

UNIVERSIDADE FEDERAL DE SÃO CARLOS - CAMPUS SOROCABA

CCGT - CIÊNCIA DA COMPUTAÇÃO

Processamento Massivo de Dados

Prof. Dra. Sahudy Montenegro González

Projeto Prático

Fase Intermediária I

João Vitor Naves Mesa - 814149

Pedro Gonçalves Correia - 813281

17/06/2025

SUMÁRIO

1. Introdução.....	3
2. Objetivos e Escopo.....	3
2.1. Objetivo geral.....	3
2.2. Objetivos por área.....	4
3. Áreas de Aplicação e Estudos de Caso.....	4
3.1 Setor Governamental.....	4
3.2 Serviços Financeiros.....	4
3.3 Telecomunicações.....	5
3.4 Varejo.....	5
4. Metodologia de Trabalho.....	6
5. Cronograma.....	6
6. Referências.....	7

1. Introdução

A modelagem de dados por meio de grafos tem se destacado como uma alternativa promissora para representar relações complexas e dinâmicas que, em muitas aplicações, se mostram difíceis de capturar com bancos de dados relacionais tradicionais. Ao invés de depender de junções dispendiosas em tabelas, os bancos de grafos permitem navegar diretamente pelas conexões entre entidades, oferecendo desempenho mais consistente em consultas que exploram múltiplos níveis de relacionamentos (ROBINSON, WEBBER & EIFREM, 2015). Essa abordagem é particularmente relevante em cenários como detecção de fraudes, análise de redes sociais e otimização de rotas, nos quais a topologia do dado é tão valiosa quanto os próprios atributos dos nós.

O Neo4j, entre as diversas soluções de mercado (por exemplo, ArangoDB, Amazon Neptune e TigerGraph), destaca-se por empregar uma engine nativa de grafos que armazena e processa diretamente arestas e vértices, sem camada de abstração relacional. Sua linguagem de consulta, Cypher, adota uma sintaxe declarativa baseada em padrões de subgrafos, o que facilita a expressão de consultas complexas de maneira legível e intuitiva (NEO4J, [s.d]). Além disso, o ecossistema da plataforma inclui extensões como APOC (coleção de procedimentos auxiliares) e o módulo Graph Data Science, que ampliam as capacidades analíticas e permitem a execução de algoritmos de centralidade, detecção de comunidades e predição de ligações.

Este relatório apresenta o Neo4j como uma das principais opções de banco de dados em grafos, avaliando seu arcabouço técnico e suas funcionalidades de modo imparcial, sem esgotar todas as soluções de mercado. Baseando-se em white papers oficiais da plataforma, serão examinados casos de aplicação em setores diversos, de modo a identificar os cenários em que o Neo4j se mostra mais adequado e apontar, para cada contexto, eventuais problemas e alternativas viáveis.

2. Objetivos e Escopo

O projeto tem como objetivo principal compreender de forma teórica o uso do Neo4j em diferentes domínios de atuação, mais especificamente dentro de grandes áreas da indústria, como: governo, serviços financeiros, telecomunicações e varejo. Com isso, busca-se entender e identificar suas principais vantagens com foco no proveito obtido com a utilização da modelagem dos relacionamentos por meio dos grafos oferecida pelo sistema de gerenciamento de banco de dados, mas, além disso, pontuar os seus desafios, ao se tratar de consistência, escalabilidade, performance e complexidade de modelagem. Portanto, o projeto centra-se no entendimento do porquê essa ferramenta é utilizada nas áreas estudadas.

3. Áreas de Aplicação e Estudos de Caso

Nesta seção, são apresentadas quatro áreas de aplicação em que o Neo4j tem sido empregado segundo relatos técnicos da própria empresa: governo, serviços financeiros, telecomunicações e varejo. Cada subseção abordará como a tecnologia foi aplicada em situações específicas, destacando os objetivos pretendidos, as funcionalidades exploradas e os principais aprendizados observados. Esta análise servirá de base para a comparação crítica entre os diferentes contextos, considerando as vantagens e as limitações associadas ao uso da plataforma.

3.1 Setor Governamental

Agências governamentais lidam frequentemente com volumes massivos de dados distribuídos em múltiplos repositórios — cadastros de cidadãos, registros financeiros, protocolos de investigação e inventários logísticos — cujas inter-relações são cruciais para o cumprimento de suas atribuições. Em cenários como detecção de fraudes, análise de redes de influência e otimização de cadeias de suprimentos, a topologia dos dados carrega valor tão relevante quanto os próprios atributos. Modelar essas conexões com junções em bancos relacionais tende a ser lento e dispendioso; por isso, cresce o interesse por soluções nativas de grafos, capazes de revelar padrões e percursos em tempo real, sem a sobrecarga de operações de join distribuídas.

O white paper “Graphs in Government” apresenta diversos casos reais, entre eles a modernização do sistema de gestão de custos de manutenção do Exército dos EUA. Antes baseado em mainframes, o sistema tornava impossível rastrear dinamicamente mais de 15 milhões de relações entre peças e equipamentos, resultando em custos imprevisíveis. Com Neo4j, foi possível construir um repositório único de dados logísticos, permitindo consultas “de minutos para milissegundos” e suportando cenários de “what-if” para projeções de logística e orçamento (ZAGALSKY, 2019).

3.2 Serviços Financeiros

O setor de serviços financeiros enfrenta um ambiente regulatório cada vez mais exigente, em que a transparência nas cadeias de ativos e a capacidade de resposta a fraudes determinam vantagens competitivas e conformidade legal. Estruturas tradicionais baseadas em bancos relacionais frequentemente se veem sobrecarregadas pela complexidade crescente de produtos financeiros — como derivativos, fundos e instrumentos compostos — cujas interdependências podem se estender por várias camadas de empacotamento e revenda. Nessa conjuntura, bancos de grafos oferecem um modelo mais natural para rastrear e analisar relacionamentos financeiros em tempo real, atendendo às demandas de risco, compliance e inovação em customer engagement.

Um estudo de caso destacado no white paper é o da Cerved, agência italiana líder em análise de risco de crédito. A aplicação inicial tinha como objetivo identificar o “real owner” de empresas, tarefa que envolvia navegar por até 15 elos de controle societário e atender a requisitos da Lei 231/2007 para prevenção à lavagem de dinheiro. A solução relacional previa

processamento em cerca de 12 segundos por consulta; ao migrar para Neo4j, o tempo foi reduzido para 67 milissegundos, permitindo respostas em tempo real e expansão do uso da tecnologia em outras áreas da empresa (MATHUR, 2020).

3.3 Telecomunicações

O setor de Telecomunicações e operações de rede tem vivenciado um crescimento exponencial de complexidade, impulsionado por iniciativas como Network Functions Virtualization (NFV), Software-Defined Networking (SDN) e automação em múltiplas camadas de infraestrutura legada. Essa evolução trouxe desafios para os provedores de serviços de comunicação (CSP) e equipes de TI, que precisam ultrapassar abordagens reativas e isoladas para alcançar uma visão holística, precisa e em tempo real das dependências entre componentes físicos, virtuais e de serviço.

O white paper “Advanced Service Assurance” descreve a adoção de soluções de Next-Generation Service Assurance (NGSA) baseadas em grafos, cujo objetivo é unificar dados de topologia, indicadores de performance e métricas de experiência do cliente em um único modelo conectado. Por meio desse repositório de grafos, é possível correlacionar eventos de falha, priorizar incidentes segundo o impacto no usuário final e executar análises preditivas de “what-if”, habilitando respostas automáticas e planejamento proativo de manutenções (HODLER, 2021).

3.4 Varejo

O setor de varejo enfrenta hoje desafios complexos, impulsionados pela concorrência de gigantes digitais que combinam baixos custos operacionais com algoritmos de recomendação em tempo real. Para se manterem competitivos, os varejistas tradicionais precisam não apenas otimizar cadeias de suprimentos e roteirização de entregas, mas sobretudo oferecer experiências altamente personalizadas. Isso requer o acesso imediato a múltiplas fontes de dados — histórico de compras, comportamento de navegação, promoções vigentes e atributos de produtos — e a capacidade de explorar suas interconexões sem atrasos, algo que ultrapassa as capacidades de bancos relacionais e sistemas de processamento em batch.

O white paper “Driving Innovation in Retail with Graph Technology” ilustra como o grupo de e-commerce da Walmart Brasil empregou o Neo4j desde 2013 para substituir um processo batch complexo por consultas em tempo real, construindo um motor de recomendações que combina dados históricos e de sessão do usuário. Com o Neo4j, o time de desenvolvimento reduziu a lógica de recomendação a consultas de baixa latência, habilitando up-sell e cross-sell dinâmicos que antes não eram possíveis com a base relacional (RATHLE, 2021).

4. Metodologia de Trabalho

A metodologia idealizada pela equipe se baseia na divisão dos white papers para análise entre os dois membros, de forma que as áreas de telecomunicações e governamental

serão estudadas por João Mesa, já as demais - serviços financeiros e varejo - serão abordadas por Pedro Correia. Desse modo, cada participante conduzirá, de maneira independente, uma leitura crítica dos materiais, mapeando os elementos alinhados aos objetivos definidos na seção anterior. Em seguida, promover-se-á uma sessão de discussão coletiva para entendimento das análises realizadas e para consolidar as conclusões que subsidiarão o relatório final.

Por fim, é relevante ressaltar que, para assegurar o rigor e a comparabilidade entre as quatro áreas, serão seguidos cinco critérios de análise ao longo da leitura dos white papers: casos de uso, recursos/funcionalidades, vantagens reportadas, limitações/desafios e indicadores de sucesso. Durante a leitura focalizada, cada membro registrará suas observações em um documento auxiliar, posteriormente, ocorrerá uma revisão cruzada - em que cada integrante valida os apontamentos do outro - antes de uma reunião de consolidação, em que os resultados serão discutidos e formalizados nas conclusões obtidas através da execução do projeto.

5. Cronograma

Nesta seção, será exposto o cronograma do projeto com o intuito de detalhar as três entregas que serão realizadas e oferecer transparência acerca do andamento e organização do projeto.

CRONOGRAMA	
Data de entrega	Objetos entregues
18/06/2025 (Fase Intermediária 1)	Proposta inicial, objetivos do projeto, áreas de aplicação analisadas e metodologia.
02/07/2025 (Fase Intermediária 2)	Refinamento da proposta e realização de dois dos estudos de caso.
16/07/2025 (Fase Final)	Relatório final, estudos de casos completos e conclusões obtidas com a realização do trabalho.

Tabela 1 - Cronograma do projeto prático

6. Referências

- [1] ROBINSON, Ian; WEBBER, Jim; EIFREM, Emil. *Graph databases: new opportunities for connected data*. Sebastopol, CA: O'Reilly Media, 2015.
- [2] NEO4J, INC. *Cypher Manual*. [S.l.]: Neo4j. Disponível em: <https://neo4j.com/docs/cypher-manual/current/>. Acesso em: 14 jun. 2025.
- [3] ZAGALSKY, Jason. *Graphs in Government: Fulfilling Your Mission with Neo4j Whitepaper*. [S.l.]: Neo4j, 2019. Disponível em: <https://neo4j.com/use-cases/government/>. Acesso em: 16 jun. 2025.
- [4] MATHUR, Nav. *Graph Technology for Financial Services: White Paper*. [S.l.]: Neo4j, 2020. Disponível em: <https://neo4j.com/use-cases/financial-services/>. Acesso em: 16 jun. 2025.
- [5] RATHLE, Phillip. *Driving Innovation in Retail with Graph Technology*. [S.l.]: Neo4j, 2021. Disponível em: <https://neo4j.com/use-cases/retail/>. Acesso em: 16 jun. 2025.
- [6] HODLER, Amy. *Advanced Service Assurance with Neo4j*. [S.l.]: Neo4j, 2021. Disponível em: <https://neo4j.com/use-cases/telecom/>. Acesso em: 16 jun. 2025.