

Persistência Poliglota: o armazenamento de dados e suas possibilidades

Lucas Souza Silveira Martins¹, João Vítor Florão da Silva², Adriana Paula Zamin Scherer³

¹Faculdade Dom Bosco – Porto Alegre – RS – Brazil

{21213001, 22221012, scherer}@faculdadedombosco.edu.br

Abstract. *The exponential growth in data generation by organizations results in the need to persist data in a way that ensures integrity, consistency, and availability. As data complexity continues to increase, the relational model no longer fully meets storage requirements. To address these new demands, NoSQL databases have emerged as an alternative. In this context, Polyglot Persistence presents itself as a strategy for managing the diversity of data and application requirements in modern systems. This multiplicity of approaches to data storage raises the guiding question of this study: how can we assist in choosing the most appropriate model? To address this question, a literature review was conducted to analyze relational and NoSQL database models, highlighting the advantages and disadvantages of each within a polyglot persistence framework. This study produced a theoretical framework to aid in selecting the most suitable model for each storage scenario.*

Resumo. *O crescimento exponencial da geração de dados pelas organizações resulta na necessidade de persisti-los de modo a garantir sua integridade, consistência e disponibilidade. Em um cenário de crescente complexidade dos dados a serem armazenados, o modelo relacional não mais atende os requisitos de armazenamento em plenitude e para atender a estas novas demandas, os bancos NoSQL surgem como alternativa. Neste contexto, a Persistência Poliglota se apresenta como uma estratégia para lidar com a diversidade de dados e requisitos das aplicações modernas. A partir desta multiplicidade de abordagens para o cenário de armazenamento de dados, emerge a pergunta norteadora deste trabalho: como auxiliar na escolha do modelo mais adequado? Para tanto, realizou-se uma revisão da literatura que analisa os modelos de bancos de dados relacionais e NoSQL, destacando as vantagens e desvantagens de cada um, baseados em um modelo de persistência poliglota. O estudo produziu um arcabouço teórico como forma de auxiliar na decisão do modelo mais adequado para cada situação de armazenamento.*

1. Introdução

O atual contexto tecnológico, marcado por um ritmo elevado no desenvolvimento de soluções inovadoras e pelo aumento expressivo na geração de dados pelas organizações, impõe novas exigências aos requisitos e limitações do modelo de armazenamento de dados relacional. Nesse cenário, emerge a necessidade de modelos de armazenamento de dados que garantem integridade, consistência e disponibilidade de forma escalável e eficiente. Como resposta a essas demandas emergentes, destacam-se as soluções baseadas em bancos de dados NoSQL, que oferecem maior flexibilidade e desempenho, superando as limitações impostas pelo modelo de banco de dados relacional (MOURA, 2022).

Tendo em vista a diversidade de modelos de banco de dados e a volatilidade do mercado com diferentes necessidades de negócios, surge o debate entre pesquisadores, desenvolvedores e profissionais da área de Tecnologia da Informação sobre a substituição do modelo de banco de dados relacional por modelos de bancos de dados NoSQL. Essa necessidade se torna evidente ao observar que, de acordo com Sadalage e Fowler (2013), o banco de dados relacional não apresenta um design voltado para uma estrutura horizontal. Por outro lado, os modelos de banco de dados NoSQL são projetados para escalonamento horizontal, lidando de forma eficiente com grandes volumes de dados.

Identifica-se o debate sobre a substituição de modelos de banco de dados para atender às diferentes necessidades de negócios. Como possível solução, apresenta-se a Persistência Poliglota, que, em linhas gerais, defende a ideia de que diferentes tipos de dados em um sistema podem ser armazenados e recuperados usando tecnologias de banco de dados específicas, conforme as necessidades e características desses dados (FOWLER, 2011).

Adotando a Persistência Poliglota como estratégia para enfrentar os desafios da diversidade de dados a serem armazenados, este trabalho tem como objetivo explorar os diferentes modelos de armazenamento de dados. Assim, examina os modelos de dados, desde o relacional até os NoSQL, incluindo os modelos orientados a documentos, chave-valor, família de colunas e grafos. Ao final, busca fornecer uma compreensão abrangente dos modelos de dados da persistência poliglota e das opções disponíveis para o armazenamento eficiente e flexível de dados.

Para isso, na seção 2, apresentam-se os modelos de dados para a persistência de dados. Na seção 3, discutem-se as principais características dos modelos de armazenamento de dados e uma comparação entre eles. Por fim, na seção 4, expõem-se as considerações finais a partir do estudo realizado.

2. Modelos de armazenamento de dados

Nesta seção, abordam-se os principais modelos de armazenamento de dados com maior relevância no contexto da Tecnologia da Informação. O conceito-chave que permeia essa discussão é o da Persistência Poliglota, que, segundo Fowler (2011), desafia a abordagem convencional de utilizar apenas um tipo de banco de dados para todas as necessidades de armazenamento. De acordo com o autor, a ideia por trás da persistência poliglota é que diferentes partes de um sistema se beneficiam de diferentes modelos de armazenamento, dependendo de suas características e requisitos específicos. Explora-se como essa abordagem diversificada oferece soluções mais eficientes e adaptáveis para uma variedade de contextos de aplicação.

Esta revisão de literatura adota o método ad hoc, que permite uma análise mais flexível e exploratória (SIMON, 1981). Esse método é escolhido por possibilitar ajustes metodológicos ao longo do estudo, favorecendo o desenvolvimento de novos insights conforme surgem novas questões. A revisão analisa os modelos de banco de dados relacional e NoSQL, destacando suas particularidades e adequações para diferentes requisitos de armazenamento de dados.

2.1. Modelo de dados relacional

O modelo relacional de banco de dados organiza as informações por meio de tabelas, nas quais devem ser definidas regras como a chave primária, utilizada para a identificação única de um registro. Além disso, incluem-se as chaves estrangeiras, que mantêm o relacionamento entre uma ou mais tabelas, o tipo de dado armazenado em cada campo, o espaço dedicado em disco para o mesmo e outras restrições, como a chave primária e a definição de aceitar ou não valores nulos (ELMASRI; NAVATHE, 2018).

Os bancos relacionais são reconhecidos por sua estabilidade como tecnologia de persistência de dados, pois, por meio das propriedades ACID (Atomicidade, Consistência, Isolamento e Durabilidade), garantem que o banco de dados se mantenha consistente mesmo diante de falhas e asseguram o cumprimento das regras definidas na criação. Além disso, garantem que, uma vez confirmados, os dados não sejam perdidos, mesmo em caso de falhas no sistema (SILBERSCHATZ; SUDARSHAN, 2020). Esses sistemas desempenham um papel fundamental em aplicações que exigem uma estrutura rigorosa e integridade referencial. E também, oferecem vantagens como escalabilidade e suporte transacional, garantindo operações atomicamente consistentes (SILBERSCHATZ; SUDARSHAN, 2020). Entretanto, conforme Silberschatz e Sudarshan (2020), a rigidez do esquema relacional pode ser um limitador em ambientes onde a estrutura dos dados é dinâmica. Apesar disso, os bancos de dados relacionais têm amplo uso em aplicações que priorizam confiabilidade e consistência, como sistemas financeiros e corporativos.

2.2. Modelo de dados orientado a documentos

O modelo orientado a documentos lida com dados semiestruturados e não estruturados. Esse modelo de dados não inclui semântica para transações, sendo caracterizado pela ausência das propriedades ACID e pela consistência eventual dos dados. Isso significa que, após uma alteração nos dados, não há garantia de que eles estejam disponíveis de forma imediata, ou seja, existe uma janela de tempo até que a atualização seja efetivamente confirmada (MOURA, 2022). A durabilidade é garantida por meio da replicação dos dados, que consiste na disponibilização de cópias idênticas do banco principal em outros servidores. Essa abordagem aumenta a disponibilidade e evita a perda de dados pela redundância. Em outras palavras, os dados permanecem disponíveis para consulta mesmo no caso de falhas em uma das cópias do banco. Contudo, podem não ser a versão mais recente devido à janela de inconsistência (MOURA, 2022; WANZELLER, 2013).

A modelagem de dados nesse modelo ocorre de duas maneiras: a primeira é a abordagem de documentos por referência, na qual os dados são armazenados de forma que um documento faça referência a outro por meio de um campo chave; a segunda é a

abordagem de documentos aninhados, onde as informações de um documento são guardadas dentro de outro (SADALAGE; FOWLER, 2013).

2.3. Modelo de dados chave-valor

O modelo chave-valor armazena os dados em uma estrutura semelhante a um dicionário, onde cada registro possui uma chave única que faz referência a um valor, sendo essa a única forma de acessá-lo (COSTA; VILAIN; MELLO, 2016). Esse tipo de banco de dados é altamente otimizado para leituras e escritas rápidas, mostrando-se particularmente adequado para casos de uso que envolvem grandes volumes de dados e requisitos de baixa latência, conforme discutem Sadalage e Fowler (2013). No entanto, é importante destacar que existem variações dentro desse modelo, dependendo do produto utilizado. Algumas soluções armazenam seus registros com base no conceito de tabelas, itens e atributos, enquanto outras permitem que as tabelas não necessitem de um esquema pré-definido, possibilitando registros com diferentes quantidades e tamanhos de atributos dentro da mesma tabela (SILVA; COELHO; CAMBRAIA; OLIVEIRA, 2019).

Embora esse modelo facilite operações rápidas de recuperação, ele apresenta limitações para consultas mais complexas, como apontam Sadalage e Fowler (2013). Bancos de dados chave-valor são amplamente utilizados em sistemas que exigem respostas rápidas para dados específicos, como em sistemas de cache e e-commerce, onde os dados são acessados frequentemente (SADALAGE; FOWLER, 2013). Essa diferenciação no referenciamento dos valores armazenados também se reflete nas consultas, com algumas soluções adotando linguagem própria, enquanto outras utilizam APIs de terceiros para realização das consultas.

2.4. Modelo de dados orientado a colunas

O modelo orientado a colunas é composto por um keyspace que contém famílias de colunas responsáveis por armazenar coleções ordenadas de linhas, as quais, por sua vez, contêm coleções ordenadas de colunas. Esse modelo de arquitetura se assemelha ao modelo relacional, onde o keyspace corresponde ao banco, as famílias de colunas às tabelas, e as colunas e linhas aos mesmos elementos, respectivamente. Sua principal característica reside nos níveis de consistência configuráveis para operações de leitura e escrita, permitindo ajustar a consistência conforme os requisitos da aplicação (ANDRADE, 2017).

De acordo com Sadalage e Fowler (2013), os dados são armazenados e organizados em colunas, o que torna o modelo ideal para a análise de grandes volumes de dados e consultas analíticas intensivas. Esse tipo de armazenamento possibilita a leitura de colunas específicas sem a necessidade de carregar linhas inteiras, promovendo eficiência significativa em operações de leitura. Além disso, o modelo orientado a colunas otimiza a compressão dos dados, reduzindo o uso de espaço e melhorando o desempenho do banco. Modelos orientados a colunas encontram ampla aplicação em sistemas de big data e análises de dados em tempo real, onde a extração rápida de

informações agregadas atende às demandas por alto desempenho e eficiência (SADALAGE; FOWLER, 2013).

2.5. Modelo de dados orientado a grafos

O modelo orientado a grafos utiliza uma estrutura de dados abstrata para armazenar o contexto de problemas por meio de grafos, empregando vértices e arestas para representar entidades e relacionamentos através de pontos e linhas, respectivamente (URQUIZA JÚNIOR; CURA, 2020). Nessa estrutura, cada vértice representa um nó, que pode possuir um ou mais rótulos e propriedades, utilizadas para otimização de consultas. As arestas representam o relacionamento entre os nós (vértices), sendo cada relacionamento caracterizado por um nome e uma direção, ou seja, um nó de origem e outro de destino (BARROS, 2017)..

Esse modelo é especialmente vantajoso para representar dados com relações complexas e interconectadas, como redes sociais, recomendações e sistemas de gestão de identidade (SADALAGE; FOWLER, 2013). Apresenta um formato particularmente útil em contextos onde a análise de conexões e interdependências é crucial, facilitando a navegação eficiente entre os dados relacionados. Além disso, esse modelo se aplica em domínios onde as relações entre os dados são tão importantes quanto os próprios dados, permitindo desenvolver sistemas que exigem uma compreensão profunda das inter-relações entre as entidades (SADALAGE; FOWLER, 2013).

3. Comparativo dos modelos de dados

Tendo em vista os modelos de armazenamento de dados apresentados na Seção 2, o Quadro 1 - Comparativo de Modelos de Dados apresenta uma breve comparação entre eles, destacando suas principais características relacionadas à consistência, suporte a transações, tolerância a partição, linguagem de consulta e escalabilidade.

Quadro 1 - Comparativo de Modelos de Dados

Modelo	Consistência	Transações	Tolerância a partições	Consulta	Escalabilidade
Relacional	Forte (ACID)	Fortemente consistentes	Não	SQL	Vertical / Limitada
Orientado a Documentos	Eventual	Transações sem as propriedades ACID, mas com alguns recursos transacionais.	Alta	BSON - Linguagem semelhante ao JSON	Horizontal / Alta
Orientado a Chave-valor	Eventual	Sem suporte ou suporte parcial dependendo da implementação	Alta	Linguagem própria ou API REST	Horizontal / Alta

Orientado a Colunas	Eventual	Sem suporte	Alta	CQL - Linguagem semelhante ao SQL	Horizontal / Alta
Orientado a Grafos	Eventual	Suporte limitado	Alta	Linguagem Cypher	Horizontal / Alta

Fonte: GARCIA; SOTTO (2019); WANZELLER (2014); SILVA; COELHO; CAMBRAIA; OLIVEIRA (2019); ROCKENBACH; ANDERLE; GRIEBLER; SOUZA (2018); COSTA; VILAIN; MELLO (2016); ANDRADE (2017); BARROS (2017)

No que tange aos casos de uso, o Quadro 2 - Casos de Uso, exemplifica o contexto no qual cada modelo de armazenamento é mais eficaz, bem como cenários onde o mesmo não é a solução mais adequada.

Quadro 2 - Casos de Uso

	Relacional	Orientado a Documentos	Orientado a Chave-valor	Orientado a Colunas	Orientado a Grafos
Casos de Uso Apropriados	Softwares que demandem das propriedades ACID e que não necessitem de alta escalabilidade	Softwares que manipulam grandes volumes de dados não estruturados ou semi-estruturados, como por exemplo Facebook, Netflix, etc.	Softwares de e-commerce, como por exemplo o e-commerce da Amazon que utiliza o DynamoDB.	Softwares onde a leitura é mais frequente que a escrita, como por exemplo aplicações de agregação e data warehouses.	Relações entre os usuários de redes sociais.
Casos de Uso Não Apropriados	Softwares que manipulam grandes volumes de dados não estruturados e semiestruturados, ou que possuam necessidades de escalabilidade	Softwares que demandam mais consistência do que disponibilidade ou escalabilidade.	Softwares que realizam consultas muito complexas ou análises de dados.	Softwares onde são necessárias transações ACID ou aplicações que possuam uma alta carga de escrita.	Softwares que envolvem apenas operações de leitura simples ou consultas relacionais tradicionais

Fonte: GARCIA; SOTTO (2019); WANZELLER (2014); SILVA; COELHO; CAMBRAIA; OLIVEIRA (2019); ROCKENBACH; ANDERLE; GRIEBLER; SOUZA (2018); COSTA; VILAIN; MELLO (2016); ANDRADE (2017); BARROS (2017)

4. Considerações finais

O presente artigo apresenta uma análise teórica sobre as possibilidades de armazenamento de dados, tanto em modelos tradicionais de bancos de dados quanto em sistemas NoSQL, com uma abordagem baseada na adoção da Persistência Poliglota como resposta à crescente diversidade de necessidades de armazenamento de dados.

O modelo relacional, embora conhecido por sua estabilidade e pela garantia das propriedades ACID, apresenta desafios em ambientes onde a estrutura dos dados é dinâmica ou em cenários que exigem escalabilidade horizontal. Por outro lado, os

modelos NoSQL oferecem soluções alternativas que se adequam melhor a esses cenários específicos.

Dito isso, é importante ressaltar que não existe um modelo de dados universalmente superior; cada um apresenta vantagens e desvantagens que devem ser consideradas de acordo com as necessidades específicas de cada aplicação. A escolha do modelo de armazenamento de dados deve ser feita de forma consciente, levando em conta os requisitos funcionais e não funcionais do sistema, bem como as características dos dados a serem armazenados. Nesse contexto, a adoção da Persistência Poliglota consolida-se como uma solução consistente para lidar com a crescente diversidade de demandas no campo da tecnologia da informação. Ao compreender os diferentes modelos de dados disponíveis e suas respectivas aplicações, é possível tomar decisões mais eficazes no projeto e desenvolvimento de software.

Este estudo contribui para ampliar o conhecimento sobre os modelos de armazenamento de dados da persistência poliglota e suas implicações práticas, impactando todas as camadas na arquitetura e no desenvolvimento de software, tanto para profissionais da área de tecnologia da informação quanto para pesquisadores interessados em explorar novas abordagens para o armazenamento eficiente e flexível de dados em sistemas de software. Embora o presente trabalho seja uma revisão da literatura e traga uma análise dos modelos e abordagens em bancos de dados, possui como limitação a ausência de experimentos práticos para validar as conclusões teóricas nos relatos apontados. Isso ocorre porque entende-se que, na prática, os sistemas de armazenamento de dados podem apresentar comportamentos e desempenhos com variações de acordo com o ambiente, infraestrutura e casos de uso específicos no contexto em que estão inseridos.

Além disso, uma potencial ameaça à validade dos resultados apresentados neste estudo está relacionada a novas demandas por armazenamento de dados que emergem do desenvolvimento contínuo dos modelos de negócio conhecidos até o momento da elaboração deste trabalho. Com o surgimento de cenários mais complexos e novos requisitos, uma análise aprofundada e contínua torna-se necessária, principalmente nas informações sintetizadas no Quadro 2, que reflete o atual panorama de necessidades no armazenamento de dados.

Como próximos passos deste trabalho, será conduzida uma pesquisa para investigar o posicionamento das organizações em relação à adoção da persistência poliglota para o armazenamento de dados. Esse aprendizado permitirá avaliar a possibilidade de utilização de vários modelos de armazenamento em diferentes contextos organizacionais, contrastando-a com a adoção de uma única solução para atender às diversas necessidades de armazenamento.

5. Referências

- ANDRADE, Jorge Luiz. Armazenamento de dados abertos com NoSQL: um estudo de caso com dados do Bolsa Família e NoSQL Cassandra. 2017. 57f. Monografia (Bacharelado em Ciência da Computação) - Universidade de Brasília Instituto de Ciências Exatas Departamento de Ciência da Computação, Brasília, 2022. Disponível em:
https://bdm.unb.br/bitstream/10483/19381/1/2017_JorgeLuizAndrade.pdf
Acesso em: Jun. 2024
- BARROS, Bruna Maria Nunes. Modelagem de dados NOSQL: uma modelagem de banco de dados de grafos para persistência de dados de pesquisa – um estudo de caso do projeto “Oficina Mímesis”. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) - Universidade do Estado do Rio Grande do Norte, 2017. Disponível em:
<https://di.uern.br/tccs/html/ltr/PDF/011001143.pdf> Acesso em: Mar. 2024
- COSTA, Augusto; VILAIN, Patricia; MELLO, Ronaldo. Uma Camada para o Mapeamento de Instruções SQL DML para o Banco de Dados NoSQL Chave-Valor Voldemort. In Brazilian Symposium on Information Systems, 12º, 2016, Florianópolis. ANAIS DO SIMPÓSIO BRASILEIRO DE SISTEMAS DE INFORMAÇÃO (SBSI), 2019. p. 224-231. Disponível em:
<https://sol.sbc.org.br/index.php/sbsi/article/view/5966/5864> Acesso em: Abr. 2024
- ELMASRI, Ramez; NAVATHE, Shamkant. Sistemas de banco de dados. 7.ed. São Paulo: Person, 2018.
- FOWLER, Martin, 2011. Polyglot Persistence. Martin Fowler's Website. Disponível em:
<https://martinfowler.com/bliki/PolyglotPersistence.html>. Acesso em: Abr. 2024.
- GARCIA, Vinicius; SOTTO, Eder. Comparativo entre os modelos de banco de dados relacional e não-relacional. *Interface Tecnológica*, Taquaritinga, v.16, n.2 p.12-24, dez.2019. Disponível em:
<https://revista.fatectq.edu.br/interfacetecnologica/article/view/673/408> Acesso em: Abr. 2024
- URQUIZA JÚNIOR, Luiz Sergio Velasques; CURA, Luis Mariano del Val. Uma Proposta de Mapeamento do Modelo Conceitual Entidade-Relacionamento Estendido para o Modelo de Dados NoSQL de Grafos. In: SIMPÓSIO BRASILEIRO DE BANCO DE DADOS (SBBD), 35. , 2020, Evento Online. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, 2020. p. 241-246. ISSN 2763-8979. DOI: <https://doi.org/10.5753/sbbd.2020.13648>.
- MOURA, Henrique. Análise de desempenho do banco de dados MongoDB quanto ao armazenamento de informações diárias. 2022, Monografia (Bacharelado em Engenharia da Computação) - Universidade Federal do Rio Grande do Sul Instituto de Informática, Porto Alegre, 2022. Disponível em:
<https://lume.ufrgs.br/bitstream/handle/10183/254821/001162347.pdf?sequence=1#:~:text=A%20durabilidade%20%C3%A9%20garantida%20por,bom%20desempenho%2C%20acessibilidade%20e%20escalabilidade>

ROCKENBACH, Dinei; ANDERLE, Nadine; GRIEBLER, Dalvan; SOUZA, Samuel. Estudo Comparativo de Bancos de Dados NoSQL. Disponível em: https://www.researchgate.net/profile/Dinei-Rockenbach/publication/324877371_Estudo_Comparativo_de_Bancos_de_Dados_NoSQL/links/5ae9a99745851588dd8214af/Estudocomparativo-de-Bancos-de-Dados-NoSQL.pdf

Acesso em: Abr. 2024

SADALAGE, Pramod; FOWLER, Martin. NoSQL Essencial: Um guia conciso para o mundo emergente de persistência poliglota. São Paulo: Editora Novatec, 2013.

SILBERSCHATZ, Abraham; KORTH, Henry; SUDARSHAN, S. Sistema de Banco de Dados. Rio de Janeiro: Editora GEN LTC, 2020.

SILVA, Jonathan; COELHO, Gimenes; CAMBRAIA, Lamon; OLIVEIRA, Vinícius. Um estudo comparativo entre bancos de dados nosql: dynamodb e redis. Revista Científic@ Universitas Itajubá v.6, n.1, p.34-41 Maio 2019. Disponível em: http://revista.fepi.br/revista/index.php/revista/article/view/669/pdf_66

Acesso em: Abr. 2024

SIMON, Herbert. As Ciências do Artificial. Lisboa: Editora Almedina, 1981.

WANZELLER, Diogo. Investigando o uso de bancos de dados orientados a documentos para gerenciar informações da administração pública. 2013, 67f. Monografia (Bacharelado em Ciência da Computação) - Universidade de Brasília Instituto de Ciências Exatas Departamento de Ciência da Computação, Brasília, 2013. Disponível em:

https://bdm.unb.br/bitstream/10483/7716/1/2013_DiogoAraujoPachecoWanzeller.pdf

Acesso em: Mar. 2024