

Centro Universitário da FEI
Bacharelado em Ciência da Computação

Seminário – Banco De Dados Orientado a Grafos (Graph Database)

Nome: Camylla Lima Dias	R.A: 22.217.005-2
Nome: Evelyn Santana	R.A: 22.219.002-7
Nome: Guilherme Brigagão Cabelo	R.A: 22.120.071-0
Nome: Kaio Henrique da Silva Souza	R.A: 22.218.036-6
Nome: Thiago Lopes de Araujo	R.A: 22.218.011-6

Orientadora: Prof. Dra. Leila Bergamasco

São Bernardo do Campo
2022

INTRODUÇÃO

O tema abordado para o seminário a ser desenvolvido é “Banco de Dados orientado a Grafos/Graph Database(GDB)”. Os estudos de grafos pertencem aos campos da matemática discreta e computacional e permeiam diversas áreas, como tecnologia através de redes de computadores, complexidade de algoritmos e sistemas de banco de dados. O tema supracitado consiste na especialização do GBD em modelos de dados que utilizam vértices e arestas como meios de relacionamentos.

GRAPH DATABASE (GDB)

A teoria dos grafos surgiu como ramo da Matemática Discreta a partir do artigo publicado pelo matemático Leonhard Euler em 1736 sobre O Problema das Pontes de Königsberg, onde a solução consiste em, partindo de qualquer vértice, tentar atravessar todas as arestas uma única vez e retornar ao vértice de origem.

Um grafo pode ser definido como uma estrutura, onde é um conjunto discreto e ordenado de pontos chamados vértices e um conjunto de linhas chamadas arestas, onde cada aresta está conectada em pelo menos um vértice. Com a ideia de pontos interligados por linhas, a representação por grafos pode facilitar o entendimento e a resolução de problemas. Desta forma, um mapa da estrutura organizacional de uma empresa, redes de rotas de transporte, redes de comunicação, rotas de distribuição de produtos ou serviços e estruturas químicas de uma molécula, podem ser expressos através de grafos. (OSTROSKI e MENONCINI, 2009).

Por se tratarem basicamente de abstrações da realidade para se resolver problemas combinatórios, se tornaram um campo de interesse crescente nos dias atuais e suas aplicações são inúmeras, passando pelos problemas mais comuns de traçado de rotas até projetos de processadores eletrônicos, planejamento de horários, estudos sobre a estrutura do DNA, engenharia molecular e também pelas interligações elétricas e projetos de novos compostos químicos, sendo estes, extensões dos trabalhos de Kirchhoff e Cayley, grandes contribuintes para a história do desenvolvimento e estudo dos grafos. (NETTO e JURKIEWICZ, 2017).

Os modelos de banco de dados orientados a grafos podem ser definidos como aqueles em que as estruturas de dados para o esquema e instâncias são modeladas como gráficos ou generalizações deles, e a manipulação de dados é expressa por operações orientadas a grafos e construtores de tipo. Esses modelos decolaram nos anos 80 e início dos anos 90 ao lado dos modelos orientados a objetos. Sua influência desaparece gradualmente com o surgimento de outros modelos de banco de dados, em particular geográfico, espacial, semiestruturado e XML. No entanto, o interesse pelo modelo de banco de dados orientado a grafos ressurgiu juntamente com a web semântica e a disseminação das redes sociais. (ANGLES e GUTIERREZ, 2008).

Nos GDBs, durante a sua modelagem mais especificamente, é comum a utilização de grafos orientados que determinam uma orientação sobre as relações de pontos e arestas pertencentes a um grafo. Um caso de uso comum de grafos orientados é em um sociograma, que consiste de representações gráficas das relações existentes em um grupo de indivíduos do qual se é possível obter fatos sociométricos e análises sobre a estrutura de uma comunidade. (VAZ, 2007)

Uma forma de exemplificar a obtenção de fatos sociométricos é o levantamento de informações numa suspeita de fraude de trabalho acadêmico, o foco é caminhar na direção de uma padronização em termos de comportamentos comuns que podem levar-nos a ter suspeitas de ações fraudulentas entre dois ou mais alunos. Pode ser importante para nós sabermos o nível de confiança entre cada um dos alunos. Por exemplo, se se conhecem circunstancialmente, se são amigos ou mesmo se são namorados. O facto de frequentarem as mesmas disciplinas ou em que grupos se encontravam durante as avaliações pode ser outro elemento importante a se levar em consideração.

A representação desses relacionamentos em um banco de dados orientado a grafos é estruturado através de nodes (nós) onde cada “node” do tipo “Student” possui a si associadas às propriedades “age” e “number”, bem como o nome, presente no “label” de cada “node” deste tipo. O número representa o número mecanográfico e é do tipo “string”. Por exemplo, o estudante “TS” tem o número mecanográfico “a28248”, que serve para o distinguir no contexto escolar dos outros todos. Estas relações são denominadas na “graph database” como “friends”, “engaged” e “fraud”, respetivamente. Uma relação de amizade tem características (“properties”) como a data em que foi formada (esta propriedade tem o nome “since”), que representa o ano em que a amizade foi formada (se for possível saber) e o nível de importância que essa relação tem entre os dois alunos. Esta importância pode ser definida por um peso, “weight”, que difere conforme o nível de afinidade de um estudante com o outro. Este é um exemplo de aplicação entre muitos outros, e mais complexos. (SANTOS, 2021)

GRAFOS DE PROPRIEDADES

As arestas em um grafo de propriedades são rotuladas e, portanto, as arestas são heterogêneas. Um grafo de propriedades pode modelar relacionamentos de amizade, parentesco, negócios, comunicação, etc., todos dentro da mesma estrutura. Além disso, vértices e arestas em um grafo de propriedades mantêm um conjunto de pares chave/valor. Estes são conhecidos como propriedades e permitem a representação de dados não gráficos como por exemplo o nome de um vértice, o peso de uma aresta, etc.

No modelo de grafo de propriedades, é comum que as propriedades dos vértices (e às vezes arestas) sejam indexadas usando uma estrutura de árvore análoga àquelas usadas por bancos de dados relacionais. Os elementos em um grafo são adjacentes uns aos outros por referência direta, um vértice é adjacente às suas arestas de entrada e saída e uma aresta é adjacente aos seus vértices de saída e de entrada. (RODRIGUEZ e NEUBAUER, 2010).

Os grafos de propriedades, também chamados de grafos rotulados por arestas ou multigrafo atribuído, são usados extensivamente na computação, pois são mais expressivos do que os objetos matemáticos simplificados estudados em teoria, é claro que devido à sua facilidade de uso e não pelos limites do que pode ser modelado.

GRAFOS RDF (Resource Description Framework)

Um documento RDF é estruturado em forma de grafo direcionado (também chamado de grafo orientado, grafo dirigido, dígrafo ou quiver), isto é, um conjunto de nós (ou vértices) que estão ligados por arestas direcionadas (setas). Em RDF, tanto os nós quanto as arestas são rotulados com identificadores que os distinguem. (FERREIRA e SANTOS, 2013)

É uma forma de representar os dados que navegam pela web, chamados de recursos. O formato de um RDF contém a seguinte estrutura: <sujeito> <predicado> <objeto>. A tags <sujeito> e <objeto> representam recursos e a tag <predicado> indica o relacionamento entre eles. Um recurso também pode estabelecer conexão com outro recurso e a sua representação gráfica de RDF é compreendida por intermédio de um grafo dirigido.

BANCOS DE DADOS ORIENTADO A GRAFOS (Softwares)

Podemos exemplificar como uma opção de bancos de dados orientados a grafos o Amazon Neptune, que é um mecanismo com projeto específico e alta performance, otimizado para armazenar bilhões de relacionamentos e consultar grafos com latência de milissegundos. O Neptune oferece suporte aos grafos de propriedades de modelos de grafos populares e ao Resource Description Framework (RDF) do W3C, bem como às suas respectivas linguagens de consulta, Apache TinkerPop Gremlin e SPARQL.

Uma das ferramentas que ficou muito famosa e muito utilizada foi a Neo4j, que funciona de uma forma simples e objetiva, funcionando a partir de nós(entidades) e relacionamentos(arestas), tendo como ferramenta para consulta a linguagem Cypher, que é muito similar ao SQL. O Cypher Query Language é a linguagem oficial de consultas do Neo4j e permite que se crie, modifique e procure dados em uma estrutura baseada em um grafo de informações e relacionamentos. Tendo como o tipo de arquivo gerado com a extensão .graphdb, também conseguindo exportar arquivos .csv, .json, .xml e etc.

Temos ferramentas que divergem em algumas coisas, como a ArangoDB que é um banco de dados NoSQL nativo e híbrido que combina os modelos chave-valor, baseado em documentos e orientados a grafos. Algumas outras ferramentas: Dgraph, DataStax, FlockDB, OrientDB, Titan.

APRESENTAÇÃO

O seminário do tópico Banco de Dados Orientado a Grafos (GDB) será feito através de uma apresentação em slides para introduzir os alunos sobre a origem, utilização e casos de uso do GDB em conjunto com um workshop onde iremos implementar uma solução simples com Neo4j. Vamos finalizar com um quiz para reforçar todo o conteúdo apresentado durante o seminário, através do Kahoot verificaremos qual aluno acertou a maior quantidade de questões e este levará um brinde.

Tópicos do seminário:

1. O que é um GDB?
2. Por que ele é utilizado?
3. Quais as suas vantagens em relação ao modelo relacional?
4. Casos de uso;
5. Quais softwares trabalham com GDB?
6. Implementação/Demonstração de GDB com Neo4j
7. Quiz sobre GraphDB utilizando Kahoot

REFERÊNCIAS BIBLIOGRÁFICAS:

ANGLES, R; GUTIERREZ, C. **Survey of graph database models**. ACM Computing Surveys, 40(1):1, 2008.

OSTROSKI, A; MENONCINI, L. **TEORIA DOS GRAFOS E APLICAÇÕES**. Synergismus scyentifica UTFPR, Pato Branco, 04 (2) . 2009.

NETTO, P. JURKIEWICZ, S. **Grafos: Introdução e prática**. Editora Blucher, 2017.

RODRIGUEZ, M. A.; NEUBAUER, P. **The graph traversal pattern**. CoRR, 2010.

FERREIRA, J. A. ; SANTOS, P. L. V. A. C. **O modelo de dados Resource Description Framework (RDF) e o seu papel na descrição de recursos**. Informação & Sociedade (UFPB. Online), v. 23, p. 13-23, 2013.

VAZ, G. J. **A Construção dos Sociogramas e a Teoria dos Grafos**. Rev. bras. psicodrama vol.17 no.2 São Paulo, 2009.

SANTOS, T. F. **Ontologias para detecção de fraude em meio acadêmico**. Instituto Politécnico de Bragança. Bragança, p. 74, 2021.