

# Apache Accumulo Installation Guide

for using Cloudera's packaging of Accumulo for CDH



## Important Notice

© 2010-2015 Cloudera, Inc. All rights reserved.

Cloudera, the Cloudera logo, Cloudera Impala, Impala, and any other product or service names or slogans contained in this document, except as otherwise disclaimed, are trademarks of Cloudera and its suppliers or licensors, and may not be copied, imitated or used, in whole or in part, without the prior written permission of Cloudera or the applicable trademark holder.

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

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

Cloudera may have patents, patent applications, trademarks, copyrights, or other intellectual property rights covering subject matter in this document. Except as expressly provided in any written license agreement from Cloudera, the furnishing of this document does not give you any license to these patents, trademarks copyrights, or other intellectual property.

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

**Cloudera, Inc.**  
**1001 Page Mill Road**  
**Palo Alto, CA 94304-1008**  
[info@cloudera.com](mailto:info@cloudera.com)  
**US: 1-888-789-1488**  
**Intl: 1-650-362-0488**  
[www.cloudera.com](http://www.cloudera.com)

## Release Information

Version: 1.6.0-cdh4.6.0

Date: July 17, 2014

Version: 1.6.0-cdh5.1.4

Date: March 19, 2014

# Table of Contents

|   |           |
|---|-----------|
| <b>ABOUT THIS GUIDE .....</b>   | <b>1</b>  |
| <b>INTRODUCING APACHE ACCUMULO .....</b>  | <b>1</b>  |
| KNOWN ISSUES .....  | 1         |
| — <i>Known issues with the Accumulo 1.6 service in Cloudera Manager</i> .....         | 1         |
| PREREQUISITES .....   | 1         |
| <b>INSTALL APACHE ACCUMULO USING CLOUDERA MANAGER .....</b>                           | <b>2</b>  |
| STEP 1: INSTALL AND CONFIGURE CLOUDERA MANAGER AND CDH .....                          | 2         |
| STEP 2: INSTALL THE ACCUMULO PARCEL.....  | 3         |
| STEP 3: CONFIGURE HDFS .....  | 3         |
| STEP 4: ADD THE ACCUMULO SERVICE .....  | 4         |
| STEP 5: OPTIONAL CONFIGURATION FOR PERFORMANCE TUNING – RELAXING WAL DURABILITY ..... | 5         |
| <b>INSTALL APACHE ACCUMULO FROM DISTRIBUTION PACKAGES .....</b>                       | <b>5</b>  |
| STEP 1: ADD OR BUILD THE ACCUMULO REPOSITORY .....                                    | 6         |
| <i>On Red Hat-compatible Systems</i> .....  | 6         |
| <i>On SLES Systems</i> .....  | 6         |
| <i>On Ubuntu or Debian Systems</i> .....  | 7         |
| STEP 2: INSTALL ACCUMULO .....  | 9         |
| STEP 3: CONFIGURE HDFS .....  | 11        |
| STEP 4: CONFIGURE ACCUMULO FOR YOUR ENVIRONMENT .....                                 | 11        |
| STEP 5: INITIALIZE ACCUMULO.....  | 13        |
| STEP 6: START ACCUMULO .....  | 14        |
| STEP 7: OPTIONAL CONFIGURATION FOR PERFORMANCE TUNING – RELAXING WAL DURABILITY ..... | 15        |
| <b>CONFIGURING CLOUDERA MANAGER FOR ACCUMULO 1.6 ON CDH 4.....</b>                    | <b>15</b> |
| STEP 1: ADD THE REMOTE PARCEL REPOSITORY FOR ACCUMULO 1.6.0-CDH4.6.0.....             | 15        |
| STEP 2: INSTALL AND CONFIGURE CLOUDERA MANAGER 5 AND CDH 4.6.0.....                   | 15        |
| STEP 3: INSTALL THE ACCUMULO PARCEL.....  | 16        |
| <b>UPGRADING FROM ACCUMULO 1.4.....</b>   | <b>17</b> |
| UPGRADING WITH CLOUDERA MANAGER.....  | 17        |
| UPGRADING VIA RPMS .....  | 18        |
| <b>UPGRADING FROM CDH 4 TO CDH 5.....</b>   | <b>18</b> |
| STEP 1: DOCUMENT AND REMOVE ACCUMULO 1.6 FOR CDH 4 .....                              | 19        |

|  |           |
|--|-----------|
| STEP 2A: UPDATE PARCELS AND UPGRADE CDH .....          | 19        |
| STEP 2B: UPDATE PACKAGES AND UPGRADE CDH.....          | 19        |
| STEP 3: ADD ACCUMULO 1.6 FOR CDH 5 .....               | 20        |
| <b>TEST THE ACCUMULO SHELL.....</b>                    | <b>20</b> |
| <b>USING SQOOP 1 WITH ACCUMULO.....</b>                | <b>22</b> |
| SQOOP 1 CLIENT UNDER CDH 5 AND CLLOUDERA MANAGER ..... | 22        |
| SQOOP 1 UNDER CDH 4 OR WITHOUT CLLOUDERA MANAGER ..... | 22        |
| <b>USING LZO COMPRESSION WITH ACCUMULO .....</b>       | <b>23</b> |
| <b>USING ACCUMULO WITH MAVEN.....</b>                  | <b>24</b> |
| <b>DEFAULT PORTS .....</b>                             | <b>26</b> |
| <b>CREATING A LOCAL YUM REPOSITORY.....</b>            | <b>27</b> |

## About this Guide

This guide describes how to install Cloudera's packaging of Apache Accumulo for use with CDH.

## Introducing Apache Accumulo

[Apache Accumulo](#)<sup>™</sup> is an ideal solution for government agencies looking for a secure, distributed NoSQL data store to serve their most performance-intensive Big Data applications. Accumulo is an open source project integrated with Hadoop and provides the ability to store data in massive tables (billions of rows / millions of columns) for fast, random access. Accumulo was created and contributed to the Apache Software Foundation by the National Security Agency (NSA). It has quickly gained adoption as a Hadoop-based key/value store for applications that have unique and stringent information security requirements.

## Known Issues

### —Known issues with the Accumulo 1.6 service in Cloudera Manager

See the [known issues document for your release of Cloudera Manager](#) for problems and workarounds specific to running an Accumulo 1.6 service.

## Prerequisites

Accumulo depends on HDFS and ZooKeeper libraries and configuration information. TabletServers should be collocated with DataNodes. Optionally, you can use Accumulo with MapReduce and Sqoop 1. Cloudera recommends that MapReduce users rely on the stand-alone service in CDH 4 and on YARN (includes MapReduce 2) in CDH 5.

The current release of Cloudera's packaging of Apache Accumulo is tested for use with CDH 4.6.0 and higher and CDH 5.1.0 and higher. Cloudera Manager has been tested for managing this release with both parcels and package (RPM/DEB) installations with CDH 5.1.0 and higher.

For full cluster installations, Cloudera strongly recommends following the [Tips and Guidelines](#) in the [CDH 5 Installation Guide](#).

## Install Apache Accumulo using Cloudera Manager

This section describes how to install Cloudera's packaging of Accumulo by using Cloudera Manager 5.1.0 or higher. If you prefer to install Accumulo from packages and manage the cluster manually, skip this section and see the [Install Apache Accumulo from Distribution Packages](#) section on page 5.

### Note : Managing a cluster installed with packages

The instructions in this section use parcels. You can also use Cloudera Manager to manage the Accumulo 1.6 service when installing packages. To do so, follow Steps 1 and 2 in [Install Apache Accumulo from Distribution Packages](#). During the package installation sub-step of Step 2, follow the instructions for **All client hosts** only. In particular, do *not* install any of the role-specific packages such as `accumulo-master`, `accumulo-monitor`, and so on.

After the package installation, return to Step 3: [Configure HDFS](#) below to continue installing the Accumulo 1.6 service.

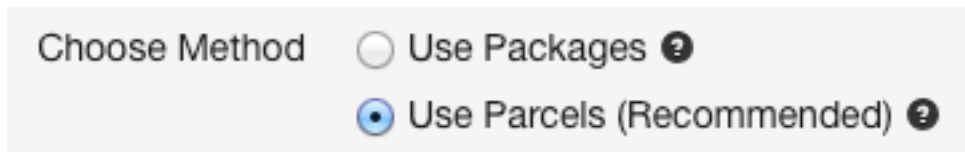
### Note: Managing a CDH 4 cluster



The instructions in this section work with CDH 5. You may also use Cloudera Manager to manage the Accumulo 1.6 service on a CDH 4 cluster. To do so, follow the instructions in [Configuring Cloudera Manager for Accumulo 1.6 on CDH 4](#) and then return to Step 3: [Configure HDFS](#) below to continue installing the Accumulo 1.6 service.

## Step 1: Install and Configure Cloudera Manager and CDH

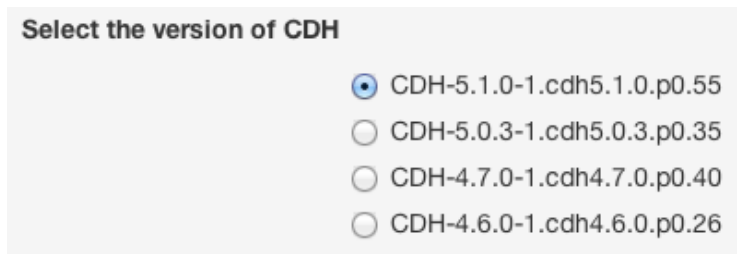
Follow the documentation to install and configure Cloudera Manager 5 with CDH. During the installation, you choose compatible CDH and Accumulo parcels while following these instructions.

1. Be sure the **Use Parcels** option is checked.



Choose Method  Use Packages   Use Parcels (Recommended) 

2. Select version **CDH-5.1.0-1.cdh5.1.0.p0.55** or higher for the CDH parcel.



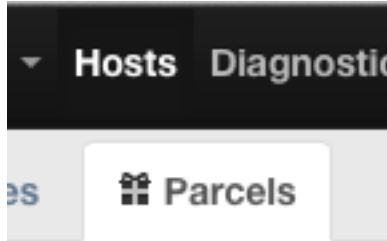
Select the version of CDH

- CDH-5.1.0-1.cdh5.1.0.p0.55
- CDH-5.0.3-1.cdh5.0.3.p0.35
- CDH-4.7.0-1.cdh4.7.0.p0.40
- CDH-4.6.0-1.cdh4.6.0.p0.26

3. Click **Continue** and follow the rest of the installation steps as described in the [documentation](#). Accumulo requires that you set up HDFS and Zookeeper. Other services are optional.

### Step 2: Install the Accumulo Parcel

1. From the **Hosts** tab, select **Parcels**.



2. Under the parcel entry for **ACCUMULO 1.6.0-1.cdh5.1.4.p0.116** or later, click **Download**.



3. Under the cluster you want to install on (for example, Cluster 1), find the Accumulo parcel and click **Distribute**.
4. Under the cluster you want to install on (for example, Cluster 1), find the Accumulo parcel and click **Activate**.

You will be prompted to restart the cluster. Because the Accumulo parcel was not previously in use, you can safely skip this step and click **Close**.

### Step 3: Configure HDFS

Cloudera strongly recommends that you establish an HDFS Name Service on the cluster that will run Accumulo. Due to the way Accumulo manages files within HDFS, doing so greatly reduces administrative tasks in the future if a NameNode needs to be replaced or moved. To set up an HDFS Name Service, follow [the instructions for enabling HDFS High Availability](#).

To guard against data loss, you must configure HDFS to durably write data on file close. If the following configuration changes are not made, Accumulo issues warning messages until the problem is corrected.

1. Navigate to your cluster's **HDFS service** page.

2. Click the "**Configuration**" tab.
3. Search for "hdfs-site.xml".
4. Search for the Service-Wide / Advanced section's property for "**HDFS Service Advanced Configuration Snippet (Safety Valve) for hdfs-site.xml**".
5. Click the field and add this snippet.

```
<property>
  <name>dfs.datanode.synconclose</name>
  <value>>true</value>
</property>
```

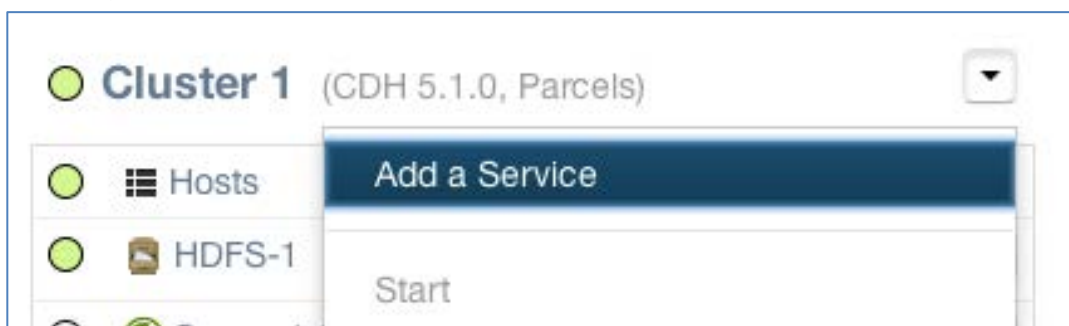
6. Search for any Gateway group properties labeled "**HDFS Client Advanced Configuration Snippet (Safety Valve) for hdfs-site.xml**".
7. Click the field and add this snippet.

```
<property>
  <name>dfs.datanode.synconclose</name>
  <value>>true</value>
</property>
```

8. Save your changes with a descriptive message, such as "HDFS changes for Accumulo."
9. Redeploy client configurations for the HDFS service.
10. Restart the HDFS service.

### Step 4: Add the Accumulo Service

1. Navigate to the Cloudera Manager **Home** page.
2. Click the actions menu for the cluster you want to add the Accumulo service to (for example, Cluster 1), and select **Add a Service**.



3. Select **Accumulo 1.6** and click **Continue**.
4. Select the dependent services and click **Continue**.
5. Assign the Accumulo roles to the hosts in your cluster. Cloudera recommends assigning a **Tablet Server** role on each host that is assigned the **DataNode** role. The **Monitor**, **Garbage Collector**, **Tracer**, and **Master** roles should all be assigned to **non-DataNodes**. The **Gateway** role should be assigned to any hosts where you want to use Accumulo that do not already have other Accumulo roles assigned.



6. Click **Continue**.
7. Configure the **Accumulo Instance Secret**. It is strongly recommended that you protect this secret, because the security of the Accumulo service relies on it.
8. Configure the **Accumulo Instance Name**.
9. Configure the **Trace User** and **Trace Password** settings. It is strongly recommended that you **not** leave the **Trace User** set to **root**, because this is not a secure configuration.
10. Click **Continue**.
11. Wait while Cloudera Manager does initial service set up.
12. Click **Continue**.
13. Click **Finish**.

Verify your installation by following the instructions in the Test the Accumulo Shell section.

### Step 5: Optional Configuration for Performance Tuning – Relaxing WAL Durability

When using the `BatchWriter` API to ingest data into Accumulo, you can attain greater write throughput at the cost of reduced data durability. Cloudera recommends using this setting only in environments with reliable UPS.

To enable this setting, perform the following configuration changes:

1. Navigate to your cluster's **Accumulo 1.6 service** page.
2. Click "**Configuration**" and then click "**View and Edit**".
3. Search for "Tablet Server accumulo-site.xml".
4. Find the Tablet Server Default Group / Advanced section's property for "**Tablet Server Advanced Configuration Snippet (Safety Valve) for accumulo-site.xml**".
5. Click the field and add the following snippet.

```
<property>
  <name>tserver.wal.sync.method</name>
  <value>hflush</value>
</property>
```

6. Save your changes with a descriptive message, such as "Accumulo WAL Durability Changes".
7. Redeploy client configurations.
8. Restart the Tablet Servers in your Accumulo 1.6 Service.

## Install Apache Accumulo from Distribution Packages

This section describes how to install Cloudera's packaging of Accumulo from packages (RPM or DEB) instead of using Cloudera Manager.

## Step 1: Add or Build the Accumulo Repository

- If you are installing Accumulo on a Red Hat system, you can download the Cloudera packages using yum or your web browser.
- If you are installing Accumulo on a SLES system, you can download the Cloudera packages using zypper, YaST, or your web browser.
- If you are installing Accumulo on an Ubuntu or Debian system, you can download the Cloudera packages using apt or your web browser.

### On Red Hat-compatible Systems

Use one of the following methods to add or build the Accumulo repository or download the packages on Red Hat-compatible systems by using the instructions in one of the following sections:

- Adding the Accumulo
- Building a yum

Do this on all systems in the cluster.

### Adding the Accumulo Repository

Follow the link in the table below that matches your Red Hat or CentOS system, navigate to the repo file for your system, and save it in the `/etc/yum.repos.d/` directory.

| For OS Version          | CDH Version | Follow this Link                                  |
|-------------------------|-------------|---|
| Red Hat/CentOS/Oracle 5 | CDH 4       | <a href="#">CDH 4 for Red Hat/CentOS/Oracle 5</a> |
|                         | CDH 5       | <a href="#">CDH 5 for Red Hat/CentOS/Oracle 5</a> |
| Red Hat/CentOS/Oracle 6 | CDH 4       | <a href="#">CDH 4 for Red Hat/CentOS/Oracle 6</a> |
|                         | CDH 5       | <a href="#">CDH 5 for Red Hat/CentOS/Oracle 6</a> |

Now continue with Step 2: Install Accumulo.

### Building a yum Repository

If you want to create your own yum repository, download the appropriate repo file, create the repo, distribute the repo file, and set up a web server, as described under Creating a Local yum Repository.

### On SLES Systems

Use one of the following methods to download the Accumulo repository or packages on SLES systems by using the instructions in one of the following sections:

- Adding the Accumulo
- Building a SLES

Do this on all systems in the cluster.

## Adding the Accumulo Repository

1. Run the command that corresponds to your CDH version:

CDH 5

```
$ sudo zypper addrepo -f http://archive.cloudera.com/accumulo-c5/sles/11/x86\_64/cdh/cloudera-accumulo.repo
```

CDH 4

```
$ sudo zypper addrepo -f http://archive.cloudera.com/accumulo/sles/11/x86\_64/cdh/cloudera-accumulo.repo
```

2. Update your system package index by running:

```
$ sudo zypper refresh
```

## Building a SLES Repository

If you want to create your own SLES repository, create a mirror of [the Accumulo SLES directory](#) by following [these instructions](#) that explain how to create a SLES repository from the mirror.

Now continue with Step 2: Install Accumulo.

## On Ubuntu or Debian Systems

Use one of the following methods to add or build the Accumulo repository or download the packages on Ubuntu or Debian systems by using the instructions in one of the following sections::

- Adding the Accumulo Repository
- Building a Debian

Do this on all the systems in the cluster.

## Adding the Accumulo Repository

Create a new file `/etc/apt/sources.list.d/cloudera-accumulo.list` with the following contents:

- CDH 5 For Ubuntu systems:

```
deb [arch=amd64] http://archive.cloudera.com/accumulo-c5/<OS-release-arch> <RELEASE>-cdh5 contrib  
deb-src http://archive.cloudera.com/accumulo-c5/<OS-release-arch>
```

## Install Apache Accumulo from Distribution Packages

```
<RELEASE>-cdh5 contrib
```

- CDH 4 For Ubuntu systems:

```
deb [arch=amd64] http://archive.cloudera.com/accumulo/<OS-release-arch> <RELEASE>-cdh4 contrib
deb-src http://archive.cloudera.com/accumulo/<OS-release-arch> <RELEASE>-cdh4 contrib
```

- CDH 5 For Debian systems:

```
deb http://archive.cloudera.com/accumulo-c5/<OS-release-arch> <RELEASE>-cdh5 contrib
deb-src http://archive.cloudera.com/accumulo-c5/<OS-release-arch> <RELEASE>-cdh5 contrib
```

- CDH 4 For Debian systems:

```
deb http://archive.cloudera.com/accumulo/<OS-release-arch> <RELEASE>-cdh4 contrib
deb-src http://archive.cloudera.com/accumulo/<OS-release-arch> <RELEASE>-cdh4 contrib
```

where: <OS-release-arch> is debian/squeeze/amd64/cdh (CDH4 only), debian/wheezy/amd64/cdh (CDH5 only), ubuntu/lucid/amd64/cdh, ubuntu/precise/amd64/cdh, or ubuntu/trusty/amd64/cdh (CDH5 only) and <RELEASE> is the name of your distribution, which you can find by running `lsb_release -c`.

For example, to install Accumulo on top of CDH 5 for 64-bit Ubuntu Lucid:

```
deb [arch=amd64] http://archive.cloudera.com/accumulo-c5/ubuntu/lucid/amd64/cdh lucid-cdh5 contrib
deb-src http://archive.cloudera.com/accumulo-c5/ubuntu/lucid/amd64/cdh lucid-cdh5 contrib
```

### Building a Debian Repository

If you want to create your own apt repository, create a mirror of [the Accumulo Debian directory](#) and then [create an apt repository from the mirror](#).

Now continue with Step 2: Install Accumulo.

## Step 2: Install Accumulo

**Important:**

Before proceeding, you need to decide where to deploy the Accumulo Master, Accumulo Monitor, Accumulo Garbage Collector, and Accumulo Tracer daemons. As a general rule:

- The Accumulo Master and Accumulo Monitor run on the same "master" host unless the cluster is large (more than a few tens of nodes), and the master host (or hosts) should not run the Accumulo TabletServer service.
- In a large cluster, it is especially important that the Accumulo Garbage Collector and Accumulo Tracer run on machines separate from the Accumulo Master.
- Each node in the cluster **except the master host(s)** should run the Accumulo TabletServer service. In particular, these services should be run on every DataNode.

1. Depending on your needs, install and deploy either CDH 5 or CDH 4. Follow instructions under the documentation for [CDH 5 Installation](#) or [CDH 4 Installation](#) as appropriate.
2. Install and deploy ZooKeeper. Follow the instructions for [CDH 5](#) or [CDH 4](#) as appropriate.
3. Install each type of daemon package on the appropriate systems(s), as follows:

| Where to install                      | Install commands  |
|---------------------------------------|---|
| <b>Accumulo Master</b> host running:  |   |
| <i>Red Hat/CentOS compatible</i>      | sudo yum clean all; sudo yum install accumulo-master          |
| <i>SLES</i>                           | sudo zypper clean --all; sudo zypper install accumulo-master  |
| <i>Ubuntu or Debian</i>               | sudo apt-get update; sudo apt-get install accumulo-master     |
| <b>Accumulo Monitor</b> host running: |   |
| <i>Red Hat/CentOS compatible</i>      | sudo yum clean all; sudo yum install accumulo-monitor         |
| <i>SLES</i>                           | sudo zypper clean --all; sudo zypper install accumulo-monitor |
| <i>Ubuntu or Debian</i>               | sudo apt-get update; sudo apt-get install accumulo-monitor    |

## Install Apache Accumulo from Distribution Packages

|   |  |
|---|--|
| <b>Accumulo Garbage Collector</b> host running:   |  |
| <i>Red Hat/CentOS compatible</i>  | sudo yum clean all; sudo yum install accumulo-gc             |
| <i>SLES</i>   | sudo zypper clean –all; sudo zypper install accumulo-gc      |
| <i>Ubuntu or Debian</i>   | sudo apt-get update; sudo apt-get install accumulo-gc        |
| <b>Accumulo Tracer</b> host running:  |  |
| <i>Red Hat/CentOS compatible</i>  | sudo yum clean all; sudo yum install accumulo-tracer         |
| <i>SLES</i>   | sudo zypper clean –all; sudo zypper install accumulo-tracer  |
| <i>Ubuntu or Debian</i>   | sudo apt-get update; sudo apt-get install accumulo-tracer    |
| <b>All cluster hosts except Accumulo Master, Accumulo Monitor, Accumulo Garbage Collector, and Accumulo Tracer</b> hosts running: |  |
| <i>Red Hat/CentOS compatible</i>  | sudo yum clean all; sudo yum install accumulo-tserver        |
| <i>SLES</i>   | sudo zypper clean –all; sudo zypper install accumulo-tserver |
| <i>Ubuntu or Debian</i>   | sudo apt-get update; sudo apt-get install accumulo-tserver   |
| <b>All client hosts</b> running:  |  |
| <i>Red Hat/CentOS compatible</i>  | sudo yum clean all; sudo yum install accumulo                |
| <i>SLES</i>   | sudo zypper clean –all; sudo zypper install accumulo         |
| <i>Ubuntu or Debian</i>   | sudo apt-get update; sudo apt-get install accumulo           |

## Step 3: Configure HDFS

Cloudera strongly recommends that you establish an HDFS Name Service on the cluster that will run Accumulo. Due to the way Accumulo manages files within HDFS, doing so greatly reduces administrative tasks in the future if a NameNode needs to be replaced or moved. To set up an HDFS Name Service, follow [the instructions for enabling HDFS High Availability](#).

To guard against data loss, you must configure HDFS to durably write data on file close. If the following configuration changes are not made, Accumulo issues warning messages until the problem is corrected.

1. Edit the `hdfs-site.xml` used in your cluster and ensure it contains the following snippet:

```
<property>
  <name>dfs.datanode.synconclose</name>
  <value>>true</value>
</property>
```

2. Synchronize the updated `hdfs-site.xml` file across your cluster.
3. Restart all HDFS DataNodes.

## Step 4: Configure Accumulo for Your Environment

After installation, follow the steps in this section to configure Accumulo for your environment.

1. On every host, configure the following properties in `/etc/accumulo/conf/accumulo-site.xml` with the proper values for your environment:

```
<property>
  <name>instance.zookeeper.host</name>
  <value>localhost:2181</value>
  <description>comma separated list of zookeeper
servers</description>
</property>

<property>
  <name>instance.secret</name>
  <value>DEFAULT</value>
  <description>A secret unique to a given instance that all servers
must know in order to communicate with one another. Change it before
initialization. To change it later use
  ./bin/accumulo org.apache.accumulo.server.util.ChangeSecret
[oldpasswd] [newpasswd],
  and then update this file.
  </description>
</property>

<property>
  <name>tserver.memory.maps.max</name>
  <value>256M</value>
```

```
</property>

<property>
  <name>tserver.cache.data.size</name>
  <value>15M</value>
</property>

<property>
  <name>tserver.cache.index.size</name>
  <value>40M</value>
</property>

<property>
  <name>trace.password</name>
  <value>trace</value>
</property>

<property>
  <name>trace.user</name>
  <value>trace</value>
</property>
```

2. Review the configured values. For example, verify that you changed the value for `instance.secret`.
3. Review the service specific options, such as Java heap size, in the `/etc/default/accumulo` file:

```
ACCUMULO_TSERVER_OPTS="-Xmx1g -Xms1g -XX:NewSize=500m -
XX:MaxNewSize=500m"
ACCUMULO_MASTER_OPTS="-Xmx2g -Xms1g"
ACCUMULO_MONITOR_OPTS="-Xmx2g -Xms256m"
ACCUMULO_GC_OPTS="-Xmx256m -Xms256m"
ACCUMULO_GENERAL_OPTS="-XX:+UseConcMarkSweepGC -
XX:CMSInitiatingOccupancyFraction=75"
ACCUMULO_OTHER_OPTS="-Xmx1g -Xms256m"
```

### Important:

On a multi-host cluster, replace `localhost` with the fully qualified domain name (FQDN) or IP address of the Accumulo Master in the `masters`, `monitor`, `gc` and `tracers` files in `/etc/accumulo/conf`, and add the FQDN or IP address of the TabletServers (one per line) to the `/etc/accumulo/conf/slaves` file.



**Important:**

On a multi-host cluster, the contents of the `/etc/accumulo/conf` directory must always be synchronized across all Accumulo servers within a cluster. This can be done using configuration management, version control, or via a utility such as `rsync`. Servers with out-of-sync configurations will not be allowed to join the cluster.

## Step 5: Initialize Accumulo

To initialize Accumulo:

1. Create the `/accumulo` and `/user/accumulo` directories in HDFS and change their ownership to the `accumulo` user:

```
$ sudo su - hdfs
$ hadoop fs -mkdir /accumulo /user/accumulo
$ hadoop fs -chown accumulo:supergroup /accumulo /user/accumulo
$ hadoop fs -chmod 751 /accumulo
$ hadoop fs -chmod 750 /user/accumulo
$ exit
```

2. On the Accumulo Master, enter the following commands to initialize Accumulo and follow the prompts to name your instance (for this example, `cloudera`) and set a root password:

```
$ sudo -i service accumulo-master init
[util.Initialize] INFO : Hadoop Filesystem is
hdfs://localhost.localdomain:8020
[util.Initialize] INFO : Accumulo data dir is /accumulo
[util.Initialize] INFO : Zookeeper server is localhost:2181
[util.Initialize] INFO : Checking if Zookeeper is available. If this
hangs, then you need to make sure zookeeper is running

Instance name : cloudera
Enter initial password for root: ****
Confirm initial password for root: ****

[conf.Configuration] WARN : dfs.replication.min is deprecated. Instead,
use dfs.namenode.replication.min
[conf.Configuration] WARN : dfs.block.size is deprecated. Instead, use
dfs.blocksize
[security.ZKAuthenticator] INFO : Initialized root user with username:
root at the request of user !SYSTEM
```

**Warnings:**

You will be warned here if you did not change your instance secret in

```
/etc/accumulo/conf/accumulo-site.xml.
```

If the "Hadoop Filesystem is" line contains "file://" or "fs://" instead of "hdfs://", HDFS is not properly configured.

## Step 6: Start Accumulo

To start Accumulo:

1. Run the following commands on the following hosts:

| For the following service  | Run this command                                    |
|--|---|
| <b>Accumulo Master</b>   | <code>sudo -i service accumulo-master start</code>  |
| <b>Accumulo Monitor</b>  | <code>sudo -i service accumulo-monitor start</code> |
| <b>Accumulo Garbage Collector</b>  | <code>sudo -i service accumulo-gc start</code>      |
| <b>Accumulo Tracer</b>   | <code>sudo -i service accumulo-tracer start</code>  |
| <b>All cluster hosts except Accumulo Master, Accumulo Monitor, Accumulo Garbage Collector, and Accumulo Tracer hosts</b> | <code>sudo -i service accumulo-tserver start</code> |

2. Connect to Accumulo on <http://localhost:50095>. You can check the status of each daemon with the following command:

```
$ sudo -i service accumulo-<service> status
```

where <service> is one of master, monitor, gc, tracer, or tserver.

3. You can stop each daemon with the following command:

```
$ sudo -i service accumulo-<service> stop
```

where <service> is one of master, monitor, gc, tracer, or tserver.

Verify your installation by following the instructions in the [Test the Accumulo Shell](#) **Error! Reference source not found.** section.

### Step 7: Optional Configuration for Performance Tuning – Relaxing WAL Durability

When using the `BatchWriter` API to ingest data into Accumulo, greater write throughput can be attained at the cost of reduced data durability. Cloudera recommends using this setting only in environments with reliable UPS.

To enable this setting, add the following text to the "accumulo-site.xml" file and distribute the change across hosts running Accumulo roles in the cluster:

```
<property>
  <name>tserver.wal.sync.method</name>
  <value>hflush</value>
</property>
```

You should restart all Tablet Servers following this change.

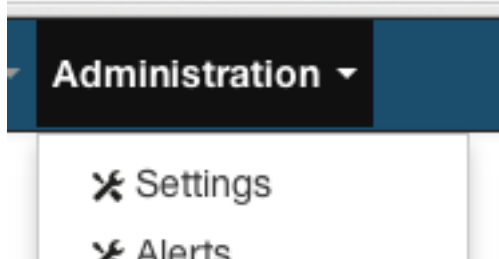
## Configuring Cloudera Manager for Accumulo 1.6 on CDH 4

Cloudera Manager can manage an Accumulo 1.6 service on CDH 4. Follow these setup steps and then continue at Step 3: Configure HDFS in Install Apache Accumulo using Cloudera Manager.

### Step 1: Add the Remote Parcel Repository for Accumulo 1.6.0-cdh4.6.0

Cloudera Manager comes with repository information for running Accumulo 1.6.0 on CDH 5. To run on CDH 4, you will need to add another remote parcel repository.

1. Click the **Administration** menu and select **Settings**.

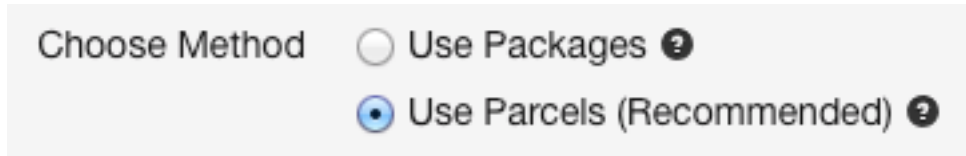


2. Select the **Parcels** category.
3. Edit the configuration item for **Remote Parcel Repository URLs**.
4. Click the + next to the last parcel repo to add a new URL.
5. Paste the following URL:  
<http://archive.cloudera.com/accumulo/parcels/latest/>
6. Click **Save Changes**.

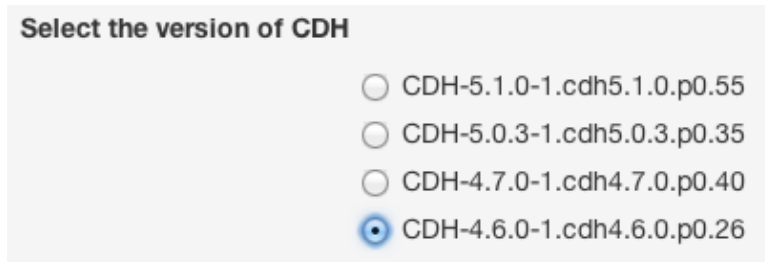
### Step 2: Install and Configure Cloudera Manager 5 and CDH 4.6.0

Follow the documentation to install and configure Cloudera Manager 5 with CDH 4.6.0. Be sure to select parcels for cluster components and choose an appropriate CDH 4 version.

1. Be sure the **Use Parcels** option is selected.



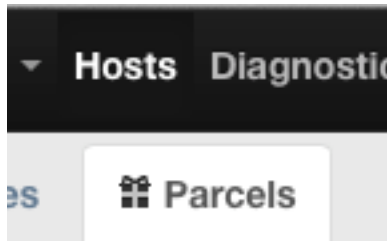
2. Select version **CDH-4.6.0-1.cdh4.6.0.p0.26** for the CDH parcel.



3. Click **Continue** and follow the rest of the installation steps as described in the [documentation](#). Accumulo requires that you set up HDFS and Zookeeper. Other services are optional.

### Step 3: Install the Accumulo Parcel

1. From the **Hosts** tab, select **Parcels**.



2. Under the **ACCUMULO 1.6.0-1.cdh4.6.0.p0.261** parcel, click **Download**.



3. Under the cluster you want to install on (for example, Cluster 1), find the Accumulo parcel and click **Distribute**.
4. Under the cluster you want to install on (for example, Cluster 1), find the Accumulo parcel and click **Activate**.

5. You will be prompted to restart the cluster. Because the Accumulo parcel was not previously in use, you can safely skip this step and click **Close**.

Now continue with the instructions in Step 3: Configure HDFS. Install Apache Accumulo using Cloudera Manager.

### Upgrading from Accumulo 1.4

If you have a cluster currently running Cloudera's packaging for Accumulo 1.4 on CDH 4, the following instructions will guide you through upgrading to Cloudera's packaging for Accumulo 1.6 on CDH 4. You should plan for down time; the Accumulo service must be stopped as a part of the upgrade.

### Upgrading with Cloudera Manager

1. Upgrade your Cloudera Manager to 5.1.0 or later, using [the instructions for updating to the latest Cloudera Manager version](#).
2. Follow the instructions in Step 1: Add the Remote Parcel Repository for Accumulo 1.6.0-cdh4.6.0 in the section on Configuring Cloudera Manager for Accumulo 1.6 on CDH 4.
3. Stop Accumulo 1.4 service in Cloudera Manager.
4. Take note of your configuration settings. In particular, you will need the Accumulo Instance Secret, Accumulo Instance Name, Logger WAL Directories, and the maximum heap sizes for both the Tablet Server and Logger roles.
5. Install the parcel for Accumulo 1.6.0-cdh4.6.0 according to Step 3: Install the Accumulo Parcel in the section on Configuring Cloudera Manager for Accumulo 1.6 on CDH 4.
6. Add Accumulo 1.6 service to your cluster.
7. In the role assignments page, you should have the same deployment as you had with 1.4. Specifically, the tablet servers should match *exactly* what they were in Accumulo 1.4.
8. In the "Review Changes" page in the wizard, set the properties to have the same values that they were in Accumulo 1.4. The **Accumulo Instance Secret** and **Accumulo Instance Name** must match exactly what they were in Accumulo 1.4.
9. The wizard will try to initialize Accumulo and fail. This is expected, because you already have an Accumulo instance defined. Cancel the wizard.
10. Browse to the Accumulo 1.6 service page.
11. Click on **Configuration** tab.
12. Search for "**logger**".
13. Click the entry for **Logger WAL Directories** and enter the same value that it was in Accumulo 1.4.
14. Search for "tserver\_max".
15. Click the value for **Tablet Server Max Heapsize**. Make sure it is set to the sum of the Accumulo 1.4 Tablet Server and Logger maximum heap sizes.

16. Enter a descriptive name for your changes, such as “Accumulo upgrade settings” and click **Save Changes**.
17. Run "**Upgrade Accumulo**" from the "Actions" menu on Accumulo 1.6 service page.
18. Start the Accumulo 1.6 service.
19. On starting, Accumulo will complete the upgrade process. This can take some time. Check the log files for the Master role for progress.
20. Run "Deploy Client Configuration" from the "Actions" menu on Accumulo 1.6 service page.
21. Once you have checked that Accumulo 1.6 is working properly, delete the Accumulo 1.4 service from your cluster.

### Upgrading via RPMs

If you are using Cloudera Manager to handle a package-based installation, follow the instructions above, substituting the Accumulo yum repository and package update steps in this section for the changes to parcels.

If you are manually managing a package-based installation, follow all of these instructions.

1. Verify that there are no outstanding FATE operations under Accumulo 1.4:  
`accumulo org.apache.accumulo.server.fate.Admin print`
2. Stop the Accumulo 1.4 cluster.
3. On every host, replace the yum repository for Accumulo according to the appropriate CDH 4 entry for your OS as seen in Step 1: Add or Build the Accumulo Repository from the section Install Apache Accumulo from Distribution Packages.
4. As root on every host, upgrade the Accumulo packages.

```
sudo yum clean all
sudo yum update 'accumulo-*
```

5. On every host that was running the Tablet Server or Logger roles, run a utility to rewrite Accumulo 1.4 WALs to HDFS in a format that Accumulo 1.6 can read.

```
sudo -u accumulo accumulo \
    org.apache.accumulo.tserver.log.LocalWALRecovery
```

6. Start the Accumulo 1.6 cluster.
7. On starting, Accumulo will complete the upgrade process. This can take some time. Check the log files for the Master role for progress.

### Upgrading from CDH 4 to CDH 5

Cloudera does not support running Accumulo 1.4 on CDH 5. If you are running Accumulo 1.4 on CDH 4, follow the instruction above to upgrade to Accumulo 1.6 before upgrading to CDH 5.

Before upgrading, you should plan for downtime. Cloudera does not support rolling upgrades from CDH 4 to CDH 5. Additionally, you will have to remove and reinstall the Accumulo service as a part of upgrading.

### Step 1: Document and Remove Accumulo 1.6 for CDH 4

1. Stop the Accumulo 1.6 service.
2. Go to the Accumulo 1.6 service page.
3. Click the **Configuration** tab.
4. Click through the configuration settings and write down any non-default settings.
5. Click the **Instances** tab.
6. Make a note of your role assignments.
7. Return to the Cloudera Manager home page.
8. Under actions for the Accumulo 1.6 Service, select "delete".

### Step 2a: Update Parcels and Upgrade CDH

If you are using parcels for your cluster, follow these instructions to update your CDH version.

1. Go to the "Hosts" and then "Parcels" page.
2. You should have an active parcel for "Accumulo 1.6.0-cdh4.6.0". You should "Deactivate", "Remove from Hosts", and "Delete" this parcel. After deactivating, CM will prompt you to restart services. Close this restart prompt without restarting.
3. You should see CDH version cdh5.1.0 in the list of downloadable parcels. "DOWNLOAD" and "DISTRIBUTE" this parcel.



4. Once the CDH 5.1.0 parcel is distributed, click the "Upgrade" button and follow the upgrade wizard according to [the documentation for parcels](#).
5. Follow the CDH 5 instructions found in Step 2: Install the Accumulo Parcel, above.

### Step 2b: Update Packages and Upgrade CDH

If you are using packages for your cluster, follow these instructions to update your CDH version.

1. Remove the Accumulo packages from all nodes in your cluster, using your system's package manager.
2. Remove the Accumulo repo from your package manager on all nodes in the cluster.
3. Upgrade CDH according to the [documentation for packages](#).
4. Follow the CDH 5 instructions found in Install Apache Accumulo from Distribution Packages, steps 1 and 2 above.

### Step 3: Add Accumulo 1.6 for CDH 5

1. Add the Accumulo 1.6 service to your cluster.
2. Assign roles to match those you wrote down in Step 1.
3. In the initial configuration, be sure the instance name and secret match those you wrote down in Step 1.
4. The wizard will try to initialize Accumulo and fail. This is expected, because you already have an Accumulo instance defined. Cancel the wizard.
5. Go to the Accumulo 1.6 Service page.
6. Click the **Configuration** tab.
7. Update the service configuration to match what you wrote down in Step 1.
8. Run "**Deploy Client Configuration**" from the "Actions" menu for the cluster on the CM Home page.
9. Start the Accumulo 1.6 service.

## Test the Accumulo Shell

You can now run the Accumulo shell on any client hosts (for CM installs, these are hosts assigned the **Gateway** role) in your cluster. By default, the user **root** is created and given the password **secret**. If you did not set a different password during install, Cloudera strongly recommends that you change the root user password.

The following steps will verify that the Accumulo shell works while allowing you to change the root user password.

1. Launch the Accumulo shell for the default root user.

```
$ accumulo shell -u root
Enter current password for 'root'@'accumulo': *****

Shell - Apache Accumulo Interactive Shell
-
- version: 1.6.0-cdh4.6.0
- instance name: accumulo
- instance id: 9863d1f1-c323-4671-9e1f-69857f0d635f
-
- type 'help' for a list of available commands
```



```
-
root@accumulo>
```

2. Use the `passwd` command to set a new password for the root user.

```
root@accumulo> passwd
Enter current password for 'root': *****
Enter new password for 'root': *****
Please confirm new password for 'root': *****
root@accumulo>
```

3. Relaunch the shell with this new password.

```
root@accumulo> exit
$ accumulo shell -u root
Enter current password for 'root'@'accumulo': *****

Shell - Apache Accumulo Interactive Shell
-
- version: 1.6.0-cdh4.6.0
- instance name: accumulo
- instance id: 9863d1f1-c323-4671-9e1f-69857f0d635f
-
- type 'help' for a list of available commands
-
root@accumulo>
```

4. Verify that you can list tables.

```
root@accumulo> tables
!METADATA
trace
root@accumulo>
```

5. If the `trace` table does not exist, make sure that you have created the `trace` user. Use the same password you used for the `trace.password` setting in `/etc/accumulo/conf/accumulo-site.xml` for a manually managed cluster or the **Trace Password** setting in Cloudera Manager installations.

```
root@cloudera> createuser trace
Enter new password for 'trace': *****
Please confirm new password for 'trace': *****
root@cloudera> grant System.CREATE_TABLE -s -u trace
root@cloudera> tables
!METADATA
```

```
trace
root@cloudera> revoke System.CREATE_TABLE -s -u trace
```

For more information on using the Accumulo shell, see the [Accumulo user manual](#).

## Using Sqoop 1 with Accumulo

CDH 4.6.0 and higher and CDH 5.1.0 and higher include Sqoop bindings for import/export of data with Accumulo. For instructions on invoking Sqoop with Accumulo as a source or sink, see [the Sqoop documentation](#).

When running the sqoop command, you may see warning messages about failing to create `/usr/lib/accumulo/logs`. These messages are safe to ignore.

## Sqoop 1 Client under CDH 5 and Cloudera Manager

To use Sqoop integration, you must perform the following configuration changes:

1. Navigate to your cluster's Sqoop 1 Client service page.
2. Click "**Configuration**".
3. Search for "sqoop-env.sh".
4. Look for the Gateway Default Group / Advanced section's property for **Sqoop 1 Client Client Advanced Configuration Snippet (Safety Valve) for sqoop-conf/sqoop-env.sh**.
5. Click the field and add the snippet appropriate for your installation, ensuring that each line ends with a bash comment hash, '#'.
  - For parcels:
 

```
export ACCUMULO_CONF_DIR=/etc/accumulo/conf #
export ACCUMULO_HOME=/opt/cloudera/parcels/ACCUMULO/lib/accumulo #
export HADOOP_CLIENT_HOME=/opt/cloudera/parcels/CDH/lib/hadoop/client #
export HADOOP_PREFIX=/opt/cloudera/parcels/CDH/lib/hadoop #
export ZOOKEEPER_HOME=/opt/cloudera/parcels/CDH/lib/zookeeper #
```
  - For packages:
 

```
export ACCUMULO_CONF_DIR=/etc/accumulo/conf #
export HADOOP_CLIENT_HOME=/usr/lib/hadoop/client #
export HADOOP_PREFIX=/usr/lib/hadoop #
export ZOOKEEPER_HOME=/usr/lib/zookeeper #
```
6. Save your changes with a descriptive message, such as "Sqoop changes for Accumulo."
7. Redeploy client configurations for the Sqoop 1 Client service

## Sqoop 1 under CDH 4 or without Cloudera Manager

To use Sqoop integration, you must perform the following configuration changes:

1. If you do not already have `/etc/sqoop/conf/sqoop-env.sh`, create it.

```
# cp /etc/sqoop/conf/sqoop-env-template.sh \
    /etc/sqoop/conf/sqoop-env.sh
```

2. Add the following exports to this `sqoop-env.sh` file; be sure they match your actual installation locations.

- For parcels:

```
export ACCUMULO_CONF_DIR=/etc/accumulo/conf
export ACCUMULO_HOME=/opt/cloudera/parcels/ACCUMULO/lib/accumulo
export HADOOP_CLIENT_HOME=/opt/cloudera/parcels/CDH/lib/hadoop/client
export HADOOP_PREFIX=/opt/cloudera/parcels/CDH/lib/hadoop
export ZOOKEEPER_HOME=/opt/cloudera/parcels/CDH/lib/zookeeper
```

- For packages:

```
export ACCUMULO_CONF_DIR=/etc/accumulo/conf
export HADOOP_CLIENT_HOME=/usr/lib/hadoop/client
export HADOOP_PREFIX=/usr/lib/hadoop
export ZOOKEEPER_HOME=/usr/lib/zookeeper
```

3. Save your changes.
4. Synchronize this file across all nodes that will run Sqoop commands.

## Using LZO Compression with Accumulo

Optionally, you may enable LZO compression for use with Accumulo. No special instructions are needed for an installation based on distribution packaging. If you are using parcels, follow these instructions:

1. [Install the GPLExtras Parcel](#).
2. Go to the HDFS service.
3. Select the **Configuration** tab.
4. Search for the **io.compression.codecs** property.
5. In the **Compression Codecs** entry, click in the field and click the + to add a new entry.
6. Add the following codecs for LZO.
  - `com.hadoop.compression.lzo.LzoCodec`
  - `com.hadoop.compression.lzo.LzopCodec`
7. Save the configuration for the HDFS Service.
8. Go to the Accumulo 1.6 service.
9. Select the **Configuration** tab.
10. Search for the **accumulo\_classpath** property.
11. In the **Additional Classpath** entry, click in the field (if needed, click + to add a new entry).
12. Add the following path that corresponds to the version you are using
  - (1.6.0-cdh4.6.0) `/opt/cloudera/parcels/HADOOP_LZO/lib/hadoop/lib/hadoop-lzo.jar`
  - (1.6.0-cdh5.1.0) `/opt/cloudera/parcels/GPLEXTRAS/lib/hadoop/lib/hadoop-lzo.jar`
13. Search for "service environment".
14. In the Service Wide / Advanced section's **Accumulo 1.6 Service Environment Advanced Configuration Snippet (Safety Valve)**, click the field.
15. Add the following line that corresponds to the version you are using:
 

```
(1.6.0-cdh4.6.0)
LD_LIBRARY_PATH=$LD_LIBRARY_PATH:/opt/cloudera/parcels/HADOOP_LZO/lib/hadoop/lib/native
```

(1.6.0-cdh5.1.0)

LD\_LIBRARY\_PATH=\$LD\_LIBRARY\_PATH:/opt/cloudera/parcels/GPLEXTRAS/lib/hadoop/lib/native

16. Search for the **accumulo\_java\_opts** property.
17. In the **Additional Java Options** entry, click the field.
18. Add to the end of the existing options:
 

```
-Dio.compression.codec.lzo.class=com.hadoop.compression.lzo.LzoCodec
```
19. Save the Accumulo 1.6 service with a description such as “Accumulo LZO”.
20. Redeploy client configurations for the cluster.
21. Restart the Accumulo 1.6 services.

If you have configured Cloudera Manager to install parcels in a location other than `/opt/cloudera/parcels/`, change the above paths to reflect that location.

## Using Accumulo with Maven

If you want to build applications or tools with Cloudera's packaging of Accumulo and you are using Maven or Ivy for dependency management, you can pull the Accumulo artifacts from the Cloudera Maven repository. The repository is available at <https://repository.cloudera.com/artifactory/cloudera-repos/>. The following is a sample snippet from a POM (pom.xml) file:

```
<repositories>
  <repository>
    <id>cloudera</id>
    <name>Cloudera Releases Repository</name>
    <url>https://repository.cloudera.com/artifactory/cloudera-repos/</url>
  </repository>
</repositories>
```

## CDH 5-Compatible Releases

### CDH 5.1-Compatible Releases

The following table lists the project name, groupId, artifactId, and version required to access each CDH 5-compatible artifact. Client applications should only require the accumulo-core artifact as a dependency and may need the accumulo-maven-plugin for running integration tests.

| Project  | groupId             | artifactId               | version        |
|----------|---------------------|--------------------------|----------------|
| Accumulo | org.apache.accumulo | accumulo                 | 1.6.0-cdh5.1.4 |
|          | org.apache.accumulo | accumulo-core            | 1.6.0-cdh5.1.4 |
|          | org.apache.accumulo | accumulo-examples-simple | 1.6.0-cdh5.1.4 |

|  |                     |                       |                |
|--|---------------------|-----------------------|----------------|
|  | org.apache.accumulo | accumulo-fate         | 1.6.0-cdh5.1.4 |
|  | org.apache.accumulo | accumulo-gc           | 1.6.0-cdh5.1.4 |
|  | org.apache.accumulo | accumulo-master       | 1.6.0-cdh5.1.4 |
|  | org.apache.accumulo | accumulo-maven-plugin | 1.6.0-cdh5.1.4 |
|  | org.apache.accumulo | accumulo-minicluster  | 1.6.0-cdh5.1.4 |
|  | org.apache.accumulo | accumulo-monitor      | 1.6.0-cdh5.1.4 |
|  | org.apache.accumulo | accumulo-proxy        | 1.6.0-cdh5.1.4 |
|  | org.apache.accumulo | accumulo-server-base  | 1.6.0-cdh5.1.4 |
|  | org.apache.accumulo | accumulo-start        | 1.6.0-cdh5.1.4 |
|  | org.apache.accumulo | accumulo-test         | 1.6.0-cdh5.1.4 |
|  | org.apache.accumulo | accumulo-trace        | 1.6.0-cdh5.1.4 |
|  | org.apache.accumulo | accumulo-tracer       | 1.6.0-cdh5.1.4 |
|  | org.apache.accumulo | accumulo-tserver      | 1.6.0-cdh5.1.4 |

## CDH 4-Compatible Releases

### CDH 4.6-Compatible Releases

The following table lists the project name, groupId, artifactId, and version required to access each CDH 4-compatible artifact. Client applications should only require the accumulo-core artifact as a dependency and may need the accumulo-maven-plugin for running integration tests.

| Project  | groupId             | artifactId               | version        |
|----------|---------------------|--------------------------|----------------|
| Accumulo | org.apache.accumulo | accumulo                 | 1.6.0-cdh4.6.0 |
|          | org.apache.accumulo | accumulo-core            | 1.6.0-cdh4.6.0 |
|          | org.apache.accumulo | accumulo-examples-simple | 1.6.0-cdh4.6.0 |
|          | org.apache.accumulo | accumulo-fate            | 1.6.0-cdh4.6.0 |

|                     |                       |                |
|---------------------|-----------------------|----------------|
| org.apache.accumulo | accumulo-gc           | 1.6.0-cdh4.6.0 |
| org.apache.accumulo | accumulo-master       | 1.6.0-cdh4.6.0 |
| org.apache.accumulo | accumulo-maven-plugin | 1.6.0-cdh4.6.0 |
| org.apache.accumulo | accumulo-minicluster  | 1.6.0-cdh4.6.0 |
| org.apache.accumulo | accumulo-monitor      | 1.6.0-cdh4.6.0 |
| org.apache.accumulo | accumulo-proxy        | 1.6.0-cdh4.6.0 |
| org.apache.accumulo | accumulo-server-base  | 1.6.0-cdh4.6.0 |
| org.apache.accumulo | accumulo-start        | 1.6.0-cdh4.6.0 |
| org.apache.accumulo | accumulo-test         | 1.6.0-cdh4.6.0 |
| org.apache.accumulo | accumulo-trace        | 1.6.0-cdh4.6.0 |
| org.apache.accumulo | accumulo-tracer       | 1.6.0-cdh4.6.0 |
| org.apache.accumulo | accumulo-tserver      | 1.6.0-cdh4.6.0 |

## Default Ports

If your cluster is running firewall software, you may need to allow communication between hosts on specific ports. The following table lists the default port for each server process and the configuration property used to change that value.

| Accumulo Process         | Port  | Property            |
|--------------------------|-------|---------------------|
| Garbage Collector        | 50091 | gc.port.client      |
| Master                   | 10010 | master.port.client  |
| Monitor (Log Forwarding) | 4560  | monitor.port.log4j  |
| Monitor (Client Port)    | 50091 | monitor.port.client |
| Tablet Server            | 10011 | tserver.port.client |

|        |       |                   |
|--------|-------|-------------------|
| Tracer | 12234 | trace.port.client |
|--------|-------|-------------------|

## Creating a Local yum Repository

This section explains how to set up a local yum repository that you can then use to install Accumulo on the machines in your cluster. There are a number of reasons you might want to do this; for example:

- The computers in your cluster may not have Internet access. You can still use yum to do an installation on those machines by creating a local yum repository.
- You may want to keep a stable local repository to ensure that any new installations (or re-installations on existing cluster members) use exactly the same bits.
- Using a local repository may be the most efficient way to distribute the software to cluster members.

To set up your own internal mirror, do the following.

### Note: Before You Start

These instructions assume you already have the appropriate Cloudera repo file on the system on which you are going to download the local repository. If this is not the case, follow the instructions in Adding the Accumulo .

1. On a computer that *does* have Internet access, install the `yum-utils` and `createrepo` packages if they are not already installed (`yum-utils` includes the `reposync` command):

```
$ sudo yum install yum-utils createrepo
```

2. On the same computer as the previous step, download the yum repository to a temporary location. On Red Hat/CentOS 6, you can use a command such as:

```
$ reposync -r cloudera-accumulo
```

### Note:

`cloudera-accumulo` is the name of the repository on your system; the name is in square brackets and usually is on the first line of the repo file, which in this example is `/etc/yum.repos.d/cloudera-accumulo.repo`.

3. Copy all of the RPMs to the machine that will serve the local repository and place them in a directory served by your web server. For this example, it is called `/var/www/html/accumulo/1.6.0/RPMS/x86_64` (or `i386` for 32-bit systems). Make sure

you can remotely access the files in the directory you just created (the URL should look like `http://<yourwebserver>/accumulo/1.6.0/RPMS/`).

4. On the server in step 3, go to `/var/www/html/accumulo/1.6.0/` and type the following command:

```
$ createrepo .
```

This will create or update the necessary metadata so yum can understand this new repository (you will see a new directory named `repodata`).

**Important:**

Check the permissions of the subdirectories under `/var/www/html/accumulo/1.6.0/`. Make sure they are all readable by your web server user.

5. Edit the repo file you downloaded previously and replace the line starting with `baseurl=` or `mirrorlist=` with `baseurl=http://<yourwebserver>/accumulo/1.6.0/`.
6. Save this modified repo file in `/etc/yum.repos.d/`, and check that you can install Accumulo through yum.

**Example:**

```
$ yum update && yum install accumulo
```

Once you have confirmed that your internal mirror works, you can distribute this modified repo file to all your machines, and they should all be able to install Accumulo without needing access to the Internet. Follow the instructions in Step 2: Install Accumulo.