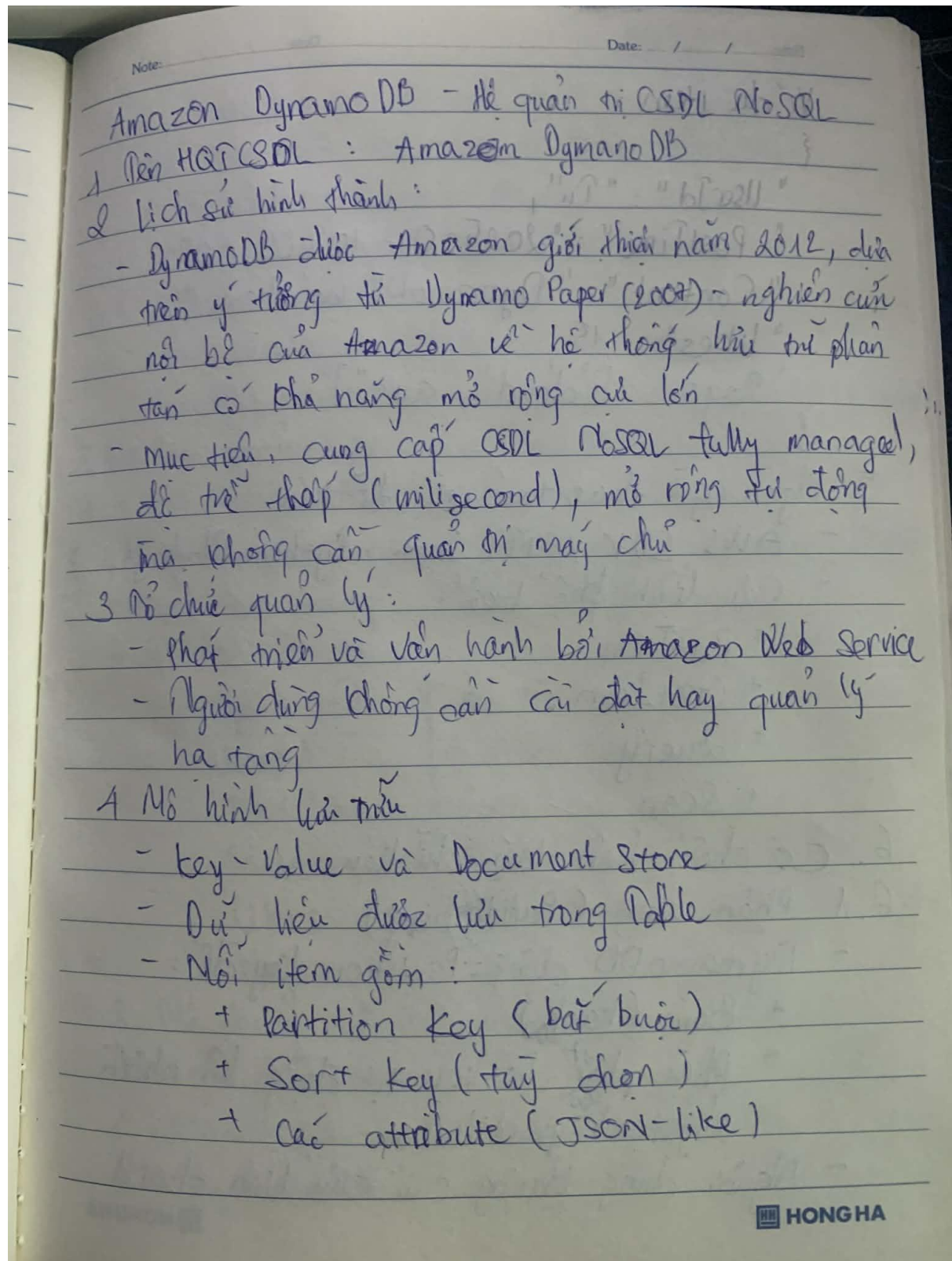


Tên: Nguyễn Quốc Tú

MSSV: 1250080219



```

Ví dụ item:
{
  "UserId": "Pú",
  "PostTime": "2025-01-10T09:00",
  "Content": "DynamoDB rất nhanh",
  "likes": 15,
  "Tags": ["cloud", "aws"]
}

```

5. Ngôn ngữ thao tác với dữ liệu

- AWS SDK / API (Java, Python, Node.js, ...)
- Câu lệnh phổ biến:
 - + PutItem
 - + GetItem
 - + Query
 - + Scan

6. Cơ chế phân tán dữ liệu

6.1 Phân vùng (Partitioning)

- DynamoDB dùng Partition key để:
 - + Băm (hash)
 - + Phân phối dữ liệu tự động lên nhiều physical partitions
- Người dùng không cần cấu hình shard

6.2 Sao chép và độ bền

- Mỗi item được replicate tự động trên nhiều Availability Zones (AZs)

- không có single point of failure

6.3 Nhãn quan

- Eventually consistent reads (mặc định)

- Strongly consistent reads (tùy chọn, trong cùng region)

7 Mô hình phòng lưu trữ và truy vấn phân tán

7.1 Bản toán mô phỏng trên giải lập

- Ứng dụng mạng xã hội lưu bài viết của người dùng
- Thiết kế bảng
- Table: Posts

Thuộc tính	Vai trò
User Id	Partition key
Post Time	Sort key
Content	Nội dung
Likes	Số lượt thích
Tags	Nhãn

7.2 Mô phỏng lưu trữ phân tán

Chỉ ghi dữ liệu

1. Client gửi PutItem lên DynamoDB
2. DynamoDB:

- Hash UserId

- Xác định partition đó

- Lưu item vào trong partition đó

3. Item được replicate sang nhiều AZ khác nhau

Vi dụ:

UserId = Tu \rightarrow Hash \rightarrow Partition A

UserId = Vu \rightarrow Hash \rightarrow Partition C

- Người dùng không thấy shard, tất cả diễn ra tự động

7.3 Mô phỏng truy vấn phân tán (Lucy)

- Truy vấn: lấy tất cả bài viết của UserId = "Tu"

response = dynamodb.query (

TableName = 'Posts',

keyConditionExpression = 'UserId = :uid',

ExpressionAttributeValues = {

:uid: { 'S': 'Tu' }

}

)

- Cách DynamoDB xử lý:

T Do truy vấn có Partition key, DynamoDB:

- Chỉ truy cập 1 partition duy nhất

- Không fan-out toàn hệ thống

+ \rightarrow Phản nhàn ($O(1)$)

7.4 Mô phỏng truy vấn phân tán toàn bộ (Scan)

- truy vấn: Tìm các bài viết có tag = "cloud"

response = dynamodb.scan (

Table Name = 'Posts',

FilterExpression = 'contains(tags, :t)',

ExpressionAttributeValues = {

:t: {'s': 'cloud'}

}

)

- DynamoDB xử lý

1. Gửi request đến tất cả dữ liệu

1. Gửi request đến tất cả partitions

2. Mỗi partition lọc dữ liệu que bộ

3. Gộp kết quả trả về client

- Scan = truy vấn phân tán toàn hệ thống

\rightarrow chậm và tốn tài nguyên

~~7.5~~

Sơ đồ kiến trúc Phân tán

