RCMOS – Revista Científica Multidisciplinar O Saber. ISSN: 2675-9128. São Paulo-SP.

Ano VII, v.2 2026 | submissão: 15/07/2026 | aceito: 20/07/2026 | publicação: 25/07/2026

Arquiteturas de dados para informação complexa: persistência, estruturas de acesso e preservação de contexto

Data architectures for complex information: persistence, access structures and context preservation

Arquitecturas de datos para información compleja: persistencia, estructuras de acceso y preservación del contexto

Ricardo José Leal dos Santos

Engenheiro Eletrônico, pós-graduado em Análise de Sistemas e mestre em Engenharia da Computação, com área de concentração em Geomática. Sua trajetória profissional concentra-se especialmente em sistemas de dados, administração de bancos de dados e Business Intelligence, combinada a atividades acadêmicas e docência.

RESUMO

A expansão de aplicações analíticas, serviços em nuvem e sistemas de inteligência artificial aumentou a dependência de infraestruturas capazes de preservar dados em diferentes formatos, escalas e contextos. Este artigo analisa a relação entre persistência, estruturas de acesso, modelagem analítica, proveniência e informação geográfica, tomando como problema central a adequação entre a natureza do dado e a arquitetura utilizada para armazená-lo e consultá-lo. A pesquisa é qualitativa, de caráter bibliográfico e documental, baseada em literatura acadêmica rastreável e documentação técnica oficial. A análise mostra que complexidade não decorre apenas do volume: concorrência, temporalidade, padrões de consulta, distribuição física, tipos especializados e necessidade de reconstruir transformações também condicionam decisões arquiteturais. Dados espaciais são empregados como caso representativo por evidenciarem a necessidade de métodos de indexação e operadores compatíveis com relações não escalares. O estudo conclui que ambientes de BI e inteligência artificial dependem de uma cadeia anterior de decisões sobre persistência, qualidade, metadados e proveniência, e que a modernização tecnológica não elimina fundamentos de bancos de dados, apenas os reposiciona em novas camadas de infraestrutura.

Palavras-chave: bancos de dados; arquitetura de dados; persistência; informação geográfica; indexação; Business Intelligence; proveniência.

ABSTRACT

The expansion of analytical applications, cloud services, and artificial intelligence systems has increased dependence on infrastructures capable of preserving data across different formats, scales, and contexts. This article analyzes the relationship among persistence, access structures, analytical modeling, provenance, and geographic information, focusing on the fit between the nature of data and the architecture used to store and query it. The research adopts a qualitative bibliographic and documentary approach based on traceable academic literature and official technical documentation. The analysis shows that complexity is not determined by volume alone: concurrency, temporality, query patterns, physical distribution, specialized data types, and the need to reconstruct transformations also shape architectural decisions. Spatial data are used as a representative case because they make explicit the need for indexing methods and operators compatible with non-scalar relationships. The study concludes that Business Intelligence and artificial intelligence environments depend on a prior chain of decisions concerning persistence, quality, metadata, and provenance, and that technological modernization does not eliminate database fundamentals but relocates them within new infrastructure layers.

Keywords: databases; data architecture; persistence; geographic information; indexing; Business Intelligence; provenance.

Este é um artigo publicado em acesso aberto (Open Access) sob a licença CreativeCommons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.

Ano VII, v.2 2026 | submissão: 15/07/2026 | aceito: 20/07/2026 | publicação: 25/07/2026

1 INTRODUÇÃO

A transformação digital ampliou o volume de informações produzido por sistemas transacionais, dispositivos, plataformas de relacionamento, aplicações móveis e ambientes analíticos. Ao mesmo tempo, a popularização da computação em nuvem e da inteligência artificial elevou a expectativa de que organizações consigam explorar esses dados com menor latência e maior capacidade de processamento. Essa expansão, entretanto, não elimina os problemas clássicos da engenharia de dados. Pelo contrário, torna mais relevante a necessidade de decidir como a informação será representada, persistida, indexada, historizada e recuperada. Stonebraker e Çetintemel (2005) já argumentavam que diferentes classes de aplicações impõem requisitos suficientemente distintos para questionar a ideia de uma única arquitetura de banco adequada a todos os cenários. A arquitetura interna de SGBDs combina armazenamento, processamento de consultas, transações e componentes compartilhados, o que impede reduzir o problema de dados a uma única camada tecnológica (HELLERSTEIN; STONEBRAKER; HAMILTON, 2007).

A complexidade de uma arquitetura não pode ser medida apenas pelo número de registros armazenados. Um conjunto de dados pode apresentar baixo volume e ainda exigir tratamento sofisticado por envolver concorrência, histórico, relações espaciais, proveniência ou consultas especializadas. A documentação do PostgreSQL (2026) evidencia essa diversidade ao oferecer diferentes métodos de índice e mecanismos próprios para controle de concorrência, processamento transacional e armazenamento físico. A presença de múltiplas estruturas em um sistema maduro mostra que desempenho e confiabilidade dependem da relação entre o padrão de uso e a forma como os dados são organizados. Fundamentos de modelagem, processamento de consultas, armazenamento, indexação e transações são tratados como dimensões articuladas da administração de bancos de dados por Ramakrishnan e Gehrke (2003).

Esse problema torna-se ainda mais relevante quando o dado atravessa a fronteira entre operação e análise. Kimball e Ross (2013) demonstram que ambientes de data warehouse exigem decisões específicas de granularidade, modelagem dimensional e tratamento histórico, diferentes das prioridades de um sistema transacional. Glavic (2021), por sua vez, mostra que a proveniência se consolidou como área importante de pesquisa em bancos de dados por permitir explicar origem e derivação da informação. O presente artigo aproxima esses fundamentos da informação geográfica, domínio que torna particularmente visível a relação entre tipo de dado, estrutura de acesso e custo computacional.

Este é um artigo publicado em acesso aberto (Open Access) sob a licença CreativeCommons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.

Ano VII, v.2 2026 | submissão: 15/07/2026 | aceito: 20/07/2026 | publicação: 25/07/2026

2 METODOLOGIA

A pesquisa adota abordagem qualitativa, exploratória e bibliográfica-documental. Foram selecionadas referências acadêmicas reconhecidas nas áreas de bancos de dados, modelagem dimensional, proveniência e gestão de dados, complementadas por documentação técnica oficial do PostgreSQL 18 e pelo Research Data Framework do National Institute of Standards and Technology. O critério de seleção privilegiou fontes com autoria, editora, DOI ou documentação institucional rastreável, de modo a permitir a conferência das afirmações utilizadas no desenvolvimento.

A análise foi organizada em cinco dimensões: persistência e processamento transacional; estruturas de acesso; passagem de sistemas operacionais para ambientes analíticos; metadados e proveniência; e tratamento de informação geográfica. A Geomática não é tratada como campo autônomo em relação ao objeto central, mas como domínio que evidencia problemas de armazenamento e consulta de dados complexos. Essa delimitação é coerente com o objetivo de discutir fundamentos de engenharia de dados, evitando transformar experiência profissional do autor em evidência científica.

O procedimento consistiu em confrontar conceitos consolidados da literatura com mecanismos documentados em sistemas contemporâneos. A modelagem dimensional é discutida a partir de Kimball e Ross (2013); a diversidade de arquiteturas é contextualizada por Stonebraker e Çetintemel (2005); proveniência é fundamentada em Glavic (2021); princípios de gestão e preservação são relacionados ao NIST Research Data Framework (HANISCH et al., 2024); e mecanismos de indexação, transações e armazenamento são verificados na documentação PostgreSQL 18 (POSTGRESQL GLOBAL DEVELOPMENT GROUP, 2026). Não são apresentados resultados experimentais próprios nem generalizações quantitativas que não possam ser rastreadas às fontes.

3 PERSISTÊNCIA, TRANSAÇÕES E CONCORRÊNCIA

Persistência é frequentemente reduzida à ideia de gravar um valor em disco, mas sistemas de bancos de dados precisam garantir propriedades mais amplas. Uma aplicação pode executar diversas alterações que fazem sentido apenas quando confirmadas em conjunto. A abstração de transação permite tratar esse conjunto como unidade lógica, evitando que falhas deixem estados intermediários indevidamente expostos. No PostgreSQL, transações e mecanismos de controle de concorrência são tratados como componentes centrais da consistência operacional (POSTGRESQL GLOBAL DEVELOPMENT GROUP, 2026). A literatura de sistemas de bancos de dados trata gerenciamento de armazenamento e processamento transacional como componentes interdependentes da arquitetura

Este é um artigo publicado em acesso aberto (Open Access) sob a licença CreativeCommons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.

Ano VII, v.2 2026 | submissão: 15/07/2026 | aceito: 20/07/2026 | publicação: 25/07/2026

de um SGBD (HELLERSTEIN; STONEBRAKER; HAMILTON, 2007; RAMAKRISHNAN; GEHRKE, 2003).

A durabilidade de uma transação confirmada também não significa que todas as páginas modificadas tenham sido imediatamente reescritas em sua posição definitiva. Bancos utilizam mecanismos de log para registrar alterações de forma compatível com recuperação após falhas. Essa separação entre confirmação lógica e reorganização física é essencial para equilibrar segurança e desempenho. A arquitetura, portanto, precisa ser analisada como combinação de abstrações oferecidas à aplicação e mecanismos internos que materializam essas garantias.

Concorrência acrescenta outro nível de dificuldade. Vários usuários e processos podem ler e modificar dados simultaneamente, e o sistema precisa impedir que essa sobreposição produza estados incompatíveis com as regras da aplicação. O controle de concorrência multiversão, MVCC, utilizado pelo PostgreSQL, permite que transações trabalhem com visões apropriadas dos dados, reduzindo a necessidade de bloquear leituras e escritas em todos os cenários (POSTGRESQL GLOBAL DEVELOPMENT GROUP, 2026). Em sistemas distribuídos, os mesmos problemas de consistência, concorrência e recuperação passam a coexistir com distribuição, replicação e coordenação entre nós (ÖZSU; VALDURIEZ, 2020).

A existência de versões, entretanto, produz custos de manutenção e armazenamento. Sistemas precisam identificar quais versões ainda podem ser observadas e quais podem ser descartadas ou reutilizadas. Esse ponto demonstra que uma propriedade lógica, como isolamento, possui consequências físicas. Em grandes ambientes, decisões sobre concorrência, duração de transações e manutenção podem influenciar diretamente o comportamento da infraestrutura.

Também é necessário separar durabilidade, backup e histórico. Durabilidade procura assegurar a sobrevivência de uma alteração confirmada diante de determinadas falhas; backup cria cópias destinadas à recuperação; histórico permite reconstruir estados ou eventos anteriores como parte do modelo informacional. Uma organização pode possuir transações duráveis e backups adequados e, ainda assim, não conseguir responder qual era o valor de um atributo em uma data passada se essa temporalidade não tiver sido preservada.

Essa distinção ganha importância em sistemas analíticos e regulados, nos quais a pergunta frequentemente não é apenas qual é o estado atual, mas como ele foi produzido. O NIST Research Data Framework inclui atividades relacionadas a armazenamento, preservação, metadados e gestão ao longo do ciclo de vida, reforçando que a confiabilidade do dado depende de processos que ultrapassam a gravação inicial (HANISCH et al., 2024).

Este é um artigo publicado em acesso aberto (Open Access) sob a licença CreativeCommons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.

Ano VII, v.2 2026 | submissão: 15/07/2026 | aceito: 20/07/2026 | publicação: 25/07/2026

4 ESTRUTURAS DE ACESSO E DADOS COMPLEXOS

Índices existem porque localizar registros por varredura completa pode se tornar inadequado à medida que tabelas crescem ou consultas se tornam mais frequentes. A documentação PostgreSQL 18 apresenta B-tree, Hash, GiST, SP-GiST, GIN e BRIN como métodos de índice incorporados, cada um com propriedades e classes de operação próprias (POSTGRESQL GLOBAL DEVELOPMENT GROUP, 2026). Essa diversidade mostra que indexação não é uma otimização genérica, mas uma escolha associada ao tipo de dado e ao padrão de consulta. A literatura clássica de bancos de dados apresenta estruturas de índice como mecanismos centrais para reduzir o custo de acesso e apoiar diferentes padrões de consulta (RAMAKRISHNAN; GEHRKE, 2003).

B-tree é uma estrutura amplamente utilizada para dados escalares e operações de igualdade e ordenação. O PostgreSQL destaca sua adequação a aplicações concorrentes que precisam indexar dados escalares. Entretanto, a própria documentação observa que dados não escalares podem exigir métodos como GiST, SP-GiST ou GIN. A escolha incorreta pode produzir índices que ocupam espaço e impõem custo de manutenção sem resolver eficientemente a consulta predominante.

GiST, Generalized Search Tree, funciona como uma infraestrutura extensível para esquemas de indexação. A documentação oficial explica que o método pode servir de base para diferentes estratégias e permite que especialistas de domínio definam métodos apropriados a tipos customizados (POSTGRESQL GLOBAL DEVELOPMENT GROUP, 2026). Essa característica é especialmente relevante quando relações entre objetos não podem ser representadas apenas por uma ordem linear.

BRIN segue lógica diferente. Em vez de manter uma entrada convencional para cada linha, trabalha com resumos associados a intervalos de blocos, sendo útil quando valores possuem correlação com a localização física. A economia de tamanho pode ser expressiva em tabelas muito grandes, embora sua eficiência dependa dessa correlação. O exemplo demonstra que a arquitetura física precisa levar em conta distribuição e padrão de acesso, não apenas semântica de alto nível.

Dados complexos também podem incluir texto, arrays, documentos, séries temporais e estruturas espaciais. Cada categoria modifica as perguntas que o sistema precisa responder. A tese de Stonebraker e Çetintemel (2005) sobre a inadequação de uma solução única para todos os workloads permanece pertinente: cargas transacionais, analíticas e especializadas apresentam prioridades distintas de latência, atualização, leitura e organização. A teoria de bancos de dados também demonstra que modelos e linguagens precisam acompanhar diferentes formas de representação e consulta da informação (ABITEBOUL; HULL; VIANU, 1995).

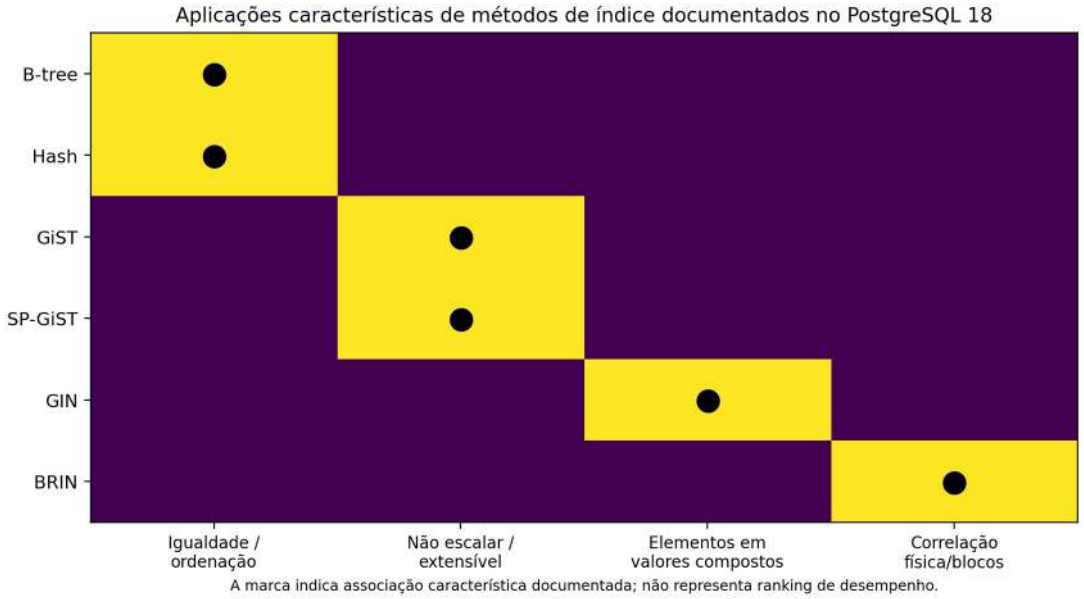
Por isso, desempenho não deve ser tratado como simples consequência de adicionar hardware.

Este é um artigo publicado em acesso aberto (Open Access) sob a licença CreativeCommons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.

Ano VII, v.2 2026 | submissão: 15/07/2026 | aceito: 20/07/2026 | publicação: 25/07/2026

Estruturas de acesso, seletividade, distribuição, concorrência e formato de armazenamento interferem na quantidade de trabalho necessária para responder uma consulta. Uma arquitetura eficiente procura reduzir trabalho desnecessário sem impor custos de manutenção incompatíveis com o padrão de escrita. A Figura 1 sintetiza, sem atribuir pontuações de desempenho, algumas associações características entre métodos de índice e classes de uso documentadas pelo PostgreSQL.

Figura 1 – Aplicações características de métodos de índice documentados no PostgreSQL 18



Fonte: elaboração do autor com base em PostgreSQL Global Development Group (2026), capítulos 11 e 65. A figura é qualitativa e não expressa comparação de desempenho.

5 INFORMAÇÃO GEOGRÁFICA COMO PROBLEMA DE DADOS

Informação geográfica oferece um caso particularmente claro de dado complexo porque incorpora posição, forma e relações espaciais. Um ponto, uma linha ou um polígono não são apenas conjuntos de números. Esses objetos podem participar de perguntas sobre proximidade, interseção, contenção e distribuição territorial. A lógica de consulta, portanto, modifica o modo como a informação deve ser representada e indexada. O trabalho clássico de Guttman (1984) introduziu a R-tree justamente para permitir busca eficiente de objetos espaciais cuja extensão multidimensional não é bem atendida por métodos tradicionais de indexação.

Do ponto de vista de bancos de dados, a relevância da Geomática está justamente nessa exigência de estruturas capazes de persistir e recuperar informação espacial. A documentação PostgreSQL inclui tipos geométricos e métodos de indexação extensíveis, enquanto o ecossistema de bancos espaciais amplia essas capacidades. O problema técnico não é desenhar um mapa, mas permitir que operações

Este é um artigo publicado em acesso aberto (Open Access) sob a licença CreativeCommons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.

Ano VII, v.2 2026 | submissão: 15/07/2026 | aceito: 20/07/2026 | publicação: 25/07/2026

sobre espaço sejam executadas de maneira consistente e eficiente.

Uma consulta por identificador pode localizar uma linha específica; uma consulta espacial pode precisar descobrir todos os objetos que se cruzam com determinada área. Sem mecanismos apropriados, o sistema pode precisar comparar grande quantidade de objetos. Estruturas como GiST oferecem uma base para estratégias de busca adequadas a relações mais complexas que a ordenação escalar (POSTGRESQL GLOBAL DEVELOPMENT GROUP, 2026). R-trees foram propostas para recuperar objetos de acordo com sua localização espacial, oferecendo uma estrutura dinâmica de busca e atualização para aplicações geográficas e de projeto assistido por computador (GUTTMAN, 1984).

Qualidade também assume características próprias. Coordenadas podem estar formalmente preenchidas e ainda representar localização incorreta. Sistemas de referência incompatíveis podem produzir deslocamentos ou comparações inadequadas. Um dado espacial, portanto, pode ser sintaticamente válido e semanticamente impróprio para a análise pretendida. Essa diferença reforça a ideia de que qualidade depende do uso. Batini e Scannapieco (2006) tratam qualidade de dados como um problema multidimensional, envolvendo definições, medição e estratégias de melhoria, o que reforça que validade sintática não é suficiente para garantir adequação ao uso.

Em ambientes analíticos, a dimensão espacial pode ser combinada a dados de negócio, população, infraestrutura ou eventos operacionais. Essa combinação amplia o potencial de análise, mas aumenta a necessidade de metadados que expliquem origem, escala, sistema de referência, período e método de produção. Sem essas informações, a visualização pode transmitir precisão que a fonte não possui.

A relação entre bancos de dados e Geomática ilustra um princípio mais amplo deste artigo: a natureza da informação condiciona a arquitetura. Não se trata de separar dado geográfico dos fundamentos de bancos, mas de mostrar como um domínio especializado exige que esses fundamentos sejam aplicados de maneira compatível com suas propriedades.

6 DO SISTEMA OPERACIONAL AO BUSINESS INTELLIGENCE

Sistemas operacionais e ambientes analíticos atendem necessidades diferentes. O primeiro grupo privilegia registro de operações, atualizações e consistência transacional. O segundo precisa percorrer históricos, consolidar fontes e executar agregações. Essa diferença motivou arquiteturas de data warehouse que reorganizam dados para consulta e tomada de decisão. Chaudhuri e Dayal (1997) mostram que data warehousing e OLAP impõem requisitos distintos daqueles encontrados em aplicações tradicionais de processamento transacional, incluindo extração, limpeza, carga, modelos

Este é um artigo publicado em acesso aberto (Open Access) sob a licença CreativeCommons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.

Ano VII, v.2 2026 | submissão: 15/07/2026 | aceito: 20/07/2026 | publicação: 25/07/2026

multidimensionais e processamento eficiente de consultas.

Kimball e Ross (2013) estruturam a modelagem dimensional em torno de fatos, dimensões, granularidade e processos de negócio. A proposta busca tornar ambientes analíticos compreensíveis e eficientes para consultas, em contraste com modelos operacionais frequentemente normalizados para reduzir redundância e preservar consistência de atualização. Não se trata de declarar um modelo superior ao outro, mas de reconhecer finalidades diferentes.

A passagem entre os ambientes exige extração e transformação. Processos ETL ou ELT podem padronizar valores, integrar fontes e construir estruturas analíticas. Cada transformação, porém, cria uma nova etapa que precisa ser documentada. Se a regra usada para calcular um indicador muda, o resultado pode mudar mesmo quando o fenômeno observado permanece estável. A literatura de data warehousing já identificava ferramentas de back-end para extração, limpeza e carga como componentes fundamentais da arquitetura analítica (CHAUDHURI; DAYAL, 1997).

Esse ponto aproxima Business Intelligence de governança e proveniência. Um painel não é apenas uma camada visual. Ele representa o resultado de uma cadeia que começa em sistemas de origem e passa por filtros, conversões, junções e regras de negócio. A confiabilidade da visualização depende da capacidade de explicar essa cadeia.

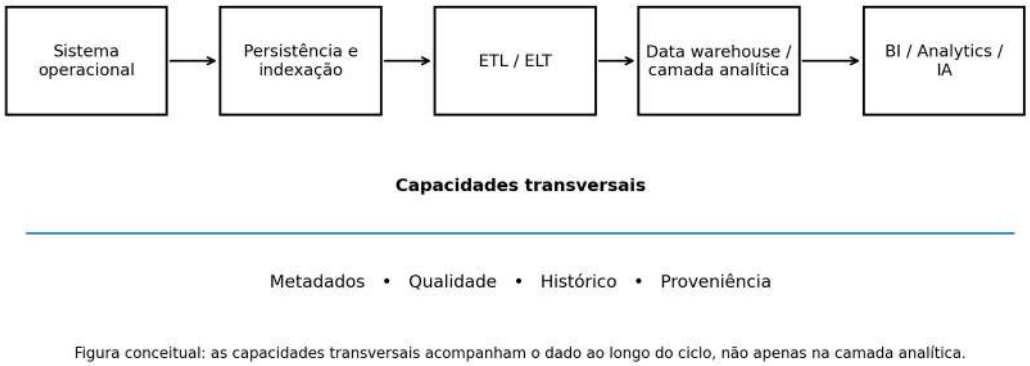
A experiência histórica de profissionais que atuaram tanto em administração de bancos quanto em BI evidencia essa continuidade técnica. Entretanto, no âmbito científico, a relação deve ser sustentada por fundamentos verificáveis. Kimball e Ross (2013) demonstram que desenho dimensional, tratamento de mudanças e definição de granularidade são decisões estruturais, não simples escolhas de ferramenta.

Com a expansão de IA, essa cadeia ganha outro consumidor. Modelos podem utilizar dados provenientes de ambientes operacionais e analíticos, mas não corrigem automaticamente inconsistências, duplicidades ou ausência de contexto. Quanto maior a automação da decisão, maior a importância de compreender como a informação foi produzida antes de chegar ao modelo. A Figura 2 resume essa continuidade entre origem operacional, persistência, transformação, ambiente analítico e consumo, destacando metadados, qualidade, histórico e proveniência como capacidades transversais. Arquiteturas lakehouse foram posteriormente propostas para combinar armazenamento aberto e diretamente acessível com recursos tradicionais de gerenciamento e desempenho de SGBDs analíticos, incluindo transações, versionamento, auditoria, indexação e otimização (ARMBRUST et al., 2021).

Este é um artigo publicado em acesso aberto (Open Access) sob a licença CreativeCommons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.

Ano VII, v.2 2026 | submissão: 15/07/2026 | aceito: 20/07/2026 | publicação: 25/07/2026

Figura 2 – Ciclo conceitual do dado entre o ambiente operacional e o uso analítico



Fonte: elaboração do autor com base em Kimball e Ross (2013), Glavic (2021), Hanisch et al. (2024) e Wilkinson et al. (2016).

7 METADADOS, PROVENIÊNCIA E PRESERVAÇÃO DE CONTEXTO

Glavic (2021) descreve proveniência de dados como área de pesquisa que busca representar e utilizar informações sobre origem e derivação. Em bancos de dados, a proveniência pode ajudar a explicar por que determinado resultado existe e quais entradas contribuíram para sua produção. Essa capacidade é útil para auditoria, depuração, reprodutibilidade e compreensão de pipelines.

Metadados cumprem função complementar. Um campo chamado valor ou status pode ser tecnicamente acessível e ainda permanecer ambíguo para quem não conhece o sistema. Descrição, domínio, responsável, unidade, periodicidade e regras de transformação transformam uma estrutura física em ativo interpretável. O NIST Research Data Framework enfatiza práticas de gestão que acompanham dados ao longo do ciclo de vida (HANISCH et al., 2024). A qualidade e a capacidade de reutilização dependem também de descrições que permitam compreender o significado e as condições de uso dos dados (BATINI; SCANNAPIECO, 2006; WILKINSON et al., 2016).

Os princípios FAIR também reforçam a importância de metadados e reutilização. Wilkinson et al. (2016) propõem que dados sejam localizáveis, acessíveis, interoperáveis e reutilizáveis, com ênfase em descrições e identificadores que permitam sua descoberta e interpretação. Embora formulados no contexto de gestão de dados científicos, os princípios oferecem uma referência útil para discutir preservação de contexto.

Lineage costuma ser empregado para descrever o fluxo entre fontes, transformações e destinos. Em um ambiente analítico, essa representação pode indicar quais tabelas alimentaram determinado

Este é um artigo publicado em acesso aberto (Open Access) sob a licença CreativeCommons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.

Ano VII, v.2 2026 | submissão: 15/07/2026 | aceito: 20/07/2026 | publicação: 25/07/2026

indicador e quais processos modificaram os valores. Proveniência pode alcançar uma visão mais ampla de origem e derivação. Na prática, a granularidade necessária depende do risco e da finalidade.

A preservação de contexto também possui dimensão temporal. Um valor correto hoje pode não representar o estado existente no passado. Data warehouses tratam mudanças dimensionais justamente porque análises históricas precisam decidir se atributos devem refletir o estado atual ou o estado vigente no momento do evento (KIMBALL; ROSS, 2013).

Esses mecanismos tornam-se ainda mais relevantes quando dados alimentam modelos de IA. Reproduzir um resultado pode exigir conhecer versão da fonte, período de extração, transformações e critérios de seleção. Sem esse registro, a organização pode conseguir executar novamente o código e ainda assim não reconstruir o mesmo conjunto informacional.

8 DISCUSSÃO

A análise das fontes permite observar que a modernização das plataformas não elimina os fundamentos de bancos de dados. Serviços gerenciados podem automatizar replicação, backup, escalabilidade e manutenção, mas as aplicações continuam dependendo de decisões sobre consistência, histórico, estruturas de acesso e qualidade. O deslocamento da implementação para a nuvem modifica responsabilidades operacionais, não a necessidade de compreender as propriedades do dado. A descrição arquitetural de Hellerstein, Stonebraker e Hamilton (2007) reforça que escalabilidade e confiabilidade emergem da combinação de vários subsistemas internos, e não de uma única tecnologia.

Também fica evidente que volume não é medida suficiente de complexidade. Dados espaciais demonstram como um conjunto relativamente menor pode exigir operadores e índices especializados. Da mesma forma, uma base analítica pode exigir histórico e proveniência mesmo sem atingir escala extrema. O problema arquitetural nasce da combinação entre natureza da informação, padrão de consulta e garantias necessárias. Essa conclusão é compatível com a literatura de qualidade de dados, que diferencia múltiplas dimensões de adequação, e com a literatura de bancos distribuídos, na qual escala acrescenta problemas de distribuição e coordenação (BATINI; SCANNAPIECO, 2006; ÖZSU; VALDURIEZ, 2020).

A literatura analisada sustenta a rejeição de uma arquitetura universal. Stonebraker e Çetintemel (2005) formularam essa crítica em relação a diferentes classes de aplicações, e a documentação PostgreSQL contemporânea materializa a diversidade por meio de múltiplos métodos de índice e mecanismos internos. A permanência dessa variedade reforça que especialização continua relevante.

Este é um artigo publicado em acesso aberto (Open Access) sob a licença CreativeCommons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.

Ano VII, v.2 2026 | submissão: 15/07/2026 | aceito: 20/07/2026 | publicação: 25/07/2026

Outro resultado da análise é a continuidade entre bancos operacionais e Business Intelligence. Kimball e Ross (2013) mostram que ambientes analíticos possuem modelagem própria, mas dependem de dados produzidos em sistemas anteriores. Assim, BI não substitui administração de dados; ele amplia a cadeia de responsabilidades sobre qualidade, histórico e interpretação.

A proveniência acrescenta uma camada essencial a essa cadeia. Glavic (2021) demonstra que compreender origem e derivação é problema consolidado de pesquisa. Em cenários de IA, essa necessidade tende a se intensificar porque resultados podem depender de conjuntos de dados e transformações que não são visíveis ao usuário final. A rastreabilidade passa a ser componente da confiança.

Por fim, a informação geográfica mostra como uma formação acadêmica em Geomática pode ser integrada ao campo de dados sem deslocar o foco para cartografia ou interoperabilidade. O armazenamento e a consulta espacial são problemas legítimos de bancos de dados. Esse enquadramento é tecnicamente coerente e evita atribuir peso excessivo a experiências periféricas quando o núcleo profissional está em dados e na relação entre prática e academia.

9 CONCLUSÃO

A engenharia de dados precisa ser compreendida como disciplina que preserva condições para que a informação continue utilizável ao atravessar sistemas, tecnologias e finalidades. Persistir um valor é apenas o início desse processo. A arquitetura também precisa lidar com concorrência, recuperação, histórico, indexação, qualidade e interpretação.

As fontes analisadas mostram que diferentes cargas exigem escolhas diferentes. A crítica de Stonebraker e Çetintemel (2005) à ideia de uma solução única permanece compatível com a variedade de estruturas observada em plataformas atuais. Métodos de índice, modelos transacionais e arquiteturas analíticas respondem a problemas específicos.

A informação geográfica reforça essa conclusão. Relações espaciais exigem representação e mecanismos de acesso compatíveis com proximidade, interseção e contenção. O caso demonstra que tipo de dado e tipo de pergunta precisam ser considerados conjuntamente. A R-tree de Guttman (1984) é um marco dessa especialização, pois foi concebida especificamente para busca espacial de objetos multidimensionais.

Business Intelligence amplia o ciclo de vida ao reorganizar dados operacionais para análise. A modelagem dimensional descrita por Kimball e Ross (2013) evidencia que granularidade, fatos, dimensões e histórico são decisões que afetam diretamente a interpretação de indicadores.

Este é um artigo publicado em acesso aberto (Open Access) sob a licença CreativeCommons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.

Ano VII, v.2 2026 | submissão: 15/07/2026 | aceito: 20/07/2026 | publicação: 25/07/2026

Proveniência e metadados acrescentam capacidade de explicação. Glavic (2021) mostra que origem e derivação constituem um campo consolidado de pesquisa, enquanto o NIST Research Data Framework oferece uma visão de gestão ao longo do ciclo de vida (HANISCH et al., 2024). Esses elementos são particularmente importantes quando múltiplas transformações separam a fonte do resultado.

Os princípios FAIR de Wilkinson et al. (2016) também lembram que utilidade depende de condições de descoberta, acesso, interpretação e reutilização. Embora tenham origem no contexto de dados científicos, ajudam a reforçar que armazenar não equivale a tornar um ativo compreensível e reutilizável.

A expansão de nuvem e inteligência artificial aumenta a capacidade computacional disponível, mas não altera esse princípio. Modelos sofisticados continuam dependendo de dados cuja origem, estrutura e qualidade possam ser compreendidas. Automatizar decisões sobre uma base mal estruturada pode apenas acelerar a propagação de inconsistências.

Conclui-se, portanto, que arquiteturas robustas resultam da adequação entre natureza da informação, finalidade de uso e mecanismos de persistência e acesso. A evolução tecnológica muda produtos e escalas, mas preserva a necessidade de fundamentos sólidos. Para ambientes de dados, BI, informação geográfica e IA, a confiança continua sendo construída antes da camada visível da aplicação.

REFERÊNCIAS

ABITEBOUL, Serge; HULL, Richard; VIANU, Victor. Foundations of Databases. Reading, MA: Addison-Wesley, 1995. ISBN 0-201-53771-0.

ARMBRUST, Michael; GHODSI, Ali; XIN, Reynold; ZAHARIA, Matei. Lakehouse: A New Generation of Open Platforms that Unify Data Warehousing and Advanced Analytics. In: CONFERENCE ON INNOVATIVE DATA SYSTEMS RESEARCH, 11., 2021. Proceedings [...]. CIDR, 2021.

BATINI, Carlo; SCANNAPIECO, Monica. Data Quality: Concepts, Methodologies and Techniques. Berlin: Springer, 2006. DOI: 10.1007/3-540-33173-5.

CHAUDHURI, Surajit; DAYAL, Umeshwar. An Overview of Data Warehousing and OLAP Technology. ACM SIGMOD Record, v. 26, n. 1, p. 65-74, 1997. DOI: 10.1145/248603.248616.

GLAVIC, Boris. Data Provenance: Origins, Applications, Algorithms, and Models. Foundations and Trends in Databases, v. 9, n. 3-4, p. 209-441, 2021. DOI: 10.1561/1900000068.

GUTTMAN, Antonin. R-Trees: A Dynamic Index Structure for Spatial Searching. In: ACM SIGMOD INTERNATIONAL CONFERENCE ON MANAGEMENT OF DATA, 1984. Proceedings [...]. New York: ACM, 1984. p. 47-57. DOI: 10.1145/602259.602266.

Este é um artigo publicado em acesso aberto (Open Access) sob a licença CreativeCommons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.

Ano VII, v.2 2026 | submissão: 15/07/2026 | aceito: 20/07/2026 | publicação: 25/07/2026

HANISCH, Robert J. et al. NIST Research Data Framework (RDaF): Version 2.0. Gaithersburg: National Institute of Standards and Technology, 2024. NIST SP 1500-18r2. DOI: 10.6028/NIST.SP.1500-18r2.

HELLERSTEIN, Joseph M.; STONEBRAKER, Michael; HAMILTON, James. Architecture of a Database System. Foundations and Trends in Databases, v. 1, n. 2, p. 141-259, 2007. DOI: 10.1561/1900000002.

KIMBALL, Ralph; ROSS, Margy. The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling. 3. ed. Indianapolis: John Wiley & Sons, 2013. ISBN 978-1-118-53080-1.

ÖZSU, M. Tamer; VALDURIEZ, Patrick. Principles of Distributed Database Systems. 4. ed. Cham: Springer, 2020. DOI: 10.1007/978-3-030-26253-2.

POSTGRESQL GLOBAL DEVELOPMENT GROUP. PostgreSQL 18 Documentation. 2026. Disponível em: https://www.postgresql.org/docs/18/.

RAMAKRISHNAN, Raghu; GEHRKE, Johannes. Database Management Systems. 3. ed. New York: McGraw-Hill, 2003. ISBN 0-07-246563-8.

STONEBRAKER, Michael; ÇETINTEMEL, Uğur. One Size Fits All: An Idea Whose Time Has Come and Gone. In: INTERNATIONAL CONFERENCE ON DATA ENGINEERING, 21., 2005. Proceedings [...]. IEEE, 2005. DOI: 10.1109/ICDE.2005.1.

WILKINSON, Mark D. et al. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, v. 3, art. 160018, 2016. DOI: 10.1038/sdata.2016.18.

Este é um artigo publicado em acesso aberto (Open Access) sob a licença CreativeCommons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.