**CDH and Cloudera Manager Networking and Security Requirements**

The hosts in a Cloudera Manager deployment must satisfy the following networking and security requirements:

* CDH requires IPv4. IPv6 is not supported and must be disabled.

See also

* Multihoming CDH or Cloudera Manager is not supported outside specifically certified Cloudera partner appliances. Cloudera finds that current Hadoop architectures combined with modern network infrastructures and security practices remove the need for multihoming. Multihoming, however, is beneficial internally in appliance form factors to take advantage of high-bandwidth InfiniBand interconnects.
* Although some subareas of the product might work with unsupported custom multihoming configurations, there are known issues with multihoming. In addition, unknown issues can arise because multihoming is not covered by the test matrix outside the Cloudera-certified partner appliances.
* Cluster hosts must have a working network name resolution system and correctly formatted/etc/hosts file. All cluster hosts must have properly configured forward and reverse host resolution through DNS. The /etc/hosts files must:
  + Contain consistent information about hostnames and IP addresses across all hosts
  + Not contain uppercase hostnames
  + Not contain duplicate IP addresses

Cluster hosts must not use aliases, either in /etc/hosts or in configuring DNS. A properly formatted/etc/hosts file should be similar to the following example:

127.0.0.1

**localhost**.localdomain

**localhost**

192.168.1.1

**cluster-01**.example.com

**cluster-01**

192.168.1.2

**cluster-02**.example.com

**cluster-02**

192.168.1.3

**cluster-03**.example.com

**cluster-03**

* In most cases, the Cloudera Manager Server must have SSH access to the cluster hosts when you run the installation or upgrade wizard. You must log in using a root account or an account that has password-less sudo permission. For authentication during the installation and upgrade procedures, you must either enter the password or upload a public and private key pair for the root or sudo user account. If you want to use a public and private key pair, the public key must be installed on the cluster hosts before you use Cloudera Manager.

Cloudera Manager uses SSH only during the initial install or upgrade. Once the cluster is set up, you can disable root SSH access or change the root password. Cloudera Manager does not save SSH credentials, and all credential information is discarded when the installation is complete.

* If [single user mode](http://www.cloudera.com/documentation/enterprise/release-notes/topics/install_singleuser_reqts.html#xd_583c10bfdbd326ba--69adf108-1492ec0ce48--7ade) is not enabled, 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 and mapred users).
* No blocking is done by Security-Enhanced Linux (SELinux).**Note:** Cloudera Enterprise is supported on platforms with Security-Enhanced Linux (SELinux) enabled. Cloudera is not responsible for policy support nor policy enforcement. If you experience issues with SELinux, contact your OS provider.
* No blocking by iptables or firewalls; port 7180 must be open because it is used to access Cloudera Manager after installation. Cloudera Manager communicates using specific ports, which must be open.
* For RHEL and CentOS, the /etc/sysconfig/network file on each host must contain the hostname you have just set (or verified) for that host.
* Cloudera Manager and CDH use several user accounts and groups to complete their tasks. The set of user accounts and groups varies according to the components you choose to install. Do not delete these accounts or groups and do not modify their permissions and rights. Ensure that no existing systems prevent these accounts and groups from functioning. For example, if you have scripts that delete user accounts not in a whitelist, add these accounts to the list of permitted accounts. Cloudera Manager, CDH, and managed services create and use the following accounts and groups:

| ***Users and Groups*** | | | |
| --- | --- | --- | --- |
| **Component (Version)** | **Unix User ID** | **Groups** | **Notes** |
| Cloudera Manager (all versions) | cloudera-scm | cloudera-scm | Cloudera Manager processes such as the Cloudera Manager Server and the monitoring roles run as this user.  The Cloudera Manager keytab file must be named cmf.keytab since that name is hard-coded in Cloudera Manager.**Note:** Applicable to clusters managed by Cloudera Manager only. |
| Apache Accumulo (Accumulo 1.4.3 and higher) | accumulo | accumulo | Accumulo processes run as this user. |
| Apache Avro |  | | No special users. |
| Apache Flume (CDH 4, CDH 5) | flume | flume | The sink that writes to HDFS as this user must have write privileges. |
| Apache HBase (CDH 4, CDH 5) | hbase | hbase | The Master and the RegionServer processes run as this user. |
| HDFS (CDH 4, CDH 5) | hdfs | hdfs, hadoop | The NameNode and DataNodes run as this user, and the HDFS root directory as well as the directories used for edit logs should be owned by it. |
| Apache Hive (CDH 4, CDH 5) | hive | hive | The HiveServer2 process and the Hive Metastore processes run as this user.  A user must be defined for Hive access to its Metastore DB (for example, MySQL or Postgres) but it can be any identifier and does not correspond to a Unix uid. This isjavax.jdo.option.ConnectionUserName in hive-site.xml. |
| Apache HCatalog (CDH 4.2 and higher, CDH 5) | hive | hive | The WebHCat service (for REST access to Hive functionality) runs as the hive user. |
| HttpFS (CDH 4, CDH 5) | httpfs | httpfs | The HttpFS service runs as this user. See[HttpFS Security Configuration](http://www.cloudera.com/documentation/enterprise/release-notes/topics/cdh_sg_httpfs_security.html#topic_7) for instructions on how to generate the merged httpfs-http.keytab file. |
| Hue (CDH 4, CDH 5) | hue | hue | Hue services run as this user. |
| Cloudera Impala (CDH 4.1 and higher, CDH 5) | impala | impala, hive | Impala services run as this user. |
| Apache Kafka (Cloudera Distribution of Kafka 1.2.0) | kafka | kafka | Kafka services run as this user. |
| Java KeyStore KMS (CDH 5.2.1 and higher) | kms | kms | The Java KeyStore KMS service runs as this user. |
| Key Trustee KMS (CDH 5.3 and higher) | kms | kms | The Key Trustee KMS service runs as this user. |
| Key Trustee Server (CDH 5.4 and higher) | keytrustee | keytrustee | The Key Trustee Server service runs as this user. |
| Kudu | kudu | kudu | Kudu services run as this user. |
| Llama (CDH 5) | llama | llama | Llama runs as this user. |
| Apache Mahout |  | | No special users. |
| MapReduce (CDH 4, CDH 5) | mapred | mapred, hadoop | Without Kerberos, the JobTracker and tasks run as this user. The LinuxTaskController binary is owned by this user for Kerberos. |
| Apache Oozie (CDH 4, CDH 5) | oozie | oozie | The Oozie service runs as this user. |
| Parquet |  | | No special users. |
| Apache Pig |  | | No special users. |
| Cloudera Search (CDH 4.3 and higher, CDH 5) | solr | solr | The Solr processes run as this user. |
| Apache Spark (CDH 5) | spark | spark | The Spark History Server process runs as this user. |
| Apache Sentry (CDH 5.1 and higher) | sentry | sentry | The Sentry service runs as this user. |
| Apache Sqoop (CDH 4, CDH 5) | sqoop | sqoop | This user is only for the Sqoop1 Metastore, a configuration option that is not recommended. |
| Apache Sqoop2 (CDH 4.2 and higher, CDH 5) | sqoop2 | sqoop, sqoop2 | The Sqoop2 service runs as this user. |
| Apache Whirr |  | | No special users. |
| YARN (CDH 4, CDH 5) | yarn | yarn, hadoop | Without Kerberos, all YARN services and applications run as this user. The LinuxContainerExecutor binary is owned by this user for Kerberos. |
| Apache ZooKeeper (CDH 4, CDH 5) | zookeeper | zookeeper | The ZooKeeper processes run as this user. It is not configurable. |

**CDH and Cloudera Manager Supported Transport Layer Security Versions**

The following components are supported by the indicated versions of Transport Layer Security (TLS):

| ***Components Supported by TLS*** | | | | |
| --- | --- | --- | --- | --- |
| **Component** | **Role** | **Name** | **Port** | **Version** |
| Cloudera Manager | Cloudera Manager Server |  | 7182 | TLS 1.2 |
| Cloudera Manager | Cloudera Manager Server |  | 7183 | TLS 1.2 |
| Flume |  |  | 9099 | TLS 1.2 |
| Flume |  | Avro Source/Sink |  | TLS 1.2 |
| Flume |  | Flume HTTP Source/Sink |  | TLS 1.2 |
| HBase | Master | HBase Master Web UI Port | 60010 | TLS 1.2 |
| HDFS | NameNode | Secure NameNode Web UI Port | 50470 | TLS 1.2 |
| HDFS | Secondary NameNode | Secure Secondary NameNode Web UI Port | 50495 | TLS 1.2 |
| HDFS | HttpFS | REST Port | 14000 | TLS 1.1, TLS 1.2 |
| Hive | HiveServer2 | HiveServer2 Port | 10000 | TLS 1.2 |
| Hue | Hue Server | Hue HTTP Port | 8888 | TLS 1.2 |
| Impala | Impala Daemon | Impala Daemon Beeswax Port | 21000 | TLS 1.2 |
| Impala | Impala Daemon | Impala Daemon HiveServer2 Port | 21050 | TLS 1.2 |
| Impala | Impala Daemon | Impala Daemon Backend Port | 22000 | TLS 1.2 |
| Impala | Impala StateStore | StateStore Service Port | 24000 | TLS 1.2 |
| Impala | Impala Daemon | Impala Daemon HTTP Server Port | 25000 | TLS 1.2 |
| Impala | Impala StateStore | StateStore HTTP Server Port | 25010 | TLS 1.2 |
| Impala | Impala Catalog Server | Catalog Server HTTP Server Port | 25020 | TLS 1.2 |
| Impala | Impala Catalog Server | Catalog Server Service Port | 26000 | TLS 1.2 |
| Oozie | Oozie Server | Oozie HTTPS Port | 11443 | TLS 1.1, TLS 1.2 |
| Solr | Solr Server | Solr HTTP Port | 8983 | TLS 1.1, TLS 1.2 |
| Solr | Solr Server | Solr HTTPS Port | 8985 | TLS 1.1, TLS 1.2 |
| Spark | History Server |  | 18080 | TLS 1.2 |
| YARN | ResourceManager | ResourceManager Web Application HTTP Port | 8090 | TLS 1.2 |
| YARN | JobHistory Server | MRv1 JobHistory Web Application HTTP Port | 19890 | TLS 1.2 |

# Configuring Network Settings

**Minimum Required Role:** [**Full Administrator**](http://www.cloudera.com/documentation/enterprise/latest/topics/cm_sg_user_roles.html#concept_wfh_tvy_qp)

To configure a proxy server thorough which data is downloaded to and uploaded from the Cloudera Manager Server, do the following:

1. Select **Administration** > **Settings**.
2. Click the **Network** category.
3. Configure proxy properties.
4. Click **Save Changes** to commit the changes.

# Specifying Racks for Hosts

**Minimum Required Role:** [**Cluster Administrator**](http://www.cloudera.com/documentation/enterprise/latest/topics/cm_sg_user_roles.html#concept_wfh_tvy_qp) (also provided by **Full Administrator**)

To get maximum performance, it is important to configure CDH so that it knows the topology of your network. Network locations such as hosts and racks are represented in a tree, which reflects the network “distance” between locations. HDFS will use the network location to be able to place block replicas more intelligently to trade off performance and resilience. When placing jobs on hosts, CDH will prefer within-rack transfers (where there is more bandwidth available) to off-rack transfers; the MapReduce and YARN schedulers use network location to determine where the closest replica is as input to a map task. These computations are performed with the assistance of rack awareness scripts.

Cloudera Manager includes internal rack awareness scripts, but you must specify the racks where the hosts in your cluster are located. If your cluster contains more than 10 hosts, Cloudera recommends that you specify the rack for each host. HDFS, MapReduce, and YARN will automatically use the racks you specify.

Cloudera Manager supports nested rack specifications. For example, you could specify the rack /rack3, or/group5/rack3 to indicate the third rack in the fifth group. All hosts in a cluster must have the same number of path components in their rack specifications.

To specify racks for hosts:

1. Click the **Hosts** tab.
2. Check the checkboxes next to the host(s) for a particular rack, such as all hosts for /rack123.
3. Click **Actions for Selected (n)** > **Assign Rack**, where n is the number of selected hosts.
4. Enter a rack name or ID that starts with a slash /, such as /rack123 or /aisle1/rack123, and then click**Confirm**.
5. Optionally [restart affected services](http://www.cloudera.com/documentation/enterprise/latest/topics/cm_mc_start_stop_service.html#cmug_topic_5_6). Rack assignments are not automatically updated for running services.

**Decommissioning Hosts**

**Minimum Required Role:** [**Limited Operator**](http://www.cloudera.com/documentation/enterprise/latest/topics/cm_sg_user_roles.html#concept_wfh_tvy_qp) (also provided by **Operator**, **Configurator**, **Cluster Administrator**, or **Full Administrator**)

You cannot decommission a DataNode or a host with a DataNode if the number of DataNodes equals the replication factor (which by default is three) of any file stored in HDFS. For example, if the replication factor of any file is three, and you have three DataNodes, you cannot decommission a DataNode or a host with a DataNode. If you attempt to decommission a DataNode or a host with a DataNode in such situations, the DataNode will be decommissioned, but the decommission process will not complete. You will have to abort the decommission and recommission the DataNode.

To decommission hosts:

1. If the host has a DataNode, perform the steps in [Tuning HDFS Prior to Decommissioning DataNodes](http://www.cloudera.com/documentation/enterprise/latest/topics/cm_mc_decomm_host.html#cmug_topic_7_8__section_wq5_ztq_nm).
2. Click the **Hosts** tab.
3. Select the checkboxes next to one or more hosts.
4. Select **Actions for Selected** > **Hosts Decommission**.

A confirmation pop-up informs you of the roles that will be decommissioned or stopped on the hosts you have selected.

1. Click **Confirm**. A Decommission Command pop-up displays that shows each step or decommission command as it is run, service by service. In the Details area, click  to see the subcommands that are run for decommissioning a given service. Depending on the service, the steps may include adding the host 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 stopped.

You can abort the decommission process by clicking the **Abort** button, but you must recommission and restart each role that has been decommissioned.

The Commission State facet in the Filters lists displays  Decommissioning while decommissioning is in progress, and  Decommissioned when the decommissioning process has finished. When the process is complete, a  is added in front of Decommission Command.

You cannot start roles on a decommissioned host.

**Note:** When a DataNode is decommissioned, the data blocks are not removed from the storage directories. You must delete the data manually.

**Recommissioning Hosts**

**Minimum Required Role:** [**Operator**](http://www.cloudera.com/documentation/enterprise/latest/topics/cm_sg_user_roles.html#concept_wfh_tvy_qp) (also provided by **Configurator**, **Cluster Administrator**, **Full Administrator**)

Only hosts that are decommissioned using Cloudera Manager can be recommissioned.

1. Click the **Hosts** tab.
2. Select one or more hosts to recommission.
3. Select **Actions for Selected** > **Recommission** and **Confirm**. A Recommission Command pop-up displays that shows each step or recommission command as it is run. When the process is complete, a  is added in front of Recommission Command. The host and roles are marked as commissioned, but the roles themselves are not restarted.

**Restarting All The Roles on a Host**

**Minimum Required Role:** [**Operator**](http://www.cloudera.com/documentation/enterprise/latest/topics/cm_sg_user_roles.html#concept_wfh_tvy_qp) (also provided by **Configurator**, **Cluster Administrator**, **Full Administrator**)

1. Click the **Hosts** tab.
2. Select one or more hosts on which to start all roles.
3. Select **Actions for Selected** > **Start Roles on Hosts**.

# Configuring Cloudera Manager for High Availability With a Load Balancer

This section provides an example of configuring Cloudera Manager 5 for high availability using a TCP load balancer. The procedures describe how to configure high availability using a specific, open-source load balancer. Depending on the operational requirements of your CDH deployment, you can select a different load balancer. You can use either a hardware or software load balancer, but must be capable of forwarding all Cloudera Manager ports to backing server instances. (See [Ports Used by Cloudera Manager and Cloudera Navigator](http://www.cloudera.com/documentation/enterprise/latest/topics/cm_ig_ports_cm.html#xd_583c10bfdbd326ba-3ca24a24-13d80143249--7eec) for more information about the ports used by Cloudera Manager.)

This topic discusses Cloudera Manager high availability in the context of active-passive configurations only;active-active configurations are currently unsupported. For information about active-active configuration options, see <http://en.wikipedia.org/wiki/High-availability_cluster>.

**Important:** Cloudera Support supports all of the configuration and modification to Cloudera software detailed in this document. However, Cloudera Support is unable to assist with issues or failures with the third-party software that is used. Use of any third-party software, or software not directly covered by Cloudera Support, is at the risk of the end user.

Continue reading:

* [Introduction to Cloudera Manager Deployment Architecture](http://www.cloudera.com/documentation/enterprise/latest/topics/admin_cm_ha_deploy_arch.html#concept_lxc_dvc_pr)
* [Prerequisites for Setting up Cloudera Manager High Availability](http://www.cloudera.com/documentation/enterprise/latest/topics/admin_cm_ha_prereqs.html#concept_tc4_nzc_pr)
* [Cloudera Manager Failover Protection](http://www.cloudera.com/documentation/enterprise/latest/topics/cm_failover_db.html#concept_n5l_kdd_bv)
* [High-Level Steps to Configure Cloudera Manager High Availability](http://www.cloudera.com/documentation/enterprise/latest/topics/admin_cm_ha_steps.html#concept_k1p_hbd_pr)
* [Database High Availability Configuration](http://www.cloudera.com/documentation/enterprise/latest/topics/admin_cm_ha_dbms.html#concept_opf_2j2_qr)
* [TLS and Kerberos Configuration for Cloudera Manager High Availability](http://www.cloudera.com/documentation/enterprise/latest/topics/admin_cm_ha_tls.html#concept_j5g_sj2_qr)

**Cloudera Manager Resource Requirements**

Cloudera Manager requires the following resources:

* **Disk Space**
  + **Cloudera Manager Server**
    - 5 GB on the partition hosting /var.
    - 500 MB on the partition hosting /usr.
    - For parcels, the space required depends on the number of parcels you download to the Cloudera Manager Server and distribute to Agent hosts. You can download multiple parcels of the same product, of different versions and different builds. If you are managing multiple clusters, only one parcel of a product/version/build/distribution is downloaded on the Cloudera Manager Server—not one per cluster. In the local parcel repository on the Cloudera Manager Server, the approximate sizes of the various parcels are as follows:
      * CDH 5 (which includes Impala and Search) - 1.5 GB per parcel (packed), 2 GB per parcel (unpacked)
      * Impala - 200 MB per parcel
      * Cloudera Search - 400 MB per parcel
  + **Cloudera Management Service** -The Host Monitor and Service Monitor databases are stored on the partition hosting /var. Ensure that you have at least 20 GB available on this partition.
  + **Agents** - On Agent hosts, each unpacked parcel requires about three times the space of the downloaded parcel on the Cloudera Manager Server. By default, unpacked parcels are located in/opt/cloudera/parcels.
* **RAM** - 4 GB is recommended for most cases and is required when using Oracle databases. 2 GB might be sufficient for non-Oracle deployments with fewer than 100 hosts. However, to run the Cloudera Manager Server on a machine with 2 GB of RAM, you must tune down its maximum heap size (by modifying -Xmx in /etc/default/cloudera-scm-server). Otherwise the kernel might kill the Server for consuming too much RAM.
* **Python** - Cloudera Manager requires Python 2.4 or higher, but Hue in CDH 5 and package installs of CDH 5 require Python 2.6 or 2.7. All supported operating systems include Python version 2.4 or higher.
* **Perl** - Cloudera Manager requires [perl](https://www.perl.org/get.html).