Thứ .....ngày.....tháng.....

Trần Đình Min
1250080257
CNTT5

Bài Tập Lý Thuyết Tuần 3

Hệ quản trị cơ sở dữ liệu NoSQL được chọn: Neo4j (Graph Database)

1. Tổng quan về Neo4j

1.1 Neo4j là gì?

Neo4j là hệ quản trị cơ sở dữ liệu NoSQL dạng đồ thị (Graph Database), lưu trữ dưới dạng:

- Node (đỉnh)
- Relationship (cạnh)
- Property (thuộc tính)

~~Neo~~ Neo4j đặc biệt phù hợp với các bài toán có mối quan hệ phức tạp

1.2 Lịch sử hình thành

- Ra đời năm 2007
- Phát triển bởi Neo4j Inc
- Được phát triển trong mạng xã hội, gợi ý, phát hiện gian lận.

2. Đặc trưng của Neo4j

2.1 Mô hình dữ liệu đồ thị

Ví dụ:

(SinhVien) - [DangKy] -> (MonHoc)

Mỗi node và relationship đều có thuộc tính

2.2 Ưu điểm

HONG HA

Scanned with CamScanner

Thứ ..... ngày ..... tháng .....

- Truy vấn quan hệ cực ~~mạnh~~ nhanh
- Trực quan, dễ mô hình hoá
- Không cần JOIN phức tạp

2.3 Nhược điểm

- Không phù hợp dữ liệu dạng bảng lớn
- Cần tư duy đồ thị

3. Ngôn ngữ truy vấn Neo4j – Cypher

Cypher là ngôn ngữ truy vấn mô tả quan hệ

Ví dụ:

```
MATCH (s: SinhVien) - [DANGKY] -> (m: MonHoc)
RETURN s, m;
```

4. Cơ chế phân tán trong Neo4j

Neo4j hỗ trợ Neo4j Fabric:

- Dữ liệu chia thành nhiều graph con
- Mỗi graph nằm trên 1 ~~clu~~ cluster / database khác nhau.
- Truy vấn có thể cross-graph

5. Mô phỏng lưu trữ và truy vấn phân tán

5.1 Bài toán mô phỏng

Hệ thống phân tích luồng di chuyển sinh viên trong trường đại học.

Mục tiêu:

- Theo dõi sinh viên
- Phân tích di chuyển giữa các khu
- Phục vụ tối ưu cơ sở vật chất

HONGHA

Scanned with CamScanner

Thứ ..... ngày ..... tháng .....

5.2 Thiết kế graph phân tán

Graph 1 - Khu học tập

- Node: SinhVien, PhongHoc
- Relationship: ~~S_HOC_TAI~~ HOC_TAI

Graph 2 - Khu dịch vụ

- Node: SinhVien, CanTin, ThuVien
- Relationship: SU_DUNG

Graph 3 - Khu ký túc xá

- Node: SinhVien, KyTucXa
- Relationship: O_TAI

=> Mỗi graph được lưu trên database khác nhau

6. Mô phỏng truy vấn phân tán Neo4j

- Truy vấn 1: Sinh viên học ở đâu?

```
MATCH (s:SinhVien {id:'SV01'})-[:HOC_TAI]->
(p:PhongHoc)
RETURN p.ten;
```

=> Truy vấn chỉ chạy trong 1 graph

- Truy vấn 2: Sinh viên sử dụng dịch vụ gì?

```
USE graph_dich_vu
MATCH (s:SinhVien {id:'SV01'})-[SU_DUNG]->(d)
-> RETURN d;
```

=> Truy vấn chạy trên graph khác

- Truy vấn 3: Truy vấn phân tán liên graph (Fabric)

HONG HA

Scanned with CamScanner

Thứ ..... ngày ..... tháng .....

```
USE (s: Sinh_vien {id: 'SV01'})
CALL {
    USE graph_hoc_tap
    MATCH (s)-[:HOC_TAI]->(p)
    RETURN p.ten AS dia_diem
}
CALL {
    USE graph_dich_vu
    MATCH (s)-[:SU_DUNG]->d
    ~~MATCH (s)~~
    RETURN d.ten AS dia_~~di~~diem
}
    RETURN dia_diem;
```

7. Đánh giá mô hình

Ưu điểm

- Truy vấn quan hệ đa chiều nhanh
- Mô hình thực tế rõ ràng
- Phù hợp hệ thống thông minh

Hạn chế

- Cài đặt ~~phức~~ phức tạp
- Cần thiết kế graph hợp lý

8. Kết luận

Neo4j là hệ quản trị NoSQL mạnh cho:

- Phân tích mối quan hệ
- Truy vấn đồ thị phân tán
- Hệ thống gợi ý, theo dõi hành vi

HONG HA

Scanned with CamScanner