation, any warranties or conditions of TITLE, NON-INFRINGEMENT, MERCHANTABILITY, or FITNESS FOR A PARTICULAR PURPOSE. You are solely responsible for determining the appropriateness of using or redistributing the Work and assume any risks associated with Your exercise of permissions under this License.

8. Limitation of Liability.

In no event and under no legal theory, whether in tort (including negligence), contract, or otherwise, unless required by applicable law (such as deliberate and grossly negligent acts) or agreed to in writing, shall any Contributor be liable to You for damages, including any direct, indirect, special, incidental, or consequential damages of any character arising as a result of this License or out of the use or inability to use the Work (including but not limited to damages for loss of goodwill, work stoppage, computer failure or malfunction, or any and all other commercial damages or losses), even if such Contributor has been advised of the possibility of such damages.

9. Accepting Warranty or Additional Liability.

While redistributing the Work or Derivative Works thereof, You may choose to offer, and charge a fee for, acceptance of support, warranty, indemnity, or other liability obligations and/or rights consistent with this License. However, in accepting such obligations, You may act only on Your own behalf and on Your sole responsibility, not on behalf of any other Contributor, and only if You agree to indemnify, defend, and hold each Contributor harmless for any liability incurred by, or claims asserted against, such Contributor by reason of your accepting any such warranty or additional liability.

END OF TERMS AND CONDITIONS

APPENDIX: How to apply the Apache License to your work

To apply the Apache License to your work, attach the following boilerplate notice, with the fields enclosed by brackets "[]" replaced with your own identifying information. (Don't include the brackets!) The text should be enclosed in the appropriate comment syntax for the file format. We also recommend that a file or class name and description of purpose be included on the same "printed page" as the copyright notice for easier identification within third-party archives.

```
Copyright [yyyy] [name of copyright owner]

Licensed under the Apache License, Version 2.0 (the "License");
you may not use this file except in compliance with the License.
You may obtain a copy of the License at

  http://www.apache.org/licenses/LICENSE-2.0

Unless required by applicable law or agreed to in writing, software
distributed under the License is distributed on an "AS IS" BASIS,
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
See the License for the specific language governing permissions and
limitations under the License.
```