



## Week 9: Cassandra Query Language

January 28

Due date: Wednesday January 29, 2024 at 6:00pm!!!

1. สร้างคีย์สเปซชื่อ `user_management`

```
cqlsh> create keyspace user_management
... with replication = {
... 'class': 'SimpleStrategy',
... 'replication_factor':1};
```

2. ตั้งค่า `user_management` เป็นคีย์สเปซเริ่มต้น (default keyspace)

```
cqlsh> use user_management;
cqlsh:user_management> |
```

### 3. สร้างตารางชื่อ `users` ดังต่อไปนี้

- 4 คอลัมน์: `last_name`, `first_name`, `email` และ `address`
- คอลัมน์ `last_name`: คีย์พาร์ติชัน
- คอลัมน์ `first_name` และ `email`: คอลัมน์คัสเตอร์
- คอลัมน์ `address`: คอลัมน์เพย์โหลด

```
cqlsh:user_management> create table users (
... last_name text,
... first_name text,
... email text,
... primary key (last_name, first_name, email));
```

### 4. ตรวจสอบโครงสร้างตาราง

```
cqlsh:user_management> select * from users
... ;
```

last_name	first_name	email	address
(0 rows)			

(0 rows)

5. ใส่แถวในตารางด้วยค่าต่อไปนี้:

Column Name	Value
last_name	Smith
first_name	Alex
email	asmith@gmail.com
address	123 Main st.

```
... values ('Smith', 'Alex', 'asmith@gmail.com', '123 Main st.');
```

```
cqlsh:user_management> select * from users ;
```

last_name	first_name	email	address
Smith	Alex	asmith@gmail.com	123 Main st.

```
(1 rows)
```

6. ใส่แถวในตารางด้วยค่าต่อไปนี้:

Column Name	Value
last_name	Smith
first_name	Bailey
email	bsmith@gmail.com
address	234 Elm st.

```
cqlsh:user_management> insert into users (last_name, first_name, email, address)
... values ('Smith', 'Bailey', 'bsmith@gmail.com', '234 Elm st.');
```

```
cqlsh:user_management> select * from users ;
```

last_name	first_name	email	address
Smith	Alex	asmith@gmail.com	123 Main st.
Smith	Bailey	bsmith@gmail.com	234 Elm st.

```
(2 rows)
```

7. ใส่แถวในตารางด้วยค่าต่อไปนี้ (Cassandra รองรับคอลัมน์แบบ sparse ซึ่งหมายความว่าคอลัมน์ที่ไม่ใช่คีย์หลักไม่จำเป็นต้องมีค่า):

Column Name	Value
last_name	Jones
first_name	Chris
email	cjones@gmail.com

```
cqlsh:user_management> insert into users (last_name, first_name, email)
values ('Jones', 'Chris', '
cjones@gmail.com');
```

8. แสดงเนื้อหาของตาราง โดยใช้คำสั่ง CQL **SELECT**

```
cqlsh:user_management> select * from users ;
```

last_name	first_name	email	address
Jones	Chris	cjones@gmail.com	null
Smith	Alex	asmith@gmail.com	123 Main st.
Smith	Bailey	bsmith@gmail.com	234 Elm st.

(3 rows)

## 9. เหตุใดจึงเลือกคอลัมน์เหล่านี้สำหรับคีย์หลัก?

ดูวิธีดึงข้อมูลพาร์ติชันเดียวโดยใช้คำสั่ง WHERE: ให้แสดงเนื้อหาของตาราง โดยคอลัมน์ `last_name` มีค่า Smith

```
cqlsh:user_management> select * from users where last_name = 'Smith';
```

last_name	first_name	email	address
Smith	Alex	asmith@gmail.com	123 Main st.
Smith	Bailey	bsmith@gmail.com	234 Elm st.

สมมติว่าเราไม่ต้องการดึงข้อมูลคอลัมน์ทั้งหมดจากแถว ให้ดึงเฉพาะคอลัมน์ `address` แทน:

```
cqlsh:user_management> select address from users where last_name = 'Smith';
```

address
123 Main st.
234 Elm st.

หากต้องการหลายฟิลด์ เราสามารถใช้ตัวค้นด้วยเครื่องหมายจุลภาค (,) ได้ ลองดึงคอลัมน์

`first_name` และ `email`:

```
cqlsh:user_management> select first_name, email from users where last_name = 'Smith';
```

first_name	email
Alex	asmith@gmail.com
Bailey	bsmith@gmail.com

(2 rows)

เกิดอะไรขึ้นถ้าต้องการดึงผู้ใช้ทั้งหมด โดยให้คอลัมน์ first\_name มีค่า Alex?

```
SELECT * FROM users WHERE first_name = 'Alex';
```

```
cqlsh:user_management> SELECT * FROM users WHERE first_name = 'Alex';
InvalidRequest: Error from server: code=2200 [Invalid query] message="Ca
nnot execute this query as it might involve data filtering and thus may
have unpredictable performance. If you want to execute this query despit
e the performance unpredictability, use ALLOW FILTERING"
```

## 10. เหตุใดการสืบค้น (queries) จึงต้องมีคีย์พาร์ติชันด้วย?

เราจะเห็นว่าการสืบค้นโดยไม่ใช้คีย์พาร์ติชันไม่ทำงาน ลองใช้การสืบค้นอื่น:

```
SELECT * FROM users WHERE last_name = 'Smith' AND email = 'asmith@gmail.com';
```

**ตอบ** Cassandra จะค้นหาจาก Partition Key แล้วถึงจะเข้าไปเป็น Clustering Column และ Payload Column ด้วยเหตุผลนี้จึงไม่สามารถหาข้อมูลโดยอิงจาก Column ที่ไม่ใช่ Partition Key โดยการหาข้อมูลจะเรียงลำดับเป็น Partition Key, Clustering Column ที่ 1, Clustering Column ที่ 2 และถัด ๆ ไป และสุดท้ายคือ Payload Column คำสั่ง

```
SELECT * FROM users WHERE last_name = 'Smith' AND email = 'asmith@gmail.com';
```

จึงไม่สามารถใช้งานได้

เพราะในชื่อก่อนหน้าสร้างตารางเรียง Clustering Column เป็น first\_name และ email

## 11. เหตุใดคอลัมน์คัสเตอร์จึงต้องเรียงตามลำดับการสืบค้น?

Cassandra เก็บแถวของพาร์ติชันที่เรียงตามคอลัมน์คัสเตอร์ ดังนั้นลองสร้างการสืบค้น email และ address ของผู้ใช้ที่คอลัมน์ last\_name มีค่า Smith และ first\_name มีค่า Bailey ได้หรือไม่

```
cqlsh:user_management> select email, address from users
... where last_name = 'Smith' and first_name = 'Bailey';
```

email	address
bsmith@gmail.com	234 Elm st.

(1 rows)

ตอบ ทำได้ เพราะ last\_name คือ partition key first\_name คือ Clustering Column ที่ 1 ไม่ได้ใช้ where แบบข้าม Clustering Column เหมือนข้อที่แล้ว

## 12. เมื่อต้องการเปลี่ยนค่าคอลัมน์ เราใช้คำสั่ง CQL UPDATE

ลองเปลี่ยนค่าคอลัมน์ address ของ Alex Smith ให้เป็น 987 Main St.:

```
cqlsh:user_management> update users set address = '987 Main St.' where last_name = 'Smith' and first_name = 'Alex' and email = 'asmith@gmail.com';
```

จากเดิม Chris Jones ไม่มีค่าคอลัมน์ address ลองให้ค่าเป็น 789 Cherry St:

```
cqlsh:user_management> update users
... set address = '789 Cherry St.'
... where last_name = 'Jones' and first_name = 'Chris' and email = 'cjones@gmail.com';
cqlsh:user_management> select * from users;
```

last_name	first_name	email	address
Jones	Chris	cjones@gmail.com	789 Cherry St.
Smith	Alex	asmith@gmail.com	987 Main St.
Smith	Bailey	bsmith@gmail.com	234 Elm st.

(3 rows)

เกิดอะไรขึ้นถ้าต้องการเปลี่ยนอีเมลของ Alex Smith เป็น alexsmith@gmail.com?

```
UPDATE users SET email = 'alexsmith@gmail.com' WHERE last_name = 'Smith'
AND first_name = 'Alex' AND email = 'asmith@gmail.com';
```

```
cqlsh:user_management> update users set email = 'alexsmith@gmail.com'
... where last_name = 'Smith' and first_name = 'Alex'
and email = 'asmith@gmail.com';
InvalidRequest: Error from server: code=2200 [Invalid query] message="PRIMARY KEY part email found in SET part"
```

ตอบ Cassandra ไม่อนุญาตให้แก้ไข Primary Key (ทั้ง Partition Key และ Clustering Column) email คือ Clustering Column จึงไม่สามารถแก้ไขได้

### 13. อัปเดตแถวที่ไม่มีอยู่:

```
UPDATE users SET address = '987 Elm St.' WHERE last_name = 'Jones'
AND first_name = 'Drew' AND email = 'djones@gmail.com';
```

ดูเหมือนว่าจะทำงานโดยไม่มีข้อผิดพลาด เกิดอะไรขึ้น ตรวจสอบโดยลองดึงผู้ใช้ทั้งหมดในพาร์ติชัน Jones?:

```
cqlsh:user_management> update users set address = '987 Elm St.'
... where last_name = 'Jones'
... and first_name = 'Drew'
... and email = 'djones@gmail.com';
cqlsh:user_management> select * from users
... where last_name = 'Jones';
```

last_name	first_name	email	address
Jones	Chris	cjones@gmail.com	789 Cherry St.
Jones	Drew	djones@gmail.com	987 Elm St.

(2 rows)

ตอบ ทำงานโดยไม่มีข้อผิดพลาด เพราะระบุ Partition Key และ Clustering Column

ครบถ้วนและถูกต้องตามลำดับ



#### 14. คำสั่ง CQL DELETE ให้เราสามารถลบค่า ทั้งค่าของคอลัมน์และทั้งแถว

ลองลบค่า address สำหรับ Chris Jones:

```
cqlsh:user_management> delete address from users
... where last_name = 'Jones'
... and first_name = 'Chris'
... and email = 'cjones@gmail.com';
cqlsh:user_management> select * from users ;
```

last_name	first_name	email	address
Jones	Chris	cjones@gmail.com	null
Jones	Drew	djones@gmail.com	987 Elm St.
Smith	Alex	asmith@gmail.com	987 Main St.
Smith	Bailey	bsmith@gmail.com	234 Elm st.

(4 rows)

ลองลบทั้งแถวโดยไม่ต้องระบุค่าใดๆ สำหรับ Drew Jones:

```
cqlsh:user_management> delete from users
... where last_name = 'Jones'
... and first_name = 'Drew'
... and email = 'djones@gmail.com'
... ;
cqlsh:user_management> select * from users ;
```

last_name	first_name	email	address
Jones	Chris	cjones@gmail.com	null
Smith	Alex	asmith@gmail.com	987 Main St.
Smith	Bailey	bsmith@gmail.com	234 Elm st.

(3 rows)

เกิดอะไรขึ้นเมื่อเราพยายามลบค่าคอลัมน์คลัสเตอร์?

```
DELETE email FROM users WHERE last_name = 'Smith' AND first_name = 'Alex'
AND email = 'asmith@gmail.com';
```

```
cqlsh:user_management> delete email from users where last_name = 'Smith'
and first_name = 'Alex' and email = 'asmith@gmail.com';
InvalidRequest: Error from server: code=2200 [Invalid query] message="In
valid identifier email for deletion (should not be a PRIMARY KEY part)"
```

ตอบ เพราะ email เป็น Primary Key

15. ลบพาร์ติชัน Smith ทั้งหมดได้ด้วยคำสั่งเดียวหรือไม่

```
cqlsh:user_management> delete from users
... where last_name = 'Smith';
```

ตอบ ลบได้

16. แสดงผลลัพธ์สุดท้ายของตาราง users

```
SELECT * FROM users;
```

```
cqlsh:user_management> select * from users;

last_name | first_name | email                | address
-----+-----+-----+-----
      Jones |      Chris | cjones@gmail.com    | null
(1 rows)
```