Cloudera Manager User Guide

Important Notice

Version: 4.5

Date: March 25, 2013

Contents

About this Guide

This User Guide is for Apache Hadoop developers and system administrators who want to automate CDH installation on a cluster using Cloudera Manager. Cloudera Manager includes an optional wizard to guide you through the easy steps of installing and configuring CDH, Cloudera Manager Server and Agents, Oracle Java Development Kit (JDK), and an embedded PostgreSQL database. The wizard then starts the Hadoop services on the cluster hosts. After completing the steps in the wizard, you can run MapReduce jobs, and use HDFS, Hue, HBase, ZooKeeper, or Oozie. You can also use the included Cloudera Manager Admin Console to change CDH configuration settings, add new services, monitor cluster usage, activities, and health, and view status and logs to troubleshoot problems.

The information in this guide is also available in the online Help included with Cloudera Manager.

Other Cloudera Manager Guides

Guide
Cloudera Manager 4.5.x Release Notes
Cloudera Manager Installation Guide
Configuring Hadoop Security with Cloudera Manager
Configuring TLS Security for Cloudera Manager
Configuring Ports for Cloudera Manager

Introducing Cloudera Manager

Deployment and ongoing administration of a Hadoop stack can be difficult and time consuming. Deciding which components and versions to deploy based on use cases; assigning roles for nodes; effectively configuring, starting and managing services across the cluster; and performing diagnostics to optimize cluster performance requires significant expertise.

Cloudera Manager is the industry's first end-to-end management application for Apache Hadoop. By delivering granular visibility into and control over the every part of the Hadoop cluster, Cloudera Manager empowers enterprise operators to improve cluster performance, enhance quality of service, increase compliance and reduce administrative costs.

Introducing Cloudera Manager

Cloudera Manager provides many useful features for monitoring the health and performance of the components of your cluster (hosts, service daemons) as well as the performance and resource demands of the user jobs running on your cluster.

With Cloudera Manager, you can easily deploy and centrally operate a complete Hadoop stack. The application automates the installation process, reducing deployment time from weeks to minutes; gives you a cluster-wide, real time view of the services running and the status of their hosts; provides a single, central place to enact configuration changes across your cluster; and incorporates a full range of reporting and diagnostic tools to help you optimize cluster performance and utilization.

Cloudera Manager provides full lifecycle management for Apache Hadoop.

- Installs the complete Hadoop stack in minutes via a wizard-based interface.

- Lets you install multiple clusters, with the choice of running CDH3 or CDH4 on a given cluster.

- Gives you complete, end-to-end visibility and control over your Hadoop clusters from a single interface.

- Correlates jobs, activities, logs, system changes, configuration changes, service and host metrics along a single timeline to simplify diagnosis.

- Lets you set server roles, configure services and manage security across the cluster.

- Lets you gracefully start, stop and restart services as needed.

- Maintains a complete record of configuration changes with the ability to roll back to previous states.

- Automatically deploys client configuration files for the services you have installed.

- Supports HDFS High Availability using either Quorum-based storage (introduced with CDH 4.1) for its shared directory, or an NFS-mounted shared edits directory.

- Monitors dozens of service performance metrics and alerts you when you approach critical thresholds.

- Lets you gather, view and search Hadoop logs collected from across the cluster.

- Creates and aggregates relevant Hadoop events pertaining to system health, log messages, user services and activities and makes them available for alerting (by email) and searching.

- Consolidates cluster activity (user jobs) into a single, real-time view.

- Lets you drill down into individual workflows and jobs at the task attempt level to diagnose performance issues.

- Shows information pertaining to hosts in your cluster including status, resident memory, virtual memory and roles.

- Monitors the available space in log and other directories used by Cloudera Manager and CDH components.

- Provides operational reports on current and historical disk usage by user, group, and directory, as well as MapReduce activity on the cluster by job or user.

- Takes a snapshot of the cluster state and automatically sends it to Cloudera support to assist with resolution.

You work primarily in the Cloudera Manager Admin Console in a web browser that is connected to the Cloudera Manager Server, where you can manage the configuration settings, monitor the health of your services, and monitor and track user activity on your cluster.

Cloudera Manager Architecture

The Cloudera Manager Server performs the following functions:

- Tracks the Cloudera Manager data model, which is stored in the Cloudera Manager Server database. The data model is a catalog of the available host machines in the cluster, and the services, roles, and configurations assigned to each host.

- Communicates with Agents to send configuration instructions and track Agents' heartbeats

- Performs command execution to do tasks

- Provides an Admin console for the operator to perform management and configuration tasks

- Creates, reads, validates, updates, and deletes configuration settings

- Calculates and displays the health of the cluster and its components

- Tracks host metrics such as disk usage, CPU, and RAM

- Provides a comprehensive set of APIs for the various features supported in Cloudera Manager

- Manages Kerberos credentials

- Monitors the health of Hadoop daemons, and dozens of service performance metrics, and alerts you when you approach critical thresholds.

- Keeps a history of activity monitoring data and configuration changes

Each Agent starts and stops Hadoop daemons on the local host machine and collects statistics (overall and per-process memory usage and CPU usage, log tailing) for health calculations and status in the Admin console.
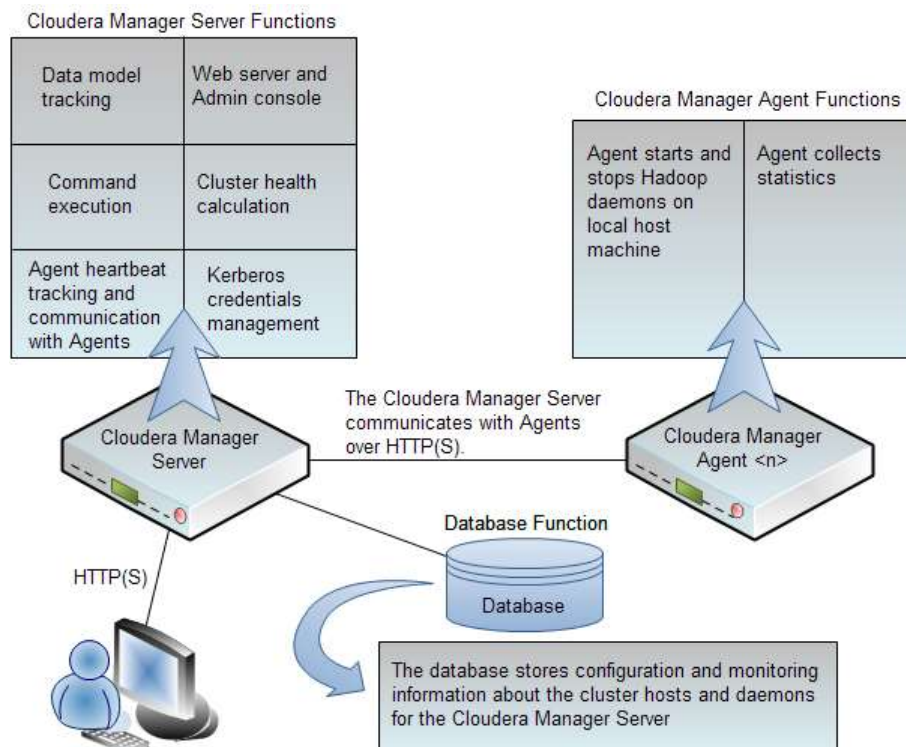
> **Note**
>
> The Cloudera Manager Agent runs as root so that it can make sure the required directories are created and that processes and files are owned by the appropriate user (for example, the `hdfs` user and `mapred` user).

What You Can Use Cloudera Manager to Do

Using Cloudera Manager, you can manage, configure and supervise Hadoop daemons on a set of host machines:

The first time you start the Cloudera Manager Admin Console, you can use the Cloudera Manager wizard to:

- Install CDH and the Oracle JDK on cluster hosts.
- Optionally install Cloudera Impala (if installing on RHEL/Centos 6).
- Configure and start services.

After First Run, you can use the Cloudera Manager Admin Console to:

- Configure CDH while seeing suggested ranges of values for parameters and illegal values highlighted; you can also configure override settings on specific hosts, and for specific role instances.

- Start and stop Hadoop daemons on hosts.

- Decommission individual roles, or all roles on a host to facilitate host maintenance.

- View the health of your system and its components.

- View the daemons that are currently running.

- Add and reconfigure services and role instances.

- Specify dependencies between services. Configuration changes for a service are propagated to its dependent service

- Generate CDH configurations for clients to use to connect to the cluster, and deploy those configurations automatically to clients.

- Manage rack locality configuration.

- With CDH4, configure HDFS High Availability or NameNode Federation.

- Download, distribute and activate a new CDH version (CDH 4.1.2 or later) all from within Cloudera Manager.

- Use the Cloudera Manager API to export or import deployment settings to and from clusters.

- Manage multiple clusters, which can be either CDH3 or CDH4 clusters.

- Display metrics about your jobs, such as the number of currently running tasks and their CPU and memory usage.

- Display metrics about your Hadoop services, such as the average HDFS I/O latency and the number of jobs running concurrently.

- Display metrics about your cluster, such as the average CPU load across all your machines.

- Get assistance with configuring Kerberos security (Cloudera Manager generates and installs the host and service key tab files for you.)

- Temporarily suppress alerting for individual roles, services, hosts, or even the entire cluster to allow maintenance/troubleshooting without generating excessive alert traffic.

Cloudera Manager also collapses several levels of CDH configuration abstraction into one. For example, you can manage Java heap usage in the same place as Hadoop-specific parameters. Cloudera Manager is internally secure, and you can configure the Admin Console and Agents to connect with the Server over TLS.

Overview of Using Cloudera Manager for Configuring Services

During Cloudera Manager installation, the first run of the Cloudera Manager Wizard will add and configure the Hadoop services you want to run on the hosts in your cluster. By default, Cloudera Manager will determine what it considers the optimal configuration of role instances on the available hosts, though you can can make these determinations yourself if you want. After the first run of the wizard, you can use Cloudera Manager to reconfigure the existing services, and add and configure additional hosts and services.

When Cloudera Manager configures a service, it configures host machines in your cluster with one or more functions (called roles in Cloudera Manager) that are required for that service. The role determines which Hadoop daemons run on a given host. For example, when Cloudera Manager configures an HDFS service instance it configures one host to run as a NameNode, another host to run as a Secondary NameNode, another host to run as a Balancer, and the remaining hosts as DataNodes. Services are named by default based on their function (e.g. the HDFS service may be named `hdfs1`) though you can optionally provide your own display names.

The associated role instances are named based on a combination of the role name and the host on which that role runs. So, if you have a set of hosts (myhost1, myhost2,... etc.) the roles associated with the HDFS service would be `namenode (myhost1)`, `secondarynamenode (myhost2)`, `datanode (myhost3)`, `datanode (myhost4)`, and so on, which all run under the `hdfs1` service instance on those same hosts.

Whenever you add and configure a service, you are creating an instance of that service on your cluster; that is, you can uniquely configure and run multiple instances of some services and roles. For most purposes, configuring and running one instance of the services may be sufficient.

Overview of Using Cloudera Manager for Monitoring Services and User Activities

Once you have installed and configured your Hadoop cluster, Cloudera Manager provides many useful features for monitoring the health and performance of the components of your cluster (hosts, service daemons) as well as the performance and resource demands of the user jobs running on your cluster. Not only does Cloudera Monitor let you monitor these events at will, but it can also notify you asynchronously when an important event of interest occurs.

Monitoring Hadoop Services
For HDFS, HBase, MapReduce, ZooKeeper, Cloudera Impala, and Flume, host agents harvest Health metrics that the Service Monitor checks constantly. You can view the results of these health checks at both the service and role instance level. Information is provided as specific metrics as well as in charts that help with problem diagnosis. Health check descriptions typically 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 your Cluster
Cloudera Manager's Host Monitoring lets you 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 User Activities

Activity Monitoring lets you see who's running what activities on the cluster, both at the current time and through views of 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.

Searching Logs and Events

Cloudera Manager provides access to logs and events 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 activity, you can easily view the relevant log or event and alert entries that occurred on the hosts used by the job while the job was running.

You can also search independently for log entries or events and alerts by time range, service, host, keyword.
The Event Server aggregates relevant Hadoop events and makes them available for alerting and for searching, giving you a view into the history of all relevant events that occur cluster-wide.

Receiving Alert Notifications

You can configure Cloudera Manager to generate Alerts from a variety of events. You can configure thresholds for certain types of events, enable and disable them, and configure email delivery of Alerts on 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.

Operational Reports

Reports provide an historical view into disk utilization by user, user group, and by directory. You can manage your HDFS directories as well, including searching and setting quotas. You can also view cluster job activity user, group, or job ID. These reports are aggregated over selected time periods (Hourly, Daily, Weekly etc.) and can be exported as XLS or CSV files.

Basics of Using Cloudera Manager

This section introduces some of the concepts and features that you will encounter throughout the Cloudera Manager Admin Console, as well as some basic information about how to access the Cloudera Manager user interface.

The following topics describe some of the basic features of the Cloudera Manager Admin Console.

- [Starting the Cloudera Manager Admin Console](#)
- [About the Cloudera Manager Admin Console](#)
- [Selecting a Time Range](#)
- [About Events and Alerts](#)
- [About Service, Role, and Host Health](#)
- [Cloudera Manager User Accounts](#)

Starting the Cloudera Manager Admin Console

The Cloudera Manager Admin console enables you to use Cloudera Manager to configure and monitor your cluster. In this release, the Cloudera Manager Admin console supports the following web browsers:

- Internet Explorer 9
- Google Chrome
- Safari 5
- Firefox 3.6 and later

To start the Cloudera Manager Admin Console:

1. In a web browser, type the following URL:

   ```
   http(s)://<Server host>:<port>
   ```

   where:

   `<Server host>` is the name or IP address of the host machine where the Cloudera Manager Server is installed.

   `<port>` is the port configured for the Cloudera Manager Server. The default port is 7180.

2. Log into the Cloudera Manager Admin Console. The `admin` user credentials are:

   Username: admin
   Password: admin

> **Note**
>
> For security, change the password for the default `admin` user account as soon as possible. You can also add user accounts and selectively assign admin privileges to them as necessary. For instructions, see the [Changing the Password for an Account](#).

About the Cloudera Manager Admin Console

When you log into the Cloudera Manager Admin console, if services have been configured, you land on the **Services** page. The **Services** page is the central point from which you can view the status of your cluster, and it is the launching point for many of the tasks you can perform through Cloudera Manager.

(If there are no services, logging in to the Admin Console will take you to the Cloudera Manager wizard where you can discover cluster hosts, install CDH packages, and map services to hosts, and configure and start Hadoop and Cloudera Manager services.)

Cloudera Manager displays time-stamped data (for example, in the Commands tab for a service) using the time zone of the host where Cloudera Manager server is running.

The top navigation bar provides access to the many functions provided by Cloudera Manager:

Services: The Services Monitoring and Services Configuration sections of this Guide describes the features under the **Services** tab. Under this tab you can:

- View the status and other details of a Service instance, or the Role instances associated with the service

- Make configuration changes to a service instance, a role, or a specific role instance

- Add a new service or role

- Stop, Start, or Restart a service or role.

- View the commands that have been run for a service or a role

- View an audit history for a service or role instance

- Deploy client Configurations

- Decommission a role

You can also perform actions unique to a specific type of service, for example:

- Enable HDFS High Availability or NameNode Federation

- Run the Balancer (HDFS)

Hosts: The Host Management section describes the features under the **Hosts** tab. Under this tab you can:

- View the status and a variety of detail metrics about individual hosts

- View all the processes running on a host, with status, access to logs and so on

- Run the Host Inspector

- Add and delete hosts

- Create and manage host templates

- Manage (download, distribute and activate) parcels.

- Decommission and Recommission hosts

- Make rack assignments

- Run the host upgrade wizard

- Make configuration changes for host monitoring

Activities: The Activities tab lets you monitor and manage MapReduce jobs running on your clusters. The Activity Monitoring section of this Guide describes these features in detail.

Logs: The **Logs** page presents log information for Hadoop services, and lets you search by service, role, host, and/or search phrase as well log level (severity).

Events: The **Events** page lets you to search for and display events and alerts about that have occurred within a time range you select, anywhere in your cluster. See Searching for Events and Alerts for more information. See Charting Time-series Data for more information.

Charts: The Charts page lets you search for metrics of interest and display them as charts. You can also create custom chart views that can act as a personalized dashboard for your cluster. See Charting Time-series Data for details.

Reports: The **Reports** tab lets you create reports about the usage of HDFS in your cluster, as well as browse files and manage quotas for HDFS directories. See Viewing Reports for details.

Search Box

The Search box on the top navigation bar lets you search by Service, Role or Host name. You can enter a partial name and it will search for all entities that match.

Running Commands Indicator

The indicator ( ) to the left of the **Support** menu shows you how many commands are currently running in the clusters you are managing. You can click on it to see a list of running commands in your cluster. See Viewing Running and Recent Commands for more information.

Configuration Validations Indicator

The Configuration Validations indicator ( ) to the left of the Running Commands indicator shows you how many actionable validation notifications (Validation Errors and Warnings) are pending for your cluster. The color of the badge with the number in it indicates the severity of the notifications – Red indicates a Validation Error, Orange indicates a Validation Warning, and Blue is informational. Click on the indicator to pop up a dialog box where you can filter for and display these notifications. See Configuration Validation Notifications for more information.

New Parcel Indicator

The New Parcel indicator ( ) to the left of the Configuration Validations indicator shows you whether a parcel for a newer version of your software is available for the clusters you are managing. Click on the indicator to see a list of parcels available for your cluster. See Managing Parcels for more information.

Support Menu

Under the **Support** menu you will find the command for sending diagnostic data to Cloudera Support if you need help solving problems when using Cloudera Manager on your cluster. See Sending Diagnostic Data to Cloudera for instructions.

The **Support Portal** link connects you to the Cloudera Support portal.

Help Menu

The **Help** menu provides access to the *Cloudera Manager User Guide* (via the **Help** option), the *Cloudera Manager Installation Guide*, and the Cloudera Manager API documentation. By default, these links open locally installed Help files, and do not require internet access. You can configure Cloudera Manager to open the latest Help files from the Cloudera web site — see Administration for more information.

The **About** link under the **Help** menu shows you the version number and build details for the version of Cloudera Manager you are running. It also shows the current date and time stamp of your Cloudera Manager server.

Logged-in User Menu

The link to the right of the **Help** menu shows the currently logged-in user. For example, if you are logged in as *admin* this menu will be named **admin** (admin). Pull down this menu to find the **Logout** command.

This is also where you can change your own password — the **Change Password** option opens the **Account Settings** page where you can change your password.

Administration

Click the gear icon to display the **Administration** page. For details of the functions available from the page, see Administration.

Selecting a Time Range

The **Time Range Selector** appears when you are viewing the Activities, Logs, and Events tabs, and when you are viewing the Status, Commands and Audits pages of individual Services, Roles, and Hosts. The Time Range Selector lets you highlight a range to time over which to view historical data. The top level Services (All Services) and Hosts tabs do not show the Time Range Selector; those pages always show status and health for a specific point in time.

Cloudera Manager displays time-stamped data using the time zone of the host where Cloudera Manager server is running. The time zone information can be found from within the application, under the **Help** -> **About** menu.

In the pages that support a time range selection, the highlighted (blue) area of the graph shows the selected time range. There are a variety of ways to change the time range in this mode.

The Reports tab does not support the Time Range Selector: the historical reports available under the Reports tab have their own time range selection mechanism.

The background chart in the Time Range Selector bar shows the percentage of CPU utilization on all machines 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. Note that the background chart appears even when the Time Range Selector handles are not available.

Current vs. Historical data

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.

Specifying a Point in Time for "Snapshot" Data

Health and Status information on pages such as the All Services page and Service status pages, reflects the state at a single point in time (a snapshot of the health and status). By default, this is the status and health at the current time. However, by moving the Time Marker (◆) to an earlier point on the time range graph, you see the status as it was at the selected point in the past.

- When the Time Marker is set to the current time, it is blue (◆). When it is set to a time in the past, it is orange (◆).
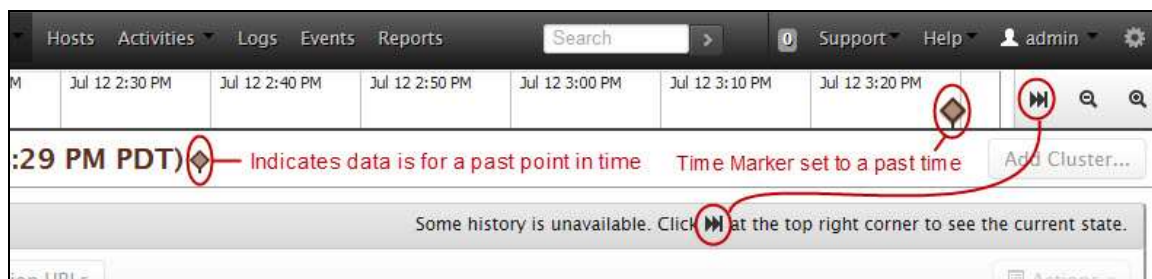
- When the Time Marker is set to a past time, you can quickly switch back to view the current status using the Current Time button (⏭).

- When displayed data is from a single point in time (a snapshot) the panel or column will display a small version of the Time Market icon (◆) in the panel. This indicates that the data corresponds to the time at the location of the Time Marker on the Time Range Selector. The Status and Health panels in the Service status pages are examples of this.

> **Note:** 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.

Specifying a Time Interval for Historical Data

Pages such as the Logs, Events, and Activities show data over a time interval 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.

- You can drag one (or both) edges of the highlighted area of the graph to expand or contract the range for which data will be displayed. (See the handles noted in the illustration below.)

- You can also grab and slide the highlighted area as a unit.

- You can zoom the time line out or in to change the time scale and make it easier to drag the time slider to an earlier or later time.

- Click the calendar icon to open the time selection widget:

  - To look backwards from the present time, you can use preselected time chunks (the past 30 minutes, the past hour, and several other intervals up to the past day). Select among the options provided in the time selection widget.

  - To enter a specific start time and stop time, enter them into the fields provided.

- When displayed data covers time interval rather than a single point in time, an icon representative of the Time Range Selector (🖼) appears in the header of the panel. This indicates that the displayed data corresponds to the time range currently selected and highlighted on the Time Range Selector. The Charts panel and the Logs and Events tabs shown on the individual Service Status pages are examples of this.

- When you are under the **Activities** 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.

> **Note:** When you are looking historical information, some functions will not be available. For example, 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.

Zooming the Time Range Selector In or Out

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

- **Zoom Out** lets you show a longer time period on the time range graph (with correspondingly less granular segmentation).

- **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 Range Selector can make it easier to use the selector to highlight a time range.

Selecting a Custom Time Range

To manually select a time range:

1. Click the small calendar icon (▦) to display the **Custom Time Range** panel.

2. Enter either a start and end time, and Click OK to put your choice into effect,
   — or —
   Select a time interval relative to the current time — such as the past 30 minutes, the past 6 hours, and so on — using the buttons provided.

About Events and Alerts

Events are 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.

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.

Alerts are events that are considered especially noteworthy and are triggered by selected events. You can configure which events trigger alerts, and you can configure the Alert Publisher to send alert notifications by email.

Events

Cloudera Manager supports four types of events managed through the Event Server:

These events are available for alerting and searching.

Thresholds for various health checks can be set under the **Configuration** tabs for HDFS, HBase, and MapReduce service instances, at both the service and role level. See Configuring Monitoring Settings for more information.

Alerts

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 services that are a part of the Cloudera Management Services.

Alerts are shown in red when they appear in a list of events.

The settings for enabling or disabling specific alerts are found under the Configuration tab for the services to which they pertain. Email configuration, to enable the receipt of alerts by email, is done

under the Configuration tab for the Alert Publisher role, found under the Cloudera Management Services service. See Configuring Monitoring Settings for more information on setting up alerting.

About Service, Role, and Host Health

Cloudera Manager monitors the state of the services and roles that are running on your cluster, and the hosts in your cluster. You can see the summary results of these under the **Services** tab, where various health results determine an overall health assessment of the service or role. The **Hosts** tab shows similar summary result for the hosts.

Cloudera Manager also monitors a number of metrics for HDFS, MapReduce, HBase, ZooKeeper, Flume and Impala service and role instances. These are reflected in the results shown in the **Health Tests** panel under the **Status** tab when you have selected an HDFS, MapReduce, HBase, ZooKeeper, Flume or Impala service or role instance to view.

The overall health of a role or service is a roll-up of its health checks; if any health check is **Bad**, the service's or role's health will be **Bad**. If any health check is **Concerning** (but none are **Bad**) the role's or service's health will be concerning.

The health check results are presented in the **Health Tests** panel. For some of these, you can also chart the associated metrics over time. Other metrics are also shown as charts over a time range. See Viewing Service Status and Viewing Status for a Role Instance for more details. See Viewing Detailed Information about Hosts for details of the health of a host. The **Health** status can be:

✓ Good	For a specific health check, the returned result is normal or within the acceptable range. For a role or service, this means all health checks for that role or service are **Good**.
⚠ Concerning	For a specific health check, 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 check is **Concerning**.
✗ Bad	For a specific health check, the check 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 check is **Bad**.
? Unknown	For a role or service, its health 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.
ⓘ Checks disabled	Health Checks have been disabled in the configuration for this service or role.

⊘ History unavailable	Cloudera Manager does not support the collection of historical information for this role or service. This is the case for services such as ZooKeeper, Oozie, or Hue — services other than HDFS, MapReduce and HBase.

There are several types of health checks that can be performed, depending on the type of service or role instance:

- Simple pass/fail checks, such as a service or role started as expected, a DataNode is connected to its NameNode, or a TaskTracker is (or is not) blacklisted. These checks result in the health of that metric being either Good or Bad.

- Metric-type tests, such as 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.

  The results of these types of checks can be compared to threshold values that determine whether everything is OK (e.g. plenty of disk space available), whether it is "Concerning" (disk space getting low), or is Bad (a critically low amount of disk space).

- HDFS (NameNode) and HBase also run a health test known as the "canary" test; it periodically does a set of simple create, write, read, and delete operations to determine the service is indeed functioning.

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

Cloudera Manager User Accounts

Cloudera Manager user accounts allow users to log into the Cloudera Manager Admin Console. User authentication can be done through a local database, through an external LDAP directory server (Active Directory or OpenLDAP-compatible), or through an external authentication program of your own choosing.

Cloudera Manager users are managed through the **Users** tab of the Administration page (accessed with the gear icon).

User accounts added from an LDAP directory or other external authentication mechanism will have **External** in the **User Type** column shown under the Users tab. Users in the local database will have **Cloudera Manager** as the user type. See Configuring External Authentication for information on configuring Cloudera Manager to use an external LDAP directory or other authentication program for user authentication.

User accounts can optionally have Administrator privileges:

- Administrator privileges: Allows the user to add, change, delete, and configure services or administer user accounts. Also, even if you are using an external authentication mechanism for user authentication, users with Administrator privileges will also be able to log in to Cloudera Manager using their local Cloudera Manager username and password. (This prevents the system from locking everyone out if the external authentication settings get misconfigured.)

- No Administrator privileges: User accounts that don't have Administrator privileges can view services and monitoring information but they cannot add services or take any actions that affect the state of the cluster.

When you are logged in to the Cloudera Manager Admin Console, the user name you are logged in as is shown on the top navigation bar — for example, if you are logged in as *admin* you will see this: `admin`.

Changing Your Password

> **Important**
>
> As soon as possible after running the installation wizard and beginning to use Cloudera Manager, you should use the following procedure to change the password for the default `admin` account, if you have not already done so.

To change the password for the logged-in account:

1. Logged in as `admin` pull down user menu and select the **Change Password** option.
2. Enter a new password twice and then click **Submit**.

Adding Cloudera Manager User Accounts

To add a Cloudera Manager user account:

1. Click the gear icon to display the **Administration** page.
2. Click the **Users** tab.
3. Click the **Add User** button.
4. Enter a username and password.
5. To grant Administrator privileges to the user account, select **Add Administrator Privileges**.
6. Click **Submit**.

Users accounts created in this way will show **Cloudera Manager** in the User Type column.

Changing the Privileges and Password for an Account

To change the privileges for an account:

1. Click the checkbox to the left to select the user account.

2. Click the **Add Administrator Privileges** or **Remove Administrator Privileges** button.

To change an account password:

1. Click the **Change Password** button.

2. Type the new password and repeat it to confirm.

3. Click the **Submit** button to make the change.

Deleting an Account

To delete an account:

1. Select the user account.

2. Click the **Delete** button.
   (Note that there is no confirmation of the action.)

Services Monitoring

Cloudera Manager's Service Monitoring feature monitors dozens of service health and performance metrics about the services and role instances running on your cluster. It 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, and maintains a complete record of service-related actions and configuration changes.

> **Note**
>
> Impala and HBase monitoring are a separately-licensed features. To determine what license capabilities you have, go to the License tab on the Administration page (see Administering Licenses).

The following topics describe how to monitor the services and role instances installed on your cluster.

- Monitoring the Health and Status of Services

- Viewing Service Status

  - Flume Metric Details

- Configuration Validation Notifications

- [Viewing Service Instance Details](#)

- [Viewing Status for a Role Instance](#)

- [Managing and Monitoring Federated HDFS](#)

- [Viewing Running and Recent Commands](#)

- [Viewing the Audit History](#)

- [Viewing Charts for Service, Role, or Host Instances](#)

- [Viewing the Processes for a Role Instance](#)

- [Viewing Heatmaps for Services and Roles](#)

- [Configuring Monitoring Settings](#)

Monitoring the Health and Status of Services

From the **Services** page, you can:

- Monitor the health and 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. (Note that these configuration files are normally deployed automatically when you install your cluster or add a service).

- View the Maintenance Mode status of your cluster.

- Install an additional cluster. After initial installation, you can use the **Add Cluster** wizard to add and configure an additional cluster. See [Managing Multiple Clusters](#) and [Adding a Cluster](#) for more information on this topic.

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 tabs.

Service Health and Status

To view the status of your services, click the **Services** tab and select **All Services**.
The Services page opens and displays an overview of the service instances currently installed on your cluster.

Services Monitoring

For each service instance, this page shows:

- The type of service

- The service status (for example, *Started*)

- The overall health of the service

- The type and number of the roles that have been configured for that service instance.

> **Note:** By default, the All Services page shows the **current** state of the services in your cluster. By moving the Time Marker (⬥), you can see what the status was at any point in the past. When you are looking at the past, the **Actions** menus and most other commands are disabled, and Role Counts information may not be accurate. Click the Current Time button (⏭) to return to the current time.
>
> See [Selecting the Time Range](#) for details of how time range selection works in Cloudera Manager.

Add a Service

After initial installation, you can use the **Add a Service** wizard to add and configure (but not start) new service instances. The **Add a Service...** command is found under the cluster **Actions** menu for the cluster where you want to add the service.

The cluster **Actions** menu, and thus the **Add a Service...** command, is not available if you are viewing status for a point of time in the past.

See [Adding Services](#) for more information on this topic.

View the URLs of the Client Configuration Files

To allow Hadoop client users to work with the HDFS, MapReduce, YARN and HBase 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 download and distribute these client configuration files manually to the users of a service, if necessary.

The **Client Configuration URLs** command on the cluster **Actions** menu 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 **Client Configuration URLs** button is not available if you are viewing status for a point of time in the past.

See [Deploying Client Configuration Files](#) for more information on this topic.

View the Health and Status of a Service Instance or Role Instance

To see the status of a service instance:

- Click the link in the **Name** column,   *OR*

- Click the health status associated with the instance,   *OR*

- From the **Services** tab, select the service instance you want to see.

This will open the **Status** page where you can view a variety of information about a service and its performance. See Viewing Service Status for details. **To see the 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 Status for a Role Instance for details.

If you are viewing a past point in time, 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

- Click the **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 placed into maintenance mode explicitly. (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 **View Maintenance Mode Status** button is not available if you are viewing status for a point of time in the past.

The Actions Menus

There are two **Actions** menu available on the **All Services** page: one for the cluster, and one for each service.

Actions for a Cluster

There are multiple actions you can take at a cluster level:

- Stop, Start, or Restart all the services in the cluster

- Deploy the client configurations onto the appropriate nodes of the cluster, or view the client configuration file URLs.

- Upgrade the cluster

- Rename the cluster

- Enter or exit Maintenance Mode for the cluster.

Actions for a Service

There is an **Actions** menu associated with each service instance installed on your cluster.
From the **Actions** menu for a service you can:

- Stop, start, restart, or delete the service (the available actions depend on the current status of the associated service – for example, you cannot Start a Started service).

- Change the display name of a service

- Enter or exit Maintenance Mode for the service.

These actions are covered in the Services Configuration section of this document:

- Starting, Stopping, and Restarting Services

- Deleting Service Instances and Role Instances

- Renaming a Service to change its display name

This action is covered in the Maintenance section of this document:

- Entering and exiting Maintenance Mode

Viewing Service Status

To see the status of a Service instance:

- Pull down the menu from the **Services** tab and select the service instance you want to see.   *OR*

- Click the **Services** tab and select **All Services**.

  o   Click the link in the **Name** column,   *OR*

  o   Click either the Status or Health value associated with the instance.

For all service types there is a **Status and Health 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. Hue, Oozie, Hive, and YARN (CDH4 only) only provide the basic Status and Health Summary.
>
> Impala also provides only basic status and health information if you do not have a license for Impala monitoring.

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 checks. for more information see Configuring Monitoring Settings.

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

Impala also provides this information if you have license installed that enables Impala Monitoring.

The Actions Menu

The Actions menu is available from the Service Status page when you are viewing Current time status. The commands function at the Service level – for example, **Restart** selected from this page will restart all the roles within this service.

Some services provide additional commands that are unique to that service, such as HDFS.

> **Note:** The Actions menu is only available when you are viewing **Current** 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 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 Current Time button ( ) is present whenever you are viewing past status, to enable 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 Selecting a Time Range for more details.

Status and Health Summary

The Status and Health Summary shows the Status and Health 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 and Health Summary.

The **Status** can be:

	Started	The service or role is running normally.
	Started	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.
	Starting	The service or role is starting up but is not yet running.
	Starting	For a service, this indicates the service is starting up, but at least one of its roles has 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.
	Stopping	The service or role is stopping: a stop command has been issued, but the service (and its roles) have not finished shutting down.
	Stopped	The role or service is not running.
	History Unavailable	Cloudera Manager does not support historical information for this role or service. This is the case for services such as ZooKeeper, Oozie, or Hue — services other than HDFS, MapReduce and HBase.
—	N/A	The service or role is not started or stopped in the same way as a regular service or role. 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.
	Unknown	Status of the service or role is not known.

The overall Health Status for a service is a roll-up of the health check results for the service and all its role instances.

The **Health** status can be:

✓ Good	For a specific health check, the returned result is normal or within the acceptable range. For a role or service, this means all health checks for that role or service are **Good**.
⚠ Concerning	For a specific health check, 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 check is **Concerning**.
⊗ Bad	For a specific health check, the check 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 check is **Bad**.
? Unknown	For a role or service, its health 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.
⏸ Checks disabled	Health Checks have been disabled in the configuration for this service or role.
⊘ History unavailable	Cloudera Manager does not support the collection of historical information for this role or service. This is the case for services such as ZooKeeper, Oozie, or Hue — services other than HDFS, MapReduce and HBase.

You can click either the **Status** or **Health** link for a role to drill down to see the details of the status and health of the role instance(s). If there is a single instance of the role type, the link takes you directly to the Role Instance page.

If there are multiple role instances (such as for DataNodes, TaskTrackers, RegionServers) a pop-up opens to allow you to select the specific instances you want to see. Furthermore, this pop-up displays the results for each health check that applies to this role type.

You can filter by an individual health check result. Click the result link — an **X** 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. (Note that in the example above, although the filter was to look at an "Unknown" result, the Health status of the instance is "Bad". This is because there is at least one "Bad" health check associated with that instance.

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 — for example:

- The HDFS Summary shows read and write latency statistics and disk space usage.

- The MapReduce Summary shows statistics on slot usage, jobs and so on.

- The HBase Summary shows statistics about get and put operations and other similar metrics.

- The Flume summary provides a link to a page of Flume metric details. See Flume Metric Details.

- The ZooKeeper Summary provides links to the ZooKeeper role instances (nodes) as will as Zxid information if you have a ZooKeeper Quorum (multiple ZooKeeper servers).

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

Move your cursor over an individual metric to pop up a definition.

Health Tests

The Health Tests panel appears 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/collapsable 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 for more information.

- The small heatmap icon (▦) to the right of some of the tests takes you to a heatmap display that lets you compare the values of the relevant test result metrics across the nodes of your cluster.

Charts

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

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

Flume Metric Details

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

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

Note that the Flume configuration can be viewed under the Configuration tab, Agent category.

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.

Note that 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). The **Actions** menu is also available from this page.

Configuration Validation Notifications

The Configuration Validations pop-up shows the validation warnings and errors that are pending for your cluster. It is located in the main navigation bar between the New Parcels indicator and the Running Commands indicator.

The indicator shows how many validation errors or warnings are currently pending. The indicator badge show the number of notifications at the highest severity level – for example, if there are configuration errors, the indicator will show a Red badge with the number of error notifications. If there are no errors but configuration warnings exist, then the badge will be orange and show the number of warnings. No badge will be shown if there are no notifications.

- Click the Configuration Validations indicator to display the Configuration Validations pop-up.

In the pop-up, the notifications at the highest severity level are shown, grouped by Service Name.

- To display other validation notifications click the button ( ) with the icon that represents that severity level (error, warning, or informational notifications). You can select the buttons so that all three types of notifications are shown together.

- Click the message associated with a warning or error to be taken to the configuration property for which the validation notification has been issued.

- You can group the notifications in a number of ways, such as Role Group, Entity type, host name, and so on — group by Service Name is the default.   Pull down the arrow int he Group By: field to select the group category.

- Click Close to close the pop-up.

Viewing Service Instance Details

For a selected service, the **Instances** tab shows all the role instances that have been instantiated for this service.

To view the details of a service instance:

1. Click the **Services** tab on the top navigation bar and select **All Services**.

2. Click the service Name (or its Health Status) to go to the **Status** tab for that service.

3. Click the **Instances** tab on the Services 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:

- Each role instance by name.
  Click the role name to <u>view the Role Status</u> for that role.

- The host on which it is running.
  Click the Host name to <u>view the Host Status Details</u> for the host.

- The rack assignment

- The role instance's status

- The role instance's health.

- 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.

To sort the Instances list:

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.

To filter the Instances list:

- To filter by Role, Status, Health, Decommissioned, or Maintenance Mode, select the value from the drop-down search field at the top of the column.

- To filter by Host or Rack, type the filter value in the search field.

From the **Actions for Selected** menu you can stop, start, restart, or delete a role, put a role into or remove it from maintenance mode, and (for HDFS or HBase roles only) decommission or recommission a role.

To take an action on one or more roles:

1. Check the checkbox next to the role instance(s) you want to act upon (or check the box to the top of the list to select all role instances).

2. From the **Actions for Selected** menu, select the appropriate action. See Services Configuration for details on these actions.
   (Note that the Decommission action only applies to HDFS DataNodes, MapReduce TaskTrackers, YARN NodeManagers, and HBase RegionServers.)

To add a role instance:

- Click the **Add** button.
  This takes you to the **Add Role Instances** page. See Adding Role Instances for further information.

Viewing Status for a Role Instance

To view status for a role instance:

- Select a Service instance to display the **Status** page for that service.

- Click the **Instances** tab.

- From the list of Roles, select one to display that role instance's **Status** page.

The Actions Menu

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 WebUI 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 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 Current Time button (⏮) is present whenever you are viewing past status, to enable 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 Selecting a Time Range for more details.

Role Summary

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

- Click the service name to return to the **Service Status** page.

- Click the Host name to view the **Host Status Details** page for that host.

All role types provide **Role Summary** and **Processes** panels.
Some role instances related to HDFS, MapReduce, and HBase also provide a Health Tests panel and associated charts. The **Status** can be:

✔	Started	The service or role is running normally.
⬛	Started	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.
⏳	Starting	The service or role is starting up but is not yet running.
⬛	Starting	For a service, this indicates the service is starting up, but at least one of its roles has 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.
⏳	Stopping	The service or role is stopping: a stop command has been issued, but the service (and its roles) have not finished shutting down.
⬤	Stopped	The role or service is not running.
⊘	History Unavailable	Cloudera Manager does not support historical information for this role or service. This is the case for services such as ZooKeeper, Oozie, or Hue — services other than HDFS, MapReduce and HBase.

—	N/A	The service or role is not started or stopped in the same way as a regular service or role. 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.
❓	Unknown	Status of the service or role is not known.

The overall Health Status for a role is a roll-up of the health check results for that role. The **Health** status can be:

✔ Good	For a specific health check, the returned result is normal or within the acceptable range. For a role or service, this means all health checks for that role or service are **Good**.
⚠ Concerning	For a specific health check, 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 check is **Concerning**.
❌ Bad	For a specific health check, the check 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 check is **Bad**.
❓ Unknown	For a role or service, its health 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.
⏸ Checks disabled	Health Checks have been disabled in the configuration for this service or role.
⊘ History unavailable	Cloudera Manager does not support the collection of historical information for this role or service. This is the case for services such as ZooKeeper, Oozie, or Hue — services other than HDFS, MapReduce and HBase.

Health Tests

The Health Tests panel is 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/collapsable 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 for more information.

- The small heatmap icon (⊞) to the right of some of the tests takes you to a heatmap display that lets you compare the values of the relevant test result metrics across the nodes of your cluster.

Charts

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

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

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 Federated Nameservices or 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 Federation or High Availability configured, the Instances page has a section at the top that provides information about the configured Nameservices. This includes information about:

- The mounts points in the namespace

- 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).

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

- Validate the Shared Edits directory if Quorum-based Storage is not enabled.

From this page you can also add a NameService via the **Add Nameservice** button at the top of the page. See Adding a NameService.

Viewing Running and Recent Commands

Viewing Running Commands

In the Admin Console main navigation bar there is an indicator ( ) that shows the number of commands that are currently running in your cluster (if any). This indicator is positioned just to the left of the **Support** link at the right hand side of the navigation bar. Unlike the Commands tab for a role or service, this indicator includes all commands running for all services or roles in the cluster.

Click on this indicator to view a list of all commands currently running in your cluster. From this window you can click the **All Recent Commands** button to view all commands that have run and finished recently. This displays information on all running and recent commands in the same form as described below in the Running Commands and Recent Commands sections.

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 Recent and Running Commands for a Specific 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.

To view the commands that are running or have run for a Service or Role instance:

1. Click the **Services** tab on the top navigation bar.

2. Click the service Name to go to the Status tab for that service.

3. To view recent commands for a role, select the role instance name to go its Status tab.

4. Click the **Commands** tab on the Services navigation bar.

Running Commands

The Running Commands area shows commands that are currently in progress.

If a command is running, the **Command Details** section at the top shows:

- The command

- The Context or Parent command if relevant.

- The time that the command was started

- A progress indicator

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

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 information status 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 other **Recent Commands**.

Recent Commands

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

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 preset time ranges at the right side of the page, the **Time Range Selector**, or the **Custom Time Range** panel (see Selecting a Time Range). Each entry shows:

- The command (and how many subcommands, if any, it has)

- The time at which it was started

- Its current status (progress)

- The command result message.

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. Click a command in the **Recent Commands** 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; its start and end times, its progress (status), and a link to its parent command. The information includes:

    o The Context, which may be a cluster, service, a host, or a role

    o The time at which it was started

    o Its current status (progress)

    o The time the command completed.

You can use the **Parent** link near the top of the page to return to the parent command's details.

> **Note:** If the parent is **First Run**, this indicates that this command was run as part of the initial startup of your cluster. Clicking on this link takes you to the command history for the startup of your cluster.

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. This section includes:

    o The Context, which may be a cluster, service, a host, or a role

    o The time at which it was started

    o Its current status, including timestamp

    o A result message

- 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.

Viewing the Audit History

The Cloudera Manager **Audits** tab lets you search for and display audit events that have occurred within a time range you select anywhere in your cluster. You can use the Time Range Selector or the time range link
(30m 1h 2h 6h 12h 1d) to set the time range for your search. (See [Selecting a Time Range](#) for details). Note that 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.

You can reach the **Audits** tab from a service, role, or host instance.

Filtering Audit Events

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

Adding Filters

To add a filter to an audit event log, do one of the following:

- Click the **✚ Add Filter** to the left of the log.
  A filter control is added to the list of filters.

  a. Choose an audit event property in the property drop-down list. You can search by properties such as Username, Service, Command, or Role. The actual properties may vary depending on the service or role you are looking at.

  b. If the property allows it, choose an operator in the operator drop-down list.

  c. Type an audit property value in the value text field.
  Note that for some properties, where the list of values is finite and known, you can start typing and then select from a list of potential matches.
  For some properties you can include multiple values in the value field. For example, you can create a filter like "SERVICE = HBASE1, HDFS1". Multiple values for a single filter property are combined using OR (i.e. SERVICE = HBASE1 OR HDFS1)

  d. Click **Add Another** to add additional filter components. A filter containing the property and its value is added to the list of filters at the left. Multiple filters are combined using AND – e.g. SERVICE = HBASE1 AND USERNAME = admin)

  e. Click **Search**.
  The log displays all events that match the filter criteria.

Removing a Filter

To remove a filter from a filter specification:

1. Click the ✖ at the right of the filter.
   The filter is removed and the audit log redisplays all events that match the remaining filters.

Modifying a Filter

To modify a filter:

1. Click the filter.
   The filter expands into separate property, operator, and value fields.

2. Modify the value of one or more fields.

3. Click **Search**.
   A filter containing the property, operation, and value is added to the list of filters at the left and the audit log redisplays all events that match the modified set of filters.

The Audit Log Display

Audit log entries are ordered (within the time range you've selected) with the most recent at the top.

- Click on a link in an event entry to add that value as an additional filter.

The **Audits** tab lets you see the actions that have been taken for a Service or Role instance, and what user performed them. The audit history includes actions such as creating a role or service, making configuration revisions for a role or service, and running commands.

To view the Audit history for a Service:

1. Click the **Services** tab on the top navigation bar, then choose the service you want to see.

2. Click the **Audits** tab on the Services navigation bar.

To view the Audit history for a Role:

1. Click the **Services** tab on the top navigation bar, then choose the service you want to see.

2. Click the **Instances** tab on the Services navigation bar to show the list of role instances.

3. Select the Role whose audit history you want to see.

4. Click the **Audits** tab on the navigation bar for the role.

The Audit History provides the following information:

- **Context:** The service or role and host affected by the action.

- **Message:** What action was taken.

- **Date:** Date and time that the action was taken.

- **By User:** The user name of the user that performed the action.

The audit history does not track the progress or results of commands it sees (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. If you want to view the progress or results of a command, you can look at **Recent Commands** under the **Commands** tab.

If no actions were taken 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 **Time Range Selector** or the **Custom Time Range** panel (see Selecting a Time Range).

Viewing Charts for Service, Role, or Host Instances

For Service or Role instances, or for individual hosts, you can see charts of various metrics relevant to the entity you are viewing.

A limited set of charts appear under the **Status** tab for a Service, a Role instance, or a host.

There is also a **Charts** tab that appears for each of these entity types, and which displays 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.

While the actual metrics displayed are different on each of these pages, the basic functionality works in the same way.

Status Tab Charts for a Service, Role, or Host

A relatively small set of charts is shown by default on the Status page of a Service, Role, or Host.

You can toggle between a default view and a Custom view for these charts.

The Custom View

The Custom view is displayed by default when you view the Status tab for one of these entities. Initially it shows the same charts as the Default charts view.

In either view, when you move your mouse over a chart, its background turns yellow, indicating that you can act upon it.

- Moving the mouse to a data point on the chart shows the details about that data point in a pop-up tooltip.

- Click on a chart to expand it into a full-page view with a legend for the individual charted entities as well more fine-grained axes divisions.

    o If there are multiple elements in the chart, you can check/uncheck the legend item to hide or show that element on the chart.

    o Click the **Close** button to return to the regular chart view.

- When the mouse is over a chart, a down-arrow icon appears at the upper right. Click this to display a menu where you can choose to edit the individual chart, or to remove the chart from the Custom view.

- When the mouse is over a chart, a **Clone** link appears at the bottom right of the chart. Click the **Clone** link to duplicate the chart, make any modifications you want, and then save back to the same page or to a different page.

The Default View

The Default view is a predefined set of charts that you cannot change. You can expand a chart by clicking on it, and can Clone the chart, but you cannot otherwise edit it or remove it.

Editing a Chart

Editing a chart lets you edit a chart from the Custom view and save it back into the same view. You cannot edit charts in the Default View. Editing a chart only affect the copy of the chart in the current view – if you have copied the chart into other views, those charts are not affected by your edits.

To edit a chart:

1. Move the cursor over the chart, and click the blue menu (down arrow) icon at the top right.

2. Select **Edit**

   This takes you to the **Edit View** page for the view you were at, with the chart you selected already displayed.

   See [Modifying Your Chart](#) for information on the editing features for a chart.

3. Click **Save** to save the revised chart back to the original view.

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 will be displayed below the query box on the charts search page. For example, you might see similar variables for the query shown below:

```
select load_1, load_5, load_15 where entityName=$HOSTID

$HOSTID = server-1.my.company.com $HOSTNAME = server-1.my.company.com
$CLUSTERID = 1
```

Notice the "$HOSTID" portion of the query string: "$HOSTID" is a variable which will be resolved to a specific value based on the page before the query is actually issued. In this case, "$HOSTID" will become "server-1.my.company.com".

The chart below shows an example of the output of a similar query.
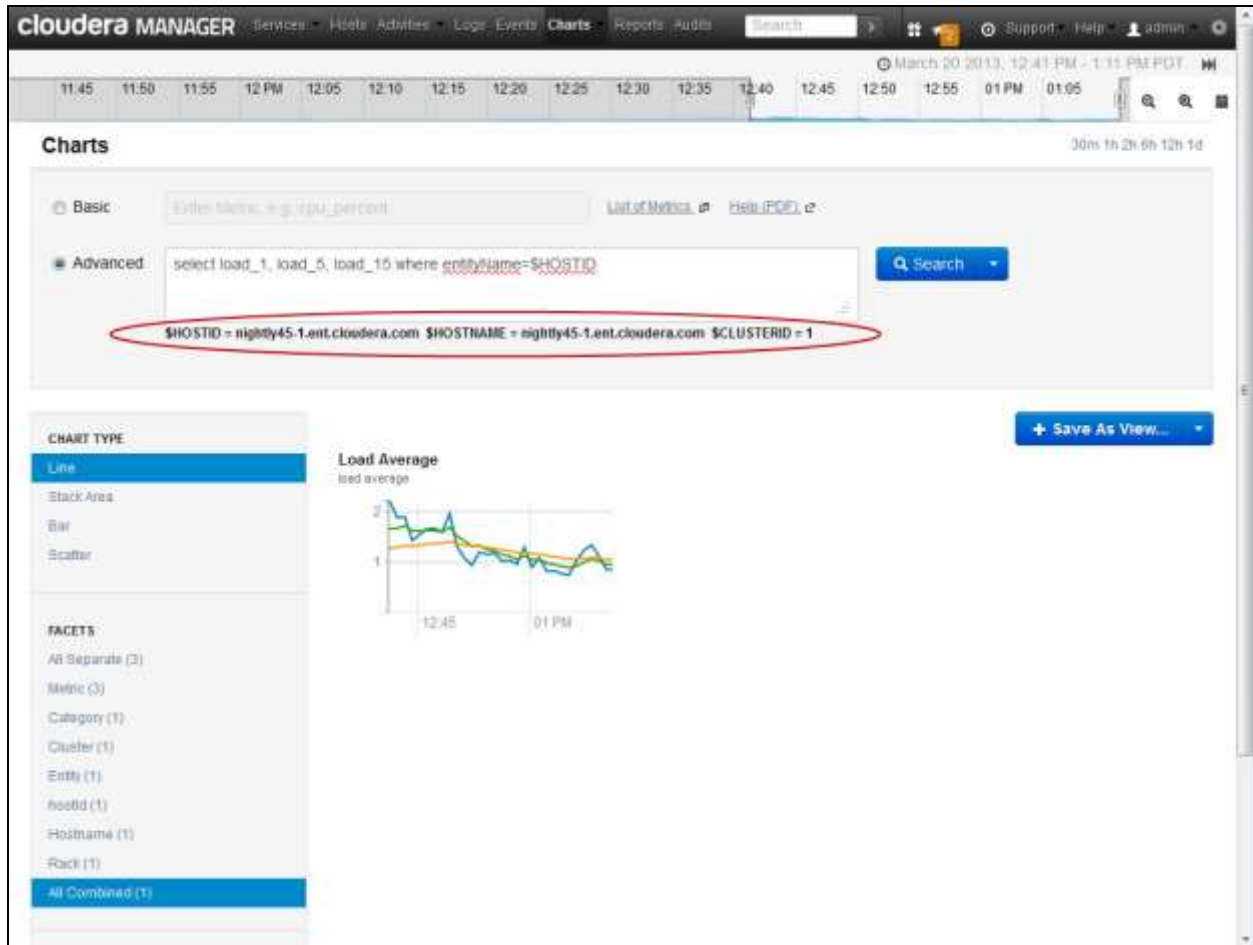
Context-sensitive variables are useful since they allow portable queries to be written. For example the query above may be on the host status page or any role status page to display the appropriate host's swap rate. Variables cannot be used in queries that are part of global views since those views have no service, role or host context.

Copying and Editing a Chart

Editing a copy of a chart is just like editing a chart in the current view, except that you can save it to another existing chart view, back to the current **Custom** view, or to a new chart view that you create. You cannot save a chart to a Status tab Default view. You can copy a chart from any existing chart view, including the Status tab **Default** chart view, and save it to any chart view *except* one of the Status tab Default views.

To make a copy of a chart:

1. Move the cursor over the chart and click the **Edit a Copy** link at the bottom right of the chart.

This opens the Chart Search page with the chart you selected already displayed.

See Modifying Your Chart below for details on how you can modify an existing chart.

2. To save your chart to an existing view*

   a. Click the down-arrow at the right of the **Save as View...** button to display a list of the existing chart views.

   b. Select the view to which you want to add the chart.

3. To save to a new view*

   a. Click the the **Save as View...** button and enter a name for the new chart.

   b. Your new chart view should appear in the menu under the top-level Charts tab.

4. Click your browser back button to return to your original chart view.

Adding a New Chart to the Custom View

From the Custom view under the Status tab of a service, host, or role, you can add new charts to the view.

To add a new chart:

1. Click the Add button at the bottom of the page. This takes you to the **Add To View** page, with variables preset for the specific service, role, or host where you want to add the view.

2. 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.

3. Click **Search.**

   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.

4. To add the new chart back to your chart view, click **Add**.

> **Note:** If the query you've chosen has resulted in multiple charts, all the charts are added to the view as a set. Although the individual charts in this set can be copied, you can only edit the set as a whole.

Modifying Your Chart

In the Charts Search page, you can change the properties of the chart you are editing — the chart type (style), how (or whether) multiple time series are group into charts, and the dimensions and axes ranges of the charts.

The charts page also shows you the tsquery that generated the chart, and you can modify that if you want. For details on modifying the query, see Charting Time-series Data.

Chart Type

Most charts are shown by default as line charts or Stack Area charts.

- To change the chart type, click one of the possible chart types on the left: **Line**, **Stack Area**, and **Bar** or **Scatter**.

Facets

A time-series plot for a service, role, or host may actually be a composite of multiple individual time-series. For charts shown under the Status tab for a service, role, or host, multiple time series may be combined on a single chart. Facets let you chose how to group the different time series in a variety of ways, based on the attributes of those time-series. For example, for a host, the Load Average chart shows you the time-series data for average load at one- five- and ten-minute intervals, by default all on a single chart. Using Facets you can choose to display each time-series as a separate chart. Depending on the query, you can combine or separate time series based on attributes such as Service, Role type, Hostname and so on.

- Click on one of the facets to change the organization of the chart data. The number in parentheses indicates how many charts will be displayed for that facet.

Dimensions and Axes

You can change the size of your charts by moving the **DIMENSION** slider. It moves in 50-pixel increments. 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.

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 one hour of data. You can change the time range for your plot using the time range sets shown at the upper right of the window (right below the Time Range Selector) or by expanding or shrinking the Time Range Selector.

The Processes Tab

To view the processes running for a role instance:

- Select a Service instance to display the Status page for that service.

- Click the Instances tab.

- From the list of Roles, select one to display that role instance's Status page.

- Click the **Processes** tab to see the Processes page.

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

When you are set to the **current** time, you can link from this panel to the Web UI for the role, and see (and view) the relevant configuration files and log files.

- 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.

If you are viewing a point in time in the past, this panel will be visible but greyed out (the data will still show the current values and won't reflect the time marker position). However, the links to the configuration and log files will still work.

Viewing Heatmaps for Services and Roles

Heat maps let you compare the status or performance of the different hosts and role instances in your cluster.

From the Health Tests panel for a Service or Role instance, you can access heat maps that show related metrics for all the nodes in your cluster. These are accessed by clicking the small heatmap icon (▦) to the right of some of the tests in the Health Tests panel for the Service or Role you are viewing.

The heatmap display shows the nodes in your cluster, one cell per node. If you were viewing the status for a specific role, the node where that role resides is outlined in black.
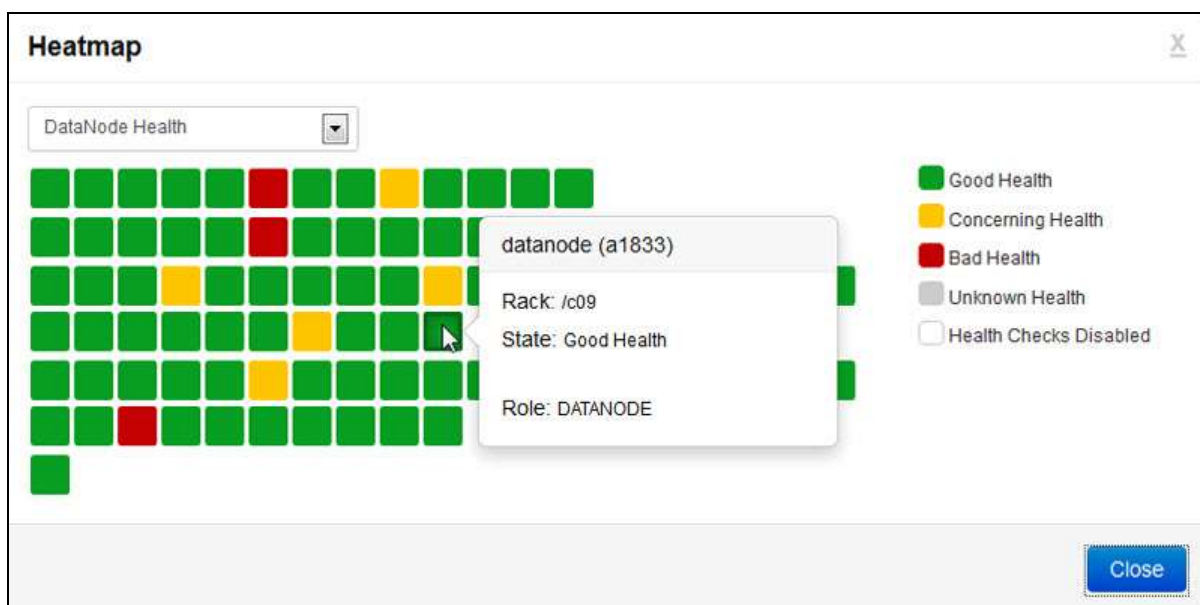
The legend at the right shows the meaning of the colors — in the example above it represents the health of a given node.

The links in the drop-down menu let you view other related metrics.

Moving the cursor over a cell in the grid (as shown above) displays the name of the node, the role, the rack assignment, and other information specific to the type of metric the map is displaying.

A second example shows metrics for a Regionserver. In this case, the color of each cell represents a range of values (duration in milliseconds).
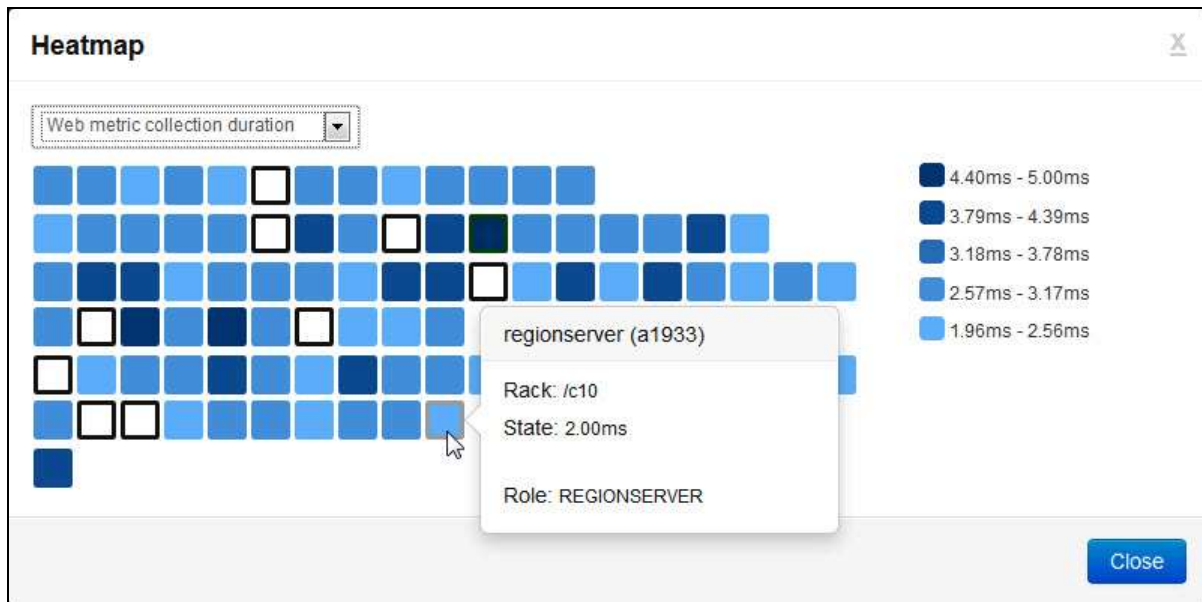
Moving the cursor over a cell displays the exact value for that cell, in addition to the node name and so on.

Configuring Monitoring Settings

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 checks and events, configure how those health checks factor into the overall health of the service, and modify thresholds for the status of certain health checks. Cloudera Manager supports this type of monitoring configuration for HDFS, MapReduce, HBase, ZooKeeper, and Flume. It is also supported for Impala with an Impala monitoring license.

- For individual hosts you can also disable or enable selected health checks, modify thresholds, and enable or disable health alerts.

- Each of the Cloudera Management Services has its own parameters that can be modified in order to modify 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 the Cloudera Management Services you can configure monitoring settings for the monitoring roles themselves — enable and disable health checks on the monitoring processes as well as configuring some general settings related to events and alerts (specifically with the Event Server and Alert Publisher).

In addition, you can configure the basic functions of Cloudera Manager's Management Services through the standard configuration settings for the various management roles. For example, the mail server and related properties for the Alerts Publisher are set under the Default set of Alert Publisher configuration properties.

This section covers the following topics:

- [Configuring Health Check Settings](#)

- [Configuring Directory Monitoring](#)

- [Configuring Activity Monitor Events](#)

- [Configuring Log Events](#)

- [Configuring Alerts](#)

- [Enabling Health Checks for Cloudera Management Services](#)

- [Configuring Cloudera Management Services Database Limits](#)

For general information about modifying configuration settings, see [Changing Service Configurations](#).

Configuring Health Check Settings

The initial 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.

> **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.

To configure a Service monitoring setting:

1. Click the **Services** tab, and select the service instance you want to modify.
   (This can be any of the services for which monitoring is provided, or the Cloudera Management Service.)

2. Click the **Configuration** tab.

3. Click the **Monitoring** category at the bottom of the left-hand **Category** panel.

4. Under the Monitoring category, select the category of properties you want to change (these are organized as Service-Wide or by role).

To configure a Host monitoring setting:

1. Click the **Hosts** tab.

2. To modify the settings for an individual host, select the host.

3. Click the **Configuration** tab.

4. Click the **Monitoring** category in the left-hand **Category** panel.

   Note that if you perform this from the Hosts page, rather than for an individual host, the settings will apply to all hosts.

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

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

You can control when an individual Health Check will be forwarded as an Event or an Alert by modifying the threshold values for the relevant Health test.

Configuring Directory Monitoring

Cloudera Manager can perform threshold-based monitoring of free space in the various directories on the hosts its 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 MiBx/GiBs etc.) 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.

These thresholds are set under the **Monitoring** section of the **Configuration** page for each service.

Configuring Activity Monitor Events

The Activity Monitor monitors the MapReduce jobs running on your cluster. This also includes the higher-level activities, such as Pig, Hive, and Oozie workflows that eventually are run as MapReduce tasks. Currently the Activity Monitor does not support MapReduce v2 (YARN).

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 you must configure Activity Duration Rules that specify what jobs to monitor, and what the limits on duration are for those jobs.

Activity Monitor-related events and alerts for MapReduce are configured via the **Monitoring** category under the **Configuration** tab of the **MapReduce** services page.

To configure Activity Monitor settings for MapReduce:

1. Click the **Services** tab.

2. Select the **MapReduce** service instance.

3. Click the **Configuration** tab.

4. Click the **Monitoring** category at the bottom of the left-hand **Category** panel.

A "slow activity" alert 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 alerts for jobs that exceed the duration defined by these rules.

An Activity Duration Rule is a regular expression (used to match an activity name (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 full Java regular expressions can be used, if the rule set is:

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

In this case, any activity with a name that starts with foo (e.g. fool, food, foot) will match the first rule (see http://download.oracle.com/javase/tutorial/essential/regex/).

If there is not a 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 alert.

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. Alerts for log events is disabled by default for all alerts.

To enable or disable log event capture:

1. Click the **Services** tab, and select the service instance you want to modify.
    You can enable disable event capture for CDH services or for the Cloudera management services.

2. Pull down the **Configuration** tab and select **Edit**.

3. Click the **Monitoring** category at the bottom of the left-hand **Category** panel.

4. Under **Service Wide** > **Events and Alerts**, modify the **Enable Log Event Capture** setting.

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.

To configure which log messages become events:

1. Click the **Services** tab, and select the service instance you want to modify.

2. Pull down the **Configuration** tab and select **Edit**.

3. Click the **Monitoring** category at the bottom of the left-hand **Category** panel.

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 to 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.

A number of useful rules are defined by default, based on Cloudera's experience supporting Hadoop clusters. For example:

- The line `{"rate": 10, "threshold":"FATAL"}`, means log entries with severity FATAL should be forwarded as alerts, 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 alert.

The syntax for these rules is defined in the **Description** field for this property: basically, the syntax lets you create rules that identify log messages based on log4j severity, message content matching, and/or the exception type. These rules must result in valid JSON. You can also specify that the event should generate an alert (by setting `"alert":true` in the rule).
Note that 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.

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

Configuring Alerts

You can configure alerts to be delivered by email, or sent as SNMP traps. These configurations are set under the Alert Publisher role of the Cloudera Manager management service. See Configuring Alert Delivery.

Note that if you just want to add to or modify the list of alert recipient email addresses, you can do this starting at the **Alerts** tab under the **Administration** page, accessed with the gear icon .

You can also send a test alert e-mail from the **Alerts** tab under the **Administration** page.

Enabling Health Checks for Cloudera Management Services

The Cloudera Manager management service provides health checks for its own roles.

You can enable or disable these health checks for each management service. (Role-based health checks are enabled by default).
You can also set a variety of thresholds for specific roles such as thresholds for log directory free space.

Configuring Cloudera Management Services Database Limits

Each Cloudera Management Service maintains a database for retaining the data it monitors. These databases (as well as the log files maintained by these services) can grow quite large. For example, the Activity Monitor maintains data at the service level, the activity level (MapReduce jobs and aggregate activities), and at the task attempt level.   Limits on these data sets are configured when you install your management services, but you can modify these parameters through the Configuration settings in the Cloudera Manager Admin console, for each management service.

For example, the Event Server lets you set a total number of events you want to store. Host Monitor and Service Monitor let you set data expiration thresholds (in hours), and Activity Monitor gives you "purge" settings (also in hours) for the data it stores. There are also settings for the logs that these various services create. You can throttle how big the logs are allowed to get and how many previous logs to retain.

To change any of the data retention or log size settings:

1. From the **Services** tab, select the **Cloudera Management Services** service instance.

2. Pull down the **Configuration** tab and click **Edit**.

3. In the left-hand column, select the role group for the role whose configurations you want to modify.
   (Note that the management services are singleton roles so there will be only a Base role group for the role.)

4. For some services, such as the Activity Monitor, Service Monitor, or Host Monitor, the purge or expiration period properties are found in the top-level settings for the role.

   Typically, Log file size settings will be under the **Logs** category under the role group.

Services Configuration

Cloudera Manager's Services Configuration features let you manage the deployment and configuration of your cluster. You can add new services and roles if needed, gracefully start, stop and restart services or roles, and decommission and delete roles or services if necessary. Further, you can modify the configuration properties for services or for individual role instances, with an audit trail that allows you to role them back if necessary. You can also generate client configuration files, enabling you to easily distribute them to the users of a service.

The following topics describe how to configure and use the services on your cluster.

- Adding Services

- Adding Role Instances

- Modifying Service Configurations

- Managing Role Groups

- Viewing and Reverting Configuration Changes

- Starting, Stopping, and Restarting Services

- Rolling Restart

- Aborting a Pending Command

- Deploying Client Configuration Files

- [Decommissioning a Role Instance](#)

- [Deleting Service Instances and Role Instances](#)

- [Configuring HDFS High Availability](#)

- [Configuring Federated NameServices](#)

- [Running the Balancer](#)

- [Renaming a Service](#)

- [Configuring Agent Heartbeat and Health Status Options](#)

- [Moving the NameNode to a Different Host](#)

Adding Services

After initial installation, you can use the **Add a Service** wizard to add and configure new service instances. For example, you may want to add a service such as Oozie that you did not select in the wizard during the initial installation.

If you have installed a CDH4 cluster, you can use **Add a Service** to add the YARN service to enable MapReduce version 2 (MRv2). By default, the initial installation configures and enables only the original MapReduce service (though you can use the **Custom** option in the wizard to include it). If you add both versions of MapReduce, the original MapReduce service will be given a higher alternatives priority by default, so that the MRv1 configuration will take priority. (You can change this by changing the values of the Alternatives Priority in the MapReduce or YARN configuration settings.)

You can also use **Add a Service** to install Flume NG. Because Flume requires the addition of a configuration file to specify the agent configuration, it must be added separately after the wizard has finished.

As of Cloudera Manager 4.5, Cloudera Impala can be added using the initial installation wizard, and does not need to be added separately.

> **Important**
> The current upstream MRv2 release is not considered stable at this time, and the current Impala release is beta software. Therefore, these are not recommended for use in production at this time.

Adding a Service

1. Click the **Services** tab, then choose **All Services**.

2. From the **Actions** menu, select **Add a Service**.

   A list of possible services are displayed. You can add one type of service at a time.

3. Follow the instructions in the Add Service wizard to add the service.

   As you go through the wizard pages, Cloudera Manager will recommend assignments of service roles to hosts based on the host properties and existing roles on the host; you can modify these assignments if necessary. Cloudera Manager will also recommend configuration settings, such as data directory paths and heap sizes. You can modify the settings as indicated before continuing.

   o If you accept the configuration settings by clicking **Continue** in the wizard's configuration settings page, Cloudera Manager will create the service and its roles and with your specified configuration settings.

   o If you click **Skip** in the wizard's configuration settings page, Cloudera Manager will create the service and its roles without the configuration settings, and you will need to configure the settings later in the **Service > Configuration** tab for the new service.

4. When the wizard is finished:

   o If the new service is not dependent on another service — for example, ZooKeeper — and if you continued with the recommended configurations, the service is configured and started automatically. If you skipped the configuration settings page, the new service will not be configured or started automatically. You must configure the settings for the new service in the **Service > Configuration** pages and then start it.

   o If you added a service that *is* dependent on another service — for example, HBase is dependent on HDFS and ZooKeeper — the new service is not started automatically if the dependent service has an outdated configuration.

   o If you added a service that is dependent on another service that was stopped at the time you used the Add Service wizard, the wizard will start the dependent service for you and perform any other steps required to prepare the cluster for the new service. When you are ready, start the new service.

   o If you added a service that is dependent on another service that was already started at the time you used the Add Service wizard, its configurations might be out of date if you continued with the recommended configurations. Or, you may need to update its configurations so that the new service works correctly. Restart the dependent service before you start the new service.

Note

For information about the order in which to start services, see Starting, Stopping, and Restarting Services.

You can verify the new service is started properly by navigating to **Services > Status** and checking the health status for the new service. If the Health Status is **Good**, then the service is started properly.

Services Configuration

Important

The current upstream MRv2 release is not yet stable, and should not be considered production-ready at this time. It is given by default a lower alternatives priority than MRv1.

Formatting the NameNode and Creating the /tmp Directory

When adding the HDFS service, the **Add Host** wizard automatically formats the NameNode and creates the `/tmp` directory on HDFS. If you quit the **Add Host** wizard or it does not finish, you can format the NameNode and create the `/tmp` directory outside the wizard by doing these steps:

1. Stop the HDFS service if it is running. See [Starting, Stopping, and Restarting Services](#).

2. In the **HDFS > Status** tab, choose **Format** from the **Actions** menu.

3. In the **HDFS > Status** tab, choose **Create /tmp Directory** from the **Actions** menu.

4. Start the HDFS service. See [Starting, Stopping, and Restarting Services](#).

Creating the HBase Root Directory

When adding the HBase service, the **Add Service** wizard automatically creates a root directory for HBase. If you quit the **Add Service** wizard or it does not finish, you can create the root directory outside the wizard by doing these steps:

1. Choose **Create Root Directory** from the **Actions** menu in the **HBase > Status** tab.

2. Click **Create Root Directory** again to confirm.

Initializing the ZooKeeper Service

When adding the ZooKeeper service, the **Add Service** wizard automatically initializes the data directories. If you quit the **Add Service** wizard or it does not finish successfully, you can initialize the directories outside the wizard by doing these steps:

1. Choose **Initialize** from the **Actions** menu in the **ZooKeeper > Status** tab.

2. Click **Initialize** again to confirm.

> **Note:** If the data directories are not initialized, the ZooKeeper nodes cannot be started. If you add additional ZooKeeper roles after the initial installation, you must initialize their data directories before you start those roles.

54 | *Cloudera Manager User Guide*

Creating Beeswax's Hive Warehouse Directory

When adding the Hue service, the **Add Service** wizard automatically creates a Hive warehouse directory for Beeswax. If you quit the **Add Service** wizard or it does not finish, you can create the Hive warehouse directory outside the wizard by doing these steps:

1. In the **Hue > Status** tab, choose **Create Beeswax's Hive Warehouse Directory** from the **Actions** menu.

2. Click **Create Beeswax's Hive Warehouse Directory** again to confirm.

Adding the YARN Service (MapReduce v2)

The YARN service is only available with CDH4.

> **Important**
>
> The current upstream MRv2 release is not yet stable, and should not be considered production-ready at this time. It is given by default a lower alternatives priority than MRv1.

By default, the original MapReduce (MRv1) is set to a higher alternatives priority than YARN. Therefore, even though you add and start the YARN service, MapReduce jobs will still be run under MRv1. If you want to use YARN to run jobs, you will need to change its alternatives priority to be higher than MRv1.

To add the YARN service and configure it to have a higher priority than MRv1, do the following:

1. Follow the steps (above) to add YARN as a service.

2. Stop the YARN service.

3. Go to the YARN service page (by selecting the YARN service from the **Services** menu or from the **All Services** page).

4. Go to the **Configuration** tab, and search for "Alternatives Priority".

5. Change the priority value to be higher than MRv1. (MRv1 by default is set to 92.) Alternatively, you can lower the MapReduce service alternatives priority, via the Configuration tab for the MapReduce service.)

6. Save your configuration change.

7. Start the YARN service.

Creating the Job History Directory

When adding the YARN service, the **Add Service** wizard automatically creates a job history directory for HBase. If you quit the **Add Service** wizard or it does not finish, you can create the directory outside the wizard by doing these steps:

1. Choose **Create Job History Dir** from the **Actions** menu in the **YARN > Status** tab.

2. Click **Create Job History Dir** again to confirm.

Adding Flume

The Flume NG service must be added separately from the wizard; the packages are installed by the installation wizard, but the agents are not configured or started as part of First Run. As part of adding Flume as a service, you should first configure your Flume agents before you start those role instances.

For details of how to modify configurations and use configuration overrides in Cloudera Manager, see Changing Service Configurations.

For detailed information about Flume agent configuration, see the Flume User Guide. **To install Flume agents on your cluster:**

1. Follow the initial steps (above) to select Flume as the service to be added.

2. Select the hosts on which you want Flume agents to be installed.

3. Click **Continue** and the Flume agents are installed on the nodes you've selected.

The Flume agents are not started automatically. You must first configure your agents appropriately before you start them, following the instructions below.

A default Flume flow configuration is provided as an example in the Configuration properties for the flume agents; you should replace this with the your own configuration. The default configuration provides configuration for a single agent.

A single configuration file can contain the configuration for multiple agents, since each configuration property is prefixed by the agent name. You can then set the agents' names using role instance configuration overrides to specify the configuration applicable to each agent. Note that different agent role instances can have the same name. The agent names do not have to be unique. You can use this to further simplify the configuration file. This is the recommended method to configure Flume.

Flume NG can be installed on a cluster running either CDH3 or CDH4. However, monitoring of Flume is only supported if your cluster is running CDH4.1 or later, or CDH3u5 (refresh 2) or later.

> **Note:** If you are using Flume to write to HDFS or HBase sinks, you must have at least one HDFS or HBase role instance on the Flume agent's host. If you do not want to run a daemon on the Flume agent's host, you can just add a Flume Gateway role on the host.

To configure your Flume agents:

1. Go to the Flume Service page (by selecting your Flume service from the **Services** menu or from the **All Services** page).

2. Pull down the **Configuration** tab, and select **View and Edit**.

3. Select the **Agent (Base)** role group in the left hand column.
   The settings you make here apply to the default role group, and thus will apply to all agent instances unless those instances are associated with a different role group, or are overridden for specific agents.

4. Set the **Agent Name** property to the name of the agent (or one of the agents) whose configuration is defined in your `flume.conf`. You can specify only one agent name here — the name you specify will be used as the default for all Flume agent instances, unless you override the name for specific agents. You can have multiple agents with the same name — they will share the same configuration based on your configuration file.

5. Copy the contents of your `flume.conf` file, in its entirety, into the **Configuration File** field. Unless overridden for specific agent instances, this `flume.conf` file will apply to all your agents. You can provide multiple agent configurations in this file and use Agent Name overrides to determine which configurations to use for each agent.
   This is the recommended procedure.

To override the agent name for one or more specific agents:

If you have specified multiple agent configurations in your `flume.conf` file, you must override the default agent name for the agent instances that should use a different (not the default) configuration.

1. Pull down the Flume service **Configuration** tab, select **Edit** and the select the **Agent (Base)** role group in the left hand column.

2. To override the Agent Name for one or more instances, move your cursor over the value area of the **Agent Name** property, and click **Override Instances**.

3. Select the agent (role) instances you want to override.

4. In the field labeled **Change value of selected instances to:** select "Other".
   (You can use the "Inherited Value" setting to return to the service-level value.)

5. In the field that appears, type the agent name you want to use for the selected agents.

6. Click **Apply** to have your change take effect.

After you have completed your configuration changes, you can start the Flume service, which will start all your Flume agents.

> **Note**
>
> If you need to modify your Flume configuration file after you have started the Flume service, you can use the **Update Config...** command from the **Actions** menu on the Flume Service Status page to update the configuration across flume agents without having to shut down the Flume service.

Adding the Cloudera Impala Service

Monitoring of the Impala Service (beyond basic status and health) is a separately-licensed feature. To obtain an upgraded license, contact Cloudera Sales here.

> **Note**
>
> - The current Impala release is **beta software** and not recommended for use in production at this time.
>
> - Cloudera Manager 4.5 supports Cloudera Impala beta version 0.6 running with CDH 4.2; earlier beta versions are not supported.

You can install Cloudera Impala through the Cloudera Manager installation wizard, using either parcels or packages, and have the service started as part of the First Run process. All configuration settings, including the Hive metastore setup, are handled by Cloudera Manager as part of the installation wizard. See Installation Path A - Automated Installation by Cloudera Manager for more information.

If you elect not to include the Impala service using the Installation Wizard, you can you the Add Service wizard to perform the installation.

Impala depends on ZooKeeper, HDFS, HBase, and Hive. All these services must be present in order to run the Impala service.

Simply follow the steps in the Add Service wizard. It will automatically configure and start the dependent services and the Impala service.

Adding Role Instances

After creating a service using one of the wizards, you can add a role instance to that service. For example, after initial installation in which you added the HDFS service, you can add a DataNode to a host where one was not previously running.

In a CDH4 cluster, some services provide roles that are not available with CDH3. For example, HDFS in CDH4 supports **HttpFS**, so that role is available as part of the HDFS service.

CDH4 HDFS also now provides a **Failover Controller** role, which is added to the HDFS service as a companion to each NameNode when you enable Automatic Failover after enabling High Availability. It is

recommended that you let Cloudera Manager add this role as appropriate, rather than adding it manually.

There is also a new role called **Gateway** available in both CDH3 and CDH4 clusters for the HDFS, MapReduce, and HBase services (and for YARN in CDH4) . You can add a Gateway role to a host that does not otherwise have a CDH service installed — this enables Cloudera Manager to install and manage client configurations on that host. This is a convenient way to manage configurations on your CDH clients. There is no process associated with a Gateway role, and its status will always be **Stopped**.

To add a role instance:

1. Click the **Services** tab.

2. Click the link for the service for which you want to add a role instance. For example, click the HDFS service link if you want to add a DataNode role instance for that HDFS service.

3. Click the **Instances** tab.

4. Click the **Add** button.

5. Follow the instructions in the wizard to add the role instance.

   During the wizard, Cloudera Manager will list the existing roles on hosts, recommend configuration settings such as data directory paths and heap sizes depending on the roles. If the new roles are assigned to the same host as roles of another service, Cloudera Manager will recommend configuration changes such that heap allocations of all the roles on the host can be accommodated. You can change some settings, such as data directory paths, before continuing. If you click **Continue** after making changes, the new roles will be created with your configuration changes and configuration settings will be made. If you click **Skip**, the new roles will be created with the recommended changes. If necessary, you can reconfigure the new roles later by navigating to the **Configurations** page of each role or of the service that these roles belong to.

6. The wizard finishes by performing any actions necessary to prepare the cluster for the new role instances. For example, new DataNodes are added to the NameNode's `dfs_hosts_allow.txt` file.

   The new role instance is configured with the Base role group for its role type, even if there are multiple role groups for the role type. If you want to use a different role group, you can go to the **Role Groups** page under the **Configuration** tab for the service, and move the role instance to a different role group. See [Managing Role Groups](#) for more information.

7. The new role instances are not started automatically. You can start them on the service's **Instances** page.

Adding ZooKeeper Roles

If you add ZooKeeper nodes to an existing ZooKeeper service, you must initialize the data directories for the new nodes (role instances) before you restart the ZooKeeper service.

1. Add new ZooKeeper role instances as described in the steps above.

2. Go to the Instances tab for the ZooKeeper service. Your newly added roles should show their status as **Stopped**.

3. From the **Actions** menu at the top of the page, select **Initialize...**.

4. Confirm that you want to perform this action. Note that the dialog will inform you that the action cannot be performed on your previously-existing ZooKeeper nodes.

5. When this action has completed, you can then restart the ZooKeeper service. This will start the new nodes as well as update the configuration for the existing nodes.

When you start the ZooKeeper service after you have added new nodes, the original node will have the datastore, but the datastores of the new nodes will be empty. Therefore, you must ensure that the original node is included when the new quorum is started up. If the new nodes are able to form a quorum without the original node being included, then the ensemble will have an empty datastore. You can avoid this by starting up just the original node plus one of the new nodes and allowing those to form a quorum, resulting a quorum with the datastore from the original node. You can then add the other new nodes.

Modifying Service Configurations

When a service is added to Cloudera Manager, either through the Installation or Upgrade Wizard or with the Add Services workflow, Cloudera Manager configures it with a default set of configuration properties, based on the needs of the service and various characteristics of the cluster in which it will run. These configuration properties include both service-wide configuration settings, as well as specific settings for each role type associated with the service, managed through **Role Groups**. A role group includes a set of configuration properties for that role type, as well as a list of role instances associated with that role group. Cloudera Manager automatically creates a base role group for each role type.

Certain role types — specifically those that allow multiple instances on multiple nodes, such as DataNodes, TaskTrackers, RegionServers — allow the creation of additional role groups that differ from the base configuration. Each role instance can be associated with only a single role group.

Note that when you run the installation or upgrade wizard, Cloudera Manager automatically creates the appropriate base configurations for the roles it adds. It may also create additional role groups for a given role type, if necessary. For example, if you have a DataNode role on the same host as the NameNode, it may require a slightly different configuration that DataNode roles running on other hosts. Therefore, Cloudera Manager will create a separate role group for the DataNode role that is running on the NameNode host, and use the base DataNode configuration for the DataNode roles running on other hosts.

You can modify the settings of the base role group, or you can create new role groups and associate role instances to whichever role group is most appropriate. This simplifies the management of role configurations when one group of role instances may require different settings than another group of instances of the same role type — for example, due to differences in the hardware the roles run on.

For information on creating a new role group, see Managing Role Groups.

Certain roles, such as the CDH3 NameNode and SecondaryNameNode, provide only a base role group, as only one instance of the role can exist in the cluster. You cannot create additional role groups for those roles.

You modify the configuration for any of the service's role groups through the **Configuration** tab for the service. You can also override the settings inherited from a role group for a given role instance, if necessary see Overriding Configuration Settings.

Role groups provide two types of properties: those that affect the configuration of the service itself (the **Default** section in the left-hand panel) and those that affect monitoring of the service, if applicable (the **Monitoring** section). (Not all services have monitoring properties). For more information about monitoring properties see Configuring Monitoring Settings.

> **Important**
>
> If you change configuration settings in the **Configuration** tab after you have started the service or instance, you may need to restart the service or instance. If you need to restart, a message is displayed at the top of the **Configuration** tab. For more information, see Restarting Services and Instances after Configuration Changes.

Changing the Configuration of a Service or Role

To change configuration settings for a service or role:

1. Click the **Services** tab and select the service you want to modify.

2. Pull down the **Configuration** tab at the top of the window and select **Edit**.

   The left hand panel organizes the configuration properties into into categories; first those that are **Service-Wide**, followed by role groups for each role type within the service. Each role group shows its own set of properties, organized by function **Advanced** properties are listed separately for each role group.

   If you have created additional role groups they will appear in this panel and you can modify them just as you can the base role group. For example, if during installation, Cloudera Manager determined that a different set of configuration values is needed for the DataNode colocated with the NameNode, you might see two categories in the Category panel — DataNode (Base)

and HDFS-1-DATANODE-1 (where HDFS-1-DATANODE-1 is the role group Cloudera Manager created for the DataNode instance colocated with the NameNode role).

3. Under the appropriate role group, select the category for the properties you want to change.

4. To search for a text string (such as "safety valve"), in a property, value, or description, enter the text string in the **Search** box at the top of the Category list.

5. Moving the cursor over the value cell highlights the cell; click anywhere in the highlighted area to enable editing of the value. Then type the new value in the field provided (or check or uncheck the box, as appropriate).

   o To facilitate entering some types of values, you can specify not only the value, but also the units that apply to the value. for example, to enter a setting that specifies bytes per second, you can choose to enter the value in bytes (B), KiBs, MiBs, or GiBs — selected from a drop-down menu that appears when you edit the value.

   o To remove the value you entered, click the **Reset to the default value** link.

   o If the property allows a list of values, click the **Plus** icon to the right of the edit field to add an additional field. An example of this is the DataNode Data Directory property, which can have a comma-delimited list of directories as its value.
   To remove an item from such a list, click the **Minus** icon to the right of the field you want to remove.

6. Click **Save Changes** to commit the changes. You can add a note that will be included with the change in the Configuration History.

   This will change the setting for the role group, and will apply to all role instances associated with that role group.

   Depending on the change you made, you may need to restart the service or roles associated with the configuration you just changed. Or, you may need to redploy your client configuration for the service. You should see a message to that effect at the top of the Configuration page, and services may show "Running with Outdated Configuration" as their status.

Validation of Configuration Settings

If you try to use a configuration value that is outside the recommended range of values, Cloudera Manager displays a warning after you click **Save Changes** at the **Configuration** tab. Cloudera Manager displays a warning at the top of the **Configuration** tab and a warning label below the text box where the out-of-range value is entered. Normally this warning value is yellow, but you may get a red validation error if the input is blatantly incorrect.

Overriding Configuration Settings

For role types that allow multiple instances, each role instance inherits its configuration settings from its associated role group. While role groups provide a convenient way to provide alternate configuration settings for selected groups of role instances, there may be situations where you want to make a one-off configuration change — for example when a node has malfunctioned and you want to temporarily reconfigure it. In this case, you can override configuration settings for a specific role instance.

To override a configuration setting for a specific role instance:

1. Go to the **Service** page for the Service whose role you want to change.

2. Click the **Instances** tab

3. Click the role instance you want to change

4. On the role instance page, click the **Configuration** tab.

5. Change the configuration values as appropriate.

6. Save your changes.

Note that you will most likely need to restart your service or role to have your configuration changes take effect. To see a list of all role instances that have an override value for a particular configuration setting, expand the **Overridden by** link in the value cell for the overridden value. To view the override values, and change them if appropriate, click the **Edit Overrides** link. This opens the **Edit Overrides** page, and lists the role instances that have override settings for the selected configuration setting.

On the **Edit Overrides** page, you can do any of the following:

- View the list of role instances that have overridden the value specified in the role group. Use the selections on the drop-down menu below the **Value** column header to view a list of instances that use the inherited value, instances that use an override value, or all instances. This view is

especially useful for finding inconsistent settings in a cluster. You can also use the **Host** and **Rack** text boxes to filter the list.

- Change the override value for the role instances to the inherited value from the associated role group. To do so, select the role instances you want to change, choose **Inherited Value** from the drop-down menu next to **Change value of selected instances to** and click **Apply**.

- Change the override value for the role instances to a different value. To do so, select the role instances you want to change, choose **Other** from the drop-down menu next to **Change value of selected instances to**. Enter the new value in the text box and then click **Apply**.

Resetting Configuration Settings to the Default Value

If you want to reset a setting back to its default value, click the **Reset to the default value** link below the text box in the value cell. The default value is inserted and both the text box and the Reset link disappear.

> **Important:** Explicitly setting a configuration to the same value as its default (inherited value) will have the same effect as using the **Reset to the default value** link. Setting a configuration to the same value as its default value will NOT result in an override.

Using a Configuration Safety Valve

Found in the **Advanced** category (usually under a role group) a Safety Valve configuration setting lets you insert an XML text string into the configuration file, such as `hdfs-site.xml` or `mapred-site.xml`, owned by a service or role. It is intended for advanced use in case there is a specific Hadoop configuration setting that you find is not exposed in Cloudera Manager; contact Cloudera Support if you are required to use it.

For example, there are several safety valves for the NameNode role under the HDFS service **Configuration** tab, **NameNode (Base)** role group, **Advanced** subcategory. There are a number of Safety Valve properties that affect various configuration files; the Description field tells you into which configuration file your additions will be placed. For example, one NameNode safety valve property is called the **NameNode Configuration Safety Valve for hdfs-site.xml**; settings you enter here will be inserted verbatim into the `hdfs-site.xml` file associated with the NameNode — thus each value you enter into that configuration safety valve must be a valid xml property definition, for example:

To see a list of safety valve settings that apply to a specific configuration file, you can enter the configuration file name in the search field and filter for all safety valves that affect that file. For example, searching for `mapred-site.xml` will show all the safety valve settings that have `mapred-site.xml` in their descriptions.

Another example of a safety valve is an *environment* safety valve, such as the `HDFS Service Environment Safety Valve` found under the Service-Wide **Advanced** settings for HDFS. The key/value pairs you specify in an environment safety valve for a service or role are inserted verbatim into the role's environment. Service-wide safety valve values apply to all roles in the service; a safety valve value for a role group apply to all instances of the role associated with that role group.

Restarting Services and Instances after Configuration Changes

If you change the configuration settings after you start a service or instance, you may need to restart the service or instance to have the configuration settings become active. If you need to restart, a message is displayed at the top of the **Configuration** tab when you save your changes.

> **Note:** If you change configuration settings at the service level that affect a particular role only (such as all DataNodes but not the NameNodes), you can restart only that role; you do not need to restart the entire service. If you changed the configuration for a particular instance only (such as one of four Datanodes), you may need to restart only that instance.

To restart a service or instance:

1. Navigate to the **Services > Status** tab for the service.
   The status for the roles whose configuration has changed will be **Started with Outdated**

Configuration; that is, not using your most recent changes.

If you made changes for some but not all instances of a role, the warning will indicate how many are affected. For example, if you changed just one of the four DataNode instances, the DataNode status on the HDFS Service Status page would appear as **3 Started, 1 Started with Outdated Configuration**.

2. To restart the entire service (such as all of HDFS), choose **Restart** from the **Actions** menu (at the upper right).

3. To restart an instance (such as all DataNodes or a particular DataNode), click the **Instances** link, select the instances you want to restart and then select **Restart** from the **Actions for Selected** menu.

4. Click **Restart** that appears in the next screen to confirm.

5. If you see a **Finished** status, the service or instances have restarted.

6. Navigate to the **Service > Status** tab. The service should show a Status of **Started** for all instances and a health status of **Good**.

Managing Role Groups

For role types that allow multiple instances on multiple nodes, such as DataNodes, TaskTrackers, RegionServers (and many others), you can create multiple role groups to allow one set of role instances to use different configuration settings than another set of instances of the same role type.

If there are multiple role groups for a role type, you can move role instances from one group to another.

Creating a New Role Group

To create a new role group for a role type:

1. Click the **Services** tab and select the service where you want to create a new role group.

2. Pull down the **Configuration** tab at the top of the window and select **Role Groups**.

3. Click **Create New Group**.

4. Provide a name for the group.

5. Select the type of role for the group.
   You will only be able to select role types that allow multiple instances, and that exist the Service you have selected.

6. In the **Copy From** field, select the source of the basic configuration information for this role group; you can use any existing role group of the appropriate type, or you can choose **None** and the role group will be set up with all generic default values.

> **Note:** If you select **None** as the source, the default values are *not* the same as the values Cloudera Manager sets in the base role group, as Cloudera Manager specifically sets the

> appropriate configuration properties for the services and roles it installs. After you create your group using **None** you must edit the configuration to set missing properties (for example the TaskTracker Local Data Directory List property, which is not populated if you select None) and clear other validation warnings and errors.

Managing an existing role group

For an existing role groups, you can rename or delete them, and move roles from one to another (of the same role type).

1. Click the **Services** tab and select the service you want to manage.

2. Pull down the **Configuration** tab at the top of the window and select **Role Groups**.

   The Role Groups panel shows the Role Groups you can manage. This will include any groups created automatically by Cloudera Manager (such as the Base group for each role type) as well as any groups of the appropriate role types that you have created.

3. Click the group you want to manage.
   If there are role instances assigned to this role group, they are listed here.

   o To rename the role group, click the **Rename** button, then enter the new name in the pop-up window, and click **Rename** to complete the action.
   You cannot rename a base group.

   o To delete the role group, click the **Delete** button, and confirm you want to perform the delete.
   You cannot delete any of the base groups, and you cannot delete a group if it has role instances associated with it. If you want to delete a group you've created, you must move any role instances to a different role group.

4. To move a role to a different role group:

   a. Select the role instance(s) you want to move

   b. Pull down the **Actions for Selected** menu and select **Move...**

   c. In the pop-up that appears, select the group to which you want to move your selected role instance, and click **Move**.

Viewing and Reverting Configuration Changes

Whenever you change and save a set of configuration settings for a service or role instance, or a host, Cloudera Manager saves a revision of the previous settings and the name of the user who made the changes. You can then view past revisions of the configuration settings, and, if desired, roll back the settings to a previous state.

To view configuration changes

1. Pull down the **Configuration** tab for a service, role, or host, and select **History and Rollback**.

   The most recent revision, currently in effect, is shown under **Current Revision**.

   Prior revisions are shown under **Past Revisions**.

   o By default, or if you click **Show All**, a list of all revisions is shown. If you are viewing a Service or Role Instance, all Service/Role Group related revisions are shown. If you are viewing a Host or All Hosts, all Host/All Hosts related revisions are shown.

   o To list only the configuration revisions that were done in a particular time period, use the Time Range Selector to select a time range. Then, click **Show within the Selected Time Range**.

2. To view the changes, select **Details...**.

Revision Details

For a service or role instance, Revision Details shows the following:

- A brief message describing the context of the changes.

- The date/time stamp of the change.

- The user who performed the change.

- The names of any Role Groups created.

- The names of any Role Groups deleted.

For a host instance, Revision Details shows just a message, date and time stamp, and the user.

The Configuration Values tab

Configuration Value changes are shown under the **Configuration Values** tab, where changes are organized under the Role Group to which they were applied. (For example, if you changed a Service-Wide property, it will affect all role groups for that service).

For each modified property, the Value column shows the new value of the property and the previous value.

The Group Membership tab

If you changed the group membership of a role instance (moved the instance from one group to another) that change is shown under the **Group Membership** tab. This tab is only shown for service and role configurations.

To revert a configuration change

1. Select the Current or Past Revision you want to roll back, and go to the Configuration Values tab.

2. Select the **Revert Configuration Changes** button.

The revert action occurs immediately. Note that you may need to restart the service or the affected roles for the change to take effect.

Note: Only configuration value changes can be reverted. You cannot revert role group actions (creating, deleting, or moving membership among groups). You must perform these actions through the **Role Groups** function under the service or role-level Configuration tab.

Starting, Stopping, and Restarting Services

Starting and Stopping All Services

It's important to start and stop services that have dependencies in the correct order. For example, because MapReduce has a dependency on HDFS, you must start HDFS before starting MapReduce. The Cloudera Management Services and Hue are the only two services that do not have a dependency on other services; although you can start and stop them at anytime, their preferred order is shown in the following procedures.

Starting All Services on All Hosts

To start all services on all hosts:

1. Choose **All Services** from the **Services** tab.

2. Choose **Start** on the **Actions** menu for the service you want to start. Click **Start** that appears in the next screen to confirm. When you see a **Finished** status, the service has started.

3. Repeat Step 2 for each service you want to start.

The order in which to start the services is:

1. HDFS

2. MapReduce

3. YARN

4. ZooKeeper

5. HBase

6. Hue

7. Oozie

8. Impala

9. Flume

10. Cloudera Management Services

> **Note**
>
> If you are unable to start the HDFS service, it's possible that one of the roles instances, such as a DataNode, was running on a host that is no longer connected to the Cloudera Manager Server host, perhaps because of a hardware or network failure. If this is the case, the Cloudera Manager Server will be unable to connect to the Cloudera Manager Agent on that disconnected host to start the role instance which will prevent the HDFS service from starting. To work around this, you can stop all services, abort the pending command to start the role instance on the disconnected host, and then restart all services again without that role instance. For information about aborting a pending command, see Aborting a Pending Command.

Stopping All Services on All Hosts

To stop all services on all hosts:

1. Choose **All Services** from the **Services** tab.

2. Choose **Stop** on the **Actions** menu for the service you want to stop. Click **Stop** that appears in the next screen to confirm. When you see a **Finished** status, the service has stopped.

3. Repeat Step 2 for each service you want to stop.

The order in which to stop the services is:

1. Cloudera Management Services

2. Flume

3. Impala

4. Oozie

5. Hue

6. HBase

7. ZooKeeper

8. YARN

9. MapReduce

10. HDFS

Restarting a Service

It is sometimes necessary to restart a service, which is essentially a combination of stopping a service and then starting it again. For example, if you change the hostname or port where the Cloudera Manager is running, or you enable TLS security, you must restart the Cloudera Management Services to update the URL to the Server.

> **Note**
>
> If you need to restart all services, you should stop them all first and then start them all again in the order described above. It is not possible to restart all of the services in the correct order using the **Restart** command.

To restart a service:

1. Choose **All Services** from the **Services** tab.

2. Choose **Restart** on the **Actions** menu for the service you want to restart. Click **Restart** that appears in the next screen to confirm. When you see a **Finished** status, the service has restarted.

Rolling Restart

Rolling restart allows you to conditionally restart the role instances of your HDFS, MapReduce, HBase, ZooKeeper, and Flume services. Note that if the service is not running, Rolling Restart is not available.

You can do a rolling restart of each of these services individually.

If you have High Availability enabled, you can also perform a cluster-level rolling restart. You cannot perform a cluster-level rolling restart unless you have High Availability enabled.

Restarting an Individual Service

You can initiate a rolling restart from either the Service page for one of the eligible services, or from the service's Instances page, where you can select individual roles to be restarted.

1. From the **Services** page (or the Services tab) select the service you want to restart.

2. From the service's **Actions** menu, select **Rolling Restart...**
   — OR —
   a. Go to the **Instances** tab.
   b. Select the roles you want to restart.
   c. Select **Rolling Restart** from the **Actions for Selected** menu.

3. In the pop-up dialog box, select the options you want:

   o You can choose to restart only roles who's configurations are stale

   o You can choose to restart only roles that are running outdated software versions.

   o Select which role types you want to restart.

4. If you have a significant number of slave roles (DataNodes, TaskTrackers, RegionServers) you can have those restarted in batches. You can configure:

> o How many roles should be included in a batch (the default is one, so individual roles will be started one at a time).
>
> o How long should Cloudera Manager wait before starting the next batch.
>
> o The number of *batch* failures that will cause the entire rolling restart to fail (this is an advanced feature).

5. Click **Confirm** to start the Rolling Restart.

Restarting HDFS

If you do not have High Availability configured, a warning appears reminding you that the service will become unavailable during the restart while the NameNode is restarted. Services that depend on that HDFS service will also be disrupted.

It is recommended that you restart the DataNodes one at a time — one host per batch, which is the default.

Restarting HBase

Administration operations such as any of the following should not be performed during the rolling restart, to avoid leaving the cluster in an inconsistent state:

1. Splits
2. Create/disable/enable/drop table
3. Metadata changes

Restarting MapReduce

If you restart the JobTracker, all current jobs will fail.

Restarting ZooKeeper or Flume

For both ZooKeeper and Flume, the option to restart roles in batches is not available. They are always restarted one by one.

Restarting a Cluster

> **Note:** Rolling Restart for a cluster is available ONLY if you have High Availability enabled. In order to avoid having your cluster go down during the restart, Cloudera Manager will force a failover to the Standby NameNode while the critical roles are being restarted.

1. If you have not already done so, enable High Availability. See Configuring HDFS High Availability for instructions. You do not need to enable Automatic Failover for rolling restart to work, though you can enable it if you wish. Automatic Failover does not affect the rolling restart operation.

2. From the Actions menu for the cluster you want to restart (found on the **All Services** page) select **Rolling Restart...**.

3. In the pop-up dialog box, select the services you want to restart.

    Please review the caveats in the preceding sections for the services you elect to have restarted.

    Note that the services that do not support rolling restart will simply be restarted, and will be unavailable during their restart.

4. If you select an HDFS or HBase service, you can also configure the following:

    o How many roles should be included in a batch (the default is one, so individual roles will be started one at a time).

    o How long should Cloudera Manager wait before starting the next batch.

    o The number of *batch* failures that will cause the entire rolling restart to fail (this is an advanced feature).

5. Click **Confirm** to start the Rolling Restart.

    While the restart is in progress, the Command Details page shows the steps for stopping and restarting the services.

Aborting a Pending Command

Commands will time out if they are unable to complete after a period of time.

If necessary, you can abort a pending command. For example, this may become necessary because of a hardware or network failure where a host running a role instance becomes disconnected from the Cloudera Manager Server host. In this case, the Cloudera Manager Server will be unable to connect to the Cloudera Manager Agent on that disconnected host to start or stop the role instance which will prevent the corresponding service from starting or stopping. To work around this, you can abort the command to start or stop the role instance on the disconnected host, and then you can start or stop the service again.

To abort any pending command:

You can click this indicator ( ) that shows the number of commands that are currently running in your cluster (if any). This indicator is positioned just to the left of the **Support** link at the right hand side of the navigation bar. Unlike the Commands tab for a role or service, this indicator includes all commands running for all services or roles in the cluster. In the Running Commands window, click **Abort** to abort the pending command. For more information, see Viewing Running and Recent Commands.

To abort a pending command for a service or role:

1. Navigate to the **Service > Instances** tab for the service where the role instance you want to stop is located. For example, navigate to the **HDFS Service > Instances** tab if you want to abort a pending command for a DataNode.

2. In the list of instances, click the link for role instance where the command is running (for example, the instance that is located on the disconnected host).

3. Go to the **Commands** tab.

4. Find the command in the list of **Running Commands** and click **Abort Command** to abort the running command.

Deploying Client Configuration Files

To allow Hadoop client users to work with the HDFS, MapReduce, YARN and HBase services you created, Cloudera Manager creates zip files that contain the relevant configuration files with the settings for your services. Each zip file contains the set of configuration files needed by the appropriate service: for example, the MapReduce client configuration zip file contains copies of `core-site.xml`, `hadoop-env.sh`, `hdfs-site.xml`, `log4j.properties` and `mapred-site.xml`.

These client configuration files are generated automatically by Cloudera Manager based on the services and roles you have installed.

Cloudera Manager deploys these configurations automatically when you install your cluster, when you add a service on a host, or when you add a Gateway role on a host. Specifically, for each host that has a service role instance installed, and for each host that is configured as a Gateway role for that service, the Deploy function downloads the configuration zip file, unzips it into the appropriate configuration directory, and uses the Linux "alternatives" mechanism to set a given, configurable priority level.

> **Note:** A Gateway is a role whose sole purpose is to designate a host that should receive a client configuration for a specific service, when the host does not otherwise have any roles running on it. Gateways are configured by going to the **Instances** tab for the service and using the **Add** command to add Gateway roles as needed. You can configure Gateway roles for HDFS, MapReduce, and HBase services (and for YARN in CDH4). See [Adding Role Instances](#) for more information on adding Gateway roles.

Note that if you are installing on a system that happens to to have pre-existing alternatives, then it is possible another alternative may have higher priority and will continue to be used. The alternatives priority of the Cloudera Manager client configuration is configurable under the **Client** section of the **Configuration** tab for the appropriate service.

You can also distribute these client configuration files manually to the users of a service.

The main circumstance that may require a redeployment of the client configuration files is when you have modified the configuration of your cluster. In this case you will typically see a message telling you to redeploy your client configurations. The affected service(s) will also typically be shown as "Running with outdated Configuration."

Viewing and Downloading the Client Configuration Files

You can view the client configuration files using the **Client Configuration URLs** button from the main **Services** tab:

1. Click the **Services** tab in the Cloudera Manager Admin Console.

2. Click the **Client Configuration URLs** button.
   This opens a popup window with links to the configuration zip files that have been created for the services you have installed: HDFS, MapReduce, YARN, and HBase.

3. Click a link to initiate a download of the configuration zip file to your local system.

To download an individual client configuration zip file:

You can download client configuration files for HDFS, MapReduce, YARN, and HBase.

1. Click the **Services** tab in the Cloudera Manager Admin Console and select the service instance whose configuration you want to download.

2. From the **Actions** menu at the top right of the service page, select **Download Client Configuration**
   This initiates a download to your local system of the configuration files for the selected service.

> **Note:** The client configuration files can be downloaded without authentication using tools like `wget` and `curl`.

Redeploying the Client Configuration Files Manually

Although Cloudera Manager will deploy client configuration files automatically in many cases, if you have modified the configurations for a service, you may need to redeploy those configuration files.

If your client configurations were deployed automatically, this command will attempt to redeploy them as appropriate.

> **Note:** If you are deploying client configurations on a node that has multiple services installed, some of the same configuration files, though with different configurations, will be installed in the `conf` directories for each service. Cloudera Manager uses the `priority` parameter in the `alternatives --install` command to ensure that the correct configuration directory is made active based on the combination of services on that node. The priority order (as of Cloudera

Manager 4.1.2) is MapReduce > YARN > HDFS.

The priority can be configured under the **Client/Advanced** section of the **Configuration** tab for the appropriate service.

To deploy all the client configuration files to all nodes on your cluster:

1. Click the **Services** tab in the Cloudera Manager Admin Console.

2. From the cluster-level **Actions** menu at the top right of the page, select **Deploy Client Configuration...**

3. If you are sure you want to proceed, click **Deploy client configuration**.

To deploy client configuration files for a specific service:

1. From the **Services** tab, click the service for which you want to deploy client configurations.

2. From the **Actions** menu at the top right of the service page, select **Deploy client Configuration...**

3. If you are sure you want to proceed, click **Deploy client configuration**.

How Client Configurations are Deployed

Client configuration files are deployed on any host that is a client for a service — i.e. that has a role for the service on that host. This includes roles such as DataNodes, TaskTrackers, RegionServers and so on as well as Gateway roles for the service.

If roles for multiple services are running on the same host (e.g. a DataNode role and a TaskTracker role on the same host) then the client configurations for both roles are deployed on that host, with the alternatives priority determining which configuration takes precedence.

For example, if we have six hosts running roles as follows:
Host H1: HDFS-NameNode
Host H2: MR-JobTracker
Host H3: HBase-Master
Host H4: MR-TaskTracker, HDFS-DataNode, HBase-RegionServer
Host H5: MR-Gateway
Host H6: HBase-Gateway

Client configuration files will be deployed on these hosts as follows:
Host H1: hdfs-clientconfig (only)
Host H2: mapreduce-clientconfig
Host H3: hbase-clientconfig
Host H4: hdfs-clientconfig, mapreduce-clientconfig, hbase-clientconfig

Host H5: mapreduce-clientconfig

Host H6: hbase-clientconfig

If the HDFS NameNode and MR JobTracker were on the same host, then that host would have both hdfs-clientconfig and mapreduce-clientconfig installed.

Decommissioning a Role Instance

If necessary, you can safely remove a role instance such as a DataNode from a cluster while it is running by decommissioning the role instance. When you decommission a role instance, Cloudera Manager performs a procedure for you to safely retire the node on a schedule to avoid data loss. Role decommissioning applies to HDFS DataNodes, MapReduce TaskTrackers, YARN NodeManagers, and HBase RegionServers.

> **Note:** A role will be decommissioned if its host is decommissioned. In that case, the role will appear as decommissioned on the role instances page for the service. You can recommission the role from this page but the host will need to be recommissioned before any roles that were running on it can be restarted. See Decommissioning a Host for more details.

To decommission a role instance:

1. Click the **Services** tab.

2. Click the link for the HDFS, MapReduce, YARN, or HBase service that contains the role instance you want to decommission.

3. Click the **Instances** tab.

4. Select the role instance(s) you want to decommission (such as a DataNode instance).

5. From the **Actions on Selected** menu, select **Decommission**, and then click **Decommission** again to start the process. When you see a **Finished** status, the decommissioning process has finished.

To recommission a role instance:

1. Under the Instances tab, select the decommissioned role instance(s) you want to recommission.

2. From the **Actions on Selected** menu, select **Recommission**, and then click **Recommission** again to start the process. When you see a **Finished** status, the recommissioning process has finished.

If the host for this role instance is currently decommissioned, you will not be able to start this role until the host has been recommissioned.

Deleting Service Instances and Role Instances

> **Note**
>
> Before deleting a service or role instance that is running, you must first stop it.

Deleting a Service Instance

To delete a service instance:

1. Click the **Services** tab.

2. Choose **Delete** from the **Actions** menu for the service instance you want to delete.

3. Click **Delete** to confirm the deletion.

Deleting a Role Instance

To delete a role instance:

1. Click the **Services** tab.

2. Click the service instance that contains the role instance you want to delete. For example, click the `hdfs` service instance if you want to delete a DataNode role instance.

3. Click the **Instances** tab.

4. Select the role instance you want to delete.

5. Choose **Actions for Selected > Delete**. Click **Delete** again to confirm the deletion.

> **Note**
>
> Deleting a service or a role does *not* clean up the associated client configurations that have been deployed in your cluster.

Configuring HDFS High Availability

You can use Cloudera Manager to configure your CDH4 cluster for HDFS High Availability (HA). High Availability is not supported for CDH3 clusters.

An HDFS HA cluster is configured with two NameNodes - an Active NameNode and a Standby NameNode. Only one NameNode can be active at any point in time. HDFS High Availability depends on maintaining a log of all namespace modifications in a location available to both NameNodes, so that in the event of a failure the Standby NameNode has up-to-date information about the edits and location of blocks in the cluster.

There are two implementations available for maintaining the copies of the edit logs:

- High Availability using Quorum-based Storage

- High Availability using an NFS-mounted shared edits directory

Quorum-based Storage relies upon a set of JournalNodes, each of which maintains a local edits directory that logs the modifications to the namespace metadata.

The other alternative is to use a NFS-mounted shared edits directory (typically a remote Filer) to which both the Active and Standby NameNodes have read/write access.

Once you have enabled High Availability, you can enable Automatic Failover, which will automatically failover to the Standby NameNode in case the Active NameNode fails.
You can also initiate a manual failover from Cloudera Manager.

See the _CDH4 High Availability Guide_ for a more detailed introduction to High Availability with CDH4.

> **Important:** Enabling or Disabling High Availability will shut down your HDFS service, and the services that depend on it – MapReduce, YARN, and HBase. Therefore, you should not do this while you have jobs running on your cluster. Further, once HDFS has been restored, the services that depend upon it must be restarted, and the client configurations for HDFS must be redeployed.

> **Important:** Enabling or Disabling High Availability will cause the previous monitoring history to become unavailable.

Enabling High Availability with Quorum-based Storage

After you have installed HDFS on your CDH4 cluster, the **Enable High Availability** workflow leads you through adding a second (Standby) NameNode and configuring JournalNodes.

1. From the **Services** tab, select your HDFS service.

2. Click the **Instances** tab.

3. Click **Enable High Availability**
   (This button does not appear if this is a CDH3 version of the HDFS service.)

4. The next screen shows the hosts that are eligible to run a Standby NameNode and the JournalNodes.

   a. Select **Enable High Availability with Quorum-based Storage** as the High Availability Type.

   b. Select the host where you want the Standby NameNode to be set up.
      The Standby NameNode cannot be on the same host as the Active NameNode, and the

host that is chosen should have the same hardware configuration (RAM, Disk space, number of cores, etc.) as the Active NameNode.

c. Select an odd number of hosts (a minimum of three) to act as JournalNodes. JournalNodes should be hosted on machines with similar hardware specification as the NameNodes. It is recommended that you put a JournalNode each on the same hosts as the Active and Standby NameNodes, and the third JournalNode on similar hardware, such as the JobTracker.

d. Click **Continue**.

5. Enter a directory location for the JournalNode edits directory into the fields for each JournalNode host.

o You may enter only one directory for each JournalNode. The names/paths do not need to be the same on every JournalNode.

o The directories you specify should be empty, and must have the appropriate permissions.

o If the directories are not empty, Cloudera Manager will not delete the contents; however, in that case the data should be in sync across the edits directories of the JournalNodes and should have the same version data as the NameNodes.

6. You can choose whether the workflow will restart the dependent services and redeploy the client configuration for HDFS. To do this manually rather than have it done as part of the workflow, uncheck these extra options.

7. Click **Continue**

Cloudera Manager proceeds to execute the set of commands that will stop the dependent services, delete, create, and configure roles and directories as appropriate, and will restart the dependent services and deploy the new client configuration if those options were selected.

8. There are some additional steps you must perform if you want to use Hive, Impala, or Hue in a cluster with High Availability configured. Follow the Post Setup Steps described below.

Enabling High Availability using NFS Shared Edits Directory

After you have installed HDFS on your CDH4 cluster, the **Enable High Availability** workflow leads you through adding a second (Standby) NameNode and configuring the shared edits directory.

The shared edits directory is what the Standby NameNode uses to stay up-to-date with all the file system changes the Active NameNode makes. Note that you must have a shared directory already configured to which both NameNode machines have read/write access. Typically, this is a remote filer which supports NFS and is mounted on each of the NameNode machines. This directory must be writable by the `hdfs` user, and must be empty before you run the Enable HA workflow.

You can enable High Availability from the **Actions** menu on the HDFS Service page in a CDH4 cluster, or from the HDFS Service **Instances** tab.

1. From the **Services** tab, select your HDFS service.

2. Click the **Instances** tab.

3. Click **Enable High Availability**
   (This button does not appear if this is a CDH3 version of the HDFS service.)

4. The next screen shows the hosts that are eligible to run a Standby NameNode.

   a. Select **Enable High Availability with NFS shared edits directory** as the High Availability Type.

   b. Select the host where you want the Standby NameNode to be installed, and click **Continue**.
      The Standby NameNode cannot be on the same host as the Active NameNode, and the host that is chosen should have the same hardware configuration (RM, Disk space, number of cores, etc.) as the Active NameNode.

5. Confirm or enter the directories to be used as the name directories for the NameNode.

6. Enter the absolute path of the local directory, on each NameNode host, that is mounted to the remote shared edits directory.

   For example, hostA has `/dfs/sharedA` mounted to `nfs:///exported/namenode`, and hostB has `/dfs/sharedB` mounted to the same NFS location. The user should enter `/dfs/sharedA` for hostA and `/dfs/sharedB` for hostB. (`/dfs/sharedA` and `/dfs/sharedB` can be the same paths).

   You should only configure one shared edits directory. This directory must be mounted read/write on both NameNode machines. This directory must be writable by the `hdfs` user, and must be empty when you run the enable HA command.

7. You can choose whether the workflow will restart the dependent services and redeploy the client configuration for HDFS. To do this manually rather than have it done as part of the workflow, uncheck these extra options.

8. Click **Continue** to proceed.

9. Cloudera Manager will now perform the steps to set up the Active and Standby NameNodes.

10. When all the steps have been completed, click **Finish**.

    If the workflow fails, inspect the error message and logs for the cause of failure. After addressing the cause of failure, click **Retry** to re-execute all the steps. Alternatively, perform the remaining steps using the commands available in the **Actions** menu. Note that **Retry** will not work for workflows that fail after the "Bootstrapping Standby NameNode" step. To revert changes made

by the failed workflow, use the **Disable High Availability** action available in the **Instances** tab.

Note that when HA is enabled, there will no longer be a Secondary NameNode role running on your cluster. However, the Secondary NameNode's checkpoint directories are not deleted from the host.

Make sure you start your services and re-deploy your client configurations before you try to run jobs on your cluster, if you did not have the Enable High Availability workflow do this automatically.

11. There are some additional steps you must perform if you want to use Hive, Impala, or Hue in a cluster with High Availability configured. Follow the Post Setup Steps described below.

Note: After you enable High Availability for the first time, there may be a time lag before the next Reports Manager re-indexing phase, which means certain reports may not be immediately available. Restarting the Reports Manager service will make those reports more quickly available.

Post Setup Steps for Hue and Hive

There are several configuration changes you must make in order to successfully enable High Availability, whether you will be using Quorum-based storage or NFS-mounted shared edits directory. Before you enable HA, you must do the following:

- Configure the HDFS Web Interface Role for Hue to be a HTTPFS role. See Configuring Hue to work with High Availability.

- Upgrade the Hive Metastore to use High Availability. You must do this for each Hive service in your cluster. See Upgrading the Hive Metastore for HDFS High Availability.

Configuring Hue to work with High Availability

1. From the **Services** tab, select your HDFS service.

2. Click the **Instances** tab.

3. Click the **Add** button.

4. Under the HttpFS column, select a host where you want to install the HttpFS role and click **Continue**.

5. After you are returned to the Instances page, select the new HttpFS role.

6. From the Actions for Selected menu, select **Start** (and confirm).

7. After the command has completed, go to the **Services** tab and select your Hue service.

8. From the **Configuration** tab, select **Edit**.

9. The HDFS Web Interface Role property will now show the httpfs role you just added. Select it instead of the namenode role, and Save your changes. (The HDFS Web Interface Role property is under the Service-Wide Configuration category.)

10. Restart the Hue service for the changes to take effect.

Upgrading the Hive Metastore for HDFS High Availability

To upgrade the Hive metastore to work with High Availability, do the following:

1. Go to the **Services** tab and select the **Hive** service.

2. From the **Actions** menu, select **Stop...**.

> **Note:** You may want to stop the Hue and Impala services first, if present, as they depend on the Hive service.

Confirm that you want to stop the service.

3. When the service has stopped, back up the Hive metastore database to persistent storage.

4. From the **Actions** menu, click **Update Hive Metastore NameNodes...** and confirm the command.

5. From the **Actions** menu on the Hive Service page, **Start...** the Hive MetaStore Service.
Also restart the Hue and Impala services if you stopped them prior to updating the metastore.

Enabling Automatic Failover

You must have HDFS High Availability enabled in order to enable Automatic Failover.

> **Important:** Enabling or Disabling Automatic Failover will shut down your HDFS service, and requires the services that depend on it to be shut down.

To enable Automatic Failover:

1. From the **Services** tab, select your HDFS service.

2. Click the **Instances** tab.

3. Click **Enable Automatic Failover...**

4. Confirm that you want to take this action.
This will stop the NameNodes for the Nameservice, create and configure Failover Controllers for

each NameNode, initialize the High Availability state in ZooKeeper, and start the NameNodes and Failover Controllers.

> **Note:** If you are using NFS-based High Availability, a fencing method must be configured in order for failover (either automatic or manual) to function — Cloudera Manager configures this automatically. This is not required with Quorum-based Storage. See Fencing Methods if you want more information.

> **Note:** If you started your services and re-deployed your client configurations after you enabled HA, you should not need to do so again now. If you did not start them after enabling HA, you must do so now, before you attempt to run any jobs on your cluster.

Disabling Automatic Failover

> **Note:** You must disable Automatic Failover before you can disable High Availability.

To disable Automatic Failover

1. From the **Services** tab, select your HDFS service.

2. Click the **Instances** tab.

3. Click **Disable Automatic Failover...**

4. Confirm that you want to take this action.
   Cloudera Manager will stop the NameNodes, remove the Failover Controllers, and restart the NameNodes, transitioning one of them to be the Active NameNode.

Disabling High Availability

> **Note:** If you have enabled Automatic Failover, you must disable it before you can disable High Availability.

To disable High Availability

1. From the **Services** tab, select your HDFS service.

2. Click the **Instances** tab.

3. Click **Disable High Availability...**

4. Confirm that you want to take this action.

> If you are using Quorum-based Storage, you will have the option of disabling the Quorum-based Storage, or leaving it enabled. If you are using NameNode Federation, you should consider leaving it enabled.

> Cloudera Manager ensures that one NameNode is active, and saves the namespace. Then it stops the Standby NameNode, creates a SecondaryNameNode, removes the Standby NameNode role, and restarts all the HDFS services.
> Note that although the Standby NameNode role is removed, its name directories are not deleted. Empty these directories after making a backup of their contents.

> As when you enabled High Availability, you have the choice to have your dependent services restarted, and your client configuration redeployed as part of the Disable High Availability workflow. If you choose not to do this, you must do this manually.

Fencing Methods

In order to ensure that only one NameNode is active at a time, a fencing method is required for the shared directory. During a failover, the fencing method is responsible for ensuring that the previous Active NameNode no longer has access to the shared edits directory, so that the new Active NameNode can safely proceed writing to it.

For details of the fencing methods supplied with CDH4, and how fencing is configured, see the Fencing Configuration section in the CDH4 High Availability Guide.

By default, Cloudera Manager configures HDFS to use a shell fencing method (`shell(./cloudera_manager_agent_fencer.py)`) that takes advantage of the Cloudera Manager agent. However, you can configure HDFS to use the `sshfence` method, or you can add your own shell fencing scripts, instead of or in addition to the one Cloudera Manager provides. .

The fencing parameters are found in the **Service-Wide** section of the Configuration tab for your HDFS service.

Converting from NFS-mounted shared edits directory to Quorum-based Storage

Converting your High Availability configuration from using a NFS-mounted shared edits directory to Quorum-based Storage just involves disabling your current High Availability configuration, then enabling High Availability using Quorum-based Storage.

1. Disable High Availability (see Disabling High Availability).

2. Although the Standby NameNode role is removed, its name directories are not deleted. Empty these directories.

3. Enable High Availability with Quorum-based Storage (see Enabling High Availability with Quorum-based Storage).

Converting from Quorum-based Storage to NFS-mounted shared edits directory

To convert your High Availability configuration from using Quorum-based Storage to using a NFS-mounted shared edits directory you disable your current High Availability configuration, configure your NFS-mounted shared edits directories, then enable High Availability using your NFS-mounted directories.

1. Disable High Availability (see Disabling High Availability).

2. Although the Standby NameNode role is removed, its name directories are not deleted. Empty these directories.

3. Enable High Availability using the NFS-mounted directory.

    Note that you must have a shared directory already configured to which both NameNode machines have read/write access.

    See Enabling High Availability using NFS Shared Edits Directory for detailed instructions.

Configuring Federated NameServices

With CDH4, Cloudera Manager supports the configuration of multiple Nameservices managing separate HDFS namespaces, all of which share the storage available on the set of DataNodes. These Nameservices are federated, meaning each Nameservice is independent and does not require coordination with other Nameservices. See HDFS Federation for more information about HDFS Federation.

It is simplest to add a second Nameservice if High Availability is already enabled. The process of enabling High Availability creates a Nameservice as part of the enable High Availability workflow.

> **Important:** Configuring a new Nameservice will shut down the services that depend upon HDFS. Once the new Nameservice has been started, the services that depend upon HDFS must be restarted, and the client configurations must be redeployed. (This can be done as part of the **Add Nameservice** workflow, as an option.)

Converting a non-Federated HDFS Service to a Federated HDFS Service

You must have one Nameservice in place before you can add a second (or additional) Nameservices. Follow the instructions below to convert your current NameNode/SecondaryNameNode setup to a Federated Setup with a single Nameservice.

1. Click the **Services** tab and select your CDH4 HDFS service.

2. Go to the Configurations Tab, and search for "nameservice". This will show you the Nameservice properties for your NameNode and SecondaryNameNode.

3. In the **NameNode Nameservice** field, type a name for the Nameservice. Note that this name must not include the underscore character.

4. In the **Mountpoints** field, change the mount point from "/" to a list of mount points that are in the namespace that this Nameservice will manage. (You can enter this as a comma-separated list — e.g. "`/hbase, /tmp, /user`" or by clicking the plus icon to add each mount point in its own field.)

    You can determine the list of mount points by running the command `hadoop fs -ls /` from the CLI on the NameNode host.

5. In the **SecondaryNameNode Nameservice** field, type the name of the Nameservice. This must be the same as you provided for the **NameNode Nameservice** property.

6. Save your changes.

7. Return to the HDFS Service page, and click the **Instances** tab. You should now see the **Federation and High Availability** section with your Nameservice listed.

8. You can use the **Edit** command under the **Actions** menu to edit the list of mount points for this Nameservice.
    In the **Mountpoints** field, change the mount point from "/" to a list of mount points that are in the namespace that this Nameservice will manage.

Adding a Nameservice

The instructions below for adding a Nameservice assume that a Nameservice is already set up. The first Nameservice cam be set up either by converting a simple HDFS service as described above (see Converting a non-Federated HDFS Service to a Federated HDFS Service or by enabling High Availability.

1. Click the **Services** tab and select your CDH4 HDFS service.

2. Click the **Instances** tab.
    At the top of this page you should see the **Federation and High Availability** section.

    > **Note:** If this section does not appear, it means you do not have any Nameservices configured. You must have one Nameservice already configured in order to add a second. You can either enable High Availability, which will create a Nameservice, or you can convert

your existing HDFS service. See Converting a non-Federated HDFS Service to a Federated HDFS Service for instructions.

3. Click the **Add Nameservice** button.

    a. Enter a name for your new Nameservice. This name must be unique.

    b. Enter at least one mount point for this Nameservice. This defines the portion of HDFS that will be managed under the new Nameservice.
        (Click the **+** to the right of the Mount Point field to add a new mount point).
        You cannot use "/" as a mount point; you must specify HDFS directories by name.

        - Note that the mount points must be unique for this Nameservice; you cannot specify any of the same mount points you have used for other Nameservices.

        - You can specify mount points that do not yet exist, and create the corresponding directories in a later step in this procedure.

        - If you want to use a mount point previously associated with another Nameservice you must first remove that mount point from that service. You can do this using the **Edit** command from the **Actions** menu for that Nameservice, and later add the mount point to the new Nameservice.

        - After you have brought up the new Nameservice, you will need to create, in the new namespace, the directories that correspond with the mount points you specified.

        - If a mount point corresponds to a directory that formerly was under a different Nameservice, you will also need to move any contents of that directory, if appropriate. Instructions for doing this are at the end of this procedure — you must have created the Nameservice before you can create directories in its namespace.

        - If an HBase service is set to depend on the federated HDFS service, make sure to edit the mount points of the existing Nameservice to reference:

            - the HBase service's *HBase root directory* (default value `/hbase`)

            - the *MapReduce System Directory* (default value `/tmp/mapred/system`) and

            - the *MapReduce JobTracker Staging Root Directory* (default value `/user`).

    c. If you want to configure High Availability for this Nameservice, leave the **Highly Available** checkbox checked.

    d. Click **Continue**.

4. Select the hosts on which the new NameNode and SecondaryNameNodes will be created. (Note that these must be hosts that are not already running other NameNode or SecondaryNameNode instances, and their `/dfs/nn` and `/dfs/snn` directories should be empty if they exist.

   Click **Continue**.

5. Enter or confirm the directory property values (these will differ depending on whether you are enabling High Availability for this Nameservice, or not).

6. Uncheck the **Start Dependent Services** checkbox if you need to create directories or move data onto the new Nameservice. Leave this checked if you want the workflow to restart services and redeploy the client configurations as the last steps in the workflow.

7. Click **Continue**.

   If the process finished successfully, click **Finish**.

   You should now see your new Nameservice in the **Federation and High Availability** section in the **Instances** tab of the HDFS service.

8. You must now create the directories you want under the new Nameservice. You need to do this in the CLI.

   a. To create a directory in the new namespace, use the command
      `hadoop fs -mkdir /nameservices/ <nameservice name> / <directory>`
      where `<nameservice name>` is the new nameservice you just created, and `<directory>` is the directory that corresponds to a mount point you specified.

   b. If you need to move data from one Nameservice to another, use `distcp` or manual export/import. `dfs -cp` and `dfs -mv` will not work.

   c. Verify that the directories and data are where you expect them to be.

9. Restart the dependent services.

> **Note:** The monitoring configurations at the HDFS level apply to **all** NameServices. So if you have two NameServices, it is not possible to disable a check on one but not the other. Likewise, it's not possible to have different thresholds for events for the two NameServices.

Nameservice and Quorum-based Storage

With Quorum-based Storage, JournalNodes are shared across Nameservices. So, if JournalNodes are present in an HDFS service, all Nameservices will have Quorum-based Storage enabled. To override this:

- the `dfs.namenode.shared.edits.dir` configuration of the two NameNodes of a High Availability Nameservice should be configured to include the NFS mount, or

- the `dfs.namenode.edits.dir` configuration of the one NameNode of a non-High Availability Nameservice should be configured to include the value of the `dfs.namenode.name.dirs` setting.

Running the Balancer

The Balancer usually shows a status of **N/A** on the **HDFS > Status** tab because the Balancer runs only temporarily.

To run the Balancer:

1. On the **HDFS > Status** tab, choose **Rebalance** on the **Actions** menu.

2. Click **Rebalance** that appears in the next screen to confirm. If you see a **Finished** status, the Balancer successfully ran and completed.

> **Note:** The Balancer role is normally added (by default) when the HDFS service is installed. If it has not been added, you must add a Balancer role instance in order to rebalance HDFS and to see the **Rebalance** action.

Renaming a Service

A service is given a name upon installation, and that name is used as an identifier internally. However, Cloudera Manager allows you to provide your own display name for a service, and that name will appear in the Cloudera Manager User Interface instead of the original (internal) name.

To provide (or change) the display name of a service:

1. Pull down the **Actions** menu for the service, and select **Rename...**

2. Type the new name you want.

3. Click **Rename Service**.

Note, however, that the original service name will still be used internally, and may appear or be required in certain circumstances, such as in log messages or in the API.

The rename action is recorded as an Audit event.

Both the display name and the original (internal) name are displayed when looking at Audit or Event search results for the renamed service, to make it easier to correlate the event or audit text, which uses the internal name.

Configuring Agent Heartbeat and Health Status Options

You can configure the Cloudera Manager Agent heartbeat interval and timeouts to trigger changes in Agent health status.

To configure Agent heartbeat and health status options:

1. Click the gear icon ⚙ to display the **Administration** page.

2. On the **Properties** tab, under the **Performance** category, set the following option:

Setting	Description
Send Agent Heartbeat Every ____ second(s)	The interval between each heartbeat that is sent from Cloudera Manager Agents to the Cloudera Manager Server.

3. On the **Properties** tab, under the **Threshold** category, set the following options:

Setting	Description
Set health status to Concerning if the Agent heartbeats fail ____ time(s)	If an Agent fails to send this number of expected consecutive heartbeats to the Server, a **Concerning** health status is assigned to that Agent.
Set health status to Bad if the Agent heartbeats fail ____ time(s)	If an Agent fails to send this number of expected consecutive heartbeats to the Server, a **Bad** health status is assigned to that Agent.

4. Click **Save Changes**.

For information about health status, see Viewing Service Status.

Moving the NameNode to a Different Host

If necessary, you can move the NameNode role instance to a different host machine. For example, the NameNode host may be having hardware problems and you need to move the NameNode to a properly functioning host. Use the instructions in this section to move the NameNode. The host where you want to move the NameNode must be managed by Cloudera Manager.

Adding a New Host

If you need to first add a new host in the cluster, follow the instructions in <u>Adding a Host to the Cluster</u> and then return to this page and proceed to next section below. If you are moving the NameNode to a host machine where you already installed CDH3 and the Cloudera Manager Agent, you can proceed directly to the next section below.

Moving the NameNode Role Instance to a Different Host

To move the NameNode to a different host machine:

1. Click the **Services** tab and then stop all services. For instructions, see <u>Stopping All Services</u>.

2. Using the command line, make a backup copy of the `dfs.name.dir` directories on the existing NameNode host. Make sure you backup the `fsimage` and `edits` files. They should be the same across all of the directories specified by the `dfs.name.dir` property. You can view the setting for this property in the **HDFS Service > Configuration** tab.

3. Using the command line, copy the files you backed up from `dfs.name.dir` directories on the old NameNode host to the new host where you want to run the NameNode.

4. In Cloudera Manager, click the **Services** tab and navigate to the **HDFS Service > Instances** tab.

5. Select the check box next to the NameNode role instance and then click the **Delete** button. Click **Delete** again to confirm.

6. In the **Review configuration changes** page that appears, click **Skip**.

7. On the same **HDFS Service > Instances** tab, click **Add** to add a NameNode role instance.

8. Select the new host where you want to run the NameNode and then click **Continue**.

9. Specify the location of the `dfs.name.dir` directories where you copied the data on the new host, and then click **Accept Changes**.

10. Click the **Services** tab and then start all services. For instructions, see <u>Starting All Services</u>. After the HDFS service has started, the Cloudera Manager Server will distribute the new configuration files to the DataNodes which will then be configured with the new IP address of the NameNode.

11. Navigate to the **HDFS Service > Status** tab. The NameNode, Secondary NameNode, and DataNode for the HDFS service should each show a process state of **Started**, and the overall HDFS service should show a health status of **Good**.

Managing Multiple Clusters

Cloudera Manager can manage multiple clusters. Once you have successfully installed your first cluster, you can add additional clusters, running the same or a different version of CDH. You can then manage each cluster and its services independently.

To install an additional cluster to be managed by Cloudera Manager, see <u>Adding a Cluster</u>.

Cloudera Manager automatically creates the cluster name and the names of the services running in the cluster – for example, *hdfs1* for the HDFS service instance in Cluster 1, *hdfs2* for the HDFS service instance in Cluster 2. Role names are typically created by combining the role name and the host name on which it is running. For example, a DataNode role instance running on host *node1.xyz.com* would be named *datanode (node1)*.

It is important to note the names of the services and roles, because in some areas within Cloudera Manager they are not organized by cluster, but rather just by service or role name. For example, on the **Logs** or **Events** pages, services are listed by name and are not organized by cluster, so you need to know which service instance is the one you want to look at.

On the **Services** page, if you view All Services, you will see each cluster in its own section, with the Cloudera Management services separately below. The Reports tab also shows the set of reports separately for each cluster. The **Hosts** page lists all hosts under management by Cloudera Manager, but also indicates the cluster to which each belongs.

The **Activities** tab lets you select which cluster's MapReduce service you want to see.

Adding a Cluster

Cloudera Manager can manage multiple clusters. Furthermore, the clusters do not need to run the same version of CDH; you can manage both CDH3 and CDH4 clusters with Cloudera Manager.

To add a cluster with new hosts:

1. From the **Services** tab, click **Add Cluster...**

    This begins the Installation Wizard, just as if you were installing a cluster for the first time. (See the Cloudera Manager Installation Guide for detailed instructions.)

2. To find new hosts, not currently managed by Cloudera Manager, where you want to install CDH, enter the host names or IP addresses, and click **Search**.

    Cloudera Manager lists the hosts you can use to configure a new cluster. Managed hosts that already have services installed will not be selectable.

3. Click **Install CDH on Selected Hosts** to install the new cluster.

    At this point the installation continues through the wizard the same as it did when you installed your first cluster. You will be asked to select the version of CDH to install, which services you want and so on, just as previously.

To add a cluster using currently managed hosts:

Alternatively, you may have hosts that are already "managed" but are not part of a cluster. You can have managed hosts that are not part of a cluster when you have added hosts to Cloudera Manager either through the Add Host wizard, or by manually installing the Cloudera Manager agent onto hosts where

you have not install any other services. This will also be the case if you remove all services from a host so that it no longer is part of a cluster.

1. From the **Services** tab, click **Add Cluster...**

2. To see the list of the currently managed hosts, click the **View Currently Managed Hosts** link under the Search field. This link does not appear if you have no currently managed hosts that are not part of a cluster.

3. To perform the installation, click **Continue**. Instead of searching for hosts, this will attempt to install onto any hosts managed by Cloudera Manager that are not already part of a cluster. It will proceed with the installation wizard as for a new cluster installation.

Moving a Host Between Clusters

Moving a host between clusters can be accomplished by:

1. Decommissioning the host (see Decommissioning a Host

2. Removing all roles from the host (except for the Cloudera Manager management roles). See Deleting Service Instances and Role Instances.

3. Deleting the host from the cluster (see Deleting Hosts), specifically the section on removing a host from a cluster but leaving it available to Cloudera Manager.

4. Adding the host to the new cluster (see Adding a Host to the Cluster).

5. Adding roles to the host (optionally using one of the Host Templates associated with the new cluster). See Adding Role Instances and Working with Host Templates.

Host Configuration and Monitoring

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

The All Hosts Status Tab

The **Status** tab of the **All Hosts** page shows summary information about the hosts under management by Cloudera Manager, and is displayed when the All Hosts list is initially selected.

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 small arrow in front of the number of roles will list 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.

- All columns in the table can be filtered either by typing text or by selecting a filter specification from the drop-down list in the filter field.

- You can change the data you see for the hosts in the host list using the **View Columns** menu:

    o **Current States** shows you basic information (Host name, IP, rack assignment, CDH version, health status, when the Last Heartbeat occurred, and the Decommission status.)

    o **Physical Attributes** shows you physical information about the host such as the number of cores, system load averages for the past 1, 5 and 15 minutes, disk usage, physical memory and swap space usage.

Under the **Actions for Selected** menu you can manage rack locality by assigning hosts to racks, delete hosts, decommission and recommission hosts, start all roles on the host, and enter or exit maintenance mode. You can also apply a Host Template to a host, if that host does not have any roles currently running on it.

You can also **Add New Hosts to Cluster**, run the **Host Inspector**, and **Re-run Host Upgrade Wizard** from this page.

You can change the data you see for the hosts in the host list using the **View Columns** menu:

- **Current States** shows you basic information (Host name, IP, Rack assignment, CDH version, Health status, Last heartbeat, Maintenance Mode status, and Decommission status.)

- **Physical Attributes** shows you physical information about the host such as the number of cores, system load averages for the past 1, 5 and 15 minutes, disk usage, physical memory and swap space usage.

Viewing Individual Hosts

You can view detailed information about an individual host — resources (CPU/memory/storage) used and available, which processes they are running, details about the host agent, and much more — by clicking on the individual host. See Viewing Detailed Information about Hosts.

Configuration Tab

The **Configuration** tab lets you set properties related to parcels and to resource management, and also monitoring properties for the hosts under management. Note that configuration settings you make here will affect all your managed hosts. You can also configure properties for individual hosts from the Host Details page (see Viewing Detailed Information about Hosts) which will override the global properties set here).

To edit the **Default** configuration properties for a host:

1. From the **Configuration** tab, pull down the menu and select **Edit**.

    o Under **Parcels**, the **Configuration** tab lets you specify how parcels will interact with your managed hosts. You can provide a "blacklist" of products that should not be distributed to these hosts.

Blacklisting a product at the All Hosts level will prevent the product from being distributed to any of the managed hosts in your cluster. You can also blacklist products at the individual host level.

o Under **Resource Management** click the checkbox to enable resource management. If this is enabled, then CPU shares and memory limits can be configured through the appropriate role groups.

To modify the monitoring properties for your hosts:

1. From the **Configuration** tab, pull down the menu and select **Edit**.

2. Click the **Monitoring** category.

o 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).

o 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 Templates Tab

The **Templates** tab lets you create and manage Host Templates, which provide a way to specify a set of role configurations that should be applied to a host. This greatly simplifies the process of adding new hosts, because it lets you specify the configuration for multiple roles on a host in a single step, and then (optionally) start all those roles. See Working with Host Templates for more information.

The Parcels Tab

Here you can download, distribute, and activate an available parcel to your cluster. For more information, see Managing Parcels.

The following topics describe how to configure and monitor hosts on your cluster.

- Adding a Host to the Cluster

- Viewing Detailed Information about Hosts

- Working with Host Templates

- Decommissioning a Host

- Deleting Hosts

- Using the Host Inspector

- Managing Parcels

- Performing a Rolling Upgrade on your Cluster

- [Re-Running the Cloudera Manager Upgrade Wizard](#)

- [Resource Management](#)

Adding a Host to the Cluster

You can add one or more hosts to your Hadoop cluster using the Add Hosts wizard, which will install the Oracle JDK, CDH, Impala (optional) and the Cloudera Manager Agent packages. After these packages are installed and the Cloudera Manager Agent is started, the Agent will connect to the Cloudera Manager Server and you will then be able to use the Cloudera Manager Admin Console to manage and monitor CDH on the new host.

The Add Hosts wizard does not create roles on the new host; once you have successfully added the host(s) you can either add roles, one service at a time, or apply a host template, which can define role configurations for multiple roles.

> **Important**
>
> All hosts in a single cluster must be running the same version of CDH, for example CDH3 Update 5 or CDH4.1.
>
> When you install the new hosts on your system, you must install the same version of CDH to enable the new host to work with the other hosts in the cluster. The installation wizard lets you select the version of CDH you want to install, and you can choose a custom repository to ensure that the version you install matches the version on your other hosts.
>
> If you are managing multiple clusters, be sure to select the version of CDH that matches the version in use on the cluster where you plan to add the new hosts.

Using the Add Hosts Wizard to Add Hosts

You can use the Add Hosts wizard to install CDH, Impala, and the Cloudera Manager Agent on a host.

Step 1: Disable TLS Encryption or Authentication

If you have enabled TLS encryption or authentication for the Cloudera Manager Agents, you must disable both of them before starting the Add Hosts wizard. Otherwise, skip to the next step.

To disable TLS:

1. Click the gear icon ⚙ to display the **Administration** page.

2. Under the **Properties** tab, select the **Security** category.

3. Disable all levels of TLS that are currently enabled by deselecting the following options: **Use TLS Encryption for Admin Console**, **Use TLS Encryption for Agents**, and **Use TLS Authentication of Agents to Server**.

4. Click **Save Changes** to save the settings.

5. Restart the Cloudera Management Server to have these changes take effect.

Step 2: Using the Add Hosts wizard

To use the Add Hosts wizard:

1. Click the **Hosts** tab.

2. Click the **Add Hosts** button.

3. Follow the instructions in the wizard to install the Oracle JDK, CDH, Impala, and Cloudera Manager Agent packages or parcels and start the Agent.

4. In the **Specify hosts for your CDH Cluster installation** page, you can search for new hosts to add under the **New Hosts** tab. However, if you have hosts that are already known to Cloudera Manager but have no roles assigned, (for example, a host that was previously in your cluster but was then removed) these will appear under the **Currently Managed Hosts** tab.

5. You will have an opportunity to add (and start) role instances to your newly-added hosts using a Host Template.

    a. You can select an existing host template, or create a new one.

    b. To create a new host template, click the **+ Create...** button. This will open the **Create New Host Template** pop-up. See Working with Host Templates for details on how you select the role groups that define the roles that should run on a host.

        When you have created the template, it will appear in the list of host templates from which you can choose.

    c. Select the host template you want to use.

    d. By default Cloudera Manager will automatically start the roles specified in the host template on your newly added hosts. To prevent this, uncheck the option to start the newly-created roles.

6. When the wizard is finished, you can verify the Agent is connecting properly with the Cloudera Manager Server by clicking the **Hosts** tab and checking the health status for the new host. If the Health Status is **Good** and the value for the Last Heartbeat is recent, then the Agent is connecting properly with the Cloudera Manager Server.

Note that if you did not specify a Host template during the Add Hosts wizard, then no roles will be present on your new hosts until you add them. You can do this by adding individual roles under the **Instances** tab for a specific service, or by using a Host Template. See Adding Role Instances for information about adding roles for a specific service. See Working with Host Templates to create a host template that specifies a set of roles (from different services) that should run on a host.

Step 3: Enable TLS Encryption or Authentication After Using the Add Hosts Wizard

If you previously enabled TLS security on your cluster, you must re-enable the TLS options on the **Administration** page and also configure TLS on each new host after using the Add Hosts wizard. Otherwise, you can ignore this step.

1. Enable and configure TLS on each new host by specifying `1` for the `use_tls` property in the `/etc/cloudera-scm-agent/config.ini` configuration file.

2. Configure the same level(s) of TLS security on the new hosts by following the instructions in Configuring TLS Security for Cloudera Manager.

Adding a Host by Installing the Packages Using Your Own Method

If you used a different mechanism to install the Oracle JDK, CDH, Cloudera Manager Agent packages, you can use that same mechanism to install the Oracle JDK, CDH, Cloudera Manager Agent packages and then start the Cloudera Manager Agent.

To add a host by installing the packages using your own method:

1. Install the Oracle JDK, CDH, Cloudera Manager Agent packages using your own method. For instructions on installing these packages, see Installation Path B - Installation Using Your Own Method.

2. After installation is complete, start the Cloudera Manager Agent. For instructions, see Start the Cloudera Manager Agents.

3. After the Agent is started, you can verify the Agent is connecting properly with the Cloudera Manager Server by clicking the **Hosts** tab and checking the health status for the new host. If the Health Status is **Good** and the value for the Last Heartbeat is recent, then the Agent is connecting properly with the Cloudera Manager Server.

4. If you have enabled TLS security on your cluster, you must enable and configure TLS on each new host. Otherwise, ignore this step.

   a. Enable and configure TLS on each new host by specifying `1` for the `use_tls` property in the `/etc/cloudera-scm-agent/config.ini` configuration file.

   b. Configure the same level(s) of TLS security on the new hosts by following the instructions in Configuring TLS Security for Cloudera Manager.

Viewing Detailed Information about Hosts

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 heart beat to the Cloudera Manager Server

- Number of cores

Host Configuration and Monitoring

- 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, 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 by clicking on the following tabs:

- [Status Tab](#)

- [Processes Tab](#)

- [Resources Tab](#)

- [Commands Tab](#)

- [Configuration Tab](#)

- [Components Tab](#)

- [Audits Tab](#)

- [Charts Tab](#)

Status Tab

This page is displayed when a Host is initially selected. This 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 link at the far right 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/collapsable 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 for more information.

- The small heatmap icon (▦) to the right of some of the tests takes you to a heatmap display that lets you compare the values of the relevant test result metrics across the nodes of your cluster.

Health History

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

- Click the arrow symbol at the left to view the description of the health test state change.

- Click the **View** link to open a new page that shows the state of the host at the time of the transition. Note that 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.

Host Configuration and Monitoring

Host machines 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. 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, or Audits tabs) as well as the status page for the parent Service of the role.

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 Adding Role Instances and Deleting Service Instances and Role Instances.

Charts

Charts are shown for each host instance in your cluster.

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

Heat Maps

Health heat maps let you compare the status or performance of the different hosts in your cluster.

From the Health Tests panel for the host, you can access heatmaps that show related metrics for all the nodes in your cluster. These are accessed by clicking the small heatmap icon (⊞) to the right of some of the tests in the Health Tests panel for the Host you are viewing.

See Viewing Heatmaps for Services and Roles for more information — heatmaps for hosts are very similar to those for roles, and the explanation there applies to hosts as well.

Processes Tab

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.

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 Processes, Commands, Configuration, or Audits 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, or Audits tabs) as well as the status page for the parent Service of the role.

- **Name** — The process name.

- **Link** — A link to the management interface for this role instance on this system. This 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 for this host log entries for this host (a file external to Cloudera Manager).

- **Stderr** — A link to the stderr log for this host (a file external to Cloudera Manager).

- **Stdout** ---A link to the stdout log for this host (a file external to Cloudera Manager).

Resources Tab

Under the **Resources** tab you can view 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:

Category	Description
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.

Host Configuration and Monitoring

Commands Tab

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

Configuration Tab

The **Configuration** tab 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. Pull down the **Configuration** Tab and select **View and Edit**.

2. Click the **Monitoring** category.

   o 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).

   o 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 All Hosts page.

For more information, see Modifying Service Configurations.

Components Tab

The Components tab 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 (CDH3 or CDH4).

- **Component Version** — The detailed version number for each component.

Audits Tab

The Audits tab lets you filter for audit events related to this host. See Viewing the Audit History for more information.

Charts Tab

The Charts tab for a Host instance provides charts for all metrics kept for that host instance, organized by category.
Each category is collapsible/expandable.

- Click on a chart to expand it to full size.

- You can edit a copy of any chart on this page and save it to a custom chart view.

To make a copy of a chart:

1. Move the cursor over the chart and click the **Edit a Copy** link at the bottom right of the chart.

   This opens the Chart Search page with the chart you selected already displayed.

   See [Editing your Time-series Data](#) below for details on how you can modify an existing chart.

2. To save your chart to an existing view*

   a. Click the down-arrow at the right of the **Save as View...** button to display a list of the existing chart views.

   b. Select the view to which you want to add the chart.

3. To save to a new view*

   a. Click the the **Save as View...** button and enter a name for the new chart.

   b. Your new chart view should appear in the menu under the top-level Charts tab.

4. Click your browser back button to return to your original chart view.

Working with Host Templates

Host Templates let you designate a set of role groups that can be applied in a single operation to a host or a set of hosts. This significantly simplifies the process of configuring new hosts when you need to expand your cluster.   Host templates are supported for both CDH4 and CDH3 cluster hosts.

> **Important:** A host template can only be applied on a host with a version of CDH that matches the CDH version running on the cluster to which the host template belongs.

You can create and manage host templates under the **Templates** tab from the **All Hosts** page.

1. Click the **Hosts** tab on the main Cloudera Manager navigation bar.

2. Click the **Templates** tab on the All Hosts page.

Templates are not required; Cloudera Manager assigns roles and role groups to the hosts of your cluster when you perform the initial cluster installation. However, if you want to add new hosts to your cluster, a host templates can make this much easier.

If there are existing host templates, they are listed on the page, along with links to each role group included in the template.

If you are managing multiple clusters, you must create separate host templates for each cluster, as the templates specify role configurations specific to the roles in a single cluster. Existing host templates are listed under the cluster to which they apply.

- You can click a role group name to be taken to the Edit page for that role group, where you can modify the role group settings.

- From the **Actions** menu associated with the template you can edit the template, or delete it.

Creating a Host Template

1. From the **Templates** tab, click **Create...**

2. In the **Create New Host Template** pop-up window that appears:

   o Type a name for the template.

   o For each role, select either the "none" option, or the appropriate role group. There may be multiple role groups for a given role type — you want to select the one with the configuration that meets your needs.
   Selecting the "none" option means no role group will be included in the template for that role type.

3. Click **Create** to create the Host Template.

Editing a Host Template

1. From the **Hosts** tab, click the **Templates** tab.

2. Pull down the **Actions** menu for the template you want to modify, and click **Edit**.

   This put you into the **Edit Host Template** pop-up window. This works exactly like the **Create New Host Template** window — you can modify they template name or any of the role group selections.

3. Click OK when you have finished.

Applying a Host Template to a Host

You can use a host template to apply configurations for multiple roles in a single operation.

You can apply a template to a host that has no roles on it, or that has roles from the same services as those included in the host template. New roles specified in the template that do not already exist on the host will be added. A role on the host that is already a member of the role group specified in the template will be left unchanged. If a role on the host matches a role in the template, but is a member of a different role group, it will be moved to the role group specified by the template.

For example, suppose you have two role groups for a DataNode (*DataNode (Base)* and *DataNode (1)*). The host has a DataNode role that belongs to the *DataNode (Base)* group. If you apply a host template that specifies the *DataNode (1)* group, the role on the host will be moved from *DataNode (Base)* to *DataNode (1)*.

However, if you have two instances of a service, such as MapReduce (for example, *mr1* and *mr2*) and the host has a TaskTracker role from service *mr2*, you cannot apply a TaskTracker role from service *mr1*.

Note that a host may have no roles on it if you have just added the host to your cluster, or if you decommissioned a managed host and removed its existing roles.

Also note that the host must have the same version of CDH installed as is running on the cluster whose host templates you are applying.

If a host belongs to a different cluster than the one for which you created the host template, you can apply the host template if the "foreign" host either has no roles on it, or has only management roles on it. When you apply the host template, the host will then become a member of the cluster whose host template you applied.   The following instructions assume you have already created the appropriate host template.

1. Go to the **All Hosts** page, **Status** tab.

2. Select the host(s) to which you want to apply your host template.

3. From the **Actions for Selected** menu, select **Apply Host Template**.

4. In the pop-up window that appears, select the host template you want to apply.

5. Optionally you can have Cloudera Manager start the roles created per the host template – check the box to enable this.

6. Click **Confirm** to initiate the action.

Decommissioning a Host

Decommissioning a host lets you decommission all roles on a single host without having to go to each service and decommission the roles individually. Once all roles on the host have been decommissioned and stopped, the host can be removed from service.

Decommissioning applies to only to HDFS DataNodes, MapReduce TaskTrackers, YARN NodeManagers, and HBase RegionServers. If the host you select has other roles running on it, those roles will simply be stopped.

Host decommissioning supports decommissioning multiple hosts in parallel.

To decommission one or more hosts:

1. Click the **Hosts** tab.

2. Select the host(s) you want to decommission.

3. From the **Actions for Selected** menu, click **Decommission**.

A confirmation pop-up informs you of the roles that will be decommissioned or stopped on the nodes you have selected. To proceed with the decommissioning, click **Confirm**.

A **Command Details** window appears that will show each stop or decommission command as it is run, service by service. You can click one of the decommission links to see the subcommands that are run for decommissioning a given role. Depending on the role, the steps may include adding the node to an "exclusions list" and refreshing the NameNode, JobTracker, or NodeManager, stopping the Balancer (if it is running), and moving data blocks or regions. Roles that do not have specific decommission actions are just stopped.

While decommissioning is in progress, the host is marked **Decommissioning** in the list under the Hosts tab. Once all roles have been decommissioned or stopped, the host is marked **Decommissioned**.

Roles on a decommissioned host cannot be restarted until the host is recommissioned.

Recommissioning a Host

Only hosts that are decommissioned using Cloudera Manager can be recommissioned.

To recommission one or more hosts:

1. Click the **Hosts** tab.

2. Select the host(s) you want to recommission.

3. From the **Actions for Selected** menu, click **recommission**.

This will recommission the host (i.e. remove it from the exclusion lists and run the appropriate refresh) so that the roles that reside on it can be restarted. The **Decommissioned** indicator is removed from the host. It also removes the **Decommissioned** indicator from the roles that reside on the host. However, the roles themselves are NOT restarted by the recommission command.

You can restart all the roles on a recommissioned host in a single command from the Hosts page:

1. Select the host(s) on which you want to start the decommissioned roles.

2. From the **Actions for Selected** menu, click **Start All Roles**.

This will start all the roles on the selected host.

Deleting Hosts

There are two ways to remove a host from a cluster:

- You can remove a host from a cluster, but leave it available to be added to a different cluster managed by Cloudera Manager.

- You can stop Cloudera Manager from managing a host and the Hadoop daemons on the host.

First, make sure there are no roles running on the Host; you can decommission the host to ensure all roles are stopped.

To remove a host from a cluster but leave it available to Cloudera Manager, you must remove all CDH roles from the host. If the host has Cloudera Manager management roles (such as the Events Server, Activity Monitor and so on), those roles can remain.

To remove a host from Cloudera Manager mangement entirely, you must stop the Cloudera Manager Agent from running on the host; if you don't stop the Agent, it will send heartbeats to the Cloudera Manager Server and show up again in the list of hosts.

To delete a host from being managed by Cloudera Manager:

First, make sure there are no roles running on the Host; you can decommission the host to ensure all roles are stopped.

Second, you must stop the Cloudera Manager Agent from running on the host; if you don't stop the Agent, it will send heartbeats to the Cloudera Manager Server and show up again in the list of hosts.

1. In the Cloudera Manager Admin Console, click the **Hosts** tab.

2. Select the host you want to delete.

3. From the **Actions for Selected** menu, select **Decommission** to ensure that all roles on the host have been stopped. For further details, see Decommissioning a Host.

4. Stop the Agent on the host. For instructions, see Stopping and Restarting Cloudera Manager Agents.

5. In the Cloudera Manager Admin Console, click the **Hosts** tab and select the host you want to delete.

6. From the **Actions for Selected** menu, select **Delete**.

To remove a host from a cluster but leave it available to Cloudera Manager:

First, make sure there are no roles running on the Host; you can decommission the host to ensure all roles are stopped.

Second, remove all CDH roles from the host. If the host has Cloudera Manager management roles (such as the Events Server, Activity Monitor and so on), those roles can remain. The Cloudera Manager Agent should continue to run on the host.

1. From the Hosts tab, select the host you want to delete, and from the **Actions for Selected** menu, select **Decommission**. For instructions, see Decommissioning a Host.

2. Remove all the roles (other than Cloudera Manager Management roles) from the host.   You must go to each Service's Instances tab, and select and delete the appropriate role instance for that service. See Deleting Service Instances and Role Instances.

3. From the **Hosts** tab, select the host you want to remove from the cluster.

4. From the **Actions for Selected** menu, select **Remove From Cluster**.

Using the 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 misconfigurations.

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

To use the Host Inspector:

1. Click the **Hosts** tab.
2. Click **Host Inspector**.
   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 (CDH3 or CDH4).   If you are running CDH3, the Version will be listed as "Unavailable". Version identification is not available with CDH3. In a CDH3 cluster, CDH4 components will be listed as "Not installed or path incorrect".

If you are running multiple clusters with both CDH3 and CDH4, the lists will be organized by distribution (CDH3 or CDH4). 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 ( ) to the left of the Support menu.

2. Click the **Recent Commands** button.

3. If the command is too far in the past, you can use the Time Range Selector to move the time range back to cover the time period you want.

4. When you find the Host Inspector command, click its name to display its subcommands.

5. Click the **Show Inspector Results** button to view the report.

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

Managing Parcels

Cloudera Manager now supports parcels as an alternate form of distribution for CDH and other system packages. Among other benefits, parcels provide a mechanism for performing upgrades to the packages installed on a cluster with minimal disruption.

To enable upgrades with minimal disruption, packages are handles in three steps: Download, Distribution, and Activation.

Downloading a parcel downloads the appropriate software to a local parcel repository on the Cloudera Manager server, where it is available for distribution to the nodes in the cluster. You can have multiple parcels downloaded to your cluster.

Distributing a parcel copies the parcel to the member hosts of the cluster. Note that distributing a parcel does not actually upgrade the components; the current services continue to run unchanged. You can have multiple parcels distributed on your cluster.

> **Note:** The Distribute process does not require Internet access; rather the parcels are downloaded by the Cloudera Manager agents on each cluster member from the local parcel repository on the Cloudera Manager server.

Activating a parcel causes the Cloudera Manager to link to the new components, ready to run the new version upon the next restart. Activating does not stop the current services or perform a restart — instead the system administrator can determine the appropriate time to perform those operations.

Cloudera Manager detects when new parcels are available. The parcel indicator in the Admin console navigation bar ( ) indicates when parcels are available for downloading or distribution. If no parcels are available, or if all parcels have been activated, then this indicator will be zero. Note that you can configure Cloudera Manager to download and distribute parcels automatically, if desired.

Downloading a parcel

1. Click the parcel indicator in the top navigation bar.

   This takes you to the **Hosts** page, **Parcels** tab. By default, any parcels available for download or distribution are shown, as well as any activated parcels.
   Parcels available for download will display a **Download** button.

2. Click **Download** to initiate the download of the parcel to the local repository. You can configure the location of the remote repository, the location of the local parcel repository, and other settings on the **Administration** page, **Properties** tab under the Parcels section. You can also access these properties by clicking the link "Click here to configure these settings" near the top of the Parcels page.

   When the parcel has been downloaded, the button changes to say **Distribute**.

> **Note:** The parcel download is done at the Cloudera Manager server, so with multiple clusters, the downloaded parcels are shown as available to *all* clusters managed by the Cloudera Manager server. However, distribution (to a specific cluster's member nodes) must be selected on a cluster-by-cluster basis.

Distributing a Parcel

Parcels that have been downloaded can be distributed to the nodes in your cluster, available for activation.

1. From the Parcels tab, click the **Distribute** button for the parcel you want to distribute.
   This will begin the process of distributing the parcel to the nodes in the cluster.

If you have a large number of nodes to which the parcels should be distributed, you can control how many concurrent uploads Cloudera Manager will perform. You can configure this setting on the **Adminstration** page, **Properties** tab under the Parcels section.

You can delete a parcel that is ready to be distributed; click the triangle at the right end of the Distribute button to access the Delete command. This will delete the downloaded parcel from the local parcel repository.

Note that distributing parcels to the nodes in the cluster does not affect the current running services.

Activating a parcel

Parcels that have been distributed to the nodes in a cluster are ready to be activated.

1. From the Parcels tab, click the **Activate** button for the parcel you want to activate.
   This will update Cloudera Manager to point to the new software, ready to be run the next time a service is restarted.

2. To start using a new version of a component, go to the **Services** tab and restart your services.

   Until you restart services, the current software will continue to run. This allows you to restart your services at a time that is convenient based on your maintenance schedules or other considerations.

Deactivating a parcel

You can deactivate an active parcel; this will update Cloudera Manager to point to the previous software version, ready to be run the next time a service is restarted.

To use the previous version of the software, go to the **Services** tab and restart your services.

> If you did your original installation from parcels, and there is only one version of your software installed (i.e. no packages, and no additional parcels have been activated and started) then when you attempt to restart after deactivating the current version, your roles will be stopped but will not be able to restart.

Parcel Configuration Settings

You can configure settings for parcels, on the **Administration** page, **Properties** tab under the **Parcels** section. You can also access these properties by clicking the link **Click here to configure these settings** near the top of the Parcels page.

The **Local Parcel Repository Path** defines the path on the Cloudera Manager server host where downloaded parcels are stored.

The **Remote Parcel Repository URLs** is a list of remote repositories the Cloudera Manager should check for parcels. Initially this points to the default repository at archive.cloudera.com (http://archive.cloudera.com/cdh4/parcels/latest/) but you can added your own repository locations to the list.

You can also:

- Set the frequency with which Cloudera Manager will check for new parcels

- Configure a proxy to access to the remote repositories

- Configure whether downloads and distribution of parcels should occur automatically whenever new ones are detected.

- Control which products can be downloaded if automatic downloading is enabled.

If automatic downloading/distribution are not enabled (the default), you must go to the **Parcels** page to initiate these actions.

You can configure the bandwidth limits and the number of concurrent uploads, to tune the load that parcel distribution puts on your network. The defaults are up to 50 concurrent parcel uploads and 50 MiB/s aggregate bandwidth.

- The concurrent upload count (**Maximum Parcel Uploads**) doesn't matter, theoretically, if all nodes have the same speed ethernet. In general, 50 concurrent uploads is an acceptable setting in most cases. However, in a scenario where the server has more bandwidth (say 10Gbe while the normal nodes are using 1Gbe), then the count is important to maximize bandwidth, and would need to be at least the difference in speeds (10x in this case).

- The bandwidth limit (**Parcel Distribution Rate Limit**) should be your ethernet speed (in MiB/seconds) divided by approximately 16. You can use a higher limit if you have QoS set up to prevent starving other services, or if you are willing accept a higher risk of higher bandwidth load.

For individual hosts (or all hosts) you can "blacklist" selected parcels; this will prevent those parcels from being distributed to or activated upon those hosts.

To blacklist a parcel:

1. Go to the Configuration tab for a host (or for All Hosts) and click Edit.

2. Under the parcels category, enter the parcel(s) you want to blacklist. Enter the name as it appears on the Parcels page — for example, 4.1.2-1.cdh4.1.2.p0.30 — and click **Save**.

If a parcel you blacklist has already been distributed to the host, it will be removed from that host. If it is already running on the host, it will continue to run until the next restart, when it will not be restarted.

Parcel File Locations

The default location for the local parcel repository is `/opt/cloudera/parcel-repo` on the Cloudera Manager server host. This is configurable as described above.

The default path for the distributed parcels on the member nodes is `/opt/cloudera/parcels`. This is configurable in the `config.ini` of the Cloudera Manager agent on each node.

Performing a Rolling Upgrade on your Cluster

Cloudera Manager's rolling upgrade feature takes advantage of parcels and the HDFS High Availability configuration to enable you to upgrade your cluster software and restart the upgraded services, without taking the entire cluster down.

> **Note:** You must have HDFS High Availability enabled to perform a Rolling Upgrade.

A rolling upgrade involves two steps:

1. Download, distribute, and activate the parcel for the new software you want to install.

2. Perform a rolling restart to restart the services in your cluster. Note that you can do a rolling restart of individual services, or if you have High Availability enabled, you can perform a restart of the entire cluster. Cloudera Manager will manually fail over your NameNode at the appropriate point in the process so that your cluster will not be without a functional NameNode.

The detailed steps to perform a rolling upgrade of a cluster are as follows:

Step 1. Ensure High Availability is enabled.

To enable High Availability see [Configuring HDFS High Availability](#) for instructions. You do not need to enable Automatic Failover for rolling restart to work, though you can enable it if you wish. Automatic Failover does not affect the rolling restart operation.

Step 2. Download, Distribute, and Activate the parcels with the new software

1. From the All Hosts page, click the **Parcels** tab, and download, distribute and activate the parcel(s) you want to upgrade to.
   For detailed instructions see [Managing Parcels](#).

   When the parcel has been activated, it is not yet running, but is staged to become the running version the next time the service is restarted.

   > **Note**
   >
   > If you are doing a major version upgrade (e.g. from CDH3 to CDH4 using parcels), after the distribution phase the button will be labeled **Upgrade** rather than **Activate**. In this case, follow the instructions at [Upgrading CDH3 to CDH4 in a Cloudera Managed Deployment](#).

2. You are asked if you want to restart the cluster. ***DO NOT RESTART THE CLUSTER AT THIS TIME***.

   o Click **Close**.
   There are several additional steps you must perform prior to restarting your services.

Step 3. Upgrade your Hive Metastore

1. (Strongly recommended) Make a backup copy of your Hive metastore database.

2. Follow the instructions at Step 4. "Upgrading Your Metastore" in the Hive Installation section on Upgrading Hive from an Earlier Version of CDH4 to run the metastore upgrade script.

   o The upgrade script is located at
     ```
     /opt/cloudera/parcels/<parcel_name>/lib/hive/scripts/metastore/up
     grade/<database>.
     ```

     `<parcel_name>` should be the name of the parcel to which you have upgraded.
     `<database>` is the type of database you are running (i.e. mysql, postgres, etc.)

     For example, if you are installing a CDH 4.2.0 parcel using the default location for the local repository, and using the default database (PostgreSQL) the script will be at:
     ```
     /opt/cloudera/parcels/CDH-4.2.0-1.cdh4.2.0.p0.10-
     e16.parcel/lib/hive/scripts/metastore/upgrade/postgres
     ```

   o You must `cd` to the directory the scripts are in.

   o Execute the script in the appropriate DB command shell.

   > **Important**
   >
   > You must know the password for the database; if you installed Cloudera Manager using the default (embedded PostgreSQL) database, the password was displayed on the Database Setup page during the Cloudera Manager installation wizard.
   >
   > If you do not know the password for your Hive metastore database, you can find it as follows:
   >
   > - `cat /etc/cloudera-scm-server/db.properties`
   >   This shows you Cloudera Manager's internal database credentials.
   >
   > - Run the following command:
   >
   >   ```
   >   psql -p 7432 -U scm scm -c "select s.display_name as
   >   hive_service_name, s.name as hive_internal_name,
   >   c.value as metastore_password from configs c, services
   >   s where attr='hive_metastore_database_password' and
   >   c.service_id = s.service_id"
   >   ```
   >
   > - Use the password shown in the `com.cloudera.cmf.db.password` property.
   >
   >   o This script will output the passwords for the hive service metastore as follows:

```
hive_service_name | hive_internal_name | metastore_password
------------------+--------------------+--------------------
 hive1            | hive1              | lF3Cv2zsvI
(1 row)
```

o   For the embedded PostgreSQL database, the database name and user
    name are both `hive`.

3. If you have multiple instances of Hive, run the upgrade script on each metastore
    database.

Step 4. (If Upgrading to CDH 4.2) Upgrade the Oozie Sharelib

1. In the Cloudera Manager Admin Console, select **Oozie** from the **Services** tab. The service should
    already be stopped.

2. From the **Actions** button choose **Install Oozie Sharelib**. The commands to perform this function
    are run.

Step 5. Restart your Cluster

1. From the **Services** tab, select **All Services** to go to the main Services page.

2. From the **Actions** menu for the Cluster, click **Rolling Restart...** to proceed with a Rolling Restart.
    Rolling Restart is available only if High Availability is enabled.

    Click **Restart** to perform a normal restart.

    Services that do not support Rolling Restart will undergo a normal restart, and will not be
    available during the restart process.

3. For a Rolling Restart, a pop-up allows you to chose which services you want to restart, and
    presents caveats to be aware of for those services that can undergo a rolling restart.

    o   For HBase and HDFS, you can configure:

        ▪   How many roles should be included in a batch (the default is one, so individual
            roles will be started one at a time).

        ▪   How long should Cloudera Manager wait before starting the next batch.

        ▪   The number of batch failures that will cause the entire rolling restart to fail (this
            is an advanced feature).

        Please see the Rolling Restart topic for more information about these choices.

4. Click **Confirm** to start the rolling restart.

Step 6. Remove the previous CDH version packages.

If your previous installation of CDH 4 (4.0.x or 4.1.x) was done using packages, you must remove those packages and refresh the symlinks so that clients will run the new software versions.

1. Uninstall the CDH and Cloudera Manager packages:

   On Red Hat systems:

   ```
   $ sudo yum remove hadoop hue oozie hbase sqoop pig flume hive impala
   zookeeper bigtop
   ```

   On SLES systems:

   ```
   $ sudo zypper remove hadoop hue oozie hbase sqoop pig flume hive
   impala zookeeper bigtop
   ```

   On Debian/Ubuntu systems:

   ```
   $ sudo apt-get remove hadoop hue oozie hbase sqoop pig flume hive
   impala zookeeper bigtop
   ```

Step 7. Force an update of the symlinks to point to the newly installed components.

There are two ways you can do this:

- Restart all the Cloudera Manager agents, *OR*

- Disable and re-enable the "Create symlinks" property through the Cloudera Manager Admin console.

Restarting the agents re-triggers symlink creation at agent start up. Unsetting/resetting the symlink property can be done via the Admin console, but it requires waiting at each step until all the agents have synced (at least one heartbeat – harder to verify) for this to be effective.

Do one of the following:

To restart the Cloudera Manager agents:

On each host:

```
$ sudo service cloudera-scm-agent restart
```

To force a re-sync via the Admin Console:

1. In the Cloudera Manager Admin Console, click the gear icon at the right end of the navigation bar to go to the **Administration** page.

2. From the **Properties** tab, **Parcels** section, uncheck the property **Create System-Wide Symlinks for Active Parcels** and **Save** your change.

3. Now, **wait for at least one heartbeat**, then restore the check to the **Create System-Wide Symlinks for Active Parcels** property and **Save** your change.

To restart just selected services:

1. From the **Services** tab, select the service you want to restart, to go to that Service's Status page.

2. From the **Actions** menu, select **Rolling Restart...** or **Restart** (not all services support rolling Restart).

For details of doing a rolling restart, see [Rolling Restart](#).

Re-Running the Cloudera Manager Upgrade Wizard

The first time you log in to the Cloudera Manager server after upgrading your Cloudera Manager software, the upgrade wizard runs.
If you did not complete the wizard at that time, or if you had hosts that were unavailable at that time and still need to be upgraded, you can re-run the upgrade wizard from the Hosts page.

To re-run the Upgrade Wizard:

1. Go to the **Hosts** tab.

2. Click **Re-run Host Upgrade Wizard**.
    This takes you back through the installation wizard to upgrade Cloudera Manager on your hosts as necessary.

3. Specify whether you want to install from the repo that matches your Cloudera Manager server, or from a custom repo. You should only specify the custom option if you have a private repo of the correct version.

4. You will need to provide the SSH credentials for the hosts you want to upgrade.

When you click **Continue** this will upgrade Cloudera Manager on all the currently-managed hosts. You cannot search for new hosts through this process. To add hosts to your cluster, use the **Add Hosts** command.

Resource Management

The 4.5 release of Cloudera Manager reinforces existing resource management techniques and introduces several new ones. These are primarily intended to isolate compute frameworks from one another. For example, MapReduce and Impala often work with the same data set and run side-by-side

on the same physical hardware. Without explicitly managing the cluster's resources, Impala queries may affect MapReduce job SLAs, and vice versa.

Resource Management via Control Groups (Cgroups)

Cloudera Manager 4.5 introduces support for the Control Groups (cgroups) Linux kernel feature. With cgroups, administrators can impose per-resource restrictions and limits on CDH processes.

Enabling Resource Management

> **Important**
>
> If you've upgraded from an older version of Cloudera Manager to Cloudera Manager 4.5, you must restart every Cloudera Manager supervisor process before using cgroups-based resource management. The easiest and safest way to do this is:
>
> 1. Stop all services, including the Cloudera Management Services.
>
> 2. On each cluster node, run as root:
>
> ```
> $ service cloudera-scm-agent hard_restart
> ```
>
> 3. Start all services.

Cgroups-based resource management can be enabled for all hosts, or on a per-host basis.

To enable cgroups for all hosts:

1. Click the **Hosts** tab.

2. Click on the **Configuration** tab, then select **View and Edit**.

3. Select the **Resource Management** category.

4. Check the box next to **Enable Cgroup-based Resource Management**.

To enable cgroups for individual hosts:

1. Click the **Hosts** tab.

2. Click the link for the host where you want to enable cgroups.

3. Click on the **Configuration** tab, then select **View and Edit**.

4. Select the **Resource Management** category.

5. Check the box next to **Enable Cgroup-based Resource Management**.

When cgroups-based resource management is enabled for a particular host, all roles on that host must be restarted for the changes to take effect.

Using Resource management

After enabling cgroups, one can restrict and limit the resource consumption of roles (or role config groups) on a per-resource basis. All of these parameters can be found in the Cloudera Manager configuration UI, under the **Resource Management** category.

CPU Shares

The more CPU shares given to a role, the larger its share of the CPU when under contention. Until processes on the host (including both roles managed by Cloudera Manager and other system processes) are contending for all of the CPUs, this will have no effect. When there is contention, those processes with higher CPU shares will be given more CPU time. The effect is linear: a process with 4 CPU shares will be given roughly twice as much CPU time as a process with 2 CPU shares.

Updates to this parameter will be dynamically reflected in the running role.

I/O Weight

Much like CPU shares, the more the I/O weight, the higher priority will be given to I/O requests made by the role when I/O is under contention (either by roles managed by Cloudera Manager or by other system processes). Note that this only affects read requests; write requests remain unprioritized.

Updates to this parameter will be dynamically reflected in the running role.

Memory Hard Limit

When a role's resident set size (RSS) exceeds the value of this parameter, the kernel will swap out some of the role's memory. If it's unable to do so, it will kill the process. Note that the kernel measures memory consumption in a manner that doesn't necessarily match what the `top` or `ps` report for RSS, so expect that this limit is a rough approximation.

After updating this parameter, the role must be restarted before changes take effect.

Known issues

1. The role config group and role override cgroup-based resource management parameters must be saved one at a time. Otherwise some of the changes that should be reflected dynamically will be ignored.

2. The role config group abstraction is an imperfect fit for resource management parameters, where the goal is often to take a numeric value for a host resource and distribute it amongst running roles. The role config group represents a "horizontal" slice: the same role across a set of hosts. However, the cluster is often viewed in terms of "vertical" slices, each being a combination of slave roles (such as TaskTracker, DataNode, Region Server, Impala daemon, etc.). Nothing in Cloudera Manager guarantees that these disparate horizontal slices are "aligned" (meaning, that the role assignment is identical across hosts). If they unaligned, some of the role config group values will be incorrect on unaligned hosts. For example, a host whose

role config groups have been configured with memory limits but that's missing a role will probably have unassigned memory.

Linux distribution support

Cgroups are a feature of the Linux kernel, and as such, support depends on the underlying host's Linux distribution and version.

Distribution	CPU Shares	I/O Weight	Memory Hard Limit
Red Hat Enterprise Linux (or CentOS) 5			
Red Hat Enterprise Linux (or CentOS) 6	✅	✅	✅
SUSE Linux Enterprise Server 11	✅	✅	✅
Ubuntu 10.04 LTS	✅		✅
Ubuntu 12.04 LTS	✅	✅	✅
Debian 6.0	✅		

If the distribution lacks support for a given parameter, changes to it will have no effect.

The exact level of support can be found in the Cloudera Manager Agent's log file, shortly after the agent has started. See Viewing the Cloudera Manager Server and Agent Logs to find the agent log. In the log file, look for an entry like this:

```
Found cgroups capabilities: {'has_memory': True,
'default_memory_limit_in_bytes': 9223372036854775807,
'writable_cgroup_dot_procs': True, 'has_cpu': True, 'default_blkio_weight':
1000, 'default_cpu_shares': 1024, 'has_blkio': True}
```

The 'has_memory' and similar entries correspond directly to support for the CPU, I/O, and Memory parameters.

Further reading

- http://www.kernel.org/doc/Documentation/cgroups/cgroups.txt

- http://www.kernel.org/doc/Documentation/cgroups/blkio-controller.txt

- http://www.kernel.org/doc/Documentation/cgroups/memory.txt

- http://access.redhat.com/knowledge/docs/en-US/Red_Hat_Enterprise_Linux/6/html/Resource_Management_Guide

Existing resource management controls

Cloudera Manager 4.5 reorganizes existing resource management parameters under the **Resource Management** configuration UI category. Affected parameters include:

- Java maximum heap sizes for all Java-based roles.

- Impala Daemon Memory Limit.

- MapReduce Child Java Maximum Heap Size (Gateway and Client Override).

- MapReduce Map Task Maximum Heap Size (Gateway and Client Override).

- MapReduce Reduce Task Maximum Heap Size (Gateway and Client Override).

- MapReduce Maximum Virtual Memory (Gateway and Client Override).

- MapReduce Map Task Maximum Virtual Memory (Gateway and Client Override).

- MapReduce Reduce Task Maximum Virtual Memory (Gateway and Client Override).

- YARN Container Memory.

- YARN Virtual to Physical Memory Ratio.

- YARN Map Task Maximum Heap Size.

- YARN Reducer Task Maximum Heap Size.

- YARN Map Task Memory.

- YARN Reduce Task Memory.

- YARN Application Master Java Maximum Heap Size.

- YARN Application Master Memory.

Examples

Protecting production MapReduce jobs from Impala queries

Suppose you have MapReduce deployed in production and want to roll out Impala without clobbering your production MapReduce jobs.

For simplicity, we will make the following assumptions:

1. The cluster is using homogenous hardware.

2. Each slave node has 8 GB of RAM.

3. Each slave node is running a Datanode, Tasktracker, and an Impala daemon.

4. Each role type is in a single role config group.

5. Cgroups-based resource management has been enabled on all hosts.

CPU:

1. Leave Datanode and Tasktracker role config group CPU shares at 1024.

2. Set Impala daemon role config group's CPU shares to 256.

Memory:

1. Set Impala daemon role config group memory limit to 1024 MB.

2. Leave Datanode maximum Java heap size at 1 GB.

3. Leave Tasktracker maximim Java heap size at 1 GB.

4. Set MapReduce Child Java Maximum Heap Size for Gateway to 5 GB.

5. Leave cgroups hard memory limits alone. We'll rely on "cooperative" memory limits exclusively, as they yield a nicer user experience than the cgroups-based hard memory limits.

I/O:

1. Leave Datanode and Tasktracker role config group I/O weight at 500.

2. Impala daemon role config group I/O weight is set to 100.

When you're done with configuration, restart all services for these changes to take effect.

The results are:

1. When MapReduce jobs are running, all Impala queries together will consume up to a fifth of the cluster's CPU resources.

2. Individual Impala daemons won't consume more than 1 GB of RAM. If this figure is exceeded, new queries will be cancelled.

3. Datanodes and TaskTrackers can consume up to 1 GB of RAM each.

4. The remainder of each host's available RAM (6 GB) is reserved for MapReduce tasks.

5. When MapReduce jobs are running, read requests issued by Impala queries will receive a fifth of the priority of either HDFS read requests or MapReduce read requests.

Activity Monitoring

Cloudera Manager's activity monitoring capability monitors the Pig, Hive, Oozie, MapReduce and streaming jobs that are running on your cluster. The Activity Monitor provides many statistics about the performance of and resources used by those jobs. You can see which users are running jobs, both at the current time and through views of historical activity. When the individual jobs are part of larger workflows (via Oozie, Hive, or Pig), these jobs are aggregated into 'activities' that can be monitored as a whole, as well as by the component MapReduce jobs.

If you are running multiple clusters, there will be an separate link under the Activities tab for each cluster's MapReduce activities.

Currently, only MapReduce v1 jobs can be monitored with Cloudera Manager's Activity Monitor. MapReduce v2 (YARN) jobs are not currently supported in the Activity Monitor.

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

- Viewing Activities

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

- Viewing a Job's Task Attempts

- Viewing Activity Details in a Report Format

- Comparing Similar Activities

- Viewing the Distribution of Task Attempts

Viewing Activities

From the Activities tab you can view information about the activities (jobs and tasks) that have run in your cluster during a selected time span.

- The list of activities provides specific metrics about the activities 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 the Custom Time Range panel to select the time interval over which job and task information is displayed in the Activities list (see Selecting a Time Range for more details).

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

- View the children (MapReduce jobs) of a Pig or Hive activity.

- View the children (Pig or Hive activities) of an Oozie job.

- View the task attempts generated by a MapReduce 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 amount of input or output data or CPU usage 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.

To view user job activities:

- Click the **Activities** tab on the navigation bar.

- Click the MapReduce service you want to see (if you are running multiple clusters, each cluster's MapReduce service will have a separate entry under the Activities tab).

  The columns in the Activities list show statistics about the performance of and resources used by each activity. By default only a subset of the possible metrics are displayed – you can modify the columns that are displayed to add or remove columns.

  o 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 commands are:

Children	For a Pig, Hive or Oozie activity, this takes you to the **Children** tab of the individual activity page. You can also go to this page by clicking the activity ID in the activity list. This command only appears for Pig, Hive or Oozie activities.
Tasks	For a MapReduce job, this takes you to the **Tasks** tab of the individual job page. You can also go to this page by clicking the job ID in the activity or activity children list. This command only appears for a MapReduce job.
Details	Takes you to the **Details** tab where you can view the activity or job statistics in report form.
Compare	Takes you to the **Compare** tab where you can see how the selected activity compares to other similar activities in terms of a wide variety of metrics.

Task Distribution	Takes you to the **Task Distribution** tab where you can view the distribution of task attempts that made up this job, by amount of data and task duration. This command is available for MapReduce and Streaming jobs.
Kill Job	A pop-up asks for confirmation that you want to kill the job. This command is available only for MapReduce and Streaming jobs.

o The second column shows a chart icon ( ). Select this to chart statistics for this job. If there are charts showing similar statistics for the cluster or for other jobs, the statistics for this job will be added to the chart. See Activity Charts for more details.

o 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.

> **Note:** If the activity is a Pig, Hive, or Oozie activity, no overall status is shown for the activity because the activity may be composed individual MapReduce jobs. Select the activity in the Activities list to view its children — the individual jobs that make up the activity work flow. Each child job shows its own status.

o 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.

To show or hide statistics in the list:

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. Note that 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.

2. Click the small arrow that appears next to the column header to reverse the sort direction.
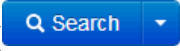
Filtering the Activities list

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 pre-defined queries:

- From the drop-down menu, select the query you want to run. There are predefined queries to search by job type (e.g. Pig jobs, MapReduce jobs and so on) or for running, failed, or long-running activities.

To create a filter:

1. Click the down arrow next to 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 will be remembered across user sessions — i.e 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: Running Tasks, 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 Reducers** 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:

- Click the plus at the top right of the chart panel to open the chart selection panel.

- 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.

To view the jobs that make up a Pig, Hive, or Oozie activity:

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 Activities 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.

Viewing a Job's Task Attempts

To view the task attempts associated with a job:

1. From the **Activities** tab select the activity you want to inspect.

2. If the activity is a MR job, the **Tasks** tab opens.

3. 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 has been 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.**To show or hide statistics in the list:**

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. Note that 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 list by any of the information displayed in the list:

1. Click the column header to initiate a sort.

2. Click the small arrow that appears next to the column header to reverse the sort direction.

Filtering the Tasks List

You can filter the list of activities based on values of any of the metrics that are available.**To create a filter:**

1. Click the down arrow next to the Search button (🔍 Search ▾) 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.

Your filter will only persist for this user session — when you log out, your tasks list filter will be 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 Activities list *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. Select a MapReduce job from the **Activities** list   *OR*
   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 creates a map of the performance of task attempts based on a number of different measures (on the Y-axis) and the length of time taken to complete the task on the X-axis. The chart shows the distribution of tasks in cells that represent the relationship of task duration to values of

the Y-axis metric. The number in each cell shows the number of tasks whose performance statistics fall within the parameters of the cell.

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

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

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

The Y-axis can show Input or Output records or bytes for Map or Reduce tasks, or the amount of CPU seconds for the user who ran the job, while the X-axis shows the task duration in seconds.

In the **Select Axis:** field you can chart the distribution of the following:

- Map Input Records vs. Duration

- Map Output Records vs. Duration

- Map Input Bytes vs. Duration

- Map Output Bytes vs. Duration

- Current User CPUs (CPU seconds) vs. Duration

- Reduce Input Records vs. Duration

- Reduce Output Records vs. Duration

- Reduce Input Bytes vs. Duration

- Reduce Output Bytes vs. Duration

TaskTracker Nodes

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 nodes that executed the tasks that are represented within the cell, and the number of task attempts run on that node.

Clicking a TaskTracker host name takes you to the <u>Role Status</u> page for that TaskTracker instance.

Searching Logs

The Logs page presents log information for Hadoop services, filtered by service, role, host, and/or search phrase as well log level (severity).

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 Current Time button (⏭) is also not available.

You can use the Time Range Selector or the Custom Date panel (▦) to set a specific start and end time, or to choose a period of time back from the current time (selections range from the past 30 minutes to the past day). See <u>Selecting the Time Range</u> for details of how time range selection works in Cloudera Manager.

Searching Logs

To perform a log search:

1. Click the Logs tab.

2. Modify any of the log search parameters as described below, if appropriate. This is optional, all the settings have defaults.

3. Click the **Search** button.

The logs for each of the selected roles are searched. If any of the hosts cannot be searched, an error message notifies you of the error and the host(s) on which it occurred.

The Log Search criteria include the following settings:

*   **Services**
    This section presents 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 Services 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.

*   **Minimum Log Level**
    This specifies the minimum severity level for messages to be included in the search results. Results will include all log entries at the selected level or higher. This defaults to WARN (i.e. a search will return log entries with severity of WARN, ERROR, or FATAL only.

*   **Hosts**
    You can also specify which hosts should be included in the search. To simplify entry, as soon as you start typing a host name, 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.

- **Search Phrase**

  You can specify 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 regex primitives are supported.

- **Search Timeout**

  Specifies a time (in seconds) after which the search will time out. The default is 10 seconds.

- **Results per Page**

  Lets you specify how many results (log entries) to be displayed per page.

Search Results

Search 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 Viewing Detailed Information about Hosts).

- **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 a message takes you to the **Log Details** 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.

Log Details

Clicking the **View Details** link takes you to the **Log Details** page, which presents a portion of the full log, showing the selected message (highlighted) and the 100 messages before and after it in the log.

The page shows you:

- The full path and name of the log file you are viewing.

- The offset in the file of the message you selected, as well as the current offset range (the range of messages shown on the page).

- The 100 messages before and after the one you selected.

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.

Events and Alerts

Events are 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.

Alerts are events that are considered especially noteworthy and are triggered by selected events. Alerts appear in red when you view a events in a list, and you can configure the Alert Publisher to send alert notifications by email or via SNMP trap to a trap receiver such as HP OpenView.

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.

See Searching for Events and Alerts for details about filtering for specific events or alerts.

See Configuring Alert Delivery for information about configuring the Alert Publisher to send email or SNMP notifications for alerts.

If you just want to view a summary of how alerting is configured (what alerts are enabled and disabled across your cluster) you can see this from the **Alerts** tab under the **Administration** page, accessed with the gear icon ![gear icon]. You can also see this from the **All Alerts Summary** link from the Alert Publisher's role page.

Searching for Events and Alerts

The **Events** page lets you search for and display events and alerts that have occurred within a time range you select anywhere in your clusters. You can use the Time Range Selector or the time range link (30m 1h 2h 6h 12h 1d) to set the time range for your search. (See Selecting a Time Range for details). Note that 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.

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.

Filtering Events

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

Adding Filters

To add a filter to an event log, do one of the following:

- Click an event property value link in one of the event entries.
  A filter containing the property and its value is added to the list of filters at the left and Cloudera Manager redisplays all events that match the filter.

- Click the **+ Add Filter** to the left of the log.
  A filter control is added to the list of filters.

  a. Choose an event property in the property drop-down list.

  b. If the property is a numeric type, choose an operator in the operator drop-down list.

  c. Type an event property value in the value text field.
      Note that for some properties, where the list of values is finite and known, you can start typing and then select from a list of potential matches.
      For some properties you can include multiple values in the value field. For example, you can create a filter like "SERVICE = HBASE1, HDFS1" which will find events from either service.

  d. Click **Add Another** to add additional filter components. A filter containing the property and its value is added to the list of filters at the left. Multiple filters are combined using AND – e.g. SERVICE = HBASE1 AND SEVERITY = CRITICAL)

  e. Click **Search**.
      The log displays all events that match the filter criteria.

Removing a Filter

To remove a filter from a filter specification:

1. Click the ✖ at the right of the filter.
    The filter is removed and the event log redisplays all events that match the remaining filters.

If there are no filters, the event log displays all events.

Modifying a Filter

To modify a filter:

1. Click the filter.
    The filter expands into separate property, operator, and value fields.

2. Modify the value of one or more fields.

3. Click **Search**.
    A filter containing the property, operation, and value is added to the list of filters at the left and the event log redisplays all events that match the modified set of filters.

The Events Log Display

Event log entries are ordered (within the time range you've selected) with the most recent at the top.

- Click on a link in an event entry to add that value as an additional filter.

- Click the **View** link to go to the **Logs** page to view the log entry for the event.

- Click the arrow at the right side of the event entry (**>**) to display details of the entry.

  - In the detail display, clicking on the **URL** link also takes you to the log entry for the event.

  - Clicking any other link adds them as additional filter specifications.

- If this event generated an Alert, that is indicated by a red Alert icon (**① Alert**) in the entry.

Configuring Alert Delivery

Under the Alert Publisher role of the Cloudera Manager Management Service, you can configure email or SNMP delivery of alert notifications.

Configuring Alert Email Delivery

When you install the Cloudera Manager Management Services, it asks you for information about the mail server you will use with the Alert Publisher. However, if you need to change these settings, you can do so under the Alert Publisher section of the management Services configuration tab.

Note that if you just want to add to or modify the list of alert recipient email addresses, you can do this starting at the **Alerts** tab under the **Administration** page, accessed with the gear icon 🔧.

You can also send a test alert e-mail from the **Alerts** tab under the **Administration** page.

You can enable and disable email alerts delivery entirely (without changing the other email settings) with the **Enable email alerts** property.

To enable, disable, or configure email alerts:

1. From the **Services** tab, select the **Cloudera Management Services** service instance.

2. Pull down the **Configuration** tab and click **Edit**.

3. Select the **Alert Publisher (Base)** role group to see the list of properties.
   In order to receive email alerts you must set (or verify) the following settings:

   - Email protocol to use.

   - Your mail server hostname and port.

   - The username and password of the email user that will be logged into the mail server as the "sender" of the alert emails.

   - A comma-separated list of email addresses that will be the recipients of alert emails.

o The format of the email alert message. Select **json** if you need the message to be parsed by a script or program.

4. Click the **Save Changes** button at the top of the page to save your settings.

5. You will need to restart the Alert Publisher role to have these changes take effect.

Configuring SNMP

Before you enable SNMP traps, make sure you have configured your trap receiver (Network Management System or SNMP server) with the Cloudera MIB.

To view the Cloudera MIB:

1. From the All Services page, go to the Cloudera Manager management service.

2. From the **Configuration** tab select **View and Edit**.

3. Expand the **Alert Publisher** category in the Category list at the left, and select **SNMP**.

4. In the **Description** column for the first property (**SNMP NMS Hostname**) there is a link to the **SMNP MIB**.

5. Click the link in the Description field to view the MIB.

To enable, disable, or configure SNMP traps:

1. From the **Services** tab, select the **Cloudera Management Services** service instance.

2. Pull down the **Configuration** tab and click **Edit**.

3. Under the **Alert Publisher (Base)** role group select **SNMP** to see the list of properties.

    o Enter the DNS name or IP address of the Network Management System (SNMP server) acting as the trap receiver in the SNMP NMS Hostname property.

    o Select the version of SNMP you are using: SNMPv2, SNMPv3 authentication with no privacy (`authNoPriv`), or SNMPv3 with no authentication and no privacy (`noAuthNoPriv`).

    o For SNMPv2, you must enter a Community String.

    o For SNMPv3, you must enter the SNMP Server Engine ID.

    o For SNMPv3 with authentication (`authNoPriv`) you must also enter the Security user name, Authentication protocol, and protocol pass phrase.

    o You can also change other settings such as the port, retry, or timeout values.

4. Click **Save Changes** when you are done.

5. You must restart the Alert Publisher role to have these changes take effect.

To disable SNMP traps, simply remove the hostname from the SNMP NMS Hostname property (alert.snmp.server.hostname).

Alert Settings

The **Alerts** tab (found on the Administration page) provides a summary of the settings for alerts in your clusters.

- Click the gear icon (⚙) to display the **Administration** page, then click the **Alerts** tab.

Alert Type

The left column lets you select by alert type (Health, Log, or Activity) and within that by service instance. In the case of Health alerts, you can look at alerts for Hosts as well. You can select an individual service to see just the alert settings for that service.

Health/Log/Activity Alert Settings

Depending on your selection in the left column, the right hand column show you the list of alerts that are enabled or disabled for the selected service type.

To change the alert settings for a service, click the edit icon (✎) next to the service name. This will take you to the Monitoring section of the Configuration tab for the service. From here you can enable or disable alerts and configure thresholds as needed.

Recipients

You can also view the list of recipients configured for the enabled alerts. Again, click the edit icon (✎) at the top of this list to go to the Alert Publisher configuration settings, where you can modify the list of recipients.

If you want to verify that the recipients will actually receive an alert, click the **Send Test Alert** link under the list of recipients. This will send a test alert to all recipients in the list.

Charting Time-series Data

The **Charts Search** page in the Cloudera Manager Admin Console enables you to search for a time series, plot the time-series data, group (facet) the individual time-series if your search produced multiple time series, and save the results as a user-defined view.

The tsquery language is used to retrieve time-series data from the Cloudera Manager time-series data store. See the Tsquery Language for more details.

To create a custom view with time-series charts of your own choosing, see Creating Custom Views.

To access the charts search page:

- In the Cloudera Manager Admin Console, pull down the **Charts** tab in the top navigation bar and select **Search**.

Terminology

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

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

Time series: A list of (time, value) pairs that is associated with some (entity, metric) pair, e.g., (`datanode-1`, `fd_open`). In more complicated cases the time-series can represent operations on other time-series, for example, (`datanode-1`, `cpu_user`) + (`datanode-1`, `cpu_system`).

Facet: A display grouping of the dataset, shown in separate charts. 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.

Searching for Time-series Data

You can search for a time series in the Cloudera Manager Admin Console in two ways: by metric, or by constructing a query in the **Advance Search** text box.

Searching by Metric

There are two ways to search by metric: type the metric name or description into the **Basic** text field, or select it from the **List of Metrics**.

To search by typing the name (or part of the name) of a metric or its description:

1. Start typing the name or description of the metric in the **Basic** Metric text box.

2. As you type, metrics that match the letters you enter will appear in a drop-down list and you can select the one you want.

3. Click the **Search** button to retrieve the time-series.
   The corresponding tsquery will appear, greyed out, in the **Advance** text box, and the chart(s) that display that time series will appear.

To select from the Metrics List:

1. Click the **Metrics List** to display a long list of metrics, organized in numerous categories.

2. Click a category to display the metrics related to that category. Note that some metrics will appear in multiple categories as they may apply to more than one entity or activity (for example, the Alerts metric applies to a large number of categories).

3. Click the metric that you want to chart to retrieve the time series.

4. Click the **Search** button to retrieve the time-series.
   The corresponding tsquery will appear, greyed out, in the **Advance** text box, and the chart(s) that display that time series will appear.

Advance search

In the **Advance** search, you can type any legal tsquery.

See the Tsquery Language for the language specification.

1. Type your query into the Advanced text box.

2. Click the **Search** button to retrieve the time-series.

The following chart shows the output for a single metric (Used Heap Memory), modified in the Advanced Search field to restrict it to a specific cluster, and displayed with a separate chart for each entity (each role instance on each host):
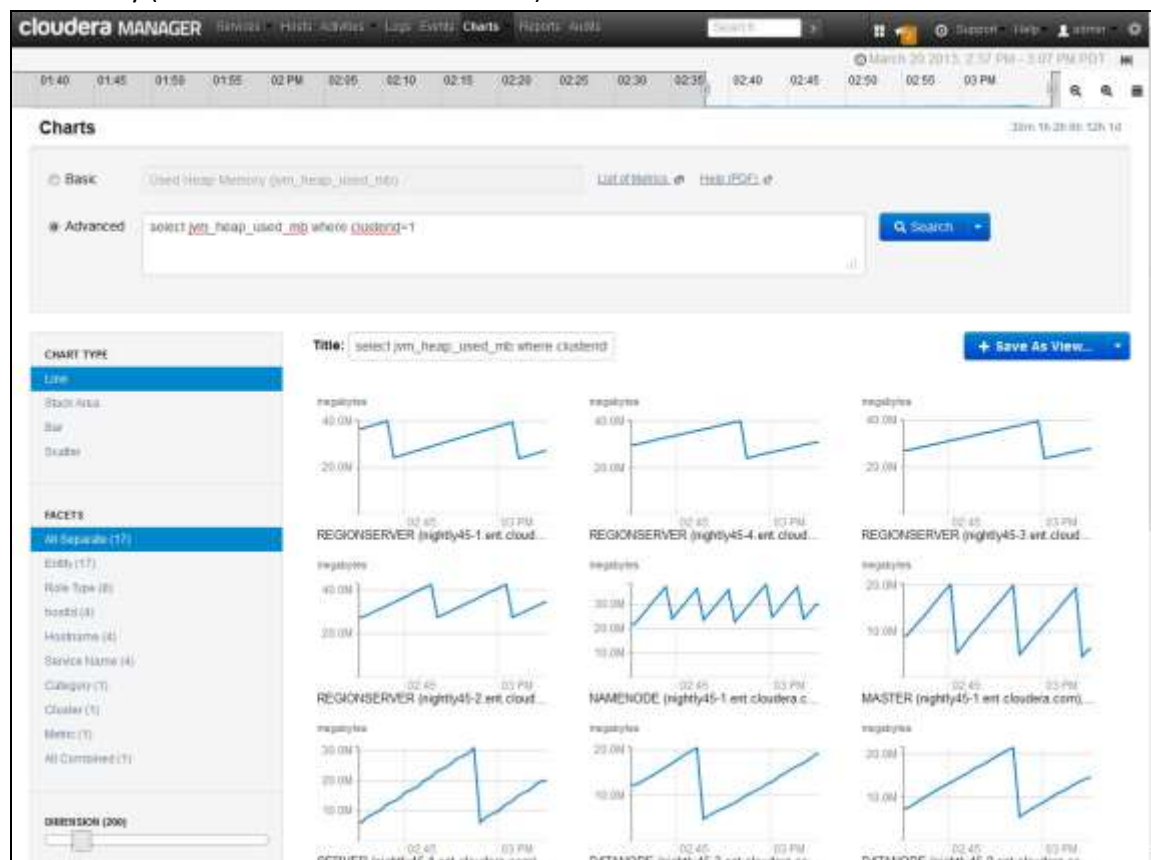
Editing Your Time-series Plot

The time-series data retrieved by the tsquery are displayed on different charts. By default, each time series is displayed on its own chart, using a **Line** style chart, a default size, and a default minimum and maximum for the y-axis.

- To change the chart-type, click one of the possible chart-types on the left: **Line**, **Stack Area**, and **Bar**.

Grouping (Faceting) Your Time-series

Every time series returned by the query has a set of attributes associated with it. In this case, each time series will have a hostname, role type, metric, and entity name attribute. The charting function can group (or *facet*) the time series into different numbers of charts by considering these attributes. By

default, all time series are plotted on their own chart (facets set to **All Separate**) as shown in the chart above. If **role type** is selected for the **Faceting** option for the above query, the time series will be grouped on eight charts, one chart each for each role type. The charts for role types with multiple instances will show multiple lines (e.g. DataNodes, RegionServers, TaskTrackers) while singleton roles (e.g. NameNode, JobTracker) will still show just a single line.
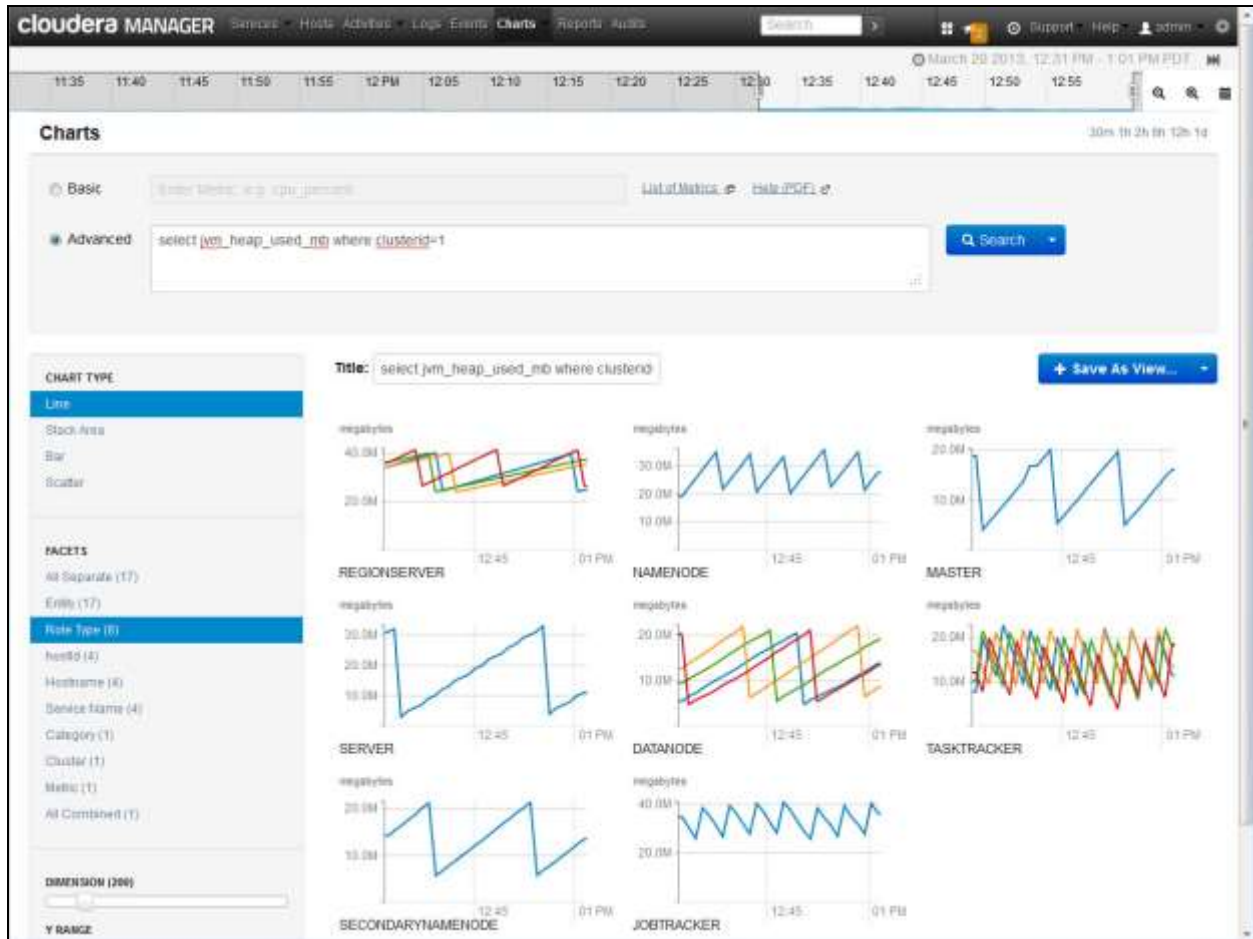
Changing Dimensions and Axes

You can change the size of your charts by moving the **DIMENSION** slider. It moves in 50-pixel increments. 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.

The following chart shows the same query as the previous charts, but with **All Combined** selected (which shows all time series in a single chart) and with the Dimension slider used to expand the chart.
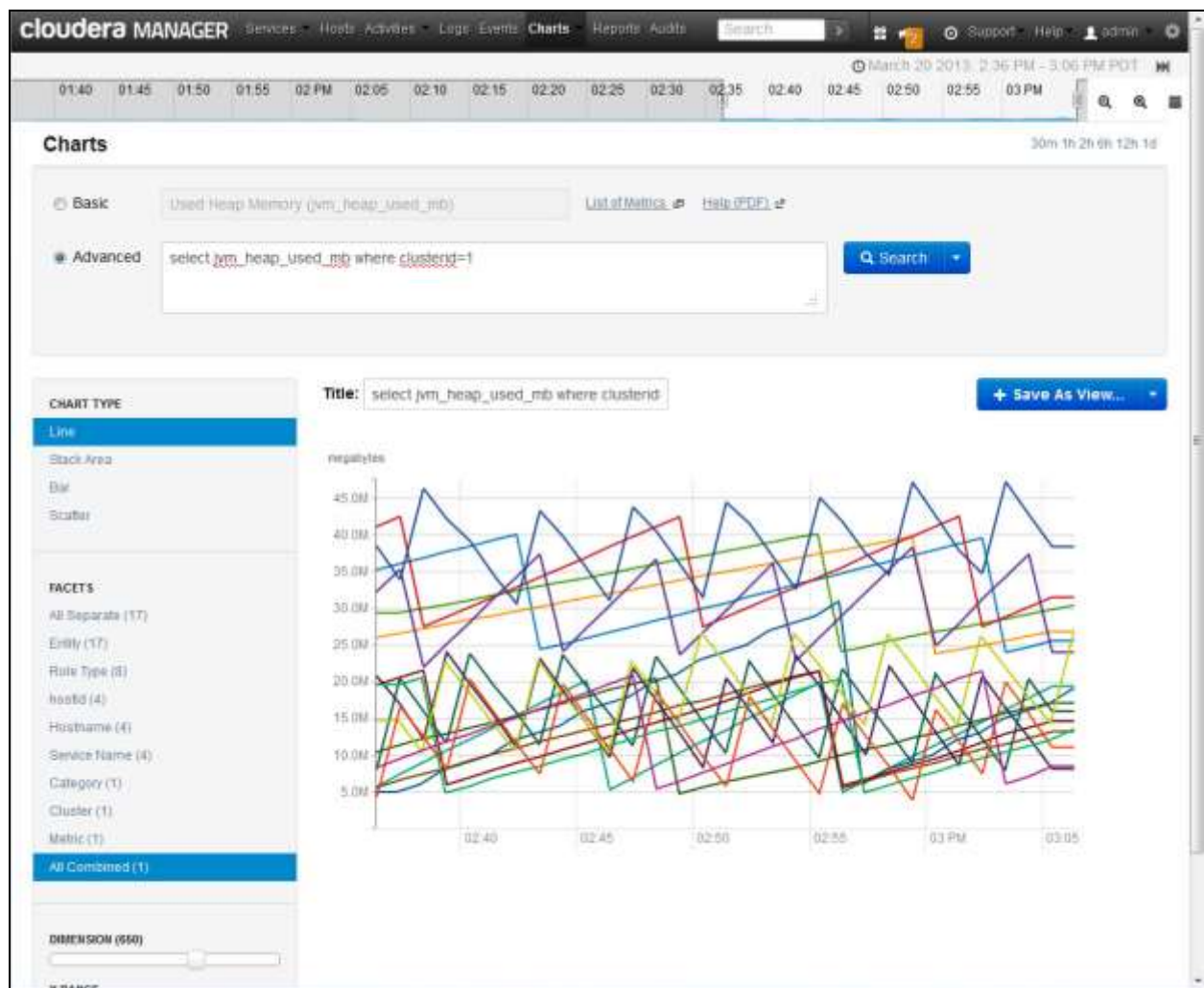
You can change the Y-axis range using the **Y RANGE** minimum and maximum fields.

The X-axis is based on clock time, and by default shows the last half hour of data. You can change the time range for your plot using the time range sets shown at the upper right of the window (right below the Time Range Selector) or by expanding or shrinking the Time Range Selector.

Saving a View

You can save the charts and their configurations (chart-type, dimension, and y-axis minimum and maximum) as a view. To save the plots as a **new** view, click the **Save as View** button, enter a view name, and then click **Create**. The new view will appear on the menu under the top-level **Charts** tab so that you can select it later. See Creating a Custom View for more information.

Managing Chart Views

The **Manage** option under the top level **Charts** menu lets you import, export, and remove views.

Export exports the specifications for the charts in the view as a json file.

Import reads an exported json file and recreates the charts based on the query.

Remove deletes the view.

Saving a view

You can save the charts and their configurations (chart-type, dimension, and y-axis minimum and maximum) as a view.

To save the charts as a *new* **view:**

1. Click the **+ Save as View** button

2. Enter a view name

3. Click **Create**

   The new view will appear on the menu under the top-level **Charts** tab so that you can select it later.

You can also **save your chart(s) to an existing view:**

1. Click the down arrow at the right of the **Save as View** button to pull down a menu of existing views.

2. Select the view name from the menu.

   Your chart will be added (appended) to the view you select.

Note that if your tsquery resulted in multiple charts, those charts will be saved as a unit (either to a new or existing view). You will not be able to edit the individual plots in that set of charts, but you will be able to edit the set as a whole. A single edit button will appear for the set that you saved — typically on the last chart in the set.

You **will** be able to edit a copy of the individual charts in the set, but the edited copy will not change the original chart in the view from which it was copied.

The Tsquery Language

General Structure

The tsquery language is a query language used to retrieve time-series data from the Cloudera Manager time-series data store. A tsquery has the following structure:

```
SELECT [metric expression] WHERE [predicate list]

examples:

(1) select * where roleType=datanode
Meaning: retrieve time-series for all metrics for all DATANODEs.

(2) select total_cpu_user where roleType=DATANODE
Meaning: retrieve the total_cpu_user metric time-series for all DATANODEs.

(3) select jvm_heap_used_mb / 1024, jvm_heap_committed_mb / 1024 where
category=ROLE and hostname="my host"
Meaning: 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".

(4) SELECT jvm_total_threads,jvm_blocked_threads
Meaning: retrieve the jvm_total_threads and jvm_blocked_threads metrics
time-series
for all entities which have these two metrics.
```

Each tsquery returns one or more time-series. The (2) tsquery example shown above returns one time-series for each DataNode. A time-series is a stream of metric data points for a specific entity. Each metric data point contains a timestamp and the value of that metric at that timestamp. See the section Entites and Predicates below for details on how modeled by Cloudera Manager.

Multiple tsqueries can be concatenated with semi-colons. The (3) example shown above can be written as:

```
select jvm_heap_used_mb / 1024 where category=ROLE and hostname=myhost;
select jvm_heap_commited_mb / 1024 where category=ROLE and hostname=myhost
```

Tsquery tokens are case insensitive: `Select`, `select` or `SeLeCt` are accepted for `SELECT`. This applies for all tsquery tokens. Tsquery attribute names and most attribute values are also case insensitive. The `displayName` and `entityName` attributes are two whose values are case sensitive.

The metric expression can be replaced with a `*` (asterisk) as shown in the (1) example above. In that case, all metrics that are applicable for selected entities, such as `DATANODEs` in the (1) example, will be returned.

Filter expressions can be omitted as shown in the (4) example above. In that case, time-series for all entities for which the metrics are appropriate will be returned. For this query you would see the jvm_new_threads metric for NameNodes, DataNodes, TaskTrackers, and so on.

The query `select *` is invalid. For any other query, a maximum of 250 time-series will be returned. This value can be configured in the SCM server settings.

Metric Expression

A metric expression is a comma-delimited list of one or more metric expression statements. A metric expression statement is either the name of a metric collected by Cloudera Manager or a scalar value. For example:

```
jvm_heap_used_mb, cpu_user, 5
```

See FAQ below to learn how to discover metrics collected by Cloudera Manager and use-cases for using scalar values in metric expressions.

Metric Expression Operators

Metrics expressions support the following binary operators:

- + (plus)

- - (minus)

- * (multiplication)

- / (division)

The following are examples of legal metric expressions:

```
total_cpu_user + total_cpu_system
1000 * (jvm_gc_time_ms / jvm_gc_count)
```

Metric Expression Functions

Metrics expressions support the following functions:

- dt: derivative with negative values. The change of the underlying metric expression, per second.

- dt0: derivative where negative values are skipped (useful for dealing with counter resets). The change of the underlying metric expression, per second.

So the following are all legal metric expressions:

```
dt(jvm_gc_count)
dt0(jvm_gc_time_ms) / 10
```

getHostFact

getHostFact(string factName, double defaultValue). Retrieves a fact about a host, for example:

```
select dt(total_cpu_user) / getHostFact(numCores, 2) where category=HOST
```

The example above will divide the results of `dt(total_cpu_user)` by the current number of cores for each host. If the number of cores cannot be determined the default will be used, '2'.

`getHostFact` currently supports one fact, `numCores`.

The chart below is an example of a query using these functions. The query computes the percentage of total user and system cpu usage each role is using on the host. It first computes the "CPU seconds per second" for the number of cores used by taking the derivative of the total user and system cpu times. It normalizes the result to the number of cores on the host by using the getHostFact function and multiply the result by 100 to get the percentage.
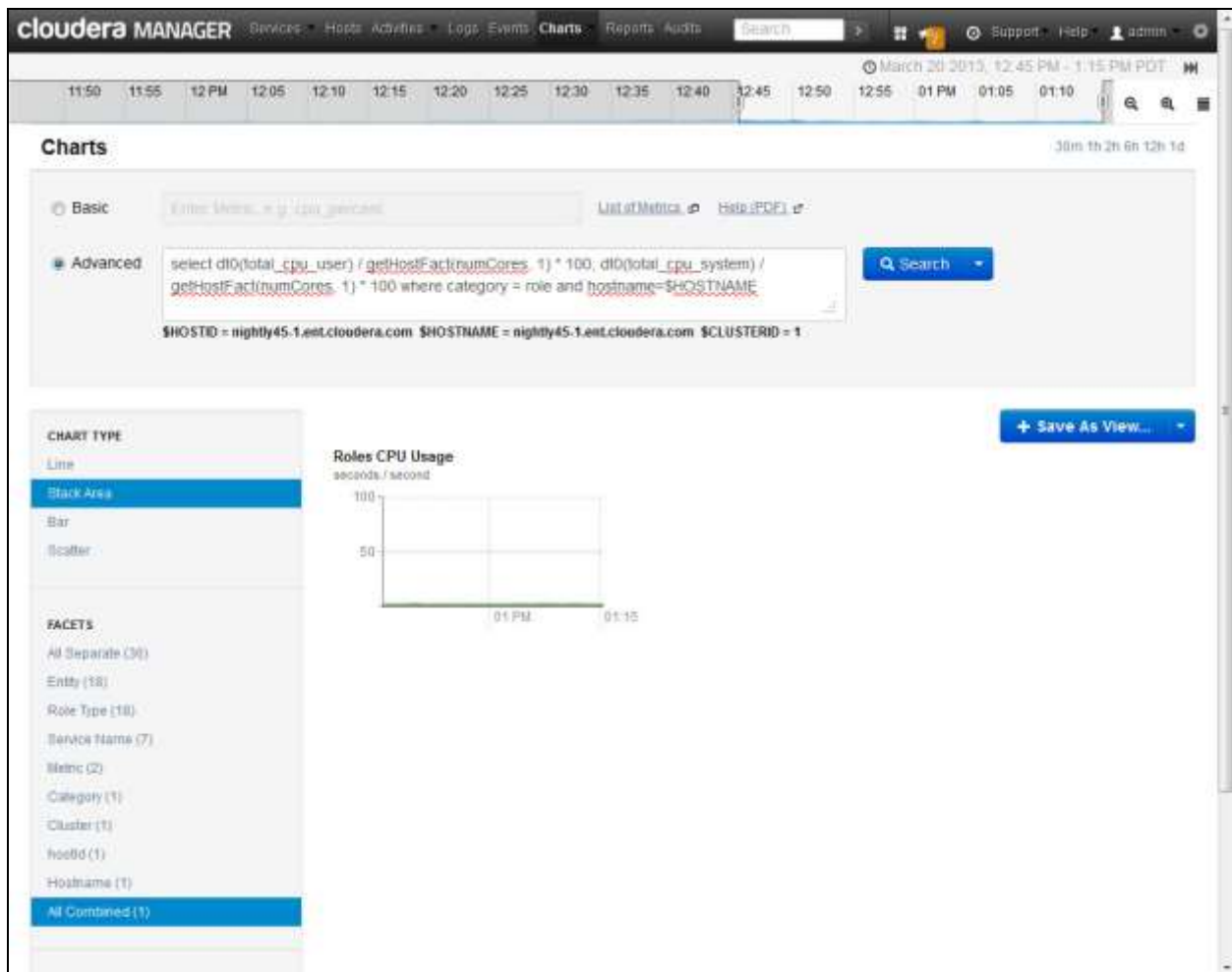
Predicate List

A predicate is a 'key operator value' pair in which `key` is a time-series attribute, `value` is a possible value, and `operator` can be either '=' or 'like'. AND and OR are logical operators that can be used to compose a complex predicate. For example:

```
(1) select * where roleType=DATANODE
Meaning: retrieve all time-series for all metrics for DATANODEs.

(2) select * where roleType=DATANODE or roleType=TASKTRACKER
Meaning: retrieve all time-series for all metrics for DATANODEs or
TASKTRACKERs.

(3) select * where (roleType=DATANODE or roleType=TASKTRACKER) and
hostname=myhost
Meaning: retrieve all time-series for all metrics for DATANODEs or
```

```
TASKTRACKERs that are running on host named myhost.

(4) select total_cpu_user where category=role and hostname like "host[0-
3]+.*"
Meaning: retrieve the 'total_cpu_metric' for all hosts with names that
match the regular expression "host[0-3]+.*".
```

The 'like' operator accepts only quoted `values`. `value` can be any regular expression as specified in regular expression constructs in the Java [Pattern](#) class documentation.

Here are some of the time-series attributes and their possible values.

Time Series Attribute Name	Possible values
roleType	NAMENODE, DATANODE, SECONDARYNAMENODE, JOURNALNODE, MASTER, REGIONSERVER, JOBTRACKER, TASKTRACKER, ACTIVITYMONITOR, SERVICEMONITOR, HOSTMONITOR, EVENTSERVER, ALERTPUBLISHER, REPORTSMANAGER, SERVER, AGENT, IMPALAD, STATESTORE
category	ROLE, DIRECTORY, HOST, FILESYSTEM, SERVICE, NETWORK_INTERFACE, DISK, CLUSTER, FLUME_SOURCE, FLUME_CHANNEL, FLUME_SINK
serviceType	HDFS, HBASE, MAPREDUCE, MGMT, ZOOKEEPER, FLUME, IMPALA
displayName	Use quoted strings to specify localized names or names that include spaces.
hostname	host name
hostId	The hostId is ithe canonical identifier for a host in Cloudera Manager. It must be unique and may not change over time. Often the hostname is used as the hostId.
rackId	rack id, e.g., '/default'
clusterId	The cluster id. To specify a cluster by its name, use filter `'where category=CLUSTER and displayName="[the display name]"'`
serviceName	The service id. To specify a service by its name use filter `'category=SERVICE and displayName="[name]"'`

Time Series Attribute Name	Possible values
device	Disk device name, e.g., 'sda'
partition	Partition name, e.g., 'sda1'
mountpoint	mount point name, .e.g., '/var', '/mnt/homes'
iface	network interface name, e.g., 'eth0'.
componentName	flume component name, .e.g., 'channel1', 'sink1'

The category attribute

The `category` attribute controls the type of the entities returned by the query. Some metrics are collected for more than one type of entities. For example, `total_cpu_user` is collected for entities of type `HOST` and for entities of type `ROLE`. To retrieve the data for all hosts in your deployment use

```
select total_cpu_user where category=HOST
```

The `ROLE` category applies to all role types (see `roleType` attribute above). The `SERVICE` category applies to all service types (see `serviceType` attribute above). For example, to retrieve the committed heap for all roles on host1 use

```
select jvm_committed_heap_mb where category=ROLE and hostname="host1"
```

FAQ

How do I compare all disk io for all the datanodes that belong to a specific hdfs service?

```
select bytes_read, bytes_written where roleType=datanode and
serviceName=hdfs1
```

Replace 'hdfs1' with the appropriate service name. You can then facet by "Metric" and compare all datanodes byte_reads and byte_writes metric at once. See the Charting Time-Series Data page for more details about faceting.

When would I use a derivative function?

Some metrics represent a counter, e.g., 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 (e.g., process restarted, host restarted, etc.). Use `dt0` for these metrics.

How do I display a threshold on a chart?

Assume 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 to retrieve the metrics and the threshold:

```
select service_time, await_time, await_read_time, await_write_time, 50
where category=disk
```

You can then facet the results to be all in one chart. The scalar threshold '50' will also be rendered on the chart. See the Charting Time-Series Data page for more details about faceting.

I get "The query hit the maximum results limit" warning. 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 try to reduce the number of metrics you are trying to retrieve. You can also use the `like` operator to limit the query to a subset of entities. For example, instead of

```
select service_time, await_time, await_read_time, await_write_time, 50
where category=disk
```

you can use

```
select service_time, await_time, await_read_time, await_write_time, 50
where category=disk and hostname like "host1[0-9]?.cloudera.com"
```

The latter query will retrieve the disks for only ten hosts.

How do I discover which metrics are available for which entities?

One way to discover which metrics are collected by Cloudera Manager is to look at the `List of Metrics` (on the Charts page next to the metric-type-ahead box) or use the metric-type-ahead box. Another way is to retrieve all metrics for the type of entity you are interested in:

```
select * where roleType=datanode and hostname=host1
```

Metric Aggregation

Overview

It is often useful to see an aggregated view of the activity on a cluster. For example, one might want to see the average number of bytes read per datanode, or they might want to see the maximum number of bytes read by any datanode. To make this easy we pre-aggregate many of these metrics and allow users to access them through our charts.

What We Aggregate

We aggregate metrics based on the category of the entity the generated them. The categories map to components in the system such as hosts, disks, regionservers, and HDFS services.

Metrics are aggregated from their generating entity to larger entities they are a part of. For example, metrics that are generated by disks, network interfaces, and filesystems are aggregated to their respective hosts and clusters. Generally, this hierarchy is defined as follows:

Disks, Network Interfaces, Filesystems -> Hosts, Clusters
Hosts -> Clusters
Roles -> Services, Clusters
HTables -> HBase Services, Clusters
Agents -> Hosts, Clusters

Aggregation Types

Each aggregate is calculated every minute and takes into account all the metrics logged over the previous minute. For example, the metric cpu_percent_host_max takes into account all cpu_percent metrics logged by all hosts in a cluster in the previous minute.

We support five types of aggregation:
1. Max - the largest value for any entity
2. Min - the smallest value for any entity
3. Average - the average value for all entities

Charting Time-series Data

4. Standard deviation - the standard deviation of the values for all entities

5. Sum - the sum total of the value for all entities

Sample Use Cases

Use Case 1:

Compare the maximum, minimum and average CPU usage across a cluster:

1. Go the the Charts tab and select search.

2. Click on advanced and enter the tsquery:

"SELECT cpu_percent_host_max, cpu_percent_host_min, cpu_percent_host_avg"

3. Click search. You should see three charts, each with cpu data.

4. Click on Facets -> All Combined in the left column. Now you should see all the data on one chart.

Use Case 2:

Compare the cpu usage of a single host to the max, min, and average for the cluster

1. Copy the instructions from above, except in #2 enter the following query instead:

"SELECT cpu_percent_host_max, cpu_percent_host_min, cpu_percent_host_avg, cpu_percent where category=cluster or hostname='MYHOST.COM'"

Aggregate Metric Names

To access aggregated metrics it helps to know how they are named. There are three components to the name:

1. The metric we're aggregating - for example "cpu_percent" or "jvm_gc_count"

2. The category of the entity generating the metric - for example "host" or "regionserver"

3. The aggregation type - for example "max" or "avg"

These parts are combined to form a full name such as "cpu_percent_host_max"

The naming of the final component, aggregation type, varies by the type of the metric. We support three types of metrics:

1. Gauges

These are metrics that can go up and down, like cpu_percent. Gauges have a straightforward naming convention:

max -> "max"

min -> "min"

average -> "avg"

standard deviation -> "std_dev"

sum -> "sum"

2. Weighted Gauges

These are probably best explained with an example. Consider the HBase regionserver metric put_avg_time. This metric tracks the average put time for each regionserver. Now consider the case where you have two regionservers, one that did 10,000 puts with an average time of one millisecond

per put, and another that did 10 puts with an average time of one second per put. In this case if you just averaged the two averages, you would get that that the average across the whole service was about half a second, but that doesn't accurately reflect reality. Instead if you calculated the average by weighting by the number of puts by the counter per regionserver you would get a more accurate number:

Total puts = 10,000 + 10 = 10,010 puts

Total time = (10000 * 1ms) + (10 * 1000ms) = 20,000 ms

Average time = (20,000ms) / (10,010 puts) = ~2 ms

To reflect this we have the concept of weighted gauges which perform this calculation. Their aggregates are named as follows:

max -> "max"

min -> "min"

average -> "weighted_avg"

standard deviation -> "weighted_std_dev"

Sum aggregations are a special case. They represent the weighted total, which would be 20,000 ms in our example and are named accordingly - "put_time_regionserver_sum". Note that we remove "avg" from the name of the metric.

3. Counters

These are metrics that track the total count since a process / host started. An example of a counter is jvm_gc_count which tracks the number of java garbage collections since a java process started. Since users are more interested in the rate of change of counters (ie how many garbage collections were there per second over the last five minutes) rather than their raw value we calculate the aggregates in terms of rate. They are named as follows:

max -> "max_rate"

min -> "min_rate"

average -> "avg_rate"

standard deviation -> "std_dev_rate"

Again like in the weighted gauges case, sum aggregations are a special case. For counters they represent the total number of times an event occurred and are NOT a rate. In this case we append the word "sum" to the end of name just like we would for gauge metrics - "jvm_gc_count_regionserver_sum"

Viewing Reports

The **Reports** tab 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.

The **Search files and Manage Directories** button takes you to a File Browser where you can search files, manage directories, and set quotas.

If you are managing multiple clusters, or have multiple NameServices configured (if High Availability and/or Federation is configured) there will be separate reports for each cluster and NameService.

To view reports of HDFS usage or MapReduce activity:

- Click the **Reports** tab in the Cloudera Manager navigation bar.

Disk Usage Reports

The following reports show HDFS disk usage statistics, either current or historical, by user, group or directory.

Note that 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.

Current Disk Usage: by User, by Group, or by 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.

To create one of these reports:

- Click the report name (link) to produce the resulting report.

Each of these reports show:

Bytes:	The logical number of bytes in the files, aggregated by user, group, or directory. This is based on the actual files sizes, not taking replication into account.
Raw Bytes	The physical number of bytes (total disk space in HDFS) used by the files aggregated by user, group, or directory. This does include replication, and so is actually Bytes times the number of replicas.
File and Directories Count	The number of files aggregated by user, group or directory.

Note that 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 Search Files and Manage Directories).

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.

Historical Disk Usage by User, by Group, or by 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 Search Files and Manage Directories).

Downloading Reports as XLS or CVS files

Any report can be downloaded to your local system as an XLS file (Microsoft Excel 97-2003 worksheet) or CSV (Comma-Separated Values) text file.

To download a report:

- 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 *or*

- From any report page, click the **Download CSV** or **Download XLS** buttons.

Either of these opens the Open file dialog where you can open or save the file locally.

Activities Reports

The following report shows metrics on the job activity in your cluster.

MapReduce Usage by User

This produces a tabular report that you can use to view aggregate job activity per hour, day, week, month, or year.

- In the Report Period field, select the period over which you want the metrics aggregated. Default is Daily.

- Click Generate Report to produce the resulting report.

For weekly reports, the Date indicates the year and week number (e.g. 2011-01 through 2011-52). For monthly reports, the Date indicates the year and month by number (2011-01 through 2011-12).

The Activity data in these reports comes from the Activity Monitor; they can include all the data currently in the Activity Monitor database. Note that Activity Monitor data will eventually expire, based on the configuration you have set for Activity Monitor.

Search Files and Manage Directories

The **Search Files and Manage Directories** page provides a file browser function that lets you browse the HDFS namespace, as well as letting you manage your files and directories.

File and Disk Space Quotas

You can set quotas for file count and disk space per directory:

1. From the Reports page, click **Search Files and Manage Directories** for the namespace you want to manage.

2. Browse the file system to find the directory for which you want to set quotas.

3. Click the **Manage Quota** button at the right of the row for the directory you want.
   A **Manage Quota** pop-up appears, where you can set file count or disk space limits for the directory you have selected.

4. When you have set the limits you want, click **OK**

Searching within the File System

The **Search Files and Manage Directories** page lets you search the file system using predefined search criteria. The default is **Search by Name** but you can also search for large files, directories with quotas, watched directories, or custom search criteria which you can construct using criteria such as filename, owner, file size and so on.

To search the file system:

1. From the Reports page, click **Search Files and Manage Directories** for the namespace you want to search.

2. In the **Search** menu at the top right of the page, select the type of query you want to use. 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 **Search by Name** will show a field where you can enter the name.
   Selecting **Large Files** will provide fields where you provide size to be used as the search criteria.
   If you select **Custom...** and enter multiple criteria, all of them must be met for a file to be considered a match.

3. Click the Search button () to execute the search.

If you search within a directory, only files within that directory will be found, so if you're browsing `/user` and do a search, you might find `/user/foo/file`, but you will not find `/bar/baz`.

Watched Directories

Search Files and Manage Directories lets you designate the HDFS directories that you want watch for inclusion in the Directory-based usage reports (**Current Disk Usage By Directory** and **Historical Disk Usage By Directory**).

To add or remove directories from the Directory-based usage reports:

1. From the main page of the **Report** tab, click the **Search Files and Manage Directories** button at the upper right of the page *or*
   From either of the Directory-based reports, click the **Search Files and Manage Directories** button.

2. Click the Star icon () at the left of the directories you want to include. The icon changes to the activated form (). You can navigate through the file system to see the directory you want to add — you can include a directory at any level without needing to include its parent.

3. To remove a directory from the list, just deactivate the Star icon.

HDFS Directory Access Permission Reports

Cloudera Navigator provides the ability to download a report that details the HDFS directories a group has permission to access. See Downloading HDFS Directory Access Permission Reports.

Administration

Click the gear icon to display the **Administration** page where you can configure settings that affect how Cloudera Manager interacts with your cluster.

Properties Tab

On the **Properties** tab, you can set:

- **Performance**: Set the Cloudera Manager Agent heartbeat.

- **Thresholds**: Set Health status parameters.
  For configuration instructions, see see [Configuring Agent Heartbeat and Health Status Options](#).

- **Security**: Set TLS encryption settings to enable TLS encryption between the Cloudera Manager Server, Agents, and clients. For configuration instructions, see [Configuring TLS Security for Cloudera Manager](#)
  You can also:

  - Set the realm for Kerberos security and point to a custom keytab retrieval script. For configuration instructions, see [Configuring Hadoop Security with Cloudera Manager](#)

  - Specify session timeout and a "Remember Me" option.

- **Ports and Addresses**: Set ports for the Cloudera Manager Admin Console and Server. For configuration instructions, see [Configuring the Ports for the Admin Console and Agents](#).

- **Other**: To enable Cloudera usage data collection For configuration instructions, see [Configuring Anonymous Usage Data Collection](#).
  You can also:

  - Set a custom header color and banner text for the Admin console.

  - Set an "Information Assurance Policy" statement – this statement will be presented to every user before they are allowed to access the login dialog. The user must click "I Agree" in order to proceed to the login dialog.

  - Disable/enable the auto-search for the Events panel at the bottom of a page.

- **Enterprise Support**: Enable access to online Help files from the Clouders web site rather than from locally-installed files. (see [Opening the Help Files from the Cloudera Web Site](#)), and enable automatic sending of diagnostic data to Cloudera when you trigger a data collection (see [Sending Diagnostic Data to Cloudera

- **External Authentication**: Specify the configuration to use LDAP, Active Directory, or an external program for authentication. See [Configuring External Authentication](#) for instructions.

- **Advanced**: Enable API debugging.

Import Tab

You can import Cloudera Manager configuration settings to transfer the settings. For instructions, see Importing Cloudera Manager Settings.

Alerts Tab

This tab provides a summary of the settings for alerts in your clusters. From this tab you can view the alerts by alert type (Health, Log, or Activity alerts) and by service within those categories. You can also see the email addresses configured as recipients for alerts, and send test messages. See Alert Settings for more information.

Users Tab

Cloudera Manager user accounts allow users to log into the Cloudera Manager Admin Console. For configuration instructions, see Cloudera Manager User Accounts.

Kerberos Tab

After enabling and configuring Hadoop security using Kerberos on your cluster, you can view and regenerate the Kerberos principals for your cluster. If you make a global configuration change in your cluster, such as changing the encryption type, you would use the Kerberos tab to regenerate the principals for your cluster.

> **Important**
>
> Do not regenerate the principals for your cluster unless you have made a global configuration change. Before regenerating, be sure to read the *Configuring Hadoop Security with Cloudera Manager* to avoid making your existing host keytabs invalid.

Server Log Tab

To help you troubleshoot problems, you can view the Cloudera Manager Server log. For more information, see Viewing the Cloudera Manager Server and Agent Logs.

License Tab

The **License** tab indicates the status of your license (for example, whether you license is currently valid) and shows you the owner, the license key, and expiration date of the license. It also shows any Add-Ons enabled by your license, such as Impala Monitoring.

You can enter a new license (for example, to enable additional add-ons) by browsing to the new license file and uploading the license.

Language Tab

You can change the language of the Cloudera Manager Admin Console User Interface through the language preference in your browser. Information on how to do this for the browsers supported by Cloudera Manager is shown under the Language tab. You can also change the language for the information provided with activity and health events, and for alert email messages.

To change the language of the activity and health event information and alert email messages, select the language you want from the drop-down list on this page, the click **Save Changes**.

Configuring External Authentication

Cloudera Manager provides several different mechanisms for authenticating users for Cloudera Manager. You can enter users into Cloudera Manager's own database (the default) or configure Cloudera Manager to authenticate against an external authentication service — an LDAP server (Active Directory or an OpenLDAP compatible directory) or you can specify another external service.

Further, you can configure Cloudera Manager so that it can use both methods of authentication (internal database vs. external service), and you can determine the order in which it performs these searches. You can also restrict login access to members of specific groups, and can specify groups whose members will automatically be given administrator access to Cloudera Manager.

For an OpenLDAP compatible directory, you have several options for searching for users and groups:

- You can specify a single base Distinguished Name (DN) and then provide a "Distinguished Name Pattern" to use to match a specific user in the LDAP directory.

- Search filter options let you search for a particular user based on somewhat broader search criteria – for example Cloudera Manager users could be members of different groups or organizational units (OUs), so a single pattern won't find all those users. Search filter options also let you find all the groups to which a user belongs, to help determine if that user should have login or admin access.

To configure an external authentication service for Cloudera Manager user authentication:

1. Click the gear icon to display the **Administration** page.
2. Click the **Properties** tab.
3. In the left-hand column, select the **External Authentication** category.
4. Select the order in which Cloudera Manager should attempt its authentication (**Authentication Backend Order**). Here you can choose to authenticate users using just one of the methods (using Cloudera Manager's own Database is the default), or you can set it so that if the user cannot be authenticated by the first method, it will attempt using the second method.

    Note that if you select **External Only**, users who are administrators in the Cloudera Manager database will still be able to log in with their database password. This is to prevent the system

from locking everyone out if the authentication settings get misconfigured — such as with a bad LDAP URL.

5. Go to the section below for the type of authentication you want to configure, and follow the steps to set the properties appropriately.

Configure User Authentication Using Active Directory

1. For **External Authentication Type** select Active Directory.

2. Provide the URL of the Active Directory server.

3. Provide the NT domain to authenticate against.

4. Optionally, provide a comma-separated list of LDAP group names in the **LDAP User Groups** property. If this list is provided, only users who are members of one or more of the groups in the list will be allowed to log into Cloudera Manager. If this property is left empty, *all* authenticated LDAP users will be able to log into Cloudera Manager.

    For example, if there is a group called "CN=ClouderaManagerUsers,OU=Groups,DC=corp,DC=com", add the group name `ClouderaManagerUsers` to the **LDAP User Groups** list to allow members of that group to log in to Cloudera Manager. The group names are case-sensitive.

5. In the **LDAP Administrator Groups** property you can provide a list of groups whose members should be given administrator access when they log in to Cloudera Manager. (Note that admin users must also be a member of at least one of the groups specified in the **LDAP User Groups** property or they will not be allowed to log in.) If this is left empty, then no users will be granted administrator access automatically at login — administrator access will need to be granted manually by another administrator.

Configure User Authentication Using an OpenLDAP-compatible Server

1. For **External Authentication Type** select **LDAP**.

2. Provide the URL of the LDAP server and (optionally) the base Distinguished Name (DN) (the search base) as part of the URL — for example `ldap://ldap-server.corp.com/dc=corp,dc=com`.

3. **If your server does NOT allow anonymous binding:**
    Provide the user DN and password to be used to bind to the directory. These are the **LDAP Bind User Distinguished Name** and **LDAP Bind Password** properties. By default, Cloudera Manager assumes anonymous binding.

4. To use a single "Distinguished Name Pattern," provide a pattern in the **LDAP Distinguished Name Pattern** property.

Use {0} in the pattern to indicate where the username should go. For example, to search for a distinguished name where the the uid attribute is the username, you might provide a pattern similar to `uid={0},ou=People,dc=corp,dc=com`. Cloudera Manager substitutes the name provided at login into this pattern and performs a search for that specific user. So if a user provides the username "foo" at the Cloudera Manager login page, Cloudera Manager will search for the DN `uid=foo,ou=People,dc=corp,dc=com`.

Note that if you provided a base DN along with the URL, the pattern only needs to specify the rest of the DN pattern. For example, if the URL you provide is `ldap://ldap-server.corp.com/dc=corp,dc=com`, and the pattern is `uid={0},ou=People`, then the search DN will be `uid=foo,ou=People,dc=corp,dc=com`.

5. You can also search using User and/or Group search filters, using the **LDAP User Search Base**, **LDAP User Search Filter**, **LDAP Group Search Base** and **LDAP Group Search Filter** settings. These allow you to combine a base DN with a search filter to allow a greater range of search targets.

For example, if you want to authenticate users who may be in one of multiple OUs, the search filter mechanism will allow this. You can specify the User Search Base DN as `dc=corp,dc=com` and the user search filter as `uid={0}`. Then Cloudera Manager will search for the user anywhere in the tree starting from the Base DN. Suppose you have two OUs — ou=Engineering and ou=Operations — Cloudera Manager will find User "foo" if it exists in either of these OUs, i.e. `uid=foo,ou=Engineering,dc=corp,dc=com` or `uid=foo,ou=Operations,dc=corp,dc=com`.

You can use a user search filter along with a DN pattern, so that the search filter provides a fallback if the DN pattern search fails.

The Groups filters let you search to determine if a DN or user name is a member of a target group. In this case, the filter you provide can be something like `member={0}` where {0} will be replaced with the **DN** of the user you are authenticating. For a filter requiring the user name, {1} may be used, as `memberUid={1}`. This will return a list of groups this user belongs to, which will be compared to the list in the **LDAP User Groups** and **LDAP Administrator Groups** properties (discussed [previously](#) in the section about Active Directory).

Configure User Authentication Using an External Program

You can configure Cloudera Manager to use an external authentication program of your own choosing. Typically, this may be a custom script that interacts with a custom authentication service. Cloudera Manager will call the external program with the user name as the first command line argument. The password is passed over `stdin`. Cloudera Manager assumes the program will return the following exit codes:

- 0 for the successful authentication of a regular user

- 1 for the successful authentication of an admin user

- a negative value for failure to authenticate.

1. For **External Authentication Type** select **External Program**.

2. Provide a path to the external program in the **External Authentication Program Path** property.

Configuring the Ports for the Admin Console and Agents

You can configure the HTTP and HTTPS ports you want to use for the Cloudera Manager Admin Console and Agents.

To configure the ports for the Cloudera Manager Admin Console and Agents:

1. Click the gear icon ⚙ to display the **Administration** page.

2. On the **Properties** tab, under the **Ports and Addresses** category, set the following options as described below:

Setting	Description
HTTP Port for Admin Console	Specify the HTTP port to use to access the Server via the Admin Console.
HTTPS Port for Admin Console	Specify the HTTPS port to use to access the Server via the Admin Console.
Agent Port to connect to Server	Specify the port for Agents to use to connect to the Server.

3. Click **Save Changes**.

4. Restart the Cloudera Manager Server by typing the following command on the Cloudera Manager Server host:

```
$ sudo service cloudera-scm-server restart
```

Configuring Anonymous Usage Data Collection

You can configure Cloudera Manager to send anonymous usage information using Google Analytics to Cloudera. The information helps Cloudera improve Cloudera Manager.

To configure anonymous usage data collection:

1. Click the gear icon ⚙ to display the **Administration** page.

2. On the **Properties** tab, under the **Other** category, set the **Allow Usage Data Collection** option to enable or disable anonymous usage data collection.

3. Click **Save Changes**.

Importing Cloudera Manager Settings

Backing up your Current Deployment

To back up your current deployment, please see the section on backing up your database in the Cloudera Manager documentation. The import feature should not be relied on for backup and recovery at this time.

Building a Cloudera Manager Deployment

You can use the Cloudera Manager API to programmatically build a Cloudera Manager Deployment — a definition of all the entities in your Cloudera Manager-managed deployment — clusters, service, roles, hosts, users and so on. See the RESTful API documentation on how to manage deployments using the `/cm/deployment` resource.

Uploading a Cloudera Manager 4.0 Configuration Script

Note: As of Cloudera Manager 4.1, the import of configuration settings through the Cloudera Manager Admin Console UI has been deprecated. If you have exported a configuration using the **Export** tab in an older version of Cloudera Manager, you can still import it following the instructions below. However, going forward, importing a deployment should be done using the Cloudera Manager API. See the documentation for `/cm/deployment` for details.

Important: You must import the configuration settings on a clean cluster that does not have existing hosts or services.

Important: When you first installed the Cloudera Manager Server, you set up a database to store the Cloudera Manager service configuration information (see Installing and Configuring Databases). That database also stores the Cloudera Manager license information. If the original database is lost (for example, the database was deleted and you recreated a new one), you must first upload your license on the **Administration > License** tab and restart the Cloudera Manager Server before importing the configuration settings. If you don't upload your license first to store the license information in the new database, the import will fail.

To import the configuration script into Cloudera Manager:

1. On every Cloudera Manager Agent host, run this command to stop the Cloudera Manager Agent:

```
$ sudo service cloudera-scm-agent stop
```

2. Delete all services on the **Services** tab by choosing **Delete** from the **Actions** menu next to each service instance.

3. Delete all hosts on the **Hosts** tab by clicking the check box at the top of list of hosts, and then click **Delete**.

4. Copy the configuration script file that you downloaded during export to the host with the new Cloudera Manager server.

5. Click the gear icon to display the **Administration** page.

6. Click the **Import** tab.

7. Click **Browse**, navigate to the file location, and click **Open**.

8. Click **Import**.

9. On every Cloudera Manager Agent host, run this command to start the Cloudera Manager Agent:

```
$ sudo service cloudera-scm-agent start
```

Opening the Help Files from the Cloudera Web Site

By default, when you click the Help link in Cloudera Manager, the locally-installed Help files are opened. These local Help files are not updated after installation. You can configure Cloudera Manager to open the latest Help files from the Cloudera web site (this option requires Internet access from the browser).

To open the Help files from the Cloudera web site:

1. Click the gear icon to display the **Administration** page.

2. On the **Properties** tab, under the **Enterprise Support** category, enable the **Open latest Help files from the Cloudera website**.

3. Click **Save Changes**.

Maintenance

There may be situations where you need to temporarily stop the Cloudera Manager server or Cloudera Manager agents on selected nodes, for example, in order to perform maintenance on a host. The following topics cover stopping or restarting the Cloudera Manager server or its agents.

- [Maintenance Mode](#)
- [Manually Failing Over Your Cluster](#)
- [Stopping and Restarting the Cloudera Manager Server](#)
- [Stopping or Restarting Cloudera Manager Agents](#)

Maintenance Mode

Maintenance mode allows you to suppress alerts for a host, service, role, or even the entire cluster. This can be useful when you need to take actions in your cluster (make configuration changes and restart various elements) and do not want/need to see the alerts that will be generated due to those actions.

Putting a component into maintenance mode does not prevent events from being logged; it only suppresses the alerts that those events would otherwise generate. As a result, you can still see a history of the events that were recorded due to your actions.

You can enable maintenance mode for a service, a role, a host, or the entire cluster.

You can view the status of Maintenance Mode in your cluster with the **View Maintenance Mode Status** button from the
All Services page. This button appears for each cluster, and separately for the Cloudera Management Services.

When you enter maintenance mode on a component (cluster, service, or host) that has subordinate components (for example, the roles for a service) the subordinate components are also put into maintenance mode. These are considered to be in "effective" maintenance mode, as they have inherited the setting from the higher-level component.

For example:

- If you set the HBase service into maintenance mode, then its roles (HBase Master and all Region Servers) are put into effective maintenance mode.
- If you set a host into maintenance mode, then any roles running on that host are put into effective maintenance mode.

Components that have been explicitly put into maintenance mode show the following icon ( ). Components that have entered effective maintenance mode as a result of inheritance from a higher-level component show a similar icon, but it is grey and yellow instead of black and red ( ).

Enabling Maintenance Mode

To put the entire cluster into Maintenance Mode:

1. Go to the main **Services** page (**All Services**)

2. From the **Actions** menu, select **Enter Maintenance Mode...**

3. Confirm that you want to do this.

The cluster is put into explicit maintenance mode, as indicated by the black maintenance mode icon. All services and roles in the cluster are entered into effective maintenance mode, as indicated by the grey/yellow maintenance mode icon.

To put a service into maintenance mode:

1. From the **Actions** menu for the individual service **Enter Maintenance Mode...**

2. Confirm that you want to do this.

The service is put into explicit maintenance mode, as indicated by the black maintenance mode icon. All roles for the service are entered into effective maintenance mode, as indicated by the grey/yellow maintenance mode icon.

To put one or more individual roles into maintenance mode:

1. Go to the **Services** page that includes the role

2. Go to the **Instances** tab

3. Select the role(s) you want to put into maintenance mode

4. From the **Actions for Selected** menu, select **Enter Maintenance Mode...**

5. Confirm that you want to do this.

The role will be put in explicit maintenance mode.

If this role instance was already in effective maintenance mode (because its service or host was put into maintenance mode) the role will now be in effective maintenance mode. This means that it will NOT exit maintenance mode automatically if its host or service exits maintenance mode. It will need to be removed from maintenance mode explicitly.

To put one or more hosts into maintenance mode:

1. Go to the **Hosts** page

2. Select the host(s) you want to put into maintenance mode

3. From the **Actions for Selected** menu, select **Enter Maintenance Mode**

4. Confirm that you want to do this.

The confirmation popup lists the role instances that will be put into effective maintenance mode when the host goes into maintenance mode.

Maintenance

Interaction between Explicit and Effective Maintenance Mode

When a component (role, host or service) is in effective maintenance mode, it can only be removed from maintenance mode when the higher-level component exits maintenance mode. For example, if you put a service into maintenance mode, then the roles associated with that service will be entered into effective maintenance mode, and will remain in effective maintenance mode until the service exits maintenance mode. You cannot remove them from maintenance mode individually.

On the other hand, a component that is in effective maintenance mode can be put into explicit maintenance mode – just select the individual component and use the **Enter Maintenance Mode** command. In this case, the component will remain in maintenance mode even when the higher-level component exits maintenance mode.

For example, suppose you put a host into maintenance mode, (which puts all the roles on that host into effective maintenance mode). You then select one of the roles on that host and put it explicitly into maintenance mode. When you have the host exit maintenance mode, that one role will remain in maintenance mode. You will need to select it individually and specifically have it exit maintenance mode.

Manually Failing Over Your Cluster

If you are running a HDFS service with High Availability enabled, you can manually cause your Active NameNode to failover to your Standby NameNode. This is useful for planned downtime — for hardware changes, configuration changes, or software upgrades of your primary host.

To perform a manual failover:

1. From the **Services** tab, select your HDFS service.

2. Click the **Instances** tab.

3. Click **Manual Failover...**

4. From the pop-up, select the NameNode that should be made active, then click **Manual Failover**.

> **Note: For advanced use only:** You can set the **Force Failover** checkbox to force the selected NameNode to be active, irrespective of its state or the other NameNode's state. Forcing a failover will first attempt to failover the selected NameNode to active mode and the other NameNode to standby mode. It will do so even if the selected NameNode is in safe mode. If this fails, it will proceed to transition the selected NameNode to active mode. To avoid having two NameNodes be active, use this only if the other NameNode is either definitely stopped, or can be transitioned to standby mode by the first failover step.

5. When all the steps have been completed, click **Finish**.

The Cloudera Manager will transition the NameNode you selected to be the Active NameNode, and the other NameNode to be the Standby NameNode. HDFS should never have two Active NameNodes.

Note: If you are using a NFS-mounted shared edits directory, a fencing method must be configured in order for failover (either automatic or manual) to function — Cloudera Manager configures this automatically. See Fencing Methods if you want more information about where this is set through Cloudera Manager.

For details of the fencing methods supplied with CDH4, and how fencing is configured, see the Fencing Configuration section in the CDH4 High Availability Guide.

Stopping and Restarting the Cloudera Manager Server

You can stop the Cloudera Manager server (for example, to perform maintenance on its host) without affecting the other services running on your cluster. Statistics data used by Activity Monitoring and Service Monitoring will continue to be collected during the time the server is down.

To stop the Cloudera Manager server without affecting other services:

```
service cloudera-scm-server stop
```

To restart it:

```
service cloudera-scm-server start
```

Note

If you are intending to perform an upgrade of Cloudera Manager, then you should stop the management service (through the Admin Console) prior to stopping the server.

Stopping or Restarting Cloudera Manager Agents

Usually (during an upgrade to a new version of Cloudera Manager, for example) you want to stop or restart the Agents while leaving the processes they manage running. To do this, use one of the following commands on every Agent host.

- To stop the Agent itself, but leave the processes it manages running:

```
$ sudo service cloudera-scm-agent stop
```

- To restart a running Agent without restarting any of the processes it manages:

```
$ sudo service cloudera-scm-agent restart
```

If you want to stop or restart the Agents themselves and the services they manage, use one of the following commands on every Agent host.

- To stop the Agent and the processes it manages:

```
$ sudo service cloudera-scm-agent hard_stop
```

- To restart the running Agent and the processes it manages:

```
$ sudo service cloudera-scm-agent hard_restart
```

When an Agent is stopped using either of the `stop` or `hard_stop` commands, you cannot use either of the `restart` or `hard_restart` commands to start it. You must use the following `start` command to start a stopped agent regardless of how you stopped it:

```
$ sudo service cloudera-scm-agent start
```

Troubleshooting Cluster Configuration and Operation

This section contains solutions to some common problems that prevent you from using Cloudera Manager.

Solutions to Common Problems

Problems	Possible Causes	Solutions
Starting Services		
After you click the **Start** button to start a service, the **Finished** status doesn't display.	The host machine is disconnected from the Server, as indicated by missing heartbeats on the **Hosts** tab.	• Look at the logs for the service for causes of the problem. • Restart the Agents on the hosts where the heartbeats are missing.

Problems	Possible Causes	Solutions
After you click **Start** to start a service, the **Finished** status displays but there are error messages. The subcommands to start service components (such as JobTracker and one or more TaskTrackers) do not start.	A port specified in the **Configuration** tab of the service is already being used in your cluster. For example, the JobTracker port is in use by another process.	Enter an available port number in the port property (such as **JobTracker port)** in the **Configuration** tab of the service.
	There are incorrect directories specified in the **Configuration** tab of the service (such as the log directory).	Enter correct directories in the **Configuration** tab of the service.

Problems	Possible Causes	Solutions
Logs include `APPARENT DEADLOCK` entries for c3p0.	These deadlock messages are cause by the c3p0 process not making progress at the expected rate. This can indicate either that c3p0 is deadlocked or that its progress is slow enough to trigger these messages. In many cases, progress is occurring and these messages should not be seen as catastrophic.	There are a variety of ways to react to these log entries. • You may ignore these messages if system performance is not otherwise affected. Because these entries often occur during slow progress, they may be ignored in some cases. • You may modify the timer triggers. If c3p0 is making slow progress, increasing the period of time during which progress is evaluated stop the log entries from occurring. The default time between Timer triggers is 10 seconds and is configurable indirectly by configuring maxAdministrativeTaskTime. For more information, see http://www.mchange.com/projects/c3p0/#maxAdministrativeTaskTime. • You may increase the number of threads in the c3p0 pool, thereby increasing the resources available to make progress on tasks. For more information, see http://www.mchange.com/projects/c3p0/#numHelperThreads.

Logs and Events

For information about problems, you can check the logs and events:

- The Cloudera Manager Server and Agent logs. See Viewing the Cloudera Manager Server Log.

- The Events tab lets you search for and display events and alerts that have occurred within a selected time range filtered by service, hosts, and/or keywords. See Events and Alerts Filtering.

- The Logs tab presents log information for Hadoop services, filtered by role, host, and/or keywords as well log level (severity). See Log Filtering.

- Event and Log search features are also provided for individual user jobs, or for specific service. See the sections on Activity Monitoring and Service Monitoring.

Viewing the Cloudera Manager Server and Agent Logs

To help you troubleshoot problems, you can view the Cloudera Manager Server log.

To view the Cloudera Manager Server log:

1. Click the gear icon ⚙ to display the **Administration** page.

2. Click the **Server Log** link.

To view the Cloudera Manager Agent log:

1. Click the **Hosts** tab.

2. Click the link for the host where you want to see the Agent logs.

3. In the **Details** panel, click the **Details** link in the **Host Agent** column.

4. Click the **Agent Log** link.

> **Note**
>
> You can also view the Cloudera Manager Server log at `/var/log/cloudera-scm-server/cloudera-scm-server.log` on the Server host or the Cloudera Manager Agent log at `/var/log/cloudera-scm-agent/cloudera-scm-agent.log` on the Agent hosts for information about the problems.

Sending Diagnostic Data to Cloudera

- Configuring the Frequency of Diagnostic Data Collection

- Collecting and Sending Diagnostic Data to Cloudera on Demand

- Disabling the Automatic Sending of Diagnostic Data

- Manually Sending Diagnostic Data to Cloudera

- What Data Does Cloudera Manager Collect?

To obtain help solving problems when using Cloudera Manager on your cluster, you can collect and send diagnostic data to Cloudera Support.

You can have Cloudera Manager send diagnostic data to Cloudera automatically whenever a data collection occurs – either on a regular schedule or when you specifically trigger a data collection. You can also send a collected data set manually. Cloudera Manager is configured by default to collect data weekly and to send it automatically. You can schedule the frequency of data collection on a daily, weekly, or monthly schedule, or disable the scheduled collection of data entirely. Separately you can disable the automatic sending of data to Cloudera — see Disabling the Automatic Sending of Diagnostic Data.

> **Note**
>
> To automatically send diagnostic data requires the Cloudera Manager Server host to have Internet access, and be configured for sending data automatically. If your Cloudera Manager server does not have Internet access, you can manually send the diagnostic data as described below.

Configuring the Frequency of Diagnostic Data Collection

By default, Cloudera Manager collects diagnostic data on a weekly basis, and sends it automatically to Cloudera. You can change the frequency to daily, weekly, monthly, or never. If you set the schedule to Never you can still collect and send data to Cloudera on demand.

To change the frequency of diagnostic data collection:

1. Click the gear icon ▓ to display the **Administration** page.

2. Under the **Properties** tab, **Enterprise Support** category, click in the field for the property **Send diagnostic Data to Cloudera Automatically** and select the frequency you want.

3. You can change the day and time of day that the collection will be performed.

4. Click **Save Changes**

You can see the setting for the current data collection frequency under the **Support** menu in the main navigation bar.

Collecting and Sending Diagnostic Data to Cloudera on Demand

By default, Cloudera Manager will automatically attempt to send your collected data to Cloudera when you trigger a data collection. If you do not want data sent automatically, you must disable that feature (see Disabling the Automatic Sending of Diagnostic Data).

To collect and automatically send diagnostic data to Cloudera:

1. Click the **Support** menu link.

2. Choose **Send Diagnostic Data**.

   This opens the Send Diagnostic Data form.

   Note that at the top of the form it tells you whether Cloudera Manager is configured to send the data automatically or not. See the instructions below to change this.

3. Fill in or change the information here as appropriate.

   o To change the System Identifier, go to the **Administration** page (via the gear icon), under the **Properties** tab, **Other** category.

   o Cloudera Manager pre-populates the End Time based on the setting of the Time Range Selector. You should change this to be a few minutes after you observed the problem or condition that you are trying to capture.

   Note that the time range is based on the time zone of the host where Cloudera Manager server is running.

   o If you have a support ticket open with Cloudera support, please include the support ticket number in the field provided.

4. Click **Collect Diagnostic Data**.

   A Running Commands window shows you the progress of the data collection steps.

   When these steps are complete, the collected data is sent to Cloudera.

Disabling the Automatic Sending of Diagnostic Data

If you do not want data sent to Cloudera automatically, you can disable this feature. The data you collect will be saved

To disable sending diagnostic data to Cloudera automatically:

1. Click the gear icon to display the **Administration** page.

2. Under the **Properties** tab, **Enterprise Support** category, uncheck the box for **Send diagnostic Data to Cloudera Automatically**.

3. Click **Save Changes**

Manually Sending Diagnostic Data to Cloudera

> **Note**
>
> Automatically sending diagnostic data may fail sometimes and return an error message of "Could not send data to Cloudera." To work around this issue, you can manually send the data to Cloudera Support as described below.

To manually send collected diagnostic data to Cloudera:

1. Click the **Support** menu link.

2. Choose **Send Diagnostic Data**.
    This opens the Send Diagnostic Data form.

    Note that at the top of the form it tells you whether Cloudera Manager is configured to send the data automatically or not. See the instructions above to change this.

3. Fill in or change the information in the form as appropriate.
    Cloudera Manager pre-populates the start and end times, but you can change them.

    If you have a support ticket open with Cloudera support, please include the support ticket number in the field provided.

4. Click **Collect Diagnostic Data**.
    A Command Details window shows you the progress of the data collection steps.

5. Click **Download Result Data** to download and save a zip file of the information collected, on a host that has Internet access.

6. Download this [script](#) and run the following command on that host to send the data to Cloudera Support:

```
python phone_home.py --user anonymous --host cops.cloudera.com --port
22 --dropdir drop --file [file-you-downloaded].
```

> **Note**
>
> If you want to send your file manually but choose not to download the script, you can follow the instructions documented on the Cloudera Customer Portal at [Get Support - Uploading Files for Cloudera Support](#).

What Data Does Cloudera Manager Collect?

Cloudera Manager collects and returns a significant amount of information about the health and performance of the cluster. It includes the following:

*   Up to 1000 Cloudera Manager Audit Events: Configuration changes, add/remove of users, roles, services, etc.

- Data about the cluster structure which includes a list of all hosts, roles, and services along with the configs that are set through Cloudera Manager. Where passwords are set in Cloudera Manager, the passwords are not returned.

- Cloudera Manager License and version number.

- One day's worth of Cloudera Manager events: This includes critical errors Cloudera Manager watches for and more.

- Current health information for hosts, service, and roles. Includes results of health tests run by Cloudera Manager.

- Heartbeat information from each host, service, and role. These include status and some information about memory/disk/processer usage.

- The results of running Host Inspector.

- One day's worth of Cloudera Manager metrics.

- A download of the debug pages for Cloudera Manager roles.

- For each machine in the cluster, the result of running a number of system-level commands on that machine.

- Logs from each role on the cluster, as well as the CM server and agent logs.

Backup and Disaster Recovery

The following sections are covered in this topic:

- Designating a Replication Source

- HDFS Replication

- Hive Replication

Cloudera Backup and Disaster Recovery (BDR) provides an integrated, easy-to-use management solution for enabling data protection in the Hadoop platform. Cloudera BDR 1.0 provides rich functionality aimed towards replicating data stored in HDFS and accessed through Hive across datacenters for Disaster Recovery scenarios. When critical data is stored on HDFS, Cloudera BDR provides the necessary capabilities to ensure that the data is available at all times, even in the face of the complete shutdown of a datacenter.

Cloudera BDR provides key capabilities that are fully integrated into the Cloudera Manager User Interface:

- Select: Choose the key datasets that are critical for your business operations.

- Schedule: Create an appropriate schedule for data replication – trigger replication as quickly as is appropriate for your business needs.

- Monitor: Track progress of your replication jobs through a central console and easily identify issues or files that failed to be transferred.

- Alert: Issue alerts when a replication job fails or is aborted so that the problem can be diagnosed expeditiously.

These capabilities work seamlessly across Hive and HDFS – replication can be setup on files or directories in the case of HDFS and on tables in the case of Hive – without any manual translation of Hive datasets into HDFS datasets or vice-versa. Hive Metastore information is also replicated which means that the applications that depend upon the table definitions stored in Hive will work correctly on the replica side as well as the source side as table definitions are updated.

Built on top of a hardened version of "distcp" – the replication uses the scalability and availability of MapReduce itself to parallelize the copying of files using a specialized MapReduce job that diffs and transfers only changed files from each Mapper to the replica side efficiently and quickly.

Also available in the new version is the ability to do a "Dry Run" to verify configuration and understand the cost of the overall operation before actually copying the entire dataset.

Since the functionality is implemented as an add-on to Cloudera Manager, all Cloudera BDR functionality is available directly through the Cloudera Manager Admin Console.

Minimum Supported Version: CDH 4.0
License: BDR requires a separate license for each node on the destination side of your replication cluster.

> **Note**
>
> You must ensure that the following ports are open to allow communication between the source and destination Cloudera Manager servers and the HDFS, MapReduce, and Hive hosts to enable replication jobs to work:
>
> - Cloudera Manager Admin Console port: Default is 7180.
>
> - HDFS NameNode port: Default is 8020.
>
> - HDFS DataNode port: Default is 50010.
>
> See Configuring Ports for Cloudera Manager for more information, including how to verify the current values for these ports.

Designating a Replication Source

From the Cloudera Manager Admin Console, you can designate a Cloudera Manager server as the source for data (files) to be replicated to a service managed by the Cloudera Manager server you are logged into.

To set up a peer relationship as a replication source:

1. From the Service page for either Hive or HDFS, select the **Replication** tab.

2. Click the link **Add Replication Source** to go to the Admin **Peers** tab.

   You can also go directly to the **Peers** page by clicking the gear icon ⚙ at the far right of the navigation bar, and selecting the **Peers** tab.

   If there are no existing peers, you will see only an **Add Peer** button in addition to a short message.
   If you have existing peers, they are listed here.

3. Click the **Add Peer** button.

4. In the Add Peer pop-up, provide a name, the URL (including the port) of the Cloudera Manager Server that will act as the source for the data to be replicated, and the login credentials for that server.

   Note that the Data Replication feature recommends that SSL be used, and a warning is shown if the URL uses http instead of https. However, you can ignore the warning and proceed if SSL is not available.

5. Click the **Add Peer** button in the pop-up to create the peer relationship.

6. To test the connectivity between your current Cloudera Manager server and the remote server select **Test Connectivity** from the **Actions** menu associated with the peer.

Note that the current Cloudera Manager system is also available as a replication source.

Modifying the Peer Configuration

To modify the peer configuration (to change the login or password):

1. Click the gear icon to go to the Admin page.

2. Select the **Peers** tab.

3. From the **Actions** menu for the peer, select **Edit**.

4. Make your changes.

5. Click **Update Peer** to save your changes.

You can also delete a peer relationship:

- From the **Actions** menu for the peer, select **Delete**.

HDFS Replication

HDFS Replication enables you to copy (replicate) your HDFS data from a remote Peer Cloudera Manager server to your local Cloudera Manager server (the server whose Admin console you are currently logged into). You can add Peers though the **Administration** > **Peers** tab (see [Designating Peer Clusters](#)).

You can also use the **Add Replication Source** link on the HDFS **Replication** page to go to the **Peers** page.

Once you have a peer relationship set up with a Cloudera Manager server, you can configure replication of your HDFS data.

1. From the **Services** tab, go to the CDH4 HDFS service where you want to host the replicated data.

2. Click the **Replication** tab at the top of the page.

3. Select the HDFS service to be the source of the replicated data. If the peer Cloudera Manager Server has multiple CDH4 HDFS services (for example, if it is managing multiple CDH4 clusters) you will be able to select the HDFS service you want to use as the source.

   Note that the local CDH4 HDFS service (being managed by the Cloudera Manager server you are logged into) is also available as a replication source.

   If the peer whose HDFS service you want is not listed, click the **Add Peer** link to go to the Peers page to add a Cloudera Manager peer.

   When you select a replication source, the **Create Replication** pop-up opens.

4. Enter the path to the directory (or file) you want to replicate (the source).

5. Enter the path where the target files should be placed.

6. Select a schedule: You can have it run immediately, run once at a scheduled time in the future, or at regularly scheduled intervals. If you select "Once" or "Recurring" you are presented with fields that let you set the date and time and (if appropriate) the interval between runs.

7. If you want to modify the parameters of the MapReduce job, click **More Options**.
Here you will be able to select a MapReduce service (if there is more than on in your cluster) and change the following parameters:

   o The MapReduce service to use

   o The scheduler pool to use.

   o The user that should run the MapReduce job. By default this is *hdfs*. If you want to run the MR job as a different user, you can enter that here. If you are using Kerberos, you MUST provide a user name here, and it must be one with an ID greater 1000.

   o An alternative path for the logs.

   o Limits for the number of map slots and for bandwidth per mapper. The defaults are unlimited.

   o Whether to abort the job on an error (default is not to do so). This means that files copied up to that point will remain on the destination, but no additional files will be copied.

   o Whether to skip checksum checks (default is to perform them).

   o Whether to remove deleted files from the target directory if they have been removed on the source.

   o Whether to preserve the block size, replication count, and permissions as they exist on the source file system, or to use the settings as configured on the target file system. The default is to preserve these settings as on the source.

   > **Note:** If you leave the setting to preserve permissions, then you must be running as a superuser. You can use the "Run as" option to ensure that is the case.

   o Whether to generate alerts for various state changes in the replication workflow.

8. Click **Save Schedule** to save the replication specs.

When saved, the replication job appears in the Replication list, with relevant information about the source and target locations, and the timestamp of the last run and the next scheduled run (if there is a recurring schedule). A scheduled job will show a calendar icon to the left of the job specification. If it is scheduled to run once, the calendar icon will disappear after the job has run.

To specify additional replication tasks, click the **Create Replication** button that appears once you have added the first replication tasks.

Note that only one replication can occur at a time; if another replication job starts before the previous one has finished, the second one is cancelled.

You can test the replication task without actually transferring data using the "Dry Run" feature:

- From the **Actions** menu for the replication task you want to test and click **Dry Run**.

From the **Actions** menu for a replication task, in addition to **Dry Run** you can also:

- Edit the job configuration

- Run the job (immediately)

- Delete the job

- Disable/Enable the job (if the job is on a recurring schedule)
  When a task is disabled, instead of the calendar icon you will see a Stopped icon, and the job entry will appear in grey. Disabling and enabling a job is only available if the job is on a recurring schedule.

Viewing Replication Job Status

While a run is in progress, the calendar icon turns into spinner, and each stage of the replication task is indicated in the message after the replication specification.

- If the replication is successful, the number of files copied is indicated.
  If there have been no changes to a file at the source since the previous replication, then that file will *not* be copied. AS a result, after the initial replication run, only a subset of the files may actually be copied, and this will be indicated in the success message.

- If the replication fails, that will be indicated and the timestamp will appear in Red text.

- To view more information about completed replication runs, click anywhere in the replication job entry row in the replication list. This displays sub-entries for each past replication run.

- To view detailed information about a particular past run, click the entry for that replication run. This opens another sub-entry that shows:

  - A result message

  - The start and end time of the replication job.

  - A link to the command details for that replication run.

  - Details about the data that was replicated.

- When viewing a sub-entry, you can dismiss the sub-entry by clicking anywhere in its parent entry, or by clicking the return arrow icon !at the top left of the sub-entry area.

Hive Replication

Hive Replication enables you to copy (backup) and keep in sync the Hive Metastore and data from clusters managed by a remote peer or local Cloudera Manager server, and keep the copy on a cluster managed by your local Cloudera Manager server (the server whose Admin console you are currently logged into). You can add Peers though the **Administration** > **Peers** tab (see [Designating Peer Clusters](#)).

You can use the **Add Peer** link on the **Replication** page to go to the **Peers** page to add a new peer Cloudera Manager server.

Once you have a peer relationship set up with a Cloudera Manager server, you can configure replication of your Hive Metastore data.

1. From the Services tab, go to the CDH4 Hive service where you want to host the replicated data.

2. Click the **Replication** tab at the top of the page.

3. Select the Hive service to be the source of the replicated data. If the peer Cloudera Manager Server has multiple CDH4 Hive services (for example, if it is managing multiple CDH4 clusters) you will be able to select the service you want to use as the source.

   Note that the local CDH4 Hive service (being managed by the Cloudera Manager server you are logged into) is also available as a replication source.

   If the peer whose Hive service you want is not listed, click the **Add Peer** link to go to the Peers page to add a Cloudera Manager peer.

   When you select a replication source, the **Create Replication** pop-up opens.

4. Leave **Replicate All** checked to replicate all the Hive metastore databases from the source.

   To replicate only selected databases, uncheck this option and enter the Database name(s) and tables you want to replicate.

5. Select the target destination. If there is only one Hive service managed by Cloudera Manager available as a target, then this will be specified as the target. If there are more than one Hive services managed by this Cloudera Manager, then you will be able to select among those.

6. Select a schedule: You can have it run immediately, run once at a scheduled time in the future, or at regularly scheduled intervals. If you select "Once" or "Recurring" you are presented with fields that let you set the date and time and (if appropriate) the interval between runs.

7. Uncheck the **Replicate HDFS Files** checkbox to skip replicating the associated data files — if you uncheck this, only the Hive metadata will be replicated. These are replicated to a default location; to specify a different location, enter the path in the **Destination** field under the **More Options** section.

8. Use the **More Options** section to specify an export location, modify the parameters of the MapReduce job that will perform the replication, and other options.
Here you will be able to select a MapReduce service (if there is more than one in your cluster) and change the following parameters:

- By default, Cloudera Manager exports the Hive Metadata to a default HDFS location (`/user/${user.name}/.cm/hive`) and then imports from this HDFS file to the target Hive Metastore. The default HDFS location for this export file can be overriden by specifying a path in the **Export Path** field.

- The **Force Overwrite** option, if checked, forces overwriting data in the target metastore if there are incompatible changes detected. For example, if the target metastore was modified and a new partition was added to a table, this option would force deletion of that partition, overwriting the table with the version found on the source.

- By default, Cloudera Manager replicates Hive's HDFS data files to a default location (`/`). To override the default, enter a path in the **Destination** field.

- Select the MapReduce service to use for this replication (if there is more than one in your cluster).

- The user that should run the MapReduce job. By default this is *hdfs*. If you want to run the MR job as a different user, you can enter that here. If you are using Kerberos, you *MUST* provide a user name here, and it must be one with an ID greater 1000.

- An alternative path for the logs.

- Limits for the number of map slots and for bandwidth per mapper. The defaults are unlimited.

- Whether to abort the job on an error (default is not to do so). This means that files copied up to that point will remain on the destination, but no additional files will be copied.

- Whether to skip checksum checks (default is to perform them).

- Whether to remove deleted files from the target directory if they have been removed on the source.

- Whether to preserve the block size, replication count, and permissions as they exist on the source file system, or to use the settings as configured on the target file system. The default is to preserve these settings as on the source.

> **Note:** If you leave the setting to preserve permissions, then you must be running as a superuser. You can use the "Run as" option to ensure that is the case.

- Whether to generate alerts for various state changes in the replication workflow.

9. Click **Save Schedule** to save the replication specs.

When saved, the replication job appears in the Replication list, with relevant information about the source and target locations, and the timestamp of the last run and the next scheduled run (if there is a recurring schedule). A scheduled job will show a calendar icon to the left of the job specification. If it is scheduled to run once, the calendar icon will disappear after the job has run.

To specify additional replication tasks, click the **Create Replication** button that appears once you have added the first replication tasks.

If the replication failed, the timestamp will appear in Red text.

Note that only one replication can occur at a time; if another replication job starts before the previous one has finished, the second one is cancelled.

You can test the replication task without actually transferring data using the "Dry Run" feature:

- From the **Actions** menu for the replication task you want to test, click **Dry Run**.

From the **Actions** menu for a replication task, in addition to **Dry Run** you can also:

- Edit the task configuration

- Run the task (immediately)

- Delete the task

- Disable/Enable the job (if the job is on a recurring schedule).
  When a task is disabled, instead of the calendar icon you will see a **Stopped** icon, and the job entry will appear in grey. Disabling and enabling a job is only available if the job is on a recurring schedule.

Viewing Replication Job Status

While a run is in progress, the calendar icon turns into spinner, and each stage of the replication task is indicated in the message after the replication specification.

- If the replication is successful, the number of files copied is indicated.
  If there have been no changes to a file at the source since the previous replication, then that file will *not* be copied. AS a result, after the initial replication run, only a subset of the files may actually be copied, and this will be indicated in the success message.

- If the replication fails, that will be indicated and the timestamp will appear in Red text.

- To view more information about completed replication runs, click anywhere in the replication job entry row in the replication list. This displays sub-entries for each past replication run.

- To view detailed information about a particular past run, click the entry for that replication run. This opens another sub-entry that shows:

  o A result message

  o The start and end time of the replication job.

- o A link to the command details for that replication run.

- o Details about the data that was replicated.

- When viewing a sub-entry, you can dismiss the sub-entry by clicking anywhere in its parent entry, or by clicking the return arrow icon !at the top left of the sub-entry area.

Getting Support

Cloudera Support

Cloudera can help you install, configure, optimize, tune, and run Hadoop for large scale data processing and analysis. Cloudera supports Hadoop whether you run our distribution on servers in your own data center, or on hosted infrastructure services such as Amazon EC2, Rackspace, SoftLayer, or VMware's vCloud.

If you are a Cloudera customer, you can:

- Create a [Cloudera Support Ticket](#).

- Visit the [Cloudera Knowledge Base](#).

- Learn how to [register for an account](#) to create a support ticket.

Community Support

Register for the [Cloudera Manager Users group](#).

Register for the [CDH Users group](#).

Report Issues

Cloudera tracks software and documentation bugs and enhancement requests for CDH on [issues.cloudera.org](#). Your input is appreciated, but before filing a request, please search the Cloudera issue tracker for existing issues and send a message to the CDH user's list, [cdh-user@cloudera.org](#), or the CDH developer's list, [cdh-dev@cloudera.org](#).

Get Announcements about New CDH and Cloudera Manager Releases

Cloudera provides the following public mailing lists that send announcements about new CDH and Cloudera Manager product releases and updates:

- To receive CDH release announcements, subscribe to the [CDH-announce](#) list.

- To receive Cloudera Manager release announcements, subscribe to the [CM-announce](#) list.