

ANO
2024



UNINTER

**CADERNO DE RESPOSTAS DA
ATIVIDADE PRÁTICA DE:**

NoSQL

ALUNO: (DARLEY LEAL DOS SANTOS - 4102341)

**Caderno de Resposta Elaborado por:
Prof. MSc. Guilherme Ditzel Patriota**

Prática 01 – JSON COM NEO4J.

Questão 01 – DESCOBERTA DA HASHTAG PRINCIPAL.

ENUNCIADO: Veja o Roteiro da Atividade Prática para mais detalhes.

- I. Apresentação dos comandos (códigos e queries) usados (não esquecer do identificador pessoal):

```
neo4j@bolt://localhost:7687/neo4j - Neo4j Browser
File Edit View Window Help Developer
1 CALL apoc.load.directory('*.*json') YIELD value
2 WITH value AS RU_4102341
3 ORDER BY RU_4102341 ASC
4 CALL apoc.load.json(RU_4102341) YIELD value
5 UNWIND value.data AS tweet
6
7 MERGE (a: Aluno {nome: 'Darley', ru: '4102341', curso: 'engenharia de software'})
8 MERGE (t: Tweet {tweet_id: tweet.id})
9 MERGE (u: Usuario {user_id: tweet.author_id})
10 WITH t, u, a, tweet.entities.hashtags AS hashtags
11
12 FOREACH (hashtag IN hashtags |
13     MERGE (h: Hashtag {tag: apoc.text.replace(apoc.text.clean(hashtag.tag), '[^a-zA-Z0-9]', '')})
14     ON CREATE SET h.count = 1
15     ON MATCH SET h.count = h.count + 1
16     MERGE (t)-[:CONTAINS]->(h)
17 )
18
19 MERGE (u)←[:CONTAINS]-(t)
20
21 RETURN t, u, a
```

Figura 1: A consulta utilizando Cypher inicia com o carregamento de todos os arquivos JSON que estavam na pasta 'import' do banco de dados. Em seguida, uma variável é criada para armazenar essas informações, utilizando o RU. Esta variável é então utilizada para importar os tweets dos arquivos. São criados vértices para representar as entidades: Tweet, Usuário, Hashtag e Aluno (sendo este último criado com dados do aluno para representação no grafo). Na próxima etapa foi criado um loop para normalizar os nomes das hashtags e contar quantas vezes elas aparecem nos tweets. Por fim, são criadas as arestas entre Tweet > Hashtag e Usuário > Tweet. Ao final, as variáveis são retornadas.

- II. Apresentação das Imagens/Print do resultado (não esquecer do identificador):

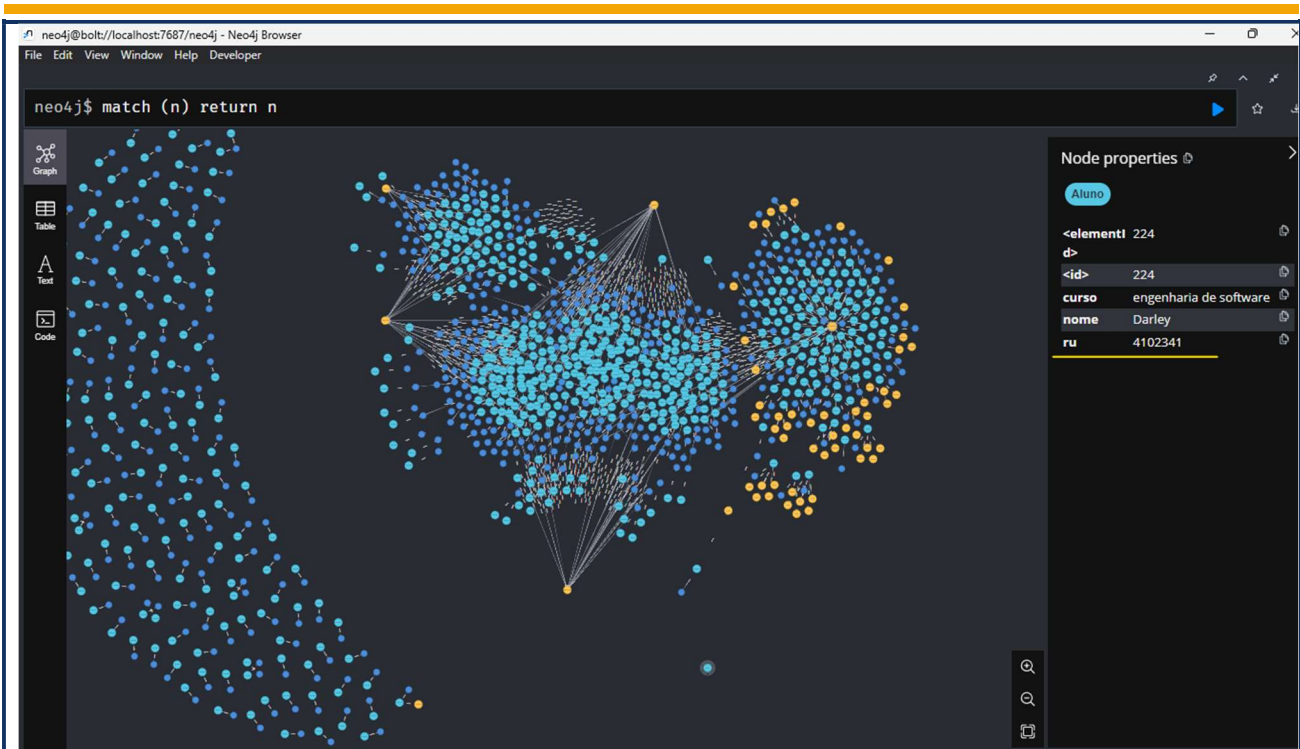


Figura 2: (Representação em grafo da consulta Cypher: a esquerda representa os vértices de usuários contendo tweets sem hashtags. A direita em maior quantidade temos os usuários ligados por arestas aos tweets e em amarelo as Hashtags. Ao centro do grafo temos as 4 hashtags com mais conexões, na parte superior encontra-se a hashtag 'depheliolopes' sendo o nó com mais relações entre as hashtags.)

III. Responda à pergunta: Qual foi a hashtag usada como filtro para coleta dos dados analisados?

Resposta: (A Hashtag utilizada foi 'depheliolopes' na consulta em 679 tweets ela foi utilizada 327 vezes.)

Prática 01 – JSON COM NEO4J.

Questão 02 – ANÁLISE DOS DADOS SEGUNDO VIÉS A SUA ESCOLHA.

ENUNCIADO: Veja o Roteiro da Atividade Prática para mais detalhes.

- I. Apresentação dos comandos (códigos e queries) usados (não esquecer do identificador pessoal):



```
1 CALL apoc.load.directory('*.*json') YIELD value
2 WITH value AS RU_4102341
3 ORDER BY RU_4102341 ASC
4 CALL apoc.load.json(RU_4102341) YIELD value
5 UNWIND value.data AS tweet
6
7 WITH tweet, tweet.entities.mentions AS mentions
8
9 FOREACH (mention IN mentions |
10   MERGE (m: Mentions {username: apoc.text.replace(apoc.text.clean(mention.username), '[^a-zA-Z0-9]', '')})
11   ON CREATE SET m.count = 1
12   ON MATCH SET m.count = m.count + 1
13 )
14 MERGE (t:Tweet)-[:CONTAINS]->(m)
15
16 RETURN t, m
17
```

Figura 3: A consulta em Cypher utiliza os dados dos arquivos na pasta 'import' do banco de dados, os quais são armazenados na variável RU_4102341 e na variável tweet após a importação. O objetivo da consulta é descobrir qual termo foi mais mencionado nos tweets coletados durante aquele período. Para isso, é acessado o campo 'mentions' dentro dos dados do JSON. Em seguida, é realizado um loop desse campo em todos os 679 tweets, normalizando os termos encontrados e contando quantas vezes eles foram utilizados. Posteriormente, são criados os nós que representam as menções e a quantidade delas. Ao analisar o resultado, observa-se que 'depheliolopes' foi mencionado 1602 vezes nos tweets.

- II. Apresentação das Imagens/Print do resultado (não esquecer do identificador):

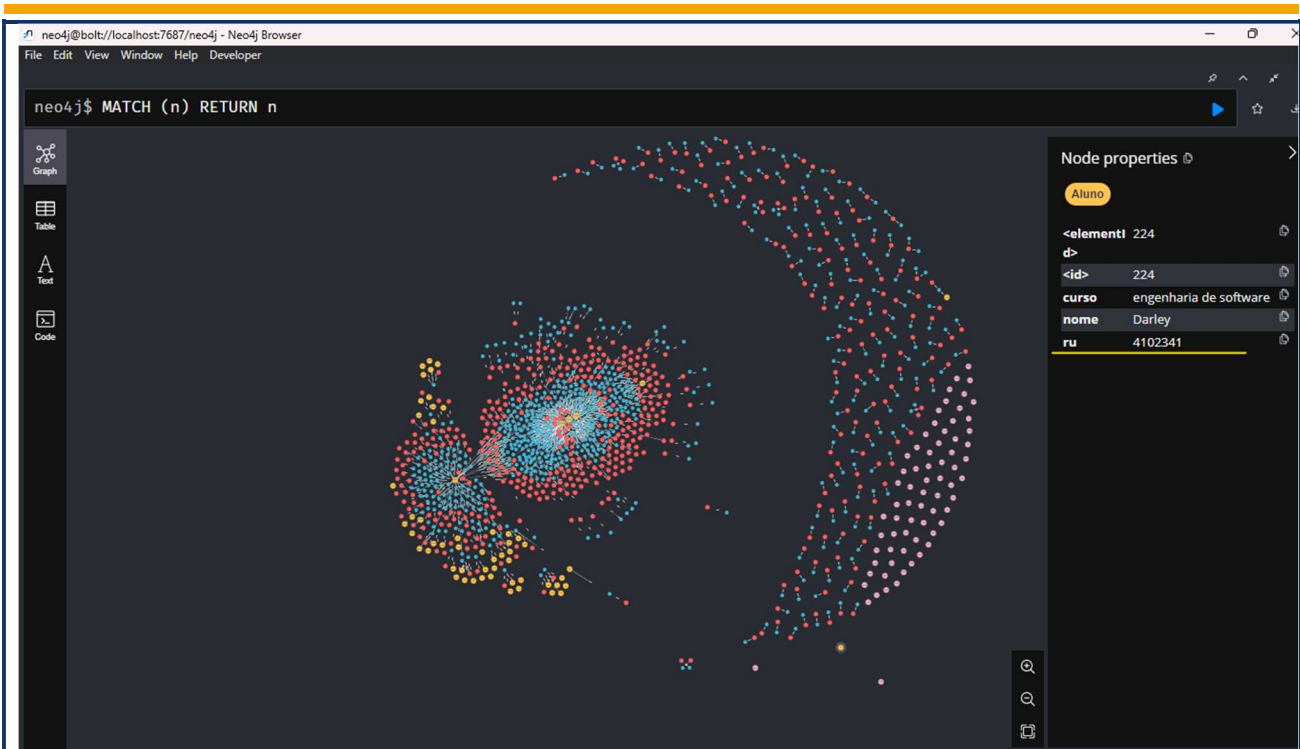


Figura 4: (Representação em grafo da consulta: à esquerda, estão os vértices dos usuários, tweets e hashtags interconectados entre si. À direita, temos os nós de usuários que possuíam tweets sem hashtags. No sudeste do grafo, destacados em tamanho, estão os nós de 'Mentions', que representam a quantidade de vezes que um termo foi mencionado nos tweets, com o termo em questão. O vértice localizado na parte inferior direita representa 'depheliolopes', mencionado 1602 vezes nos tweets.)

III. Responda à pergunta: Qual foi o comando usado por você para dar entrada dos dados em JSON no seu banco de dados Neo4j?

Resposta: 1. CALL apoc.load.directory('*.*json') YIELD value
2. WITH value AS RU_4102341
3. CALL apoc.load.json(RU_4102341) YIELD value

Na linha 1, é feita a chamada da função `apoc.load.directory`, que foi utilizada para importar os arquivos armazenados no diretório do banco de dados, com a extensão `.json`.

Na linha 2, foi utilizado a cláusula `WITH` para passar o resultado da chamada anterior para a próxima etapa do processo, renomeado como `'RU_4102341'` para referenciar as informações do aluno.

Na linha 3, é feita a chamada da função `apoc.load.json` para carregar o conteúdo do arquivo cujo caminho foi armazenado na variável `'RU_4102341'`. O resultado é armazenado na variável `value` que contém os dados importados e que poderá ser utilizado conforme necessário.