

# INFSCI 2750 - Cloud Computing Mini Project 3 - Queries and Snapshots of Outputs Part 2 and Part 3

## Members:

1. Aishwarya Jakka, [aij12@pitt.edu](mailto:aij12@pitt.edu)
2. Shreyank Ranganath, [shr74@pitt.edu](mailto:shr74@pitt.edu)
3. Kwesi Aguilera, [kra40@pitt.edu](mailto:kra40@pitt.edu)

## PART - 2

# Step 1 - start Cassandra on all 3 VMs using command: `sudo cassandra -Rf`

# Step 2 - start CQL shell: `cqlsh CC-MON-3`

### Create Keyspace

```
student@CC-MON-3:~$ cqlsh CC-MON-3
Connected to Test Cluster at CC-MON-3:9042.
[cqlsh 5.0.1 | Cassandra 3.11.6 | CQL spec 3.4.4 | Native protocol v4]
Use HELP for help.
cqlsh> CREATE KEYSPACE online_accesslog_ks WITH replication = { 'class': 'SimpleStrategy', 'replication_factor' : 2};
cqlsh>
```

### Create Table

```
Shreyank — student@CC-MON-3: ~ — ssh -i ~/.ssh/key_student student@165.227.73.164 — 177x28
cqlsh> CREATE TABLE online_accesslog_ks.table1 (
... ID int,
... IP text,
... PATH text,
... PRIMARY KEY((ID), IP, PATH)
... );
cqlsh>
```

### Import Data from CSV

```
Shreyank — student@CC-MON-3: ~ — ssh -i ~/.ssh/key_student student@165.227.73.164 — 177x18
student@CC-MON-3:~$ cqlsh CC-MON-3
Connected to Test Cluster at CC-MON-3:9042.
[cqlsh 5.0.1 | Cassandra 3.11.6 | CQL spec 3.4.4 | Native protocol v4]
Use HELP for help.
cqlsh> COPY online_accesslog_ks.table1 (ID,IP,PATH) FROM '/home/student/Mini_proj_3/access_logs_paths.csv' WITH DELIMITER= ',' AND HEADER = TRUE and NULL = 'null';
Reading options from the command line: {'header': 'TRUE', 'delimiter': ',', 'null': 'null'}
Using 1 child processes

Starting copy of online_accesslog_ks.table1 with columns [id, ip, path].
Processed: 4477843 rows; Rate: 2920 rows/s; Avg. rate: 6494 rows/s
4477843 rows imported from 1 files in 11 minutes and 29.493 seconds (0 skipped).
cqlsh>
```

## Create Index

```
student@CC-MON-3:~$ cqlsh CC-MON-3
Connected to Test Cluster at CC-MON-3:9042.
[cqlsh 5.0.1 | Cassandra 3.11.6 | CQL spec 3.4.4 | Native protocol v4]
Use HELP for help.
cqlsh> CREATE INDEX ip_index ON online_accesslog_ks.table1 (IP);
cqlsh> CREATE INDEX path_index ON online_accesslog_ks.table1(PATH);
cqlsh>
```

## PART - 3

Part 3.1: How many hits were made to the website item `"/assets/img/release-schedule-logo.png"`?

```
student@CC-MON-3:~$ cqlsh CC-MON-3
Connected to Test Cluster at CC-MON-3:9042.
[cqlsh 5.0.1 | Cassandra 3.11.6 | CQL spec 3.4.4 | Native protocol v4]
Use HELP for help.
cqlsh> SELECT COUNT(*)
...      FROM online_accesslog_ks.table1
...      WHERE PATH = '/assets/img/release-schedule-logo.png'
...      ALLOW FILTERING;

count
-----
24292

(1 rows)

Warnings :
Aggregation query used without partition key
```

Part 3.2: How many hits were made from the IP: 10.207.188.188?

```
Shreyank — student@CC-MON-3: ~ — ssh -i ~/.ssh/key_student student@165.227.73.164 — 177x35
student@CC-MON-3:~$ cqlsh CC-MON-3
Connected to Test Cluster at CC-MON-3:9042.
[cqlsh 5.0.1 | Cassandra 3.11.6 | CQL spec 3.4.4 | Native protocol v4]
Use HELP for help.
cqlsh> SELECT COUNT(*)
... FROM online_accesslog_ks.table1
... WHERE IP = '10.207.188.188'
... ALLOW FILTERING;

count
-----
398

(1 rows)

Warnings :
Aggregation query used without partition key
```

=====

## Part 3.3 and 3.4

=====

=====

### Queries:

=====

#### User-defined function 1:

```
Shreyank — student@CC-MON-3: ~ — ssh -i ~/.ssh/key_student student@165.227.73.164 — 118x47
student@CC-MON-3:~$ cqlsh CC-MON-3
Connected to Test Cluster at CC-MON-3:9042.
[cqlsh 5.0.1 | Cassandra 3.11.6 | CQL spec 3.4.4 | Native protocol v4]
Use HELP for help.
cqlsh> USE online_accesslog_ks;
cqlsh:online_accesslog_ks> CREATE OR REPLACE FUNCTION online_accesslog_ks.input_count(state map<text, int>, path text)
... CALLED ON NULL INPUT
... RETURNS map<text, int>
... LANGUAGE java
... AS $$
... Integer count = (Integer) state.get(path);
... if (count == null)
...     count = 1;
... else
...     count++;
... state.put(path, count);
... return state;
... $$;
cqlsh:online_accesslog_ks>
```

#### User-defined function 2:

```

Shreyank — student@CC-MON-3: ~ — ssh -i ~/.ssh/key_student student@165.227.73.164 — 118x29
cqlsh:online_accesslog_ks> CREATE OR REPLACE FUNCTION online_accesslog_ks.find_max_count(input map<text, int>)
... RETURNS NULL ON NULL INPUT
... RETURNS map<text, int>
... LANGUAGE java
... AS $$
... Integer max_value = Integer.MIN_VALUE;
... String data="";
... for (String temp1 : input.keySet())
... {
...     Integer temp2 = input.get(temp1);
...     if (temp2 > max_value)
...     {
...         max_value = temp2;
...         data = temp1;
...     }
... }
... Map<String,Integer> max_map = new HashMap<String,Integer>();
... max_map.put(data,max_value);
... return max_map;
... $$;
cqlsh:online_accesslog_ks>

```

Aggregate function 3:

```

Shreyank — student@CC-MON-3: ~ — ssh -i ~/.ssh/key_student student@165.227.73.164 — 118x9
cqlsh:online_accesslog_ks> CREATE OR REPLACE AGGREGATE max_hits(text)
... SFUNC input_count
... STYPE map<text, int>
... FINALFUNC find_max_count
... INITCOND {};
cqlsh:online_accesslog_ks>

```

## Results 3.3 and 3.4

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

```

Shreyank — student@CC-MON-3: ~ — ssh -i ~/.ssh/key_student student@165.227.73.164 — 118x12
cqlsh:online_accesslog_ks> select max_hits(PATH) from table1 allow filtering;

online_accesslog_ks.max_hits(path)
-----
{'/assets/css/combined.css': 117348}

(1 rows)

Warnings :
Aggregation query used without partition key
cqlsh:online_accesslog_ks>

```

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

```
Shreyank — student@CC-MON-3: ~ — ssh -i ~/.ssh/key_student student@165.227.73.164 — 118x12
cqlsh:online_accesslog_ks> select max_hits(ip) from table1 allow filtering;

online_accesslog_ks.max_hits(ip)
-----
{'10.216.113.172': 158614}

(1 rows)

Warnings :
Aggregation query used without partition key

cqlsh:online_accesslog_ks> █
```