

Cloud Computing and Cloud Native Systems

ECE 573

Project 3

Abhilash Kashyap Balasubramanyam

abalasubramanyam@hawk.iit.edu

A20566944

Fall 2024

Illinois Institute of Technology

Date: October 20, 2024

Table of Contents

Solutions.....	2
1 Cassandra Cluster Management.....	2
1.1 Question 1.....	2
1.2 Question 2.....	2
1.3 Question 3.....	3
1.4 Question 4.....	4
1.5 Screenshots of task completion.....	5
2 Tunable Consistency.....	9
2.1 Question 1.....	9
2.2 Question 2.....	11
2.3 Question 3.....	12
2.4 Question 4.....	13
2.5 Question 5.....	15
2.6 Question 6.....	16
3 Availability	18
3.1 Question 1.....	18
3.2 Question 2.....	19
3.3 Question 3.....	21
3.4 Question 4.....	23

List of Figures

Figure 1: docker-compose.yml file used for project 3.	5
Figure 2: Running docker compose build command.	6
Figure 3: Running docker compose create command.	6
Figure 4: Running docker compose start command.	7
Figure 5: Running docker compose ps command for status.	7
Figure 6: Running docker compose exec db1 nodetool status command.	8
Figure 7: Running docker compose exec db1 cqlsh command and exiting the CQL command line with exit command.	8
Figure 8: Running docker compose down command.	9
Figure 9: The two terminal windows running each the reader and the writer.	10
Figure 10: The reader in the terminal window on the right stopped and restarted again immediately.	11
Figure 11: The run of different topic with QUORUM consistency.	12

Figure 12: The run of a writer with ALL consistency and a reader with ONE consistency for a different topic on both.	14
Figure 13: Modified code segment for the edited writer.go file to implement retry on a failed write operation.	18
Figure 14: Successful execution of build of docker after editing writer.go file.	19
Figure 15: Code edit to achieve uninterrupted writer.	20
Figure 16: Initial writer run and observation.	20
Figure 17: Uninterrupted writer even after kill of db2.	21
Figure 18: Writer continuing its operation after db2 was killed.	22
Figure 19: Uninterrupted writer even after db3 kill.	22
Figure 20: Writer doing its job with QUORUM consistency.	23
Figure 21: Writer recovering after db2 kill in QUORUM consistency.	24
Figure 22: Writer failing to recover after db3 kill.	24

Solutions

1 Cassandra Cluster Management

1.1 Question 1

Q: What happens if one misconfigures the CASSANDRA_CLUSTER_NAME variable in the docker-compose.yml?

A cluster name mismatch issue may occur if the CASSANDRA_CLUSTER_NAME is incorrectly set in the docker-compose.yml file. Cassandra will raise an exception such as "saved name cluster A!= configured name cluster B" when a node attempts to join a cluster with a different name. This misconfiguration would make it impossible for the Cassandra nodes to correctly form a cluster in a Docker environment. If the problem is not identified early, each node would effectively think it is part of a distinct cluster, which might result in communication failures, the inability to replicate data across nodes, and possible data discrepancies.

Making ensuring that every node in the docker-compose.yml file has the same CASSANDRA_CLUSTER_NAME value is going to rectify this. It might be necessary to delete the data and begin over if the cluster has already been formed with names that don't match. It was also advised to wipe out the /var/lib/cassandra directory.

1.2 Question 2

Q: What is the purpose of the CASSANDRA_SEEDS variable in the docker-compose.yml? Is it necessary to list all the servers?

The creation and upkeep of Cassandra clusters depend heavily on the CASSANDRA_SEEDS setting in docker-compose.yml. It identifies the Cassandra cluster's seed nodes. Special nodes known as "seed nodes" are employed during the bootstrap procedure when a node first boots up and needs to learn the cluster structure.

The following are the main functions of seed nodes:

1. To assist new nodes in learning the topology of the cluster
2. To help with the process of cluster bootstrapping
3. To make the gossip protocol easier to use to preserve cluster state data

It is crucial to remember that listing every server as a seed is neither required nor advised. A relatively modest number of seed nodes is advised by best practices.

Regardless of the cluster's overall size, one would normally designate two to three seed nodes.

Because too many seed nodes might slow down the gossip process and make cluster formation more complex, the number is limited. Initial contact and discovery are the main functions of seed nodes. Regardless of whether they were identified as seeds or not, a node will use the gossip protocol to learn about every other node after joining the cluster.

Typically, the first two to three nodes in the cluster could be designated as seeds. For instance,

```
CASSANDRA_SEEDS: "cass1,cass2"
```

This setup instructs every node to join the cluster by using cass1 and cass2 as initial contact points. These two seed nodes will be enough to build and sustain the cluster, even if it includes many more nodes.

1.3 Question 3

Q: What does the UN mean for the first column of the output from **nodetool status**?

The state of every node in the Cassandra cluster is represented by a two-letter code that appears in the first column of the output of the nodetool status command. One of the most popular and sought-after statuses one will come across is "UN". "Up" and "Normal" are what it stands for.

To break it down even more:

- U: Up indicates that the node is online and answering inquiries.
- N: Normal, signifying that the node is operating as intended and actively taking part in cluster activities.

A node is in an ideal state when it displays "UN" in the nodetool status report. It is completely functional, taking part in all cluster operations, including data replication and repair procedures, and responding to read and write requests.

Other status codes you may come upon are:

- DN: Down and Normal (the node was in a normal condition before going down but is now offline).
- UJ: Up and Joining (the cluster is being joined by the node, which is online).
- UL: Up and Leaving (the node is online but is leaving the cluster in a gentle manner).

- UM: Up and Moving (the node is traveling to a new location within the token ring while still being online).
- DL: Going down and departing (the node was going out when it went down and is now offline).

All nodes in a healthy Cassandra cluster should normally have the "UN" status. It shows that the cluster is completely functional and that every node is enhancing the overall dependability and performance of the system.

1.4 Question 4

Q: In our cluster consisting of containers, where does Cassandra store data?

The data storage method in a containerized Cassandra cluster is made to strike a compromise between the necessity for dependable, long-lasting data storage and the transient nature of containers. Cassandra saves its data in the file system of the container by default. However, when containers are stopped or removed, this method alone would cause data loss, which is inappropriate for a database system.

To overcome this, Cassandra deployments in containers usually employ volume mappings to preserve data for longer than the lifespan of a single container. These volume mappings link Docker-managed named volumes or directories on the host system to directories within the container.

The data storage technique is different from standard settings in the Cassandra cluster configuration that is given, as specified by the docker-compose.yml file. Volume mappings for the Cassandra containers (db1, db2, and db3) are not specified in the setup. This implies that Cassandra would not have any permanent storage on the host computer; instead, all its data would be stored in the container's file system.

Figure 1 shows the given configuration of the docker-compose.yml file used for project 3. Under this setup, each container's /var/lib/cassandra directory would house Cassandra's data files, such as SSTables, commit logs, and other structures. Each container's /etc/cassandra directory would include configuration files. These folders would only be present in the container's file system because there are no volume mappings.

This arrangement has important ramifications. All data contained in a container would be lost if it were stopped or removed. This covers any configuration modifications and database content made after the container was launched. In essence, a new Cassandra instance would be launched each time the containers are restarted.

using docker compose up. There would be no access to any data that was written during the prior cycle.

```
1 services:
2   client:
3     build: .
4   db1:
5     image: cassandra:4.1.3
6     environment:
7       - HEAP_NEWSIZE=128M
8       - MAX_HEAP_SIZE=512M
9       - CASSANDRA_CLUSTER_NAME=ece573-prj03
10      - CASSANDRA_SEEDS=db1,db2,db3
11   db2:
12     extends: db1
13   db3:
14     extends: db1
```

Figure 1: docker-compose.yml file used for project 3.

While the containers are operating, the Cassandra cluster would continue form and function correctly even without persistent storage. According to Cassandra's settings, data would be duplicated between nodes, but only inside the containers that are now operating.

Since there would be no mapping to external volumes and data would be written directly to the container's file system, this configuration could provide somewhat improved I/O performance. It's crucial to remember that this setup would work better in testing or development settings when it's preferable to start from scratch every time, but it wouldn't work in any circumstance where data persistence is necessary.

It would be advised to include volume mappings to the docker-compose.yml file for production environments or any situation where data permanence is required. By doing this, data would be permanently saved on the host system or a networked storage solution, enabling it to withstand container removals and restarts.

1.5 Screenshots of task completion

The following set of figures portrays the completion of tasks necessary for project 3 until the end of the main heading Cassandra Cluster Management.

Figure 2: Running docker compose build command.

```
Activities Terminal Oct 20 09:13
ubuntu@ece573:~/ece573-prj03$ 
=> sha256:0e5e4a57c2499249aaafc3b40fcdf541e9a456aab7296 6.69kB 0.05
=> extracting sha256:6414378b647780fee8fd903db9541d134a1947ce092d08b 4.65
=> CACHED [client build 1/6] FROM docker.io/library/golang:1.21@sha256:4 0.05
=> [client build 2/6] COPY . /go/src 0.35
=> [client build 3/6] WORKDIR /go/src/reader 0.35
=> [client build 4/6] RUN go build -o reader 20.05
=> [client image 2/4] WORKDIR /client 1.05
=> [client build 5/6] WORKDIR /go/src/writer 0.25
=> [client build 6/6] RUN go build -o writer 1.25
=> [client image 3/4] COPY --from=build /go/src/reader/reader . 0.35
=> [client image 4/4] COPY --from=build /go/src/writer/writer . 0.35
=> [client] exporting to image 0.25
=> exporting layers 0.15
=> writing image sha256:496895ebca8602198b2c910d0518db16e2f5b0ad1c2d2 0.05
=> naming to docker.io/library/ece573-prj03-client 0.05
=> [client] resolving provenance for metadata file 0.05
ubuntu@ece573:~/ece573-prj03$ docker compose create
[+] Running 13/13
  ✓ db2 Pulled 19.65
  ✓ db3 Pulled 19.65
    ✓ 31bd5f451a84 Pull complete 0.35
    ✓ 32b311b06c8 Pull complete 10.15
    ✓ 23f2664f4576 Pull complete 13.05
    ✓ e028f15ee70b Pull complete 13.65
    ✓ 66b307664f73 Pull complete 13.75
    ✓ 908bad5be6da Pull complete 13.85
    ✓ dbd0abab0896e Pull complete 15.65
    ✓ 6d69b54fc45b Pull complete 15.85
    ✓ 7ff89e14c6b6 Pull complete 17.35
    ✓ b4e4e3cfaf8e0 Pull complete 17.45
  ✓ db1 Pulled 19.65
[+] Creating 5/5
  ✓ Network ece573-prj03 default Created 0.25
  ✓ Container ece573-prj03-db2-1 Created 2.05
  ✓ Container ece573-prj03-db3-1 Created 2.05
  ✓ Container ece573-prj03-dbi-1 Created 2.05
  ✓ Container ece573-prj03-client-1 Created 2.05
ubuntu@ece573:~/ece573-prj03$
```

Figure 3: Running docker compose create command.

```

Activities Terminal Oct 20 09:13
ubuntu@ece573: ~/ece573-prj03$ docker compose create
[+] Running 13/13
  ✓ db2 Pulled
  ✓ db3 Pulled
    ✓ 31bdf451a84 Pull complete   1.0s
    ✓ 32b311b806c8 Pull complete  0.2s
    ✓ 23f2664f4576 Pull complete  1.2s
    ✓ e028f15ee70b Pull complete  0.3s
    ✓ 66b307664f73 Pull complete  0.3s
    ✓ 908bad5be6da Pull complete  0.2s
    ✓ dbd0aba0896e Pull complete  0.0s
    ✓ dd69b54fc45b Pull complete  0.0s
    ✓ 7ff89e14c6b6 Pull complete  0.0s
    ✓ b4e4e3cfa8e0 Pull complete  0.0s
  ✓ db1 Pulled
[+] Creating 5/5
  ✓ Network ece573-prj03_default     Created   0.2s
  ✓ Container ece573-prj03-db2-1     Created   19.6s
  ✓ Container ece573-prj03-db3-1     Created   6.3s
  ✓ Container ece573-prj03-db1-1     Created   10.1s
  ✓ Container ece573-prj03-client-1 Created   13.6s
ubuntu@ece573:~/ece573-prj03$ docker compose start
[+] Running 4/4
  ✓ Container ece573-prj03-db1-1     Started   1.5s
  ✓ Container ece573-prj03-db2-1     Started   1.4s
  ✓ Container ece573-prj03-db3-1     Started   1.4s
  ✓ Container ece573-prj03-client-1 Started   1.5s
ubuntu@ece573:~/ece573-prj03$ 

```

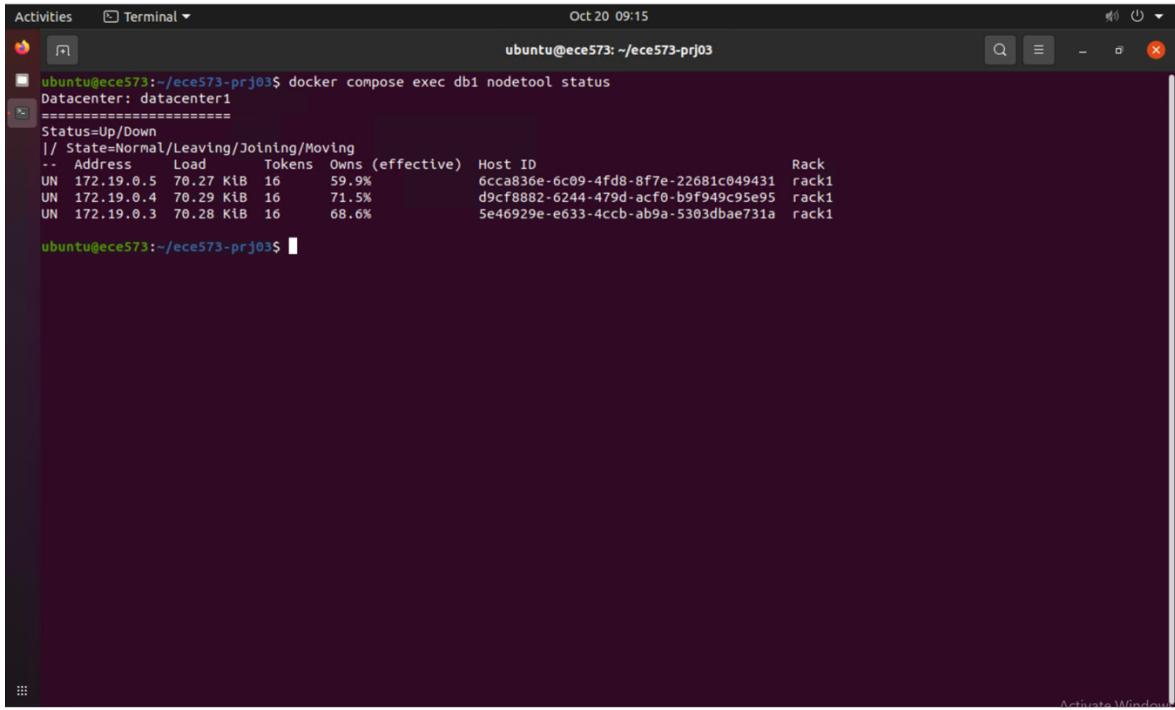
Figure 4: Running docker compose start command.

```

Activities Terminal Oct 20 09:14
ubuntu@ece573: ~/ece573-prj03$ docker compose create
[+] Running 13/13
  ✓ db2 Pulled
  ✓ db3 Pulled
    ✓ 31bdf451a84 Pull complete   0.0s
    ✓ 32b311b806c8 Pull complete  19.6s
    ✓ 23f2664f4576 Pull complete  6.3s
    ✓ e028f15ee70b Pull complete  10.1s
    ✓ 66b307664f73 Pull complete  13.6s
    ✓ 908bad5be6da Pull complete  13.6s
    ✓ dbd0aba0896e Pull complete  13.7s
    ✓ dd69b54fc45b Pull complete  13.7s
    ✓ 7ff89e14c6b6 Pull complete  13.7s
    ✓ b4e4e3cfa8e0 Pull complete  13.7s
  ✓ db1 Pulled
[+] Creating 5/5
  ✓ Network ece573-prj03_default     Created   0.2s
  ✓ Container ece573-prj03-db2-1     Created   19.6s
  ✓ Container ece573-prj03-db3-1     Created   6.3s
  ✓ Container ece573-prj03-db1-1     Created   10.1s
  ✓ Container ece573-prj03-client-1 Created   13.6s
ubuntu@ece573:~/ece573-prj03$ docker compose start
[+] Running 4/4
  ✓ Container ece573-prj03-db1-1     Started   1.5s
  ✓ Container ece573-prj03-db2-1     Started   1.4s
  ✓ Container ece573-prj03-db3-1     Started   1.4s
  ✓ Container ece573-prj03-client-1 Started   1.5s
ubuntu@ece573:~/ece573-prj03$ docker compose ps
      NAME           IMAGE        COMMAND       SERVICE  CREATED          STATUS          PORTS
  ece573-prj03-client-1  ece573-prj03-client  "/bin/sh -c 'tail -f'"  client  About a minute ago Up 19 seconds  7000-7001/tcp, 719
  9/tcp, 9042/tcp, 9160/tcp
  ece573-prj03-db1-1    cassandra:4.1.3    "docker-entrypoint.s..."  db1    About a minute ago Up 18 seconds  7000-7001/tcp, 719
  9/tcp, 9042/tcp, 9160/tcp
  ece573-prj03-db2-1    cassandra:4.1.3    "docker-entrypoint.s..."  db2    About a minute ago Up 19 seconds  7000-7001/tcp, 719
  9/tcp, 9042/tcp, 9160/tcp
  ece573-prj03-db3-1    cassandra:4.1.3    "docker-entrypoint.s..."  db3    About a minute ago Up 18 seconds  7000-7001/tcp, 719
  9/tcp, 9042/tcp, 9160/tcp
ubuntu@ece573:~/ece573-prj03$ 

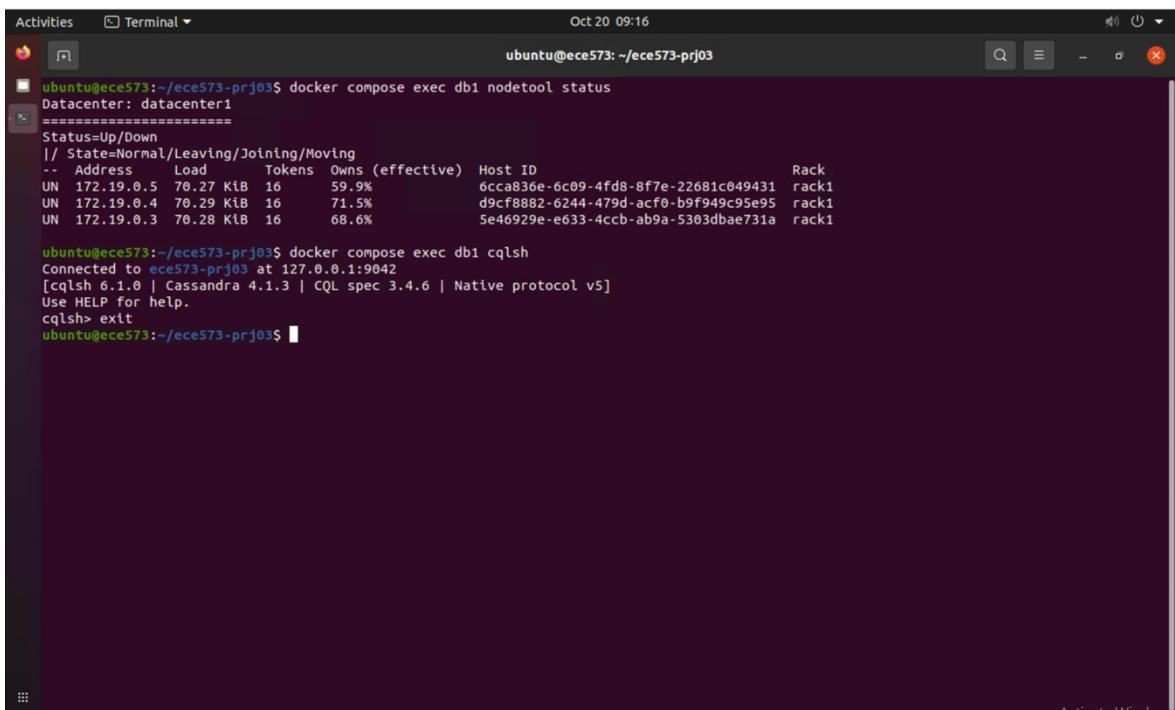
```

Figure 5: Running docker compose ps command for status.



```
Activities Terminal Oct 20 09:15
ubuntu@ece573:~/ece573-prj03$ docker compose exec db1 nodetool status
Datacenter: datacenter1
=====
Status=Up/Down
|/ State=Normal/Leaving/Joining/Moving
-- Address   Load   Tokens  Owns (effective)  Host ID            Rack
UN 172.19.0.5  70.27 KtB  16      59.9%          6cca836e-6c09-4fd8-8f7e-22681c049431  rack1
UN 172.19.0.4  70.29 KtB  16      71.5%          d9cf8882-6244-479d-acf0-b9f949c95e95  rack1
UN 172.19.0.3  70.28 KtB  16      68.6%          5e46929e-e633-4ccb-ab9a-5303dbae731a  rack1
ubuntu@ece573:~/ece573-prj03$
```

Figure 6: Running docker compose exec db1 nodetool status command.



```
Activities Terminal Oct 20 09:16
ubuntu@ece573:~/ece573-prj03$ docker compose exec db1 nodetool status
Datacenter: datacenter1
=====
Status=Up/Down
|/ State=Normal/Leaving/Joining/Moving
-- Address   Load   Tokens  Owns (effective)  Host ID            Rack
UN 172.19.0.5  70.27 KtB  16      59.9%          6cca836e-6c09-4fd8-8f7e-22681c049431  rack1
UN 172.19.0.4  70.29 KtB  16      71.5%          d9cf8882-6244-479d-acf0-b9f949c95e95  rack1
UN 172.19.0.3  70.28 KtB  16      68.6%          5e46929e-e633-4ccb-ab9a-5303dbae731a  rack1
ubuntu@ece573:~/ece573-prj03$ docker compose exec db1 cqlsh
Connected to ece573-prj03 at 127.0.0.1:9042
[cqlsh 6.1.0 | Cassandra 4.1.3 | CQL spec 3.4.6 | Native protocol v5]
Use HELP for help.
cqlsh> exit
ubuntu@ece573:~/ece573-prj03$
```

Figure 7: Running docker compose exec db1 cqlsh command and exiting the CQL command line with exit command.

```

Activities Terminal Oct 20 09:18
ubuntu@ece573: ~/ece573-prj03$ docker compose exec db1 nodetool status
Datacenter: datacenter1
=====
Status=Up/Down
|/ State=Normal/Leaving/Joining/Moving
-- Address Load Tokens Owns (effective) Host ID Rack
UN 172.19.0.5 70.27 KtB 16 59.9% 6cca836e-6c09-4fd8-8f7e-22681c049431 rack1
UN 172.19.0.4 70.29 KtB 16 71.5% d9cf882-6244-479d-acf0-b9f949c95e95 rack1
UN 172.19.0.3 70.28 KtB 16 68.6% 5e46929e-e633-4ccb-ab9a-5303dbae731a rack1

ubuntu@ece573:~/ece573-prj03$ docker compose exec db1 cqlsh
Connected to ece573-prj03 at 127.0.0.1:9042
[cqlsh 6.1.0 | Cassandra 4.1.3 | CQL spec 3.4.6 | Native protocol v5]
Use HELP for help.
cqlsh> exit
ubuntu@ece573:~/ece573-prj03$ docker compose down
[+] Running 5/5
✓ Container ece573-prj03-db3-1 Removed
✓ Container ece573-prj03-client-1 Removed
✓ Container ece573-prj03-db2-1 Removed
✓ Container ece573-prj03-db1-1 Removed
✓ Network ece573-prj03_default Removed
ubuntu@ece573:~/ece573-prj03$ 

```

Figure 8: Running docker compose down command.

2 Tunable Consistency

2.1 Question 1

Q: Keep both reader and writer running. In a few minutes you should be able to see that reader reports missing data. Provide screenshots and briefly explain what is happening.

When the commands,

```
docker compose exec client writer test ONE db1
docker compose exec client reader test ONE db2
```

are run in separate terminal windows and left for a while, the following is observed.

In the below Figure 9, the terminal window on the left is running the writer code and rows are being inserted. Similarly, the right terminal window is running the reader code and after a while of the code running, there is evidence of missing data which the reader is pointing out to.

```

Activities Terminal Oct 21 01:32
ubuntu@ece573:~/ece573-prj03 2024/10/21 01:30:59 test: inserted 2000 rows
2024/10/21 01:31:03 test: inserted 3000 rows
2024/10/21 01:31:06 test: inserted 4000 rows
2024/10/21 01:31:09 test: inserted 5000 rows
2024/10/21 01:31:13 test: inserted 6000 rows
2024/10/21 01:31:16 test: inserted 7000 rows
2024/10/21 01:31:19 test: inserted 8000 rows
2024/10/21 01:31:21 test: inserted 9000 rows
2024/10/21 01:31:24 test: inserted 10000 rows
2024/10/21 01:31:27 test: inserted 11000 rows
2024/10/21 01:31:30 test: inserted 12000 rows
2024/10/21 01:31:33 test: inserted 13000 rows
2024/10/21 01:31:36 test: inserted 14000 rows
2024/10/21 01:31:39 test: inserted 15000 rows
2024/10/21 01:31:42 test: inserted 16000 rows
2024/10/21 01:31:45 test: inserted 17000 rows
2024/10/21 01:31:48 test: inserted 18000 rows
2024/10/21 01:31:51 test: inserted 19000 rows
2024/10/21 01:31:54 test: inserted 20000 rows
2024/10/21 01:31:57 test: inserted 21000 rows
2024/10/21 01:32:00 test: inserted 22000 rows
2024/10/21 01:32:03 test: inserted 23000 rows
2024/10/21 01:32:06 test: inserted 24000 rows
2024/10/21 01:32:09 test: inserted 25000 rows
2024/10/21 01:32:12 test: inserted 26000 rows
2024/10/21 01:32:14 test: inserted 27000 rows
2024/10/21 01:32:17 test: inserted 28000 rows
2024/10/21 01:32:20 test: inserted 29000 rows
2024/10/21 01:32:23 test: inserted 30000 rows
2024/10/21 01:32:26 test: inserted 31000 rows
2024/10/21 01:32:29 test: inserted 32000 rows
2024/10/21 01:32:32 test: inserted 33000 rows
2024/10/21 01:32:35 test: inserted 34000 rows
2024/10/21 01:32:38 test: inserted 35000 rows
2024/10/21 01:32:41 test: inserted 36000 rows
2024/10/21 01:32:44 test: inserted 37000 rows
2024/10/21 01:32:47 test: inserted 38000 rows
2024/10/21 01:32:06 test: seq 23968 with 23968 rows
2024/10/21 01:32:06 test: seq 23972 with 23972 rows
2024/10/21 01:32:06 test: seq 23974 with 23974 rows
2024/10/21 01:32:06 test: seq 23975 with 23975 rows
2024/10/21 01:32:06 test: seq 23985 with 23985 rows
2024/10/21 01:32:06 test: seq 24002 with 24002 rows
2024/10/21 01:32:06 test: seq 24003 with 24003 rows
2024/10/21 01:32:06 test: seq 24004 with 24004 rows
2024/10/21 01:32:06 test: seq 24005 with 24005 rows
2024/10/21 01:32:06 test: seq 24005 with 24005 rows
2024/10/21 01:32:06 test: no more data, wait 10s
2024/10/21 01:32:16 test: seq 27501 with 27501 rows
2024/10/21 01:32:16 test: seq 27507 with 27507 rows
2024/10/21 01:32:16 test: seq 27508 with 27508 rows
2024/10/21 01:32:16 test: seq 27510 with 27510 rows
2024/10/21 01:32:16 test: seq 27513 with 27513 rows
2024/10/21 01:32:16 test: seq 27516 with 27516 rows
2024/10/21 01:32:16 test: seq 27516 with 27516 rows
2024/10/21 01:32:16 test: no more data, wait 10s
2024/10/21 01:32:26 test: seq 30867 with 30867 rows
2024/10/21 01:32:26 test: seq 30873 with 30873 rows
2024/10/21 01:32:26 test: seq 30875 with 30875 rows
2024/10/21 01:32:26 test: seq 30875 with 30875 rows
2024/10/21 01:32:26 test: no more data, wait 10s
2024/10/21 01:32:36 test: seq 34206 with 34206 rows
2024/10/21 01:32:36 test: seq 34210 with 34210 rows
2024/10/21 01:32:36 test: seq 34211 with 34211 rows
2024/10/21 01:32:36 test: seq 34213 with 34213 rows
2024/10/21 01:32:36 test: seq 34215 with 34215 rows
2024/10/21 01:32:36 test: seq 34215 with 34215 rows
2024/10/21 01:32:36 test: no more data, wait 10s
2024/10/21 01:32:46 test: seq 37514 with 37514 rows
2024/10/21 01:32:46 test: seq 37527 with 37526 rows, missing 1
2024/10/21 01:32:46 test: seq 37532 with 37531 rows, missing 1
2024/10/21 01:32:46 test: seq 37534 with 37533 rows, missing 1
2024/10/21 01:32:46 test: seq 37534 with 37533 rows, missing 1
2024/10/21 01:32:46 test: no more data, wait 10s

```

Figure 9: The two terminal windows running each the reader and the writer.

The reader would probably indicate missing data if both the writer and the reader were using consistency level ONE. The relaxed consistency level and Cassandra's eventual consistency mechanism are directly responsible for this behavior.

As soon as a write operation is acknowledged by even one duplicate, the writer, with consistency level ONE, deems it successful. This implies that initially, the data may only be written to one cluster node. In the meantime, the reader obtains information from any one copy that answers its query first, also with consistency level ONE.

A race situation is created by this configuration in which:

1. The writer may write to node A.
2. Before the data has been duplicated from A to B, the reader may read from node B.
3. The freshly written content is thus overlooked by the reader.

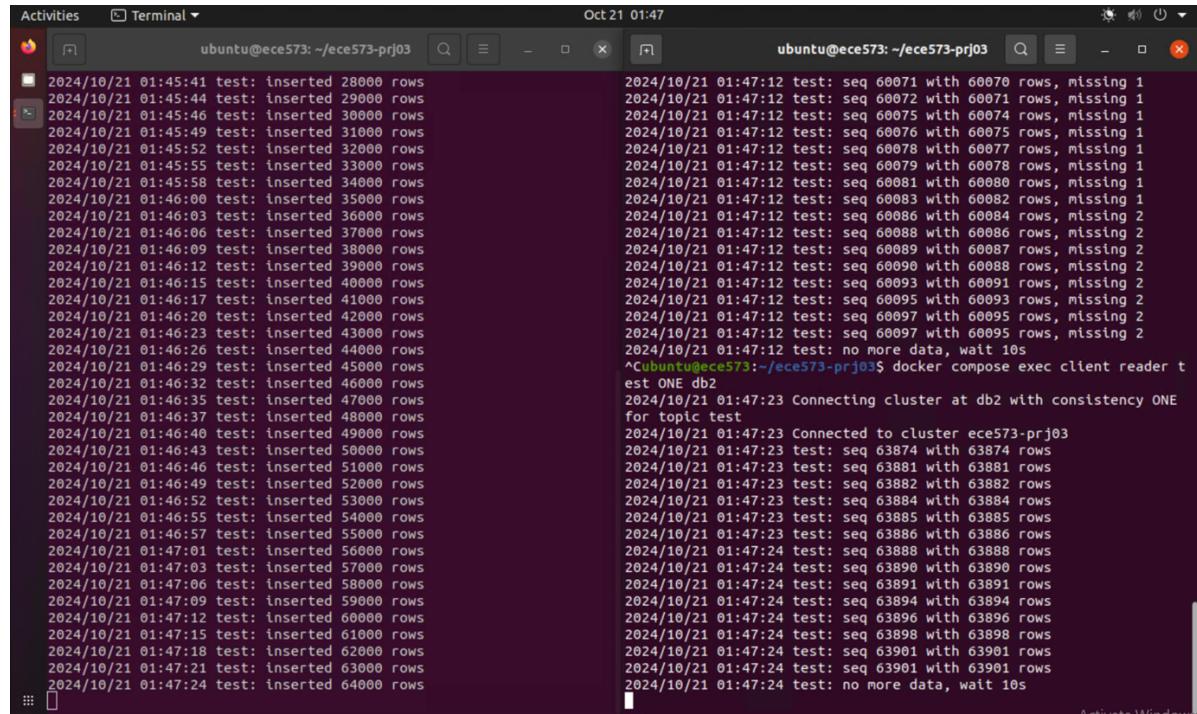
The gaps in the reader's reported value sequence would represent the missing data. These inconsistencies won't completely disappear under continuous write operations with this consistency level, but they should gradually diminish as Cassandra's internal processes (such as read repair and implied handoff) strive to spread the data across all replicas.

2.2 Question 2

Q: You can terminate the reader by Ctrl+C now and then restart it with the same arguments. Does it report any missing data? Provide screenshots and briefly explain what is happening.

The following Figure 10 shows the stopping of the reader using the Ctrl+C command and then restarting it with the command,

```
docker compose exec client reader test ONE db2
```



A screenshot of a Linux desktop environment showing two terminal windows. The left terminal window has the title 'Activities' and 'Terminal'. It contains a log of database operations from October 21, 2024, at 01:45:41, showing rows being inserted into a 'test' table with consistency level 'ONE' on a 'db2' cluster. The right terminal window also has the title 'Activities' and 'Terminal'. It shows the same log of insertions. A command is entered in the right terminal: '^Cubuntu@ece573:~/ece573-prj03\$ docker compose exec client reader test ONE db2'. The right terminal then outputs '2024/10/21 01:47:23 Connecting cluster at db2 with consistency ONE for topic test' followed by '2024/10/21 01:47:23 Connected to cluster ece573-prj03'. Both terminals show the log continuing from the point where the reader was restarted.

Figure 10: The reader in the terminal window on the right stopped and restarted again immediately.

The reader may not detect missing data right away after restarting. There are several reasons for this behavior:

1. **Catch-up period:** Cassandra's background actions, such as read repair and suggested handoff, have a chance to spread data among all replicas during the reader's halt, free from the disruption of continuous read operations.
2. **Starting point:** Since the reader most likely picks up where it left off, it is now reading latest data that has had more time to spread throughout the cluster.
3. **Random node selection:** When consistency level ONE is used, the reader may unintentionally reconnect to a more recent replica.

It's crucial to remember, though, that this seeming stability is frequently fleeting. New inconsistencies could still appear as the reader keeps working, particularly if write operations are still going on. This is because the reader is still only querying

one copy for each read operation, which may or may not have the most current data, which is the basic problem with the ONE consistency level.

With time, one may notice:

1. At certain points there seems to be consistency and no missing data.
2. Suddenly missing data appears when the reader queries less recent replicates.
3. Missing data gradually disappears as the cluster converges to consistency.

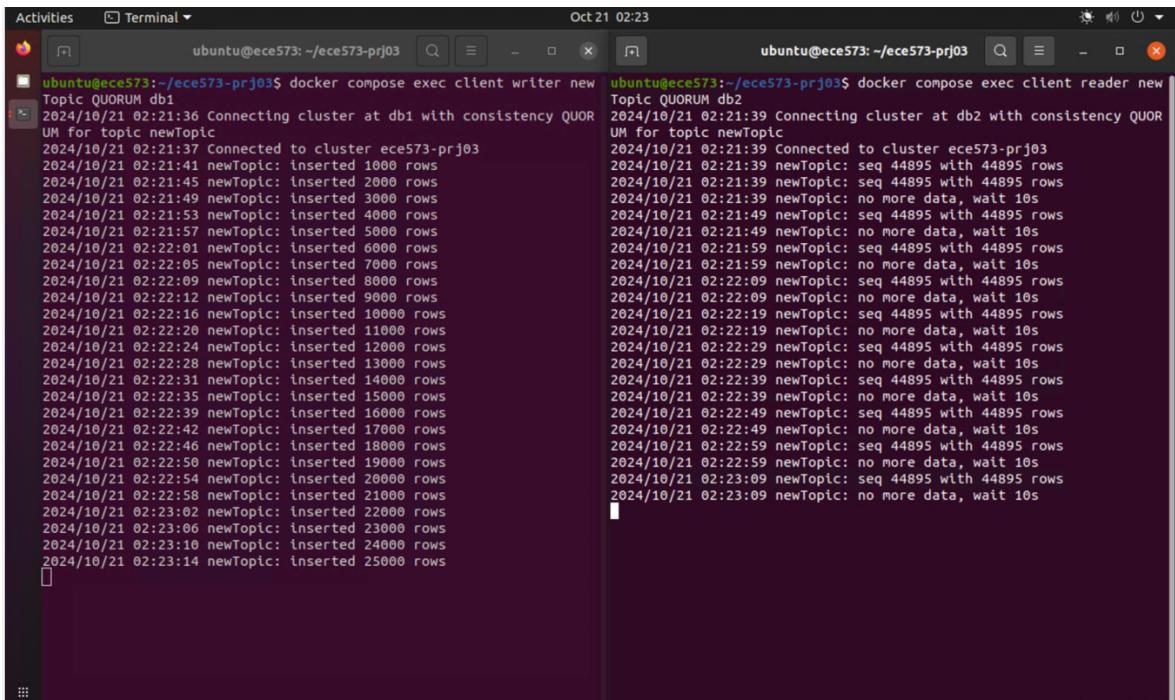
This behavior highlights Cassandra's ultimate consistency mechanism and the availability vs. consistency trade-off when employing lower consistency levels.

2.3 Question 3

Q: Stop both the reader and writer and restart them with a different topic and the QUORUM consistency. Let them run for a few minutes. Is the reader reporting any missing data and does that match your expectation? Provide screenshots and briefly explain why.

After stopping both the reader and writer executions in both the terminals, the following commands in each different terminal are run,

```
docker compose exec client writer newTopic QUORUM db1
docker compose exec client reader newTopic QUORUM db2
```



The screenshot shows two terminal windows side-by-side. Both windows are running on an Ubuntu system (version 22.04) with the terminal tab selected. The left terminal window has the title "ubuntu@ece573: ~/ece573-prj03\$". It displays the command "docker compose exec client writer newTopic QUORUM db1" followed by its execution log. The log shows the connection to the cluster at db1, the insertion of 1000 rows, and then a series of log entries from 2024/10/21 02:21:37 to 2024/10/21 02:23:14, indicating the insertion of 25000 rows in increments of 1000. The right terminal window also has the title "ubuntu@ece573: ~/ece573-prj03\$". It displays the command "docker compose exec client reader newTopic QUORUM db2" followed by its execution log. This log shows the connection to the cluster at db2, the insertion of 44895 rows, and then a series of log entries from 2024/10/21 02:21:39 to 2024/10/21 02:23:09, indicating the insertion of 20000 rows in increments of 1000. Both logs include timestamped messages such as "Connected to cluster", "newTopic: inserted [rows] rows", and "no more data, wait 10s".

Figure 11: The run of different topic with QUORUM consistency.

The above Figure 11 illustrates through a screenshot of the way that different topic and QUORUM consistency is handled.

The reader should report little or no missing data when there is QUORUM consistency for both the writer and the reader. The stronger assurances offered by the QUORUM consistency level are to blame for this notable increase in data consistency:

1. Writer behavior: To be deemed effective with QUORUM consistency, a writer must receive recognition from most replicas. This implies that a minimum of two nodes in a cluster consisting of three must verify the write.
2. Reader behavior: In a similar vein, a reader who practices QUORUM consistency will contact most replicas, compare their answers, and then provide the outcome. Based on timestamps, it will provide the most recent version of the information.

A better likelihood of consistent readings is ensured by this configuration because:

- Before validating the write, the writer made sure that most of the cluster had the data.
- Before delivering a result, the reader is consulting with most of the cluster.

It's crucial to remember that although QUORUM greatly minimizes discrepancies, it doesn't always totally eradicate them. Rarely, you could still notice slight discrepancies because of things like:

- Network partitions.
- Issues with timing in a highly concurrent setting.
- Temporary node failures.

Because the system must wait for replies from most nodes rather than just one, there is a minor increase in latency for both read and write operations as a trade-off for this enhanced consistency.

2.4 Question 4

Q: Repeat the above experiment using writer with the ALL consistency and reader with the ONE consistency. Don't forget to change to a different topic. Provide a screen shot and briefly explain your observations.

The following Figure 12 represents the tow terminals running each a writer and a reader operation by starting a different topic with the ALL consistency by the writer and the ONE consistency by the reader.

```

ubuntu@ece573:~/ece573-prj03$ docker compose exec client writer anotherTopic ALL db1
2024/10/21 02:34:24 Connecting cluster at db1 with consistency ALL
for topic anotherTopic
2024/10/21 02:34:24 Connected to cluster ece573-prj03
2024/10/21 02:34:29 anotherTopic: inserted 1000 rows
2024/10/21 02:34:34 anotherTopic: inserted 2000 rows
2024/10/21 02:34:38 anotherTopic: inserted 3000 rows
2024/10/21 02:34:43 anotherTopic: inserted 4000 rows
2024/10/21 02:34:47 anotherTopic: inserted 5000 rows
2024/10/21 02:34:52 anotherTopic: inserted 6000 rows
2024/10/21 02:34:57 anotherTopic: inserted 7000 rows
2024/10/21 02:35:01 anotherTopic: inserted 8000 rows
2024/10/21 02:35:06 anotherTopic: inserted 9000 rows
2024/10/21 02:35:10 anotherTopic: inserted 10000 rows
2024/10/21 02:35:15 anotherTopic: inserted 11000 rows
2024/10/21 02:35:20 anotherTopic: inserted 12000 rows
2024/10/21 02:35:24 anotherTopic: inserted 13000 rows
2024/10/21 02:35:29 anotherTopic: inserted 14000 rows
2024/10/21 02:35:33 anotherTopic: inserted 15000 rows

ubuntu@ece573:~/ece573-prj03$ docker compose exec client reader anotherTopic ONE db2
2024/10/21 02:34:48 Connecting cluster at db2 with consistency ONE
for topic anotherTopic
2024/10/21 02:34:49 Connected to cluster ece573-prj03
2024/10/21 02:34:49 anotherTopic: seq 5219 with 5219 rows
2024/10/21 02:34:49 anotherTopic: seq 5219 with 5219 rows
2024/10/21 02:34:49 anotherTopic: no more data, wait 10s
2024/10/21 02:34:59 anotherTopic: seq 7419 with 7419 rows
2024/10/21 02:34:59 anotherTopic: seq 7420 with 7420 rows
2024/10/21 02:34:59 anotherTopic: seq 7421 with 7421 rows
2024/10/21 02:34:59 anotherTopic: seq 7422 with 7422 rows
2024/10/21 02:34:59 anotherTopic: seq 7423 with 7423 rows
2024/10/21 02:34:59 anotherTopic: seq 7424 with 7424 rows
2024/10/21 02:34:59 anotherTopic: seq 7426 with 7426 rows
2024/10/21 02:34:59 anotherTopic: seq 7427 with 7427 rows
2024/10/21 02:34:59 anotherTopic: seq 7428 with 7428 rows
2024/10/21 02:34:59 anotherTopic: seq 7428 with 7428 rows
2024/10/21 02:34:59 anotherTopic: no more data, wait 10s
2024/10/21 02:35:09 anotherTopic: seq 9645 with 9645 rows
2024/10/21 02:35:09 anotherTopic: seq 9657 with 9657 rows
2024/10/21 02:35:09 anotherTopic: seq 9658 with 9658 rows
2024/10/21 02:35:09 anotherTopic: seq 9658 with 9658 rows
2024/10/21 02:35:09 anotherTopic: no more data, wait 10s
2024/10/21 02:35:19 anotherTopic: seq 11847 with 11847 rows
2024/10/21 02:35:19 anotherTopic: seq 11849 with 11849 rows
2024/10/21 02:35:19 anotherTopic: seq 11850 with 11850 rows
2024/10/21 02:35:19 anotherTopic: seq 11851 with 11851 rows
2024/10/21 02:35:19 anotherTopic: seq 11852 with 11852 rows
2024/10/21 02:35:19 anotherTopic: seq 11852 with 11852 rows
2024/10/21 02:35:19 anotherTopic: no more data, wait 10s
2024/10/21 02:35:29 anotherTopic: seq 13995 with 13995 rows
2024/10/21 02:35:29 anotherTopic: seq 13995 with 13995 rows
2024/10/21 02:35:29 anotherTopic: no more data, wait 10s

```

Figure 12: The run of a writer with ALL consistency and a reader with ONE consistency for a different topic on both.

In this case, when the reader uses ONE consistency and the writer uses ALL, you should see that the reader has not reported any missing data. Strong write consistency and relaxed read consistency are produced by this combination:

- Writer behavior (ALL): Before deeming a write operation successful, the writer needs confirmation from every replica in the cluster. This guarantees that, prior to the write being validated, all nodes inside the cluster possess the most recent data.
- Reader behavior (ONE): To satisfy a read request, the reader only must query one single replica.

There are no missing data with this setup because:

- Before starting any write operation, make sure all nodes have the data.
- The reader may query any single node, and it will always provide the most recent information.

Principal findings:

1. Because the writer must wait for every node to acknowledge the write, the ALL consistency level will result in a greater write latency.
2. As the reader only must contact one node, read latency is still minimal.
3. Strong consistency for writes and high availability for readers are provided by this configuration.

Nevertheless, it's important to remember that not all use cases will benefit from this configuration:

- Increased write latency may result from it, particularly in bigger clusters or when there are network problems.
- If a node disconnects and then reconnects, it doesn't prevent it from reading stale data (the ONE read consistency may read from the stale node in this scenario).

This configuration is frequently used in situations where read availability and speed take precedence over read consistency but write consistency is crucial.

2.5 Question 5

Q: What may happen if the writer uses the QUORUM consistency, and the reader uses the ONE consistency?

Although this situation offers more consistency than using ONE for both operations, there is still a chance that the reader may miss some data if the writer employs QUORUM consistency, and the reader uses ONE. This is a thorough explanation:

1. Writer (QUORUM):
 - a. Before deeming a writing successful, it makes sure that most replicas recognize it.
 - b. This implies that a minimum of two nodes in a cluster consisting of three must verify the write.
2. Reader (ONE):
 - a. For every read operation, only one duplicate should be questioned.
 - b. Whichever replica answers first, returns the data from that replica.

Potential scenarios:

1. Consistent reads: The reader will always receive the most recent data if it queries one of the nodes that took part in the QUORUM write.
2. Inconsistent reads: The reader will return stale data if it queries the one node (the non-QUORUM node) that hasn't received the most recent write.
3. Eventual consistency: As data is propagated to all nodes over time using Cassandra's internal methods (such as read repair and implied handoff), inconsistencies will be reduced.

Consistency-affecting factors include:

- Partitions and network latency.
- The particular nodes that take part in every QUORUM write.
- The chosen node for every ONE read.
- The background consistency processing speed of Cassandra.

Although it doesn't provide instant consistency across all reads, this setup provides a good compromise between write consistency and read efficiency. Strong write consistency is necessary in some use scenarios, although stale reads on occasion are acceptable in exchange for reduced read latency.

2.6 Question 6

Q: Since the source code of both writer and reader are available in our VM, can we run them from the VM directly without using the client container? Briefly explain why or why not.

Although the writer and reader programs' source code is included in the virtual machine (VM), it is not recommended or practical to execute them straight from the VM without utilizing the client container. Here's a thorough justification for why:

1. Setting Up a Network:
 - a. The network of the host virtual machine is segregated from the Cassandra cluster by means of a Docker network.
 - b. The host virtual machine (VM) cannot directly access the internal IP addresses of the containers (db1, db2, db3).
 - c. Most likely, the writer and reader programs are set up to use hostnames that can only be resolved within the Docker network—container names, such as "db1" and "db2"—to establish a connection to the Cassandra nodes.
2. Dependencies and the Environment:
 - a. Pre-configured environment variables, library versions, and dependencies needed by the writer and reader applications are included in the client container.
 - b. The environment of the virtual machine might not include these precise setups and dependencies.
 - c. Variations in library versions or configurations may result in unexpected errors or behavior.
3. Authentication and the Connection String:
 - a. Connection strings or authentication techniques unique to the containerized environment may be used by the applications.
 - b. Internal Docker network addresses or authentication tokens unique to each containers may be included in this connection information.
4. Configuration of the Cassandra driver:
 - a. It's possible that the applications' Cassandra driver is set up especially for the containerized environment.
 - b. This might include load balancing, retry policy optimizations, and connection pooling settings tailored to the Docker configuration.

5. Continuity in the Testing Setting:
 - a. By using the client container, you can make sure that every user or student on the system is working in the same, uniform environment.
 - b. By doing this, factors that can occur from variations in individual virtual machine settings are eliminated, guaranteeing consistent outcomes across various systems.
6. Access control and security:
 - a. For security purposes, the Cassandra cluster may be set up to only allow connections from within its Docker network.
 - b. Programs run from the virtual machine (VM) may circumvent planned security precautions.
7. Mapping Ports and Exposure:
 - a. It's possible that the host virtual machine's ports aren't open to the Cassandra nodes, which inhibits applications from running directly on the VM.
8. Docker-specific Improvements:
 - a. The applications may be tuned to operate in a Docker environment, utilizing the networking and inter-container communication capabilities of Docker.

It would need a lot of changes to get the applications to run straight out of the virtual machine (VM):

- Changing the network configuration will enable the virtual machine to connect to the Docker network.
- Adjusting the applications' connection strings to use the appropriate hostnames or IP addresses.
- Installing and setting up all required dependencies in the virtual machine to correspond with the environment of the client container.
- Changing Cassandra's security configuration to permit connections from outside.
- Cassandra driver settings must be updated to support the non-containerized setup.
- These changes would not only take a lot of time, but they would also not follow the project's original plan. They could produce discrepancies in outcomes across various learning settings and perhaps lead to security flaws.

Additionally, a major advantage of adopting Docker in this project is that it offers a consistent, isolated environment that closely resembles a distributed system configuration found in the real world. This advantage would be lost if the programs were run straight from the virtual machine, which would result in a less realistic learning environment.

In summary, even while it could be theoretically feasible to change the configuration such that the programs run straight from the virtual machine, doing so would be laborious, unrealistic, and at odds with the project's instructional objectives. To maintain consistency, security, and conformity with the project's learning objectives, it is highly suggested to execute the writer and reader programs inside the client container that is provided, just as intended.

3 Availability

3.1 Question 1

Q: Modify writer.go to retry for a failed write operation. Use a retry policy to wait 10 seconds before retrying but allow infinite number of retries.

The modification necessary for the retry of a failed write operation is shown in the following Figure 13.

```
File Edit Selection View Go Run Terminal Help
Restricted Mode is intended for safe code browsing. Trust this window to enable all features. Manage Learn More
File writer.go - Visual Studio Code
home > ubuntu > ece573-prj03 > writer > == writer.go
13 func main() {
14     if err := session.Query(
15         "CREATE KEYSPACE IF NOT EXISTS ece573
16             WITH replication = (
17                 'class':'SimpleStrategy',
18                 'replication factor':3)");
19     Exec(); err != nil {
20         log.Fatalf("Cannot create keyspace ece573: %v", err)
21     }
22
23     if err := session.Query(
24         "CREATE TABLE IF NOT EXISTS ece573.prj03 (
25             topic text, seq int, value double,
26             PRIMARY KEY (topic, seq))".
27     Exec(); err != nil {
28         log.Fatalf("Cannot create table ece573.prj03: %v", err)
29     }
30
31     for seq := 1; ; seq++ {
32         value := rand.Float64()
33         for {
34             err := session.Query(
35                 "INSERT INTO ece573.prj03 (topic, seq, value) VALUES (?, ?, ?)", topic, seq, value).
36             Exec();
37             if err != nil {
38                 break //successful write, exit the retry loop
39             }
40             log.Printf("Cannot write %d to table ece573.prj03: %v. Retrying in 10 seconds...", seq, err)
41             time.Sleep(10 * time.Second)
42         }
43         if seq1000 == 0 {
44             log.Printf("%s: inserted %d rows", topic, seq)
45         }
46     }
47 }
```

Figure 13: Modified code segment for the edited writer.go file to implement retry on a failed write operation.

Post the edit of the code the following Figure 14 shows the build of docker and its successful execution without errors.

```

Activities Terminal Oct 21 03:32
ubuntu@ece573:~/ece573-prj03$ docker compose down
[+] Running 5/5
  ✓ Container ece573-prj03-db1-1 Removed          10.6s
  ✓ Container ece573-prj03-client-1 Removed        10.6s
  ✓ Container ece573-prj03-db2-1 Removed          6.05s
  ✓ Container ece573-prj03-db3-1 Removed          6.05s
  ✓ Network ece573-prj03_default Removed          0.25s
ubuntu@ece573:~/ece573-prj03$ docker compose up -d
[+] Running 5/5
  ✓ Network ece573-prj03_default Created          0.25s
  ✓ Container ece573-prj03-db1-1 Started          1.1s
  ✓ Container ece573-prj03-db2-1 Started          1.3s
  ✓ Container ece573-prj03-client-1 Started        1.4s
  ✓ Container ece573-prj03-db3-1 Started          1.4s

```

Figure 14: Successful execution of build of docker after editing writer.go file.

A retry mechanism for unsuccessful writing operations would be implemented as part of the writer.go change. This modification would greatly increase the writer's resistance to transient network problems or node failures. Most likely, the implementation would make use of a loop that tries the write operation repeatedly while identifying any mistakes. When the code encountered an error, it would pause for ten seconds before retrying the task.

In situations when node recovery is anticipated but the timeframe is unknown, the infinite retry policy would guarantee that the writer keeps attempting writes indefinitely. It is important to remember, though, that in a production setting, it may be wise to use an exponential backoff technique or a maximum retry limit to avoid using too many resources in the event of extended failures.

This change would show that the author understands the significance of robust client-side code when working with distributed databases such as Cassandra, as well as error handling in distributed systems.

3.2 Question 2

Q: Build, create, and start the client again and demonstrate that writer can keep adding rows when db2 is killed. Provide screenshots as needed.

The following edits was made to the code in writer.go to achieve the result of writer that continues to add data even after db2 being killed. The same is shown in Figure 15.

```

// writer.go
package main

import (
    "log"
    "math/rand"
    "mysql"
    "time"
)

func main() {
    session, err := mysql.NewSession(mysql.SessionConfig{
        Host:     "127.0.0.1",
        Port:     3306,
        User:     "root",
        Password: "password",
        Database: "test",
    })
    if err != nil {
        log.Fatalf("Cannot connect to database: %v", err)
    }

    for seq := 1; ; seq++ {
        value := rand.Float64()
        var err error
        for retries := 0; retries < 5; retries++ {
            err = session.Query(
                "INSERT INTO test (topic, seq, value) VALUES (?, ?, ?)",
                "topic", seq, value,
            )
            if err == nil {
                break
            }
            log.Printf("Error writing %d: %v. Retrying...", seq, err)
            time.Sleep(2 * time.Second)
        }
        if err != nil {
            log.Printf("Failed to write %d to after 5 retries: %v", seq, err)
        }
        if seq%1000 == 0 {
            log.Printf("%d: inserted %d rows", topic, seq)
        }
    }
}

```

Figure 15: Code edit to achieve uninterrupted writer.

The initial run of the writer code is shown in the following screenshot Figure 16

```

ubuntu@ece573:~/ece573-prj03$ docker compose exec client writer ava il ONE db1
2024/10/21 04:40:21 Connecting cluster at db1 with consistency ONE
2024/10/21 04:40:21 Connected to cluster ece573-prj03
2024/10/21 04:40:31 avail: inserted 1000 rows
2024/10/21 04:40:37 avail: inserted 2000 rows
2024/10/21 04:40:42 avail: inserted 3000 rows
2024/10/21 04:40:47 avail: inserted 4000 rows
2024/10/21 04:40:52 avail: inserted 5000 rows
2024/10/21 04:40:57 avail: inserted 6000 rows
2024/10/21 04:41:01 avail: inserted 7000 rows
2024/10/21 04:41:05 avail: inserted 8000 rows
2024/10/21 04:41:09 avail: inserted 9000 rows
2024/10/21 04:41:13 avail: inserted 10000 rows
2024/10/21 04:41:16 avail: inserted 11000 rows
2024/10/21 04:41:20 avail: inserted 12000 rows
2024/10/21 04:41:23 avail: inserted 13000 rows
2024/10/21 04:41:27 avail: inserted 14000 rows

```

Figure 16: Initial writer run and observation.

Once the db2 is killed in the secondary terminal, the writer continues to do its job, which is shown in the following screenshot Figure 17.

The screenshot shows two terminal windows side-by-side. The left terminal window has the title 'Activities Terminal' and shows a log of a 'writer' process inserting rows into a database named 'db1'. The right terminal window has the title 'ubuntu@ece573: ~/ece573-prj03' and shows the command 'docker compose kill db2' being run, which kills the 'db2' container. Despite this, the writer process continues to insert rows into 'db1'.

```

ubuntu@ece573:~/ece573-prj03$ docker compose exec client writer avail ONE db1
2024/10/21 04:40:21 Connecting cluster at db1 with consistency ONE
for topic avail
2024/10/21 04:40:21 Connected to cluster ece573-prj03
2024/10/21 04:40:31 avail: inserted 1000 rows
2024/10/21 04:40:37 avail: inserted 2000 rows
2024/10/21 04:40:42 avail: inserted 3000 rows
2024/10/21 04:40:47 avail: inserted 4000 rows
2024/10/21 04:40:52 avail: inserted 5000 rows
2024/10/21 04:40:57 avail: inserted 6000 rows
2024/10/21 04:41:01 avail: inserted 7000 rows
2024/10/21 04:41:05 avail: inserted 8000 rows
2024/10/21 04:41:09 avail: inserted 9000 rows
2024/10/21 04:41:13 avail: inserted 10000 rows
2024/10/21 04:41:16 avail: inserted 11000 rows
2024/10/21 04:41:20 avail: inserted 12000 rows
2024/10/21 04:41:23 avail: inserted 13000 rows
2024/10/21 04:41:27 avail: inserted 14000 rows
2024/10/21 04:41:32 avail: inserted 15000 rows
2024/10/21 04:41:36 avail: inserted 16000 rows
2024/10/21 04:41:37 Error writing i6059: read tcp 172.19.0.4:43484->172.19.0.3:9042: read: connection reset by peer. Retrying...
2024/10/21 04:41:42 avail: inserted 17000 rows
2024/10/21 04:41:45 avail: inserted 18000 rows

ubuntu@ece573:~/ece573-prj03$ docker compose kill db2
[+] Killing 1/1
✓ Container ece573-prj03-db2-1 Killed
ubuntu@ece573:~/ece573-prj03$ 0.7s

```

Figure 17: Uninterrupted writer even after kill of db2.

This demonstration effectively shows that even when db2 is stopped, the writer may continue to add rows. The momentary disruption that was followed by continuous functioning demonstrates the robustness of both the Cassandra cluster and the updated writer software. It emphasizes how crucial it is for client applications to have appropriate error handling and retry methods when interacting with distributed databases.

The outcomes demonstrate how Cassandra can continue writing operations even in the event of a node loss, confirming the system's well-known fault tolerance and high availability. For applications like log collecting systems, Internet of Things data ingestion, or heavily trafficked web apps, this behavior is essential.

3.3 Question 3

Q: What if db3 is also killed? Provide screenshots and briefly explain your findings.

The robustness of the edits made to the code in `writer.go` is once again viewed here when db3 was also killed.

The following `<figure>` shows the writer continuing with its operation after db2 was killed.

The screenshot shows two terminal windows side-by-side. The left terminal window displays a log of insert operations from a writer program, showing rows being inserted at various times between 04:49:35 and 04:51:06. The right terminal window shows the command `docker compose kill db2` being run, which kills the db2 container. Despite this, the writer process continues to operate and insert rows.

```

ubuntu@ece573:~/ece573-prj03$ docker compose kill db2
[+] Killing 1/1
✓ Container ece573-prj03-db2-1 Killed
ubuntu@ece573:~/ece573-prj03$ 0.7s

```

Figure 18: Writer continuing its operation after db2 was killed.

The following Figure 19 shows the kill of db3 and still writer manages to continue with its operation being totally unhinged.

This screenshot is similar to Figure 18, showing two terminal windows. The left window shows the writer's insert log. The right window shows the commands `docker compose kill db2` and `docker compose kill db3` being run sequentially. The writer process continues to run and insert rows even after both db2 and db3 have been killed.

```

ubuntu@ece573:~/ece573-prj03$ docker compose kill db2
[+] Killing 1/1
✓ Container ece573-prj03-db2-1 Killed
ubuntu@ece573:~/ece573-prj03$ docker compose kill db3
[+] Killing 1/1
✓ Container ece573-prj03-db3-1 Killed
ubuntu@ece573:~/ece573-prj03$ 0.8s

```

Figure 19: Uninterrupted writer even after db3 kill.

There is just one operational node (db1) in the Cassandra cluster when db3 and db2 are eliminated together. The writer program ought to stay in operation despite this for the following reasons:

1. Consistency Level 1: A write operation at this level can only be deemed successful if it receives acknowledgment from a single node. This condition may still be satisfied with db1 still operational.
2. The distributed nature of Cassandra lets the surviving node to continue accepting write operations even if two nodes fail.
3. Client Connection Management: If the writer program is set up correctly, it ought to be able to direct all requests to the node that is still active.

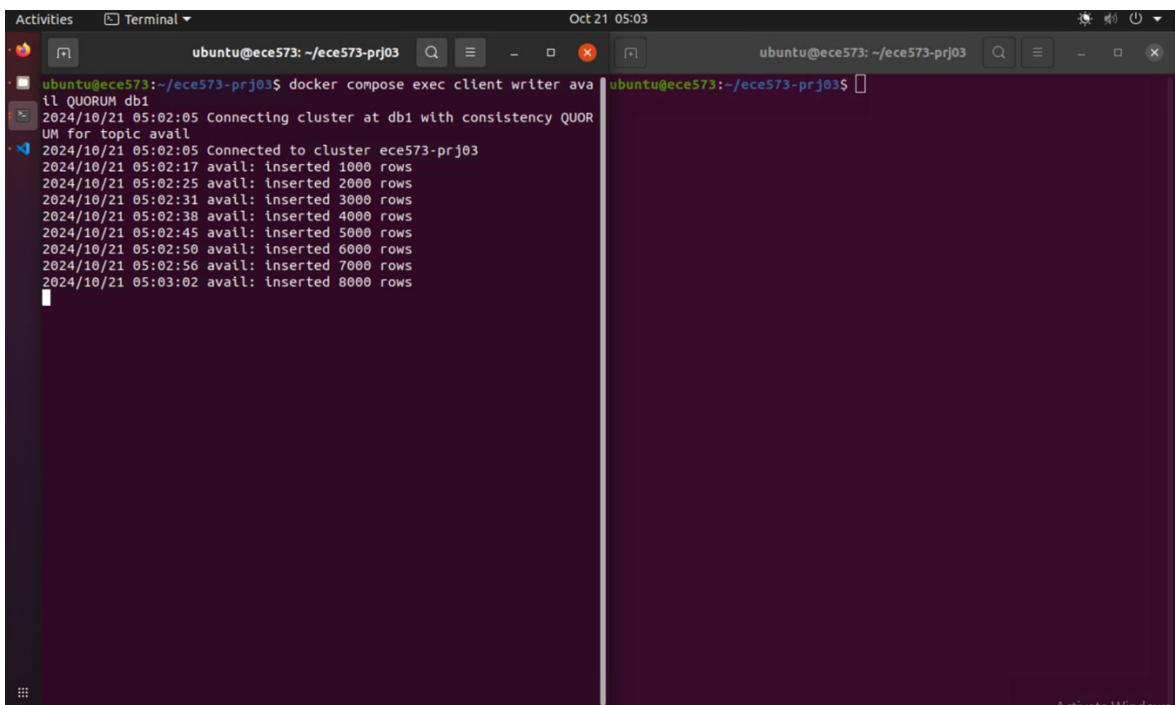
But one could note:

- A momentary stop or error messages as the system comes to terms with losing a node.
- Possible rise in delay brought on by a heavier burden on the lone surviving node.
- Write operations are ongoing, showcasing Cassandra's high availability feature.

This case illustrates how Cassandra can keep write availability even when most of the cluster is unavailable, although at the risk of possible inconsistent data once the downed nodes have recovered.

3.4 Question 4

Q: Restart writer with consistency QUORUM. Does the writer perform differently when one server is killed and when two servers are killed? Provide screenshots and briefly explain your findings.

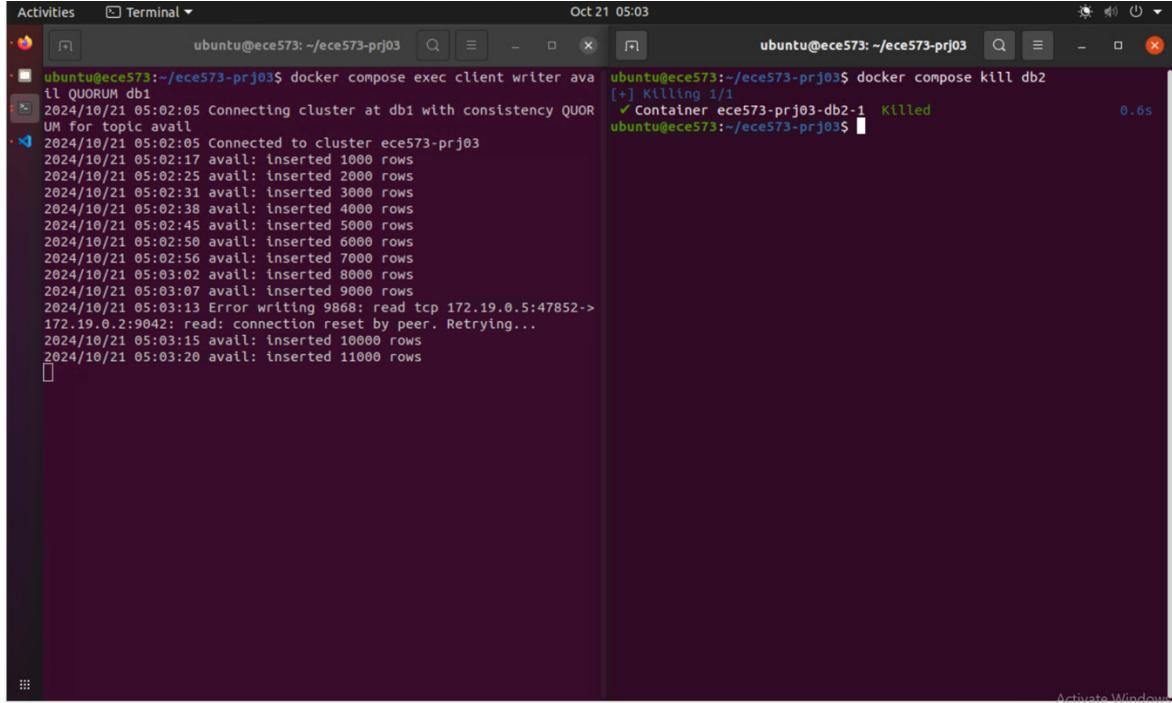


```
Activities Terminal Oct 21 05:03
ubuntu@ece573:~/ece573-prj03$ docker compose exec client writer ava
il QUORUM db1
2024/10/21 05:02:05 Connecting cluster at db1 with consistency QUORUM
UM for topic avail
2024/10/21 05:02:05 Connected to cluster ece573-prj03
2024/10/21 05:02:17 avail: inserted 1000 rows
2024/10/21 05:02:25 avail: inserted 2000 rows
2024/10/21 05:02:31 avail: inserted 3000 rows
2024/10/21 05:02:38 avail: inserted 4000 rows
2024/10/21 05:02:45 avail: inserted 5000 rows
2024/10/21 05:02:50 avail: inserted 6000 rows
2024/10/21 05:02:56 avail: inserted 7000 rows
2024/10/21 05:03:02 avail: inserted 8000 rows
```

Figure 20: Writer doing its job with QUORUM consistency.

The above Figure 20 shows the writer with QUORUM consistency doing its job in writing to the cluster.

The following Figure 21 shows the kill of db2 and writer recovering after the same.



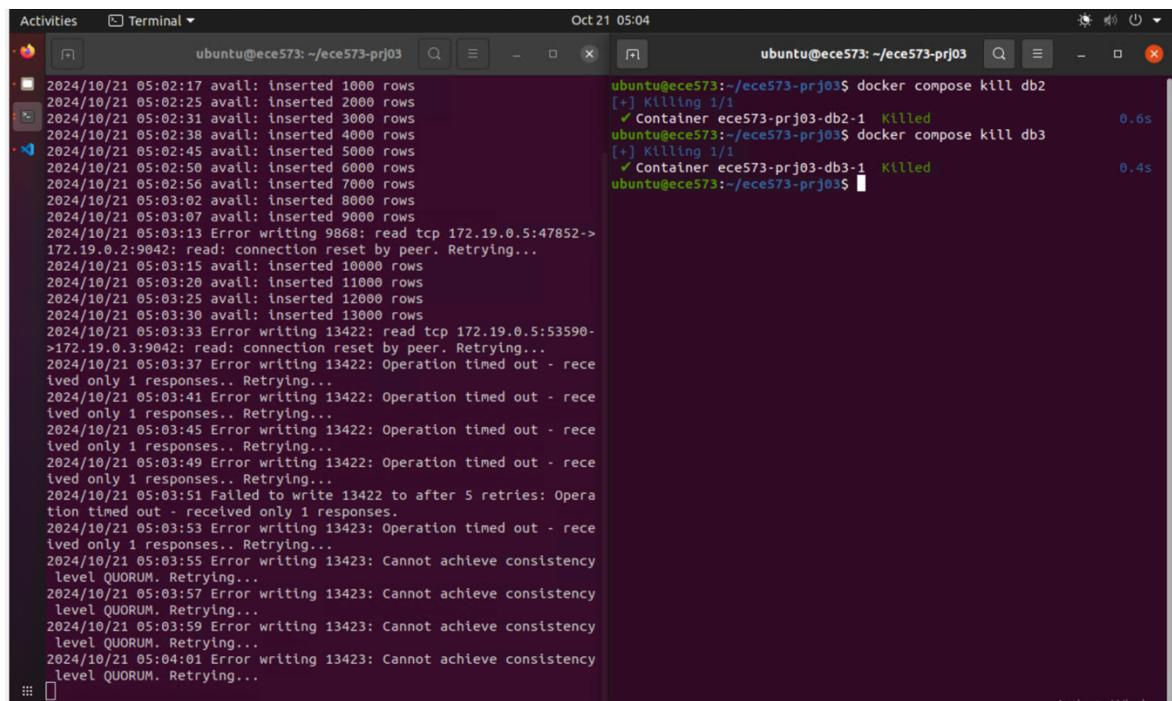
The screenshot shows two terminal windows side-by-side. The left terminal window shows the writer performing insertions into the 'avail' topic. The right terminal window shows the db2 instance being killed and then the writer recovering.

```
ubuntu@ece573:~/ece573-prj03$ docker compose exec client writer avail QUORUM db1
2024/10/21 05:02:05 Connecting cluster at db1 wth consistency QUORUM for topic avail
2024/10/21 05:02:05 Connected to cluster ece573-prj03
2024/10/21 05:02:17 avail: inserted 1000 rows
2024/10/21 05:02:25 avail: inserted 2000 rows
2024/10/21 05:02:31 avail: inserted 3000 rows
2024/10/21 05:02:38 avail: inserted 4000 rows
2024/10/21 05:02:45 avail: inserted 5000 rows
2024/10/21 05:02:50 avail: inserted 6000 rows
2024/10/21 05:02:56 avail: inserted 7000 rows
2024/10/21 05:03:02 avail: inserted 8000 rows
2024/10/21 05:03:07 avail: inserted 9000 rows
2024/10/21 05:03:13 Error writing 9868: read tcp 172.19.0.5:47852->172.19.0.2:9042: read: connection reset by peer. Retrying...
2024/10/21 05:03:15 avail: inserted 10000 rows
2024/10/21 05:03:20 avail: inserted 11000 rows

ubuntu@ece573:~/ece573-prj03$ docker compose kill db2
[+] Killing 1/1
✓ Container ece573-prj03-db2-1 Killed
0.6s
ubuntu@ece573:~/ece573-prj03$
```

Figure 21: Writer recovering after db2 kill in QUORUM consistency.

The following Figure 22 shows the interesting output where the writer fails to recover once db3 was also killed after the kill of db2.



The screenshot shows two terminal windows side-by-side. The left terminal window shows the writer performing insertions into the 'avail' topic. The right terminal window shows both db2 and db3 instances being killed, and the writer failing to recover due to lack of a quorum.

```
ubuntu@ece573:~/ece573-prj03$ docker compose exec client writer avail QUORUM db1
2024/10/21 05:02:17 avail: inserted 1000 rows
2024/10/21 05:02:25 avail: inserted 2000 rows
2024/10/21 05:02:31 avail: inserted 3000 rows
2024/10/21 05:02:38 avail: inserted 4000 rows
2024/10/21 05:02:45 avail: inserted 5000 rows
2024/10/21 05:02:50 avail: inserted 6000 rows
2024/10/21 05:02:56 avail: inserted 7000 rows
2024/10/21 05:03:02 avail: inserted 8000 rows
2024/10/21 05:03:07 avail: inserted 9000 rows
2024/10/21 05:03:13 Error writing 9868: read tcp 172.19.0.5:47852->172.19.0.2:9042: read: connection reset by peer. Retrying...
2024/10/21 05:03:15 avail: inserted 10000 rows
2024/10/21 05:03:20 avail: inserted 11000 rows
2024/10/21 05:03:25 avail: inserted 12000 rows
2024/10/21 05:03:30 avail: inserted 13000 rows
2024/10/21 05:03:33 Error writing 13422: read tcp 172.19.0.5:53590->172.19.0.3:9042: read: connection reset by peer. Retrying...
2024/10/21 05:03:37 Error writing 13422: Operation timed out - received only 1 responses.. Retrying...
2024/10/21 05:03:41 Error writing 13422: Operation timed out - received only 1 responses.. Retrying...
2024/10/21 05:03:45 Error writing 13422: Operation timed out - received only 1 responses.. Retrying...
2024/10/21 05:03:49 Error writing 13422: Operation timed out - received only 1 responses.. Retrying...
2024/10/21 05:03:51 Failed to write 13422 to after 5 retries: Operation timed out - received only 1 responses.
2024/10/21 05:03:53 Error writing 13423: Operation timed out - received only 1 responses.. Retrying...
2024/10/21 05:03:55 Error writing 13423: Cannot achieve consistency level QUORUM. Retrying...
2024/10/21 05:03:57 Error writing 13423: Cannot achieve consistency level QUORUM. Retrying...
2024/10/21 05:03:59 Error writing 13423: Cannot achieve consistency level QUORUM. Retrying...
2024/10/21 05:04:01 Error writing 13423: Cannot achieve consistency level QUORUM. Retrying...

ubuntu@ece573:~/ece573-prj03$ docker compose kill db2
[+] Killing 1/1
✓ Container ece573-prj03-db2-1 Killed
0.6s
ubuntu@ece573:~/ece573-prj03$ docker compose kill db3
[+] Killing 1/1
✓ Container ece573-prj03-db3-1 Killed
0.4s
ubuntu@ece573:~/ece573-prj03$
```

Figure 22: Writer failing to recover after db3 kill.

In a three-node Cassandra cluster with QUORUM consistency, the writer's behavior alters dramatically when nodes are eliminated:

1. Every node is operational:

At first, the writer ought to operate as usual. In a cluster of three nodes, QUORUM requires acknowledgment from most replicas, meaning that two nodes must acknowledge each write.

2. After killing db2 (one node down):

- a. The writer should be able to continue working, albeit there could be some momentary hiccups or more delay.
- b. It is still possible to reach QUORUM using two of the three nodes (db1 and db3).
- c. While retry notifications may occasionally appear, most operations should be successful.

3. Following db3's death (two nodes down):

- a. The writer is unable to produce work and keeps trying in vain.
- b. It is shown by the error "Cannot achieve consistency and retrying" that QUORUM is not met.
- c. Getting acknowledgment from most copies is impossible with only one active node (db1).

This behavior illustrates the Cassandra trade-offs between availability and consistency:

- Reliability versus Accessibility:
 - Compared to ONE, QUORUM consistency offers higher guarantees for data consistency.
 - But availability is decreased because when most nodes in the cluster are unavailable, writes cannot be served by the cluster.
- The ability to tolerate faults:
 - In a cluster of three nodes, QUORUM allows the system to withstand the failure of one node.
 - Two node failures are intolerable to it since they obstruct reaching a quorum.
- Data Accuracy:
 - Compared to ONE, QUORUM offers superior durability and consistency by guaranteeing that data is written to at least two nodes.
 - Reduced availability in the event of numerous node failures is the price for this.
- Impact on Performance:
 - You may experience greater latency even with one node offline as the system attempts to preserve QUORUM consistency.
- Error Resolution:
 - The client's persistent retrying behavior demonstrates their efforts to uphold the desired consistency level.
 - This would necessitate appropriate error handling and maybe a fallback plan in a production setting.

Cassandra's consistency paradigm is eloquently illustrated in the experiment with QUORUM consistency. It illustrates how system availability may be impacted by raising consistency requirements. Although QUORUM offers more robust assurances for data consistency, it increases the system's susceptibility to multiple node failures. This highlights the need of selecting the appropriate consistency level in accordance with your unique needs for data consistency in comparison to system availability.