

# Configuring Cassandra Cluster

Cassandra

March 2023

## Contact

Jangsoo Lee

Technology Innovation Team, Samsung SDS

[jangsoo.lee@samsung.com](mailto:jangsoo.lee@samsung.com)

He is a former developer with a background in Opensource RDB, NoSQL, and is currently in charge of designing/developing SCP DB service and providing IT support.

## Revision History

| Version | Revision Date | Revised Details         |
|---------|---------------|-------------------------|
| 1.0     | March, 2023   | First version published |
|         |               |                         |
|         |               |                         |

# Contents

---

|    |                               |   |
|----|-------------------------------|---|
| 1. | OVERVIEW                      | 1 |
| 2. | INSTALLED CASSANDRA           | 1 |
| 3. | CONFIGURING CASSANDRA CLUSTER | 2 |
| 4. | STARTING CASSANDRA CLUSTER    | 4 |

# 1. Overview

Apache Cassandra is a distributed NoSQL database that can configure clusters using multiple nodes. Cassandra cluster distributes and stores data and assigns data areas for each node. You can adjust the size of the Cassandra cluster by adding and deleting nodes, and at this time, data domain for each node is automatically updated and data transfer process will begin.

## 2. Installed Cassandra

### 2.1 Cassandra installation information

Installed Cassandra is equipped with Apache Cassandra and OpenJDK, and OS account and systemd service for Cassandra are registered. To stop Cassandra, use the command `systemctl stop cassandra`, use `systemctl start cassandra` to start, and use `systemctl status cassandra` or `ps` command to check the Cassandra process.

#### [Cassandra Installation Information]

OS account: Cassandra  
Cassandra Engine Directory: /home/cassandra/apache-cassandra-4.0.6  
Cassandra version: 4.0.6  
OpenJDK version: 11.0.2  
OpenJDK Directory: /home/cassandra/jdk-11.0.2  
Cassandra system file path : /usr/lib/systemd/system/cassandra.service

### 2.2 Cassandra Configuration Information

Configuration of Apache Cassandra is based on the `Cassandra.yaml` file and once the setting has been changed, it will have to be restarted. For configuration file, go to `/home/cassandra/apache-cassandra-4.0.6/conf`. Users have to change the setting for cluster configuration as it is set to a single environment by default.

#### [cassandra.yaml basic setting information]

Seed : localhost:7000  
=> Node information for managing cluster status  
  
native\_port : 9042  
=> Port for client connection  
  
rpc\_address : localhost  
=> Client connection ip address  
  
listen\_address : localhost

=> ip address that communicates between nodes in the cluster

commitlog\_directory : /data/cassandra/commitlog

=> Cassandra computer storage path

data\_file\_directories : /data/cassandra/data

=> Cassandra data file storage path

## 2.3 JMX Setting and Log Path Enable

In order to check the metric information for Apache Cassandra, you need to enable JMX setting and it is currently set to 7199 port. The path for Cassandra log file is set to /home/cassandra/apache-cassandra-4.0.6/logs. The above information is available in the systemd file for Cassandra.

# 3. Configuring Cassandra Cluster

## 3.1 Creating Cassandra Cluster and Registering Security Group Rule

At least three nodes are required for Cassandra cluster configuration for high availability. The current guide configured clusters based on three nodes and added rules to the Security Group to allow communication between nodes. If the user applied setting to allow internal communication to another port, port number should be registered.

### [Applying for Installed Cassandra]

The screenshot shows the 'Virtual Server - Virtual Server Request' page. The 'Select image' section is active, with tabs for Standard, Custom, Migration, and DBMS. Under the DBMS tab, 'CASSANDRA' is selected. Below this, the 'Version select' section shows 'Image version' as 'Cassandra 4.0.6 + RHEL 8.5(64bit)'.

### [Security Group Rule Information (For Registration)]

Register the following rules for each IP in the new Cassandra

Inbound/Outbound TCP 7000, 7001, 9042

### Add rule ×

Direction ⓘ  
☒ Inbound rule ☐ Outbound rule

Target IP \* ⓘ  

192.168.20.10/32,192.168.20.11/32,192.168.20.12/32

Protocol  

TCPUDPICMPALL

Allowed ports \* ⓘ  

Direct input ▼ CUSTOM Add

TCP 

7000,7001,9042

UDP

ICMP

Description  

Enter within 100 characters.

0/100

Cancel

Confirm

## 3.2 Applying Cluster Settings

For smooth communication between Cassandras in the cluster, register the domain information for each IP and define the seed server. A seed server is responsible for checking the status of cluster and adding/deleting nodes. In the example, the Cassandra-001 server is defined as a seed server.

### [Registering Domain In /etc/hosts File]

```
192.168.20.10 cassandra-001
192.168.20.11 cassandra-002
192.168.20.12 cassandra-003
```

### [Configuring Cluster In cassandra.yaml File]

```
cassandra-001 Server Configuration
- Seeds:
  192.168.20.10
listen_address: 192.168.20.10
rpc_address: IP for client

cassandra-002 Server Configuration
```

Copyright 2023. Samsung SDS Inc. All rights reserved

3

```
- Seeds:
    192.168.20.10
listen_address: 192.168.20.11
rpc_address: IP for client

cassandra-003 Server Configuration
- Seeds:
    192.168.20.10
listen_address: 192.168.20.12
rpc_address: IP for client
```

## 4. Starting Cassandra Cluster

### 4.1 Initializing Cassandra

Initially, installed Cassandra was started for the purpose of single configuration. Therefore, Cassandra instance has to be stopped and initialize the data. Stop the Cassandra for each node and delete data files altogether.

#### [Stopping Cassandra and Initializing Data]

```
[root@cassandra-003 data]$ systemctl stop cassandra
[root@cassandra-003 data]$ rm -rf /data/cassandra/data/*

[root@cassandra-002 data]$ systemctl stop cassandra
[root@cassandra-002 data]$ rm -rf /data/cassandra/data/*

[root@cassandra-001 data]$ systemctl stop cassandra
[root@cassandra-001 data]$ rm -rf /data/cassandra/data/*
```

### 4.2 Starting Cassandra Sequentially

#### [Starting Cassandra]

```
[root@cassandra-001 data]$ systemctl start cassandra

[root@cassandra-002 data]$ systemctl start cassandra

[root@cassandra-003 data]$ systemctl start cassandra
```

### 4.3 Checking Cassandra Cluster Status

Check the status of Cassandra cluster and the system table to see if it is normal. If the node is abnormal, the nodetool status result does not appear as "Up."

#### [nodetool status]

```
[cassandra@cassandra-001 data]$ $CASSANDRA_HOME/bin/nodetool status
Status=Up/Down
State=Normal/Leaving/Joining/Moving
-- Address      Load          Tokens         Owns    Host ID                               Rack
UN 192.168.1.10  156.32 KB    256          ?      83asdfa3-dfa3-7adf-24kj-dafjk34lc903 rack1
UN 192.168.20.11 163.29 KB    256          ?      55jkdfa2-7dz9-dpq0-a716-9zkdmw8dfla2 rack1
UN 192.168.20.12 190.19 KB    256          ?      19daid0g-zdk4-44kd-00di-pozqdf934ka1 rack1
```

#### [cqlsh connection]

```
[root@cassandra-001 data]# $CASSANDRA_HOME/bin/cqlsh localhost
Python 2.7 support is deprecated. Install Python 3.6+ or set CQLSH_NO_WARN_PY2 to suppress this message.

Connected to Cluster at localhost:9042
[cqlsh 6.0.0 | Cassandra 4.0.6 | CQL spec 3.4.5 | Native protocol v5]
Use HELP for help.
cqlsh> desc keyspaces;

system      system_distributed  system_traces  system_virtual_schema
system_auth  system_schema       system_views
```