Installing and configuring Cassandra on Linux

This page provides the steps you need to install Cassandra and configure the Cassandra node on your machine. You can also download the rpm package first and install Cassandra offline. Please note that Cassandra data and commit logs should not be located as per the default *cassandra.yaml* settings for a production environment, but rather in their own respective disks. We recommend mounting a disk as */data* and one as */logs.*

Note

- If a firewall is running on your machine, you need to open the following Cassandra client ports: **9042** and **9160**. See the detail here from this link.
- Make sure that both zone_reclaim_mode and swap are disabled. Failure to do so can cause severe performance issues. For detailed instructions on how to disable them, see this link.
- The Linux instructions provided on this page are for CentOS.

Installing Cassandra on Linux

How to install Cassandra on Linux

1. Check which version of Java is installed by running the following command:

   ```
   $ java -version
   ```

   Note
   Use Oracle JDK 1.8.0_151.

2. Add the Apache repository of Cassandra to */etc/yum.repos.d*

   ```
   $sudo vi /etc/yum.repos.d/cassandra.repo
   ```

   In this file, add the following lines for the Apache Cassandra repository:

   ```
   [cassandra]
   name = Apache Cassandra
   baseurl=https://www.apache.org/dist/cassandra/redhat/311x/
   gpgcheck=1
   repo_gpgcheck=1
   gpgkey=https://www.apache.org/dist/cassandra/KEYS
   ```

3. Install the packages by using the following command line:

   ```
   $sudo yum install cassandra
   ```

4. Make Cassandra starts automatically after reboot by typing the following.

   ```
   $chkconfig cassandra on
   ```

5. Configure Cassandra as follows.
   5.1 Locate the keys **- seeds**, **listen_address:**, and **broadcast_rpc_address:** in the file in */etc/cassandra/ conf/cassandra.yaml* (they are at different locations in the file).
   If, for example, the node's IP address was *10.1.1.123*, the following values would apply:

   Note
   The IP address 10.1.1.123 is just an example. You need to change it to the IP address of your server.

   - `seeds: "10.1.1.123"`
   - `listen_address: 10.1.1.123`
   - `broadcast_rpc_address: 10.1.1.123`

> **Warning**
>
> - There is a space before each IP address for parameters **listen_address** and **broadcast_rpc_address**. The space is required for Cassandra to start.

> **Tip**
>
> Cassandra nodes exchange information about one another using a mechanism called Gossip. A Seed is a node used as a Gossip contact point for information regarding ring topology. There must be one or more Seed elements for a working cluster.

name, for example, '#**broadcast_rpc_address: 10.1.1.123**', you will get an error after starting Cassandra.

5.2 Use the following keys to change the existing ones:

- `thrift_framed_transport_size_in_mb: 100`
- `commitlog_segment_size_in_mb: 128`
- `read_request_timeout_in_ms: 600000`
- `range_request_timeout_in_ms: 600000`
- `write_request_timeout_in_ms: 600000`
- `cas_contention_timeout_in_ms: 1000`
- `truncate_request_timeout_in_ms: 600000`
- `request_timeout_in_ms: 600000`
- `start_rpc: true`
- `rpc_address: 0.0.0.0`
- `batch_size_warn_threshold_in_kb: 3000`
- `batch_size_fail_threshold_in_kb: 5000`

5.3 Modify the data locations as per below:

```
data_file_directories:
  - /data/data
```

- `saved_caches_directory: /data/saved_caches`
- `commitlog_directory: /logs/commitlog`

6. Verify the installation of Cassandra.

6.1 When installed as above, you can start Cassandra using the following command:

```
$ sudo service cassandra start
```

6.2 Issue the following command to verify that Cassandra is ready:

```
$ tail /var/log/cassandra/cassandra.log
```

6.3 Verify that it contains lines similar to the following:

```
INFO 15:51:58,644 Node/10.1.1.123 state jump to normal
INFO 15:51:58,650 Waiting for gossip to settle before accepting client requests...
INFO 15:52:06,650 No gossip backlog; proceeding
```

> **Tip**
>
> If you get an out of memory error when starting Cassandra, you need to increase the Java stack size. The instructions for increasing the stack size are given in the section Starting Cassandra on Linux.

6. Verify that Cassandra is running:

```
$ nodetool status
```

```
Datacenter: datacenter1
=======================
Status=Up/Down
|/ State=Normal/Leaving/Joining/Moving
--  Address     Load       Tokens  Owns    Host ID                               Ra
UN  10.1.1.123  71.75 MB   256     51.4%   336308f4-27d2-4805-9da5-09f7515dd6c8  ra
```

Cassandra status.

Installing Cassandra offline

How to install Cassandra using a predownloaded rpm package (we have duplication of data - all of the post-install configuration can be moved to a common section. Also, if the package is downloaded, the sudo yum localinstall can be used)

---

1. Check which version of Java is installed by running the following command:

   ```
   $ java -version
   ```

   Note
   Use Oracle JDK 1.8.0_151.

2. Download the rpm package of Cassandra 3.11.2 from https://www.apache.org/dist/cassandra/redhat/311x/cassandra-3.11.2-1.noarch.rpm
3. Install the package using the following command line

   ```
   $ sudo rpm -ivh cassandra-3.11.2-1.noarch.rpm
   ```

   If you already have an older version of cassandra22 installed, use the following command instead.

   ```
   $ sudo rpm -Uvh cassandra-3.11.2-1.noarch.rpm
   ```

4. Configure Cassandra:
   4.1 Locate the keys **- seeds**, **listen_address:**, and **broadcast_rpc_address:** in the file in */etc/cassandra/ conf/cassandra.yaml* (they are at different locations in the file). If, for example, the node's IP address was *10.1.1.123*, the following values would apply:

   Note
   The IP address 10.1.1.123 is just an example. You need to change it to the IP address of your server.

   - `seeds: "10.1.1.123"`
   - `listen_address: 10.1.1.123`
   - `broadcast_rpc_address: 10.1.1.123`

   Warning
   - There is a space before each IP address for parameters **listen_address** and **broadcast_rpc_address**. The space is required for Cassandra to start.

   Tip
   Cassandra nodes exchange information about one another using a mechanism called Gossip. A Seed is a node used as a Gossip contact point for information regarding ring topology. There must be one or more Seed elements for a working cluster.

   before the parameter name, for example, <space>**#broadcast_rpc_address: 10.1.1.123**, you will get an error after starting Cassandra.

   4.2 Use the following keys' values to change the existing ones:

   - `thrift_framed_transport_size_in_mb: 100`
   - `commitlog_segment_size_in_mb: 128`
   - `read_request_timeout_in_ms: 600000`
   - `range_request_timeout_in_ms: 600000`
   - `write_request_timeout_in_ms: 600000`

- cas_contention_timeout_in_ms: 1000
- truncate_request_timeout_in_ms: 600000
- request_timeout_in_ms: 600000
- start_rpc: true
- rpc_address: 0.0.0.0
- batch_size_warn_threshold_in_kb: 3000
- batch_size_fail_threshold_in_kb: 5000

5. Verify the installation of Cassandra.
   5.1 When installed as above, you can start Cassandra using the following command:

   $ sudo service cassandra start

   5.2 Issue the following command to verify that Cassandra is ready.

   $ tail /var/log/cassandra/cassandra.log

   5.3 Verify that it contains lines similar to the following.

```
INFO 15:51:58,644 Node/10.1.1.123 state jump to normal
INFO 15:51:58,650 Waiting for gossip to settle before accepting client requests...
INFO 15:52:06,650 No gossip backlog; proceeding
```

> **Tip**
> If you get an out of memory error when starting Cassandra, you need to increase the Java stack size. The instructions for increasing the stack size are given in the section Starting Cassandra on Linux.

6. Verify the Cassandra status shows that it is running.

```
Datacenter: datacenter1
=======================
Status=Up/Down
|/ State=Normal/Leaving/Joining/Moving
--  Address     Load       Tokens  Owns   Host ID                               Rack
UN  10.1.1.123  71.75 MB   256     51.4%  336308f4-27d2-4805-9da5-09f7515dd6c8  rack1
```

Post installation configuration

Upon completion of the installation, we must edit **/etc/init.d/cassandra** to resolve the service control issue.

To edit **/etc/init.d/cassandra** to resolve the service control issue

---

1. Issue the following command
   `sudo nano /etc/init.d/cassandra`
2. Locate the line starting with

   `# chkconfig:`
3. Edit it to contain the following

   `# chkconfig: 2345 80 80`

   This will delay the execution to the appropriate point in time.
4. Next, locate the line starting with

   `CASSANDRA_PROG=/usr/sbin/cassandra`
5. Insert the following below the line:

```
#-------  Beginning of Centos7 modifications for startup script
# Note start priority changed from 20 to 80 in chkconfig definition
# create run dir for pid file
[ -d /var/run/cassandra ] || mkdir /var/run/cassandra
chown cassandra /var/run/cassandra
#------  End of Centos7 modifications for startup script
```

6. Save the file. Now we must add the service to the boot process:

   ```
   chkconfig --add cassandra
   ```

7. Now, proceed to edit */etc/cassandra/default.conf/cassandra.yaml*

   ```
   sudo nano /etc/cassandra/default.conf/cassandra.yaml
   ```

The first items we will be editing relate to the IP address of the Cassandra node and communications settings. In our diagram above, this IP address is **192.168.130.10**. You will need to search for 3 keys in the configuration file and modify them accordingly. The seeds parameter is a comma-delimited list containing all of the seeds in the Cassandra cluster. Since our cluster consists of only the single node, it contains only one entry - our IP address. The other 2 parameters contain the IP address on which Cassandra listens for connections and the IP address to broadcast to other Cassandra nodes in the cluster. The **broadcast_rpc_address** may be commented out using a # character. If so, remove the "#" and make sure there are no leading spaces.

Additionally, we need to set **rpc_address** to **0.0.0.0** (meaning, it will listen to rpc requests on all interfaces), and **start_rpc** to **true** (so it will process rpc requests).

- `seeds: "192.168.130.10"`
- `listen_address: 192.168.130.10`
- `broadcast_rpc_address: 192.168.130.10`
- `rpc_address: 0.0.0.0`
- `start_rpc: true`

The next set of parameters control thresholds to ensure that the data being sent is processed properly.

- `thrift_framed_transport_size_in_mb: 100`
- `commitlog_segment_size_in_mb: 128`
- `read_request_timeout_in_ms: 600000`
- `range_request_timeout_in_ms: 600000`
- `write_request_timeout_in_ms: 600000`
- `cas_contention_timeout_in_ms: 1000`
- `truncate_request_timeout_in_ms: 600000`
- `request_timeout_in_ms: 600000`
- `batch_size_warn_threshold_in_kb: 3000`
- `batch_size_fail_threshold_in_kb: 5000`

If you have installed your commit log in its own partition, the default commit log size will be the lesser of ¼ of the partition size or 8GB. In order to ensure that the recommended 8GB is used, you must uncomment the **commitlog_total_space_in_mb**, such that it will show as below. However, if you are uncommenting this value, please ensure that the partition has enough space to accommodate an 8GB commit log.

- `commitlog_total_space_in_mb: 8192`

The next step is to point the data to the new locations. There are 3 entries which will be modified: **data_file_directories**, **commitlog_directory**, and **saved_caches_directory**. Search for these keys and edit them as follows:

- `data_file_directories:`

  `- /data/data`
- `commitlog_directory: /logs/commitlog`
- `saved_caches_directory: /data/saved_caches`

After you have made these changes, save the **cassandra.yaml** file. Now, start the related services, as follows:

```
sudo service cassandra start
```

Now, proceed to check if Cassandra is running. To do this, issue the following command:

```
nodetool status
```

If the service is running, you will receive output such as below:

```
Datacenter: datacenter1
=======================
Status=Up/Down
```

```
|/ State=Normal/Leaving/Joining/Moving
--  Address    Load       Tokens   Owns (effective)  Host ID                               Rack
UN  127.0.0.1  128.4 KB   256      100.0%            ea3f99eb-c4ad-4d13-95a1-80aec71b750f  rack1
```

What's next?
Starting Cassandra on Linux

Related pages

- FAQs and troubleshooting