

**cloudera<sup>®</sup>**

# Cloudera Operation

## **Important Notice**

© 2010-2021 Cloudera, Inc. All rights reserved.

Cloudera, the Cloudera logo, and any other product or service names or slogans contained in this document are trademarks of Cloudera and its suppliers or licensors, and may not be copied, imitated or used, in whole or in part, without the prior written permission of Cloudera or the applicable trademark holder. If this documentation includes code, including but not limited to, code examples, Cloudera makes this available to you under the terms of the Apache License, Version 2.0, including any required notices. A copy of the Apache License Version 2.0, including any notices, is included herein. A copy of the Apache License Version 2.0 can also be found here: <https://opensource.org/licenses/Apache-2.0>

Hadoop and the Hadoop elephant logo are trademarks of the Apache Software Foundation. All other trademarks, registered trademarks, product names and company names or logos mentioned in this document are the property of their respective owners. Reference to any products, services, processes or other information, by trade name, trademark, manufacturer, supplier or otherwise does not constitute or imply endorsement, sponsorship or recommendation thereof by us.

Complying with all applicable copyright laws is the responsibility of the user. Without limiting the rights under copyright, no part of this document may be reproduced, stored in or introduced into a retrieval system, or transmitted in any form or by any means (electronic, mechanical, photocopying, recording, or otherwise), or for any purpose, without the express written permission of Cloudera.

Cloudera may have patents, patent applications, trademarks, copyrights, or other intellectual property rights covering subject matter in this document. Except as expressly provided in any written license agreement from Cloudera, the furnishing of this document does not give you any license to these patents, trademarks copyrights, or other intellectual property. For information about patents covering Cloudera products, see <http://tiny.cloudera.com/patents>.

The information in this document is subject to change without notice. Cloudera shall not be liable for any damages resulting from technical errors or omissions which may be present in this document, or from use of this document.

### **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 5.3.x  
Date: February 3, 2021

# Table of Contents

<b>About Cloudera Operation.....</b>	<b>5</b>
<b>Monitoring and Diagnostics.....</b>	<b>6</b>
Introduction to Cloudera Manager Monitoring.....	6
Starting and Logging into the Admin Console.....	7
Time Line.....	7
Health Tests.....	8
Cloudera Manager Admin Console Home Page.....	9
Viewing Charts for Cluster, Service, Role, and Host Instances.....	12
Configuring Monitoring Settings.....	13
Monitoring Clusters.....	20
Monitoring Services.....	21
Monitoring Service Status.....	21
Viewing Service Status.....	22
Viewing Service Instance Details.....	26
Viewing Role Instance Status.....	27
Running Diagnostic Commands for Roles.....	29
Period Stacks Collection.....	29
Managing and Monitoring Federated HDFS.....	30
Viewing Running and Recent Commands.....	31
Monitoring Resource Management.....	32
Monitoring Hosts.....	34
Host Details.....	35
Host Inspector.....	38
Monitoring Activities.....	39
MapReduce Jobs.....	39
Monitoring Impala Queries.....	46
YARN Applications.....	54
Events.....	59
Viewing Events.....	60
Filtering Events.....	61
Alerts.....	61
Triggers.....	62
Audit Events.....	64
Viewing Audit Events.....	64
Filtering Audit Events.....	64
Downloading Audit Event Logs.....	65

Charting Time-Series Data.....	65
Terminology.....	66
Building a Chart with Time-Series Data.....	66
Configuring Time-Series Query Results.....	67
Using Context-Sensitive Variables in Charts.....	67
Chart Properties.....	68
Displaying Chart Details.....	71
Editing a Chart.....	72
Saving a Chart.....	73
Obtaining Time-Series Data Using the API.....	73
Dashboards.....	73
tsquery Language.....	76
Metric Aggregation.....	85
Logs.....	87
Viewing Logs.....	88
Logs List.....	88
Filtering Logs.....	88
Log Details.....	89
Viewing Cloudera Manager Server and Agent Logs.....	89
Reports.....	90
Disk Usage Reports.....	90
Activity, Application, and Query Reports.....	91
The File Browser.....	92
Downloading HDFS Directory Access Permission Reports.....	93
Troubleshooting Cluster Configuration and Operation.....	93
Solutions to Common Problems.....	93
Logs and Events.....	95

<b>Appendix: Apache License, Version 2.0.....</b>	<b>96</b>
---	-----------

## 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 21](#) - 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 34](#) - 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 39](#) - 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 59](#) - 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 on page 61](#) - 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.
- [Audit Events on page 64](#) - describes how to view service, role, and host life cycle 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 65](#) - 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 87](#) - 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 90](#) - 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 93](#) - 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.

## Starting and Logging into the Admin Console

1. In a web browser, enter `http://Server host:7180`, where *Server host* is the fully-qualified domain name 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.

## 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.

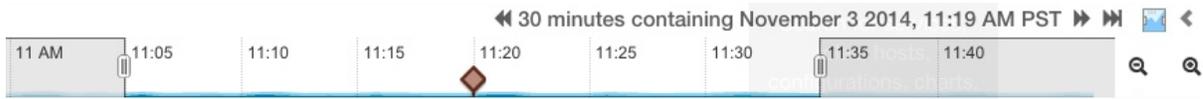
### Selecting a Point in Time

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.

## Monitoring and Diagnostics

- 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 **Apply**.  
 , choosing the date and time, and clicking

### 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



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 via **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 13 for more information.

For detailed information on each supported health test, see [Cloudera Manager Health Tests](#).

### Viewing Health Test Results

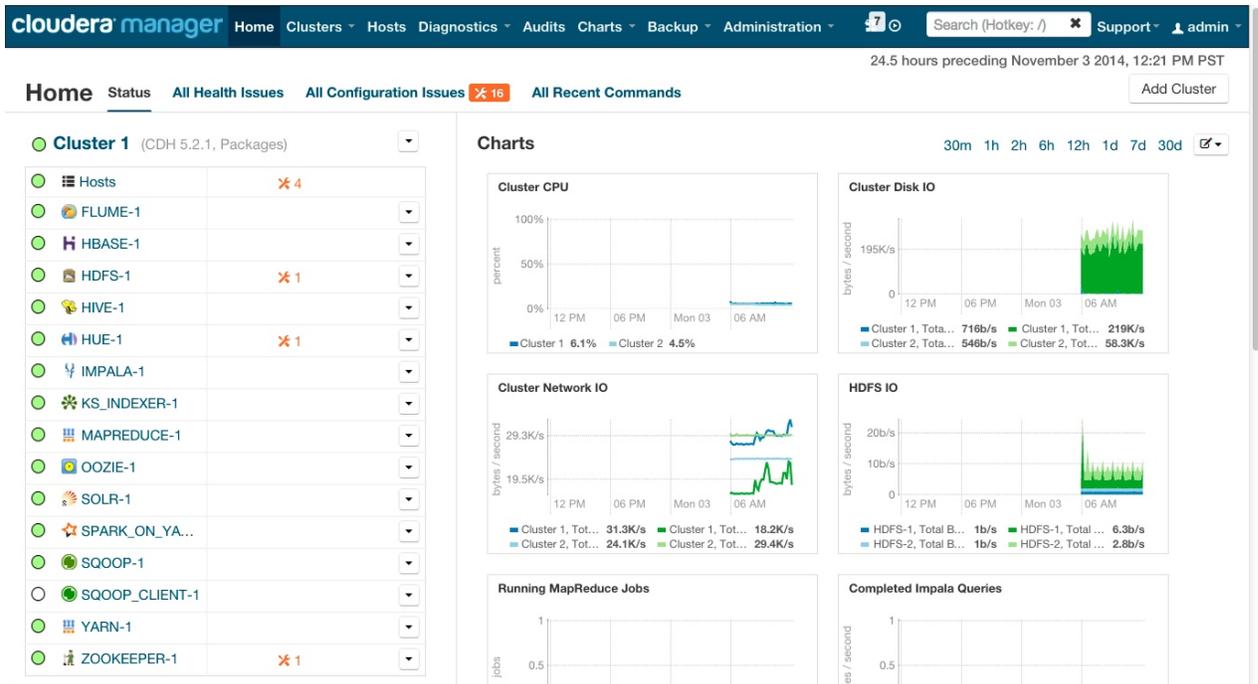
Health test results are available in the following locations:

- **Home** page 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.

For some health test results, you can chart the associated metrics over a time range. See [Viewing Service Status](#) on page 22, [Viewing Role Instance Status](#) on page 27, and [Host Details](#) on page 35 for details.

### Cloudera Manager Admin Console Home Page

When you start the [Cloudera Manager Admin Console](#), the Home page displays.



You can also navigate to the Home page by clicking **Home** in the top navigation bar.

### Status

The default tab displayed when the Home page displays. It 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	<p>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.</p> <div style="border: 1px solid orange; padding: 10px; margin: 10px 0;"> <p><b>Important:</b> 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.</p> </div> <p>Click the indicator to display the <b>Health Issues</b> pop-up dialog.</p> <p>By default only Bad health test results are shown in the dialog. To display Concerning health test results, click the <b>Also show <i>n</i> concerning issue(s)</b></p>

Indicator	Meaning	Description
		link. Click the link to display the Status page containing with details about the health test result.
	Configuration issue	<p>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.</p> <div style="border: 1px solid yellow; padding: 5px; margin: 10px 0;"> <p> <b>Important:</b> If there is one configuration error and two configuration warnings, there will be three configuration issues, but the number will be one.</p> </div> <p>Click the indicator to display the <b>Configuration Issues</b> pop-up dialog.</p> <p>By default only notifications at the Error severity level are listed, grouped by service name are shown in the dialog. To display Warning notifications, click the <b>Also show <i>n</i> warning(s)</b> 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 <a href="#">Managing Services</a>.</p>
 Restart Needed  Refresh Needed	Configuration modified	<p>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.</p> <p>Click the indicator to display the <a href="#">Stale Configurations</a> page. To bring the cluster up-to-date, click the Refresh or Restart button on the Stale Configurations page or follow the instructions in <a href="#">Refreshing a Cluster</a>, <a href="#">Restarting a Cluster</a>, or <a href="#">Restarting Services and Instances after Configuration Changes</a>.</p>
	Client configuration redeployment required	<p>Indicates that the client configuration for a service should be redeployed.</p> <p>Click the indicator to display the <a href="#">Stale Configurations</a> page. To bring the cluster up-to-date, click the Deploy Client Configuration button on the Stale Configurations page or follow the instructions in <a href="#">Manually Redeploying Client Configuration Files</a>.</p>

- **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

[30m](#) [1h](#) [2h](#) [6h](#) [12h](#) [1d](#) [7d](#) [30d](#) 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. 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** toggle.
- 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. 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 31.

## 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 page 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 toggle 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 on a 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 doesn't allow removing the original charts. Use the edit button



to the upper right of the chart to toggle 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 62.

## 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:

- 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. Cloudera Manager supports this type of monitoring configuration for HDFS, MapReduce, YARN, HBase, Impala, ZooKeeper, and Flume. For hosts you can disable or enable selected health tests, modify thresholds, and enable or disable health alerts.
- For hosts, you can set threshold-based monitoring of free space in the various directories on the hosts Cloudera Manager monitors.
- For MapReduce, YARN, and Impala services, you can configure aspects of how Cloudera Manager monitors activities, applications, and queries.

## Monitoring and Diagnostics

- For the Cloudera Management Service you can configure monitoring settings for the monitoring roles—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 in order 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](#).

This section covers the following topics:

### 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 tests whose result is considered "Concerning" or "Bad" will be 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 will also be published when the health test result returns to normal.

You can control when an individual health test will be forwarded as an event or an [alert](#) by modifying the threshold values for the relevant health test.

### Configuring Service Monitoring

1. Select **Clusters > Service**.
2. Click the **Configuration** tab.
3. Expand the **Service-Wide** category.
4. Click the **Monitoring** category.
5. Select the category of properties to change (these are organized as service-wide and by role group).
6. Configure the property.
7. Click **Save Changes** to commit the changes.

### Configuring Host Monitoring

1. Click the **Hosts** tab.
2. Select a host.
3. Click the **Configuration** tab.
4. Expand the **Service-Wide** category.
5. Click the **Monitoring** category.
6. Configure the property.
7. Click **Save Changes** to commit the changes. You can add a note that will be included with the change in the Configuration History.

### 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 will trigger events at both the Warning and Critical levels.

If you set both thresholds, the Absolute Threshold setting will be 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. Click the **Clusters** tab.
2. Select the **MapReduce** service instance.
3. Click the **Configuration** tab.
4. Expand the **Service-Wide** category.
5. Click the **Monitoring** category.
6. Specify one or more [activity duration rules](#).
7. Click **Save Changes** to commit the changes.

## 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) will match the first rule.

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

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

In this case, any job that runs longer than 60 minutes will be marked slow and will generate an event.

## Configuring YARN Application Monitoring

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

## Monitoring and Diagnostics

### Configuring Application Visibility

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

1. Click the **Clusters** tab.
2. Select the **YARN** service instance.
3. Click the **Configuration** tab.
4. Expand the **Service-Wide** category.
5. Click the **Monitoring** category.
6. Set the **Visibility Settings** properties for admin and non-admin users.
7. Click **Save Changes** to commit the changes. You can add a note that will be included with the change in the Configuration History.

### 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. Click the **Clusters** tab.
2. Select the **Impala** service instance.
3. Click the **Configuration** tab.
4. Expand the **Service-Wide** category.
5. Click the **Monitoring** category.
6. Set the **Visibility Settings** properties for admin and non-admin users.
7. Click **Save Changes** to commit the changes.

### Configuring Impala Query Data Store Maximum Size

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

1. Click the **Clusters** tab.
2. Do one of the following:
  - Select **Clusters > Cloudera Management Service > Cloudera Management Service**.
  - On the Status tab of the Home page, in **Cloudera Management Service** table, click the **Cloudera Management Service** link.
3. Click the **Configuration** tab.
4. Click the **Service Monitor Default Group** 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. Restart the Service Monitor.

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

The following topics describe how to configure when alerts are raised and how they are delivered:

- [Enabling Activity Monitor Alerts](#) on page 17
- [Enabling Configuration Change Alerts](#) on page 17
- [Enabling HBase Alerts](#) on page 17
- [Configuring Health Alerts](#) on page 17

- [Configuring Log Alerts](#) on page 18
- [Configuring Alert Delivery](#) on page 18

### Enabling Activity Monitor Alerts

You can enable alerts when an activity runs [too slowly](#) or fails.

1. Click the **Clusters** tab.
2. Select the **MapReduce** service instance.
3. Click the **Configuration** tab.
4. Expand the **Service-Wide** category.
5. Click the **Monitoring** category.
6. Check the **Alert on Slow Activities** or **Alert on Activity Failure** checkboxes.
7. Click **Save Changes** to commit the changes. You can add a note that will be included with the change in the Configuration History.

### 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. Check the **Enable Configuration Change Alerts** checkbox.
4. Click **Save Changes** to commit the changes. You can add a note that will be included with the change in the Configuration History.

### Enabling HBase Alerts

1. Click the **Clusters** tab.
2. Select the **HBase** service instance.
3. Click the **Configuration** tab.
4. Expand the **Service-Wide** category.
5. Click the **Monitoring** category.
6. 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
7. Click **Save Changes** to commit the changes.

### Configuring Health Alerts

#### Enabling Health Alerts

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

1. Select **Clusters > Service**.
2. Click the **Configuration** tab.
3. Expand the **Service-Wide** category.
4. Click the **Monitoring** category.
5. Check the **Enable Health Alerts for this Role** checkbox.
6. Click **Save Changes** to commit the changes.

### 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. In the category **Event Server Default Group**, click the **Health Alert Threshold** value, and choose the **Bad** or **Concerning** radio button.
4. Click **Save Changes** to commit the changes. You can add a note that will be included with the change in the Configuration History.

### 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. In the category **Event Server Default Group**, check the **Alert on Transitions out of Alerting Health** checkbox.
4. Click **Save Changes** to commit the changes. You can add a note that will be included with the change in the Configuration History.

### 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 20.

### 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 e-mail. See [Managing Alerts](#).



**Note:** If alerting is enabled for events, you will be able to 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 Default Group > Logs**.
4. Edit a log property.
5. Click **Save Changes** to commit the changes.
6. Restart the role.

### Configuring Log Directories

1. Do one of the following:
  - 1. On the Home page, click a cluster name.

2. Select **Configuration > All Log Directories**.
  3. Edit a **Role Log Directory** property.
- 1. Go to a service.
    2. Click the **Configuration** tab.
    3. Select **RoleName Default Group > Logs**.
    4. Edit the **Log Directory** property.

2. Click **Save Changes** to commit the changes.
3. Restart the role.

#### Enabling and Disabling Log Event Capture

1. Select **Clusters > Service**.
2. Click the **Configuration** tab.
3. Expand the **Service-Wide** category.
4. Click the **Monitoring** category.
5. Modify the **Enable Log Event Capture** setting.
6. Click **Save Changes** to commit the changes. You can add a note that will be included with the change in the Configuration History.

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 for each HDFS, MapReduce and HBase role, and for ZooKeeper, Flume agent, and monitoring roles as well.

#### Configuring Which Log Messages Become Events

1. Select **Clusters > Service**.
2. Expand the **Service-Wide** category.
3. Click the **Monitoring** category.
4. 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 will show all copies of the property that match the search filter.
5. Edit these rules as needed.
6. Click **Save Changes** to commit the changes.

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 page 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 page 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	<p>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.</p> <div style="border: 1px solid #f0e68c; padding: 10px; margin: 10px 0;"> <p> <b>Important:</b> 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.</p> </div> <p>Click the indicator to display the <b>Health Issues</b> pop-up dialog.</p> <p>By default only Bad health test results are shown in the dialog. To display Concerning health test results, click the <b>Also show <i>n</i> concerning issue(s)</b> link. Click the link to display the Status page containing with details about the health test result.</p>
	Configuration issue	<p>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.</p> <div style="border: 1px solid #f0e68c; padding: 10px; margin: 10px 0;"> <p> <b>Important:</b> If there is one configuration error and two configuration warnings, there will be three configuration issues, but the number will be one.</p> </div> <p>Click the indicator to display the <b>Configuration Issues</b> pop-up dialog.</p>

Indicator	Meaning	Description
		By default only notifications at the Error severity level are listed, grouped by service name are shown in the dialog. To display Warning notifications, click the <b>Also show n warning(s)</b> 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 <a href="#">Managing Services</a> .
 Restart Needed  Refresh Needed	Configuration modified	<p>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.</p> <p>Click the indicator to display the <a href="#">Stale Configurations</a> page. To bring the cluster up-to-date, click the Refresh or Restart button on the Stale Configurations page or follow the instructions in <a href="#">Refreshing a Cluster</a>, <a href="#">Restarting a Cluster</a>, or <a href="#">Restarting Services and Instances after Configuration Changes</a>.</p>
	Client configuration redeployment required	<p>Indicates that the client configuration for a service should be redeployed.</p> <p>Click the indicator to display the <a href="#">Stale Configurations</a> page. To bring the cluster up-to-date, click the Deploy Client Configuration button on the Stale Configurations page or follow the instructions in <a href="#">Manually Redeploying Client Configuration Files</a>.</p>

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

## 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.

## Monitoring and Diagnostics

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 page, 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 22 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 27 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 page Status tab, if the cluster is displayed in full form, click *ServiceName* in a *ClusterName* table.
- In the Home page 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 23.

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 13.

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 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 and the Logs and Events 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 (🕒) allows 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.

### 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 1: 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 <a href="#">Stale Configurations</a> 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 <b>Good</b> .
	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 <b>Concerning</b> .

Indicator	Status	Description
	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 <b>Bad</b> .
	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.

You can click either the **Status** link for a role to drill down to see the details of the status of the role instance(s). If there is a single instance of the role type, the link takes you directly to the [role instance status](#).

If there are multiple role instances (such as for DataNodes, TaskTrackers, RegionServers) clicking  to the right of the role's health

DataNode  [3 Good Health](#) 

displays a pop-up that allows you to select specific instances. Furthermore, this pop-up displays the results for each health test that applies to this role type.

### 3 DataNodes: Good Health ✕

Review the following health check results from these 3 DataNodes.

- **GC Duration:** 3 Good
- **Web Server Status:** 3 Good
- **Free Space:** 3 Good
- **Host Health:** 3 Good
- **Data Directory Status:** 3 Good
- **NameNode Connectivity:** 3 Good
- **Unexpected Exits:** 3 Good
- **Block Count:** 3 Good
- **Log Directory Free Space:** 3 Good
- **Process Status:** 3 Good
- **File Descriptors:** 3 Good

Name	Host	Status
<a href="#">datanode (nightly-2)</a>	<a href="#">nightly-2.ent.cloudera.com</a>	 Good Health
<a href="#">datanode (nightly-3)</a>	<a href="#">nightly-3.ent.cloudera.com</a>	 Good Health
<a href="#">datanode (nightly-4)</a>	<a href="#">nightly-4.ent.cloudera.com</a>	 Good Health

You can filter by an individual health test result. Click the result link - an  appears by the link (as shown in the illustration above) and only the instance(s) with that specific health status will appear in the instances list.

### Service Summary

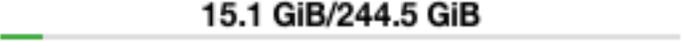
Some services (specifically HDFS, MapReduce, HBase, Flume, and ZooKeeper) provide additional statistics about their operation and performance. These are shown in a Summary panel at the left side of the page. The contents of this panel depend on the service:

- 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 25.

- 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:

## HDFS Summary

Configured Capacity	 15.1 GiB/244.5 GiB
Quick Links	<a href="#">Replication</a> , <a href="#">Reports</a> , <a href="#">Browse Filesystem</a> , <a href="#">NameNode Web UI (Active)</a> 
Event Search	<a href="#">Alerts</a>  , <a href="#">Critical</a>  , <a href="#">All</a> 

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 13 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 12 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.

## Monitoring and Diagnostics

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 page Status tab, if the cluster is displayed in full form, click *ServiceName* in a *ClusterName* table.
- In the Home page 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 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 <a href="#">Stale Configurations</a> 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 <b>Good</b> .
	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 <b>Concerning</b> .
	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 <b>Bad</b> .
	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**)

## Monitoring and Diagnostics

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 28 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 13 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 12 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 page 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 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.

### Period 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 Period 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, and expand the corresponding role category in the left-hand pane, and click the **Stacks Collection** category.
- From the service page in Cloudera Manager, click the **Instances** tab, click the role type, click the **Configuration** tab on the role page, expand the role name in the left-hand pane, and click the **Stacks Collection** category.

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. On the HDFS service page, click **Configuration**.
2. Click the arrow to expand **NameNode Default Group** in the left-hand pane.
3. Click **Stacks Collection**.
4. Modify the configuration settings if desired.
5. 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.

## Managing and Monitoring Federated HDFS

The HDFS service has some unique functions that may result in additional information on its Status and Instances pages. Specifically, if you have configured HDFS with high availability, these two pages will contain additional information.

### The HDFS Status Page with Multiple Nameservices

If your HDFS configuration has multiple nameservices, the HDFS Service Status page will have separate tabs for each nameservice. Your HDFS configuration will have multiple nameservices if you have configured federated nameservices to manage multiple namespaces.

Each tab shows the same types of status information as for an HDFS instance with a single namespace.

### The HDFS Instances Page with Federation and High Availability

If you have high availability configured, the Instances page has a section at the top that provides information about the configured nameservices. This includes information about:

- Whether high availability and automatic failover are enabled
- Links to the active and standby NameNodes and SecondaryNameNode (depending on whether high availability is enabled or not).

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

There is also an **Actions** menu for each nameservice. From this menu you can:

- Edit the list of mount points for the nameservice (using the **Edit...** command)
- Enable or disable high availability and automatic failover

## 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 hand 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 **Home** link in the Admin Console main navigation bar and click the **All Recent Commands** tab.

If you are managing multiple clusters, 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've 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 lets you view what 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 needed directories, started the HDFS role instances (the NameNode, Secondary NameNode and DataNode instances) and even the command that initially formatted HDFS on the NameNode. This may be particularly useful if a service or role seems to be taking a long time to start up or shut down, or if certain 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 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 currently in progress.

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

## Monitoring and Diagnostics

If the command generates subcommands, this is indicated; click the command link to display the subcommands in a **Child Commands** section as they are started. Each child command also has an **Abort** button that is present as long as the subcommand is in progress.

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

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

### Recent Commands

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

Select a value from the **Showing last *n*** drop-down list to control how many commands are listed.

If no commands were run during the selected time range, you can click the **Try expanding the time range selection** link. Each time you click the link it doubles the time range selection. 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).

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, there may be multiple subcommands that result from the top level command.

The **First Run** command is run as part of the initial startup of your cluster. Click this link to view the command history of the startup of your cluster.

### Command Details

Click a command in the **Command** list to display its command details, and its child commands (subcommands), if there are any. The **Command Details** section at the top shows information about the command:

- The command (and how many subcommands, if any, it has)
- The context, which may be a cluster, service, host, or role
- The current status
- The time the command started
- The time the command ended
- A message about the command completion
- If the context was a role, links to role instance logs

If the command included multiple steps, a **Command Progress** section may appear showing the steps within the command and whether they succeeded.

The **Child Commands** section lists any subcommands of the selected command. You can perform the following actions:

- Filter the child commands to display all, only failed, or only active commands.
- If you are displaying a child command, use the **Parent Command** link near the top of the page to return to the parent command's details.
- Click the **Command** link to display further command details (and any subcommands) of this command. You can continue to drill down through a tree of subcommands this way.
- 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 was related.

## Monitoring Resource Management

Cloudera Manager 4 introduced the ability to partition resources across HBase, HDFS, Impala, MapReduce, and YARN services by allowing you to set configuration properties that were enforced by Linux control groups (Linux cgroups).

With Cloudera Manager 5, the ability to statically allocate resources using cgroups is configurable through a single *static service pool wizard*. You allocate services 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** > **Resource Management** > **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 given to other pools. Otherwise, a pool receives a share of resources in accordance with the pool's weight. Dynamic resource pools have ACLs that restrict who can submit work to and administer them.

### Viewing Dynamic Resource Pools

1. Select **Clusters** > **Cluster name** > **Resource Management** > **Dynamic Resource Pools**. If the cluster has a YARN service, the Dynamic Resource Pools Status tab displays. If the cluster has only an Impala service enabled for dynamic resource pools, the Dynamic Resource Pools Configuration tab displays.

### Dynamic Resource Pools Status

For the **YARN Independent RM** and **YARN and Impala Integrated RM** resource management scenarios described in [Managing Resources with Cloudera Manager](#), the Dynamic Resource Pools Status tab displays a summary of YARN scheduler status, a list of current allocations for the pools that can or have been allocated resources, and a set of pool resource usage charts for the cluster. If the **Impala Independent RM** scenario is in effect, there is no Status 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.
- **Pools Status** - a list of pools that have been explicitly configured and pools created by YARN, properties of the pools, and an action menu.
  - **Allocated Memory** - The memory assigned to the pool that is currently allocated to applications and queries.

## Monitoring and Diagnostics

- **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.
- Click  and select
  - **YARN Applications** to display the [YARN Applications](#) on page 54 page and list the applications that are running or have run in that pool.
  - **Impala Queries** to display the [Monitoring Impala Queries](#) on page 46 page and list the queries that are running or have run in that pool.

## 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 includes the version of CDH running on the host, the cluster to which the host belongs, and the number of roles running on the host.

- Clicking the ► to the left of the number of roles lists all the role instances running on that host. The balloon annotation that appears when you move the cursor over a link indicates the service instance to which the role belongs.
- Filter the hosts by typing a property value in the Search box or selecting a value from the facets at the left of the page.

### Disks Overview

Click the **Disks Overview** button 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"`. Click a histogram to drill down and identify outliers.

### Viewing the Hosts in a Cluster

Do one of the following:

- Select **Clusters** > **Cluster name** > **General** > **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 35.

## 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 13 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.

## Monitoring and Diagnostics

- 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 12 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 31 for more information.

### Configuration

**Minimum Required Role:** [Full Administrator](#)

The Configuration page for a host lets you set monitoring properties for the selected host. In addition, for parcel upgrades, you can blacklist specific products - specify products that should not be distributed or activated on the host.

To modify the monitoring properties for the selected host:

1. Click the **Configuration** tab.
2. Click the **Monitoring** category.
  - Under **Thresholds** you can configure the thresholds for monitoring the free space in the Agent Log and Agent Process Directories for all your hosts. You can set these thresholds as either or both a percentage and an absolute value (in bytes).
  - Under **Other** you can set health check thresholds for a variety of conditions related to memory usage and other properties. Here is where you can enable Alerting for health check events for all your managed hosts.

The monitoring settings you make on this page will override the global host monitoring settings from the Configuration tab of the Hosts page.

For more information, see [Modifying Configuration Properties](#).

### 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 [Audit Events](#) on page 64 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 12 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.
2. Click the **Host Inspector** 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 (CDH 3 or CDH 4). If you are running CDH 3, the Version will be listed as "Unavailable". Version identification is not available with CDH 3. In a CDH 3 cluster, CDH 4 components will be listed as "Not installed or path incorrect".

If you are running multiple clusters with both CDH 3 and CDH 4, the lists will be organized by distribution (CDH 3 or CDH 4). The hosts running that version are shown at the top of each list.

#### 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 () located 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 sub-commands.
5. Click the **Show Inspector Results** button to view the report.

See [Viewing Running and Recent Commands](#) 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 under the Activities section of 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.

### 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](#) on page 40
- [Selecting Columns to Show in the Activities List](#) on page 41
- [Sorting the Activities List](#) on page 41
- [Filtering the Activities List](#) on page 41
- [Activity Charts](#) on page 42

## Viewing MapReduce Activities

1. Select **Clusters > Cluster name > Activities > MapReduce service name Jobs**. The *MapReduce service name* page displays a list of activities. The columns (by default only a subset of the possible metrics are displayed – you can [modify the columns](#) that are displayed to add or remove columns) in the Activities list show statistics about the performance of and resources used by each activity:

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

<a href="#">Children</a>	For a Pig, Hive or Oozie activity, takes you to the <b>Children</b> 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.
<a href="#">Tasks</a>	For a MapReduce job, takes you to the <b>Tasks</b> 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.
<a href="#">Details</a>	Takes you to the <b>Details</b> tab where you can view the activity or job statistics in report form.
<a href="#">Compare</a>	Takes you to the <b>Compare</b> tab where you can see how the selected activity compares to other similar activities in terms of a wide variety of metrics.
<a href="#">Task Distribution</a>	Takes you to the <b>Task Distribution</b> 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.
<b>Kill Job</b>	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 42 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 **x** in the upper right corner to close the panel.



**Note:** You cannot hide the context 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 **X** 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 39 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 **x** 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.

## Monitoring and Diagnostics

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 in seconds. From the drop-down you can choose 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 the 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 Cloudera Impala 1.0.1 and higher.

Contents:

### Viewing Queries

1. Do one of the following:

- Select **Clusters** > *Cluster name* > **Activities** > *Impala service name* **Queries**.
- On the **Home** > **Status** tab, select *Impala service name* and click the **Queries** tab.

The Impala queries run during the last day are listed in the [Queries list](#).

### 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 16.

Query information is stored in-memory in a ring buffer. This has two consequences: if you restart Service Monitor, all queries are lost and older queries eventually will be dropped. For information on how to configure the query store, see [Configuring Impala Query Data Store Maximum Size](#) on page 16.

### Impala Best Practices

To open the Impala Best Practices page, click the **Best Practices** tab on the Impala service page. The page contains charts to help identify whether Impala best practices are being followed. See the individual charts for a descriptions of each best practice and how to determine if it is being followed. Consult the Impala documentation for further detail on each best practice and for additional best practices.

Adjust the time range to see data on queries run at different times. Click on the charts to get more detail on individual queries. Use the filter box at the top right of the Best Practices page to adjust what 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.

### Queries List

Queries appear in the list 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:

10/17/2013 4:32 PM - [▶](#) `select sample_07.description, sample_07.salary FROM sample_07 WHERE ( sample_07.salary > 100000) ORDER BY sample_07.salary DE...` Actions ▾  
 10/17/2013 4:32 PM User: root Coordinator: [tcdn5-2.ent.cloudera.com](#) Database: default Query Type: QUERY HDFS Bytes Read: 45.0 KiB

You can add additional attributes to the summary with the [Attribute Selector](#). In each query summary, the query statement is truncated if it is too long to display. To display the entire statement, hover over a query. The query entry will expand to display the entire query string. To collapse the query display, move the mouse cursor. To display information about query attributes and possible values, hover over a field in a query. For example:

The type of this query. Possible values are QUERY, DDL and DML. Called "queryType" in searches.

Query Type: QUERY

A running query displays [▶ Running](#) under the start timestamp. If an error occurred while processing the query, [! Error](#) displays under the complete timestamp.

To kill a running query, select **Actions > Cancel**. Only an administrator can cancel queries and killing a query creates an audit event. When you cancel a query, [⊘ Canceled](#) replaces the [▶ Running](#) label. Once the page is refreshed, the entry is removed from the list.

When a job fails, `asdf` replaces the [▶ Running](#) label.

To display [query details](#), select **Actions > Details**.

To display all the queries run by the same user, select **Actions > User's Impala queries**.

To display all the queries that used the same [resource pool](#), select **Actions > Queries in the same YARN pool**.

### Filtering Queries

You filter queries by selecting a time range and specifying a filter expression in the text 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 is made up of three components:

- **Attribute** - the query language name of the attribute.
- **Operator** - the type of comparison between the attribute and the attribute value. Cloudera Manager supports the standard comparator operators: =, !=, >, <, >=, <=, and `RLIKE`, which does 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 compare 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, surround subexpressions with 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**

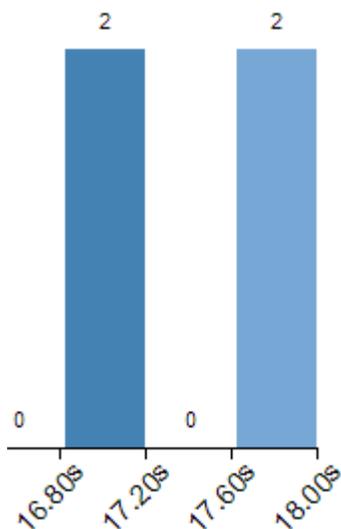
1. Click the



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 Attribute Histograms**

1. Optionally, click the **Select Attributes** link to display a dialog where you can chose which attributes to display in histograms. Check the checkbox next to one or more attributes, and click **Close**.
2. Click the **Enhance Filter** link. Histograms of the selected attributes display with the number of results that match each value of the selected attributes.
3. Click a histogram bar that represents the range of attribute values to filter on. The color of the histogram bar gets lighter (on the right below)



and a filter with the attribute value set to the range of the histogram bucket is added to the text box. The range includes the lower bound of the bucket and excludes the upper bound of the bucket. For example:

```
(<x>_duration >= 17600.0 AND <x>_duration < 18000.0)
```

where <x> is query or application.

If you click the same histogram bar again, the color reverts to the darker blue and the filter is removed from the text box.

If you click another histogram bar, another filter is OR'd with the existing filter:

```
(<x>_duration >= 17600.0 AND <x>_duration < 18000.0 OR <x>_duration >= 16000.0 AND <x>_duration < 16400.0)
```

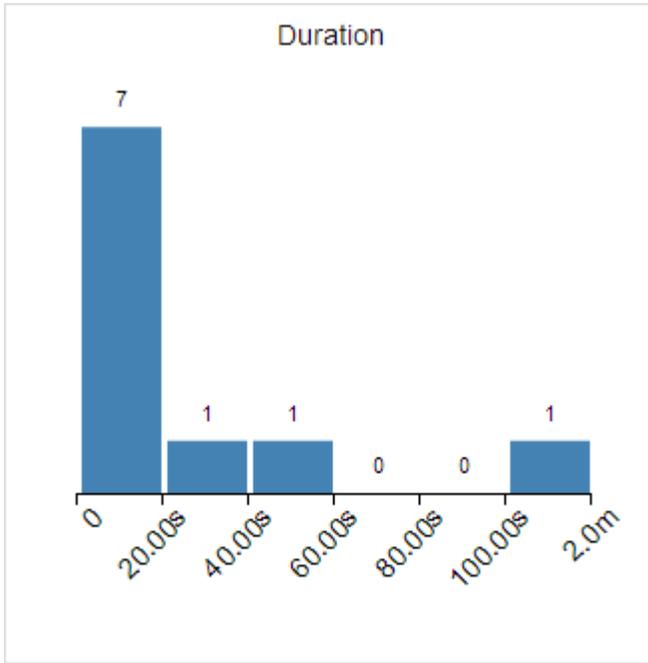
- **Type a Filter**

1. Start typing or press **Spacebar** in the text box. As you type, filter attributes matching the letter you type 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. Put the cursor on 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.

#### Example: Drilling into Query Results

Suppose we have a set of results with the following duration distribution :

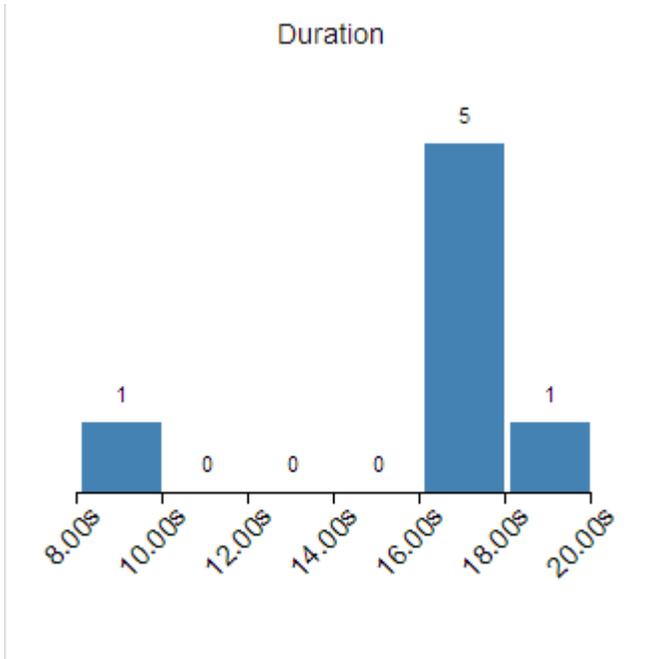


The 0-20.00s bucket has 7 results, but with the current distribution we cannot discriminate between the results in that bucket.

Selecting the left-most histogram bar adds the following filter to the text box:

```
<x>_duration >= 0.0 AND <x>_duration < 20000.0
```

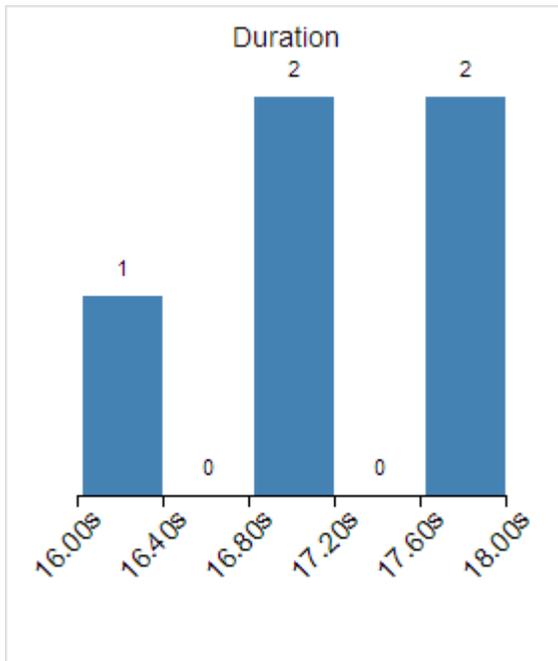
where <x> is query or application. After clicking **Search** again, the histogram appears as follows:



Selecting the histogram with 5 results again refines the filter to:

```
<x>_duration >= 16000.0 AND <x>_duration < 18000.0
```

After clicking **Search** again, the histogram appears as follows:



### Filter Attributes

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

Attribute	Display Name	Value Type	Description
bytes_streamed	Bytes Streamed	BYTES	The total number of bytes sent between Impala daemons while processing the query.
coordinator_host_id	Coordinator	STRING	The host coordinating the query.
database	Database	STRING	The database on which the query was run.
ddl_type	DDL Type	STRING	The type of DDL query.
executing	Executing	BOOLEAN	Whether the query is currently executing.
file_formats	File Formats	STRING	The file formats used in the query. A file format is a string of the form: <i>File Type/Compression Type</i> , where <i>File Type</i> can take the values: TEXT, PARQUET, SEQUENCE_FILE, and RC_FILE, and <i>Compression Type</i> can take the values: NONE, DEFAULT, BZIP2. For further information, see <a href="#">How Impala Works with Hadoop File Formats</a> .
hbase_bytes_read	HBase Bytes Read	BYTES	The total number of bytes read from HBase by the query.
hbase_bytes_read_per_second	HBase Read Throughput	BYTES_PER_SECOND	The overall HBase read throughput (in B/s) of the query.
hdfs_bytes_read	HDFS Bytes Read	BYTES	The total number of bytes (in GiB) read from HDFS by the query.

Attribute	Display Name	Value Type	Description
hdfs_bytes_read_local	HDFS Local Bytes Read	BYTES	The total number of local bytes read (in GiB) from HDFS by the query.
hdfs_bytes_read_local_percentage	HDFS Local Bytes Read Percentage	NUMBER	The percentage of all bytes read from HDFS by the query that were local.
hdfs_bytes_read_per_second	HDFS Read Throughput	BYTES_PER_SECOND	The overall HDFS read throughput (in B/s) of the query.
hdfs_bytes_read_remote	HDFS Remote Bytes Read	BYTES	The total number of remote bytes read from HDFS by this query.
hdfs_bytes_read_remote_percentage	HDFS Remote Bytes Read Percentage	NUMBER	The percentage of all bytes read from HDFS by this query that were remote.
hdfs_bytes_read_short_circuit	HDFS Short Circuit Bytes Read	BYTES	The total number of bytes (in GiB) read from HDFS by the query that used short-circuit reads.
hdfs_bytes_read_short_circuit_percentage	HDFS Short Circuit Bytes Read Percentage	NUMBER	The percentage of all bytes (in GiB) read from HDFS by the query that used short-circuit reads.
hdfs_bytes_skipped	HDFS Bytes Skipped	BYTES	The total number of bytes that had to be skipped by the query while reading from HDFS. Any number above zero may indicate a problem.
memory_accrual	Memory Accrual	BYTE_SECONDS	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.
memory_aggregate_peak	Aggregate Peak Memory Usage	BYTES	The highest amount of memory allocated by this query at a particular time across all nodes.
memory_per_node_peak	Per Node Peak Memory Usage	BYTES	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.
memory_per_node_peak_node	Node With Peak Memory Usage	STRING	The node with the highest peak memory usage for this query.
network_address	Network Address	STRING	The network address that issued this query.
pool	Pool	STRING	The name of the YARN pool to which this query was issued. Within YARN, a pool is referred to as a queue.
pool_wait_time	Pool Wait Time	MILLISECONDS	The total amount of time the query spent waiting for pool resources to become available.
query_duration	Duration	MILLISECONDS	The duration of the query in milliseconds.
query_id	Query ID	STRING	The ID of the query.

Attribute	Display Name	Value Type	Description
query_state	Query State	STRING	The current state of the query: CREATED, INITIALIZED, COMPILED, RUNNING, FINISHED, UNKNOWN, EXCEPTION. If the query has failed or been canceled, queryState will be EXCEPTION.
query_status	Query Status	STRING	The status of the query. If the query failed, queryStatus will contain diagnostic info such as Memory limit exceeded, Failed to write row ... If canceled, queryStatus is Canceled. Otherwise, queryStatus is OK.
query_type	Query Type	STRING	The type of the query's SQL statement: DML, DDL, QUERY, UNKNOWN.
rows_produced	Rows Produced	NUMBER	The number of rows returned by the query.
service_name	Service Name	STRING	The name of the Impala service.
session_id	Session ID	STRING	The ID of the session that issued this query.
session_type	Session Type	STRING	The type of the session that issued this query.
stats_missing	Stats Mission	BOOLEAN	Whether the query was flagged with a missing table or column statistics warning during the planning process.
statement	Statement	STRING	The query's SQL statement.
thread_cpu_time	Threads: CPU Time	MILLISECONDS	The sum of the CPU time used by all threads of the query.
thread_cpu_time_percentage	Threads: CPU Time Percentage	NUMBER	The sum of the CPU time used by all threads of the query divided by the total thread time.
thread_network_wait_time	Threads: Network Wait Time	MILLISECONDS	The sum of the time spent waiting for the network by all threads of the query.
thread_network_wait_time_percentage	Threads: Network Wait Time Percentage	NUMBER	The sum of the time spent waiting for the network by all threads of the query divided by the total thread time.
thread_storage_wait_time	Threads: Storage Wait Time	MILLISECONDS	The sum of the time spent waiting for storage by all threads of the query.
thread_storage_wait_time_percentage	Threads: Storage Wait Time Percentage	NUMBER	The sum of the time spent waiting for storage by all threads of the query divided by the total thread time.
thread_total_time	Threads: Total Time	MILLISECONDS	The sum of thread CPU, storage wait, and network wait times used by all threads of the query.
user	User	STRING	The user who issued the query.

### Examples

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

## Monitoring and Diagnostics

- 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 Cloudera 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:

- [Query Plan](#)
- [Query Info](#)
- [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

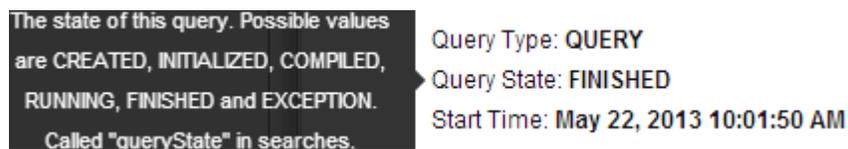
The Query Plan section can help you diagnose and tune performance issues with queries. This information is especially useful to understand performance issues with join queries, such as inefficient order of tables in the SQL statement, lack of table and column statistics, and the need for query hints 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

#### 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 fields and their names might change as Impala internal processing is enhanced.

## 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* > **Activities** > *YARN service name* **Applications**.
- On the **Home** > **Status** tab, select *YARN service name* and click the **Applications** tab.

The YARN jobs run during the last day are listed in the [Applications list](#).

## Configuring YARN Application Monitoring

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

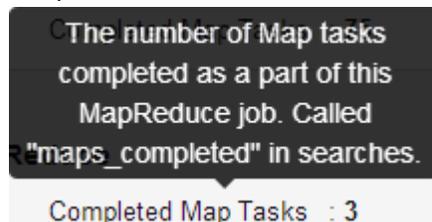
For information on how to configure whether admin and non-admin users can view all applications, only that user's applications, or no applications, see [Configuring Application Visibility](#) on page 16.

## Jobs List

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

10/16/2013 11:52 AM - 10/16/2013 11:52 AM	Query: SELECT s07.description, s07.total_emp, s08.total_emp, s07.salary FROM sample_07 s07 JOIN sample_08 s08 ON ( s07.code = s08.code ) WHERE ( s07.total_emp > s08.total_emp AND s07.salary > 100000 ) ORDER BY s07.salary DESC LIMIT 1000 Name: SELECT s07.description, s07.total_emp...1000(Stage-2) Pool: default	Actions
Type: Hive	ID: job_1381947538718_0003	Duration: 14.93s
User: hdfs		

You can add additional attributes to the summary with the [Attribute Selector](#). To display information about job attributes and possible values, hover over a field in an entry. For example:



A running job displays **▶ Running** under the start timestamp.

To kill a running job, select **Actions** > **Kill**. Only an administrator can kill jobs and killing a job creates an audit event.

When you kill a job **⊗ Killed** replaces the **▶ Running** label. Once the page is refreshed, the entry is removed from the list.

To view a completed job in the JobHistory server, select **Actions** > **View on JobHistory Server**.

To display all the jobs run by the same user, select **Actions** > **User's YARN applications**.

## 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 is made up of three components:

- **Attribute** - the query language name of the attribute.
- **Operator** - the type of comparison between the attribute and the attribute value. Cloudera Manager supports the standard comparator operators: =, !=, >, <, >=, <=, and `RLIKE`, which does 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 compare 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, surround subexpressions with 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**

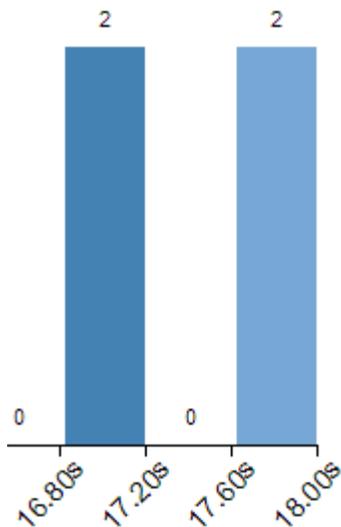
1. Click the



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 Attribute Histograms**

1. Optionally, click the **Select Attributes** link to display a dialog where you can chose which attributes to display in histograms. Check the checkbox next to one or more attributes, and click **Close**.
2. Click the **Enhance Filter** link. Histograms of the selected attributes display with the number of results that match each value of the selected attributes.
3. Click a histogram bar that represents the range of attribute values to filter on. The color of the histogram bar gets lighter (on the right below)



and a filter with the attribute value set to the range of the histogram bucket is added to the text box. The range includes the lower bound of the bucket and excludes the upper bound of the bucket. For example:

```
(<x>_duration >= 17600.0 AND <x>_duration < 18000.0)
```

where <x> is query or application.

If you click the same histogram bar again, the color reverts to the darker blue and the filter is removed from the text box.

If you click another histogram bar, another filter is OR'd with the existing filter:

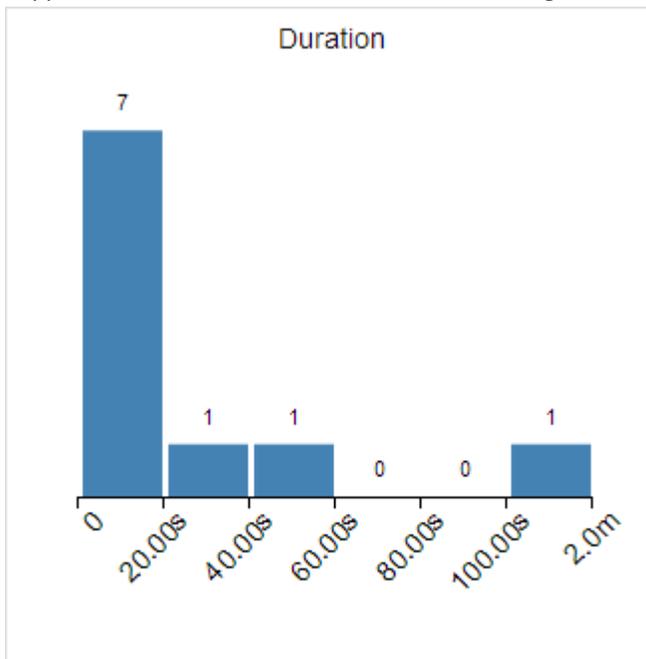
```
(<x>_duration >= 17600.0 AND <x>_duration < 18000.0 OR <x>_duration >= 16000.0 AND <x>_duration < 16400.0)
```

- **Type a Filter**

1. Start typing or press **Spacebar** in the text box. As you type, filter attributes matching the letter you type 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. Put the cursor on 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.

#### Example: Drilling into Query Results

Suppose we have a set of results with the following duration distribution :



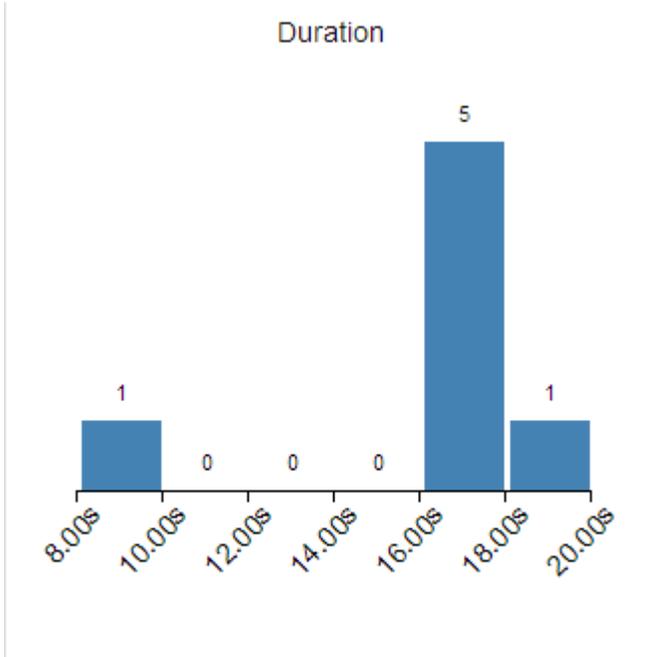
The 0-20.00s bucket has 7 results, but with the current distribution we cannot discriminate between the results in that bucket.

## Monitoring and Diagnostics

Selecting the left-most histogram bar adds the following filter to the text box:

```
<x>_duration >= 0.0 AND <x>_duration < 20000.0
```

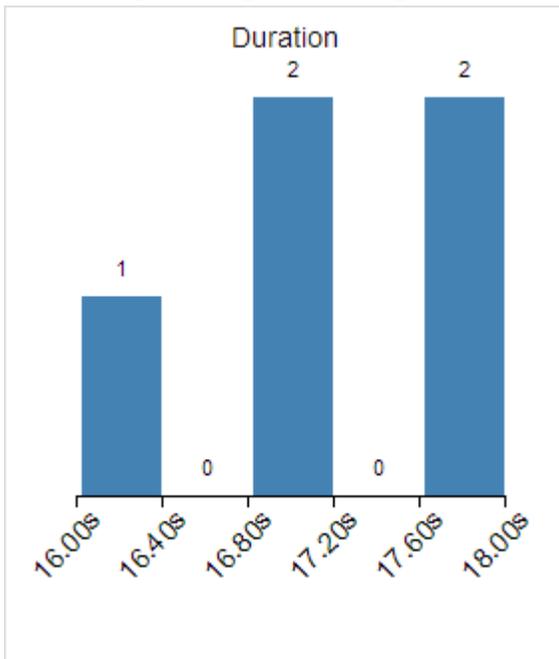
where <x> is query or application. After clicking **Search** again, the histogram appears as follows:



Selecting the histogram with 5 results again refines the filter to:

```
<x>_duration >= 16000.0 AND <x>_duration < 18000.0
```

After clicking **Search** again, the histogram appears as follows:



## Filter Attributes

Commonly-used filter attributes, their names as they are displayed in Cloudera Manager, their types, and descriptions, are enumerated below.

Attribute	Display Name	Value Type	Description
application_duration	Duration	integer (milliseconds)	How long YARN took to execute this application.
application_id	Application ID	string	The ID of the YARN application.
cpu_time	Total CPU Time	integer (milliseconds)	The total amount of CPU time used by the tasks for this YARN application.
disk_input_bytes	Disk Input Bytes	integer (bytes)	The number of bytes read from local files by this YARN application.
disk_output_bytes	Disk Output Bytes	integer (bytes)	The number of bytes written to local files by this YARN application.
mapper_class	Map Class	string	The class used by the map tasks in this YARN application.
name	Name	string	Name of the YARN application.
pool	Pool	string	The name of the pool that this application was submitted to. Within YARN a pool is referred to as a queue.
reducer_class	Reduce Class	string	The class used by the reduce tasks in this YARN application.
service_name	Service Name	string	The name of the YARN service.
shuffle_bytes	Shuffle Bytes	integer (bytes)	The number of bytes fetched from mappers over HTTP during the reduce phase.
state	Application State	string	The state of this YARN application. This reflects the ResourceManager state while the application is executing and the Job History Server state after the application has completed.
user	User	string	The user who ran the YARN application.

## 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`.

## 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 <a href="#">Activity Duration Rules</a> on page 15.
AUDIT_EVENT	Generated by actions performed <ul style="list-style-type: none"> <li>In Cloudera Manager, such as creating, configuring, starting, stopping, and deleting services or roles</li> <li>By services that are being audited by Cloudera Navigator.</li> </ul>
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 <b>Configuration</b> tabs for HBase, HDFS, Impala, and MapReduce service instances, at both the service and role level. See <a href="#">Configuring Health Monitoring</a> on page 14 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 <a href="#">matches a set of rules</a> 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.
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 () 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 redisplay 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 **x** 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.

## Alerts

An **alert** is an event that is considered especially noteworthy and is triggered by a selected event. Alerts are shown with an  badge when they appear in a list of [events](#). You can configure the Alert Publisher to send alert notifications by email or via SNMP trap to a trap receiver.

Service instances of type HDFS, MapReduce, and HBase (and their associated roles) can generate alerts if so configured. Alerts can also be configured for the monitoring roles that are a part of the Cloudera Management Service.

The settings to enable or disable specific alerts are found under the Configuration tab for the services to which they pertain. See [Configuring Alerts](#) on page 16 and for more information on setting up alerting.

## Monitoring and Diagnostics

For information about configuring the Alert Publisher to send email or SNMP notifications for alerts, see [Configuring Alert Delivery](#) on page 18.

### Viewing What Alerts are Enabled and Disabled

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

Do one of the following:

- Select **Administration > Alerts**.
- Display the All Alerts Summary page:
  1. Do one of the following:
    - Select **Clusters > Cloudera Management Service > Cloudera Management Service**.
    - On the Status tab of the Home page, in **Cloudera Management Service** table, click the **Cloudera Management Service** link.
  2. Click the **Instances** tab.
  3. Click an **Alert Publisher** role.
  4. Click the **All Alerts Summary** tab.

## 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. You can create a trigger in either of the following ways:

- By directly editing the configuration for the service, role (or role configuration group), or host configuration
- By 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 page.



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

### 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. 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 will 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 condition's 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`, that is, triggers are enabled by default.

### 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"}]
```

## Audit Events

**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 service, role, or host instance. In Cloudera Manager, audit event logs display service, role, and host life cycle (create, delete, start, stop, and so on) and security-related (add and delete user) events recorded by Cloudera Manager management services and service access events recorded by Cloudera Navigator. For information on the latter, see [Audit Events and Audit Reports](#).

The 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 31.

### Viewing Audit Events

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

Object	Procedure
<b>Cluster</b>	<ol style="list-style-type: none"> <li>1. Click the <b>Audits</b> tab on the top navigation bar.</li> </ol>
<b>Service</b>	<ol style="list-style-type: none"> <li>1. Click the <b>Clusters</b> tab on the top navigation bar.</li> <li>2. Select a service.</li> <li>3. Click the <b>Audits</b> tab on the service navigation bar.</li> </ol>
<b>Role</b>	<ol style="list-style-type: none"> <li>1. Click the <b>Clusters</b> tab on the top navigation bar.</li> <li>2. Select a service.</li> <li>3. Click the <b>Instances</b> tab on the Services navigation bar.</li> <li>4. Select a role.</li> <li>5. Click the <b>Audits</b> tab on the role navigation bar.</li> </ol>
<b>Host</b>	<ol style="list-style-type: none"> <li>1. Click the <b>Hosts</b> tab on the top navigation bar.</li> <li>2. Select a host.</li> <li>3. Click the <b>Audits</b> tab on the host navigation bar.</li> </ol>

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 redisplay 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 Event Logs

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 Cloudera Impala - `database:tablename`
  - HBase - `table family:qualifier`

For Hive, Hue, and Cloudera 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. For a list of all the metrics supported by Cloudera Manager, see [Cloudera Manager Metrics](#).

#### 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



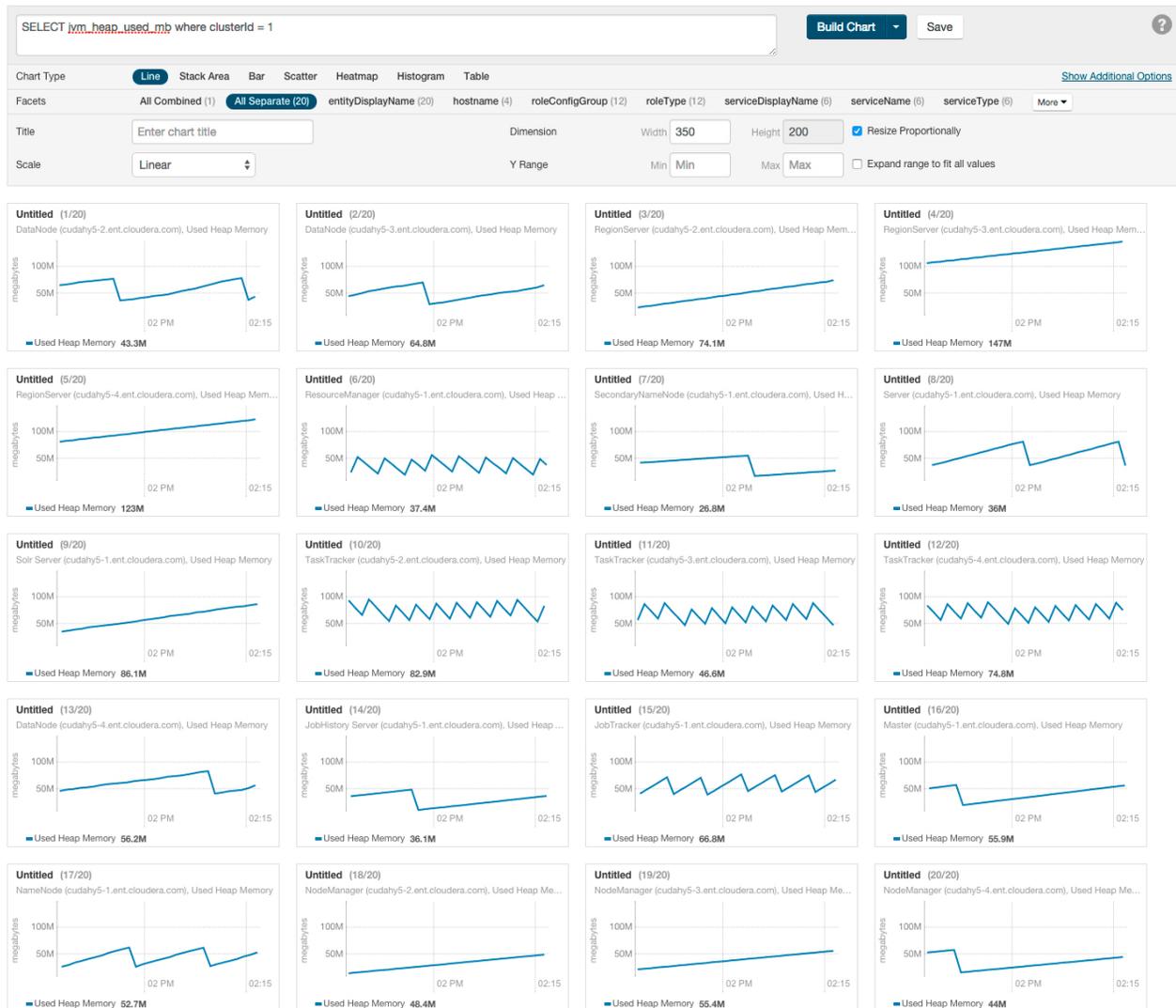
to the right of the **Build Chart** button to display a list of examples with descriptions.

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**)

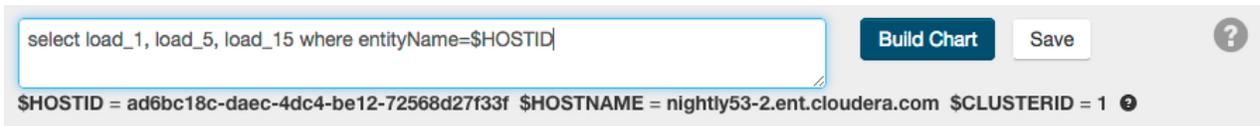
A time-series query returns one or more time series or scalar values. By default a maximum of 250 time series will be returned.

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**.

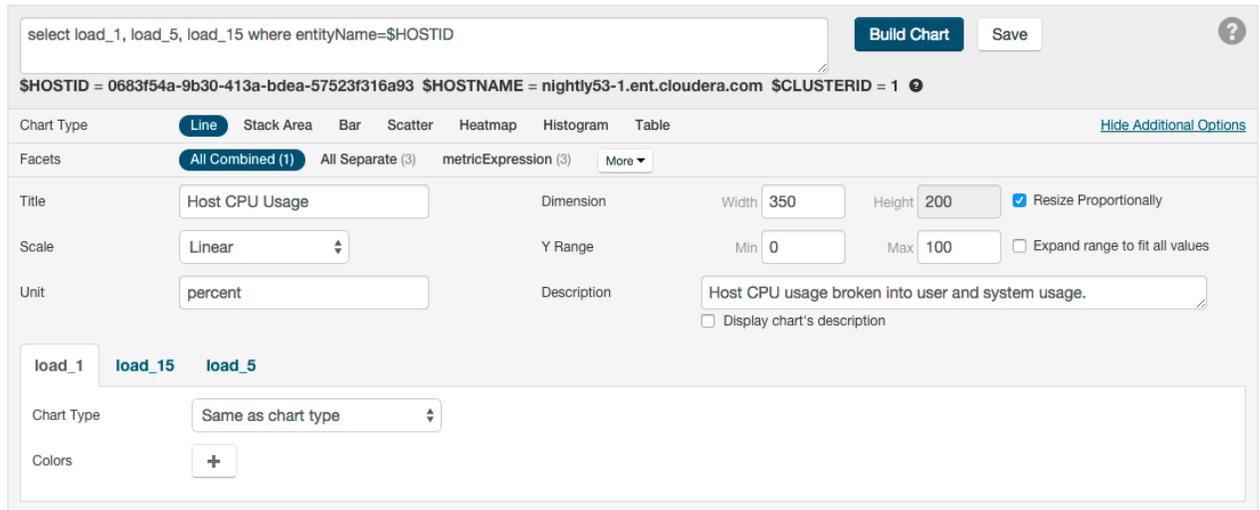
## Using Context-Sensitive Variables in Charts

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.

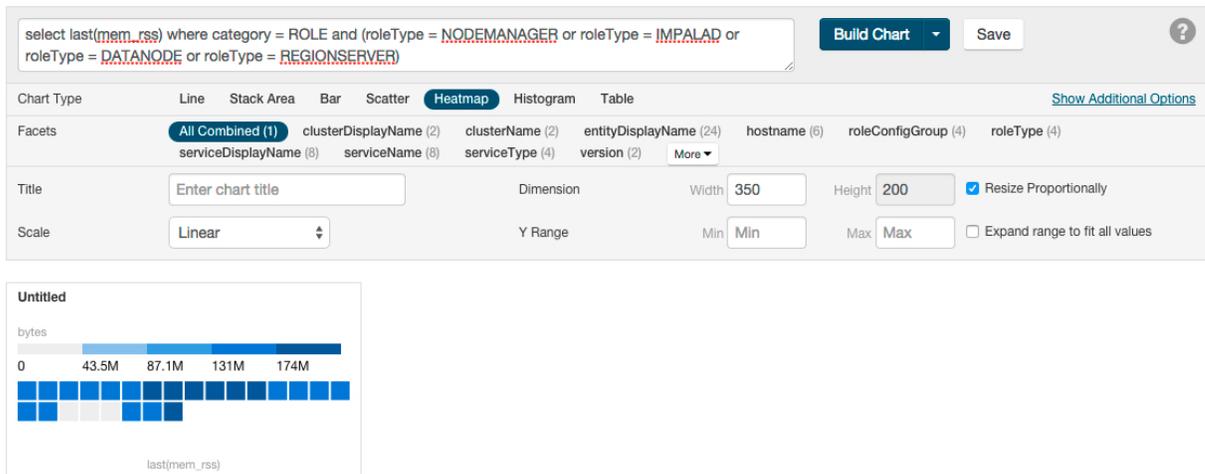
### Chart Properties

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.

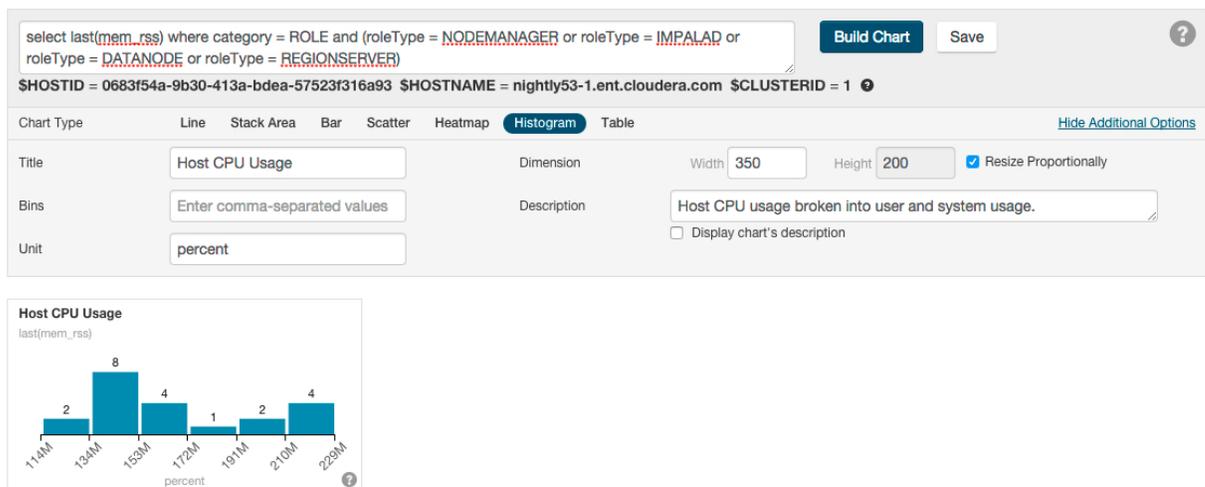
#### Changing the Chart Type

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.



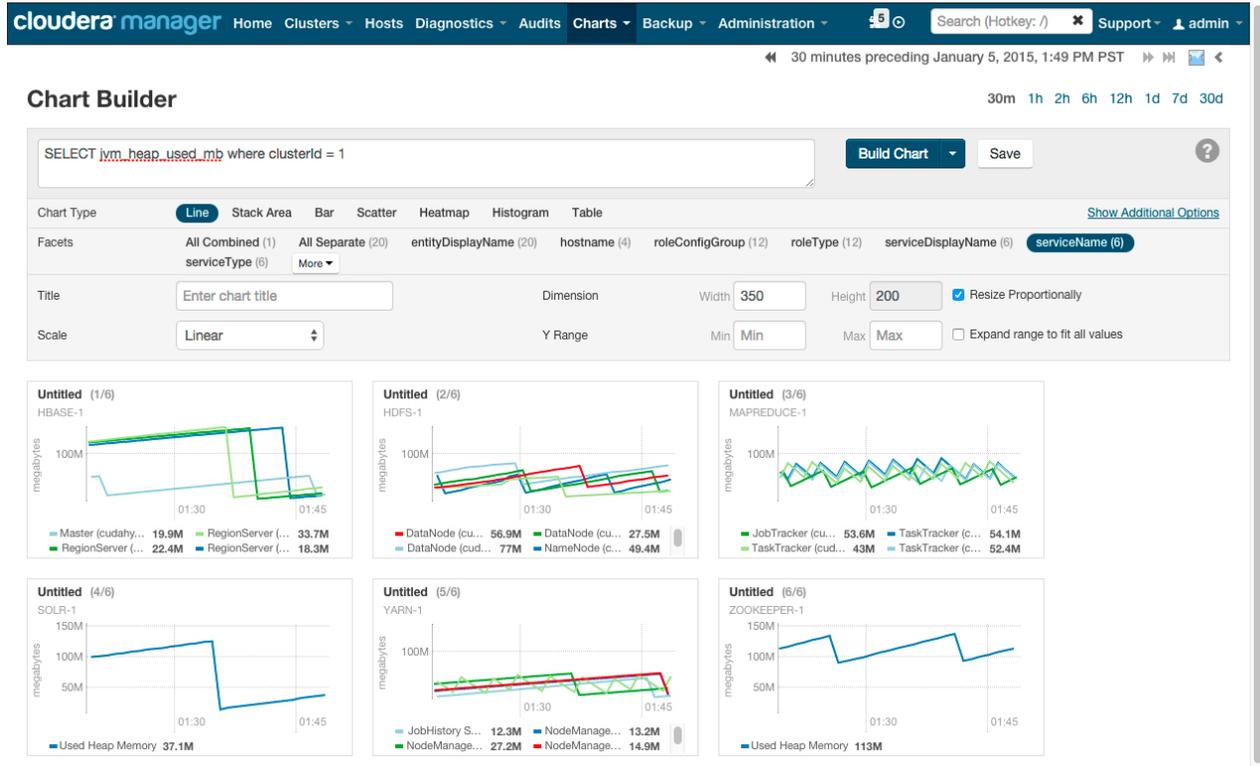
**Note:** Heatmaps and histograms render charts for a single point as opposed to time series charts that render a series of points. For queries that return time series, Cloudera Manager will generate the heatmap or histogram based on the last recorded point in the series, and will issue the warning: "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.



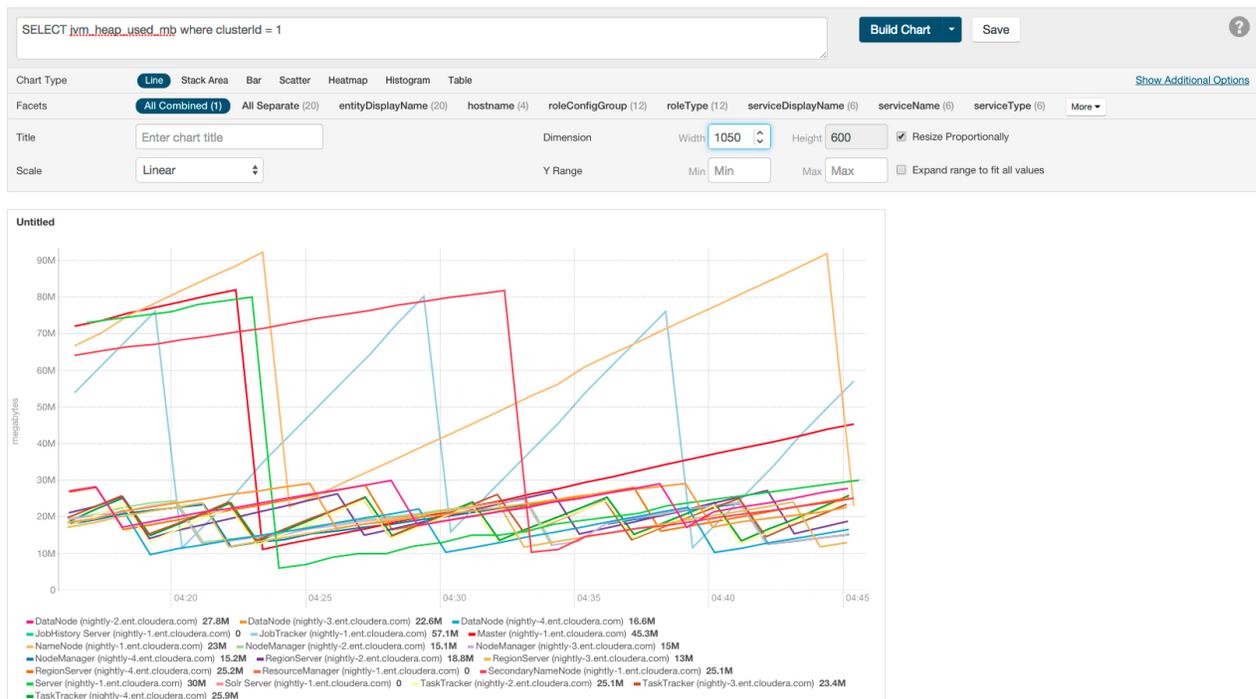
### Changing Scale

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** -

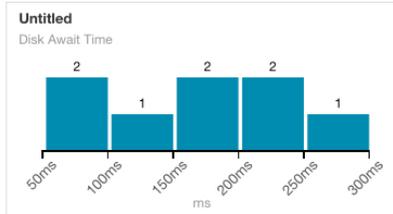
- Mousing over the upper right corner of a histogram and clicking the expand arrows opens a pop-up containing the query that generated the chart, an expanded view of the chart, a list of entity names and links to the entities whose metrics are represented by the histogram bars, and the value of the metric for each entity. For example, clicking the following histogram

select await\_time where category = DISK and device = vda

Build Chart Save

Chart Type: Line Stack Area Bar Scatter Heatmap **Histogram** Table [Show Additional Options](#)

Title: Enter chart title Dimension Width: 350 Height: 200  Resize Proportionally



displays the following:

Query: select last(await\_time) where (category = "DISK" and device = "vda")

30 minutes preceding January 7, 2015, 4:36 PM PST

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

last(await\_time)

Legend **Statistics**

All Entities

Entity Name	Value
<a href="#">vda (nightly-2.ent.cloudera.com)</a>	276ms
<a href="#">vda (nightly-3.ent.cloudera.com)</a>	219ms
<a href="#">vda (nightly-7.ent.cloudera.com)</a>	200ms
<a href="#">vda (nightly-8.ent.cloudera.com)</a>	185ms
<a href="#">vda (nightly-1.ent.cloudera.com)</a>	180ms
<a href="#">vda (nightly-6.ent.cloudera.com)</a>	119ms
<a href="#">vda (nightly-5.ent.cloudera.com)</a>	82ms
<a href="#">vda (nightly-4.ent.cloudera.com)</a>	71ms

Close

- Clicking a bar in the expanded histogram displays a line chart of the time series from which the histogram was generated: Consult the documentation for your database for details.

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 74
- [Saving Charts to an Existing Dashboard](#) on page 75

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.

## Monitoring and Diagnostics

### 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.
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 page 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. 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. 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 and drop charts on 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 data store.

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_committed_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.

For a list of all the supported metrics, see [Cloudera Manager Metrics](#). Also see the [FAQ](#) on page 83 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
<code>avg(metric expression)</code>	N	Computes a simple average for a time series.
<code>count_service_roles()</code>	Y	Returns the number of roles. There are three variants of this function: <ul style="list-style-type: none"> <li>• <code>count_service_roles(roleType, roleState)</code> - Returns the number of roles of the specified <a href="#">roleType</a> and <a href="#">roleState</a>. For example, <code>count_service_roles(datanode, running)</code> returns the number of running DataNodes.</li> <li>• <code>count_service_roles(roleType)</code> - Returns the number of roles with the specified <code>roleType</code>.</li> <li>• <code>count_service_roles()</code> - Return the number of roles. For example, <code>select events_critical where count_service_roles() &gt; 100</code> returns the <code>event_critical</code> metric when the number of roles is greater than 100.</li> </ul>
<code>dt(metric expression)</code>	N	Derivative with negative values. The change of the underlying metric expression per second. For example: <code>dt(jvm_gc_count)</code> .
<code>dt0(metric expression)</code>	N	Derivative where negative values are skipped (useful for dealing with counter resets). The change of the underlying metric expression per second. For example: <code>dt0(jvm_gc_time_ms) / 10</code> .
<code>getClusterFact(string factName, double defaultValue)</code>	Y	Retrieves a fact about a cluster. Currently supports one fact: <code>numCores</code> . If the number of cores cannot be determined, <code>defaultValue</code> is returned.

Function	Returns Scalar?	Description
<code>getHostFact(string factName, double defaultValue)</code>	Y	Retrieves a fact about a host. Currently supports one fact: <code>numCores</code> . If the number of cores cannot be determined, <code>defaultValue</code> is returned.  For example, <code>select dt(total_cpu_user) / getHostFact(numCores, 2)</code> where <code>category=HOST</code> divides the results of <code>dt(total_cpu_user)</code> by the current number of cores for each host.  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 <code>getHostFact</code> function and multiplies the result by 100 to get the percentage.  <code>select dt0(total_cpu_user)/getHostFact(numCores,1)*100, dt0(total_cpu_system)/getHostFact(numCores,1)*100 where category=ROLE and clusterId=1</code>
<code>greatest(metric expression, scalar metric expression)</code>	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 <code>max(point, scalar metric expression)</code> .
<code>integral(metric expression)</code>	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, <code>select integral(maps_failed_rate)</code> will return the count of the failed number of maps.
<code>last(metric expression)</code>	Y	Returns the last point of a time series. For example, to use the last point of the <code>cpu_percent</code> metric time series, use the expression <code>select last(cpu_percent)</code> .
<code>least(metric expression, scalar metric expression)</code>	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 <code>min(point, scalar metric expression)</code> .
<code>max(metric expression)</code>	Y	Computes the maximum value of the time series. For example, <code>select max(cpu_percent)</code> .
<code>min(metric expression)</code>	Y	Computes the minimum value of the time series.
<code>moving_avg(metric expression, time_window_sec)</code>	N	Computes the moving average for a time series over a time window <code>time_window_sec</code> specified in seconds (2, 0.1, and so on)
<code>stats(metric expression, stats name)</code>	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: <code>max</code> , <code>min</code> , <code>avg</code> , <code>std_dev</code> , and <code>sample</code> . For rate metrics, the underlying counter value is available using the "counter" statistics. For example, <code>stats(fd_open_across_datanodes, max)</code> or <code>stats(swap_out_rate, counter)</code> .
<code>sum(metric expression)</code>	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 `xxxName` 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
```

## 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
<code>active</code>	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, <code>active=true</code> ). To access time series for deleted or removed entities, specify <code>active=false</code> in the query. For example: <pre>SELECT fd_open WHERE roleType=DATANODE and active=false</pre>
<code>agentName</code>	A Flume agent name.
<code>applicationName</code>	One of the Cloudera Manager monitoring daemon names.
<code>cacheId</code>	The HDFS cache directive ID.
<code>category</code>	The category of the entities returned by the query: <code>CLUSTER</code> , <code>DIRECTORY</code> , <code>DISK</code> , <code>FILESYSTEM</code> , <code>FLUME_SOURCE</code> , <code>FLUME_CHANNEL</code> , <code>FLUME_SINK</code> , <code>HOST</code> , <code>HTABLE</code> , <code>IMPALA_QUERY_STREAM</code> , <code>NETWORK_INTERFACE</code> , <code>ROLE</code> , <code>SERVICE</code> , <code>USER</code> , <code>YARN_APPLICATION_STREAM</code> , <code>YARN_QUEUE</code> .  Some metrics are collected for more than one type of entity. For example, <code>total_cpu_user</code> is collected for entities of category <code>HOST</code> and <code>ROLE</code> . To retrieve the data only for hosts use: <pre>select total_cpu_user where category=HOST</pre> The <code>ROLE</code> category applies to all role types (see <code>roleType</code> attribute). The <code>SERVICE</code> category applies to all service types (see <code>serviceType</code> attribute). For example, to retrieve the committed heap for all roles on <code>host1</code> use: <pre>select jvm_committed_heap_mb where category=ROLE and hostname="host1"</pre>
<code>clusterDisplayName</code>	The user-defined display name of a cluster.
<code>clusterName</code>	The cluster ID. To specify the cluster by its display name, use the <code>clusterDisplayName</code> attribute.
<code>componentName</code>	A Flume component name. For example, <code>channel1</code> , <code>sink1</code> .
<code>device</code>	A disk device name. For example, <code>sda</code> .
<code>displayName</code>	The display name of an entity. <div style="border: 1px solid black; padding: 5px; margin-top: 10px;"> <b>Note:</b> The <code>displayName</code> attribute was removed in Cloudera Manager 5. Older queries should be modified to use the cluster or service display name attributes, <code>clusterDisplayName</code> and <code>serviceDisplayName</code>.</div>
<code>entityName</code>	A display name plus unique identifier. For example: <code>HDFS-1-DATANODE-692d141f436ce70aac080aedbe83f887</code> .
<code>expired</code>	A Boolean that indicates whether an HDFS cache directive expired.
<code>groupName</code>	A user group name.
<code>hbaseNamespace</code>	The name of the HBase namespace.
<code>hostId</code>	The canonical identifier for a host in Cloudera Manager. It is unique and immutable. For example: <code>3d645222-2f7e-4895-ae51-cd43b91f1e7a</code> .

Name	Description
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.
mountpoint	A mount point name. For example, /var, /mnt/homes.
nameserviceName	The name of the HDFS nameservice.
ownerName	The owner user name.
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.

Name	Description
<code>solrShardName</code>	The Solr shard name. For example, <code>shard1</code> .
<code>systemTable</code>	A boolean indicating whether the HBase table is a system table or not.
<code>tableName</code>	The name of a table.
<code>userName</code>	The name of the user.
<code>version</code>	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.

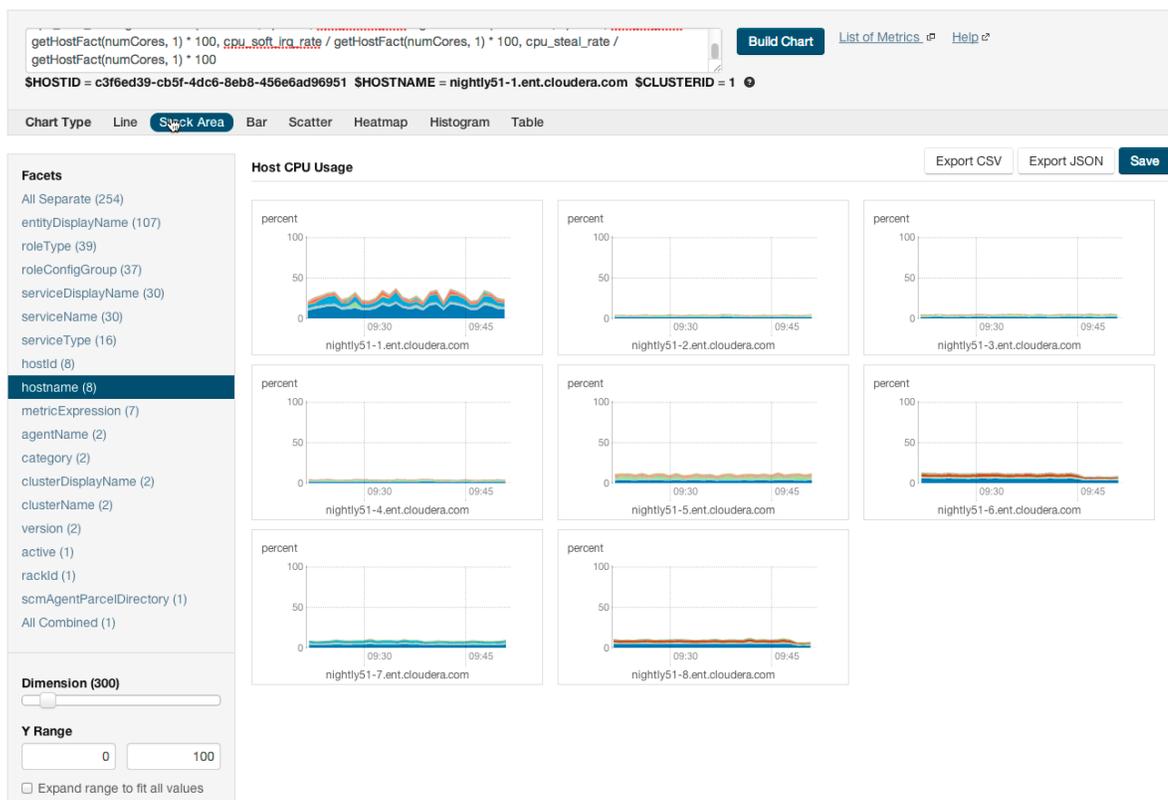
Entity	Attributes
All Roles	<code>roleType</code> , <code>hostId</code> , <code>hostname</code> , <code>rackId</code> , <code>serviceType</code> , <code>serviceName</code>
All Services	<code>serviceName</code> , <code>serviceType</code> , <code>clusterId</code> , <code>version</code> , <code>serviceDisplayName</code> , <code>clusterDisplayName</code>
Agent	<code>roleType</code> , <code>hostId</code> , <code>hostname</code> , <code>rackId</code> , <code>serviceType</code> , <code>serviceName</code> , <code>clusterId</code> , <code>version</code> , <code>agentName</code> , <code>serviceDisplayName</code> , <code>clusterDisplayName</code>
Cluster	<code>clusterId</code> , <code>version</code> , <code>clusterDisplayName</code>
Directory	<code>roleName</code> , <code>hostId</code> , <code>path</code> , <code>roleType</code> , <code>hostname</code> , <code>rackId</code> , <code>serviceType</code> , <code>serviceName</code> , <code>clusterId</code> , <code>version</code> , <code>agentName</code> , <code>hostname</code> , <code>clusterDisplayName</code>
Disk	<code>device</code> , <code>logicalPartition</code> , <code>hostId</code> , <code>rackId</code> , <code>clusterId</code> , <code>version</code> , <code>hostname</code> , <code>clusterDisplayName</code>
File System	<code>hostId</code> , <code>mountpoint</code> , <code>rackId</code> , <code>clusterId</code> , <code>version</code> , <code>partition</code> , <code>hostname</code> , <code>clusterDisplayName</code>
Flume Channel	<code>serviceName</code> , <code>hostId</code> , <code>rackId</code> , <code>roleName</code> , <code>flumeComponent</code> , <code>roleType</code> , <code>serviceType</code> , <code>clusterId</code> , <code>version</code> , <code>agentName</code> , <code>serviceDisplayName</code> , <code>clusterDisplayName</code>
Flume Sink	<code>serviceName</code> , <code>hostId</code> , <code>rackId</code> , <code>roleName</code> , <code>flumeComponent</code> , <code>roleType</code> , <code>serviceType</code> , <code>clusterId</code> , <code>version</code> , <code>agentName</code> , <code>serviceDisplayName</code> , <code>clusterDisplayName</code>
Flume Source	<code>serviceName</code> , <code>hostId</code> , <code>rackId</code> , <code>roleName</code> , <code>flumeComponent</code> , <code>roleType</code> , <code>serviceType</code> , <code>clusterId</code> , <code>version</code> , <code>agentName</code> , <code>serviceDisplayName</code> , <code>clusterDisplayName</code>
HDFS Cache Pool	<code>serviceName</code> , <code>poolName</code> , <code>nameserviceName</code> , <code>serviceType</code> , <code>clusterId</code> , <code>version</code> , <code>groupName</code> , <code>ownerName</code> , <code>serviceDisplayName</code> , <code>clusterDisplayName</code>
HNamespace	<code>serviceName</code> , <code>namespaceName</code> , <code>serviceType</code> , <code>clusterId</code> , <code>version</code> , <code>serviceDisplayName</code> , <code>clusterDisplayName</code>
Host	<code>hostId</code> , <code>rackId</code> , <code>clusterId</code> , <code>version</code> , <code>hostname</code> , <code>clusterDisplayName</code>
HRegion	<code>htableName</code> , <code>hregionName</code> , <code>hregionStartTimeMs</code> , <code>namespaceName</code> , <code>serviceName</code> , <code>tableName</code> , <code>serviceType</code> , <code>clusterId</code> , <code>version</code> , <code>roleType</code> , <code>hostname</code> , <code>roleName</code> , <code>hostId</code> , <code>rackId</code> , <code>serviceDisplayName</code> , <code>clusterDisplayName</code>
HTable	<code>namespaceName</code> , <code>serviceName</code> , <code>tableName</code> , <code>serviceType</code> , <code>clusterId</code> , <code>version</code> , <code>serviceDisplayName</code> , <code>clusterDisplayName</code>
Network Interface	<code>hostId</code> , <code>networkInterface</code> , <code>rackId</code> , <code>clusterId</code> , <code>version</code> , <code>hostname</code> , <code>clusterDisplayName</code>
Rack	<code>rackId</code>
Service	<code>serviceName</code> , <code>serviceType</code> , <code>clusterId</code> , <code>serviceDisplayName</code>
Solr Collection	<code>serviceName</code> , <code>serviceType</code> , <code>clusterId</code> , <code>version</code> , <code>serviceDisplayName</code> , <code>clusterDisplayName</code>

Entity	Attributes
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 69 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?**

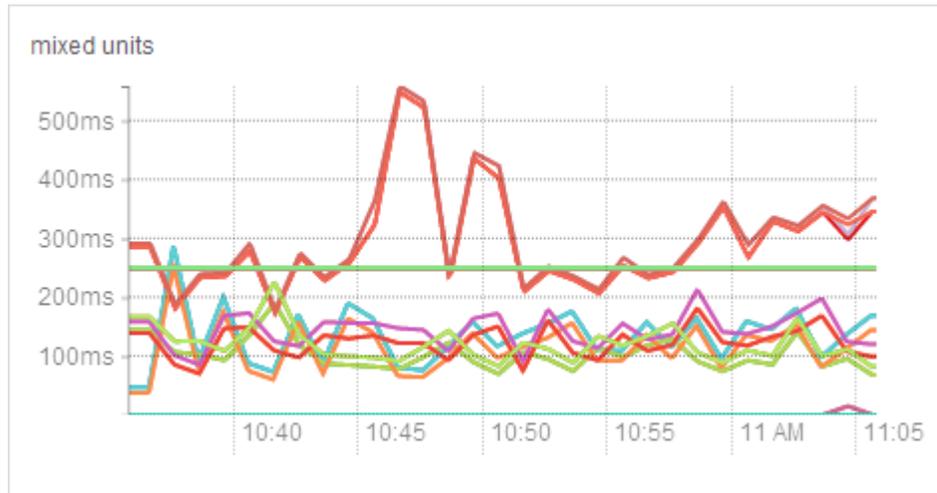
Some metrics, like `bytes_read` represent a counter that always grows. For such metrics a negative rate means that the counter has been reset (for example, process restarted, host restarted, and so on). Use `dt0` for these metrics.

**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 69 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 67.

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.
- Click the **List of Metrics** link on the Charts page to the right of the Build Chart button.
- 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
  }
}
```

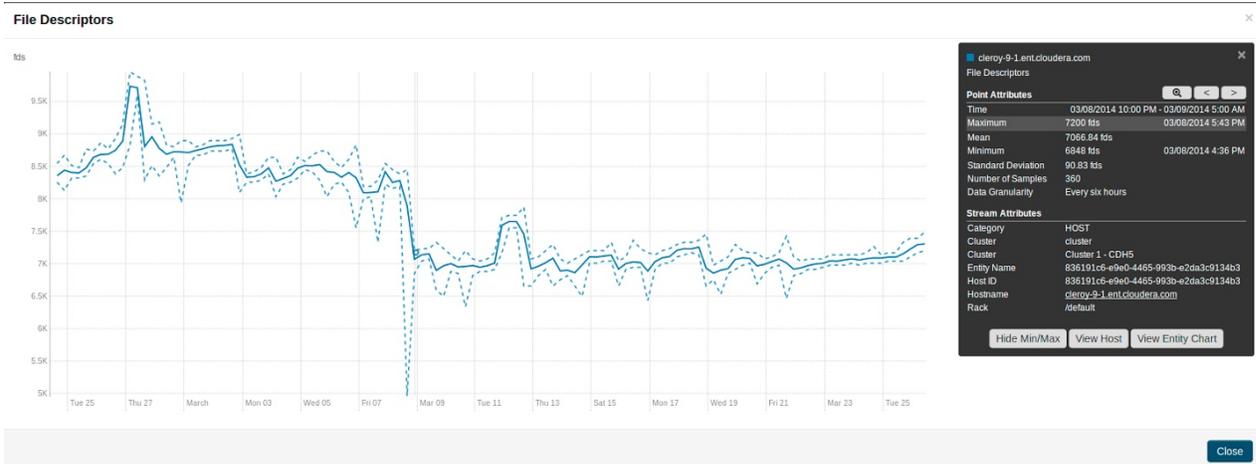
A raw cross-entity aggregate:

```
{
  "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 aggregateStatistics 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 18.

### 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 35).
- **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 89 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 deselect all services and roles in one operation. You can expand each service and limit the search to specific roles by selecting or deselecting 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 Cloudera Manager Server and Agent Logs

To help you troubleshoot problems, you can view the Cloudera Manager Server and Agent logs. You can view these logs in the Logs page or in specific pages for the logs.

### Viewing Cloudera Manager Server and 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. Check the **Cloudera Manager** checkbox to view both Agent and Server logs, or click ▶ to the left of Cloudera Manager, and check either the **Agent** or **Server** checkbox.
5. Click **Search**.

For more information about the Logs page, see [Logs](#) on page 87.

### 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 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.



**Note:** You can also view the Cloudera Manager Agent log at `/var/log/cloudera-scm-agent/cloudera-scm-agent.log` on the Agent hosts.

## Reports



**Important:** This feature is available only with a Cloudera Enterprise license.

For other licenses, the following applies:

- Cloudera Express- The feature is not available.
- Cloudera Enterprise Data Hub Edition Trial - The feature is available until you end the trial or the trial license expires.

To obtain a license for Cloudera Enterprise, fill in this [form](#) or call 866-843-7207. After you install a Cloudera Enterprise license, the feature will be available.

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 > General > 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.

## 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.

### 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:

<b>Bytes</b>	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.
<b>Raw Bytes</b>	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.
<b>File and Directory Count</b>	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 92).

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 92).

### 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

## Monitoring and Diagnostics

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

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 on 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. See [Downloading HDFS Directory Access Permission Reports](#).

## 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
<b>Cloudera Manager</b>		
<p>The Cloudera Manager service will not be running as it exited abnormally.</p> <p>Running <code>service cloudera-scm-server status</code> will print following message "cloudera-scm-server dead but pid file exists".</p> <p>The Cloudera Manager Server log file <code>/var/log/cloudera-scm-server/cloudera-scm-server.log</code> will have a stacktrace with "java.lang.OutOfMemoryError" logged.</p>	<p>Out of memory.</p>	<p>Examine the heap dump that the Cloudera Manager Server creates when it runs out of memory. The heap dump file is created in the <code>/tmp</code> directory, has file extension <code>.hprof</code> and file permission of 600. Its owner and group will be the owner and group of the Cloudera Manager server process, normally <code>cloudera-scm:cloudera-scm</code>.</p>
<p>You are unable to start service on the Cloudera Manager server, that is, <code>service cloudera-scm-server start</code> does not work and there are errors in the log file located at <code>/var/log/cloudera-scm-server/cloudera-scm-server.log</code></p>	<p>The server has been disconnected from the database or the database has stopped responding or has shut down.</p>	<p>Go to <code>/etc/cloudera-scm-server/db.properties</code> and make sure the database you are trying to connect to is listed there and has been started.</p>
<p>Logs include APPARENT DEADLOCK entries for c3p0.</p>	<p>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.</p>	<p>There are a variety of ways to react to these log entries.</p> <ul style="list-style-type: none"> <li>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.</li> <li>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 <code>maxAdministrativeTaskTime</code>. For more information, see <a href="#">maxAdministrativeTaskTime</a>.</li> </ul>

Symptom	Reason	Solution
		<ul style="list-style-type: none"> <li>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 <a href="#">numHelperThreads</a>.</li> </ul>
<b>Starting Services</b>		
<p>After you click the <b>Start</b> button to start a service, the <b>Finished</b> status doesn't display.</p> <p>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.</p>	<p>The host is disconnected from the Server, as will be indicated by missing heartbeats on the <b>Hosts</b> tab.</p> <p>Subcommands failed resulting in errors in the log file indicating that either the command timed out or the target port was already occupied</p>	<ul style="list-style-type: none"> <li>Look at the logs for the service for causes of the problem.</li> <li>Restart the Agents on the hosts where the heartbeats are missing.</li> <li>Look at the log file at <code>/var/log/cloudera-scm-server/cloudera-scm-server.log</code> for more details on the errors. For example, if the port is already occupied you should see an "Address in use" error.</li> <li>Navigate to the <b>Hosts &gt; Status</b> tab. Click on the <b>Name</b> of the host you want to inspect. Now go to the <b>Processes</b> tab and check the <b>Stdout/Stderr</b> logs to diagnose the cause of the failure. For example, if any binaries are missing or if Java could not be found.</li> </ul>
<p>After you click <b>Start</b> to start a service, the <b>Finished</b> status displays but there are error messages. The subcommands to start service components (such as JobTracker and one or more TaskTrackers) do not start.</p>	<p>A port specified in the <b>Configuration</b> tab of the service is already being used in your cluster. For example, the JobTracker port is in use by another process.</p> <p>There are incorrect directories specified in the <b>Configuration</b> tab of the service (such as the log directory).</p>	<p>Enter an available port number in the port property (such as <b>JobTracker port</b>) in the <b>Configuration</b> tab of the service.</p> <p>Enter correct directories in the <b>Configuration</b> tab of the service.</p>
<b>Job is Failing</b>	No space left on device.	<p>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:</p> <p>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:  <a href="http://JTHost:50030/taskdetails.jsp?tipid=TaskID">http://JTHost:50030/taskdetails.jsp?tipid=TaskID</a>.  You can see on which host the task is failing in the Machine column.</p> <p>In the NameNode Web UI, inspect the % used column on the NameNode Live Nodes page:  <a href="http://namenode:50070/dfsnodelist.jsp?whatNodes=LIVE">http://namenode:50070/dfsnodelist.jsp?whatNodes=LIVE</a>.</p>
<b>Send Test Alert and Diagnose SMTP Errors</b>		

Symptom	Reason	Solution
<p>You have enabled sending alerts from the Cloudera Manager Admin Console, however, Cloudera Manager does not seem to be sending any alerts.</p> <p>Using the <b>Send Test Alert</b> link under <b>Administration &gt; Alerts</b> shows success even though you do not receive an alert email.</p>	<p>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.</p>	<p>Use the following steps to make changes to the Alert Publisher configuration:</p> <ol style="list-style-type: none"> <li>1. In the Cloudera Manager Admin Console, click the Cloudera Management Service.</li> <li>2. Click the <b>Configuration</b> tab.</li> <li>3. Expand the <b>Alert Publisher Default Group</b> category and change <b>Alerts: Mail Server Protocol</b> to <code>smtp</code> (or <code>smtps</code>).</li> <li>4. Click the <b>Ports and Addresses</b> category and change <b>Alerts: Mail Server TCP Port</b> to 25 (or to 465 for SMTPS)</li> <li>5. Click <b>Save Changes</b> and restart the Alert Publisher.</li> </ol>

## Logs and Events

For information about problems, check the logs and events:

- [Logs](#) on page 87 present log information for services, filtered by role, host, or keywords as well log level (severity).
- [Viewing Cloudera Manager Server and Agent Logs](#) on page 89 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.

## Appendix: Apache License, Version 2.0

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.
```