# **Kafka và ZookeeperIntroduction to Kafka and Zookeeper**

Cần tối thiểu 03 VM/Server và phải đảm bảo mở firewall cho các cổng 8080, 2888, 3888, 2181 và 9092 để cho phép kết nối giữa các máy trong group:

## Giới thiệu về Kafka

Kafka là một nền tảng truyền phát các sự kiện trực tuyến. Đây là một hệ thống phân tán bao gồm các máy chủ và máy khách giao tiếp qua giao thức mạng TCP hiệu suất cao. Nó có thể được triển khai trên phần cứng bare-metal, máy ảo và container trong môi trường tại chỗ cũng như môi trường đám mây.

## Giới thiệu về Zookeeper

Zookeeper là một dự án nguồn mở cung cấp dịch vụ tập trung để duy trì thông tin cấu hình, đặt tên và nhóm các dịch vụ lại với nhau. Nó được sử dụng bởi một cụm hoặc một nhóm các nút để chia sẻ dữ liệu. Kafka được xây dựng trên cơ sở sử dụng Zookeeper để điều phối các tác vụ.

Phần dưới đây sẽ hướng dẫn cách cài đặt cụm máy chủ 03 node Kafka, Zookeeper và một máy client;

# Setup 3 node Zookeeper Cluster

In this section, you will setup 3 Arm machines as a Zookeeper cluster. Each of the machines is referred to as a node.

## Node 1

Run the commands shown to download and install Zookeeper on node 1:

mkdir Zookeeper\_node1

cd Zookeeper\_node1

wget https://dlcdn.apache.org/zookeeper/zookeeper-3.9.2/apache-zookeeper-3.9.2-bin.tar.gz

tar -xzf apache-zookeeper-3.9.2-bin.tar.gz

cd apache-zookeeper-3.9.2-bin

Use a file editor of you choice and create a file named conf/zoo.cfg with the content shown below: Replace zk\_2\_ip and zk\_3\_ip with the IP addresses of the node 2 and node 3 respectively.

tickTime=2000

dataDir=/tmp/zookeeper

clientPort=2181

maxClientCnxns=60

initLimit=10

syncLimit=5

4lw.commands.whitelist=\*

server.1=0.0.0.0:2888:3888

server.2=zk\_2\_ip:2888:3888

server.3=zk\_3\_ip:2888:3888

### Create the zookeeper ID for this node:

mkdir /tmp/zookeeper

echo 1 >> /tmp/zookeeper/myid

### Start Zookeeper server on node 1:

bin/zkServer.sh start

The output from this command will look like:

/usr/bin/java

ZooKeeper JMX enabled by default

Using config: /home/ubuntu/apache-zookeeper-3.8.0-bin/bin/../conf/zoo.cfg

Starting zookeeper ... STARTED

## Node 2:

Run the following commands to download and install Zookeeper node 2:

mkdir Zookeeper\_node2

cd Zookeeper\_node2

wget https://dlcdn.apache.org/zookeeper/zookeeper-3.9.2/apache-zookeeper-3.9.2-bin.tar.gz

tar -xzf apache-zookeeper-3.9.2-bin.tar.gz

cd apache-zookeeper-3.9.2-bin

Use a file editor of you choice and create a file named conf/zoo.cfg with the content shown below: Replace zk\_1\_ip and zk\_3\_ip with the IP addresses of the node 1 and node 3 respectively.

tickTime=2000

dataDir=/tmp/zookeeper

clientPort=2181

maxClientCnxns=60

initLimit=10

syncLimit=5

4lw.commands.whitelist=\*

server.1=zk\_1\_ip:2888:3888

server.2=0.0.0.0:2888:3888

server.3=zk\_3\_ip:2888:3888

### Create the Zookeeper ID for this node:

mkdir /tmp/zookeeper

echo 2 >> /tmp/zookeeper/myid

### Start Zookeeper server on node 2:

bin/zkServer.sh start

The output from this command will look like:

/usr/bin/java

ZooKeeper JMX enabled by default

Using config: /home/ubuntu/apache-zookeeper-3.8.0-bin/bin/../conf/zoo.cfg

Starting zookeeper ... STARTED

## Node 3:

Run the following commands to download and install Zookeeper node 3:

mkdir Zookeeper\_node3

cd Zookeeper\_node3

wget https://dlcdn.apache.org/zookeeper/zookeeper-3.9.2/apache-zookeeper-3.9.2-bin.tar.gz

tar -xzf apache-zookeeper-3.9.2-bin.tar.gz

cd apache-zookeeper-3.9.2-bin

Use a file editor of you choice and create a file named conf/zoo.cfg with the content shown below: Replace zk\_1\_ip and zk\_2\_ip with the IP addresses of the node 1 and node 2 respectively.

tickTime=2000

dataDir=/tmp/zookeeper

clientPort=2181

maxClientCnxns=60

initLimit=10

syncLimit=5

4lw.commands.whitelist=\*

server.1=zk\_1\_ip:2888:3888

server.2=zk\_2\_ip:2888:3888

server.3=0.0.0.0:2888:3888

### Create the Zookeeper ID for this node:

mkdir /tmp/zookeeper

echo 3 >> /tmp/zookeeper/myid

### Start Zookeeper server on node 3:

bin/zkServer.sh start

The output from this command will look like:

/usr/bin/java

ZooKeeper JMX enabled by default

Using config: /home/ubuntu/apache-zookeeper-3.8.0-bin/bin/../conf/zoo.cfg

Starting zookeeper ... STARTED

## Verify the Zookeeper Cluster:

The Zookeeper server is now running on all 3 nodes.

Run the command below using each of the Zookeeper nodes IP address to check the leader/follower mode:

Replace zk\_1\_ip with the IP address of each of the Zookeeper nodes.

echo stat | nc zk\_1\_ip 2181

The example output from this command is shown:

Latency min/avg/max: 0/0.0/0

Received: 2

Sent: 1

Connections: 1

Outstanding: 0

Zxid: 0x100000000

Mode: leader

Node count: 5

Proposal sizes last/min/max: -1/-1/-1

### Start the Zookeeper CLI to connect to the cluster.

This command can be executed on any of the 3 Zookeeper nodes:

bin/zkCli.sh

### Write a message into the Zookeeper cluster:

create /FirstZnode "message written to database"

Output for this command:

[zk: localhost:2181(CONNECTED) 0] create /FirstZnode “message written to database”

Created /FirstZnode

You can read the previously written messages by running the command below from the Zookeeper CLI on any of the nodes:

get /FirstZnode

Output for this command:

[zk: localhost:2181(CONNECTED) 1] get /FirstZnode

message written to database

# Setup a 3 node Kafka Cluster

Ensure that the 3 Node Zookeeper cluster is running.

## Node 1:

Run the commands shown to download and setup Kafka on node 1:

mkdir kafka\_node1

cd kafka\_node1

wget https://dlcdn.apache.org/kafka/3.7.0/kafka\_2.13-3.7.0.tgz

tar -xzf kafka\_2.13-3.7.0.tgz

cd kafka\_2.13-3.7.0

Use a file editor of your choice and replace the contents in config/server.properties with the contents below:. Replace zk\_1\_ip,zk\_2\_ip and zk\_3\_ip with the IP addresses of the 3 Zookeeper nodes you setup.

broker.id=1

listeners=PLAINTEXT://:9092

log.dirs=/tmp/kafka-logs

zookeeper.connect=zk\_1\_ip:2181,zk\_2\_ip:2181,zk\_3\_ip:2181

### Create a directory for the log files:

mkdir /tmp/kafka-logs

### Start Kafka server on node 1:

bin/kafka-server-start.sh config/server.properties

## Node 2:

Run the commands shown to download and setup Kafka on node 2:

mkdir kafka\_node2

cd kafka\_node2

wget https://dlcdn.apache.org/kafka/3.7.0/kafka\_2.13-3.7.0.tgz

tar -xzf kafka\_2.13-3.7.0.tgz

cd kafka\_2.13-3.7.0

Use a file editor of your choice and replace the contents in config/server.properties with the contents below:. Replace zk\_1\_ip,zk\_2\_ip and zk\_3\_ip with the IP addresses of the 3 Zookeeper nodes you setup.

broker.id=2

listeners=PLAINTEXT://:9092

log.dirs=/tmp/kafka-logs

zookeeper.connect=zk\_1\_ip:2181,zk\_2\_ip:2181,zk\_3\_ip:2181

### Create a directory for the log files:

mkdir /tmp/kafka-logs

### Start Kafka server on node 2:

bin/kafka-server-start.sh config/server.properties

## Node 3:

Run the commands shown to download and setup Kafka on node 3:

mkdir kafka\_node3

cd kafka\_node3

wget https://dlcdn.apache.org/kafka/3.7.0/kafka\_2.13-3.7.0.tgz

tar -xzf kafka\_2.13-3.7.0.tgz

cd kafka\_2.13-3.7.0

Use a file editor of your choice and replace the contents in config/server.properties with the contents below:. Replace zk\_1\_ip,zk\_2\_ip and zk\_3\_ip with the IP addresses of the 3 Zookeeper nodes you setup.

broker.id=3

listeners=PLAINTEXT://:9092

log.dirs=/tmp/kafka-logs

zookeeper.connect=zk\_1\_ip:2181,zk\_2\_ip:2181,zk\_3\_ip:2181

### Create a directory for the log files:

mkdir /tmp/kafka-logs

### Start Kafka server on node 3:

bin/kafka-server-start.sh config/server.properties

# Verify that the Kafka cluster is working.

Sau khi thiết lập thành công cụm 3 node Kafka, tiếp theo cần kiểm tra để xác minh cụm này hoạt động OK hay không bằng cách tạo topic và lưu trữ các sự kiện. Thực hiện theo các bước dưới đây để tạo topic, viết một số sự kiện vào topic, sau đó đọc các sự kiện.

## Install Kafka on Client machine

Run the commands shown to download and setup Kafka on client machine:

mkdir kafka\_client

cd kafka\_client

wget https://dlcdn.apache.org/kafka/3.7.0/kafka\_2.13-3.7.0.tgz

tar -xzf kafka\_2.13-3.7.0.tgz

cd kafka\_2.13-3.7.0

## Create a topic

Open a terminal on the client machine and run the command shown below. Replace kf\_1\_ip, kf\_2\_ip and kf\_3\_ip with the IP addresses of the 3 nodes running the Kafka server.

./bin/kafka-topics.sh --create --topic test-topic --bootstrap-server kf\_1\_ip:9092,kf\_2\_ip:9092,kf\_3\_ip:9092 --replication-factor 3 --partitions 64

## Describe the topic created:

Run this command in the same client terminal where the topic was created. Replace kf\_1\_ip, kf\_2\_ip and kf\_3\_ip with the IP addresses of the 3 nodes running the Kafka server.

./bin/kafka-topics.sh --topic test-topic --bootstrap-server kf\_1\_ip:9092,kf\_2\_ip:9092,kf\_3\_ip:9092 --describe

The output from this command is shown below:

ubuntu@ip-172-31-19-179:~/kafka\_node/kafka\_2.13-3.2.3$ ./bin/kafka-topics.sh --topic test-topic --bootstrap-server 3.144.181.100:9092,3.15.19.197:9092,18.191.61.20:9092 --describe

Topic: test-topic TopicId: WMy9lruTQC6uuuuyep-C\_Q PartitionCount: 64 ReplicationFactor: 3 Configs: segment.bytes=1073741824

Topic: test-topic Partition: 0 Leader: 3 Replicas: 3,1,2 Isr: 3,1,2

Topic: test-topic Partition: 1 Leader: 1 Replicas: 1,2,3 Isr: 1,2,3

Topic: test-topic Partition: 2 Leader: 2 Replicas: 2,3,1 Isr: 2,3,1

## Run the producer client to write events into the created topic:

Run this command in the same client terminal where the topic was created. Replace kf\_1\_ip, kf\_2\_ip and kf\_3\_ip with the IP addresses of the 3 nodes running the Kafka server.

./bin/kafka-console-producer.sh --topic test-topic --bootstrap-server kf\_1\_ip:9092,kf\_2\_ip:9092,kf\_3\_ip:9092

Write a message, example shown below:

ubuntu@ip-172-31-19-179:~/kafka\_node/kafka\_2.13-3.2.3$ ./bin/kafka-console-producer.sh --topic test-topic --bootstrap-server 3.144.181.100:9092,3.15.19.197:9092,18.191.61.20:9092

>This is the first message written on producer

>

## Run the consumer client to read all the events created:

Open a new terminal on the client machine to run the consumer client. Replace kf\_1\_ip, kf\_2\_ip and kf\_3\_ip with the IP addresses of the 3 nodes running the Kafka server.

./bin/kafka-console-consumer.sh --topic test-topic --bootstrap-server kf\_1\_ip:9092,kf\_2\_ip:9092,kf\_3\_ip:9092

The same message you wrote in the producer client terminal should appear on the consumer client. Example shown below:

ubuntu@ip-172-31-19-179:~/kafka\_node/kafka\_2.13-3.2.3$ ./bin/kafka-console-consumer.sh --topic test-topic --bootstrap-server 3.144.181.100:9092,3.15.19.197:9092,18.191.61.20:9092

Đây là tin nhắn đầu tiên được viết trên producer

Bây giờ ta có thể xác minh rằng cụm Kafka đang hoạt động thành công.