

PHÂN TÍCH DỮ LIỆU LỚN

HỌ TÊN: TRẦN TẤN PHÁT
MSSV: 2274802010644

ĐỀ BÀI: NEO4J

BÀI LÀM

1. Tạo dữ liệu mẫu (20 điểm)

Dùng các bảng dưới đây để tạo dữ liệu trong Neo4j. Mỗi người và phim phải có ít nhất 1 quan hệ. Thêm ít nhất 3 quan hệ "WORKED_WITH" dựa trên hình "Network Diagram" với thuộc tính `projects` và `lastYear`.

Tên	Năm sinh	Quốc gia	Số năm làm việc
Tom Hanks	1956	USA	40
Meg Ryan	1961	USA	35
Steven Spielberg	1946	USA	50
Leonardo DiCaprio	1974	USA	30
Kate Winslet	1975	UK	28
James Cameron	1954	Canada	45
Robert Zemeckis	1952	USA	42
Natalie Portman	1981	Israel	25
Christopher Nolan	1970	UK	35
Emma Watson	1990	UK	15

```

1
2 MATCH (n) DETACH DELETE n;
3
4 CREATE (:Person {name: "Tom Hanks", birthYear: 1956, country:
   "USA", yearsWorked: 40}),
5 (:Person {name: "Meg Ryan", birthYear: 1961, country: "USA",
   yearsWorked: 35}),
6 (:Person {name: "Steven Spielberg", birthYear: 1946, country:
   "USA", yearsWorked: 50}),
7 (:Person {name: "Leonardo DiCaprio", birthYear: 1974, country:
   "USA", yearsWorked: 30}),
8 (:Person {name: "Kate Winslet", birthYear: 1975, country: "UK",

```

```

8 (:Person {name: "Kate Winslet", birthYear: 1975, country: "UK",
   yearsWorked: 28}),
9 (:Person {name: "James Cameron", birthYear: 1954, country:
   "Canada", yearsWorked: 45}),
10 (:Person {name: "Robert Zemeckis", birthYear: 1952, country: "USA",
   yearsWorked: 42}),
11 (:Person {name: "Natalie Portman", birthYear: 1981, country:
   "Israel", yearsWorked: 25}),
12 (:Person {name: "Christopher Nolan", birthYear: 1970, country:
   "UK", yearsWorked: 35}),
13 (:Person {name: "Emma Watson", birthYear: 1990, country: "UK",
   yearsWorked: 15});

```

```

14
15 CREATE (:Movie {title: "Forrest Gump", year: 1994, type: "Drama",
   budget: 50}),
16 (:Movie {title: "Sleepless in Seattle", year: 1993, type:
   "Romance", budget: 20}),
17 (:Movie {title: "Saving Private Ryan", year: 1998, type: "War",
   budget: 70}),
18 (:Movie {title: "Titanic", year: 1997, type: "Romance", budget:
   200}),
19 (:Movie {title: "The Terminal", year: 2004, type: "Drama", budget:
   60}),
20 (:Movie {title: "Catch Me If You Can", year: 2002, type: "Crime",
   budget: 50});

```

```

20 (:Movie {title: "Catch Me If You Can", year: 2002, type: "Crime",
    budget: 52}),
21 (:Movie {title: "Inception", year: 2010, type: "Sci-Fi", budget:
    160}),
22 (:Movie {title: "Harry Potter 1", year: 2001, type: "Fantasy",
    budget: 125}),
23 (:Movie {title: "Interstellar", year: 2014, type: "Sci-Fi", budget:
    165}),
24 (:Movie {title: "Black Swan", year: 2010, type: "Thriller", budget:
    13});
25
26 MATCH (p:Person {name: "Tom Hanks"}), (m:Movie {title: "Forrest
    Gump"})

```

Status: success

```

27 CREATE (p)-[:ACTED_IN {role: "Forrest", salary: 10}]->(m);
28
29 MATCH (p:Person {name: "Tom Hanks"}), (m:Movie {title: "Sleepless
    in Seattle"})
30 CREATE (p)-[:ACTED_IN {role: "Sam", salary: 5}]->(m);
31
32 MATCH (p:Person {name: "Meg Ryan"}), (m:Movie {title: "Sleepless in
    Seattle"})
33 CREATE (p)-[:ACTED_IN {role: "Annie", salary: 4}]->(m);
34
35 MATCH (p:Person {name: "Steven Spielberg"}), (m:Movie {title:
    "Saving Private Ryan"})

```

```

36 CREATE (p)-[:DIRECTED {yearsSpent: 3}]->(m);
37
38 MATCH (p:Person {name: "Tom Hanks"}), (m:Movie {title: "Saving
    Private Ryan"})
39 CREATE (p)-[:ACTED_IN {role: "Captain Miller", salary: 12}]->(m);
40
41 MATCH (p:Person {name: "Kate Winslet"}), (m:Movie {title:
    "Titanic"})
42 CREATE (p)-[:ACTED_IN {role: "Rose", salary: 20}]->(m);
43
44 MATCH (p:Person {name: "Leonardo DiCaprio"}), (m:Movie {title:
    "Inception"})

```

Status: success

```

45 CREATE (p)-[:ACTED_IN {role: "Jack", salary: 25}]->(m);
46
47 MATCH (p:Person {name: "James Cameron"}), (m:Movie {title:
  "Titanic"})
48 CREATE (p)-[:DIRECTED {yearsSpent: 5}]->(m);
49
50 MATCH (p:Person {name: "Christopher Nolan"}), (m:Movie {title:
  "Inception"})
51 CREATE (p)-[:DIRECTED {yearsSpent: 4}]->(m);
52
53 MATCH (p:Person {name: "Leonardo DiCaprio"}), (m:Movie {title:
  "Inception"})

```

Status: success

```

54 CREATE (p)-[:ACTED_IN {role: "Cobb", salary: 30}]->(m);
55
56 MATCH (p:Person {name: "Natalie Portman"}), (m:Movie {title:
  "Inception"})
57 CREATE (p)-[:ACTED_IN {role: "Mal", salary: 5}]->(m);
58
59 MATCH (p:Person {name: "Natalie Portman"}), (m:Movie {title: "Black
  Swan"})
60 CREATE (p)-[:ACTED_IN {role: "Nina", salary: 8}]->(m);
61
62 MATCH (p:Person {name: "Christopher Nolan"}), (m:Movie {title:
  "Interstellar"})

```

Status: success

```

"Interstellar"})
63 CREATE (p)-[:DIRECTED {yearsSpent: 5}]->(m);
64
65 MATCH (p:Person {name: "Robert Zemeckis"}), (m:Movie {title:
  "Forrest Gump"})
66 CREATE (p)-[:DIRECTED {yearsSpent: 3}]->(m);
67
68 MATCH (p:Person {name: "Emma Watson"}), (m:Movie {title: "Harry
  Potter 1"})
69 CREATE (p)-[:ACTED_IN {role: "Hermione", salary: 10}]->(m);
70
71 MATCH (p:Person {name: "Tom Hanks"}), (m:Movie {title: "The

```

Status: success

```

80 MATCH (p:Person {name: "Leonardo DiCaprio"}), (m:Movie {title:
    "Catch Me If You Can"})
81 CREATE (p)-[:ACTED_IN {role: "Frank", salary: 20}]->(m);
82
83 MATCH (p:Person {name: "Steven Spielberg"}), (m:Movie {title:
    "Catch Me If You Can"})
84 CREATE (p)-[:DIRECTED {yearsSpent: 3}]->(m); Status: success
85
86 MATCH (p:Person {name: "Leonardo DiCaprio"}), (m:Movie {title:
    "Interstellar"})
87 CREATE (p)-[:ACTED_IN {role: "Cooper", salary: 35}]->(m);
88

```

```

71 MATCH (p:Person {name: "Tom Hanks"}), (m:Movie {title: "The
    Terminal"})
72 CREATE (p)-[:ACTED_IN {role: "Viktor", salary: 15}]->(m);
73
74 MATCH (p:Person {name: "Steven Spielberg"}), (m:Movie {title: "The
    Terminal"})
75 CREATE (p)-[:DIRECTED {yearsSpent: 3}]->(m);
76
77 MATCH (p:Person {name: "Tom Hanks"}), (m:Movie {title: "Catch Me If
    You Can"})
78 CREATE (p)-[:ACTED_IN {role: "Carl", salary: 10}]->(m);
79

```

```

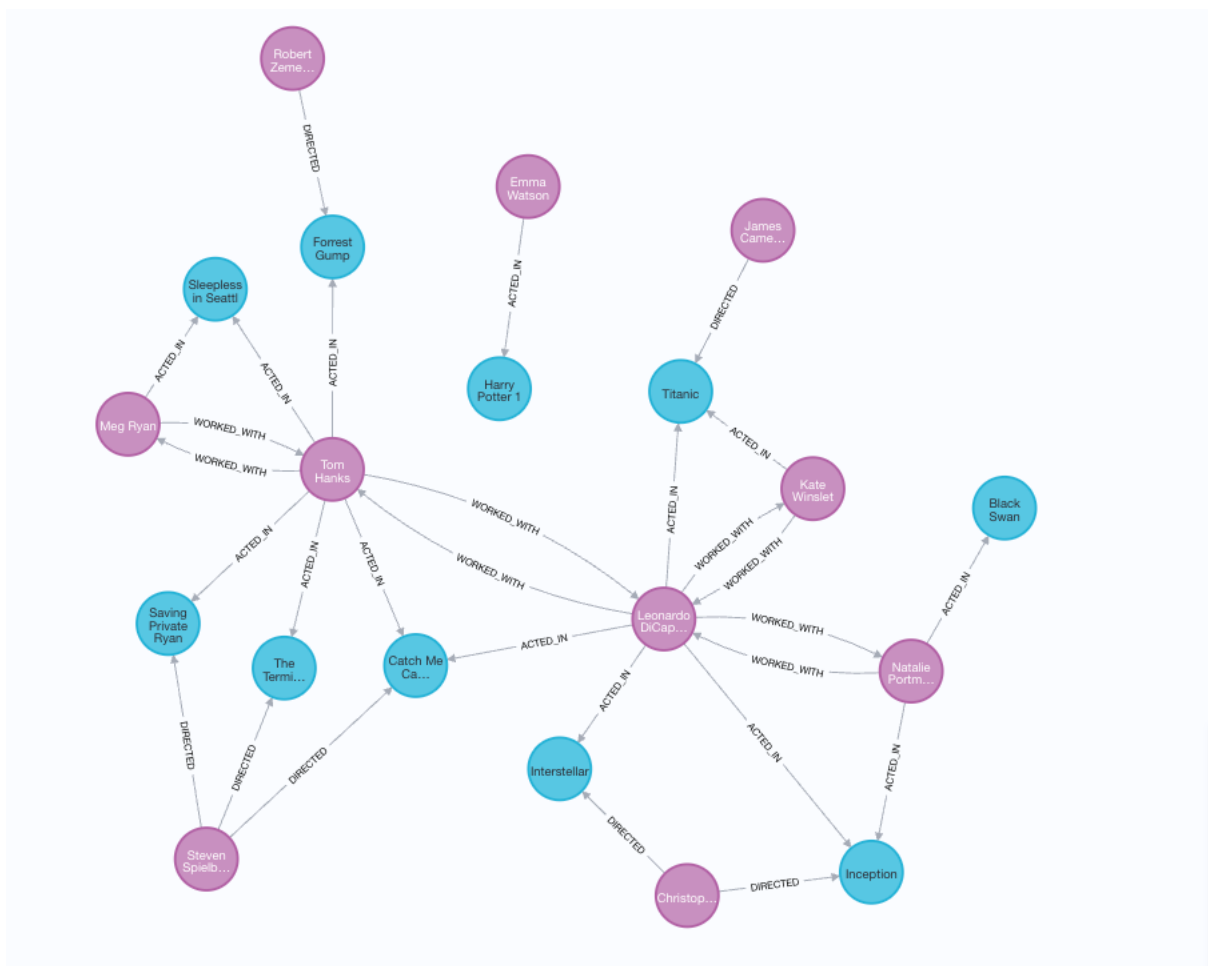
87 CREATE (p)-[:ACTED_IN {role: "Cooper", salary: 35}]->(m);
88
89 MATCH (p1:Person)-[:ACTED_IN]->(m:Movie)<-[:ACTED_IN]-(p2:Person)
90 WHERE p1 < p2
91 MERGE (p1)-[:WORKED_WITH {projects: 1, lastYear: 2024}]->(p2);
92

```

neo4j\$ MATCH (n) DETACH DELETE n	✓
neo4j\$ CREATE (:Person {name: "Tom Hanks", birthYear: 1956, country: "...	✓
neo4j\$ CREATE (:Movie {title: "Forrest Gump", year: 1994, type: "Drama...	✓
neo4j\$ MATCH (p:Person {name: "Tom Hanks"}), (m:Movie {title: "Forrest...	✓
neo4j\$ MATCH (p:Person {name: "Tom Hanks"}), (m:Movie {title: "Sleeple...	✓
neo4j\$ MATCH (p:Person {name: "Meg Ryan"}), (m:Movie {title: "Sleeples...	✓
neo4j\$ MATCH (p:Person {name: "Steven Spielberg"}), (m:Movie {title: "...	✓
neo4j\$ MATCH (p:Person {name: "Tom Hanks"}), (m:Movie {title: "Saving ...	✓
neo4j\$ MATCH (p:Person {name: "Kate Winslet"}), (m:Movie {title: "Tita...	✓
neo4j\$ MATCH (p:Person {name: "Leonardo DiCaprio"}), (m:Movie {title: ...	✓

neo4j\$ MATCH (p:Person {name: "Christopher Nolan"}), (m:Movie {title: ...	✓
neo4j\$ MATCH (p:Person {name: "Leonardo DiCaprio"}), (m:Movie {title: ...	✓
neo4j\$ MATCH (p:Person {name: "Natalie Portman"}), (m:Movie {title: "I...	✓
neo4j\$ MATCH (p:Person {name: "Natalie Portman"}), (m:Movie {title: "B...	✓
neo4j\$ MATCH (p:Person {name: "Christopher Nolan"}), (m:Movie {title: ...	✓
neo4j\$ MATCH (p:Person {name: "Robert Zemeckis"}), (m:Movie {title: "F...	✓
neo4j\$ MATCH (p:Person {name: "Emma Watson"}), (m:Movie {title: "Harry...	✓
neo4j\$ MATCH (p:Person {name: "Tom Hanks"}), (m:Movie {title: "The Ter...	✓
neo4j\$ MATCH (p:Person {name: "Steven Spielberg"}), (m:Movie {title: "...	✓
neo4j\$ MATCH (p:Person {name: "Tom Hanks"}), (m:Movie {title: "Catch M...	✓

neo4j\$ MATCH (p:Person {name: "Natalie Portman"}), (m:Movie {title: "I...	✓
neo4j\$ MATCH (p:Person {name: "Natalie Portman"}), (m:Movie {title: "B...	✓
neo4j\$ MATCH (p:Person {name: "Christopher Nolan"}), (m:Movie {title: ...	✓
neo4j\$ MATCH (p:Person {name: "Robert Zemeckis"}), (m:Movie {title: "F...	✓
neo4j\$ MATCH (p:Person {name: "Emma Watson"}), (m:Movie {title: "Harry...	✓
neo4j\$ MATCH (p:Person {name: "Tom Hanks"}), (m:Movie {title: "The Ter...	✓
neo4j\$ MATCH (p:Person {name: "Steven Spielberg"}), (m:Movie {title: "...	✓
neo4j\$ MATCH (p:Person {name: "Tom Hanks"}), (m:Movie {title: "Catch M...	✓
neo4j\$ MATCH (p:Person {name: "Leonardo DiCaprio"}), (m:Movie {title: ...	✓
neo4j\$ MATCH (p:Person {name: "Steven Spielberg"}), (m:Movie {title: "...	✓
neo4j\$ MATCH (p:Person {name: "Leonardo DiCaprio"}), (m:Movie {title: ...	✓
neo4j\$ MATCH (p1:Person)-[:ACTED_IN]→(m:Movie)←[:ACTED_IN]-(p2:Perso...	✓



2. Tìm phim và người (20 điểm)

2. Tìm phim và người (20 điểm)

Dùng bảng và hình "Network Diagram" để trả lời:

1. Tìm tất cả phim Tom Hanks tham gia (diễn hoặc hợp tác qua "WORKED_WITH").
Trả về `title`, `year`, `type`, và vai trò ("Actor" hoặc "Collaborator").

3

2. Tìm người có `yearsWorked` trên 40 đã đạo diễn ít nhất 2 phim. Trả về `name`, `country`, danh sách phim (`titles`).

```
1 MATCH (p:Person {name: "Tom Hanks"})-[:ACTED_IN]->(m:Movie)
2 RETURN m.title AS title, m.year AS year, m.type AS type, "Actor" AS
   role
3 UNION
4 MATCH (p:Person {name: "Tom Hanks"})-[:WORKED_WITH]-(coActor)-
   [:ACTED_IN]->(m:Movie)
5 RETURN m.title AS title, m.year AS year, m.type AS type,
   "Collaborator" AS role;
6
7
```

	title	year	type	role
1	"Forrest Gump"	1994	"Drama"	"Actor"
2	"Sleepless in Seattle"	1993	"Romance"	"Actor"
3	"Saving Private Ryan"	1998	"War"	"Actor"
4	"The Terminal"	2004	"Drama"	"Actor"
5	"Catch Me If You Can"	2002	"Crime"	"Actor"
6	"Sleepless in Seattle"	1993	"Romance"	"Collaborator"

6	"Sleepless in Seattle"	1993	"Romance"	"Collaborator"
7	"Titanic"	1997	"Romance"	"Collaborator"
8	"Catch Me If You Can"	2002	"Crime"	"Collaborator"
9	"Inception"	2010	"Sci-Fi"	"Collaborator"
10	"Interstellar"	2014	"Sci-Fi"	"Collaborator"

1	MATCH (d:Person)-[:DIRECTED]→(m:Movie)
2	WHERE d.yearsWorked > 40
3	WITH d, COLLECT(m.title) AS movies, COUNT(m) AS movieCount
4	WHERE movieCount ≥ 2
5	RETURN d.name AS name, d.country AS country, movies;
6	

	name	country	movies
1	"Steven Spielberg"	"USA"	["Saving Private Ryan", "The Terminal", "Catch Me If You Can"]

3. Tìm bạn diễn và phim (30 điểm)

3.1 Tìm người diễn chung với Tom Hanks trong phim ngân sách trên 50 triệu USD, có "WORKED_WITH" ít nhất 2 dự án. Trả về name, title, role, budget, projects, sắp xếp theo budget giảm dần.

1	MATCH (p:Person {name: "Tom Hanks"})-[:WORKED_WITH]-(coActor),
2	(coActor)-[:ACTED_IN]→(m:Movie)
3	WHERE m.budget > 50
4	WITH coActor, m, COUNT(coActor) AS projects
5	WHERE projects ≥ 2
6	RETURN coActor.name AS name, m.title AS title,
7	m.budget AS budget, projects
8	ORDER BY m.budget DESC;
9	

	name	title	budget	projects
1	"Leonardo DiCaprio"	"Titanic"	200	2
2	"Leonardo DiCaprio"	"Interstellar"	165	2
3	"Leonardo DiCaprio"	"Inception"	160	2
4	"Leonardo DiCaprio"	"Catch Me If You Can"	52	2

3.2 Tìm người diễn trong phim Sci-Fi ngân sách trên 100 triệu USD, hợp tác với Leonardo DiCaprio ít nhất 1 dự án. Trả về name, title, budget.

```
1 MATCH (p:Person {name: "Leonardo DiCaprio"})-[:WORKED_WITH]-
  (coActor),
2 |      (coActor)-[:ACTED_IN]→(m:Movie)
3 WHERE m.type = "Sci-Fi" AND m.budget > 100
4 RETURN coActor.name AS name, m.title AS title, m.budget AS budget;
5
```

	name	title	budget
1	"Natalie Portman"	"Inception"	160
2	"Natalie Portman"	"Inception"	160

3.3 Đếm số phim Romance ngân sách trên 150 triệu USD mà Kate Winslet tham gia. Trả về số lượng

```
1 MATCH (p:Person {name: "Kate Winslet"})-[:ACTED_IN]→(m:Movie)
2 WHERE m.type = "Romance" AND m.budget > 150
3 RETURN COUNT(m) AS num_of_movies;
4
```

	num_of_movies
1	1

4. Đường đi và phân tích (30 điểm)

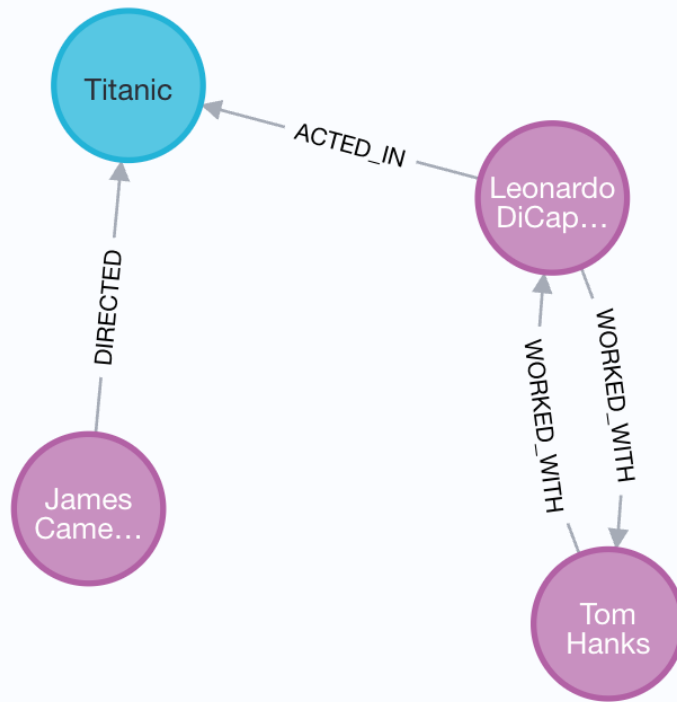
1. Tìm đường ngắn nhất từ Tom Hanks đến James Cameron qua "ACTED_IN", "DIRECTED", "WORKED_WITH". Tính tổng salary của Tom Hanks từ 3 phim ngân sách cao nhất, kiểm tra phim nào trong số đó nằm trên đường đi. Trả về path, totalSalary, titles, budgets.

Dùng bảng sau và hình "Network Diagram":

Từ	Đến	Phim top	Yếu tố kiểm tra
Tom Hanks	James Cameron	3 phim ngân sách cao nhất	Đường đi
Steven Spielberg	Natalie Portman	2 phim lương cao nhất	Thể loại

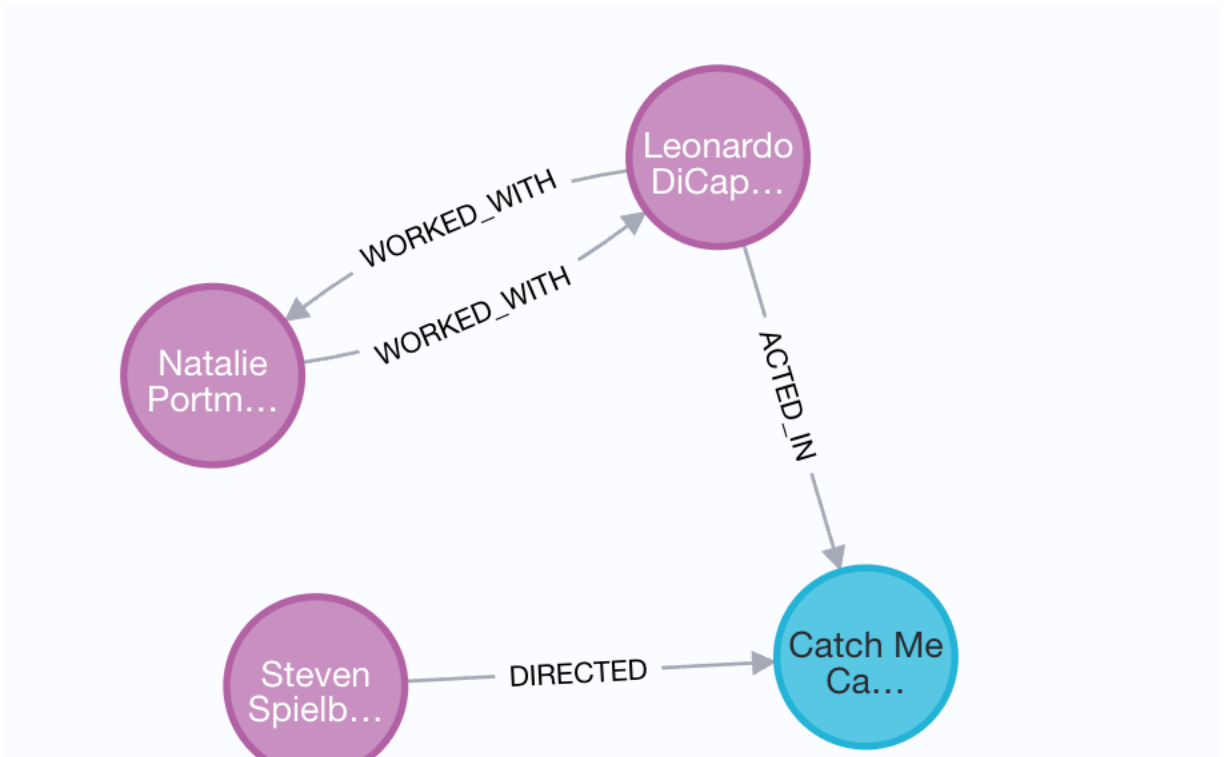
1. Tìm đường ngắn nhất từ Tom Hanks đến James Cameron qua "ACTED_IN", "DIRECTED", "WORKED_WITH". Tính tổng **salary** của Tom Hanks từ 3 phim ngân sách cao nhất, kiểm tra phim nào trong số đó nằm trên đường đi. Trả về **path**, **totalSalary**, **titles**, **budgets**.
2. Tìm đường ngắn nhất từ Steven Spielberg đến Natalie Portman qua các quan hệ trên. Tính tổng **salary** của Natalie Portman từ 2 phim lương cao nhất, kiểm tra thể loại của chúng. Trả về **path**, **totalSalary**, **titles**, **types**.
3. Đếm số người có "WORKED_WITH" với ít nhất 3 người khác trong hình. Trả về số lượng.

```
1 MATCH path = shortestPath((p1:Person {name: "Tom Hanks"})-  
[:ACTED_IN|DIRECTED|WORKED_WITH*]-(p2:Person {name: "James  
Cameron"}))  
2 WITH path  
3 MATCH (p1)-[:ACTED_IN]-(m:Movie)  
4 WHERE p1.name = "Tom Hanks"  
5 WITH path, m ORDER BY m.budget DESC LIMIT 3  
6 RETURN path, SUM(m.salary) AS totalSalary, COLLECT(m.title) AS  
titles, COLLECT(m.budget) AS budgets;  
7
```



2. Tìm đường ngắn nhất từ Steven Spielberg đến Natalie Portman qua các quan hệ trên. Tính tổng salary của Natalie Portman từ 2 phim lương cao nhất, kiểm tra thể loại của chúng. Trả về path, totalSalary, titles, types.

```
1 MATCH path = shortestPath((p1:Person {name: "Steven Spielberg"})-[:ACTED_IN|DIRECTED|WORKED_WITH*]-(p2:Person {name: "Natalie Portman"}))
2 WITH path
3 MATCH (p2)-[:ACTED_IN]->(m:Movie)
4 WHERE p2.name = "Natalie Portman"
5 WITH path, m ORDER BY m.salary DESC LIMIT 2
6 RETURN path, SUM(m.salary) AS totalSalary, COLLECT(m.title) AS titles, COLLECT(m.type) AS types;
7
```



3. Đếm số người có "WORKED_WITH" với ít nhất 3 người khác trong hình. Trả về số lượng.

```
1 MATCH (p:Person)-[:WORKED_WITH]-(coActor)
2 WITH p, COUNT(coActor) AS numConnections
3 WHERE numConnections ≥ 3
4 RETURN COUNT(p) AS numberOfPeople;
5
```

	numberOfPeople
1	2

