

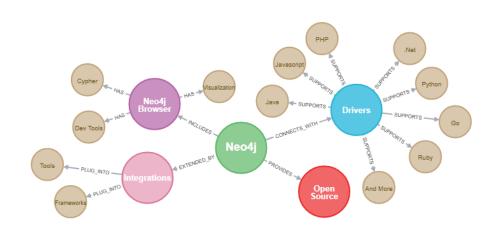


Cơ sở dữ liệu đồ thị

- Cơ sở dữ liệu đồ thị là cấu trúc được thiết kế để lưu trữ và xử lý dữ liệu theo mô hình đồ thi.
- Các hệ quản trị cơ sở dữ liệu đồ thị có cơ chế xử lý dữ liệu hiệu quả với nhiều mối liên kết phức tạp trên đồ thị mà các mô hình dữ liệu khác không thể đáp ứng.

3 Bộ môn HTTT

Giới thiệu Neo4J



4 Bộ môn HTTT

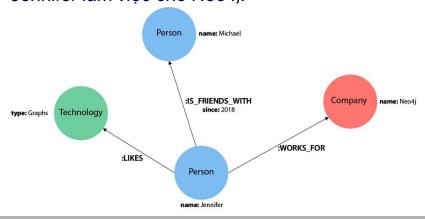
Giới thiệu Neo4J

- Neo4J là một hệ quản trị CSDL đồ thị được phát triển bởi công ty Neo4J
- Sử dụng ngôn ngữ Cypher để tạo đồ thị, cập nhật dữ liệu và truy vấn.
- Cung cấp các driver để kết nối với các ứng dụng lập trình trên nhiều ngôn ngữ khác nhau: Java, .Net, Python,...
- ❖Neo4J có 2 phiên bản:
 - Neo4J Community (mã nguồn mở)
 - Neo4J Enterprise (thương mại)

5 Bộ môn HTTT

Cơ sở dữ liệu đồ thị trong Neo4J

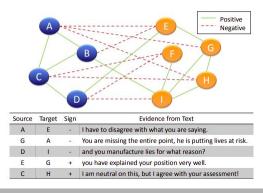
❖ Ví dụ cơ sở dữ liệu đồ thị được mô tả sau: Jennifer thích Graphs. Jennifer là bạn của Michael. Jennifer làm viêc cho Neo4i.



6 Bộ môn HTTT

Ứng dụng của đồ thị

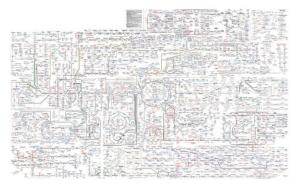
Khoa học xã hội: Để tìm những người Nổi tiếng X trong mạng xã hội, chúng ta có thể thống kê số lượng Follow của X và đưa ra những người có số lượng Follow cao nhất. Nói theo cách toán học thì ta tìm bậc của đỉnh hay số lượng cạnh nối với đỉnh đó (X) và tìm ra những đỉnh có bậc cao nhất.



7 Bộ mấm HTTT

Ứng dụng của đồ thị

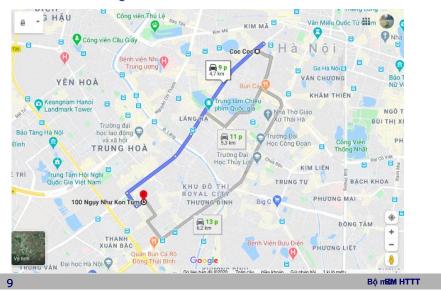
Nghiên cứu sinh học: Các thành phần sinh học(protein, phân tử, gen) và các tương tác của chúng cũng tạo nên một đồ thị sinh học. Dựa vào đó người ta có thể tìm hiểu được quá trình trao đổi chất trong cơ thể, sự tương tác giữa các bộ phận khác nhau trên cơ thể



8 Bộ m**ga HTTT**

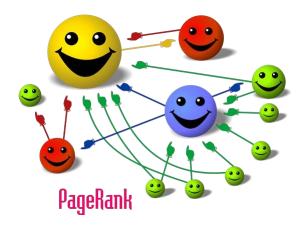
Ứng dụng của đồ thị

Bài toán tìm đường đi



Ứng dụng của đồ thị

WebSearch: Pagerank là thuật toán phân tích các liên kết được dùng trong Google Search để xếp hạng các trang web.



10 BM HTTT



Thiết kế cấu trúc đồ thị cho hệ thống mua bán hàng hoá

1 1 Bộ môn HTTT

Ngôn ngữ Cypher



- ❖Quy cách đặt tên
- ❖Sử dụng ghi chú
- ❖Biểu diễn nút, biến nút
- ❖Biểu diễn mối quan hệ, biến mối quan hệ
- ❖Biểu diễn một mẫu (Pattern)
- ❖ Tạo cơ sở dữ liệu mẫu
- ❖Truy vấn dữ liệu

12 Bộ môn HTTT



| Property Type | Structural types | Composite types |
|---|--|--|
| Number: Integer, FloatStringBoolean | Node: • Id • Label(s) • Map (of properties) | List, a heterogeneous, ordered collection of values, each of which has any property, structural or composite type. |
| Temporal: • Date • Time • LocalTime • DateTime • LocalDateTime • Duration | Relationship: Id Type Map (of properties) Id of the start node Id of the end node | Map, a heterogeneous, unordered collection of (Key, Value) pairs. Key is a String Value has any property, structural or composite type |
| Point | Path, an alternating sequence of nodes and relationships | |

13 Bộ môn HTTT



Quy cách đặt tên



- Tên đối tượng (nút, quan hệ, biến,...) bắt đầu bằng ký tự alphabet.
- Được chứa dấu gạch dưới "_"
- Không được bắt đầu bằng số.
- Không được chứa các ký tự đặt biệt như \$,
 @,...Ngoại trừ ký tự \$ đứng đầu là tham số.
- Có phân biệt chữ hoa, chữ thường

L4 Bộ môn HTTT





❖Ví du:

```
//data stored with this direction
CREATE (p:Person)-[:LIKES]->(t:Technology)
//query relationship backwards will not return results
MATCH (p:Person)<-[:LIKES]-(t:Technology)
//better to query with undirected relationship unless sure of direction
MATCH (p:Person)-[:LIKES]-(t:Technology)
```

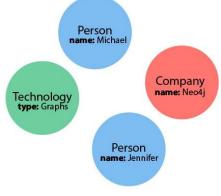
Bộ môn HTTT 15

Biểu diễn các nút của đồ thị

 Mỗi nút được biểu diễn đồ họa như hình dưới. Ký hiệu (node).

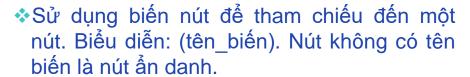
Trong hình có 4 nút có tên là: Jennifer, Michael,





Bộ môn HTTT

Biến nút (Node Variables)



❖Ví dụ:

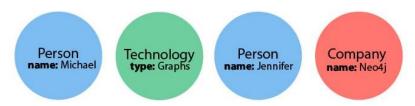
- (): nút ẩn danh
- (p): biến có tên là p
- (person): biến có tên là person
- (t): biến có tên là t
- (thing): biến có tên là thing

17 Bộ môn HTTT

Nhãn nút (Node Labels)



- Trong hình, các nhãn nút là: Person, Technology và Company
- Tương quan với SQL, mỗi nút tương tự như một dòng, mỗi nhãn nút tương tự như một bảng.
- Nếu không chỉ định nhãn thì truy vấn sẽ duyệt qua hết tất cả các nút trong đồ thị



18 Bộ môn HTTT

Biểu diễn mối quan hệ

- Mối quan hệ biểu diễn sự quan hệ giữa 2 nút trên đồ thi.
- Tên mối quan hệ thường là động từ.
- ❖Có 2 loại quan hệ:
 - Có hướng:
 - -[:RELATIONSHIP]-> hoặc <-[: RELATIONSHIP]-
 - Vô hướng:
 - -[: RELATIONSHIP]-
 - Trong đó: RELATIONSHIP là tên mối quan hệ.

19 Bộ môn HTTT

Ví dụ quan hệ giữa 2 nút

(p:Person)-[:IS_FRIEND_WITH]->(q:Person) (p:Person)<-[: IS_FRIEND_WITH]-(q:Person)

(p:Person)-[: IS_FRIEND_WITH]-(q:Person)

20 Bộ môn HTTT

Biến mối quan hệ

- Được dùng để tham chiếu đến một mối quan hệ.
 - Ký hiệu -[variable]->, <-[variable]hoặc –[variable]-
 - Có thể chỉ định kiểu quan hệ như:
 –[variable:TYPE]->
- ❖Ví dụ:
 - -[r]-> hoặc -[r:LIKES]->

21 Bộ môn HTTT

Thuộc tính nút hoặc mối quan hệ

- Thuộc tính nút hoặc thuộc tính quan hệ được biểu diễn: {name:value}
- ❖Ví dụ:
 - Thuôc tính nút:

(p:Person {name: 'Jennifer'})

Thuộc tính mối quan hệ:

-[rel:IS_FRIENDS_WITH {since: 2018}]->

22 Bộ môn HTTT

Patterns trong Cypher



- Một mẫu phức tạp có thể được tách thành các mẫu nhỏ đơn giản hơn.
- ❖Ví dụ:

(p:Person {name: "Jennifer"})-[rel:LIKES]->(g:Technology {type: "Graphs"})

23 Bộ môn HTTT

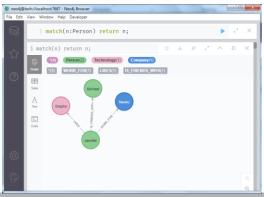
Các lệnh thao tác cơ bản

- ❖Thao tác trên cửa sổ lệnh
- Thao tác trên nút
 - Thêm/xóa/sửa nút
- Thao tác trên mối liên kết
 - Thêm/xóa/sửa liên kết

D∕I Bộ mộn HTTT

Thao tác trên cửa sổ lệnh

- Start cơ sở dữ liệu cần thao tác trên Neo4J Desktop, sau đó mở công cụ Neo4J Browser từ Neo4J Desktop.
- Gố lệnh thao tác ở khung nhập phía trên. Xuống dòng lệnh bằng phím Shift+Enter. Thực thi bằng phím Ctrl+Enter hoặc click vào nút thực thi màu xanh bên phải.



25 Bộ môn HTTT

Thao tác trên nút



- \$ Create (a) //Thêm một nút mới ko thuộc tính, ko nhãn
- \$ Create (b),(c) //Thêm 2 nút
- \$ create (p:Person) //Thêm nút với nhãn Person
- \$ Create(:Person) //Thêm nút với nhãn Person không chỉ định biến
- \$ create (p:Person:VietNam) return p //Thêm nút với 2 nhãn là Person và VietNam, hiển thị nút vừa tạo bằng lệnh return
- \$ create (p:Person {name: 'Andy', title: 'Developer'}) return p
 //Thêm nút với nhãn Person và 2 thuộc tính name và title, hiển thị nút p vừa tạo

26 Bộ môn HTTT





- \$ MATCH (p:Person{name:'Andy'}) DELETE p;//xóa nút thuộc nhãn Person có tên là Andy
- \$ MATCH (p:Person {name: 'Andy'}) DETACH DELETE n //xóa nút thuộc nhãn Person có tên là Andy và bất kỳ liên kết nối với nút.
- \$ MATCH (n) DETACH DELETE n //Xóa tất cả các nút và mối liên kết
- \$ MATCH (p) WHERE NOT (EXISTS(p.name)) DETACH DELETE p; //Xóa các nút không có tên
- \$ MATCH(p) WHERE ID(p)=4 DELETE p //Xóa nút có thuộc tính ID = 4

27 Bộ môn HTTT

Thao tác trên nút (tt)

- Thêm/sửa thuộc tính nút:
 - \$ MATCH (n {name: 'Andy'}) //Thêm thuộc tính surname = 'Taylor'

 SET n.surname = 'Taylor' //cho nút có tên là Andy

 RETURN n.name, n.surname;
 - \$ MATCH (p:Person {name: 'Jennifer'})

 SET p.birthdate = date('1980-01-01'), p.gender='female'

 RETURN p
- Xóa thuộc tính nút
 - \$ MATCH (n {name: 'Andy'})

 REMOVE n.surname, n.age //Xóa 2 thuộc tính là surname và age
 RETURN n

28 Bộ môn HTTT





- Thêm ràng buộc duy nhất:
- \$ CREATE CONSTRAINT ON (p:Person)
 ASSERT p.name is UNIQUE
- Xóa ràng buộc:
- \$ DROP CONSTRAINT ON (p:Person)
 ASSERT p.name IS UNIQUE

29 Bộ môn HTTT



- Thêm liên kết:
- \$ MATCH (a:Person), (b:Person)
 WHERE a.name = 'Andy' AND b.name = 'Michael'
 CREATE (a)-[r:KNOWS{since:2018}]->(b)
 RETURN a,b
- Xóa liên kết
 - \$ MATCH (n {name: 'Andy'})-[r:KNOWS]->()
 DELETE r

30 Bộ môn HTTT

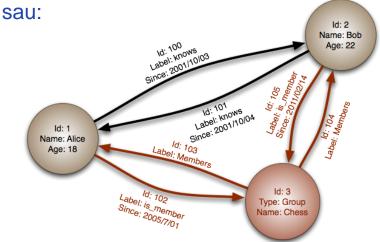
Thao tác trên mối liên kết (tt)

- Thêm/sửa thuộc tính của liên kết:
- \$ MATCH (n{name:'Andy'})-[r:KNOWS]->(m{name:'Michael'})
 set r.since=2018
 return n,m
- Xóa thuộc tính
- \$ MATCH (n{name:'Andy'})-[r:KNOWS]->(m{name:'Michael'})
 REMOVE r.since
 return n,m

31 Bộ môn HTTT

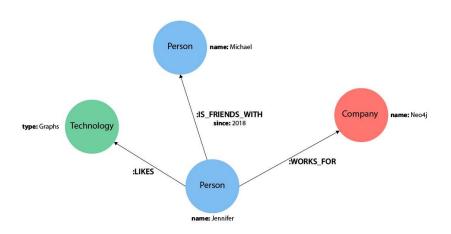
Ví dụ tạo cơ sở dữ liệu đồ thị

❖Dùng lệnh tạo cơ sở dữ liệu đồ thị như hình



32 Bộ môn HTTT

Ví dụ tạo cơ sở dữ liệu đồ thị



33 Bộ môn HTTT

Tạo cơ sở dữ liệu đồ thị

- ❖Tạo cơ sở dữ liệu đồ thị như ví dụ trên
 - Tạo node tên là Jennifer có nhãn Person
 - \$ Create(p:Person{name: 'Jennifer'})
 return p;
 - Thiết lập id là khóa:
 - \$ create constraint on (p:Person) assert p.id is unique
 - Tạo node tên là Michael có nhãn Person
 - \$ Create(p:Person{name: 'Michael'})
 return p;

34 Bộ môn HTTT

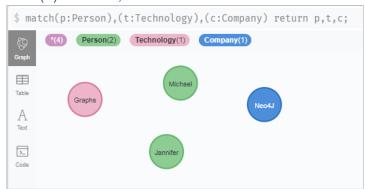
Tạo cơ sở dữ liệu đồ thị (tt)

- Tạo cơ sở dữ liệu đồ thị như ví dụ trên
 - Tạo node type là Graphs, có nhãn Technology
 - \$ Create(p:Technology{type: 'Graphs'}) return p;
 - Thiết lập id là khóa cho nhãn Technology:
 - \$ create constraint on (p:Technology) assert p.id is unique
 - Tạo node tên là Neo4J có nhãn Company
 - \$ Create(p:Company{name: 'Neo4J'}) return p;
 - Thiết lập id là khóa cho nhãn Company
 - \$ create constraint on (p:Company) assert p.id is unique

35 Bộ môn HTTT



- ❖Hiển thị các nút đã tạo
 - \$ match(p:Person),(t:Technology),(c:Company) return p,t,c;
- Hoặc
 - \$ match(n) return n;



36 Bộ môn HTTT



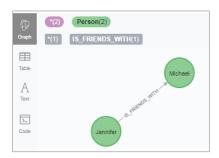
Tạo liên kết giữa 2 nút Jennifer và Michael có thuộc tính since:2018

MATCH (a:Person),(b:Person)

WHERE a.name='Jennifer' and b.name='Michael'

CREATE (a)-[r:IS_FRIENDS_WITH{since:2018}]->(b)

RETURN a,b;



37 Bộ môn HTTT

Tạo các liên kết (tt)

 Tạo liên kết giữa 2 nút Jennifer và Graphs không có thuộc tính.

```
match(a:Person),(b:Technology)
where a.name='Jennifer' and b.type='Graphs'
create (a)-[r:LIKES]->(b)
return a,b;
```

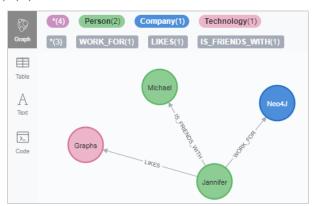
 Tạo liên kết giữa 2 nút Jennifer và Neo4J không thuộc tính

```
match(a:Person),(b:Company)
where a.name='Jennifer' and b.name='Neo4J'
create (a)-[r:WORK_FOR]->(b)
return a,b;
```

38 Bộ môn HTTT

Cơ sở dữ liệu kết quả

- Xem các nút và liên kết
 - \$ match(a:Person),(b:Company),(c:Technology) return a,b,c;



39 Bộ môn HTTT

Truy vấn dữ liệu

//Hiển thi tất cả các nút

\$ MATCH(p) RETURN p;

//Hiển thị các nút thuộc nhãn Person

\$ MATCH(p:Person) RETURN p;

//Hiển thị nút thuộc nhãn Person có tên là Jennifer

\$ MATCH(j:Person{name:'Jennifer'}) RETURN j

//Hiển thị tên công ty mà người có tên Jennifer làm việc.

\$ MATCH(:Person{name:'Jennifer'})-[:WORK_FOR]-> (c:Company)

RETURN c;

≿Có thể dùng RETURN c.name để chỉ trả về tên công ty

40 Bộ môn HTTT

Truy vấn label nút và type relationship

- //Hiển thị label của nút
- \$MATCH (p:Person{name:"Jennifer"}) return labels(p);
- //Hiển thị tất cả các label có trong CSDL
- \$CALL db.labels();
- //Hiển thị type của mối liên kết
- \$MATCH (p)-[r]->(q) return type(r);

41 Bộ môn HTTT



Truy vấn với mệnh đề WHERE



//Hiển thi nút thuộc nhãn Person có tên là Jennifer

- \$ MATCH(j:Person{name:'Jennifer'}) RETURN j
- Tương đương khi sử dụng WHERE
- \$ MATCH(p:Person)
 - WHERE p.name='Jennifer'
 - RETURN p;

//Hiến thị tên công ty mà người có tên Jennifer làm việc.

\$ MATCH(p:Person),(c:Company)
WHERE (p)-[:WORK_FOR]->(c)
RETURN c;

42 Bộ môn HTTT

Truy vấn với mệnh đề WHERE (tt)

//Hiển thị những người không có tên là Jennifer

\$ MATCH (j:Person)
WHERE NOT j.name = 'Jennifer'
RETURN j;

//Hiển thị những người có tuổi từ 20 đến 30

\$ MATCH(p:Person)
WHERE 20<=p.age<=30
RETURN p;

//Hiển thị tên những người có thuộc tính birthdate

\$ MATCH (p:Person)
WHERE exists(p.birthdate)
RETURN p.name

Bộ môn HTTT



❖Xử lý chuỗi:

//Điều kiện thuộc tính bắt đầu bằng ký tự 'M'

\$ MATCH (p:Person)
WHERE p.name STARTS WITH 'M'
RETURN p.name;

//Thuộc tính có chứa ký tự 'a'

\$ MATCH (p:Person)
WHERE p.name CONTAINS 'a'
RETURN p.name;

//Thuộc tính kết thúc với ký tự 'n'

\$ MATCH (p:Person)
WHERE p.name ENDS WITH 'n'
RETURN p.name;

14 Bộ môn HTTT

Truy vấn với mệnh đề WHERE (tt)

❖Xử lý chuỗi:

//Thuộc tính có 2 ký tự đầu là Je

\$ MATCH (p:Person)
WHERE p.name =~ 'Je.*'
RETURN p.name;

//Sử dụng từ khóa IN

\$ MATCH (p:Person)
WHERE p.yearsExp IN [1, 5, 6]
RETURN p.name, p.yearsExp;

45 Bộ môn HTTT



Exists và Not Exists với pattern



//Tìm những người là bạn của những người làm việc cho công ty có tên là Neo4J

\$ MATCH (p)-[:IS_FRIENDS_WITH]-(q)
WHERE EXISTS(
 (p)-[:WORK_FOR]->(:Company{name:'Neo4J'}))
return q;

//Tìm những người bạn của Jennifer mà không làm việc cho công ty nào.

\$ MATCH (p:Person)-[r:IS_FRIENDS_WITH]->(friend:Person)
WHERE p.name = 'Jennifer'
AND NOT exists((friend)-[:WORKS_FOR]->(:Company))
RETURN friend.name

46 Bộ môn HTTT



Một số truy vấn phức tạp



//Tìm những người thích Graphs ngoài Jennifer

\$ MATCH (j:Person {name: 'Jennifer'})-[:LIKES] (graph:Technology {type: 'Graphs'})-[:LIKES]-(p:Person)
 RETURN p.name;

// Tìm những người thích Graphs ngoài Jennifer và Jennifer cũng là bạn với những người đó.

47

Bộ môn HTTT



Aggregation trong Cypher



Các hàm count, sum,

//Đếm số lượng người có email

- \$ MATCH(p:Person)
- \$ RETURN count(p.email);

Nếu dùng count(*) sẽ đếm hết các đối tượng kể cả có thuộc tính email bằng null

//Cho biết độ tuổi trung bình

\$ MATCH(p:Person)

RETURN avg(p.age);

Sử dụng tương tự cho các hàm sum, min, max

Bộ môn HTTT



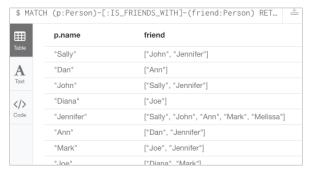


//Hiển thị tên của mỗi người và danh sách bạn bè của người đó

\$ MATCH (p:Person)-[:IS_FRIENDS_WITH]>(friend:Person)

RETURN p.name, collect(friend.name) AS friend

Hàm collect sẽ đưa các kết quả vào danh sách



49 Bộ môn HTTT



Kết hợp giá trị với collect(tt)



//Hiển thị tên của mỗi người và số lượng bạn bè của người đó.

\$ MATCH (p:Person)-[:IS_FRIENDS_WITH]->(friend:Person)
RETURN p.name, size(collect(friend.name)) AS numberOfFriends



//Hiển thị tên của mỗi người, danh sách bạn bè tương ứng sao cho mỗi người trong danh sách này có số lượng bạn bè >1

MATCH (p:Person)-[:IS_FRIENDS_WITH]->(friend:Person)

WHERE count{(friend)-[:IS_FRIENDS_WITH]-(:Person)} > 1

RETURN p.name, collect(friend.name) AS friends, size((friend)-[:IS_FRIENDS_WITH]-(:Person)) AS numberOfFoFs



50 Bộ môn HTTT





\$ MATCH (a:Person)-[r:LIKES]-(t:Technology)
WITH a.name AS name, collect(t.type) AS technologies
RETURN name, technologies

//Sử dụng WITH viết lại truy vấn:Hiển thị tên của mỗi người, danh sách bạn bè tương ứng sao cho mỗi người trong danh sách này có số lượng bạn bè >1

\$ MATCH (p:Person)-[:IS_FRIENDS_WITH]->(friend:Person) WITH p, collect(friend.name) AS friendsList, size((friend)-[:IS_FRIENDS_WITH]-(:Person)) AS numberOfFoFs WHERE numberOfFoFs > 1 RETURN p.name, friendsList, numberOfFoFs

51 Bộ môn HTTT

Sử dụng WITH (tt)

 WITH còn được sử dụng để thiết lập giá trị tham số trước câu truy vấn.

\$ WITH 2 AS experienceMin, 6 AS experienceMax MATCH (p:Person)

WHERE experienceMin <= p.yrsExperience <= experienceMax

RETURN p

52 Bộ môn HTTT



 UNWIND được dùng để tách một mảng thành các giá trị riêng lẻ để xử lý

//Với danh sách kỹ năng cho trước, cần tìm những người có kỹ năng tương ứng trong danh sách.

\$ WITH ['Graphs','Query Languages'] AS techRequirements
UNWIND techRequirements AS technology
MATCH (p:Person)-[r:LIKES]-(t:Technology {type: technology})
RETURN t.type, collect(p.name) AS potentialCandidates

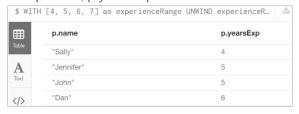


Bộ môn HTTT

Sử dụng UNWIND (tt)

//Với dan sách số năm kinh nghiệm, tìm các ứng viên có số năm kinh nghiệm thuộc danh sách.

\$ WITH [4, 5, 6, 7] AS experienceRange
UNWIND experienceRange AS number
MATCH (p:Person) WHERE p.yearsExp = number
RETURN p.name, p.yearsExp



54 Bộ môn HTTT



 ORDER BY được dùng để sắp xếp kết quả truy vấn. Mặc định sắp tặng dần, Sử dụng DESC để sắp giảm dần.

//Sử dụng truy vấn trước: Với dan sách số năm kinh nghiệm, tìm các ứng viên có số năm kinh nghiệm thuộc danh sách.

\$ WITH [4, 5, 6, 7] AS experienceRange UNWIND experienceRange AS number MATCH (p:Person) WHERE p.yearsExp = number

RETURN p.name, p.yearsExp

ORDER BY p.yearsExp DESC

55 Bộ môn HTTT



Distinct được dùng để loại bỏ kết quả trùng

//Tìm những người có tài khoản email hoặc thích graphs hoặc thích query languages

\$ MATCH (user:Person)

WHERE user.email IS NOT null

WITH user

MATCH (user)-[:LIKES]-(t:Technology)

WHERE t.type IN ['Graphs', 'Query Languages']

RETURN DISTINCT user.name

56 Bộ môn HTTT

Giới hạn kết quả trả về với LIMIT

- Ví dụ: Tìm 3 người đầu tiên có nhiều bạn nhất
 - \$ MATCH (p:Person)-[r:IS_FRIENDS_WITH]-(other:Person)
 RETURN p.name, count(other.name) AS numberOfFriends
 ORDER BY numberOfFriends DESC
 LIMIT 3

57 Bộ môn HTTT

