

**UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO TECNOLÓGICO
DEPARTAMENTO DE ENGENHARIA ELÉTRICA
PROJETO DE GRADUAÇÃO**



MATEUS BENTO ALVES VASCONCELLOS

**ANÁLISE DE SISTEMAS GERENCIADORES DE
BANCOS DE DADOS (SGBD) PARA
ARMAZENAMENTO DE UMA QUANTIDADE
VOLUMOSA DE GRAFOS**

VITÓRIA-ES

MARÇO/2022

Mateus Bento Alves Vasconcellos

ANÁLISE DE SISTEMAS GERENCIADORES DE BANCOS DE DADOS (SGBD) PARA ARMAZENAMENTO DE UMA QUANTIDADE VOLUMOSA DE GRAFOS

Parte manuscrita do Projeto de Graduação do aluno Mateus Bento Alves Vasconcellos, apresentado ao Departamento de Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do grau de Engenheiro Eletricista.

Vitória-ES

Março/2022

Mateus Bento Alves Vasconcellos

ANÁLISE DE SISTEMAS GERENCIADORES DE BANCOS DE DADOS (SGBD) PARA ARMAZENAMENTO DE UMA QUANTIDADE VOLUMOSA DE GRAFOS

Parte manuscrita do Projeto de Graduação do aluno Mateus Bento Alves Vasconcellos, apresentado ao Departamento de Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do grau de Engenheiro Eletricista.

**Profa. Dra. Marcia Helena Moreira
Paiva**

Universidade Federal do Espírito Santo
Professor da Disciplina

**Profa. Dra. Marcia Helena Moreira
Paiva**

Universidade Federal do Espírito Santo
Orientador

Mateus Bento Alves Vasconcellos

Universidade Federal do Espírito Santo
Aluno

Vitória-ES

Março/2022

RESUMO

Diversas indústrias, comércios e a área acadêmica utilizam bancos de dados para armazenar as mais variadas formas de dados, gerando uma demanda cada vez maior por métodos de melhoras de performance de todos os tipos. Surge, portanto, o desafio de encontrar as otimizações mais eficazes e os melhores gerenciadores de bancos de dados para reduzir o tempo de consulta e volume para cada uso específico.

Ao mesmo tempo, a Teoria de Grafos emerge como um modelo eficaz, capaz de resolver, visualizar e armazenar problemas matemáticos, da engenharia, da computação e da indústria pela suas diversas aplicações práticas.

Este trabalho busca estudar e aplicar testes de performance nos principais SGBD (Sistemas Gerenciadores de Bancos de Dados) realizando a comparação de performance de consulta e volume utilizado para os principais tipos de consultas a bancos de dados de grafos.

Palavras-chave : Base de Dados. Performance. Grafos. Sistemas Gerenciadores.

LISTA DE FIGURAS

Figura 1 – Pontes de <i>Königsberg</i>	8
Figura 2 – Exemplo de Grafo com seis nós	11

LISTA DE TABELAS

Tabela 1 – Lista de atividades.	13
Tabela 2 – Diagrama de tempo das atividades a serem efetuadas	14

LISTA DE ABREVIATURAS E SIGLAS

CPU	<i>Central Processing Unit</i>
CPID	Centro de Pesquisa, Inovação e Desenvolvimento
GPU	<i>Graphics Processing Unit</i>
SGBD	Sistemas Gerenciadores de Banco de Dados
SQL	<i>Structured Query Language</i>
UFES	Universidade Federal do Espírito Santo

SUMÁRIO

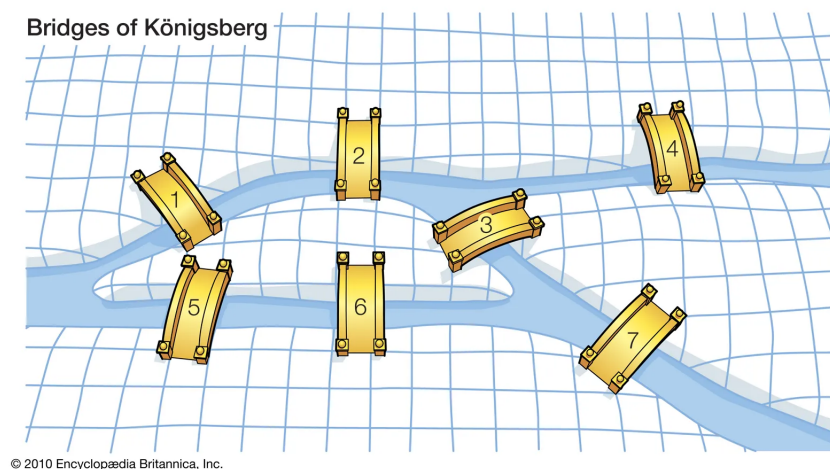
1	INTRODUÇÃO	8
2	JUSTIFICATIVAS	9
3	OBJETIVOS	10
3.1	Objetivo Geral	10
3.2	Objetivos Específicos	10
4	REFERENCIAL TEÓRICO	11
4.1	Teoria de Grafos	11
4.2	Bancos de Dados	12
5	METODOLOGIA E ETAPAS DE DESENVOLVIMENTO	13
5.1	Metodologia Adotada	13
5.2	Cronograma de Trabalho	13
6	ALOCAÇÃO DE RECURSOS	15
6.1	Recursos Materiais	15
6.2	Recursos Computacionais	15
	REFERÊNCIAS	16

1 INTRODUÇÃO

A pesquisa e a otimização de bancos de dados são assuntos em alta na área da computação há décadas, estando presentes em todos os tipos de sistemas e aplicativos utilizados mundialmente. No entanto, conforme a indústria se solidifica e amadurece, a competitividade e a necessidade de sistemas mais eficientes também aumentam.

Um problema de bancos de dados pode ser entendido como um problema de grafos quando analisa-se não só os valores a serem armazenados, mas suas conexões e relações (GUBICHEV, 2015). O primeiro artigo sobre grafos foi escrito pelo matemático Leonard Euler, em 1736, enquanto tentava descobrir se era possível atravessar as sete pontes da cidade de *Königsberg*, na Prússia, sem repetir nenhuma (BIGGS E. KEITH LLOYD, 1986). A Figura 1 ilustra o problema. Euler abstraiu os possíveis caminhos das pontes em retas e suas intersecções em pontos, criando, talvez, o primeiro Grafo da história. Desde então, o tema tem ganhado cada vez mais tração, por ser altamente utilizado para abstração de relações entre objetos, principalmente no âmbito da computação.

Figura 1 – Pontes de *Königsberg*



© 2010 Encyclopædia Britannica, Inc.

Fonte: Enciclopédia Britannica, 2010

Dentro destes cenários, surge a necessidade de armazenar dados de grafos em bancos de dados. No entanto, quando se lida com grafos com grande volume de dados, é preciso identificar o melhor SGBD para consulta, visando a rapidez do sistema.

Sendo assim, este trabalho propõe implementar e testar os principais SGBD para consultas costumeiras de bancos de grafos, analisar os resultados e definir o mais eficiente e adequado à situação.

2 JUSTIFICATIVAS

Com a popularização da internet, que se tornou uma necessidade tanto industrial quanto doméstica, e, conseqüentemente, sua adoção mundial em massa, crescem as redes de telecomunicação e a quantidade de dados armazenados. De acordo com Barnett (2016), o tráfego global da internet em 2016 chegava a 1,1 Zettabytes (10^{12} GB) por ano, chegando a 2 Zettabytes no fim de 2019.

O crescente uso da internet levou à necessidade de criação de sistemas de armazenamento de dados eficientes, capazes de oferecer rapidez e confiabilidade. Problemas de lentidão para consultas a bancos de dados podem acarretar em perdas expressivas tanto para as empresas quanto para os usuários. Atualmente, a internet pressupõe rapidez quase que instantânea para acesso a bancos de dados em sites, aplicativos, jogos, etc.

O processamento destes dados só é possível se o resgate, registro e remoção das principais informações forem rápidos e eficientes. É no processo de extração de informações que os bancos de dados revelam-se fundamentais.

Ter conhecimento sobre o SGBD mais adequado a cada tipo de dado, portanto, é fundamental para garantir a rapidez e a efetividade dos sistemas. Neste trabalho, o objetivo é encontrar o SGDB com melhor desempenho para utilização nas diferentes consultas a bancos de dados de grafos altamente volumosos.

3 OBJETIVOS

3.1 Objetivo Geral

O objetivo geral deste trabalho é encontrar o SGBD mais performante para consultas de um banco de dados de grafos altamente volumoso.

3.2 Objetivos Específicos

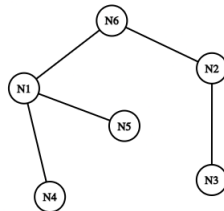
- Pesquisar e implementar *drivers* para diversos SGBD.
- Pesquisar e aplicar melhorias de performance para os *drivers* dos SGBD.
- Aferir e comparar a performance de diferentes SGBD.
- Estabelecer as vantagens e desvantagens de cada SGBD.
- Analisar os dados obtidos.

4 REFERENCIAL TEÓRICO

4.1 Teoria de Grafos

Um grafo G qualquer é composto de um conjunto de nós (também chamados de pontos ou vértices) ligados a outros nós por meio de arestas, que representam uma ligação atrelada a uma relação entre os nós (WEST, 2018). A figura Figura 2 ilustra um Grafo de seis nós $n1, n2, n3, n4, n5, n6$ cujas ligações podem ser representadas pelos nós percententes àquela aresta, portanto, $n4n1, n1n5, n1n6, n6n2, n2n3$ são as arestas do Grafo ilustrado.

Figura 2 – Exemplo de Grafo com seis nós



Fonte: Produção do próprio autor.

Alguns conceitos fundamentais de grafos, de acordo com Diestel (2017)

- **Ordem** : indica o número de vértices de um grafo.
- **Extremos** : dois vértices que formam uma aresta são chamados de extremos.
- **Adjacência** : dois vértices x, y de G são adjacentes quando a aresta xy pertence a G ; duas arestas são adjacentes se possuem um fim em comum.
- **Incidência** : se um vértice é o extremo de um nó, eles são incidentes.
- **Caminho** : é uma sequência de nós que une uma sequência distinta de arestas.
- **Distância** : é o menor número de arestas entre dois nós.
- **Diâmetro** : é o maior número de arestas distintas que se atravessa entre dois nós.
- **Loop** : é uma aresta que conecta um nó a si mesmo.

Na ciência da computação, grafos são utilizados em estruturas de dados para a representação de redes de telecomunicações, problemas de logística, design de circuitos elétricos, diagramas de árvore, árvore de decisões, algoritmo de Dijkstra, etc.

4.2 Bancos de Dados

De acordo com Silberschatz, Korth e Sudarshan (2011), bancos de dados são estruturas que contêm dados inter-relacionados, tipicamente armazenados eletronicamente em sistemas de computadores, gerenciados por SGBD. Constituem parte essencial de qualquer comércio na atualidade, pois a todo momento usuários interagem com bancos de dados ao navegar na internet, mesmo que inconscientemente.

Os SGBD são úteis para proporcionar formas de armazenar e retirar informações de bancos de forma simples e eficiente, ou seja, são responsáveis pelo gerenciamento dos bancos de dados. Permitem ao usuário criar e especificar esquemas (forma na qual os dados estão organizados) para os dados, fazer consultas ao banco, fazer consultas concorrentes aos mesmos dados, etc (GARCIA-MOLINA; ULLMAN; WIDOM, 2014). Em suma, provê ao usuário uma interface mais abstrata entre os dados e a aplicação.

Em 1970, Ted Codd, um matemático e pesquisador da IBM propôs que os sistemas de bancos de dados apresentassem aos usuários dados organizados em forma de tabelas chamadas relações (Codd, 1970), que são conexões lógicas entre diferentes tabelas, baseadas em suas interações. Esse sistema é conhecido atualmente como bancos de dados relacionais. Este modelo contém uma ou mais tabelas com linhas e colunas; cada linha possui uma identificação (id) única e cada coluna representa um atributo. Também é possível associar linhas de diferentes tabelas gerando uma chave estrangeira.

A linguagem de programação dominante, utilizada por praticamente todos os bancos de dados relacionais, é a linguagem SQL (*Structured Query Language*), responsável pela administração de permissões, consulta e manipulação dos dados.

5 METODOLOGIA E ETAPAS DE DESENVOLVIMENTO

5.1 Metodologia Adotada

De acordo com Gil (2002), este trabalho pode ser classificado como aplicado, por tentar solucionar um problema específico em uma circunstância particular. É classificado como descritivo, visto que seus objetivos são a coleta, análise e interpretação dos dados. Do ponto de vista dos procedimentos técnicos, é classificado como experimental, visto que, o trabalho busca identificar as variáveis do processo e suas dependências, e, mediante análise quantitativa, obter conclusões acerca dos dados coletados.

A princípio, serão feitos estudos de Teoria de Grafos, das diversas SGDB e suas peculiaridades e de métodos de aumento de performance em bancos de dados. Em seguida, serão escolhidos os SGDB mais interessantes e pertinentes ao uso de dados de Grafos volumosos. Após a escolha, serão realizados testes para avaliar o desempenho de cada SGDB a partir do tempo de consulta de operações comuns de Grafos. Finalmente, será feita a análise dos dados coletados para levantamento de uma conclusão sobre o SGDB mais performante.

5.2 Cronograma de Trabalho

Apresenta-se nest seção uma previsão do cronograma do plano de trabalho. O tempo total previsto para a conclusão é de 5 meses.

Na Tabela 1, são detalhadas as atividades que se pretendem realizar para o desenvolvimento do plano de trabalho. Assim, na primeira coluna da tabela são definidos os rótulos de cada atividade, e na segunda coluna, é feita uma descrição da atividade que se pretende realizar. Finalmente, na Figura 2, é apresentado o diagrama de tempo das atividades indicadas na Tabela 1.

Rótulo	Atividade
ATV 1	Estudo sobre bancos de dados e suas otimizações
ATV 2	Estudo sobre grafos
ATV 3	Programação da comunicação com o SGDB
ATV 4	Testes de performance
ATV 5	Avaliação dos testes de performance
ATV 6	Escrita do projeto de graduação
ATV 7	Defesa do projeto de graduação

Tabela 1 – Lista de atividades.

Tabela 2 – Diagrama de tempo das atividades a serem efetuadas

Meses	Abril				Maio				Junho				Julho				Agosto			
Semanas	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Atividade 1																				
Atividade 2																				
Atividade 3																				
Atividade 4																				
Atividade 5																				
Atividade 6																				
Atividade 7																				

Fonte: Produção do próprio autor.

6 ALOCAÇÃO DE RECURSOS

6.1 Recursos Materiais

O material bibliográfico utilizado é composto principalmente por periódicos científicos, artigos e livros. Consultados fisicamente, através da Biblioteca Central da UFES, pertencentes ao próprio autor, e, eletronicamente, via rede de internet da UFES.

6.2 Recursos Computacionais

As etapas de desenvolvimento do software serão realizadas utilizando a linguagem de programação *Python*, na sua versão mais atual, 3.10.2. Também será utilizada a tecnologia de containerização de código aberto *Docker*, para executar a aplicação num contêiner isolado. Serão testados os principais SGBD, como PostgreSQL, MySQL, MongoDB e MariaDB.

Além disso, o computador a ser utilizado nos experimentos é de propriedade do autor e possui a seguinte configuração: *(i)* sistema operacional Linux, distribuição Arch, Kernel versão 5.16; *(ii)* processador AMD 3600x, 3.80GHz com 6 núcleos físicos; *(iii)* memória RAM de 16 GB 3200MHz; *(iv)* armazenamento de 256GB (Unidade de estado sólido); *(v)* placa de vídeo Nvidia Geforce RTX 2060, com 6 GB de memória de vídeo dedicada.

REFERÊNCIAS

- BARNETT, T. The Zettabyte Era Officially Begins. Cisco, 2016. Disponível em: <<https://blogs.cisco.com/sp/the-zettabyte-era-officially-begins-how-much-is-that>>. Citado na página 9.
- BIGGS E. KEITH LLOYD, R. J. W. N. Graph Teory 1736-1936. [S.l.]: Clarendon Press, 1986. Citado na página 8.
- CODD, E. F. A relational model of data for large shared data banks. Communications of the ACM, v. 13, n. 6, p. 377–387, 1970. Citado na página 12.
- DIESTEL, R. Graph theory. [S.l.]: Springer, 2017. Citado na página 11.
- GARCIA-MOLINA, H.; ULLMAN, J. D.; WIDOM, J. Database systems: The complete book. 2. ed. [S.l.]: Pearson, 2014. Citado na página 12.
- GIL, A. C. Como Elaborar Projetos de Pesquisa. 4. ed. [S.l.]: Atlas, 2002. Citado na página 13.
- GUBICHEV, A. Query processing and optimization in graph databases. In: . [S.l.: s.n.], 2015. Citado na página 8.
- SILBERSCHATZ, A.; KORTH, H. F.; SUDARSHAN, S. Database system concepts. [S.l.]: McGraw-Hill, 2011. Citado na página 12.
- WEST, D. B. Introduction to graph theory. [S.l.]: Pearson, 2018. Citado na página 11.