REPORT

- Navodita Mathur, Allie Azzarello, Naveena Nagaraju

Extract the files from the attached zip. Open terminal in the extracted folder and follow the steps in report and .txt file.

PART-1:

1. Install Cassandra on Ubuntu VMs:
- First, update the package index on each VM:

```
sudo apt-get update
```

- Install Java if it's not already installed:

```
sudo apt-get install default-jdk
```

- Add the Apache Cassandra repository keys:

```
sudo curl -o /etc/apt/keyrings/apache-cassandra.asc
https://downloads.apache.org/cassandra/KEYS
```

- Add the Apache Cassandra repository to the package sources list:

```
echo "deb [signed-by=/etc/apt/keyrings/apache-cassandra.asc]
https://debian.cassandra.apache.org 41x main" | sudo tee -a
/etc/apt/sources.list.d/cassandra.sources.list
```

- Update the package index again:

```
sudo apt-get update
```

- Install Cassandra

```
sudo apt-get install Cassandra
```

2. Configure Cassandra nodes:
- Navigate to the Cassandra configuration directory:

```
cd /etc/cassandra/
```

- Edit the cassandra.yaml file:

```
sudo nano cassandra.yaml
```

Adjust the following settings:
- listen_address: Set the IP address for this node to listen on.
- rpc_address: Set the IP address for remote procedure calls.
- seeds: Add the IP addresses of all nodes in the cluster.
- modify the following in the files:

```
materialized_views_enabled: true
sasi_indexes_enabled: true
user_defined_functions_enabled: true
scripted_user_defined_functions_enabled: true
```

- modify the limits:

```
read_request_timeout: 100000ms
range_request_timeout: 100000ms
```

```
write_request_timeout: 20000ms
counter_write_request_timeout: 50000ms
cas_contention_timeout: 10000ms
truncate_request_timeout: 600000ms
request_timeout: 100000ms
slow_query_log_timeout: 50000ms
```

- Save and exit the file.

3. Start Cassandra nodes:

- Start Cassandra on each node:

```
sudo service start cassandra
```

- Check the status to ensure it's running:

```
sudo systemctl status cassandra
```

4. Verify cluster status:

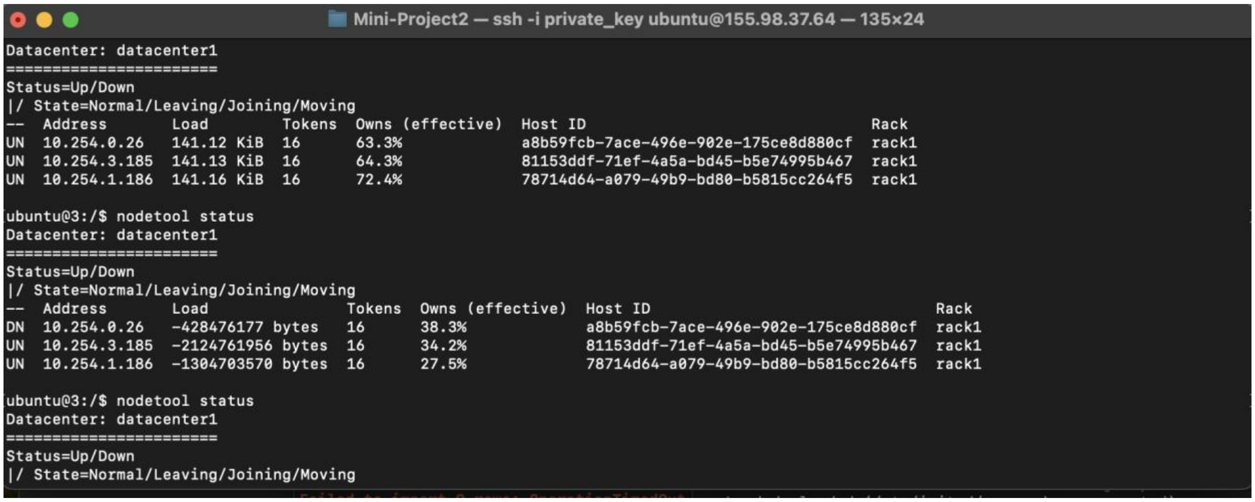
- Check the status of the Cassandra cluster:

```
nodetool status
```

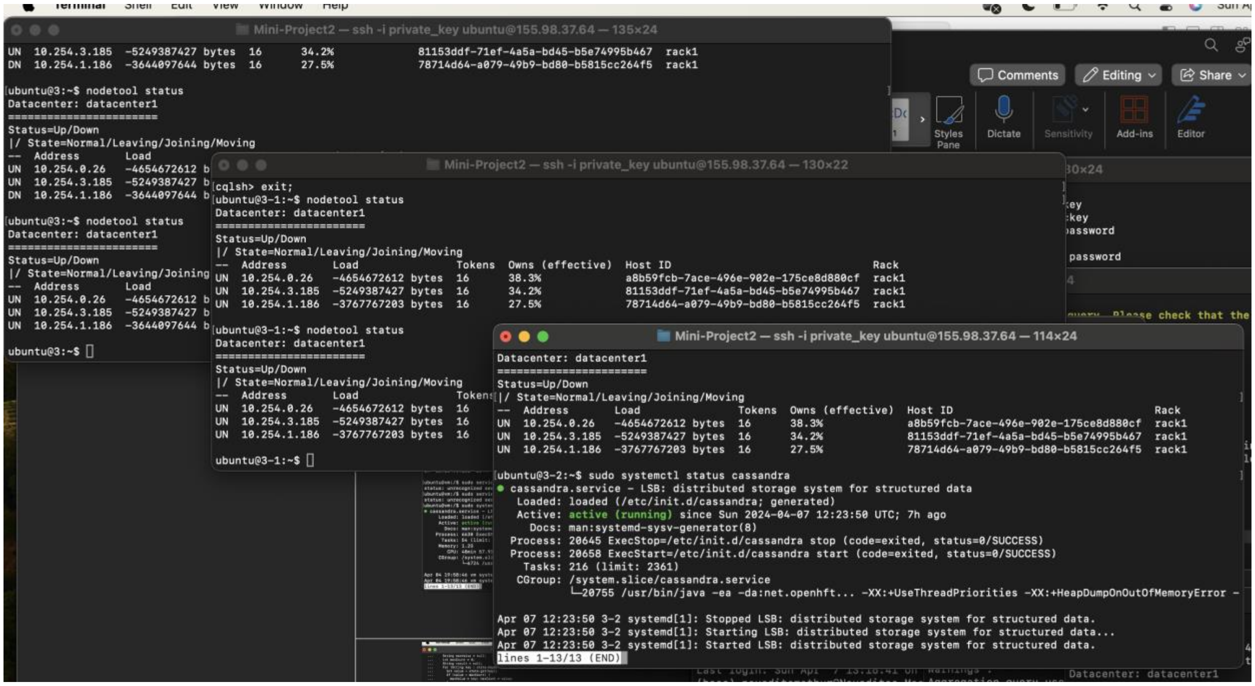
- You should see all nodes listed and 'UN' (up and normal) status for each.
- Log into one of the nodes:

```
cqlsh 10.254.0.26 9042 --request-timeout=60000
```

Output:


Mini-Project2 — ssh -i private_key ubuntu@155.98.37.64 — 135×24

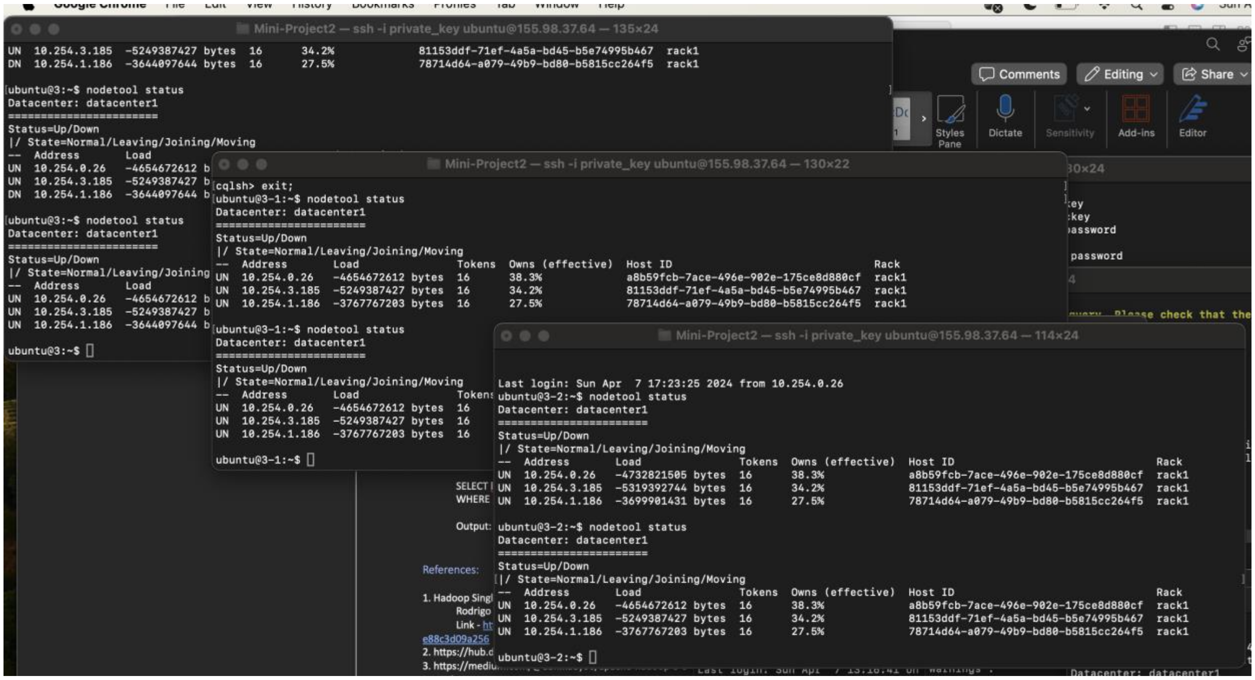
```
Datacenter: datacenter1
=======================
Status=Up/Down
|/ State=Normal/Leaving/Joining/Moving
--  Address       Load        Tokens  Owns (effective)  Host ID                               Rack
UN  10.254.0.26   141.12 KiB  16      63.3%             a8b59fcb-7ace-496e-902e-175ce8d880cf  rack1
UN  10.254.3.185  141.13 KiB  16      64.3%             81153ddf-71ef-4a5a-bd45-b5e74995b467  rack1
UN  10.254.1.186  141.16 KiB  16      72.4%             78714d64-a079-49b9-bd80-b5815cc264f5  rack1

ubuntu@3:/$ nodetool status
Datacenter: datacenter1
=======================
Status=Up/Down
|/ State=Normal/Leaving/Joining/Moving
--  Address       Load               Tokens  Owns (effective)  Host ID                               Rack
DN  10.254.0.26   -428476177 bytes   16      38.3%             a8b59fcb-7ace-496e-902e-175ce8d880cf  rack1
UN  10.254.3.185  -2124761956 bytes  16      34.2%             81153ddf-71ef-4a5a-bd45-b5e74995b467  rack1
UN  10.254.1.186  -1304703570 bytes  16      27.5%             78714d64-a079-49b9-bd80-b5815cc264f5  rack1

ubuntu@3:/$ nodetool status
Datacenter: datacenter1
=======================
Status=Up/Down
|/ State=Normal/Leaving/Joining/Moving
```


Terminal
Shell
Edit
View
Window
Help
Mini-Project2 — ssh -i private_key ubuntu@155.98.37.64 — 135×24
Mini-Project2 — ssh -i private_key ubuntu@155.98.37.64 — 130×22
Mini-Project2 — ssh -i private_key ubuntu@155.98.37.64 — 114×24
Comments
Editing
Share
Styles Pane
Dictate
Sensitivity
Add-ins
Editor

```
UN  10.254.3.185  -5249387427 bytes  16      34.2%             81153ddf-71ef-4a5a-bd45-b5e74995b467  rack1
DN  10.254.1.186  -3644097644 bytes  16      27.5%             78714d64-a079-49b9-bd80-b5815cc264f5  rack1

ubuntu@3:~$ nodetool status
Datacenter: datacenter1
=======================
Status=Up/Down
|/ State=Normal/Leaving/Joining/Moving
--  Address       Load
UN  10.254.0.26   -4654672612 b
UN  10.254.3.185  -5249387427 b
DN  10.254.1.186  -3644097644 b

ubuntu@3:~$ nodetool status
Datacenter: datacenter1
=======================
Status=Up/Down
|/ State=Normal/Leaving/Joining
--  Address       Load
UN  10.254.0.26   -4654672612 b
UN  10.254.3.185  -5249387427 b
UN  10.254.1.186  -3644097644 b

ubuntu@3:~$ 
```

```
cqlsh> exit;
ubuntu@3-1:~$ nodetool status
Datacenter: datacenter1
=======================
Status=Up/Down
|/ State=Normal/Leaving/Joining/Moving
--  Address       Load               Tokens  Owns (effective)  Host ID                               Rack
UN  10.254.0.26   -4654672612 bytes  16      38.3%             a8b59fcb-7ace-496e-902e-175ce8d880cf  rack1
UN  10.254.3.185  -5249387427 bytes  16      34.2%             81153ddf-71ef-4a5a-bd45-b5e74995b467  rack1
UN  10.254.1.186  -3767767203 bytes  16      27.5%             78714d64-a079-49b9-bd80-b5815cc264f5  rack1

ubuntu@3-1:~$ nodetool status
Datacenter: datacenter1
=======================
Status=Up/Down
|/ State=Normal/Leaving/Joining/Moving
--  Address       Load               Tokens
UN  10.254.0.26   -4654672612 bytes  16
UN  10.254.3.185  -5249387427 bytes  16
UN  10.254.1.186  -3767767203 bytes  16

ubuntu@3-1:~$ 
```

```
Datacenter: datacenter1
=======================
Status=Up/Down
|/ State=Normal/Leaving/Joining/Moving
--  Address       Load               Tokens  Owns (effective)  Host ID                               Rack
UN  10.254.0.26   -4654672612 bytes  16      38.3%             a8b59fcb-7ace-496e-902e-175ce8d880cf  rack1
UN  10.254.3.185  -5249387427 bytes  16      34.2%             81153ddf-71ef-4a5a-bd45-b5e74995b467  rack1
UN  10.254.1.186  -3767767203 bytes  16      27.5%             78714d64-a079-49b9-bd80-b5815cc264f5  rack1

ubuntu@3-2:~$ sudo systemctl status cassandra
● cassandra.service - LSB: distributed storage system for structured data
   Loaded: loaded (/etc/init.d/cassandra; generated)
   Active: active (running) since Sun 2024-04-07 12:23:50 UTC; 7h ago
     Docs: man:systemd-sysv-generator(8)
  Process: 20645 ExecStop=/etc/init.d/cassandra stop (code=exited, status=0/SUCCESS)
  Process: 20658 ExecStart=/etc/init.d/cassandra start (code=exited, status=0/SUCCESS)
    Tasks: 216 (limit: 2361)
   CGroup: /system.slice/cassandra.service
           └─20755 /usr/bin/java -ea -da:net.openhft... -XX:+UseThreadPriorities -XX:+HeapDumpOnOutOfMemoryError -

Apr 07 12:23:50 3-2 systemd[1]: Stopped LSB: distributed storage system for structured data.
Apr 07 12:23:50 3-2 systemd[1]: Starting LSB: distributed storage system for structured data...
Apr 07 12:23:50 3-2 systemd[1]: Started LSB: distributed storage system for structured data.
lines 1-13/13 (END)
```


Google Chrome
File
Edit
View
History
Bookmarks
Profiles
Tab
Window
Help
Mini-Project2 — ssh -i private_key ubuntu@155.98.37.64 — 135×24
Mini-Project2 — ssh -i private_key ubuntu@155.98.37.64 — 130×22
Mini-Project2 — ssh -i private_key ubuntu@155.98.37.64 — 114×24
Comments
Editing
Share
Styles Pane
Dictate
Sensitivity
Add-ins
Editor

```
UN  10.254.3.185  -5249387427 bytes  16      34.2%             81153ddf-71ef-4a5a-bd45-b5e74995b467  rack1
DN  10.254.1.186  -3644097644 bytes  16      27.5%             78714d64-a079-49b9-bd80-b5815cc264f5  rack1

ubuntu@3:~$ nodetool status
Datacenter: datacenter1
=======================
Status=Up/Down
|/ State=Normal/Leaving/Joining/Moving
--  Address       Load
UN  10.254.0.26   -4654672612 b
UN  10.254.3.185  -5249387427 b
DN  10.254.1.186  -3644097644 b

ubuntu@3:~$ nodetool status
Datacenter: datacenter1
=======================
Status=Up/Down
|/ State=Normal/Leaving/Joining
--  Address       Load
UN  10.254.0.26   -4654672612 b
UN  10.254.3.185  -5249387427 b
UN  10.254.1.186  -3644097644 b

ubuntu@3:~$ 
```

```
cqlsh> exit;
ubuntu@3-1:~$ nodetool status
Datacenter: datacenter1
=======================
Status=Up/Down
|/ State=Normal/Leaving/Joining/Moving
--  Address       Load               Tokens  Owns (effective)  Host ID                               Rack
UN  10.254.0.26   -4654672612 bytes  16      38.3%             a8b59fcb-7ace-496e-902e-175ce8d880cf  rack1
UN  10.254.3.185  -5249387427 bytes  16      34.2%             81153ddf-71ef-4a5a-bd45-b5e74995b467  rack1
UN  10.254.1.186  -3767767203 bytes  16      27.5%             78714d64-a079-49b9-bd80-b5815cc264f5  rack1

ubuntu@3-1:~$ nodetool status
Datacenter: datacenter1
=======================
Status=Up/Down
|/ State=Normal/Leaving/Joining/Moving
--  Address       Load               Tokens
UN  10.254.0.26   -4654672612 bytes  16
UN  10.254.3.185  -5249387427 bytes  16
UN  10.254.1.186  -3767767203 bytes  16

ubuntu@3-1:~$ 
```

```
Last login: Sun Apr  7 17:23:25 2024 from 10.254.0.26
ubuntu@3-2:~$ nodetool status
Datacenter: datacenter1
=======================
Status=Up/Down
|/ State=Normal/Leaving/Joining/Moving
--  Address       Load               Tokens  Owns (effective)  Host ID                               Rack
UN  10.254.0.26   -4732821505 bytes  16      38.3%             a8b59fcb-7ace-496e-902e-175ce8d880cf  rack1
UN  10.254.3.185  -5319392744 bytes  16      34.2%             81153ddf-71ef-4a5a-bd45-b5e74995b467  rack1
UN  10.254.1.186  -3699901431 bytes  16      27.5%             78714d64-a079-49b9-bd80-b5815cc264f5  rack1

ubuntu@3-2:~$ nodetool status
Datacenter: datacenter1
=======================
Status=Up/Down
|/ State=Normal/Leaving/Joining/Moving
--  Address       Load               Tokens  Owns (effective)  Host ID                               Rack
UN  10.254.0.26   -4654672612 bytes  16      38.3%             a8b59fcb-7ace-496e-902e-175ce8d880cf  rack1
UN  10.254.3.185  -5249387427 bytes  16      34.2%             81153ddf-71ef-4a5a-bd45-b5e74995b467  rack1
UN  10.254.1.186  -3767767203 bytes  16      27.5%             78714d64-a079-49b9-bd80-b5815cc264f5  rack1

ubuntu@3-2:~$ 
```

PART-2:

1. Data preparation:

- Run the notebook to generate csv file “access_logs_comma.csv”
- Copy the csv fike to VM

```
scp -i private_key access_logs_comma.csv ubuntu@<ip_address>:acces_logs_comma.csv
```

2. Create a Keyspace and Table:

- Create a keyspace to contain our table:

```
    CREATE KEYSPACE IF NOT EXISTS logs_keyspace WITH replication = {'class': 'SimpleStrategy', 'replication_factor': 1};
```

- Create a table to store log data

```
    CREATE TABLE IF NOT EXISTS logs_keyspace.logs_table (
        log_id INT PRIMARY KEY,
        ip_address TEXT,
        path TEXT,
        protocol TEXT,
        request_method TEXT,
        request_timestamp TIMESTAMP,
        response_code TEXT,
        size INT,
        user_agent TEXT,
    );
```

3.. Import data into cassandra

Using CQL COPY command:

```
    COPY logs_keyspace.logs_table(log_id, timestamp, ip_address, request_method, path, protocol, response_code, size, user_agent)
    FROM 'access_logs.csv'
    WITH HEADER = true ;
```

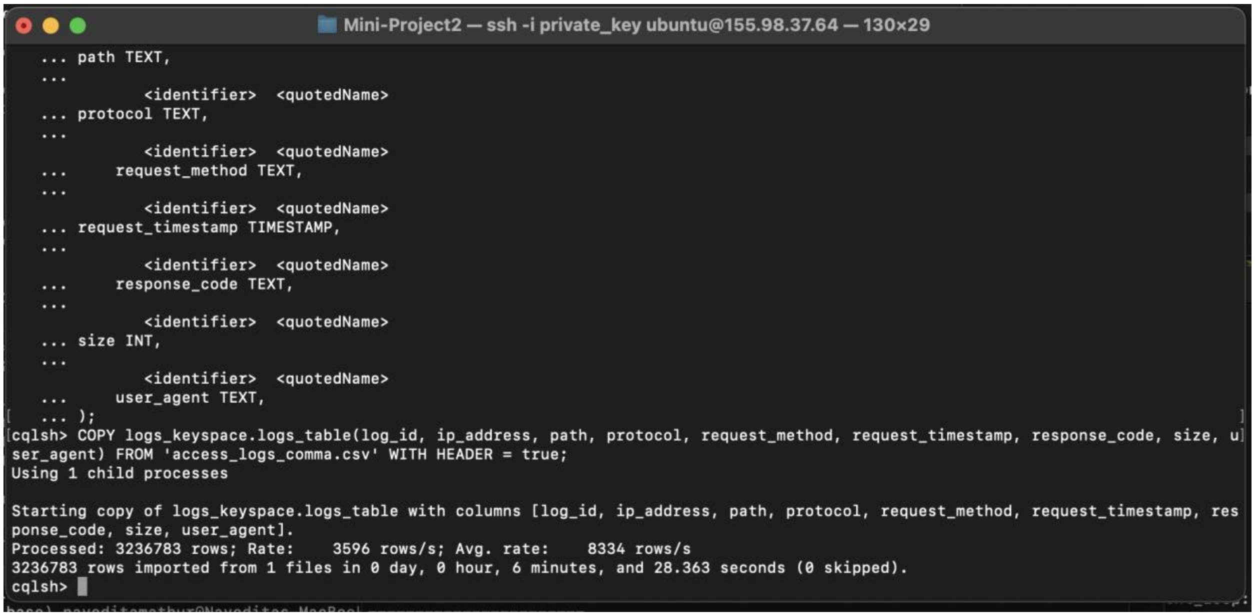
3. Verify Data Import:

- Connect to your Cassandra instance using cqlsh.

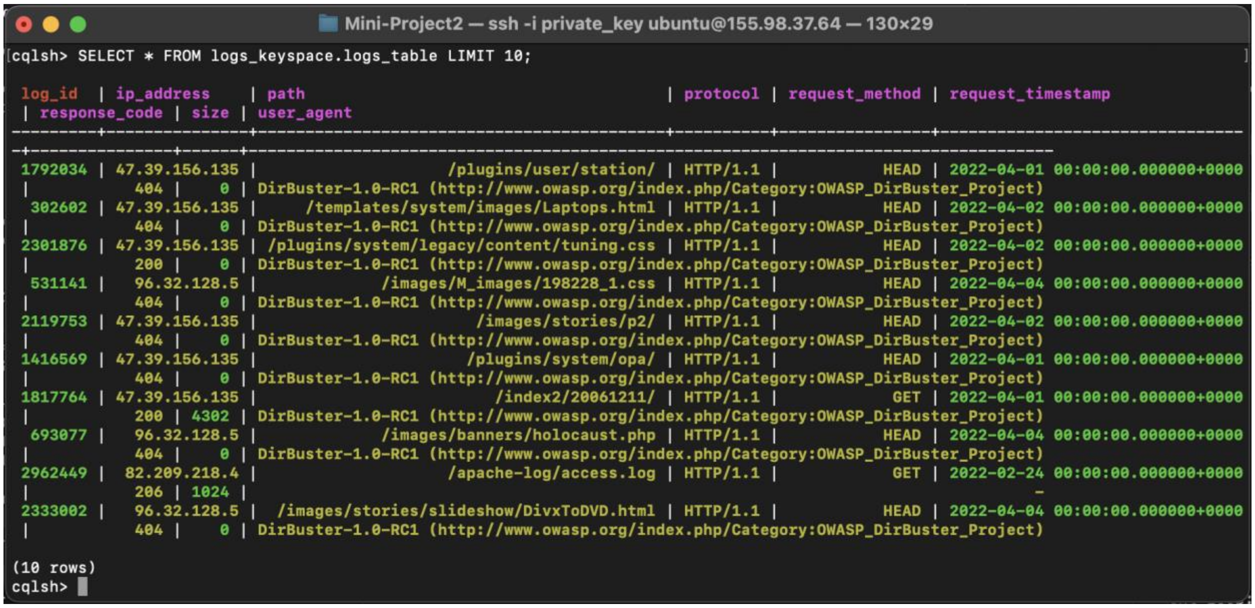
```
    cqlsh 10.254.0.26 9042 --request-timeout=60000
```

- Query the data to ensure it has been imported successfully:

```
    SELECT * FROM logs_keyspace.logs_table LIMIT 10;
```


Mini-Project2 — ssh -i private_key ubuntu@155.98.37.64 — 130×29

```
  ... path TEXT,
  ...
             <identifier>  <quotedName>
  ... protocol TEXT,
  ...
             <identifier>  <quotedName>
  ...     request_method TEXT,
  ...
             <identifier>  <quotedName>
  ... request_timestamp TIMESTAMP,
  ...
             <identifier>  <quotedName>
  ...     response_code TEXT,
  ...
             <identifier>  <quotedName>
  ... size INT,
  ...
             <identifier>  <quotedName>
  ...     user_agent TEXT,
  ... );
cqlsh> COPY logs_keyspace.logs_table(log_id, ip_address, path, protocol, request_method, request_timestamp, response_code, size, u
ser_agent) FROM 'access_logs_comma.csv' WITH HEADER = true;
Using 1 child processes

Starting copy of logs_keyspace.logs_table with columns [log_id, ip_address, path, protocol, request_method, request_timestamp, res
ponse_code, size, user_agent].
Processed: 3236783 rows; Rate:    3596 rows/s; Avg. rate:    8334 rows/s
3236783 rows imported from 1 files in 0 day, 0 hour, 6 minutes, and 28.363 seconds (0 skipped).
cqlsh> 
```


Mini-Project2 — ssh -i private_key ubuntu@155.98.37.64 — 130×29

```
cqlsh> SELECT * FROM logs_keyspace.logs_table LIMIT 10;

 log_id  | ip_address     | path                                      | protocol | request_method | request_timestamp
 | response_code | size | user_agent
---------+----------------+-------------------------------------------+----------+----------------+---------------------------------
-+---------------+------+--------------------------------------------------------------------------------------
 1792034 | 47.39.156.135 |                     /plugins/user/station/ | HTTP/1.1 |           HEAD | 2022-04-01 00:00:00.000000+0000
 |           404 |    0 | DirBuster-1.0-RC1 (http://www.owasp.org/index.php/Category:OWASP_DirBuster_Project)
  302602 | 47.39.156.135 |      /templates/system/images/Laptops.html | HTTP/1.1 |           HEAD | 2022-04-02 00:00:00.000000+0000
 |           404 |    0 | DirBuster-1.0-RC1 (http://www.owasp.org/index.php/Category:OWASP_DirBuster_Project)
 2301876 | 47.39.156.135 | /plugins/system/legacy/content/tuning.css | HTTP/1.1 |           HEAD | 2022-04-02 00:00:00.000000+0000
 |           200 |    0 | DirBuster-1.0-RC1 (http://www.owasp.org/index.php/Category:OWASP_DirBuster_Project)
  531141 |   96.32.128.5 |              /images/M_images/198228_1.css | HTTP/1.1 |           HEAD | 2022-04-04 00:00:00.000000+0000
 |           404 |    0 | DirBuster-1.0-RC1 (http://www.owasp.org/index.php/Category:OWASP_DirBuster_Project)
 2119753 | 47.39.156.135 |                        /images/stories/p2/ | HTTP/1.1 |           HEAD | 2022-04-02 00:00:00.000000+0000
 |           404 |    0 | DirBuster-1.0-RC1 (http://www.owasp.org/index.php/Category:OWASP_DirBuster_Project)
 1416569 | 47.39.156.135 |                       /plugins/system/opa/ | HTTP/1.1 |           HEAD | 2022-04-01 00:00:00.000000+0000
 |           404 |    0 | DirBuster-1.0-RC1 (http://www.owasp.org/index.php/Category:OWASP_DirBuster_Project)
 1817764 | 47.39.156.135 |                          /index2/20061211/ | HTTP/1.1 |            GET | 2022-04-01 00:00:00.000000+0000
 |           200 | 4302 | DirBuster-1.0-RC1 (http://www.owasp.org/index.php/Category:OWASP_DirBuster_Project)
  693077 |   96.32.128.5 |              /images/banners/holocaust.php | HTTP/1.1 |           HEAD | 2022-04-04 00:00:00.000000+0000
 |           404 |    0 | DirBuster-1.0-RC1 (http://www.owasp.org/index.php/Category:OWASP_DirBuster_Project)
 2962449 |  82.209.218.4 |                     /apache-log/access.log | HTTP/1.1 |            GET | 2022-02-24 00:00:00.000000+0000
 |           206 | 1024 |                                                                                     -
 2333002 |   96.32.128.5 |   /images/stories/slideshow/DivxToDVD.html | HTTP/1.1 |           HEAD | 2022-04-04 00:00:00.000000+0000
 |           404 |    0 | DirBuster-1.0-RC1 (http://www.owasp.org/index.php/Category:OWASP_DirBuster_Project)

(10 rows)
cqlsh> 
```

PART-3

Create the following functions:

```
CREATE OR REPLACE FUNCTION logs_keyspace.state_group_count( state map<text, int>, type
text )
CALLED ON NULL INPUT
RETURNS map<text, int>
LANGUAGE java AS '
```

```
Integer count = (Integer) state.get(type);  if (count == null) count = 1; else count++;
state.put(type, count); return state; ' ;

CREATE OR REPLACE AGGREGATE logs_keyspace.group_count(text)
SFUNC state_group_count
STYPE map<text, int>
INITCOND {};


CREATE CUSTOM INDEX user_agent_idx ON logs_keyspace.logs_table(user_agent) USING
'org.apache.cassandra.index.sasi.SASIIndex' WITH OPTIONS = {'mode':'CONTAINS',
'case_sensitive':'false'};

CREATE OR REPLACE FUNCTION logs_keyspace.state_group_and_count( state map<text, int>,
type text )
CALLED ON NULL INPUT
RETURNS map<text, int>
LANGUAGE java AS '
Integer count = (Integer) state.get(type);  if (count == null) count = 1; else count++;
state.put(type, count); return state; ' ;


CREATE OR REPLACE FUNCTION logs_keyspace.state_max_group_count (state map<text, int>)
CALLED ON NULL INPUT
RETURNS text LANGUAGE JAVA AS '
    String maxValue = null;
    int maxCount = 0;
    String result = null;
    for (String key : state.keySet()) {
      int value = state.get(key);
      if (value > maxCount) {
        maxValue = key; maxCount = value;
      }
    }
    result = maxValue + ": " + Integer.toString(maxCount);
    return result; ' ;


CREATE OR REPLACE AGGREGATE logs_keyspace.max_group_count(text)
SFUNC state_group_and_count
STYPE map<text, int>
FINALFUNC state_max_group_count
INITCOND {};
```

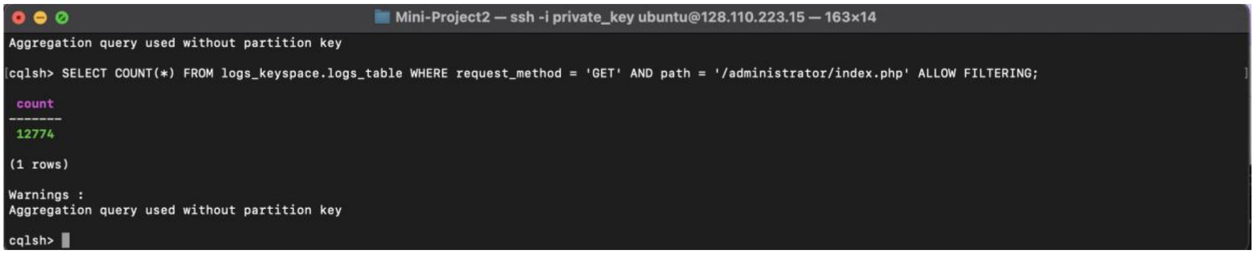
```
CREATE OR REPLACE FUNCTION logs_keyspace.state_group_count_having( state map<text,
int>, type text )
CALLED ON NULL INPUT
RETURNS map<text, int>
LANGUAGE java AS '
Integer count = (Integer) state.get(type);  if (count == null) count = 1; else count++;
state.put(type, count); return state; ' ;

CREATE OR REPLACE FUNCTION logs_keyspace.group_count_having(state map<text, int>)
CALLED ON NULL INPUT
RETURNS text LANGUAGE JAVA AS '
    String result = "";
    for (String key : state.keySet()) {
      int value = state.get(key);
      if (value > 10) {
        result += key + ": " + Integer.toString(value) +";";
      }
    }
    return result; ' ;

CREATE OR REPLACE AGGREGATE logs_keyspace.max_group_count_having(text)
SFUNC state_group_count_having
STYPE map<text, int>
FINALFUNC group_count_having
INITCOND {};
```

1. How many hits were made to the website item “/administrator/index.php”?

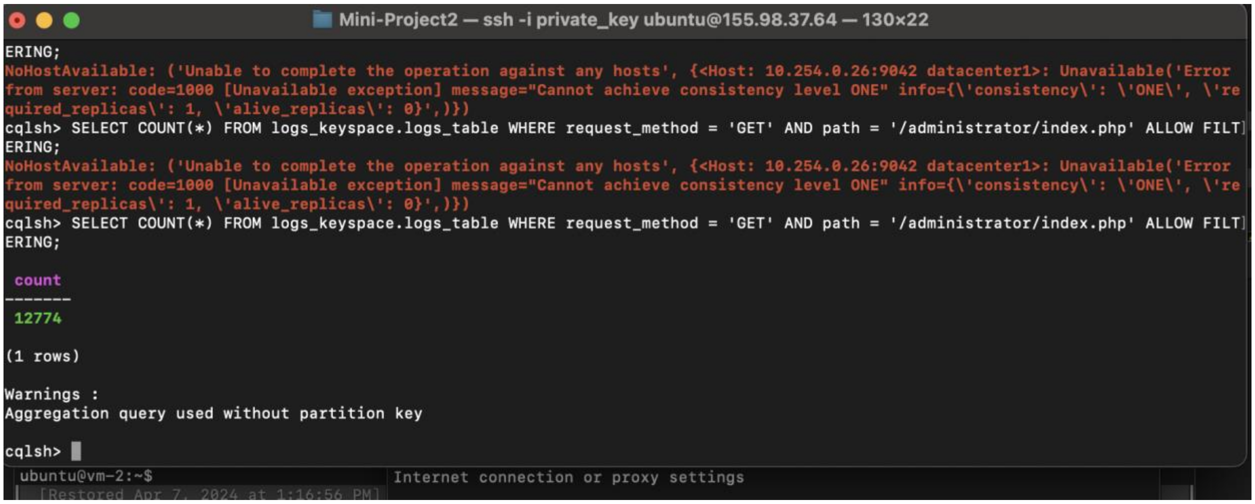
Query :

SELECT COUNT(*) FROM logs_keyspace.logs_table WHERE request_method = 'GET' AND path = '/administrator/index.php' ALLOW FILTERING;

Output:


Mini-Project2 — ssh -i private_key ubuntu@128.110.223.15 — 163×14

```
Aggregation query used without partition key

cqlsh> SELECT COUNT(*) FROM logs_keyspace.logs_table WHERE request_method = 'GET' AND path = '/administrator/index.php' ALLOW FILTERING;

 count
-------
 12774

(1 rows)

Warnings :
Aggregation query used without partition key

cqlsh> 
```



Mini-Project2 — ssh -i private_key ubuntu@155.98.37.64 — 130×22

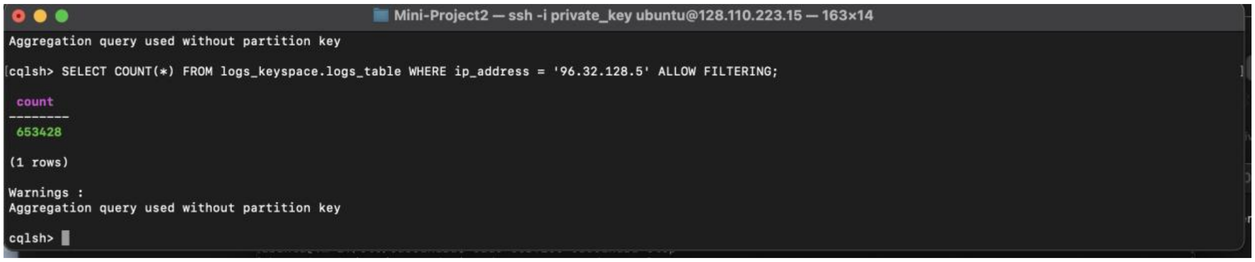
```
ERING;
NoHostAvailable: ('Unable to complete the operation against any hosts', {<Host: 10.254.0.26:9042 datacenter1>: Unavailable('Error
from server: code=1000 [Unavailable exception] message="Cannot achieve consistency level ONE" info={\'consistency\': \'ONE\', \'re
quired_replicas\': 1, \'alive_replicas\': 0}',)})
cqlsh> SELECT COUNT(*) FROM logs_keyspace.logs_table WHERE request_method = 'GET' AND path = '/administrator/index.php' ALLOW FILT
ERING;
NoHostAvailable: ('Unable to complete the operation against any hosts', {<Host: 10.254.0.26:9042 datacenter1>: Unavailable('Error
from server: code=1000 [Unavailable exception] message="Cannot achieve consistency level ONE" info={\'consistency\': \'ONE\', \'re
quired_replicas\': 1, \'alive_replicas\': 0}',)})
cqlsh> SELECT COUNT(*) FROM logs_keyspace.logs_table WHERE request_method = 'GET' AND path = '/administrator/index.php' ALLOW FILT
ERING;

 count
-------
 12774

(1 rows)

Warnings :
Aggregation query used without partition key

cqlsh> 
  ubuntu@vm-2:~$                            Internet connection or proxy settings
    [Restored Apr 7, 2024 at 1:16:56 PM]
```

2. How many hits were made from the IP: 96.32.128.5

Query:

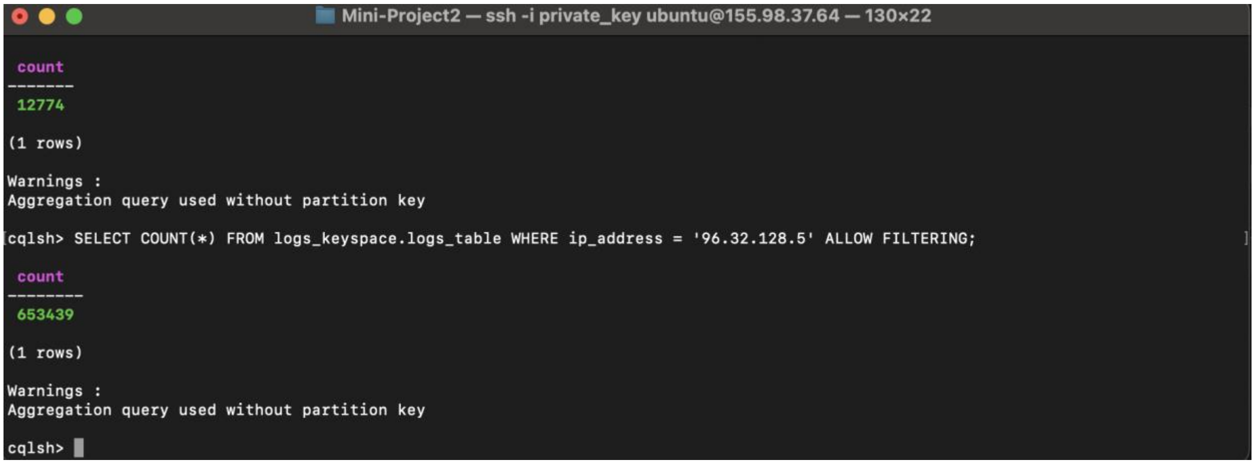
SELECT COUNT(*) FROM logs_keyspace.logs_table WHERE ip_address = '96.32.128.5' ALLOW FILTERING;

Output:



Mini-Project2 — ssh -i private_key ubuntu@128.110.223.15 — 163×14

```
Aggregation query used without partition key

cqlsh> SELECT COUNT(*) FROM logs_keyspace.logs_table WHERE ip_address = '96.32.128.5' ALLOW FILTERING;

 count
--------
 653428

(1 rows)

Warnings :
Aggregation query used without partition key

cqlsh> 
```



Mini-Project2 — ssh -i private_key ubuntu@155.98.37.64 — 130×22

```

 count
-------
 12774

(1 rows)

Warnings :
Aggregation query used without partition key

cqlsh> SELECT COUNT(*) FROM logs_keyspace.logs_table WHERE ip_address = '96.32.128.5' ALLOW FILTERING;

 count
--------
 653439

(1 rows)

Warnings :
Aggregation query used without partition key

cqlsh> 
```

3. Which path in the website has been hit most? How many hits were made to the path?

Query:

SELECT logs_keyspace.group_count(path) FROM logs_keyspace.logs_table;

Output:

4. Which IP accesses the website most? How many accesses were made by it?

Query:

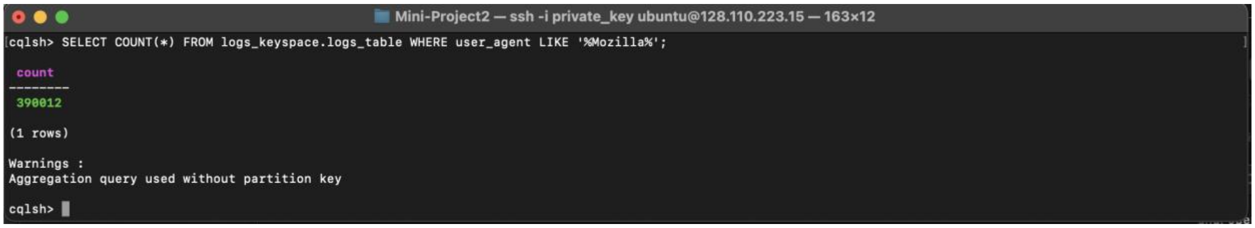
SELECT logs_keyspace.group_count(ip_address) FROM logs_keyspace.logs_table;

Output:

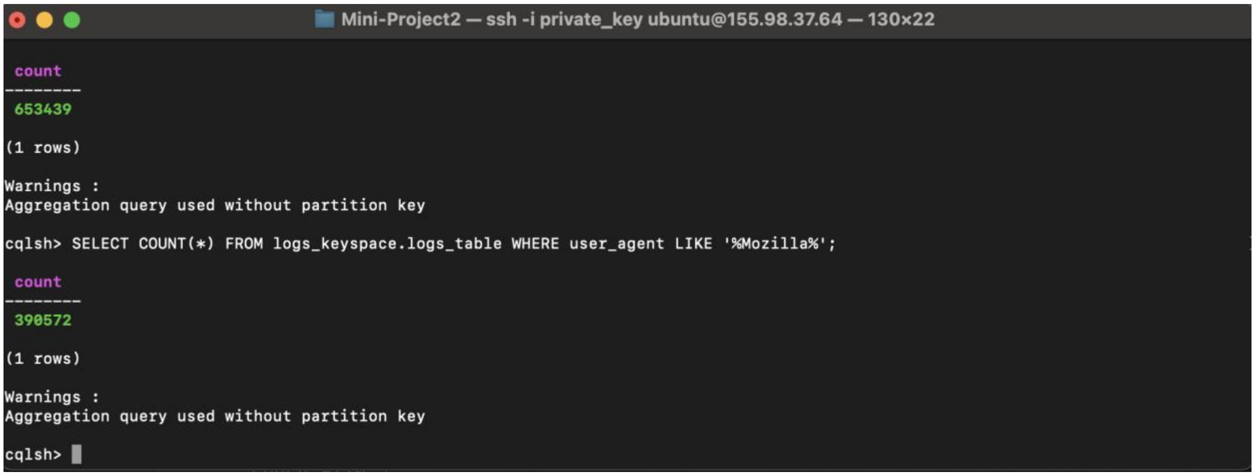
5. How many accesses were made by Firefox(Mozilla)?

Query:
SELECT COUNT(*) FROM logs_keyspace.logs_table WHERE user_agent LIKE '%Mozilla%';

Output:


Mini-Project2 — ssh -i private_key ubuntu@128.110.223.15 — 163×12

```
cqlsh> SELECT COUNT(*) FROM logs_keyspace.logs_table WHERE user_agent LIKE '%Mozilla%';

 count
--------
 390012

(1 rows)

Warnings :
Aggregation query used without partition key

cqlsh> 
```



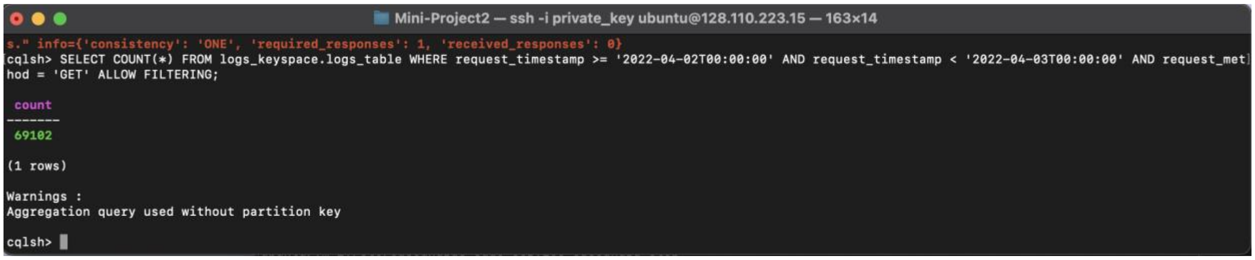
Mini-Project2 — ssh -i private_key ubuntu@155.98.37.64 — 130×22

```
 count
--------
 653439

(1 rows)

Warnings :
Aggregation query used without partition key

cqlsh> SELECT COUNT(*) FROM logs_keyspace.logs_table WHERE user_agent LIKE '%Mozilla%';

 count
--------
 390572

(1 rows)

Warnings :
Aggregation query used without partition key

cqlsh>
```

6. For all requests on 02/Apr/2022, what is the ratio of GET request?

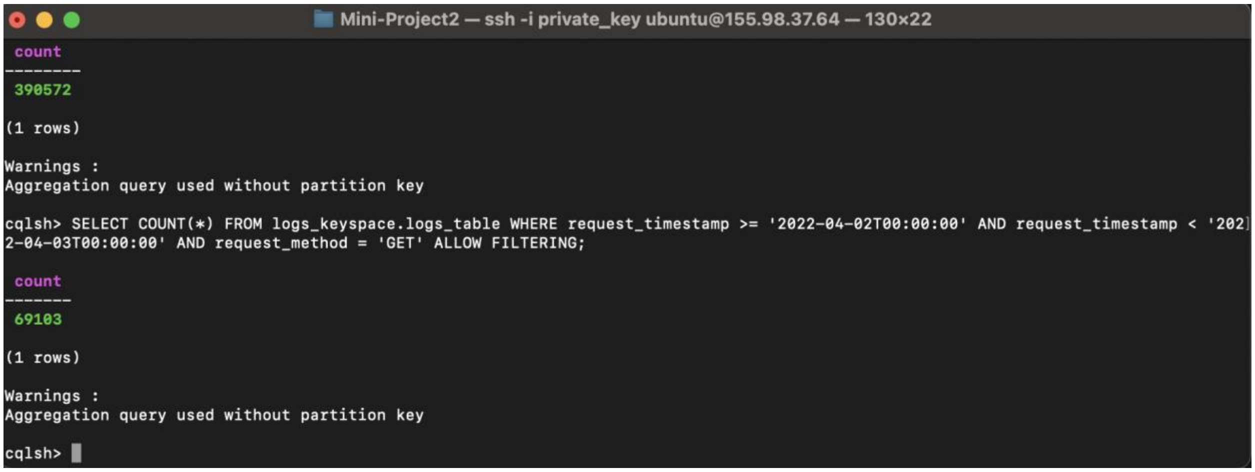
Query :

SELECT COUNT(*) FROM logs_keyspace.logs_table WHERE request_timestamp >= '2022-04-02T00:00:00' AND request_timestamp < '2022-04-03T00:00:00' AND request_method = 'GET' ALLOW FILTERING;



Mini-Project2 — ssh -i private_key ubuntu@128.110.223.15 — 163×14

```
s." info={'consistency': 'ONE', 'required_responses': 1, 'received_responses': 0}
cqlsh> SELECT COUNT(*) FROM logs_keyspace.logs_table WHERE request_timestamp >= '2022-04-02T00:00:00' AND request_timestamp < '2022-04-03T00:00:00' AND request_met
hod = 'GET' ALLOW FILTERING;

 count
-------
 69102

(1 rows)

Warnings :
Aggregation query used without partition key

cqlsh>
```

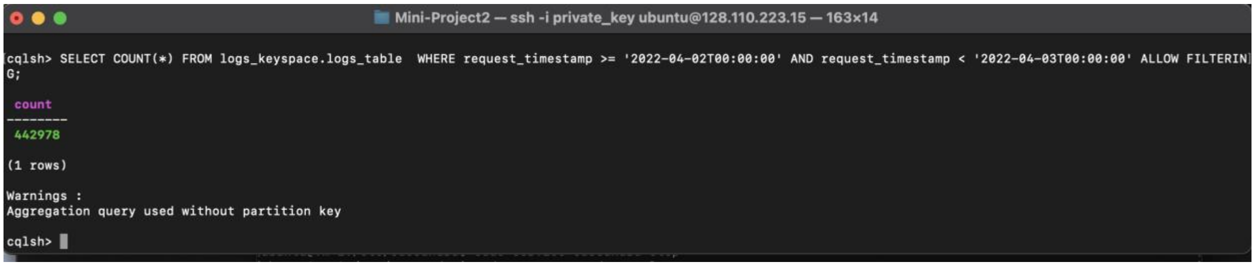


Mini-Project2 — ssh -i private_key ubuntu@155.98.37.64 — 130×22

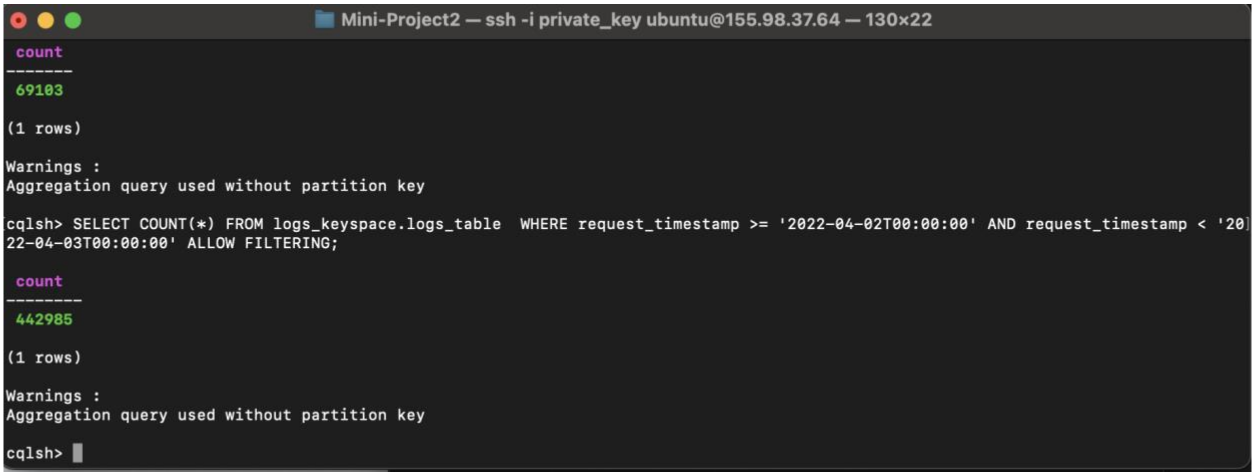
```
 count
--------
 390572

(1 rows)

Warnings :
Aggregation query used without partition key

cqlsh> SELECT COUNT(*) FROM logs_keyspace.logs_table WHERE request_timestamp >= '2022-04-02T00:00:00' AND request_timestamp < '202
2-04-03T00:00:00' AND request_method = 'GET' ALLOW FILTERING;

 count
-------
 69103

(1 rows)

Warnings :
Aggregation query used without partition key

cqlsh>
```

SELECT COUNT(*) FROM logs_keyspace.logs_table WHERE request_timestamp >= '2022-04-02T00:00:00' AND request_timestamp < '2022-04-03T00:00:00' ALLOW FILTERING;

Output:

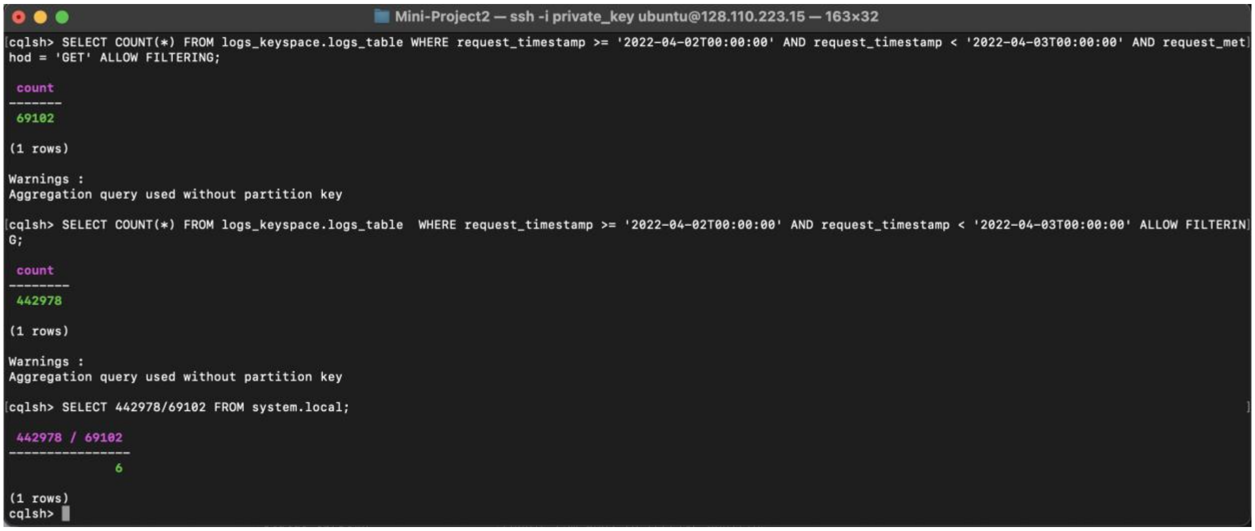

Mini-Project2 — ssh -i private_key ubuntu@128.110.223.15 — 163×14

```
cqlsh> SELECT COUNT(*) FROM logs_keyspace.logs_table  WHERE request_timestamp >= '2022-04-02T00:00:00' AND request_timestamp < '2022-04-03T00:00:00' ALLOW FILTERIN
G;

 count
--------
 442978

(1 rows)

Warnings :
Aggregation query used without partition key

cqlsh> 
```

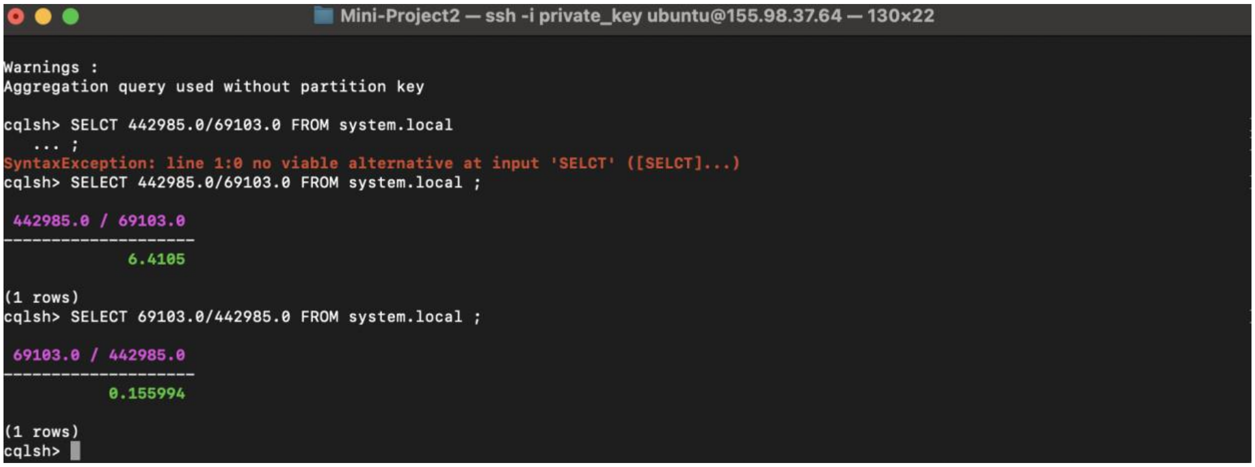

Mini-Project2 — ssh -i private_key ubuntu@155.98.37.64 — 130×22

```
 count
-------
 69103

(1 rows)

Warnings :
Aggregation query used without partition key

cqlsh> SELECT COUNT(*) FROM logs_keyspace.logs_table  WHERE request_timestamp >= '2022-04-02T00:00:00' AND request_timestamp < '20
22-04-03T00:00:00' ALLOW FILTERING;

 count
--------
 442985

(1 rows)

Warnings :
Aggregation query used without partition key

cqlsh> 
```

So, 69102/442978 = 1/6


Mini-Project2 — ssh -i private_key ubuntu@128.110.223.15 — 163×32

```
cqlsh> SELECT COUNT(*) FROM logs_keyspace.logs_table WHERE request_timestamp >= '2022-04-02T00:00:00' AND request_timestamp < '2022-04-03T00:00:00' AND request_met
hod = 'GET' ALLOW FILTERING;

 count
-------
 69102

(1 rows)

Warnings :
Aggregation query used without partition key

cqlsh> SELECT COUNT(*) FROM logs_keyspace.logs_table  WHERE request_timestamp >= '2022-04-02T00:00:00' AND request_timestamp < '2022-04-03T00:00:00' ALLOW FILTERIN
G;

 count
--------
 442978

(1 rows)

Warnings :
Aggregation query used without partition key

cqlsh> SELECT 442978/69102 FROM system.local;

 442978 / 69102
----------------
              6

(1 rows)
cqlsh> 
```


Mini-Project2 — ssh -i private_key ubuntu@155.98.37.64 — 130×22

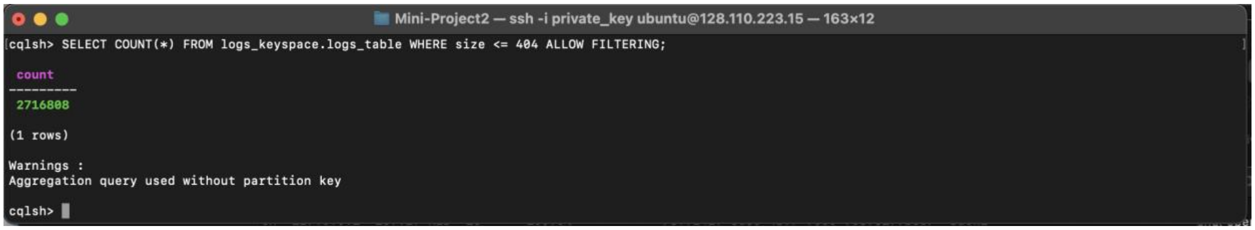
```
Warnings :
Aggregation query used without partition key

cqlsh> SELCT 442985.0/69103.0 FROM system.local
   ... ;
SyntaxException: line 1:0 no viable alternative at input 'SELCT' ([SELCT]...)
cqlsh> SELECT 442985.0/69103.0 FROM system.local ;

 442985.0 / 69103.0
--------------------
             6.4105

(1 rows)
cqlsh> SELECT 69103.0/442985.0 FROM system.local ;

 69103.0 / 442985.0
--------------------
           0.155994

(1 rows)
cqlsh> 
```

7. How many requests are lower than or equal to 404 bytes?

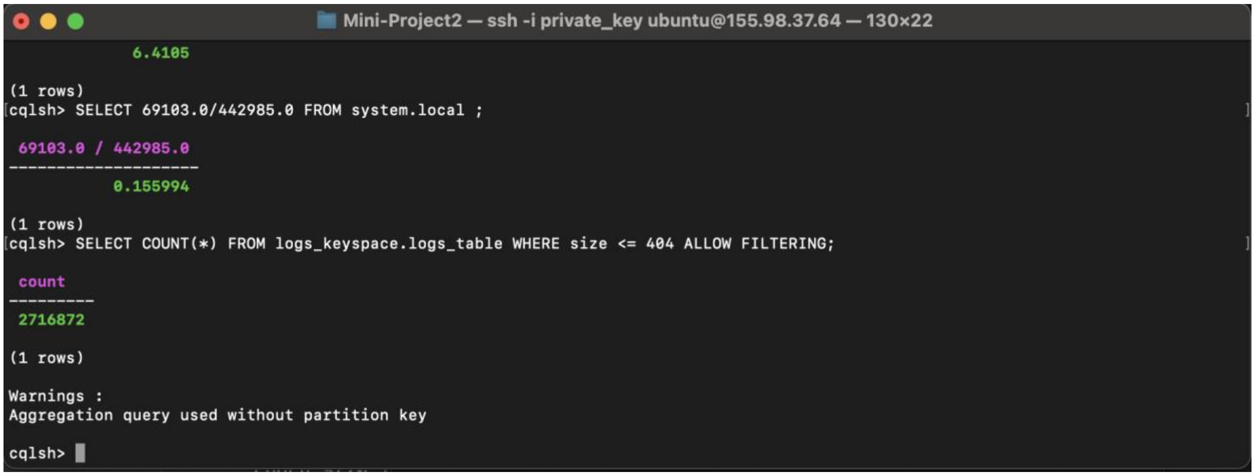
Query:

SELECT COUNT(*) FROM logs_keyspace.logs_table WHERE size <= 404 ALLOW FILTERING;

Output


Mini-Project2 — ssh -i private_key ubuntu@128.110.223.15 — 163×12

```
cqlsh> SELECT COUNT(*) FROM logs_keyspace.logs_table WHERE size <= 404 ALLOW FILTERING;

 count
---------
 2716808

(1 rows)

Warnings :
Aggregation query used without partition key

cqlsh> 
```


Mini-Project2 — ssh -i private_key ubuntu@155.98.37.64 — 130×22

```
             6.4105

(1 rows)
cqlsh> SELECT 69103.0/442985.0 FROM system.local ;

 69103.0 / 442985.0
--------------------
           0.155994

(1 rows)
cqlsh> SELECT COUNT(*) FROM logs_keyspace.logs_table WHERE size <= 404 ALLOW FILTERING;

 count
---------
 2716872

(1 rows)

Warnings :
Aggregation query used without partition key

cqlsh> 
```

8. List the IPs that have more than ten 404 requests. If no ip fulfills, print the ip that has most 404 requests and the number of requests

Query:

```
SELECT logs_keyspace.max_group_count_having(ip_address) FROM
logs_keyspace.logs_table WHERE response_code='404'ALLOW FILTERING;
```

Output:

If no IP fulfills the condition, then:

```
SELECT logs_keyspace.max_group_count(ip_address) FROM logs_keyspace.logs_table
WHERE response_code='404'ALLOW FILTERING;
```

Output:

References:

1. Hadoop Single Node Cluster on Docker
   Rodrigo Ancavil
   Link - https://medium.com/analytics-vidhya/hadoop-single-node-cluster-on-docker-e88c3d09a256
2. https://hub.docker.com/_/eclipse-temurin
3. https://medium.com/@abhikdey06/apache-hadoop-3-3-6-installation-on-ubuntu-22-04-14516bceec85
4. https://github.com/amephraim/nlp (input text)