

O PAPEL DOS BANCOS DE DADOS EM PROJETOS DE *DATA SCIENCE*: COMO A ESCOLHA DO BANCO DE DADOS IMPACTA A EFICIÊNCIA DAS ANÁLISES

THE ROLE OF DATABASES IN DATA SCIENCE PROJECTS: HOW DATABASE CHOICE IMPACTS THE EFFICIENCY OF ANALYSIS

335

João Manoel da Silva¹, Luiz Eduardo de Carvalho Chaves², Joaquim M. F. Antunes
Neto³

1- Discente do CST em Gestão da Tecnologia da Informação da FATEC Itapira, Monitoria Iniciação em Desenvolvimento Tecnológico e Inovação; 2- Mestre em Engenharia de Produção, Universidade Paulista, UNIP, Brasil. Docente e orientador da FATEC Itapira; 3- Doutor em Biologia Funcional e Molecular, IB, UNICAMP, Campinas, SP. MBA em Gestão de Estratégia Empresarial e Especialista em Tecnologias para a Indústria 4.0 (Faculdade São Luís, Jaboticabal, SP), graduado em Biologia. Docente na FATEC Itapira.

Contato: fercris.du@gmail.com

RESUMO

Este artigo de revisão explora a importância dos bancos de dados em projetos de *Data Science*, destacando como a escolha do sistema de gerenciamento de banco de dados (SGBD) pode impactar significativamente a eficiência das análises. Com o crescimento exponencial de dados provenientes de diversas fontes, a capacidade de armazenamento e processamento se tornou um fator crítico para o sucesso de iniciativas de análise. O artigo examina diferentes tipos de bancos de dados, incluindo relacionais, NoSQL e em coluna, analisando suas características, vantagens e desvantagens em contextos de *Data Science*. A revisão aborda também a relevância da integração de ferramentas e tecnologias que facilitam o fluxo de dados, permitindo que equipes de *Data Science* acessem informações de maneira ágil e eficaz. Além disso, são discutidos desafios de interoperabilidade e a importância da governança e segurança de dados para garantir a conformidade e a proteção das informações. A análise de estudos de caso reais ilustra como escolhas inadequadas de SGBD podem levar a compromissos na qualidade das análises, enquanto decisões bem fundamentadas podem resultar em ganhos significativos em eficiência e precisão. Por fim, o artigo propõe recomendações para a seleção de bancos de dados em projetos de *Data Science*, enfatizando a necessidade de considerar não apenas as características técnicas, mas também as necessidades específicas da organização e dos projetos em questão. A conclusão reafirma que, em um ambiente orientado por dados, a escolha do banco de dados é uma decisão estratégica que pode determinar o sucesso ou fracasso de iniciativas de análise.

Palavras-chave: banco de dados, *data science*, análise de dados, escalabilidade, integração de dados.

ABSTRACT

This review article explores the importance of databases in *Data Science* projects, highlighting how the choice of database management system (DBMS) can significantly impact the efficiency of analytics. With the exponential growth of data from a variety of sources, storage and processing capacity has become a critical factor in the success of analytics initiatives. The article examines different types of databases, including relational, NoSQL, and column, analyzing their characteristics, advantages, and disadvantages in Data Science contexts. The review also addresses the relevance of integrating tools and technologies that facilitate the flow of data, allowing Data Science teams to access information in an agile and effective way. In addition, interoperability challenges and the importance of data governance and security to ensure compliance and information protection are discussed. The analysis of real case studies illustrates how inappropriate DBMS choices can lead to compromises in the quality of analyses, while well-founded decisions can result in significant gains in efficiency and accuracy. Finally, the article proposes recommendations for the selection of databases in Data Science projects, emphasizing the need to consider not only the technical characteristics, but also the specific needs of the organization and the projects in question. The conclusion reaffirms that in a data-driven environment, database choice is a strategic decision that can determine the success or failure of analytics initiatives.

Keywords: database, data science, data analysis, scalability, data integration.

1 INTRODUÇÃO

Nos últimos anos, o campo da *Data Science* tem experimentado um crescimento exponencial, impulsionado pela explosão de dados gerados em diversos setores (METELO; BERNARDINO; PEDROSA, 2021). Com a ascensão das tecnologias digitais, as empresas estão cada vez mais buscando maneiras de extrair informações a partir desses dados. Nesse contexto, a escolha do banco de dados se torna um fator decisivo, pois influencia diretamente a eficiência das análises e a qualidade das informações extraídas (ESCOVEDO; KOSHIYAMA, 2020).

Os bancos de dados desempenham um papel fundamental no gerenciamento e armazenamento de grandes volumes de dados. Eles não apenas permitem a organização dos dados, mas também facilitam o acesso e a manipulação necessária para realizar análises complexas. A diversidade de tipos de bancos de dados disponíveis - desde relacionais a não relacionais - oferece uma gama de opções para os cientistas de dados, cada uma com suas próprias características e capacidades (LANDIM PINTO, 2023).

A seleção do banco de dados adequado pode impactar a performance das consultas, a escalabilidade das aplicações e a facilidade de integração com ferramentas analíticas (MELO JÚNIOR, 2023). Por exemplo, bancos de dados relacionais, como MySQL e PostgreSQL, são frequentemente escolhidos para aplicações que exigem transações complexas e integridade referencial. Por outro lado, bancos de dados NoSQL, como MongoDB e Cassandra, são preferidos para cenários que demandam alta escalabilidade e flexibilidade na estrutura dos dados (ZIMMER, 2023).

Além disso, a arquitetura de dados é outro aspecto relevante que pode influenciar o sucesso de um projeto de *Data Science*. O design do banco de dados, que inclui a normalização, a escolha de índices e a definição de esquemas, pode afetar significativamente a rapidez e a eficiência das análises. Portanto, entender as necessidades específicas do projeto e as características do banco de dados é essencial para otimizar o desempenho (FERREIRA, 2024).

Adicionalmente, a gestão de dados em ambientes de *Data Science* não se limita apenas à escolha do banco. Fatores como segurança, governança de dados e conformidade com regulamentações também são essenciais. Um banco de dados mal gerenciado pode comprometer a integridade dos dados e, conseqüentemente, a validade das análises realizadas (SILVA, 2024b).

Por fim, este artigo de tem como objetivo explorar o papel dos bancos de dados em projetos de *Data Science*, discutindo como a escolha do banco pode impactar a eficiência das análises. Através de uma análise delimitada, pretende-se fornecer informações contextualizadas para profissionais da área de Tecnologia da Informação, auxiliando na tomada de decisões informadas ao selecionar as tecnologias mais adequadas para suas necessidades analíticas.

2 METODOLOGIA

Trata-se de um estudo de revisão bibliográfica narrativa, que visa explorar processos metodológicos que possibilitem compreender o papel dos bancos de dados em projetos de *Data Science*. Esta abordagem permite uma compreensão ampla e detalhada

das principais descobertas, teorias e aplicações práticas relacionadas ao tema, utilizando uma variedade de fontes acadêmicas e científicas para sustentar os argumentos apresentados. A estratégia de busca de literatura envolveu a utilização das bases de dados acadêmicas disponibilizadas no Portal de Periódicos da CAPES, pesquisando palavras-chave relevantes e suas diferentes combinações: banco de dados, *data Science*, análise de dados, escalabilidade, integração de dados. A seleção de estudos incluiu a análise de artigos que explorassem diretamente a aplicação das metodologias de análise de cenários na perspectiva da questão norteadora, para que no final houvesse a proposição de implementação da metodologia viável neste cenário tecnológico, garantindo assim a qualidade e a relevância dos dados obtidos. Cada fonte selecionada foi submetida a uma leitura detalhada, permitindo a extração de informações essenciais, como conceitos-chave, metodologias utilizadas e principais descobertas. Essa abordagem facilitou a síntese dos resultados, identificando padrões emergentes na literatura e lacunas de conhecimento que merecem atenção adicional.

3 RESULTADOS

A escolha do banco de dados é fundamental em projetos de *Data Science*, pois impacta diretamente a eficiência e a qualidade das análises realizadas. Com o aumento dos volumes de dados e a crescente diversidade de fontes, é essencial que um sistema de gerenciamento de banco de dados (SGBD) consiga atender a essas demandas. Este artigo explora como diferentes tipos de bancos de dados—sejam relacionais, NoSQL ou em coluna—afetam o desempenho das análises e, por consequência, os resultados em projetos de *Data Science*.

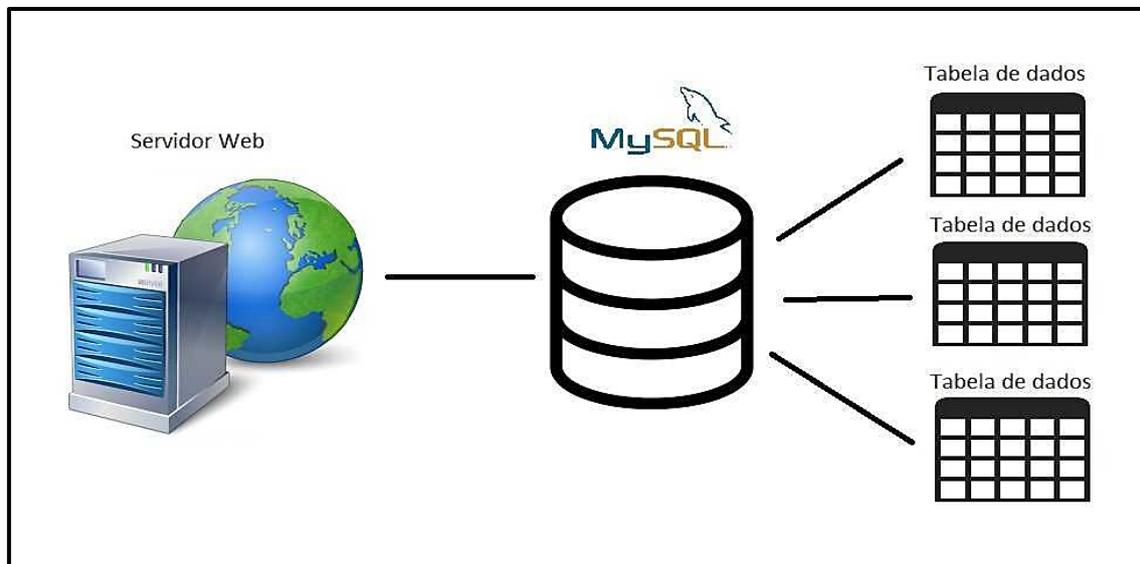
3.1 Fundamentos dos Bancos de Dados

Os fundamentos dos bancos de dados são essenciais para entender como a informação é armazenada, manipulada e recuperada em sistemas computacionais. Um banco de dados é, essencialmente, uma coleção estruturada de dados que pode ser

acessada e gerenciada de maneira eficiente. A estrutura básica de um banco de dados é composta por tabelas, que organizam os dados em linhas e colunas. Cada linha representa um registro único, enquanto cada coluna representa um atributo específico desse registro. Essa organização facilita a realização de operações como consultas, atualizações e exclusões, permitindo que os usuários acessem informações de forma rápida e intuitiva.

A modelagem de dados é um aspecto decisivo no *design* de bancos de dados, pois define como os dados são organizados e relacionados entre si (BRANDT; VIDOTTI, 2024). Modelos como o modelo relacional utilizam tabelas para representar entidades e suas relações, garantindo integridade e consistência (GUERRA, 2022). Além disso, a normalização é uma prática importante que visa minimizar redundâncias e dependências indesejadas entre dados, organizando-os de maneira que cada entidade tenha seu próprio conjunto de atributos. Essa prática não só melhora a eficiência do armazenamento, mas também simplifica as operações de manutenção e atualização dos dados (BRANDT; VIDOTTI, 2024).

A gestão de bancos de dados envolve uma série de tecnologias e práticas que garantem a segurança, integridade e disponibilidade dos dados (MARQUES; CARDOSO, 2021). Sistemas de gerenciamento de bancos de dados (SGBDs) oferecem ferramentas para a criação, manipulação e recuperação de dados, utilizando linguagens como SQL para interação com o usuário (MORANDI; PAIXÃO, 2021). Aspectos como controle de concorrência, recuperação de falhas e *backup* são fundamentais para assegurar que os dados permaneçam consistentes e acessíveis, mesmo diante de falhas de sistema ou de rede. Assim, os fundamentos dos bancos de dados não apenas estabelecem a base para o armazenamento de informações, mas também garantem que essas informações sejam geridas de forma eficaz e segura (SILVA, 2024b). A **Figura 1** esquematiza os elementos componentes de um banco de dados:

Figura 1. Funcionamento de um banco de dados.Fonte: adaptado de HomeHost¹

3.1.1 Definição e tipos de bancos de dados

Os bancos de dados são sistemas que permitem o armazenamento, recuperação e manipulação de dados de forma estruturada. Eles podem ser classificados em várias categorias, sendo as mais comuns:

3.1.1.1 Introdução aos bancos de dados relacionais

Os bancos de dados relacionais (BDRs) surgiram na década de 1970 como uma resposta à crescente necessidade de gerenciar e organizar grandes volumes de dados de forma eficiente (SCARDOELLI; SCOMBATTI PINTO, 2020). Baseados no modelo relacional proposto por Edgar F. Codd, conforme citado acima, esses bancos utilizam tabelas para armazenar dados, onde cada tabela representa uma entidade e as colunas correspondem aos atributos dessa entidade. A estrutura tabular permite que dados sejam organizados de maneira lógica, facilitando consultas e manipulações através da

¹ Disponível em: <https://www.homehost.com.br/blog/tutoriais/mysql/o-que-e-um-banco-de-dados/>

linguagem SQL (*Structured Query Language*). Essa abordagem tem se mostrado eficaz em diversas aplicações, desde sistemas de gestão empresarial até plataformas de *e-commerce* (LEAL, 2024).

3.1.1.2 Estrutura e Normalização

Uma das características fundamentais dos bancos de dados relacionais é a normalização, um processo que visa reduzir a redundância e melhorar a integridade dos dados (SOUZA, 2024). A normalização envolve a organização de dados em tabelas de tal forma que cada tabela armazene informações sobre uma única entidade, minimizando a duplicação de dados. Esse processo é estruturado em várias formas normais, cada uma com suas próprias regras. A normalização não apenas melhora a eficiência do armazenamento, mas também facilita a manutenção e atualização dos dados, garantindo que as alterações em uma tabela não resultem em inconsistências em outras (RAGO; NORBERTO, 2024).

3.1.1.3 Linguagem SQL e Manipulação de Dados

A linguagem SQL é a principal ferramenta utilizada para interagir com bancos de dados relacionais. Ela permite a realização de operações como inserção, atualização, exclusão e consulta de dados. As instruções SQL são divididas em várias categorias, incluindo DDL (*Data Definition Language*) para definição de estruturas de dados, DML (*Data Manipulation Language*) para manipulação de dados e DCL (*Data Control Language*) para controle de acesso (GOUVEA, 2022). Essa versatilidade torna os BDRs altamente eficientes para aplicações que exigem acesso rápido e confiável a dados, permitindo que usuários e desenvolvedores realizem operações complexas com facilidade (BASTOS, 2024).

3.1.1.4 Desempenho e Escalabilidade

Apesar de sua robustez, os bancos de dados relacionais podem enfrentar desafios em termos de desempenho e escalabilidade, especialmente em cenários com grandes volumes de dados ou alta concorrência de acessos (SCARDOELLI; SCOMBATTI PINTO, 2020). A necessidade de garantir a integridade dos dados em transações complexas pode levar a gargalos, especialmente em ambientes com múltiplos usuários. Para amenizar esses problemas, diversas técnicas têm sido desenvolvidas, como particionamento, replicação e otimização de consultas (GALVÃO FILHO; SILVA, 2022). A integração de BDRs com soluções de *caching* e sistemas distribuídos tem se mostrado uma abordagem eficaz para melhorar a performance e escalar horizontalmente.

Os bancos de dados relacionais continuam a desempenhar um papel vital em muitas aplicações modernas, apesar do crescimento de alternativas NoSQL e outras arquiteturas de dados (SANTOS, 2022). Sua capacidade de manter integridade, realizar transações complexas e suportar consultas robustas os torna uma escolha preferencial em muitos setores. Com a evolução contínua da tecnologia, os BDRs estão se adaptando às novas necessidades, incorporando características como suporte a dados não estruturados e integração com ferramentas de análise de dados (SCHREINER, 2021). Assim, os bancos de dados relacionais não apenas permanecem relevantes, mas também estão em constante evolução para atender às demandas dinâmicas do mercado. Portanto, tem-se:

- **Bancos de Dados NoSQL:** Desenvolvidos para lidar com grandes volumes de dados não estruturados ou semiestruturados, oferecem flexibilidade na modelagem dos dados. Dividem-se em quatro categorias principais: chave-valor, documentos, colunas e grafos. Exemplos incluem MongoDB (documento), Cassandra (coluna) e Redis (chave-valor). A escolha de um banco NoSQL geralmente é motivada pela necessidade de escalabilidade horizontal e agilidade no desenvolvimento.
- **Bancos de Dados de Grafos:** Focados na representação de relações complexas, esses bancos utilizam nós e arestas para modelar informações. São especialmente úteis em aplicações que envolvem redes sociais, recomendações e análise de conexões, com o Neo4j como um exemplo proeminente.

3.2 Arquiteturas de Banco de Dados

As arquiteturas de banco de dados referem-se à estrutura organizacional e ao design que suportam o armazenamento, a manipulação e a recuperação de dados em sistemas de gerenciamento de banco de dados (SGBDs). Existem várias arquiteturas, cada uma projetada para atender a diferentes requisitos de desempenho, escalabilidade e complexidade de dados. As arquiteturas mais comuns incluem a arquitetura de banco de dados unificada, a arquitetura em camadas, a arquitetura distribuída e a arquitetura em nuvem, cada uma oferecendo vantagens específicas em contextos variados (BRANDT; VIDOTTI, 2024).

A arquitetura unificada, também conhecida como monolítica, é uma das mais tradicionais e se caracteriza pela centralização dos dados em um único sistema de gerenciamento (BASTOS, 2020). Nesse modelo, todos os componentes do banco de dados, como armazenamento, manipulação e apresentação de dados, são geridos em um único local. Essa abordagem é eficaz para aplicações menores ou de médio porte, onde a simplicidade e a facilidade de manutenção são prioridades. No entanto, à medida que o volume de dados e as necessidades de acesso aumentam, essa arquitetura pode se tornar um gargalo, limitando a escalabilidade e o desempenho (CONCEIÇÃO; SCOMBATTI PINTO, 2021).

Em contrapartida, a arquitetura em camadas divide as funções de gerenciamento de dados em diferentes níveis ou camadas, como apresentação, lógica de negócios e armazenamento. Essa abordagem permite uma maior modularidade e flexibilidade, facilitando a manutenção e a atualização de cada camada de forma independente (REIS, 2024). A separação de responsabilidades melhora a segurança, pois cada camada pode implementar controles de acesso específicos. Essa arquitetura é particularmente útil em sistemas complexos, onde diferentes equipes podem gerenciar componentes distintos sem interferir nas operações globais do banco de dados (CASTRO, 2022).

A arquitetura distribuída envolve a distribuição dos dados em várias localizações físicas, permitindo que diferentes servidores ou nós gerenciem partes do banco de dados. Essa abordagem é vantajosa para aplicações que exigem alta disponibilidade e tolerância

a falhas, pois, mesmo que um nó falhe, os dados podem ser acessados a partir de outros nós. Além disso, a distribuição permite escalar horizontalmente, aumentando a capacidade de armazenamento e processamento de forma mais eficiente. Contudo, essa arquitetura requer técnicas avançadas de gerenciamento de dados e sincronização para garantir a consistência e integridade dos dados entre diferentes locais (CATARINO, 2023).

Por fim, a arquitetura em nuvem tem se tornado cada vez mais popular devido à sua flexibilidade e escalabilidade. Nessa abordagem, os serviços de banco de dados são hospedados em plataformas de nuvem, permitindo que as organizações acessem e gerenciem seus dados sem a necessidade de infraestrutura física própria. Essa arquitetura oferece benefícios significativos, como o provisionamento sob demanda, a redução de custos operacionais e a possibilidade de escalar rapidamente os recursos de acordo com as necessidades do negócio. No entanto, também traz desafios em termos de segurança e controle de dados, exigindo que as organizações implementem estratégias eficazes para proteger suas informações em ambientes de nuvem (CALIXTO, 2024).

A escolha da arquitetura de banco de dados é uma decisão crítica que deve ser baseada nas necessidades específicas da aplicação, na escala do projeto e nas considerações de desempenho. Cada arquitetura tem suas vantagens e desvantagens, e entender essas características é fundamental para garantir que os dados sejam geridos de forma eficaz e segura. À medida que a tecnologia evolui e as demandas do mercado mudam, a arquitetura de banco de dados continuará a se adaptar, proporcionando soluções cada vez mais eficientes para o gerenciamento de dados.

O **Quadro 1** apresenta de forma resumida as características principais dos tipos de bancos de dados:

Quadro 1. Tipos de banco de dados.

Tipo de Banco de Dados	Definição	Exemplos	Características Principais
Relacional	Estruturas de dados organizadas em tabelas com relações.	MySQL, PostgreSQL, Oracle	- Utilizam SQL - Suporte a transações ACID - Dados em linhas e colunas
Não Relacional (NoSQL)	Projetados para grandes volumes de dados não estruturados.	MongoDB, Redis, Cassandra	- Diversos modelos (documento, chave-valor, colunar, grafos) - Escalabilidade horizontal
Orientado a Objetos	Armazenam dados em forma de objetos.	db4o, ObjectDB	- Suporte a herança e encapsulamento - Integração com linguagens OO
Em Nuvem	Operam na nuvem, podendo ser relacionais ou não.	Amazon RDS, Google Firestore	- Escalabilidade automática - Alta disponibilidade - Acesso via APIs
Distribuído	Dados armazenados em múltiplos locais físicos.	Apache Cassandra, Google Spanner	- Alta resiliência - Desempenho em geolocalizações distribuídas
Em Tempo Real	Processam e analisam dados em tempo real.	Apache Kafka, Apache Flink	- Baixa latência - Processamento contínuo de dados
Multi-Modelo	Suportam múltiplos modelos de dados.	ArangoDB, OrientDB	- Flexibilidade para diferentes aplicações - Interações entre modelos

Fonte: elaborado pelos autores.

3.3 Comparação entre Bancos de Dados Tradicionais e Modernos

A comparação entre bancos de dados tradicionais e modernos revela diferenças significativas em termos de estrutura, desempenho e flexibilidade. Os bancos de dados tradicionais, geralmente baseados no modelo relacional, utilizam uma arquitetura rígida em que os dados são organizados em tabelas, e cada tabela possui um esquema definido. Esse modelo é eficaz para aplicações que exigem integridade de dados e transações complexas, como sistemas bancários e de gestão empresarial. No entanto, à medida que as demandas por volume e diversidade de dados aumentam, a rigidez do modelo relacional pode se tornar uma limitação, dificultando a adaptação a novas necessidades de negócios (FERRARAZZO et al., 2024).

Em contraste, os bancos de dados modernos, incluindo sistemas NoSQL e orientados a grafos, oferecem maior flexibilidade na modelagem de dados. Esses bancos são projetados para lidar com dados não estruturados e semiestruturados, permitindo que as organizações armazenem informações de diversas fontes sem a necessidade de um esquema rígido. Essa abordagem é especialmente útil em contextos como redes sociais, onde os relacionamentos entre dados são complexos e em constante mudança. Além disso, a capacidade de escalar horizontalmente em bancos de dados modernos permite que as empresas gerenciem grandes volumes de dados de forma mais eficiente, adaptando-se rapidamente a picos de demanda (SILVA, 2021).

Outra diferença importante entre bancos de dados tradicionais e modernos é a forma como eles lidam com consultas e análise de dados. Enquanto os bancos de dados tradicionais utilizam SQL como a linguagem padrão para consultas, os bancos de dados modernos frequentemente oferecem APIs e outras interfaces que facilitam a interação com os dados. Isso permite que os desenvolvedores integrem facilmente novos tipos de dados e realizem análises em tempo real, o que é crucial em um ambiente de negócios dinâmico. A agilidade proporcionada por essas ferramentas permite que as organizações respondam mais rapidamente às mudanças do mercado e aproveitem novas oportunidades (SILVA DIAS, 2023).

Além disso, a escalabilidade é um fator crítico a ser considerado na comparação entre esses dois tipos de bancos de dados. Os bancos de dados tradicionais costumam escalar verticalmente, o que significa que melhorias de desempenho são alcançadas através do aumento da capacidade do servidor. Esse processo pode ser limitado por questões de custo e por limites físicos de hardware. Por outro lado, os bancos de dados modernos frequentemente adotam uma arquitetura distribuída que permite escalar horizontalmente, adicionando mais servidores para compartilhar a carga. Essa capacidade de escalar de forma mais eficiente torna os bancos de dados modernos uma escolha atrativa para organizações que lidam com grandes volumes de dados (BRANDT; VIDOTTI, 2024).

Por fim, a escolha entre bancos de dados tradicionais e modernos deve ser baseada nas necessidades específicas da aplicação e nos objetivos de negócios. Embora os bancos

de dados tradicionais continuem a ser uma escolha sólida para aplicações que exigem integridade e transações complexas, os bancos de dados modernos oferecem soluções mais flexíveis e escaláveis que se adaptam melhor às demandas atuais de dados. À medida que a tecnologia avança e as organizações buscam maneiras de otimizar o gerenciamento de dados, a comparação entre esses dois tipos de bancos de dados continuará a ser um tema relevante e em evolução.

3.4 Data Science e o Ciclo de Vida dos Dados

O ciclo de vida dos dados em projetos de *Data Science* abrange várias etapas essenciais, desde a coleta até a visualização dos dados. Cada uma dessas etapas é crucial para garantir que as análises sejam eficazes e que os insights obtidos sejam significativos. Os bancos de dados desempenham um papel vital em cada fase desse ciclo, proporcionando as ferramentas e estruturas necessárias para gerenciar os dados de forma eficiente (BARROS et al., 2023).

3.4.1 Etapas do Ciclo de Vida dos Dados

O ciclo de vida dos dados é um conceito fundamental que descreve as diversas etapas pelas quais os dados passam desde a sua criação até o seu arquivamento ou descarte. Esse ciclo pode ser dividido em várias fases, que incluem a coleta, o armazenamento, a utilização, a manutenção e a exclusão dos dados (ARAÚJO; RAMOS; MARINO, 2023; BARROS; RUSCHEL, 2024). A primeira etapa, a coleta, envolve a aquisição de dados de diversas fontes, que podem incluir sistemas internos, sensores, redes sociais ou outras plataformas. A qualidade da coleta é fundamental, pois dados imprecisos ou incompletos podem comprometer as análises subsequentes.

Após a coleta, os dados são armazenados em bancos de dados ou repositórios de dados, onde são organizados de forma a facilitar o acesso e a manipulação. Essa fase é importante para garantir a integridade e a segurança dos dados, bem como para assegurar que eles possam ser facilmente recuperados quando necessário. O armazenamento

adequado envolve a escolha de tecnologias apropriadas, como bancos de dados relacionais ou NoSQL, dependendo das características e dos requisitos dos dados. Durante essa fase, também é essencial considerar aspectos de governança de dados, garantindo que políticas de acesso e privacidade sejam implementadas de acordo com as normas regulatórias.

Na fase de utilização, os dados são analisados e interpretados para gerar insights e apoiar a tomada de decisões. Essa etapa envolve o uso de ferramentas de análise de dados e técnicas estatísticas para extrair valor das informações armazenadas. A manutenção dos dados é uma etapa contínua que envolve atualizações, limpeza e auditorias regulares para garantir que os dados permaneçam precisos e relevantes. Por fim, a exclusão de dados ocorre quando as informações não são mais necessárias ou quando o ciclo de vida dos dados chega ao fim. Essa etapa deve ser realizada de maneira cuidadosa, seguindo políticas de retenção e conformidade legal, para garantir que dados sensíveis sejam eliminados de forma segura e que não haja riscos de violação de privacidade.

O ciclo de vida dos dados pode ser dividido em cinco etapas principais:

Quadro 2. Ciclo de vida dos dados.

Etapa	Descrição	Objetivos	Desafios
1. Coleta	Aquisição de dados de diversas fontes (internas e externas).	Garantir a qualidade e relevância dos dados.	Dados imprecisos ou incompletos.
2. Armazenamento	Organização e armazenamento dos dados em bancos de dados ou repositórios.	Facilitar o acesso e a manipulação dos dados.	Escolha da tecnologia adequada e segurança dos dados.
3. Utilização	Análise e interpretação dos dados para gerar dados e apoiar decisões.	Extrair valor dos dados e informar a tomada de decisão.	Uso de ferramentas de análise e técnicas estatísticas.
4. Manutenção	Atualizações, limpeza e auditorias regulares para assegurar a precisão dos dados.	Garantir a integridade e relevância contínua dos dados.	Recursos necessários para auditorias e atualizações.
5. Exclusão	Descarte seguro de dados que não são mais necessários ou que completaram seu ciclo de vida.	Cumprir políticas de retenção e conformidade legal.	Risco de violação de privacidade e segurança.

Fonte: elaborado pelos autores.

3.4.2 O Papel dos Bancos de Dados em Cada Etapa

Os bancos de dados desempenham um papel fundamental em cada etapa do ciclo de vida dos dados, garantindo que as informações sejam geridas de forma eficiente e segura. Na fase de coleta, os bancos de dados são utilizados para armazenar dados à medida que são adquiridos de diversas fontes. Essa etapa envolve a integração de dados de sistemas internos, dispositivos IoT, redes sociais e outras plataformas. A utilização de tecnologias adequadas de banco de dados desde o início é crucial para garantir que os dados sejam organizados de maneira a facilitar o acesso e a manipulação em etapas subsequentes. A estruturação eficiente dos dados na coleta permite que as informações sejam utilizadas de forma mais eficaz nas fases seguintes.

Na etapa de armazenamento, os bancos de dados garantem a segurança, integridade e acessibilidade das informações. Sistemas de gerenciamento de bancos de dados (SGBDs) oferecem mecanismos robustos para a organização e a proteção dos dados, implementando políticas de controle de acesso e garantindo a consistência dos dados através de transações. Além disso, essa fase inclui a escolha de modelos de dados, como bancos de dados relacionais ou NoSQL, que se alinham com as características dos dados armazenados e as necessidades da organização. A capacidade de escalar e realizar backups regulares é outro aspecto crítico que os bancos de dados facilitam durante esta etapa (VILLAS BÔAS JUNIOR; SILVA; GONÇALVES, 2023).

Durante a utilização dos dados, os bancos de dados desempenham um papel central na análise e na geração de insights. Ferramentas de análise de dados são frequentemente integradas a SGBDs, permitindo que usuários e analistas realizem consultas complexas e visualizações para interpretar as informações armazenadas. A eficiência na execução de consultas e a capacidade de manipular grandes volumes de dados são fundamentais para que as organizações possam tomar decisões informadas. O uso de linguagens como SQL e a implementação de algoritmos de análise avançada são exemplos de como os bancos de dados contribuem para maximizar o valor dos dados disponíveis (NEVES, 2024).

A manutenção dos dados é outra área onde os bancos de dados mostram sua importância. Os SGBDs permitem a realização de operações de atualização, limpeza e auditoria de maneira sistemática e eficiente. Isso é crucial para garantir que os dados permaneçam precisos e relevantes ao longo do tempo. Além disso, ferramentas de monitoramento e relatórios ajudam a identificar problemas de qualidade de dados, permitindo que as organizações implementem medidas corretivas rapidamente. A automação de tarefas de manutenção, como a remoção de dados redundantes e a atualização de registros, também é facilitada por sistemas de banco de dados modernos (MARTINS et al., 2023).

Por fim, na etapa de exclusão, os bancos de dados oferecem mecanismos para a remoção segura de dados que não são mais necessários. A exclusão deve ser realizada de acordo com políticas de retenção e normas legais, especialmente em setores onde a privacidade e a segurança de dados são críticas. Os SGBDs garantem que os dados sejam descartados de forma eficaz, evitando a recuperação não intencional e minimizando os riscos de violação de segurança. Essa funcionalidade é essencial para manter a conformidade regulatória e proteger informações sensíveis ao longo do ciclo de vida dos dados.

Quadro 3. O papel dos bancos de dados em cada etapa do ciclo de vida dos dados.

Etapa	Papel dos Bancos de Dados
1. Coleta	Armazenar dados adquiridos de diversas fontes e integrá-los eficientemente.
2. Armazenamento	Proteger e organizar os dados, garantindo segurança e acessibilidade.
3. Utilização	Facilitar a análise de dados através de consultas complexas e visualizações.
4. Manutenção	Permitir atualizações, limpeza e auditorias sistemáticas para garantir a qualidade.
5. Exclusão	Realizar a remoção segura de dados obsoletos, mantendo conformidade e segurança.

Fonte: elaborado pelos autores.

Essa estrutura resume a função essencial dos bancos de dados em cada etapa do ciclo de vida dos dados, enfatizando sua importância para a gestão eficiente e segura das informações. A interdependência entre as etapas do ciclo de vida dos dados e o papel dos bancos de dados não pode ser subestimada. Uma escolha inadequada de banco de dados pode comprometer a qualidade das análises e a eficácia das decisões. Portanto, é fundamental que profissionais de *Data Science* compreendam as necessidades específicas de cada etapa e selecionem as ferramentas e tecnologias adequadas para garantir uma gestão eficaz dos dados ao longo de todo o processo. Essa abordagem integrada é crucial para maximizar o valor dos dados e transformar informações brutas em insights acionáveis.

3.5 Impacto da Escolha do Banco de Dados na Eficiência das Análises

A escolha do banco de dados impacta a eficiência das análises de diversas maneiras, refletindo diretamente na capacidade de uma organização em extrair insights valiosos de seus dados. Primeiramente, a estrutura e o modelo do banco de dados determinam como os dados são armazenados, acessados e manipulados. Bancos de dados relacionais, por exemplo, são ideais para dados estruturados e permitem consultas complexas, mas podem se tornar ineficientes quando lidam com grandes volumes de dados não estruturados. Em contraste, bancos de dados NoSQL são projetados para escalar horizontalmente e lidar com formatos de dados variados, oferecendo maior flexibilidade e velocidade em operações de leitura e gravação (FRANÇA, 2023).

Tal escolha afeta a latência nas consultas, uma vez que sistemas otimizados para análise, como bancos de dados em coluna, podem realizar consultas mais rápidas em cenários analíticos, permitindo que as equipes de *Data Science* obtenham resultados em tempo real. Isso é essencial em contextos onde decisões ágeis são necessárias, como em marketing ou operações financeiras. Por outro lado, bancos de dados mal dimensionados ou mal configurados podem levar a gargalos, resultando em tempos de resposta mais longos e impactando negativamente a produtividade da equipe.

Por fim, a integração com outras ferramentas e tecnologias também é influenciada pela escolha do banco de dados. A compatibilidade com linguagens de programação, ferramentas de visualização e plataformas de *machine learning* determina quão facilmente os dados podem ser processados e analisados. Bancos de dados que suportam APIs modernas e integrações com serviços de nuvem tendem a facilitar um fluxo de trabalho mais eficiente, permitindo que as equipes acessem dados de forma rápida e utilizem análises avançadas sem complicações adicionais. Portanto, a seleção do banco de dados é uma decisão estratégica que pode maximizar ou limitar a eficiência das análises em projetos de *Data Science* (CEPLUKI JUNIOR; GONÇALVES; MENDES, 2024).

3.5.1 Performance e Escalabilidade

A performance de um banco de dados é medida pelo tempo que leva para executar operações, como consultas e atualizações. Essa performance é vital em cenários onde decisões em tempo real são necessárias (BOENTE et al., 2024). A escolha do banco de dados pode afetar significativamente a latência e a eficiência das operações.

- **Tempo de Resposta:** Bancos de dados relacionais, como MySQL e PostgreSQL, são otimizados para transações complexas e garantem a integridade dos dados, mas podem enfrentar desafios em consultas que envolvem grandes volumes de dados ou operações de junção múltipla. Em contrapartida, bancos NoSQL, como MongoDB e Cassandra, oferecem alta performance em operações de leitura e escrita, especialmente quando lidam com dados não estruturados.
- **Capacidade de Lidar com Grandes Volumes de Dados:** A escalabilidade é a capacidade de um sistema de crescer e gerenciar aumentos no volume de dados sem comprometer a performance. Bancos de dados tradicionais geralmente escalam verticalmente, aumentando os recursos de um único servidor, enquanto soluções NoSQL e de grafos são projetadas para escalar horizontalmente, permitindo a adição de novos nós a um cluster de forma eficiente. Essa abordagem horizontal é especialmente valiosa em ambientes com grandes quantidades de dados em constante crescimento.

3.5.2 Flexibilidade e Agilidade

A flexibilidade e a agilidade são essenciais para adaptar modelos e análises a novas demandas e mudanças no ambiente de negócios. A escolha do banco de dados pode impactar significativamente a capacidade de um time de *Data Science* de responder rapidamente a novas necessidades (FRANÇA, 2023).

- **Modelagem de Dados:** Bancos de dados relacionais exigem um *schema* rígido, o que pode dificultar a adaptação a novas fontes de dados ou alterações na estrutura. Já os bancos NoSQL oferecem uma modelagem de dados mais flexível, permitindo que os cientistas de dados alterem rapidamente o esquema ou adicionem novos tipos de dados sem interrupções significativas.
- **Agilidade no Desenvolvimento:** Em ambientes de desenvolvimento ágeis, onde as iterações são frequentes, a capacidade de fazer alterações rápidas nos dados e nas análises é crucial. Ferramentas que suportam integração contínua e desenvolvimento contínuo (CI/CD) se beneficiam de bancos de dados que oferecem APIs dinâmicas e suporte para dados em tempo real.

3.5.3 Casos de Uso

Diversos casos de uso demonstram como a escolha do banco de dados pode impactar os resultados de projetos de *Data Science*. Aqui estão alguns exemplos práticos:

- **E-commerce:** Uma plataforma de e-commerce pode optar por um banco de dados NoSQL, como MongoDB, para gerenciar catálogos de produtos que mudam frequentemente. A flexibilidade do NoSQL permite a adição rápida de novos atributos aos produtos, melhorando a agilidade do negócio. Além disso, a capacidade de escalar horizontalmente permite lidar com picos de tráfego durante eventos de vendas.
- **Redes Sociais:** Para uma aplicação de rede social, um banco de dados de grafos, como Neo4j, é ideal para modelar as interações complexas entre usuários. Esse tipo de banco permite consultas rápidas sobre conexões e relações, facilitando a análise de dados sociais e recomendações personalizadas em tempo real.
- **Análise Financeira:** Em um ambiente bancário, a integridade e a segurança dos dados são cruciais. Aqui, um banco de dados relacional, como Oracle, pode ser a escolha certa, pois oferece transações seguras e suporte a operações complexas. No

entanto, a capacidade de escalar verticalmente pode ser um desafio em épocas de alta demanda, onde soluções de cache e particionamento podem ser necessárias para manter a performance.

A escolha do banco de dados em projetos de *Data Science* não deve ser subestimada. A performance e a escalabilidade afetam diretamente a experiência do usuário e a eficiência das análises, enquanto a flexibilidade e agilidade são essenciais para se adaptar a um ambiente em constante mudança (SILVA et al., 2024). Casos de uso do mundo real demonstram que as decisões sobre a infraestrutura de dados podem impactar o sucesso de um projeto. Portanto, é fundamental que os profissionais considerem as características específicas de cada banco de dados e suas implicações para as necessidades do projeto ao tomar decisões informadas sobre a gestão de dados.

3.5.4 Integração e Interoperabilidade

A integração de bancos de dados em pipelines de *Data Science* é uma etapa crítica que determina a eficiência e a eficácia das análises (SILVA, 2024b). Tem-se a necessidade de explorar as ferramentas e tecnologias disponíveis para facilitar essa integração, além dos desafios comuns que surgem ao lidar com sistemas variados e as soluções para garantir a interoperabilidade.

3.5.4.1 Desafios de interoperabilidade

A integração de diferentes sistemas e bancos de dados, embora facilitada por várias ferramentas, enfrenta desafios significativos de interoperabilidade. Um dos principais obstáculos é a diversidade de formatos de dados e estruturas utilizadas por diferentes sistemas. Por exemplo, enquanto alguns bancos de dados podem armazenar informações em formatos estruturados, como tabelas relacionais, outros podem utilizar formatos não estruturados, como JSON ou XML. Essa disparidade dificulta a troca de dados, pois é necessário transformar e mapear os dados entre os diferentes formatos, o que pode resultar em perda de informação ou inconsistências (SILVA, 2024b).

Outro desafio importante é a questão da compatibilidade entre diferentes tecnologias e plataformas. Sistemas legados muitas vezes utilizam tecnologias antigas que não se comunicam facilmente com soluções modernas. Isso pode exigir investimentos significativos em atualizações ou adaptações, além de aumentar a complexidade do processo de integração. A falta de padrões comuns para a troca de dados também contribui para esse problema, já que diferentes organizações podem adotar práticas distintas para gerenciar suas informações, dificultando a colaboração e a integração (FERREIRA, 2024).

Além disso, questões de segurança e privacidade se tornam críticas quando se trata de interoperabilidade. A integração de dados entre diferentes sistemas pode expor informações sensíveis a riscos de segurança, especialmente se os protocolos de segurança não forem adequadamente implementados. Isso exige que as organizações adotem medidas robustas de governança de dados, garantindo que os dados sejam transmitidos e armazenados de maneira segura. Assim, a interoperabilidade não só exige soluções técnicas, mas também um compromisso com a conformidade regulatória e a proteção de dados, para assegurar que a integração não comprometa a integridade e a segurança das informações.

- **Diversidade de Formatos de Dados:** Dados podem vir em formatos variados (JSON, XML, CSV, etc.), o que pode complicar a extração e a transformação. A falta de um formato padrão pode levar a inconsistências e erros durante a integração.
- **Diferenças nos Modelos de Dados:** Diferentes bancos de dados podem ter estruturas e esquemas variados, dificultando a harmonização dos dados. Isso é especialmente desafiador quando se trabalha com bancos de dados relacionais e NoSQL.
- **Problemas de Performance:** A integração de dados de múltiplas fontes pode causar lentidão nas operações, especialmente se os sistemas não estiverem otimizados para trabalhar juntos.
- **Segurança e Conformidade:** A comunicação entre diferentes sistemas pode expor dados sensíveis a riscos de segurança. Além disso, as regulamentações, como GDPR e LGPD, impõem requisitos estritos sobre como os dados devem ser manipulados e armazenados.

3.5.4.2 Soluções para garantir a interoperabilidade

Para superar os desafios de interoperabilidade na integração de diferentes sistemas e bancos de dados, é essencial adotar uma abordagem baseada em padrões abertos e protocolos de comunicação amplamente aceitos. O uso de APIs RESTful, por exemplo, permite que diferentes aplicações se comuniquem de maneira padronizada, facilitando a troca de dados entre sistemas diversos. A adoção de formatos de dados como JSON e XML também contribui para a interoperabilidade, pois são amplamente suportados por várias linguagens de programação e plataformas. Ao definir interfaces de comunicação claras e consistentes, as organizações podem reduzir as complexidades associadas à integração e promover uma colaboração mais eficaz entre sistemas (TELLES, 2024).

Outra solução eficaz é a implementação de ferramentas de middleware, que atuam como intermediárias entre diferentes sistemas, facilitando a troca de informações e a tradução de formatos de dados. Essas ferramentas podem simplificar a integração ao encapsular as complexidades das diferentes tecnologias e permitir que as organizações conectem facilmente sistemas legados a novas soluções. Além disso, o uso de plataformas de integração como Apache Camel ou MuleSoft pode proporcionar uma abordagem visual e baseada em fluxos para gerenciar a comunicação entre sistemas, permitindo que equipes técnicas se concentrem na lógica de negócios em vez de se preocupar com os detalhes técnicos da interoperabilidade.

Por fim, a governança de dados e a segurança são aspectos cruciais para garantir a interoperabilidade (SHIMIZU et al., 2024). As organizações devem estabelecer políticas claras sobre a gestão e o compartilhamento de dados, garantindo que todas as partes envolvidas compreendam as normas de conformidade e proteção de dados. A implementação de criptografia para dados em trânsito e em repouso, juntamente com autenticação e autorização robustas, ajuda a mitigar os riscos de segurança durante a integração. Assim, ao combinar padrões abertos, ferramentas de middleware e uma forte governança de dados, as organizações podem garantir uma interoperabilidade eficaz e segura entre seus sistemas e bancos de dados (BRANDT; VIDOTTI, 2024).

- **Uso de Padrões Abertos:** Adotar padrões abertos, como JSON ou XML, pode facilitar a troca de dados entre sistemas. A utilização de esquemas de dados bem definidos também pode ajudar a garantir que os dados sejam interpretados corretamente em diferentes plataformas.
- **Camadas de Abstração:** Implementar uma camada de abstração, como um middleware, pode simplificar a integração. Essa camada pode transformar os dados em um formato comum, independentemente da fonte original, facilitando a comunicação entre diferentes sistemas.
- **Ferramentas de Governança de Dados:** Utilizar ferramentas de governança de dados pode ajudar a garantir a qualidade e a integridade dos dados durante a integração. Essas ferramentas ajudam a monitorar e auditar o fluxo de dados, garantindo que as regulamentações de conformidade sejam seguidas.
- **Testes e Validação:** Implementar uma estratégia de testes robusta para verificar a integridade dos dados após a integração é crucial. Testes automatizados podem ajudar a identificar inconsistências e erros antes que os dados sejam utilizados nas análises.

O **Quadro 4** resume os principais impactos determinados pela escolha de um banco de dados:

Quadro 4. Principais impactos quanto a escolha de um banco de dados.

Critério	Banco de Dados Relacional	Banco de Dados Não Relacional (NoSQL)	Banco de Dados em Nuvem	Banco de Dados em Tempo Real
Desempenho em Consultas	Alta eficiência para consultas complexas com JOINS.	Excelente para consultas simples e escaláveis, mas pode ser menos eficiente para operações complexas.	Desempenho otimizado com escalabilidade, mas depende da configuração.	Muito rápido para consultas em tempo real, ideal para dados em movimento.
Escalabilidade	Geralmente escalável verticalmente, o que pode ser limitante.	Escalabilidade horizontal, permitindo lidar com grandes volumes de dados.	Alta escalabilidade com recursos sob demanda.	Alta escalabilidade para processar fluxos contínuos de dados.
Flexibilidade de Dados	Estrutura rígida, requer esquema pré-definido.	Estruturas flexíveis, permitindo dados não estruturados ou semi-estruturados.	Flexível, mas depende do modelo escolhido (relacional ou não).	Geralmente flexível, dependendo da tecnologia utilizada.
Facilidade de Integração	Boas ferramentas para integração com outras aplicações.	Integração simples, especialmente em arquiteturas de microserviços.	Boa integração com serviços de nuvem e APIs.	Integração eficaz com sistemas de fluxo e eventos.
Manutenção e Administração	Requer administração constante para otimização.	Menor necessidade de manutenção, mas pode exigir atenção em escalas grandes.	Administração gerenciada pelo provedor, reduzindo a carga.	Requer monitoramento constante para garantir performance.
Custo de Implementação	Pode ser alto devido a licenças e hardware.	Geralmente mais econômico em grandes volumes de dados.	Custos variáveis, mas podem ser otimizados com uso eficiente.	Pode ter custos altos dependendo do volume de dados e requisitos de latência.
Suporte a Análises Avançadas	Ferramentas robustas para análises complexas.	Análises em tempo real mais fáceis com alguns modelos.	Ferramentas analíticas integradas disponíveis.	Especializado em análises de dados em tempo real.

Fonte: elaborado pelos autores.

A integração de bancos de dados em pipelines de *Data Science* é uma tarefa complexa, mas essencial para o sucesso das análises. As ferramentas e tecnologias disponíveis oferecem uma ampla gama de opções para facilitar esse processo, mas os desafios de interoperabilidade não devem ser ignorados. Ao adotar boas práticas e soluções adequadas, é possível garantir uma integração eficiente e segura, permitindo que os cientistas de dados se concentrem na extração de insights valiosos a partir dos dados. A compreensão dessas dinâmicas é crucial para maximizar o valor dos dados e melhorar a tomada de decisões informadas nas organizações.

3.6 Governança e Segurança de Dados

A governança de dados é um componente fundamental na gestão eficaz de informações dentro de uma organização, assegurando que os dados sejam tratados de maneira adequada e em conformidade com regulamentações e políticas internas (ARRUDA; LIM, 2024). Um dos principais objetivos da governança de dados é estabelecer diretrizes claras sobre como os dados devem ser coletados, armazenados, acessados e compartilhados. Isso envolve a definição de papéis e responsabilidades para a gestão de dados, a criação de políticas de qualidade de dados e a implementação de práticas de monitoramento. Com uma estrutura de governança robusta, as organizações podem não apenas garantir a integridade e a precisão dos dados, mas também fomentar uma cultura de responsabilidade em relação ao uso das informações (SILVA, 2024a).

A segurança de dados é outra dimensão crítica da governança, especialmente em um cenário onde as ameaças cibernéticas estão em constante evolução. As organizações precisam adotar uma abordagem multifacetada para proteger seus dados, que inclua a implementação de controles de acesso rigorosos, criptografia e autenticação forte. Além disso, é vital realizar auditorias regulares e testes de segurança para identificar vulnerabilidades e garantir que as medidas de proteção estejam atualizadas. A segurança não se limita apenas à proteção contra acessos não autorizados, mas também abrange a prevenção de perdas de dados, vazamentos e outras ameaças que possam comprometer a confidencialidade, integridade e disponibilidade das informações (OLIVEIRA, 2024).

A intersecção entre governança e segurança de dados também destaca a importância de uma cultura organizacional voltada para a conscientização sobre a proteção de dados. Treinamentos regulares e campanhas de sensibilização ajudam a educar os colaboradores sobre as melhores práticas de segurança e as implicações legais do manuseio inadequado de dados. Ao promover um ambiente onde a governança e a segurança são priorizadas, as organizações não apenas protegem suas informações, mas também estabelecem confiança junto a clientes e parceiros. Essa confiança é fundamental em um mercado cada vez mais orientado por dados, onde a reputação de uma organização pode ser seriamente afetada por falhas na proteção de informações sensíveis.

3.6.1 Importância da Governança de Dados

A governança de dados é essencial para garantir que as informações sejam geridas de forma eficaz e estratégica dentro das organizações. Com o aumento exponencial da quantidade de dados disponíveis, a governança proporciona um *framework* que assegura a qualidade, integridade e confiabilidade das informações. Ao estabelecer políticas e processos claros, as organizações podem evitar problemas relacionados a dados imprecisos ou desatualizados, que podem levar a decisões erradas e prejuízos financeiros. Além disso, a governança de dados facilita a identificação e a resolução de problemas, permitindo que as empresas melhorem continuamente a gestão de suas informações (SILVA, 2024a).

Outro aspecto fundamental da governança de dados é a conformidade com regulamentações e normas legais. Em um ambiente cada vez mais regulamentado, as organizações precisam garantir que suas práticas de coleta, armazenamento e uso de dados estejam em conformidade com legislações como a LGPD no Brasil e o GDPR na União Europeia. A falta de conformidade pode resultar em multas significativas e danos à reputação, tornando a governança não apenas uma prática recomendada, mas uma necessidade para a sustentabilidade e a continuidade dos negócios. Com uma estrutura de governança bem definida, as organizações podem mitigar riscos legais e proteger a privacidade dos dados de seus clientes (CAVENAGHI et al., 2024).

Além disso, a governança de dados desempenha um papel fundamental na promoção de uma cultura organizacional centrada em dados. Quando as equipes entendem a importância da qualidade e da gestão de dados, elas tendem a adotar práticas mais rigorosas no manuseio das informações. Isso não apenas melhora a qualidade dos dados, mas também empodera os colaboradores a utilizarem informações precisas e relevantes em suas decisões diárias. Ao integrar a governança de dados nas operações e na estratégia de negócios, as organizações podem transformar dados em ativos valiosos, que impulsionam a inovação e a vantagem competitiva.

360

Em projetos de *Data Science*, a governança é fundamental por várias razões:

- **Qualidade dos Dados:** A governança de dados assegura que os dados utilizados nas análises sejam precisos, consistentes e atualizados. Isso é essencial para a confiança nos insights gerados, uma vez que decisões baseadas em dados imprecisos podem levar a resultados desastrosos.
- **Responsabilidade e Transparência:** Estabelecer políticas claras de governança ajuda a definir responsabilidades sobre os dados. Isso inclui quem pode acessar os dados, como eles devem ser manipulados e quais são os procedimentos para garantir a conformidade. A transparência nas práticas de manejo de dados também é vital para a reputação da organização.
- **Conformidade Regulatória:** Em um ambiente cada vez mais regulado, a governança de dados ajuda as organizações a se manterem em conformidade com legislações como GDPR, LGPD e HIPAA. A falta de conformidade pode resultar em multas significativas e danos à reputação.
- **Facilitação da Colaboração:** Uma governança de dados eficaz promove a colaboração entre equipes, permitindo que diferentes departamentos compartilhem e utilizem dados de maneira eficiente, respeitando diretrizes de segurança e qualidade.

3.6.2 Práticas Recomendadas para Garantir a Segurança e a Conformidade Regulatória

Para garantir a segurança e a conformidade regulatória, é fundamental que as organizações implementem práticas recomendadas que abrangem desde a coleta de dados até o seu armazenamento e compartilhamento. Uma das primeiras etapas é realizar uma avaliação de riscos, identificando quais dados são sensíveis e quais podem estar sujeitos

a regulamentações específicas, como a LGPD ou o GDPR. Essa avaliação deve incluir uma análise das vulnerabilidades e das ameaças potenciais, permitindo que a organização desenvolva uma estratégia de segurança que aborde essas questões de forma proativa. Além disso, é crucial criar um inventário de dados, documentando onde e como as informações são armazenadas, quem tem acesso a elas e como estão sendo utilizadas.

A implementação de controles de acesso rigorosos é outra prática essencial para garantir a segurança dos dados. Isso inclui o uso de autenticação multifator (MFA) para proteger o acesso a sistemas sensíveis e a definição de permissões baseadas em funções, de modo que os colaboradores tenham acesso apenas às informações necessárias para desempenhar suas funções. Também é importante monitorar constantemente os acessos e as atividades nos sistemas, utilizando ferramentas de auditoria e relatórios para detectar comportamentos suspeitos ou não autorizados. A conscientização dos colaboradores sobre a importância da segurança de dados e a realização de treinamentos regulares também contribuem para minimizar os riscos de violações e garantir que todos estejam alinhados com as políticas de segurança da organização.

Por fim, a manutenção de uma política de segurança de dados e conformidade atualizada é essencial para que as organizações permaneçam em conformidade com as regulamentações em constante evolução. Isso inclui revisões regulares das políticas existentes, bem como a adaptação às novas normas e leis à medida que elas surgem. Implementar um programa de governança de dados que inclua processos de revisão, auditoria e documentação pode ajudar a garantir que a organização não apenas atenda às exigências regulatórias, mas também esteja preparada para responder a incidentes de segurança de forma eficaz. Dessa forma, ao adotar essas práticas recomendadas, as organizações podem proteger suas informações sensíveis e construir a confiança dos clientes, ao mesmo tempo em que se mantêm em conformidade com as normas legais aplicáveis.

Quadro 5. Práticas recomendadas para garantir a segurança e a conformidade regulatória.

PRÁTICA RECOMENDADA	DESCRIÇÃO	BENEFÍCIOS
Avaliação de Riscos	Identificação de dados sensíveis e análise de vulnerabilidades e ameaças potenciais.	Desenvolve uma estratégia de segurança proativa.
Inventário de Dados	Documentação sobre onde e como os dados são armazenados, quem tem acesso e como são utilizados.	Facilita a gestão e o controle de dados sensíveis.
Controles de Acesso Rigorosos	Implementação de autenticação multifator (MFA) e permissões baseadas em funções.	Minimiza o risco de acessos não autorizados.
Criptografia de Dados	Aplicação de técnicas de criptografia para proteger dados em trânsito e em repouso.	Garante a confidencialidade e a integridade dos dados.
Monitoramento Contínuo	Uso de ferramentas de auditoria e relatórios para detectar comportamentos suspeitos.	Aumenta a segurança e a capacidade de resposta a incidentes.
Treinamento de Colaboradores	Conscientização sobre a importância da segurança de dados e as políticas da organização.	Reduz riscos de violações devido a erros humanos.
Política de Segurança Atualizada	Manutenção de políticas que se adaptem a novas regulamentações e leis.	Garante conformidade contínua e proteção das informações.

Fonte: elaborado pelos autores.

Este quadro atualizado inclui a criptografia como uma prática essencial para a segurança e a conformidade regulatória, destacando sua importância na proteção dos dados.

A governança e a segurança de dados são pilares essenciais para o sucesso de projetos de *Data Science*. A implementação de práticas de governança robustas não apenas melhora a qualidade e a integridade dos dados, mas também garante que as organizações permaneçam em conformidade com regulamentos rigorosos. Ao adotar práticas recomendadas de segurança, as empresas podem proteger dados sensíveis e mitigar riscos, permitindo que os cientistas de dados se concentrem na análise e na extração de insights valiosos. Em um ambiente de negócios cada vez mais orientado por dados, a governança e a segurança não são apenas uma responsabilidade, mas sim um fator crítico para a inovação e a competitividade.

3.7 Tendências Emergentes em Bancos de Dados e *Data Science*

O campo dos bancos de dados e da *Data Science* está em constante evolução, impulsionado por novas tecnologias e abordagens que transformam a forma como os dados são armazenados, gerenciados e analisados. Este capítulo examina as tendências emergentes, incluindo bancos de dados em memória, soluções em nuvem, e o impacto da inteligência artificial (IA) e do *machine learning* (ML) na evolução dos sistemas de gerenciamento de dados.

3.7.1 Novas Tecnologias e Abordagens

3.7.1.1 Bancos de dados em memória

Os bancos de dados em memória, como SAP HANA e Redis, estão se tornando cada vez mais populares devido à sua capacidade de fornecer alta performance em tempo real. Ao armazenar dados diretamente na memória RAM em vez de em disco, essas soluções permitem acessos extremamente rápidos, tornando-as ideais para aplicações que exigem processamento em tempo real, como análise de grandes volumes de dados e operações de *machine learning* (CATARINO, 2023).

- **Vantagens:** Os bancos de dados em memória são capazes de realizar operações de leitura e escrita muito mais rapidamente do que os bancos de dados tradicionais. Isso é crucial em cenários onde decisões precisam ser tomadas instantaneamente, como em sistemas de recomendação e fraudes em tempo real.
- **Desafios:** Apesar de suas vantagens, a implementação de bancos de dados em memória pode ser custosa, especialmente em relação à capacidade de armazenamento. Além disso, a volatilidade dos dados na RAM exige estratégias de persistência para garantir que as informações não sejam perdidas em caso de falhas.

3.7.1.2 Soluções em nuvem

As soluções em nuvem, como Amazon Web Services (AWS), Google Cloud Platform (GCP) e Microsoft Azure, estão revolucionando a forma como os dados são

gerenciados e analisados. A computação em nuvem oferece escalabilidade, flexibilidade e custo-efetividade, permitindo que as organizações ajustem seus recursos de acordo com a demanda (CEPLUKI JUNIOR; GONÇALVES; MENDES, 2024).

- **Escalabilidade e Acessibilidade:** As soluções em nuvem permitem que as organizações escalem rapidamente seus bancos de dados, ajustando a capacidade de armazenamento e processamento conforme necessário. Isso é especialmente útil em ambientes dinâmicos, onde a carga de trabalho pode variar significativamente.
- **Integração com Ferramentas de Data Science:** Muitos provedores de nuvem oferecem integrações nativas com ferramentas de *Data Science*, como Jupyter Notebooks e plataformas de *machine learning*, facilitando a análise de dados em larga escala.

3.7.1.3 O impacto da inteligência artificial e *machine learning*

A inteligência artificial e o *machine learning* estão desempenhando um papel cada vez mais importante na evolução dos bancos de dados. Esses avanços tecnológicos não apenas melhoram a eficiência das operações de banco de dados, mas também possibilitam novas abordagens para a análise de dados (FIGUEIREDO; CABRAL, 2020).

- **Otimização de Consultas:** Algoritmos de *machine learning* podem ser usados para otimizar consultas e melhorar o desempenho geral dos bancos de dados. Técnicas de aprendizado supervisionado podem ajudar a prever padrões de acesso a dados, permitindo que o sistema ajuste automaticamente índices e caches para maximizar a eficiência.
- **Análise Preditiva:** A combinação de bancos de dados com algoritmos de IA permite a análise preditiva em tempo real, oferecendo insights valiosos que podem informar decisões de negócios. Por exemplo, organizações podem usar dados históricos para prever comportamentos futuros dos clientes, ajustando suas estratégias de marketing de forma proativa.
- **Automação da Governança de Dados:** A IA também está sendo utilizada para automatizar processos de governança de dados, como a classificação de dados e a detecção de anomalias. Isso ajuda as organizações a manterem a qualidade e a segurança dos dados, reduzindo a carga sobre as equipes de TI.

4 CONSIDERAÇÕES FINAIS

As tendências emergentes em bancos de dados e *Data Science* estão moldando o futuro do gerenciamento de dados, tornando-o mais rápido, flexível e inteligente. Com a adoção crescente de bancos de dados em memória e soluções em nuvem, as organizações têm a oportunidade de otimizar suas operações e extrair insights significativos de grandes volumes de dados. Além disso, a integração da inteligência artificial e do *machine learning* não só melhora a eficiência dos bancos de dados, mas também abre novas possibilidades para análises preditivas e automação de processos. À medida que essas tecnologias continuam a evoluir, a capacidade das organizações de se adaptarem e inovarem em suas estratégias de dados será crucial para sua competitividade no mercado.

365

4.1 Conclusão e Perspectivas Futuras

Este artigo explorou a importância da escolha do banco de dados no desempenho e na eficácia das análises em diversas áreas, por meio de estudos de caso reais. Os casos analisados demonstraram que a seleção de um sistema de gerenciamento de banco de dados (SGBD) apropriado pode impactar significativamente os resultados, desde a performance em tempo real até a capacidade de integrar e analisar dados de diferentes fontes.

Os principais achados incluem:

- **Escalabilidade e Performance:** A migração para bancos de dados NoSQL e em coluna proporcionou melhorias consideráveis na performance e na escalabilidade, especialmente em ambientes com grandes volumes de dados.
- **Flexibilidade na Modelagem de Dados:** A adoção de bancos de dados que permitem a inclusão de dados não estruturados possibilitou análises mais abrangentes e precisas, como evidenciado no estudo sobre saúde pública.
- **Decisões Baseadas em Dados:** O uso de tecnologias mais adequadas para a análise de risco de crédito resultou em decisões mais informadas, reduzindo significativamente as taxas de inadimplência.

4.2 Implicações para a Prática

As implicações destes achados são profundas para profissionais que trabalham com análise de dados. É essencial que organizações:

- **Avaliem suas Necessidades:** Antes de escolher um banco de dados, é fundamental entender as necessidades específicas do projeto, incluindo o volume e a variedade de dados.
- **Considerem a Escalabilidade:** A capacidade de um SGBD de escalar com o crescimento dos dados deve ser uma prioridade, garantindo que o sistema não se torne um gargalo.
- **Integrar Fontes de Dados:** A utilização de soluções que permitam a integração de dados de múltiplas fontes é vital para obter insights completos e oportunos.

366

4.3 Recomendações para Futuros Estudos

Para avançar na pesquisa sobre a escolha de bancos de dados e seu impacto nas análises, sugerimos:

- **Comparações Longitudinais:** Realizar estudos que comparem a performance de diferentes SGBDs ao longo do tempo em diversos setores.
- **Análises Multidisciplinares:** Investigar como diferentes disciplinas (como saúde, finanças e marketing) se beneficiam de abordagens específicas de banco de dados.
- **Desenvolvimento de Diretrizes:** Criar frameworks ou diretrizes que ajudem profissionais a selecionar o SGBD mais adequado com base nas características do projeto.
- **Exploração de Novas Tecnologias:** Investigar o impacto de tecnologias emergentes, como bancos de dados baseados em nuvem e inteligência artificial, na eficiência das análises.

A escolha do banco de dados é uma decisão crítica que pode moldar não apenas a eficiência operacional, mas também a capacidade de uma organização de inovar e se adaptar a um ambiente em rápida mudança. Com a crescente complexidade e volume de dados, a atenção a esse aspecto se torna mais relevante do que nunca. A contínua investigação e adaptação das práticas de gerenciamento de dados são essenciais para maximizar o valor que esses dados podem oferecer.

REFERÊNCIAS

ARAÚJO, L. S.; RAMOS, L. C.; MARINO, R. C. C. Estudos sobre *data science*: um enfoque em storytelling. **Revista Interface Tecnológica**, v. 20, n. 2, p. 195-207, 2023.

ARRUDA, C; M; M.; LIMA, P. A. L. Proteção de dados: experiência internacional e o caso brasileiro-relação com a segurança da informação e a governança cibernética. **Relações Internacionais**, n. 82, p. 45-59, 2024.

BARROS, S. T. et al. O ciclo de vida dos dados no contexto da pesquisa em user experience. **Encontros Bibli**, v. 28, p. e83434, 2023.

BARROS, N. K.; RUSCHEL, R. C. Estrutura de referência para a medicação da Avaliação do Ciclo de Vida pela Internet das Coisas. **Gestão & Tecnologia de Projetos**, v. 19, n. 2, p. 105-125, 2024.

BASTOS, J. P. S. **Desenvolvimento Web Orientado a Micro Frontends**. 2020. 245 f. Dissertação (Engenharia de Software) – Instituto Superior de Engenharia do Porto. Porto, Portugal, 2020.

BASTOS, M. P. S. S. **Banco de dados NOSQL vs. banco de dados relacional: uma análise comparativa**, 2024. 12 f. Trabalho de conclusão de curso. (Curso superior de tecnologia em tecnologia da informação). Faculdade de Tecnologia de Assis, Prof. Dr. José Luiz Guimarães. Assis, 2024.

BOENTE, K. P. et al. Modelo Fuzzy de caracterização de aplicações em ambiente de computação na nuvem. **Revista de Gestão e Secretariado**, v. 15, n. 5, p. e3710-e3710, 2024.

BRANDT, M. B. VIDOTTI, S. A. B. G. Arquitetura da informação para processos de negócio e modelagem de banco de dados: aproximações possíveis. **Em Questão**, v. 30, p. e-131304, 2024.

CALIXTO, G. M. **Computação em nuvem e tecnologias emergentes**. São Paulo: Editora SENAC., 2024. São Paulo, 162 p.

CASTRO, L. H. M. **Desenvolvimento de um módulo de help desk para o sistema Sisgera**. 2022. 48 f. Monografia (Graduação em Sistemas de Informação) - Instituto de Ciências Exatas e Aplicadas, Universidade Federal de Ouro Preto, João Monlevade, 2022.

CATARINO, M. H. **Estrutura de banco de dados para big data**. São Paulo: Editora Senac São Paulo, 2023. 145 p.

CAVENAGHI, M. G. C. et al. Lei geral de proteção de dados e a segurança da informação: atuação do gestor de tecnologia da informação. **Prospectus**, v. 6, n. 1, p. 97-115, 2024.

CEPLUKI JUNIOR, D.; GONÇALVES, D. R.; MENDES, J. S. *Machine learning* e visão computacional: análise de uma base de dados utilizando métodos estatísticos. **Conhecimento Interativo**, v. 18, n. 1, p. 327-346, 2024.

CONCEIÇÃO, M. T.; SCOMBATTI PINTO, G. Arquitetura de microsserviços. **Revista Interface Tecnológica**, v. 18, n. 2, p. 53-64, 2021.

ESCOVEDO, T.; KOSHIYAMA, A. **Introdução a Data Science**: Algoritmos de *Machine Learning* e métodos de análise. São Paulo: Casa do Código, 2020. 334 p.

FERRARAZZO, G. A. C. **Construção de back-end de hub de análise de metadados sobre repositórios de dados**. 2023. 47 f. Monografia (Engenharia de Controle e Automação) - Universidade Federal de Santa Catarina, Centro Tecnológico, Departamento de Automação e Sistemas, Curso de Graduação em Engenharia de Controle e Automação. Florianópolis, SC, 2023.

FERREIRA, Poliana N. **Introdução à ciência de dados**. São Paulo: Editora Senac São Paulo, 2024. 320 p.

FIGUEIREDO, Carla Regina Bortolaz; CABRAL, Flávio Garcia. Inteligência artificial: machine learning na Administração Pública. **International Journal of Digital Law**, v. 1, n. 1, p. 79-96, 2020.

FRANÇA, V. L. S. **Integração de dados para a construção de KPIs em uma empresa de varejo**: um estudo comparativo entre uma abordagem distribuída e uma centralizada. 2023. 45 f. TCC (Graduação) - Curso de Sistemas de Informação, Centro de Informática, Universidade Federal de Pernambuco, Recife, 2023.

GALVÃO FILHO, C. R.; SILVA, E. O. Migração de Bases de Dados Relacionais para Banco de Dados NoSQL Utilizando Técnicas de ETL. **Caderno de Estudos em Sistemas de Informação**, v. 7, n. 2, p. 1-21, 2022.

GOUVEA, L. V. **A Tuning de Banco de Dados**. 2022. 37 f. Trabalho de Conclusão de Curso (Graduação em Ciência da Computação) – Anhanguera, Taubaté, 2022.

GUERRA, C. J. S. **Avaliação de desempenho entre banco de dados NoSQL e modelo relacional**. 2022. 48 f. Trabalho de Conclusão de Curso (Bacharelado em Sistemas de Informação) – Unidade Universitária de Itaberaí, Universidade Estadual de Goiás, Itaberaí, GO, 2022.

LANDIM PINTO, E. F. **Classificação de dados quentes e frios em bancos de dados em memória utilizando técnicas de Aprendizado de Máquina**. 2023. 19 f. Trabalho

de Conclusão de Curso (Bacharelado em Sistemas de Informação) - Campus de Crateús, Universidade Federal do Ceará, Crateús, 2023.

LEAL, C. C. A. **Oracle vs. PostgreSQL na perspectiva de administração de usuários e objetos.** 2024. 75 f. Trabalho de Conclusão de Curso (Graduação em Sistemas de Informação) – Universidade Federal Rural da Amazônia, Campus Belém, PA, 2024.

MARQUES, G. F.; CARDOSO, R. A importância da segurança em banco de dados. **Revista Eletrônica da Faculdade Invest de Ciências e Tecnologia**, v. 5, n. 1, p. 13-13, 2021.

MARTINS, A. C. et al. Parametrização de Danos e Reparos para Orçamentação da Manutenção de Pontes de Concreto: Uma Abordagem Conceitual. In: **XIV Congresso Brasileiro de Pontes e Estruturas.** 2023.

MELO JÚNIOR, J. T. **Modelo de banco de dados para gerenciamento energético remoto em microrredes:** estudo de caso da microrrede do laboratório de armazenamento e mobilidade (LAM). 2023. 65 f. TCC (Graduação) - Curso de Engenharia de Controle e Automação, Departamento de Engenharia Elétrica, Centro de Tecnologia e Geociências, Universidade Federal de Pernambuco, Recife, 2023.

METELO, M.; BERNARDINO, J.; PEDROSA, I. Avaliação de Ferramentas Open Source para Data Science usando a Metodologia OSSpal. **Revista Ibérica de Sistemas e Tecnologias de Informação**, n. E46, p. 588-606, 2021.

MORANDI, A.; PAIXÃO, C. Z. F. Estudo comparativo dos sistemas de gerenciamento de bancos de dados oracle e MySQL. **Periódico de Pesquisa e Trabalhos de Conclusão de Curso**, v. 7, p. 6-25, 2021.

NEVES, C. M. F. R. **CampusHubAI:** Uma integração inovadora do aluno à universidade com GenAI.2024. 51 f. Trabalho de Conclusão de Curso (Graduação em Ciências e Tecnologia com Ênfase em Computação Aplicada), Escola de Ciências e Tecnologia, Universidade Federal do Rio Grande do Norte, Natal, 2024.

OLIVEIRA, T. C. O. Governança em Segurança da Informação: por que ela é tão importante para os negócios? **Advances of Knowledge Representation**, v. 1, n. 2, p. 142-159, 2024.

RAGO, R. A. P.; NORBERTO, P. B. **Estrutura de bancos de dados para saúde.** São Paulo: Editora Senac São Paulo, 2024. 147 p.

REIS, M. S. **Desenvolvimento de uma infraestrutura de dados como código e com capacidade analítica para monitoramento de bombas industriais utilizando internet das coisas industrial.** 2024. 94 f. Dissertação (Mestrado em Inovação Tecnológica) - Universidade Federal de Minas Gerais, Programa de Pós-Graduação em Inovação Tecnológica. Belo Horizonte. MG, 2024.

SANTOS, T. **Introdução à tecnologia**. São Paulo: Editora Senac São Paulo, 2022. 123 p.

SCARDOELLI, K. A.; SCOMBATTI PINTO, G. Banco de dados NOSQL: uma alternativa para grandes empresas. **Revista Interface Tecnológica**, v. 17, n. 2, p. 219-230, 2020.

SCHREINER, G. A. **Adaptador: uma abordagem para particionamento de dados híbrida e autônoma em bancos de dados NewSQL**. 2021. 113 f. Tese (doutorado) - Universidade Federal de Santa Catarina, Centro Tecnológico, Programa de Pós-Graduação em Ciência da Computação, Florianópolis, 2021.

SHIMIZU, C. F. et al. Aplicação de governança de dados em solução de software para pesquisa de placas em uma empresa do setor de indústria e comércio de peças automotivas. **Conhecimento Interativo**, v. 17, n. 2, p. 211-244, 2024.

SILVA, M. V. **Governança de dados e cibersegurança**. São Paulo: Editora Senac São Paulo, 2024a. 170 p.

SILVA, L. B. B. **Análise Comparativa de Benchmarks para Banco de Dados NoSQL Orientados a Grafos de Propriedades**. 2021. 92 f. Dissertação (Mestre em Ciência da Computação) - Universidade Federal de São Carlos, Centro de Ciências Exatas e de Tecnologia, Programa De Pós-Graduação Em Ciência Da Computação. São Carlos, SP, 2021.

SILVA, L. N. **Transformando a saúde com Big Data e Inteligência Artificial: uma análise abrangente**. 2024. 47 f. Trabalho de Conclusão de Curso (Graduação) - Curso de Engenharia Biomédica, Departamento de Engenharia Biomédica, Centro de Tecnologia e Geociências, Universidade Federal de Pernambuco, Recife, 2024b.

SILVA, J. R. C. et al. O processo de gestão e a transformação digital nas instituições de ensino superior. **Caderno Pedagógico**, v. 21, n. 9, p. e8106-e8106, 2024.

SILVA DIAS, Ariel. **Bancos de dados não relacionais**. São Paulo: Editora Senac São Paulo, 2023. 154 p.

SOUZA, B. S. **Aplicação da teoria dos grafos na análise de similaridade e complexidade de consultas em bancos de dados relacionais**. 2024. 63f. Dissertação (Mestrado em Engenharia Elétrica e de Computação) - Centro de Tecnologia, Universidade Federal do Rio Grande do Norte, Natal, 2024.

TELLES, André. **Protocolos e interoperabilidade para IoT**. São Paulo: Editora Senac São Paulo, 2024. 126 p.

VILLAS BÔAS JUNIOR, M.; SILVA, A. J. D.; GONÇALVES, L. F. V. Modelos de linguagem avançados, bancos de dados vetoriais e o novo paradigma para analytics. **Projectus**, v. 8, n. 3, p. 66-81, 2023.

ZIMMER, N. R. **Análise Comparativa de Desempenho de um Banco de Dados Relacional em Diferentes Linguagens de Programação**. 2-23. 52 f. Monografia (Bacharelado em Ciência da Computação) - Universidade Federal do Rio Grande do Sul, Instituto de Informática. Porto Alegre, RS, 2023.

Os autores declararam não haver qualquer potencial conflito de interesses referente a este artigo