



UNIVERSIDADE FEDERAL RURAL DO SEMI-ÁRIDO
PRÓ-REITORIA DE GRADUAÇÃO
DEPARTAMENTO DE ENGENHARIAS E TECNOLOGIA
CURSO BACHARELADO INTERDISCIPLINAR EM TECNOLOGIA DA INFORMAÇÃO

ROBERTO FERNANDES ROCHA

ESQUEMA: PROPOSTA DE UMA SOLUÇÃO COMPUTACIONAL COM APOIO DE
INTELIGÊNCIA ARTIFICIAL PARA DIAGNÓSTICO ESTRUTURAL DE BANCOS DE
DADOS RELACIONAIS

PAU DOS FERROS - RN

2026

ROBERTO FERNANDES ROCHA

ESQUEMIA: PROPOSTA DE UMA SOLUÇÃO COMPUTACIONAL COM APOIO DE
INTELIGÊNCIA ARTIFICIAL PARA DIAGNÓSTICO ESTRUTURAL DE BANCOS DE
DADOS RELACIONAIS

Monografia apresentada à Universidade Federal
Rural do Semi-Árido como requisito para
obtenção do título de Bacharel Interdisciplinar
em Tecnologia da Informação.

Orientador: Bruno Borges da Silva, Prof.
Esp.

PAU DOS FERROS - RN

2026

© Todos os direitos estão reservados a Universidade Federal Rural do Semi-Árido. O conteúdo desta obra é de inteira responsabilidade do (a) autor (a), sendo o mesmo, passível de sanções administrativas ou penais, caso sejam infringidas as leis que regulamentam a Propriedade Intelectual, respectivamente, Patentes: Lei nº 9.279/1996 e Direitos Autorais: Lei nº 9.610/1998. O conteúdo desta obra tomar-se-á de domínio público após a data de defesa e homologação da sua respectiva ata. A mesma poderá servir de base literária para novas pesquisas, desde que a obra e seu (a) respectivo (a) autor (a) sejam devidamente citados e mencionados os seus créditos bibliográficos.

R672e Rocha, Roberto Fernandes.
Esquemia: proposta de uma solução computacional
com apoio de inteligência artificial para
diagnóstico estrutural de bancos de dados
relacionais / Roberto Fernandes Rocha. - 2026.
93 f. : il.

Orientador: Bruno Borges da Silva.
Monografia (graduação) - Universidade Federal
Rural do Semi-árido, Curso de Tecnologia da
Informação, 2026.

1. Bancos de dados relacionais. 2. Metadados.
3. Inteligência artificial. 4. Modelos de
linguagem de grande porte. 5. Engenharia de
dados. I. Silva, Bruno Borges da, orient. II.
Título.

Ficha catalográfica elaborada por sistema gerador automático em conformidade
com AACR2 e os dados fornecidos pelo(a) autor(a).

Biblioteca Campus Pau dos Ferros
Bibliotecária: Erlanda Maria Lopes da Silva
CRB: 15/966

O serviço de Geração Automática de Ficha Catalográfica para Trabalhos de Conclusão de Curso (TCC's) foi desenvolvido pelo Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo (USP) e gentilmente cedido para o Sistema de Bibliotecas da Universidade Federal Rural do Semi-Árido (SISBI-UFERSA), sendo customizado pela Superintendência de Tecnologia da Informação e Comunicação (SUTIC) sob orientação dos bibliotecários da instituição para ser adaptado às necessidades dos alunos dos Cursos de Graduação e Programas de Pós-Graduação da Universidade.

ROBERTO FERNANDES ROCHA

ESQUEMIA: PROPOSTA DE UMA SOLUÇÃO COMPUTACIONAL COM APOIO DE
INTELIGÊNCIA ARTIFICIAL PARA DIAGNÓSTICO ESTRUTURAL DE BANCOS DE
DADOS RELACIONAIS

Monografia apresentada à Universidade Federal
Rural do Semi-Árido como requisito para
obtenção do título de Bacharel Interdisciplinar
em Tecnologia da Informação.

Defendida em: 06 / 07 / 2026

BANCA EXAMINADORA



Documento assinado digitalmente

BRUNO BORGES DA SILVA

Data: 11/07/2026 11:29:54-0300

Verifique em <https://validar.iti.gov.br>

Bruno Borges da Silva, Prof. Esp.
Presidente



Documento assinado digitalmente

REUDISMAM ROLIM DE SOUSA

Data: 11/07/2026 11:12:16-0300

Verifique em <https://validar.iti.gov.br>

Reudismam Rolim de Sousa, Prof. Dr.
Membro Examinador



Documento assinado digitalmente

GEORGE FELIPE FERNANDES VIEIRA

Data: 11/07/2026 10:06:03-0300

Verifique em <https://validar.iti.gov.br>

George Felipe Fernandes Vieira, Prof. Bel.
Membro Examinador

DEDICATÓRIA

*Dedico este trabalho à minha querida e guerreira mãe, que sempre esteve ao meu lado, fortalecendo meu caráter e me ensinando, pelo exemplo, a ser um homem melhor a cada dia. Dedico também aos meus amigos **ufersianos**, companheiros de jornada e fonte de alegria nos momentos mais difíceis (presentes).*

AGRADECIMENTOS

Primeiramente, agradeço a Deus, que me guiou e nunca me desamparou ao longo desta jornada acadêmica.

Agradeço à minha família, em especial à minha mãe, Antonia Risonelia Rocha Fernandes, que, ao longo de toda esta trajetória, esteve incondicionalmente ao meu lado, sendo meu maior suporte, minha fonte de força e meu maior exemplo de coragem e perseverança. Seus conselhos, seus sacrifícios e sua confiança em mim foram fundamentais para que eu continuasse firme, mesmo diante das dificuldades. Ao meu pai, Cleilton Fernandes Filgueira, por ter me apresentado, ainda cedo, ao universo da tecnologia, ao me colocar diante de um computador pela primeira vez e, posteriormente, por ter me dado meu primeiro notebook. Aquele equipamento representou muito mais do que uma ferramenta: foi nele que iniciei e desenvolvi minha curiosidade pela área e dei os primeiros passos de uma trajetória que hoje culmina neste trabalho. À minha irmã, Isabella Fernandes Rocha, pelo carinho, apoio e presença ao longo desta caminhada, e ao meu tio, Francisco Cleomar Rocha Silva, que sempre me deu suporte quando precisei e acreditou que esta conquista seria possível.

Ao grande mestre, amigo e sensei Emanuel da Rocha Arruda, por sempre ter me escutado, apoiado e orientado nos momentos em que mais precisei.

Agradeço ao meu orientador, professor Bruno Borges da Silva, por ter aceitado o desafio de orientar este trabalho, oferecendo auxílio, disponibilidade e contribuições importantes nos momentos de dúvida. Estendo meus agradecimentos aos professores do curso de Tecnologia da Informação, que, por meio de seus ensinamentos, contribuíram diretamente para minha formação acadêmica e profissional, permitindo que eu chegasse à conclusão deste trabalho. Agradeço também aos servidores Railson Barreto e Lucimar Queiroz, da UFERSA, pelo trabalho de excelência desenvolvido e por sempre estarem dispostos a auxiliar não apenas a mim, mas também a toda a comunidade ufersiana ao longo do curso.

Por fim, agradeço aos amigos que a universidade me presenteou: Enzo Morgan, Renato Alves, Matheus Irlan e Vinicius Anacleto, amigos, irmãos e fiéis companheiros. Agradeço também aos demais amigos ufersianos que levarei para a vida, por tornarem esta caminhada mais leve, especial e inesquecível. Aos amigos que ficaram pelo caminho, deixo igualmente minha gratidão, pois cada convivência, conversa, apoio e experiência compartilhada contribuiu, de alguma forma, para a construção da pessoa que sou hoje e para a conclusão desta etapa tão importante da minha vida.

EPÍGRAFE

*“Mais do que máquinas, precisamos de
humanidade.”*

(Charlie Chaplin, O Grande Ditador, 1940.)

RESUMO

O presente trabalho tem como tema a análise estrutural automatizada de bancos de dados relacionais com apoio de inteligência artificial, tendo como foco a proposta do Esquemia, uma solução computacional voltada ao diagnóstico estrutural de bases relacionais pouco documentadas ou de difícil compreensão. A pesquisa parte da constatação de que profissionais da área de dados frequentemente enfrentam dificuldades para compreender bases desconhecidas, especialmente quando não há documentação técnica adequada, quando os relacionamentos entre tabelas não estão claramente definidos ou quando a nomenclatura dos elementos do banco não é suficientemente descritiva. Nesse contexto, propõe-se uma arquitetura conceitual capaz de extrair e organizar metadados estruturais, como tabelas, atributos, chaves primárias, chaves estrangeiras e relacionamentos, apresentando essas informações por meio de representações visuais, grafos e recursos de apoio à interpretação. A proposta também prevê a utilização de modelos de linguagem de grande porte para auxiliar na geração de descrições automáticas, na sugestão de significados para atributos e na identificação de possíveis padrões ou inconsistências estruturais. Para fundamentar a relevância do problema, foi realizado um levantamento com profissionais e estudantes da área de Tecnologia da Informação, cujos resultados indicaram dificuldades recorrentes na compreensão de bancos de dados pouco documentados e receptividade à solução proposta. Destaca-se que o escopo deste trabalho limita-se à proposição e modelagem da solução, não contemplando sua implementação prática ou validação experimental. Como contribuição, o estudo apresenta uma arquitetura conceitual voltada ao apoio à redução do esforço inicial de exploração de bases relacionais, à documentação de metadados e à compreensão estrutural de ambientes de dados complexos.

Palavras-chave: bancos de dados relacionais; metadados; inteligência artificial; modelos de linguagem de grande porte; engenharia de dados.

ABSTRACT

This work has as its theme the automated structural analysis of relational databases with the support of artificial intelligence, focusing on the proposal of Esquemia, a computational solution aimed at the structural diagnosis of poorly documented or difficult-to-understand relational databases. The research is based on the observation that data professionals often face difficulties in understanding unknown databases, especially when adequate technical documentation is unavailable, when relationships between tables are not clearly defined, or when the names of database elements are not sufficiently descriptive. In this context, a conceptual architecture is proposed to extract and organize structural metadata, such as tables, attributes, primary keys, foreign keys, and relationships, presenting this information through visual representations, graphs, and interpretation support resources. The proposal also includes the use of large language models to assist in the generation of automatic descriptions, the suggestion of meanings for attributes, and the identification of possible structural patterns or inconsistencies. To support the relevance of the problem, a survey was conducted with professionals and students in the field of Information Technology, whose results indicated recurring difficulties in understanding poorly documented databases and receptiveness to the proposed solution. It is important to emphasize that the scope of this work is limited to the proposal and modeling of the solution, without including its practical implementation or experimental validation. As a contribution, the study presents a conceptual architecture aimed at supporting the reduction of the initial effort required to explore relational databases, the documentation of metadata, and the structural understanding of complex data environments.

Keywords: relational databases; metadata; artificial intelligence; large language models; data engineering.

LISTA DE FIGURAS

Figura 1 – Percurso metodológico adotado na pesquisa	41
Figura 2 – Perfil dos participantes da pesquisa	46
Figura 3 – Tempo necessário para compreender a estrutura de um banco de dados desconhecido	47
Figura 4 – Principais dificuldades encontradas na exploração de bancos de dados relacionais	48
Figura 5 – Impacto do esforço necessário para compreender a estrutura dos dados no desenvolvimento das atividades	49
Figura 6 – Percepção dos participantes sobre a utilidade da solução proposta	51
Figura 7 – Funcionalidades consideradas mais relevantes pelos participantes	52
Figura 8 – Preocupação dos participantes quanto ao envio de metadados para serviços externos	53
Figura 9 – Influência da execução local na adoção da solução proposta	54
Figura 10 – Tela inicial conceitual da solução proposta	59
Figura 11 – Tela conceitual de visualização do grafo estrutural	60
Figura 12 – Tela conceitual de detalhamento de tabela e metadados extraídos	61
Figura 13 – Tela conceitual de relacionamentos inferidos por similaridade semântica	62
Figura 14 – Tela conceitual de diagnóstico estrutural	63
Figura 15 – Diagrama de Casos de Uso da solução proposta	65
Figura 16 – Diagrama de fluxo de dados da solução proposta	69

LISTA DE QUADROS

Quadro 1 – Exemplo de prompt estruturado para análise de DDL	28
Quadro 2 – Comparação entre soluções relacionadas e a proposta apresentada	37
Quadro 3 – Métricas sugeridas para avaliação futura da solução proposta	45
Quadro 4 – Exemplo conceitual de consulta para extração de metadados	70

LISTA DE ABREVIATURAS E SIGLAS

API	<i>Application Programming Interface</i>
BI	<i>Business Intelligence</i>
DDL	<i>Data Definition Language</i>
ERD	<i>Entity Relationship Diagram</i>
IA	Inteligência Artificial
IEC	<i>International Electrotechnical Commission</i>
ISO	<i>International Organization for Standardization</i>
LGPD	Lei Geral de Proteção de Dados
LLM	<i>Large Language Model</i>
PDF	<i>Portable Document Format</i>
PLN	Processamento de Linguagem Natural
SGBD	Sistema Gerenciador de Banco de Dados
SQL	<i>Structured Query Language</i>
TI	Tecnologia da Informação

SUMÁRIO

1	INTRODUÇÃO	15
1.1	Justificativa	17
1.2	Objetivos	19
1.2.1	Objetivo Geral	19
1.2.2	Objetivos Específicos	19
2	FUNDAMENTAÇÃO TEÓRICA	21
2.1	Bancos de Dados e sua Estrutura	21
2.2	Modelagem de Dados e Complexidade Estrutural	22
2.3	Engenharia de Dados e Desafios na Compreensão de Dados	23
2.4	Business Intelligence e Geração de Insights	24
2.5	Metadados e Data Catalog	24
2.6	Ferramentas de Apoio à Análise Estrutural de Dados	25
2.7	Teoria de Grafos Aplicada à Análise de Esquemas Relacionais	25
2.8	Inteligência Artificial Aplicada à Compreensão Estrutural de Bancos de Dados	26
2.9	Prompt Engineering e Janela de Contexto em Large Language Models	27
2.10	Embeddings e Busca Semântica para Inferência de Relacionamentos	28
2.11	Privacidade de Dados e Uso de Inteligência Artificial	30
2.12	Tecnologias e Conceitos Computacionais Relacionados	31
3	TRABALHOS RELACIONADOS	33
3.1	Ferramentas de Visualização e Modelagem de Bancos de Dados Relacionais	33
3.2	Catálogos de Dados e Governança de Metadados	34
3.3	Inteligência Artificial Aplicada à Compreensão de Dados	35
3.4	Comparação entre as Soluções Existentes e a Proposta	36
3.5	Síntese dos Trabalhos Relacionados	38

4	METODOLOGIA	40
4.1	Levantamento com profissionais e estudantes da área de TI	41
4.2	Modelagem da solução proposta	43
4.3	Diretrizes para avaliação futura	44
5	RESULTADOS	46
5.1	Levantamento e Análise dos Dados	46
5.1.1	Perfil dos Participantes	46
5.1.2	Tempo Necessário para Compreensão de Bases de Dados Desconhecidas	47
5.1.3	Principais Dificuldades Encontradas	48
5.1.4	Impactos do Esforço de Compreensão Estrutural	49
5.1.5	Percepção sobre a Utilidade da Solução	50
5.1.6	Funcionalidades Consideradas Relevantes	51
5.1.7	Privacidade e Segurança da Informação	52
5.1.8	Execução Local dos Modelos	53
5.1.9	Síntese dos Resultados	54
6	PROPOSTA DA SOLUÇÃO	56
6.1	Visão Geral da Solução	56
6.2	Protótipo Visual da Interface Proposta	57
6.3	Fluxo de Funcionamento	64
6.4	Diagrama de Casos de Uso	65
6.5	Arquitetura da Solução	66
6.5.1	Arquitetura de Processamento	69
6.5.2	Geração de Descrições por IA	71
6.6	Tecnologias Propostas para a Solução	73
6.7	Contribuições Esperadas	74
6.8	Aderência da Solução aos Requisitos Identificados	75

6.9	Limitações da Solução	76
7	CONCLUSÕES E TRABALHOS FUTUROS	79
7.1	Conclusões	79
7.2	Trabalhos Futuros	82
	REFERÊNCIAS	85
	APÊNDICE A – QUESTIONÁRIO APLICADO AOS PARTICIPANTES	88

1 INTRODUÇÃO

A crescente geração e utilização de dados nas organizações contemporâneas tem impulsionado a adoção de bancos de dados cada vez mais complexos, compostos por múltiplas tabelas, relacionamentos interdependentes e regras de negócio que refletem a dinâmica dos sistemas organizacionais. Essa complexidade estrutural, embora necessária para atender às demandas crescentes por armazenamento e processamento de informações, representa um desafio significativo para profissionais como cientistas de dados, engenheiros de dados e analistas, sobretudo quando se deparam com bases de dados desconhecidas ou desprovidas de documentação técnica adequada (Silberschatz; Korth; Sudarshan, 2019).

No cenário profissional contemporâneo, a compreensão da estrutura de um banco de dados constitui uma etapa prévia e indispensável a qualquer atividade analítica. Antes de construir consultas, desenvolver *dashboards* ou gerar *insights*, o profissional precisa compreender como as informações estão organizadas, quais são os relacionamentos entre as entidades e de que maneira a lógica de negócio está incorporada à estrutura da base. Esse processo, frequentemente conduzido de forma manual e exploratória, pode consumir parcela significativa do tempo disponível para o projeto, impactando diretamente a produtividade e a eficiência das equipes, especialmente quando não há documentação disponível ou quando o analista ingressa em um ambiente de dados que lhe é completamente desconhecido (Kimball; Ross, 2013).

Conforme apontam Batini, Ceri e Navathe (1992), fatores como a elevada quantidade de tabelas, a ausência de padronização na nomenclatura, a existência de relacionamentos implícitos não formalizados no esquema e a documentação incompleta ou desatualizada contribuem para ampliar consideravelmente a dificuldade de análise e interpretação dessas bases. Nesse sentido, Heuser (2009) destaca que uma modelagem bem estruturada e documentada é essencial para garantir a clareza, a consistência e a manutenção eficiente dos dados ao longo do ciclo de vida dos sistemas. Contudo, na prática, é comum que alterações estruturais ocorram sem a devida atualização da documentação, comprometendo progressivamente a compreensão das bases e aumentando o esforço necessário para sua análise e evolução.

A área de Tecnologia da Informação (TI) tem buscado, de forma contínua, soluções que automatizem processos e reduzam a complexidade associada à análise de dados. Nesse sentido, ferramentas de *Business Intelligence* (BI), plataformas de modelagem e soluções de engenharia de dados têm evoluído com o objetivo de facilitar a geração de *insights* e apoiar a tomada de decisão (Sharda; Delen; Turban, 2017). Entretanto, conforme observam Elmasri e

Navathe (2016), a maioria dessas ferramentas ainda exige conhecimento prévio da estrutura do banco ou depende de documentação existente para funcionar de maneira eficaz, o que limita sua aplicabilidade em cenários nos quais os dados são heterogêneos, incompletos ou pouco documentados. Esse conjunto de limitações evidencia uma lacuna tecnológica relevante no que se refere à compreensão automatizada de esquemas relacionais.

No contexto da governança de dados, surgem os *Data Catalogs* como ferramentas projetadas para organizar, documentar e facilitar o acesso às informações por meio de metadados estruturados. De acordo com Nargesian *et al.* (2019), essas soluções possibilitam a descoberta, a compreensão e a governança dos dados em ambientes orientados a dados. Contudo, muitas dessas ferramentas ainda dependem de processos manuais de catalogação ou de intervenção humana para que a análise estrutural seja realmente útil, o que reforça a necessidade de abordagens capazes de automatizar essa interpretação de forma mais abrangente e independente.

Diante desse cenário, a inteligência artificial emerge como um recurso promissor para apoiar a interpretação automatizada de estruturas de dados. Segundo Russell e Norvig (2021), sistemas baseados em Inteligência Artificial (IA) são capazes de realizar tarefas que normalmente exigiriam inteligência humana, como reconhecimento de padrões, inferência lógica e interpretação semântica de informações. No contexto de bancos de dados, essa capacidade possibilita a identificação automatizada de estruturas, relacionamentos e padrões implícitos a partir da análise de metadados, sem a necessidade de documentação prévia (Géron, 2019). Em especial, os modelos de linguagem de grande porte, conhecidos como *Large Language Models* (LLMs), têm demonstrado capacidade relevante para interpretar estruturas técnicas e gerar descrições em linguagem natural, tornando-se particularmente adequados ao propósito de apoiar a compreensão de bases de dados complexas (Hong *et al.*, 2024).

Cabe destacar, todavia, que a utilização de sistemas baseados em IA no contexto da análise de dados impõe atenção rigorosa aos aspectos relacionados à privacidade e à segurança da informação. No Brasil, a Lei Geral de Proteção de Dados (LGPD) (Lei nº 13.709/2018) (Brasil, 2018) estabelece diretrizes para o tratamento de dados pessoais, orientando que apenas as informações estritamente necessárias sejam utilizadas. Nesse sentido, o presente trabalho adota, como princípio de projeto, o uso exclusivo de metadados estruturais, como nomes de tabelas, atributos e relacionamentos, no processo de análise e integração com serviços de IA, evitando o compartilhamento de dados pessoais ou sensíveis e promovendo conformidade com a legislação vigente.

Considerando o exposto, o presente trabalho propõe uma solução para análise automatizada da estrutura de bancos de dados relacionais com o auxílio de IA, com o objetivo de facilitar a compreensão dos dados, de seus relacionamentos e de sua organização. A proposta fundamenta-se na utilização de tecnologias amplamente adotadas na área de dados, como a linguagem *Python* para processamento e análise, *Structured Query Language* (SQL) para extração de metadados em bancos *PostgreSQL* e *MySQL*, as bibliotecas *Pandas* e *NetworkX* (Hagberg; Schult; Swart, 2008) para manipulação e modelagem estrutural dos dados, e *React* com *Tailwind CSS* para a construção de uma interface visual interativa. Adicionalmente, prevê-se a integração com LLMs, acessados por meio de *Application Programming Interfaces* (APIs) externas, visando à geração automática de descrições e ao apoio na identificação de padrões e possíveis inconsistências nas estruturas analisadas. Dessa forma, a solução proposta busca contribuir para a redução da lacuna existente entre a complexidade técnica dos esquemas relacionais e a necessidade de compreensão ágil por parte dos profissionais da área, especialmente em cenários que envolvam bases legadas, pouco documentadas ou de elevada complexidade estrutural.

1.1 Justificativa

A relevância deste trabalho decorre da combinação entre três fatores principais: a complexidade crescente dos esquemas relacionais, a dependência de documentação frequentemente desatualizada e o esforço manual exigido para interpretação de bases de dados desconhecidas. Em ambientes organizacionais, bancos de dados relacionais podem reunir grande quantidade de tabelas, atributos, chaves e relacionamentos, refletindo regras de negócio que se modificam ao longo do tempo. Quando essa estrutura não é devidamente documentada, a compreensão inicial da base torna-se uma atividade demorada, sujeita a interpretações equivocadas e dependente da experiência prévia dos profissionais envolvidos (Silberschatz; Korth; Sudarshan, 2019; Kimball; Ross, 2013).

Embora ferramentas de BI, modelagem de dados e administração de bancos tenham evoluído significativamente, ainda persistem limitações quanto à interpretação automatizada da estrutura dos dados. Soluções como *DBeaver* (DBeaver Corporation, s.d.a) e *MySQL Workbench* (Oracle Corporation, 2026), por exemplo, oferecem recursos importantes para visualização de tabelas, atributos e relacionamentos, mas não têm como finalidade principal interpretar semanticamente os elementos do banco, sugerir significados para atributos ou identificar possíveis inconsistências estruturais em bases pouco documentadas. De modo semelhante, ferramentas de

Data Catalog contribuem para a organização e governança dos metadados, porém sua efetividade depende, em muitos casos, da qualidade da documentação existente e da intervenção humana no processo de catalogação (Nargesian *et al.*, 2019; Elmasri; Navathe, 2016).

Nesse contexto, identifica-se uma lacuna relacionada à compreensão estrutural automatizada de bancos de dados relacionais. A proposta deste trabalho busca atuar justamente sobre essa lacuna, ao especificar uma solução baseada na extração de metadados, na representação estrutural por meio de grafos e no uso de LLMs para apoiar a geração de descrições, a sugestão de significados para atributos e a identificação de possíveis padrões ou inconsistências. Assim, a IA não é tratada apenas como uma interface para geração de consultas, mas como um recurso de apoio à interpretação da estrutura lógica dos dados (Russell; Norvig, 2021; Hong *et al.*, 2024).

A pertinência do problema também foi observada no levantamento realizado com profissionais e estudantes da área de TI. Os resultados indicaram que parte significativa dos participantes já enfrentou dificuldades ao lidar com bancos de dados sem documentação adequada, especialmente na identificação de relacionamentos, compreensão de atributos e interpretação de regras de negócio implícitas. Além disso, os respondentes demonstraram receptividade à possibilidade de uma solução capaz de automatizar parte desse processo de análise estrutural. Esses achados reforçam que o problema abordado não é apenas teórico, mas recorrente em contextos práticos de análise, engenharia de dados e desenvolvimento de soluções orientadas a dados.

Outro aspecto que justifica a proposta é a preocupação com privacidade e segurança da informação. Como a solução proposta se baseia na análise de metadados estruturais, como nomes de tabelas, atributos, tipos de dados e relacionamentos, evita-se o envio de registros sensíveis ou dados pessoais para serviços de IA. Essa decisão está alinhada aos princípios da LGPD (Lei nº 13.709/2018), especialmente quanto à necessidade de limitar o tratamento de dados ao mínimo necessário para atingir a finalidade proposta (Brasil, 2018). Além disso, a adoção do princípio de *Privacy by Design* reforça a preocupação em incorporar aspectos de proteção de dados desde a concepção da arquitetura (Cavoukian, 2009).

Do ponto de vista acadêmico, este trabalho contribui para a discussão sobre a aplicação de IA à engenharia de dados, especialmente no apoio à compreensão de esquemas relacionais pouco documentados. Ao propor e modelar uma arquitetura voltada à análise estrutural automatizada, a pesquisa estabelece uma base conceitual para trabalhos futuros que envolvam geração automática de documentação, governança de metadados, visualização estrutural e apoio à exploração inicial de bancos de dados relacionais.

Portanto, a justificativa deste trabalho está relacionada à necessidade de reduzir o esforço manual de compreensão de bases desconhecidas, aumentar a clareza sobre estruturas relacionais complexas e apoiar profissionais da área de dados na etapa inicial de exploração e interpretação dos dados. Dessa forma, a solução proposta busca contribuir para análises mais confiáveis, documentação mais acessível e maior eficiência em atividades que dependem do entendimento prévio da estrutura dos bancos de dados.

1.2 Objetivos

A presente pesquisa propõe a modelagem de uma solução capaz de analisar e facilitar a compreensão estrutural de bancos de dados relacionais, considerando a complexidade inerente a esses ambientes, bem como a frequente ausência de documentação técnica adequada. Nesse contexto, busca-se especificar uma arquitetura baseada na extração de metadados e no uso de IA para apoiar a identificação, organização e apresentação de informações relevantes sobre a estrutura dos dados, em consonância com os objetivos estabelecidos para o estudo.

1.2.1 Objetivo Geral

Propor e modelar uma solução capaz de apoiar a análise automática da estrutura de bancos de dados relacionais, com o auxílio de IA, com o objetivo de facilitar a compreensão dos dados, de seus relacionamentos e de sua organização, de modo a auxiliar profissionais na exploração e na utilização eficiente das informações, especialmente em cenários que envolvam bases legadas ou pouco documentadas.

1.2.2 Objetivos Específicos

- Propor e modelar mecanismos para identificar tabelas, atributos, chaves primárias, chaves estrangeiras e relacionamentos presentes em bancos de dados relacionais;
- Modelar uma representação da estrutura lógica do banco de dados, evidenciando como os dados estão organizados e interconectados por meio de grafos e diagramas estruturais;
- Especificar mecanismos de visualização que favoreçam a interpretação da estrutura do banco de dados por parte dos usuários;
- Propor, em nível arquitetural, a utilização de LLMs como recurso de apoio à geração de descrições automáticas de tabelas, à sugestão de significados para atributos e à identificação

de possíveis relacionamentos implícitos;

- Avaliar a percepção de profissionais e estudantes da área de TI quanto às dificuldades enfrentadas na compreensão de esquemas relacionais e à aceitação da solução proposta, por meio de levantamento aplicado junto ao público-alvo;
- Investigar formas de apoiar a identificação de padrões e possíveis inconsistências na estrutura dos dados;
- Discutir o potencial da solução proposta para apoiar atividades posteriores de análise, criação de *dashboards* e geração de *insights*, a partir de uma compreensão estrutural mais clara dos dados.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, apresentam-se os principais conceitos teóricos que fundamentam o desenvolvimento do presente trabalho. A Seção 2.1 aborda os fundamentos dos bancos de dados relacionais e sua estrutura, destacando seus principais componentes e formas de organização. Em seguida, a Seção 2.2 discute a modelagem de dados e os desafios associados à complexidade estrutural desses ambientes. A Seção 2.3 trata da engenharia de dados e dos desafios relacionados à compreensão de bases de dados complexas. Na Seção 2.4, apresenta-se o conceito de BI e sua relação com a geração de *insights*. Posteriormente, a Seção 2.5 aborda os conceitos de metadados e *Data Catalog*, enfatizando sua importância para a organização, documentação e governança dos dados. A Seção 2.6 apresenta ferramentas de apoio à análise estrutural de dados, destacando suas funcionalidades e limitações.

Na sequência, a Seção 2.7 discute a teoria de grafos aplicada à análise de esquemas relacionais, evidenciando como tabelas e relacionamentos podem ser representados por meio de vértices e arestas. A Seção 2.8 aborda a aplicação da IA na compreensão estrutural de bancos de dados relacionais. A Seção 2.9 apresenta conceitos relacionados a *Prompt Engineering* e à janela de contexto em LLMs, enquanto a Seção 2.10 discute o uso de *embeddings* e busca semântica para apoiar a inferência de relacionamentos implícitos. Por fim, a Seção 2.11 apresenta aspectos relacionados à privacidade de dados e ao uso responsável da IA, e a Seção 2.12 apresenta tecnologias e conceitos computacionais relacionados à concepção de soluções voltadas à análise estrutural automatizada de bancos de dados relacionais.

2.1 Bancos de Dados e sua Estrutura

Os bancos de dados desempenham um papel central nos sistemas de informação modernos, sendo responsáveis pelo armazenamento, pela organização e pela recuperação eficiente de grandes volumes de dados. Segundo Silberschatz, Korth e Sudarshan (2019), um banco de dados consiste em um conjunto de dados inter-relacionados, estruturados de forma a atender às necessidades específicas de aplicações e usuários, garantindo consistência, integridade e disponibilidade das informações.

A estrutura de um banco de dados é composta por elementos como tabelas, atributos, chaves primárias e estrangeiras, além de restrições que asseguram a integridade referencial. Esses componentes definem não apenas a forma como os dados estão organizados, mas também

como se relacionam entre si. Em ambientes corporativos, essa estrutura tende a se tornar progressivamente mais complexa, em decorrência do crescimento contínuo do volume de dados e da necessidade de atender, de forma simultânea, a múltiplas demandas de negócio (Silberschatz; Korth; Sudarshan, 2019).

Outro ponto relevante é que a evolução dos sistemas de informação tem impulsionado a adoção e o aprimoramento dos bancos de dados relacionais, ampliando o nível de complexidade estrutural desses sistemas. Nesse cenário, a compreensão da organização dos dados torna-se uma atividade essencial, porém desafiadora, especialmente em situações nas quais não há documentação adequada ou quando os sistemas foram desenvolvidos de forma incremental ao longo do tempo (Batini; Ceri; Navathe, 1992; Heuser, 2009).

2.2 Modelagem de Dados e Complexidade Estrutural

A modelagem de dados constitui uma etapa fundamental no desenvolvimento de sistemas de informação, pois define a forma como os dados serão organizados e representados. Segundo Batini, Ceri e Navathe (1992), a modelagem conceitual possibilita a abstração da realidade em entidades, atributos e relacionamentos, facilitando tanto o entendimento quanto a construção de bancos de dados. De acordo com Heuser (2009), uma modelagem bem estruturada é essencial para garantir a clareza, a consistência e a manutenção eficiente dos dados ao longo do ciclo de vida dos sistemas.

Ainda assim, ao longo do ciclo de vida dos sistemas, é comum ocorrerem alterações estruturais sem a devida atualização da documentação. Esse cenário compromete a clareza da estrutura dos dados, dificultando a identificação de relações e dependências entre tabelas. A ausência de um modelo bem documentado compromete diretamente a compreensão e a manutenção dos bancos de dados, aumentando o esforço necessário para análise e evolução dos sistemas (Heuser, 2009).

Além disso, fatores como o crescimento desordenado das bases, a falta de padronização na nomenclatura e a ausência de boas práticas de modelagem contribuem significativamente para a complexidade estrutural. Tais aspectos tornam o processo de interpretação dos dados mais demorado e suscetível a erros, impactando diretamente a produtividade dos profissionais envolvidos (Batini; Ceri; Navathe, 1992).

2.3 Engenharia de Dados e Desafios na Compreensão de Dados

A engenharia de dados tem como objetivo central estruturar, processar e disponibilizar informações de forma que estas possuam valor analítico e confiabilidade. Contudo, na prática profissional, engenheiros e cientistas de dados raramente encontram ambientes ideais. Frequentemente, lidam com bases heterogêneas, extensas e, em sua maioria, desprovidas de documentação técnica atualizada (Silberschatz; Korth; Sudarshan, 2019; Kimball; Ross, 2013).

Nesse contexto, o profissional frequentemente assume uma atividade investigativa de reconstrução do conhecimento estrutural da base, interpretando esquemas legados e documentação incompleta. Conforme apontam Silberschatz, Korth e Sudarshan (2019), a complexidade de bancos de dados relacionais e heterogêneos torna a compreensão estrutural das informações uma tarefa desafiadora, especialmente na ausência de documentação adequada. Paralelamente, Kimball e Ross (2013) destacam que uma parcela significativa do tempo em projetos de BI é consumida pelas etapas de compreensão, limpeza e preparação dos dados.

Esse gargalo exploratório ocorre porque a interpretação da semântica dos dados e dos relacionamentos implícitos entre tabelas continua sendo uma atividade predominantemente manual. Sem ferramentas que automatizem a análise estrutural, o engenheiro fica suscetível a problemas como:

- Perda de linhagem (*Data Lineage*): dificuldade em rastrear a origem e as transformações dos dados em tabelas com nomes pouco descritivos;
- Inferências técnicas equivocadas: suposições incorretas sobre chaves primárias e estrangeiras que não foram explicitamente declaradas no *Data Definition Language* (DDL) do banco;
- Inconsistência analítica: geração de relatórios incorretos devido à má interpretação da lógica de negócio incorporada à estrutura do banco.

Dessa forma, a complexidade inerente à engenharia de dados moderna exige que o entendimento da base não dependa exclusivamente da memória institucional ou de documentos estáticos, mas de uma capacidade de análise dinâmica e automatizada da estrutura disponível (Kimball; Ross, 2013).

2.4 Business Intelligence e Geração de Insights

O BI abrange um conjunto de metodologias, processos e ferramentas que permitem transformar dados brutos em informações relevantes para a tomada de decisão. Conforme Sharda, Delen e Turban (2017), os avanços nas tecnologias de BI têm possibilitado análises cada vez mais rápidas, interativas e acessíveis.

Entretanto, a qualidade das análises depende diretamente da compreensão dos dados utilizados. Sem um entendimento claro da estrutura do banco de dados, há risco de interpretações equivocadas, inconsistências nos resultados e falhas na geração de *insights* (Kimball; Ross, 2013; Sharda; Delen; Turban, 2017).

Dessa forma, a dificuldade em compreender a organização dos dados pode comprometer todo o processo analítico, evidenciando a relevância de ferramentas que apoiem a interpretação estrutural dos bancos de dados (Kimball; Ross, 2013).

2.5 Metadados e Data Catalog

Os metadados podem ser definidos como informações que descrevem outros dados, fornecendo contexto adicional que não está presente no conteúdo dos próprios dados. Em termos simples, metadados são "dados sobre dados". Eles registram atributos como origem, formato, autor, data de criação, tipo e estrutura de um conjunto de informações, facilitando a busca, organização e uso desses dados em sistemas de informação. Em ambientes de bancos de dados, os metadados permitem identificar e compreender aspectos relevantes das tabelas, relacionamentos e elementos de dados, contribuindo para a interpretação e utilização eficiente das informações (Badman; Kosinski, 2024).

Nesse cenário, surgem os *Data Catalogs*, ferramentas projetadas para organizar, documentar e facilitar o acesso aos dados por meio de metadados estruturados. De acordo com Nargesian *et al.* (2019), os catálogos de dados possibilitam a descoberta, a compreensão e a governança dos dados, sendo fundamentais em ambientes orientados a dados.

Apesar disso, muitas dessas ferramentas ainda dependem de documentação manual ou de processos parcialmente automatizados, o que limita sua eficiência em ambientes altamente dinâmicos. Nesse sentido, a proposição de sistemas capazes de interpretar automaticamente a estrutura dos bancos de dados se alinha diretamente aos avanços e às necessidades dessa área (Nargesian *et al.*, 2019).

2.6 Ferramentas de Apoio à Análise Estrutural de Dados

Diversas ferramentas têm sido desenvolvidas com o objetivo de auxiliar na análise e na visualização de dados, incluindo plataformas de modelagem de banco de dados, sistemas de BI e soluções de engenharia de dados. Exemplos notáveis incluem *DBeaver* (DBeaver Corporation, s.d.a), *MySQL Workbench* (Oracle Corporation, 2026) e ferramentas comerciais de *Data Catalog* (Apache Software Foundation, s.d.; DataHub Project, s.d.a; Microsoft Corporation, 2026). Essas soluções permitem a representação gráfica das estruturas de dados, facilitando a interpretação por parte dos usuários e oferecendo suporte à construção de consultas, diagramas de relacionamento e *dashboards* analíticos.

À luz dos conceitos de modelagem e administração de bancos relacionais discutidos por Elmasri e Navathe (2016), observa-se que tais ferramentas fornecem suporte técnico relevante, mas a interpretação semântica dos elementos permanece dependente da análise do usuário. Por exemplo, ferramentas como *DBeaver* e *MySQL Workbench* fornecem visualização de esquemas e relacionamentos, mas não conseguem automaticamente inferir a semântica dos dados ou detectar inconsistências em bancos não documentados. Da mesma forma, *Data Catalogs* comerciais auxiliam na catalogação e organização dos dados, porém sua automação é limitada, exigindo configuração manual e intervenção do usuário para que a análise estrutural seja realmente útil.

Essa dependência de intervenção humana limita a aplicabilidade dessas soluções em cenários nos quais os dados são heterogêneos, incompletos ou pouco documentados. Tais limitações evidenciam uma lacuna tecnológica, reforçando a relevância do desenvolvimento de soluções capazes de automatizar a compreensão estrutural de bancos de dados, reduzindo o esforço manual e aumentando a confiabilidade das análises (Elmasri; Navathe, 2016).

2.7 Teoria de Grafos Aplicada à Análise de Esquemas Relacionais

A representação de um banco de dados relacional como um grafo constitui uma abordagem poderosa para a análise estrutural de seus elementos e relacionamentos. Formalmente, um grafo $G = (V, E)$ é definido por um conjunto de vértices V e um conjunto de arestas E que conectam pares de vértices (Cormen *et al.*, 2009). No contexto da solução proposta, cada tabela do banco de dados é representada como um vértice, e cada relacionamento entre tabelas, seja por chave estrangeira explícita ou por relacionamento inferido por similaridade semântica, é representado como uma aresta direcionada entre os vértices correspondentes.

Essa modelagem poderia ser implementada por meio da biblioteca *NetworkX*, amplamente utilizada em projetos científicos e de engenharia de dados para a criação, manipulação e análise de grafos e redes complexas (Hagberg; Schult; Swart, 2008). A partir dessa representação, torna-se possível aplicar métricas de análise estrutural que vão além da simples visualização dos relacionamentos.

Uma das métricas mais relevantes nesse contexto é a *centralidade de grau* (*Degree Centrality*), que mensura a proporção de tabelas do esquema às quais uma determinada tabela está diretamente conectada (Freeman, 1978). Na prática, quanto maior esse valor, mais central é a tabela dentro do esquema relacional, característica típica de entidades *core* do negócio, como cliente, pedido ou produto, que naturalmente se relacionam com um grande número de outras tabelas. A biblioteca *NetworkX* permitiria calcular essa métrica automaticamente a partir do grafo construído, bastando uma única chamada de função para obter os valores de centralidade de todas as tabelas simultaneamente (Hagberg; Schult; Swart, 2008).

Dessa forma, a aplicação da centralidade de grau sobre o grafo do banco de dados poderia permitir a identificação, sem necessidade de conhecimento prévio do esquema, das entidades estruturalmente mais relevantes. Essa informação poderia orientar o usuário na exploração inicial do banco, priorizando a compreensão das tabelas de maior impacto antes de avançar para entidades periféricas de menor conectividade (Freeman, 1978; Hagberg; Schult; Swart, 2008).

Além da centralidade de grau, outras métricas de grafos podem ser exploradas em trabalhos futuros, como a *centralidade de intermediação* (*Betweenness Centrality*), que identifica tabelas que funcionam como pontes entre diferentes partes do esquema, e a *detecção de comunidades*, que agrupa tabelas com alta densidade de relacionamentos entre si, sugerindo módulos funcionais ou domínios de negócio presentes na base de dados (Cormen *et al.*, 2009).

2.8 Inteligência Artificial Aplicada à Compreensão Estrutural de Bancos de Dados

A utilização de IA tem se expandido significativamente no contexto da análise de dados, contribuindo para a automação de processos complexos e para a extração de conhecimento a partir de grandes volumes de informação. Segundo Russell e Norvig (2021), a IA pode ser definida como a capacidade de sistemas computacionais realizarem tarefas que normalmente exigiriam inteligência humana, como reconhecimento de padrões, inferência e interpretação de dados.

No contexto de bancos de dados, a aplicação de técnicas de IA pode apoiar a identificação

automatizada de estruturas, relacionamentos e padrões implícitos nas bases de dados. Conforme discutido por Géron (2019), técnicas de IA podem contribuir para a interpretação de informações complexas e para a identificação de padrões em conjuntos de dados. No contexto deste trabalho, essa capacidade é considerada como apoio à interpretação de metadados estruturais, oferecendo hipóteses sobre o significado de tabelas, atributos e relacionamentos quando a documentação é limitada.

Nesse sentido, a solução proposta neste trabalho prevê a utilização de recursos de IA com o objetivo de apoiar a compreensão estrutural de bancos de dados. A partir da análise de metadados estruturais, a IA poderia apoiar a geração de descrições resumidas das tabelas, a identificação de possíveis relacionamentos entre entidades, a sugestão de significados para atributos e a indicação de padrões ou inconsistências estruturais. Essa abordagem poderia contribuir para reduzir o esforço manual necessário para interpretar bases desconhecidas (Russell; Norvig, 2021; Géron, 2019).

Dessa forma, a integração de IA ao sistema proposto configura-se como um diferencial relevante, pois poderia contribuir para uma compreensão potencialmente mais ágil e estruturada das bases de dados. Essa funcionalidade, quando implementada e validada experimentalmente, poderia apoiar a produtividade de profissionais da área de dados, facilitando atividades como exploração de dados, construção de *dashboards* e geração de *insights*, especialmente em contextos nos quais não há documentação disponível ou conhecimento prévio da estrutura analisada (Hong *et al.*, 2024).

2.9 Prompt Engineering e Janela de Contexto em Large Language Models

A utilização de LLMs no contexto da análise estrutural de bancos de dados exige uma etapa prévia de preparação das informações que serão submetidas ao modelo. Esse processo, denominado *Prompt Engineering*, consiste na construção deliberada e estruturada de instruções textuais que orientam o LLM a produzir respostas precisas e relevantes para um determinado problema (White *et al.*, 2023).

No âmbito da solução proposta neste trabalho, a estrutura do banco de dados é representada por meio do DDL, que contém informações sobre tabelas, atributos, tipos de dados, chaves primárias e chaves estrangeiras. Esse conjunto de metadados estruturais é serializado e incorporado ao prompt enviado ao LLM, permitindo que o modelo compreenda a organização do banco e gere descrições semânticas dos seus elementos. Um exemplo simplificado de prompt

estruturado é apresentado no Quadro 1.

Quadro 1 – Exemplo de prompt estruturado para análise de DDL

```
1  Você é um especialista em banco de dados relacionais.
2
3  A seguir está o DDL de um banco de dados:
4
5  CREATE TABLE cliente (
6      id_cliente INT PRIMARY KEY,
7      nome VARCHAR(100),
8      data_cadastro DATE
9  );
10
11 Com base nessa estrutura, descreva o propósito desta tabela e
12 o significado de cada atributo.
```

Fonte: Autoria própria (2026).

Contudo, um desafio relevante surge quando o banco de dados analisado é de grande porte, composto por centenas ou milhares de tabelas. Nesse cenário, a serialização completa do DDL pode exceder a *janela de contexto* (*context window*) do modelo, que corresponde ao limite máximo de *tokens* que o LLM é capaz de processar em uma única requisição (Anthropic, 2023). Modelos contemporâneos apresentam janelas de contexto que variam de 8 mil a mais de 200 mil *tokens*, o que, embora expressivo, pode ser insuficiente para esquemas relacionais muito extensos.

Para contornar essa limitação, estratégias como a segmentação do DDL em subconjuntos temáticos de tabelas, a priorização das entidades mais centrais do esquema e a sumarização progressiva dos metadados podem ser adotadas. Dessa forma, é possível manter a coerência semântica das descrições geradas mesmo em ambientes de elevada complexidade estrutural (White *et al.*, 2023; Anthropic, 2023).

2.10 Embeddings e Busca Semântica para Inferência de Relacionamentos

Em bancos de dados relacionais, nem sempre os relacionamentos entre tabelas são explicitamente formalizados por meio de chaves estrangeiras declaradas no DDL. Frequentemente, associações implícitas existem entre colunas de tabelas distintas, reconhecíveis por profissionais experientes, mas não identificadas diretamente por ferramentas convencionais de análise estrutural. Para lidar com esse desafio, a solução proposta incorpora o conceito de *embeddings*,

que consistem em representações vetoriais densas de elementos textuais em um espaço de alta dimensionalidade (Mikolov *et al.*, 2013).

Um *embedding* é gerado por um modelo ao processar um texto, como o nome de uma coluna, seu tipo de dado ou uma descrição associada à tabela, mapeando esse conteúdo para um vetor numérico. A partir dessa representação, elementos semanticamente semelhantes podem ocupar regiões próximas no espaço vetorial. No contexto deste trabalho, essa abordagem pode auxiliar na identificação de atributos potencialmente relacionados, como colunas que representem conceitos semelhantes em diferentes tabelas, ainda que utilizem nomenclaturas distintas.

Entretanto, é importante destacar que a similaridade obtida por meio de *embeddings* não garante, por si só, a existência de um relacionamento real entre os atributos analisados. Abreviações, homônimos, diferenças de idioma, padrões locais de nomenclatura e ausência de contexto de negócio podem produzir falsos positivos ou falsos negativos. Por exemplo, nomes como `id_cliente` e `cod_cli` podem indicar conceitos semelhantes, mas essa correspondência depende do domínio analisado e deve ser interpretada com cautela.

A medida utilizada para quantificar essa proximidade semântica é a *similaridade de cosseno*, que compara a direção de dois vetores no espaço multidimensional, produzindo valores entre -1 e 1 . Quanto mais próximo de 1 for o resultado, maior tende a ser a semelhança semântica entre os elementos comparados (Manning; Raghavan; Schütze, 2008). Na prática, essa métrica pode ser calculada por bibliotecas especializadas, como o *scikit-learn* e o *sentence-transformers*, dispensando a implementação manual do cálculo pelo desenvolvedor.

Ao aplicar essa medida entre os *embeddings* das colunas do banco de dados, torna-se possível identificar pares de atributos com maior proximidade semântica, mesmo na ausência de chaves estrangeiras explícitas. Esses pares, contudo, devem ser tratados como candidatos a relacionamento, e não como relacionamentos confirmados, pois abreviações, homônimos e convenções específicas do domínio podem produzir correspondências incorretas.

Dessa forma, a solução proposta deve apresentar essas associações ao usuário como sugestões de possíveis relacionamentos implícitos, permitindo validação humana antes de qualquer confirmação. Essa etapa de validação é essencial para reduzir interpretações equivocadas e garantir que a inferência semântica seja utilizada como apoio à análise estrutural, e não como substituição ao julgamento técnico do profissional responsável. Assim, o uso de *embeddings* amplia a capacidade de exploração inicial do esquema relacional, ao mesmo tempo em que preserva a necessidade de análise crítica e validação contextual (Mikolov *et al.*, 2013; Manning;

Raghavan; Schütze, 2008).

2.11 Privacidade de Dados e Uso de Inteligência Artificial

A crescente adoção de sistemas baseados em IA no contexto da análise de dados impõe a necessidade de atenção rigorosa aos aspectos relacionados à privacidade e à segurança da informação. No Brasil, a LGPD, Lei nº 13.709/2018, estabelece diretrizes para o tratamento de dados pessoais, visando à proteção dos direitos fundamentais de liberdade, privacidade e livre desenvolvimento da personalidade (Brasil, 2018).

A LGPD fundamenta-se em princípios como finalidade, adequação, necessidade e segurança, orientando que o tratamento de dados deve se restringir ao mínimo necessário para o atendimento de seus objetivos. Nesse contexto, destaca-se o princípio da minimização de dados, segundo o qual apenas as informações estritamente essenciais devem ser utilizadas, reduzindo riscos associados à exposição indevida de dados sensíveis (Brasil, 2018).

No âmbito de aplicações que utilizam serviços de IA, especialmente por meio de APIs externas, torna-se fundamental avaliar criteriosamente quais informações são compartilhadas com tais serviços. Recomenda-se, como boa prática, evitar o envio de dados pessoais ou sensíveis, priorizando o uso de informações abstratas, agregadas ou estruturais (Brasil, 2018; Cavoukian, 2009).

Nesse sentido, a utilização de metadados, tais como nomes de tabelas, atributos e relacionamentos, apresenta-se como uma abordagem adequada para reduzir a exposição direta dos dados armazenados nas tabelas, pois permite realizar análises estruturais sem a necessidade de enviar registros reais do banco de dados. Tal estratégia encontra respaldo tanto nas diretrizes da LGPD quanto nas práticas contemporâneas de desenvolvimento de sistemas inteligentes, que buscam conciliar capacidade analítica e segurança da informação, em conformidade com normas internacionais como a *International Organization for Standardization (ISO)/International Electrotechnical Commission (IEC) 27001* (ISO; IEC, 2022).

Entretanto, embora os metadados não correspondam aos registros armazenados nas tabelas, eles também podem revelar aspectos sensíveis ou estratégicos do domínio analisado, como nomes de entidades, processos internos, categorias de dados pessoais ou informações relacionadas à lógica de negócio da organização. Dessa forma, a arquitetura proposta deve prever mecanismos de anonimização, filtragem, mascaramento e controle de acesso antes do envio de metadados a serviços externos de IA.

Assim, o uso de metadados deve ser compreendido como uma medida de redução de risco, e não como uma garantia absoluta de privacidade. A adoção de filtros prévios, políticas de acesso, revisão dos elementos enviados e, quando necessário, execução local dos modelos de linguagem são medidas complementares importantes para fortalecer a segurança da solução proposta e ampliar sua aderência às boas práticas de privacidade desde a concepção.

2.12 Tecnologias e Conceitos Computacionais Relacionados

A concepção de soluções voltadas à análise estrutural de bancos de dados relacionais envolve diferentes tecnologias e conceitos computacionais, relacionados à extração de metadados, ao processamento estrutural, à representação de relacionamentos, à integração com modelos de IA e à apresentação visual das informações ao usuário.

No contexto de bancos de dados relacionais, a linguagem SQL desempenha papel fundamental por permitir a consulta, manipulação e extração de informações estruturais armazenadas nos sistemas gerenciadores de bancos de dados. Além de possibilitar o acesso aos dados propriamente ditos, a SQL também permite consultar metadados do esquema, como tabelas, atributos, tipos de dados, chaves primárias, chaves estrangeiras e restrições estruturais (Silberschatz; Korth; Sudarshan, 2019; Elmasri; Navathe, 2016).

Para o processamento dessas informações, linguagens de programação de uso geral, como *Python*, são amplamente empregadas em projetos de ciência de dados, engenharia de dados e automação de processos analíticos. Sua adoção nesse tipo de contexto está relacionada à disponibilidade de bibliotecas para manipulação de dados tabulares, integração com bancos de dados, análise estrutural e comunicação com serviços externos.

Outro conceito relevante para este trabalho é a representação de estruturas relacionais por meio de grafos. Nesse tipo de representação, tabelas podem ser modeladas como vértices, enquanto relacionamentos entre tabelas podem ser representados como arestas. Essa abordagem favorece a visualização da estrutura do banco de dados e permite a aplicação de métricas capazes de identificar entidades centrais, dependências e agrupamentos estruturais (Cormen *et al.*, 2009; Hagberg; Schult; Swart, 2008).

Além disso, os LLMs podem ser utilizados como apoio à interpretação de metadados estruturais. A partir de nomes de tabelas, atributos, tipos de dados e relacionamentos, esses modelos podem auxiliar na geração de descrições em linguagem natural, na sugestão de significados para atributos e na identificação de possíveis padrões ou inconsistências estruturais (White *et al.*,

2023; Hong *et al.*, 2024).

Por fim, tecnologias de interface web podem ser empregadas para apresentar os resultados da análise de forma visual e interativa. Esse tipo de recurso é importante para tornar a exploração da estrutura do banco de dados mais acessível, permitindo que o usuário visualize relacionamentos, consulte descrições geradas automaticamente e navegue pelos elementos identificados.

Dessa forma, as tecnologias e conceitos apresentados nesta seção constituem a base teórica e computacional necessária para compreender a arquitetura proposta neste trabalho. A justificativa específica das tecnologias selecionadas para cada componente da solução é apresentada posteriormente, no capítulo referente à proposta da solução.

3 TRABALHOS RELACIONADOS

3.1 Ferramentas de Visualização e Modelagem de Bancos de Dados Relacionais

Diversas ferramentas foram desenvolvidas com o objetivo de auxiliar profissionais na exploração, administração e compreensão de bancos de dados relacionais. Entre as soluções mais utilizadas destacam-se o *DBeaver* (DBeaver Corporation, s.d.a) e o *MySQL Workbench* (Oracle Corporation, 2026), amplamente empregados em ambientes acadêmicos e corporativos para gerenciamento de estruturas de dados, execução de consultas SQL e visualização de esquemas relacionais.

O *DBeaver* é uma ferramenta multiplataforma que oferece suporte a diversos sistemas gerenciadores de bancos de dados. Entre suas funcionalidades estão a navegação por tabelas, visualização de atributos, inspeção de chaves primárias e estrangeiras, além da geração de diagramas de relacionamento. Essas características permitem ao usuário compreender a estrutura física e lógica do banco de dados, facilitando atividades de manutenção e análise (DBeaver Corporation, s.d.a).

De forma semelhante, o *MySQL Workbench* disponibiliza recursos voltados à modelagem de dados, engenharia reversa e visualização de esquemas relacionais. A ferramenta permite gerar diagramas entidade-relacionamento, também conhecidos como *Entity Relationship Diagram* (ERD), a partir de bancos existentes, contribuindo para a documentação e compreensão das estruturas armazenadas (Oracle Corporation, 2026).

Apesar de suas contribuições, ambas as ferramentas apresentam limitações relevantes quando utilizadas em ambientes pouco documentados. Embora sejam capazes de exibir tabelas, atributos e relacionamentos explicitamente definidos, elas não realizam interpretações semânticas dos elementos identificados. Consequentemente, cabe ao usuário inferir o significado das tabelas, compreender regras de negócio implícitas e identificar possíveis inconsistências estruturais.

Outro aspecto importante refere-se à dependência de relacionamentos formalmente definidos no banco de dados. Em cenários nos quais as chaves estrangeiras não foram corretamente implementadas ou documentadas, a capacidade dessas ferramentas de representar a estrutura do sistema torna-se limitada. Dessa forma, o entendimento do banco permanece fortemente dependente da experiência e do conhecimento prévio do profissional responsável pela análise.

Além disso, ferramentas como *DBeaver* e *MySQL Workbench* são bastante eficientes para inspeção técnica da estrutura física do banco, mas não possuem como finalidade principal

interpretar o contexto de negócio associado às tabelas e atributos. Assim, mesmo que consigam apresentar visualmente a organização do esquema relacional, a compreensão do propósito de cada entidade, da função de cada coluna e da existência de possíveis padrões estruturais continua sendo responsabilidade do usuário.

Observa-se, portanto, que embora ferramentas de modelagem e visualização sejam essenciais para a exploração de bancos de dados relacionais, ainda existe uma lacuna relacionada à interpretação automatizada das estruturas analisadas. Essa limitação constitui uma das principais motivações para a proposta apresentada neste trabalho, que busca complementar a visualização estrutural com recursos de interpretação semântica, documentação automática e apoio por IA.

3.2 Catálogos de Dados e Governança de Metadados

Com a crescente adoção de estratégias orientadas a dados nas organizações, surgiram soluções voltadas à catalogação e governança de ativos informacionais. Nesse contexto, os *Data Catalogs* desempenham papel fundamental ao fornecer mecanismos para organização, documentação e descoberta de dados por meio da utilização de metadados estruturados (Nargesian *et al.*, 2019).

Ferramentas como *Apache Atlas* (Apache Software Foundation, s.d.), *DataHub* (DataHub Project, s.d.a) e *Microsoft Purview* (Microsoft Corporation, 2026) permitem registrar informações relacionadas a tabelas, colunas, processos de transformação e linhagem dos dados. Esses recursos auxiliam profissionais na localização de informações relevantes e contribuem para iniciativas de governança e conformidade regulatória.

Além disso, os *Data Catalogs* possibilitam a centralização do conhecimento organizacional sobre os dados, reduzindo a dependência de conhecimento informal distribuído entre equipes. Em ambientes corporativos complexos, essa característica representa uma importante contribuição para a manutenção e reutilização de ativos de informação.

Entretanto, apesar dos avanços proporcionados por essas soluções, grande parte de sua eficácia depende da qualidade e da atualização dos metadados cadastrados. Em muitos casos, a documentação precisa ser alimentada manualmente pelos usuários ou por equipes especializadas, tornando o processo suscetível a inconsistências e desatualizações.

Outro ponto relevante é que os *Data Catalogs* normalmente atuam como repositórios de documentação e governança, mas nem sempre realizam uma interpretação automática e aprofundada da estrutura relacional. Embora possam armazenar informações sobre tabelas, colunas,

linhagem e responsáveis pelos dados, a extração de significado semântico e a identificação de relações não explicitamente documentadas ainda dependem, em grande parte, da intervenção humana.

Dessa forma, embora os *Data Catalogs* contribuam significativamente para a organização e governança dos dados, eles não eliminam completamente a necessidade de interpretação manual das estruturas presentes nos bancos de dados relacionais. Essa limitação reforça a necessidade de abordagens capazes de analisar e interpretar automaticamente os metadados disponíveis, especialmente em cenários nos quais a documentação existente é incompleta, desatualizada ou insuficiente para apoiar a compreensão inicial da base.

3.3 Inteligência Artificial Aplicada à Compreensão de Dados

Nos últimos anos, os avanços em IA, especialmente na área de Processamento de Linguagem Natural (PLN), impulsionaram o surgimento de novas aplicações voltadas à análise e interpretação de informações estruturadas. Nesse contexto, os LLMs passaram a ser utilizados em tarefas relacionadas à engenharia de software, análise de dados e geração automática de documentação.

Segundo Hong *et al.* (2024), os LLMs vêm sendo amplamente estudados como interfaces em linguagem natural para bancos de dados, especialmente em tarefas relacionadas à geração automática de consultas SQL, documentação e exploração de esquemas relacionais. Entre as aplicações mais comuns destacam-se a geração automática de consultas SQL, a documentação de esquemas relacionais e a assistência à exploração de bases de dados complexas. Essas abordagens demonstram que modelos de linguagem possuem capacidade significativa para interpretar estruturas técnicas e produzir descrições em linguagem natural.

Apesar desses avanços, observa-se que a maior parte das pesquisas concentra-se em problemas específicos, como tradução de linguagem natural para SQL ou geração automática de consultas. Em geral, os estudos não possuem como objetivo principal a compreensão estrutural completa de bancos de dados relacionais nem a identificação de relacionamentos implícitos entre entidades.

Além disso, muitas soluções baseadas em IA operam apenas sobre consultas ou descrições fornecidas pelo usuário, sem realizar uma análise aprofundada dos metadados estruturais do banco de dados. Consequentemente, aspectos como interpretação de atributos, identificação de padrões estruturais e detecção de inconsistências permanecem pouco explorados na literatura.

Nesse cenário, a proposta apresentada neste trabalho busca ampliar a utilização da IA para além da simples geração de consultas, empregando modelos de linguagem como mecanismo de apoio à compreensão estrutural de bancos de dados relacionais. A integração entre análise de metadados, representação em grafos e LLMs pretende fornecer uma visão mais abrangente da estrutura dos dados, contribuindo para reduzir o esforço necessário para sua exploração inicial.

Outro diferencial da abordagem proposta está na utilização da IA como mecanismo de apoio interpretativo, e não apenas como interface de consulta. Enquanto muitas aplicações baseadas em modelos de linguagem buscam transformar perguntas em comandos SQL, a proposta deste trabalho considera o uso desses modelos para auxiliar na leitura da estrutura do banco, na geração de descrições automáticas, na sugestão de significados para atributos e na identificação de possíveis relações ou inconsistências estruturais. Dessa forma, a IA passa a atuar como componente complementar à análise técnica do esquema relacional.

Essa perspectiva é relevante porque bancos de dados relacionais pouco documentados exigem mais do que a capacidade de consultar dados: exigem compreensão estrutural. Nesse sentido, a utilização de LLMs combinada com metadados e grafos pode contribuir para transformar informações técnicas dispersas em explicações mais acessíveis, favorecendo tanto usuários experientes quanto profissionais que estejam iniciando a exploração de uma base desconhecida.

3.4 Comparação entre as Soluções Existentes e a Proposta

Com base na análise das ferramentas e abordagens apresentadas nas seções anteriores, elaborou-se uma comparação entre as principais funcionalidades oferecidas pelas soluções existentes e aquelas previstas na proposta desenvolvida neste trabalho. O objetivo é evidenciar as características atualmente disponíveis no mercado e destacar os diferenciais da arquitetura proposta para a compreensão estrutural automatizada de bancos de dados relacionais.

Quadro 2 – Comparação entre soluções relacionadas e a proposta apresentada

Critério de comparação	Soluções / ferramentas analisadas			
	DBeaver / MySQL Workbench	Data Catalogs Tradicionais	Pesquisas Text-to-SQL com LLM	Esquemia
Abordagem principal	Administração e modelagem visual de bancos de dados relacionais.	Governança, catalogação e linhagem de dados.	Uso de linguagem natural para geração de consultas SQL.	Arquitetura conceitual para diagnóstico estrutural e semântico de bancos relacionais.
Mapeamento estrutural	Visualização por meio de diagramas entidade-relacionamento, quando os relacionamentos estão definidos.	Mapeamento parcial, com foco no inventário de ativos e no dicionário de dados.	Não possui foco principal no mapeamento estrutural do banco.	Representação da estrutura do banco por meio de grafo interativo e métricas de centralidade.
Interpretação semântica dos dados	Não realiza interpretação semântica automática, dependendo da análise manual do usuário.	Não realiza interpretação semântica automática, pois os metadados precisam ser cadastrados ou mantidos.	Realiza interpretação parcial do esquema, geralmente voltada à execução de consultas.	Uso de engenharia de prompts e LLMs para apoiar a inferência de significados.
Geração automática de documentação	Não gera documentação automática, apenas exibe a estrutura física atual do banco.	Centraliza a documentação, mas depende de processos manuais ou parcialmente automatizados.	Não possui como objetivo principal a geração de documentação estrutural.	Previsão de geração automática de descrições sobre tabelas, atributos e relacionamentos.
Diagnóstico estrutural e inconsistências	Não possui foco na detecção de inconsistências, nomenclaturas confusas ou chaves implícitas.	Não possui foco em diagnóstico estrutural, pois sua ênfase está na governança e rastreabilidade.	Não possui foco em diagnóstico estrutural, pois geralmente assume que o banco já está corretamente modelado.	Apoio à identificação de inconsistências, relacionamentos implícitos e problemas estruturais.
Relação com a proposta	Referência para visualização técnica de esquemas relacionais.	Referência para organização e governança de metadados.	Referência para o uso de LLMs na interação com bancos de dados.	Integra visualização estrutural, interpretação semântica e diagnóstico conceitual em uma única proposta.

Fonte: Autoria própria (2026).

O quadro 2 demonstra que as soluções atualmente disponíveis oferecem recursos im-

portantes para visualização e documentação de estruturas de dados. Entretanto, observa-se que funcionalidades relacionadas à interpretação automatizada de metadados, geração automática de descrições, identificação de inconsistências estruturais e utilização integrada de IA ainda são limitadas ou inexistentes na maioria das ferramentas analisadas.

Nesse contexto, a arquitetura proposta neste trabalho diferencia-se conceitualmente por prever a combinação de análise estrutural, representação baseada em grafos e apoio de LLMs. Essa combinação é apresentada como uma possibilidade arquitetural voltada à redução do esforço necessário para compreender bancos de dados relacionais pouco documentados e apoiar profissionais na exploração e interpretação das estruturas analisadas.

A comparação também evidencia que as ferramentas tradicionais são voltadas principalmente à visualização e administração dos bancos de dados, enquanto os *Data Catalogs* concentram-se na organização e governança dos metadados. A proposta deste trabalho, por sua vez, situa-se conceitualmente em uma etapa intermediária entre esses dois grupos: a etapa de compreensão estrutural automatizada. Essa etapa envolveria identificar a organização do banco, interpretar o significado dos elementos estruturais, representar os relacionamentos de forma visual e fornecer explicações que auxiliem o usuário na análise inicial.

Portanto, a proposta não busca substituir ferramentas consolidadas de administração, modelagem ou governança de dados. Ao contrário, apresenta uma arquitetura conceitual que poderia complementar essas soluções ao oferecer uma camada adicional de interpretação, documentação automática e apoio à compreensão de bases relacionais, especialmente quando a documentação disponível é insuficiente ou inexistente.

3.5 Síntese dos Trabalhos Relacionados

A análise das soluções estudadas demonstra que existem diferentes abordagens voltadas ao suporte de atividades relacionadas à administração, documentação e exploração de bancos de dados relacionais. Ferramentas de modelagem, como *DBeaver* e *MySQL Workbench*, oferecem recursos relevantes para visualização estrutural e gerenciamento dos esquemas, enquanto os *Data Catalogs* concentram-se na organização e governança dos metadados.

Observa-se também que algumas ferramentas relacionadas passaram a incorporar recursos de IA em seus ambientes. O *DBeaver*, por exemplo, disponibiliza recursos de assistência por IA voltados ao trabalho com bancos de dados em linguagem natural, incluindo criação e edição de consultas, resposta a dúvidas, apoio à correção de erros e exploração de objetos do banco

(DBeaver Corporation, s.d.b). De forma semelhante, soluções de catálogo e governança de dados têm avançado no uso de IA para apoiar a documentação e o enriquecimento de metadados. O *DataHub*, por exemplo, apresenta recursos de documentação gerada por IA para tabelas e colunas, enquanto o *Microsoft Purview* posiciona seus recursos de governança como apoio à visibilidade, confiança e inovação responsável em ambientes orientados por IA (DataHub Project, s.d.b; Microsoft Corporation, 2025).

Entretanto, observa-se que essas soluções ainda apresentam limitações significativas no que se refere à compreensão automatizada da estrutura dos dados. Em geral, a interpretação dos elementos identificados permanece dependente da experiência do usuário, da existência de documentação prévia ou da realização de análises manuais complementares.

Paralelamente, os avanços recentes em IA demonstram elevado potencial para auxiliar atividades relacionadas à engenharia de software e à análise de dados. Contudo, a literatura ainda apresenta quantidade limitada de estudos especificamente direcionados à compreensão estrutural automatizada de bancos de dados relacionais.

A comparação realizada evidencia que, embora algumas soluções já utilizem recursos de IA para tarefas específicas, nenhuma das soluções analisadas integra simultaneamente interpretação de metadados, geração automática de descrições, identificação de inconsistências estruturais, representação baseada em grafos e utilização de LLMs em uma arquitetura voltada especificamente à compreensão estrutural de bancos de dados relacionais.

Nesse contexto, a proposta apresentada neste trabalho diferencia-se por combinar análise estrutural, representação em grafos e IA em uma única arquitetura. Espera-se que essa integração contribua para reduzir o esforço necessário à exploração de bases desconhecidas, ampliar a compreensão dos relacionamentos existentes e facilitar atividades relacionadas à engenharia de dados, análise de dados e BI.

Dessa forma, os trabalhos relacionados analisados reforçam a pertinência da proposta desenvolvida nesta pesquisa. As ferramentas existentes oferecem contribuições relevantes, mas ainda não contemplam de forma integrada todos os aspectos necessários para apoiar a compreensão estrutural automatizada de bancos de dados relacionais. Assim, a proposta deste trabalho posiciona-se como uma alternativa conceitual voltada a preencher essa lacuna, articulando recursos de visualização, documentação, análise de metadados e apoio por modelos de linguagem.

4 METODOLOGIA

O presente trabalho caracteriza-se como uma pesquisa de natureza aplicada, uma vez que tem como objetivo propor e modelar uma solução tecnológica voltada à análise e à compreensão estrutural de bancos de dados relacionais. Esse tipo de pesquisa concentra-se na geração de conhecimentos com potencial de aplicação prática, direcionados à resolução de problemas específicos, alinhando-se diretamente à proposta deste estudo (Prodanov; Freitas, 2013).

Quanto aos objetivos, a pesquisa apresenta caráter exploratório e descritivo. É exploratória por buscar compreender as dificuldades enfrentadas por profissionais e estudantes da área de TI na interpretação de bancos de dados complexos e pouco documentados. Simultaneamente, assume caráter descritivo ao propor, estruturar e detalhar uma arquitetura voltada à análise automatizada de estruturas de bancos de dados relacionais, evidenciando seus componentes, funcionalidades previstas e formas de interação (Prodanov; Freitas, 2013).

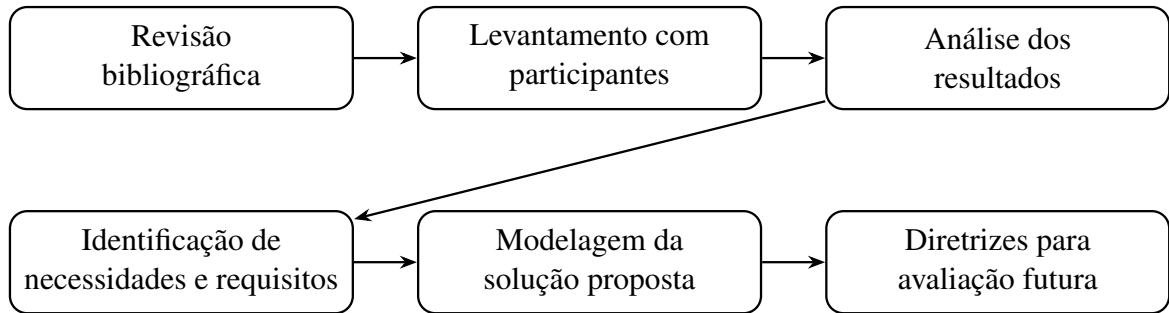
A abordagem adotada é mista, contemplando aspectos qualitativos e quantitativos. Sob a perspectiva qualitativa, o trabalho fundamenta-se na revisão bibliográfica e na modelagem da solução proposta, buscando compreender como técnicas de IA podem contribuir para a interpretação estrutural de bancos de dados relacionais. Sob a perspectiva quantitativa, são considerados os dados obtidos por meio do levantamento realizado com profissionais e estudantes da área de TI, permitindo analisar a relevância da problemática investigada e a percepção dos participantes quanto à utilidade potencial da solução proposta (Prodanov; Freitas, 2013).

A pesquisa fundamenta-se em revisão bibliográfica, abrangendo temas como bancos de dados relacionais, modelagem de dados, engenharia de dados, metadados, BI, teoria de grafos, IA, LLMs e privacidade de dados. Esses conceitos fornecem a base teórica necessária para a definição da arquitetura proposta e para a compreensão do problema investigado (Silberschatz; Korth; Sudarshan, 2019; Batini; Ceri; Navathe, 1992; Heuser, 2009; Kimball; Ross, 2013; Badman; Kosinski, 2024; Sharda; Delen; Turban, 2017; Cormen *et al.*, 2009; Russell; Norvig, 2021; Hong *et al.*, 2024; Brasil, 2018).

Com o objetivo de tornar mais claro o percurso metodológico adotado, a Figura 1 apresenta as principais etapas seguidas durante a pesquisa. Inicialmente, foi realizada a revisão bibliográfica, responsável por fundamentar teoricamente o problema e os conceitos utilizados no trabalho. Em seguida, foi aplicado o levantamento com profissionais e estudantes da área de TI, cujos resultados contribuíram para a identificação das principais dificuldades e requisitos relacionados à solução proposta. A partir dessa base teórica e empírica, foi realizada a modelagem

da solução, contemplando sua arquitetura, seus módulos e seus fluxos de funcionamento. Por fim, foram definidas diretrizes para avaliação futura da proposta.

Figura 1 – Percurso metodológico adotado na pesquisa



Fonte: Autoria própria (2026).

4.1 Levantamento com profissionais e estudantes da área de TI

A fim de avaliar a relevância da problemática investigada e compreender a percepção dos potenciais usuários da solução proposta, foi realizado um levantamento com 30 participantes, entre profissionais e estudantes da área de TI. Esse levantamento teve como finalidade complementar a revisão bibliográfica com evidências empíricas obtidas junto ao público-alvo da solução, permitindo relacionar os desafios discutidos na literatura com experiências práticas vivenciadas por pessoas que lidam direta ou indiretamente com dados.

O levantamento foi aplicado por meio de formulário eletrônico, utilizando amostragem por conveniência, contemplando participantes com vínculo acadêmico ou profissional relacionado à área de TI. A divulgação do formulário ocorreu por meio de canais digitais, incluindo grupos de alunos da instituição, e-mails acadêmicos, aplicativos de mensagens e contatos acadêmicos e profissionais do autor. Foram consideradas válidas apenas as respostas completas de participantes que declararam atuar, estudar ou possuir familiaridade com temas relacionados a bancos de dados, análise de dados, desenvolvimento de sistemas, BI, engenharia de dados ou áreas correlatas. Foram desconsideradas respostas incompletas ou que não apresentassem relação mínima com o público-alvo da pesquisa.

Antes do preenchimento do questionário, os participantes foram informados sobre a finalidade acadêmica da pesquisa, o caráter voluntário da participação, a ausência de identificação pessoal obrigatória e o uso dos dados exclusivamente para fins de análise agregada. Dessa forma, o envio das respostas foi considerado como manifestação de consentimento livre e esclarecido

para participação no estudo. Os dados foram coletados de forma anônima no período de 26 de maio de 2026 a 1º de junho de 2026, sendo analisados em conformidade com os princípios estabelecidos pela LGPD (Brasil, 2018).

O instrumento de coleta contemplou questões relacionadas ao perfil dos participantes, à experiência com bancos de dados relacionais, às dificuldades encontradas durante a exploração de bases de dados desconhecidas, ao tempo necessário para compreender estruturas pouco documentadas e à percepção acerca da utilidade da solução proposta neste trabalho. As questões também buscaram identificar funcionalidades consideradas relevantes pelos respondentes, bem como preocupações relacionadas à privacidade, segurança da informação e possibilidade de execução local dos modelos de IA.

O questionário foi elaborado no âmbito desta pesquisa, com o objetivo de levantar percepções, dificuldades e expectativas de profissionais e estudantes da área de TI em relação à compreensão estrutural de bancos de dados relacionais. Sua construção foi orientada pelos objetivos do trabalho e pelos temas discutidos na revisão bibliográfica, especialmente aqueles relacionados à ausência de documentação, à compreensão de metadados, à interpretação de relacionamentos entre tabelas, à governança de dados, à qualidade estrutural das bases e ao uso de IA como apoio à análise. Estudos relacionados à gestão de metadados, catálogos de dados e qualidade de dados também reforçam a importância de investigar dificuldades de acesso, compreensão, documentação e confiabilidade das informações em ambientes orientados a dados (Subramaniam *et al.*, 2021; Ehrlinger; Rusz; Wöss, 2019; Papastergios; Gounaris, 2024).

Cabe destacar que o questionário aplicado não apresentou telas, protótipos ou representações visuais da solução aos participantes. O levantamento teve como foco identificar dificuldades, percepções e funcionalidades consideradas relevantes, enquanto o protótipo visual da interface foi elaborado posteriormente, com base nos requisitos observados a partir dos resultados obtidos.

O questionário completo aplicado aos participantes encontra-se disponível no Apêndice A, incluindo perguntas complementares que contribuíram para a contextualização geral da pesquisa, mas que não foram detalhadas individualmente no capítulo de resultados. Embora o instrumento contemple um conjunto mais amplo de perguntas, a análise discutida posteriormente concentra-se nas questões mais diretamente relacionadas à problemática da pesquisa e aos requisitos da solução proposta. Essa delimitação tem como objetivo manter o foco nos dados que contribuem diretamente para a fundamentação da proposta, sem comprometer a apresentação integral do instrumento utilizado.

As respostas fechadas foram analisadas por meio de estatística descritiva simples, considerando frequências absolutas e percentuais, com o objetivo de identificar tendências gerais nas percepções dos participantes. Já as respostas abertas foram utilizadas de forma complementar, permitindo observar percepções recorrentes, justificativas e comentários qualitativos relacionados às dificuldades enfrentadas na compreensão de bancos de dados pouco documentados, sem identificação individual dos respondentes.

A caracterização detalhada do perfil dos participantes e a análise dos resultados obtidos são apresentadas no Capítulo 5.

4.2 Modelagem da solução proposta

No que se refere à concepção da solução proposta, adotou-se uma abordagem fundamentada nos princípios da Engenharia de Software, estruturada em etapas que incluem levantamento de requisitos, modelagem arquitetural e especificação das funcionalidades previstas. Inicialmente, foram identificadas as principais necessidades dos usuários, especialmente profissionais da área de dados, com ênfase nas dificuldades relacionadas à compreensão estrutural de bases complexas e pouco documentadas.

A proposta apresentada neste trabalho consiste na modelagem de uma solução tecnológica voltada à análise estrutural automatizada de bancos de dados relacionais. Para isso, foi definida uma arquitetura capaz de realizar a extração e o processamento de metadados estruturais, incluindo tabelas, atributos, chaves primárias, chaves estrangeiras e relacionamentos. A partir dessas informações, a solução proposta busca organizar e representar a estrutura do banco de dados de forma clara e intuitiva, utilizando abordagens baseadas em grafos e diagramas estruturais para apoiar a compreensão dos dados por parte dos usuários (Silberschatz; Korth; Sudarshan, 2019; Cormen *et al.*, 2009; Hagberg; Schult; Swart, 2008).

Na etapa de modelagem, são utilizados diagramas para representar a organização geral da solução e as interações previstas entre o usuário e o sistema proposto. O Diagrama de Fluxo de Dados é utilizado para demonstrar o caminho percorrido pelos metadados desde a extração até a apresentação dos resultados, enquanto o Diagrama de Casos de Uso representa as principais funcionalidades previstas para o usuário. Esses artefatos auxiliam na visualização da arquitetura proposta e tornam mais clara a relação entre os componentes responsáveis pela extração, processamento, interpretação e apresentação das informações.

Como parte da arquitetura proposta, prevê-se a utilização de técnicas de IA para auxiliar

na interpretação das estruturas identificadas. Nesse contexto, LLMs poderão ser empregados para gerar descrições automáticas das tabelas, sugerir significados para atributos, identificar padrões estruturais e apontar possíveis inconsistências a partir da análise dos metadados extraídos (Russell; Norvig, 2021; Géron, 2019; White *et al.*, 2023; Hong *et al.*, 2024). Essa abordagem tem como objetivo ampliar a capacidade de compreensão de bases de dados complexas e pouco documentadas, reduzindo o esforço necessário para sua exploração inicial.

No que se refere à privacidade e à segurança da informação, a arquitetura proposta foi concebida em conformidade com os princípios estabelecidos pela LGPD. Nesse contexto, prevê-se a adoção de estratégias que limitem a análise aos metadados estruturais do banco de dados, evitando o envio de registros de dados, informações pessoais ou dados sensíveis para serviços externos de IA (Brasil, 2018; Cavoukian, 2009; ISO; IEC, 2022).

4.3 Diretrizes para avaliação futura

Embora o escopo deste trabalho esteja concentrado na proposição e modelagem da solução, considera-se relevante estabelecer diretrizes para sua validação em estudos futuros. Nesse contexto, propõe-se a realização de avaliações comparativas entre a abordagem tradicional de exploração manual de bancos de dados e a utilização da arquitetura proposta neste trabalho, com o objetivo de analisar sua contribuição para a compreensão estrutural de bases de dados relacionais.

A avaliação futura poderá considerar métricas quantitativas e qualitativas, permitindo verificar não apenas ganhos de tempo, mas também aspectos relacionados à clareza das informações apresentadas, à identificação de relacionamentos, à interpretação dos atributos e à percepção dos usuários quanto à utilidade da solução. O Quadro 3 apresenta uma síntese das principais métricas sugeridas para essa etapa de validação.

Quadro 3 – Métricas sugeridas para avaliação futura da solução proposta

Métrica	Forma de avaliação	Resultado esperado
Tempo de compreensão da base	Comparação entre o tempo gasto na análise manual e o tempo gasto com o apoio da solução proposta.	Redução do tempo necessário para compreender a estrutura inicial do banco de dados.
Relacionamentos identificados	Verificação da quantidade e da qualidade dos relacionamentos identificados entre tabelas, incluindo chaves explícitas e possíveis relações implícitas.	Maior facilidade na identificação da organização estrutural da base.
Interpretação de atributos	Avaliação da clareza das descrições geradas para tabelas e atributos, por meio da percepção dos usuários.	Melhor compreensão do significado dos elementos presentes no banco de dados.
Clareza das descrições geradas	Análise qualitativa das descrições produzidas pela inteligência artificial, considerando coerência, objetividade e utilidade prática.	Geração de explicações mais acessíveis e úteis para a exploração da base.
Identificação de inconsistências	Verificação da capacidade da solução em indicar possíveis padrões inadequados, nomenclaturas ambíguas ou ausências de relacionamentos esperados.	Apoio à identificação de problemas estruturais ou pontos que exigem revisão técnica.
Usabilidade percebida	Aplicação de questionário ou formulário de avaliação com usuários após o uso da solução.	Verificação da facilidade de uso, compreensão da interface e aceitação da solução proposta.
Potencial de apoio à criação de dashboards	Avaliação da percepção dos usuários quanto à utilidade da solução para apoiar atividades de análise, modelagem e construção de indicadores.	Maior agilidade nas etapas iniciais de exploração dos dados e geração de insights.

Fonte: Autoria própria (2026).

A definição dessas diretrizes busca estabelecer uma base metodológica para futuras validações experimentais da proposta, possibilitando a obtenção de evidências quantitativas e qualitativas acerca de sua aplicabilidade. Dessa forma, estudos posteriores poderão verificar, de maneira mais objetiva, se a solução contribui para reduzir o esforço inicial de exploração de bancos de dados relacionais, melhorar a compreensão estrutural das bases analisadas e apoiar profissionais que atuam em atividades relacionadas à análise, documentação e interpretação de dados.

5 RESULTADOS

5.1 Levantamento e Análise dos Dados

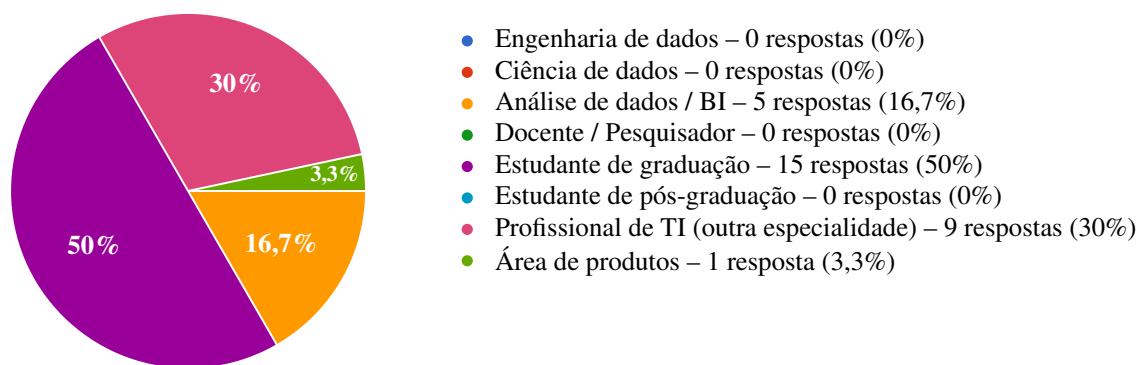
A relevância da solução proposta foi analisada por meio de um levantamento realizado com profissionais e estudantes da área de TI. Os resultados obtidos permitiram compreender a dimensão do problema investigado, identificar as principais dificuldades enfrentadas na exploração de bancos de dados relacionais e avaliar a percepção dos participantes quanto à utilidade da solução proposta.

Neste capítulo, são discutidos os resultados das questões mais diretamente relacionadas à problemática investigada e aos requisitos da solução proposta.

5.1.1 Perfil dos Participantes

A pesquisa contou com a participação de estudantes e profissionais com diferentes perfis de atuação relacionados à área de TI. Conforme apresentado na Figura 2, a maior parte dos participantes corresponde a estudantes de graduação, representando 50% da amostra. Em seguida, aparecem profissionais de TI de outras especialidades, com 30%, profissionais de análise de dados e BI, com 16,7%, e participantes da área de produtos, com 3,3%.

Figura 2 – Perfil dos participantes da pesquisa



Fonte: Autoria própria (2026).

Essa composição demonstra que o levantamento contemplou um público misto, formado tanto por indivíduos em processo de formação quanto por participantes com experiência profissional em áreas relacionadas à tecnologia. Entretanto, observa-se que a amostra é majoritariamente composta por estudantes de graduação e profissionais de TI de outras especialidades, enquanto

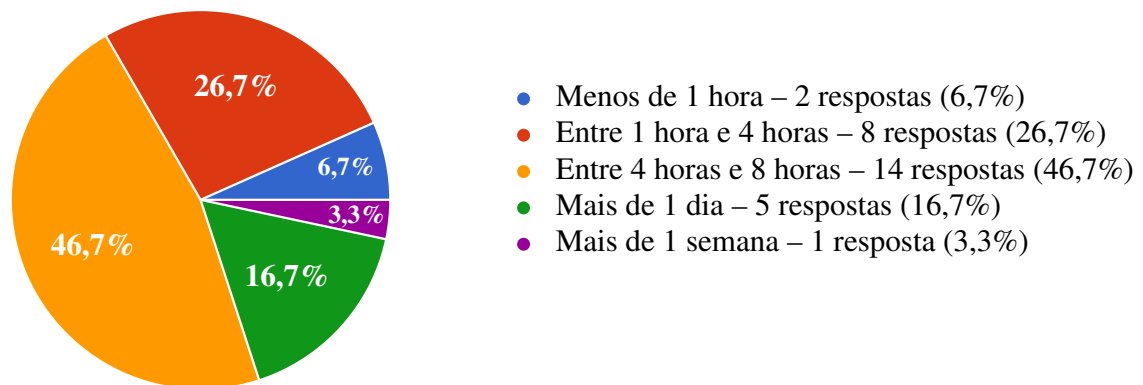
os participantes diretamente vinculados à análise de dados e BI representam uma parcela menor do total.

A diversidade da amostra contribui para observar a problemática sob diferentes níveis de maturidade técnica, envolvendo tanto usuários em formação quanto profissionais que atuam em contextos tecnológicos. Essa característica reforça a relevância exploratória do levantamento, ao indicar que desafios como ausência de documentação, dificuldade de interpretação de relacionamentos e compreensão de estruturas pouco padronizadas podem ser percebidos por diferentes perfis dentro da área de TI.

5.1.2 Tempo Necessário para Compreensão de Bases de Dados Desconhecidas

Um dos principais desafios identificados está relacionado ao tempo necessário para compreender a estrutura de uma base de dados desconhecida. Conforme ilustrado na Figura 3, 46,7% dos participantes afirmaram demandar entre quatro e oito horas para realizar essa atividade, enquanto 16,7% relataram necessitar de mais de um dia de trabalho.

Figura 3 – Tempo necessário para compreender a estrutura de um banco de dados desconhecido



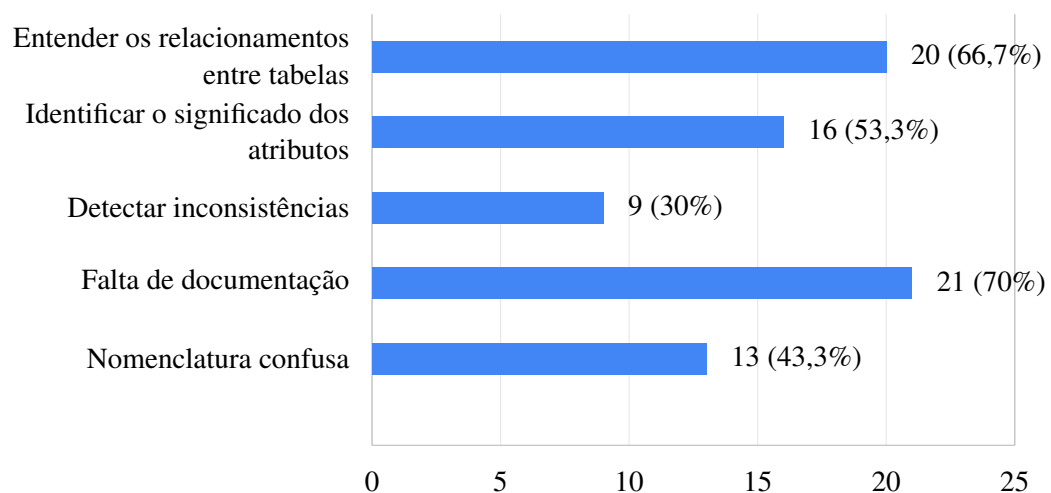
Fonte: Autoria própria (2026).

Esses dados indicam que a compreensão inicial de uma base desconhecida não representa uma tarefa pontual ou de baixo custo operacional. Ao contrário, trata-se de uma etapa que pode consumir parcela significativa do tempo disponível em projetos de dados, comprometendo prazos e reduzindo o tempo disponível para atividades analíticas de maior valor agregado.

5.1.3 Principais Dificuldades Encontradas

Os resultados evidenciam que a ausência de documentação adequada representa a principal dificuldade enfrentada pelos participantes durante a exploração de bancos de dados relacionais. Conforme apresentado na Figura 4, destacam-se a ausência de documentação adequada (70%), a complexidade na compreensão dos relacionamentos entre tabelas (66,7%) e a dificuldade em identificar o significado dos atributos presentes na base (53,3%).

Figura 4 – Principais dificuldades encontradas na exploração de bancos de dados relacionais



Fonte: Autoria própria (2026).

A proximidade entre os percentuais relacionados à ausência de documentação adequada (70%) e à dificuldade de compreender os relacionamentos entre tabelas (66,7%) indica que ambos os problemas são recorrentes na experiência dos participantes. Esses resultados sugerem que a exploração de bancos de dados desconhecidos envolve tanto dificuldades relacionadas à disponibilidade de informações estruturais quanto desafios associados à interpretação das conexões existentes entre as tabelas. Entretanto, como não foram realizados cruzamentos estatísticos entre as respostas individuais, não é possível afirmar que exista uma relação causal ou uma correlação direta entre essas dificuldades.

Quando esses resultados são considerados juntamente com o tempo relatado para a compreensão de bases desconhecidas, observa-se que a exploração inicial da estrutura dos dados pode representar uma atividade relevante no desenvolvimento de projetos. Os dados não permitem determinar exatamente quais fatores são responsáveis pelo tempo despendido por cada participante, mas indicam que a ausência de documentação e a dificuldade de interpretar relacionamentos

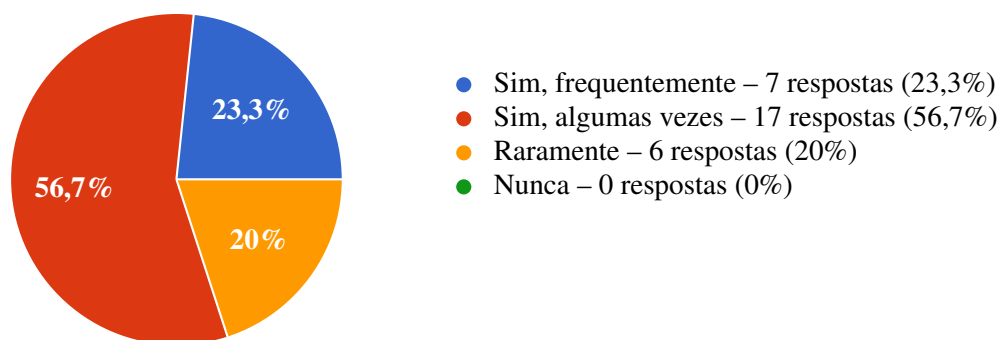
estão entre os principais obstáculos percebidos. Esse cenário reforça a pertinência de soluções que auxiliem na apresentação e na compreensão da estrutura de bancos de dados relacionais, reduzindo potencialmente o esforço necessário durante sua exploração inicial.

Além disso, 53,3% dos participantes relataram dificuldade em compreender o significado dos atributos presentes na base. Quando esses resultados são analisados em conjunto, percebe-se que os desafios enfrentados pelos profissionais não estão restritos à estrutura dos dados, mas também à sua interpretação semântica. Isso sugere que soluções voltadas apenas para visualização gráfica não seriam suficientes, sendo necessário fornecer mecanismos complementares de documentação e interpretação automática dos elementos identificados — o que justifica a inclusão do módulo de IA na arquitetura proposta.

5.1.4 Impactos do Esforço de Compreensão Estrutural

Além das dificuldades técnicas identificadas, os resultados indicaram que o esforço necessário para compreender a estrutura dos dados pode estar associado a impactos no desenvolvimento das atividades dos participantes. Conforme ilustrado na Figura 5, 80% dos respondentes relataram que a necessidade de compreender previamente a estrutura dos dados já ocasionou atrasos em suas demandas.

Figura 5 – Impacto do esforço necessário para compreender a estrutura dos dados no desenvolvimento das atividades



Fonte: Autoria própria (2026).

Esse resultado reforça a percepção de que as dificuldades técnicas identificadas no levantamento podem produzir efeitos práticos no desenvolvimento das atividades dos participantes. Observa-se que 70% dos respondentes apontaram a ausência de documentação como uma das principais dificuldades encontradas na exploração de bancos de dados relacionais. De forma

complementar, 80% afirmaram que já sofreram atrasos em suas atividades devido à necessidade de compreender previamente a estrutura dos dados.

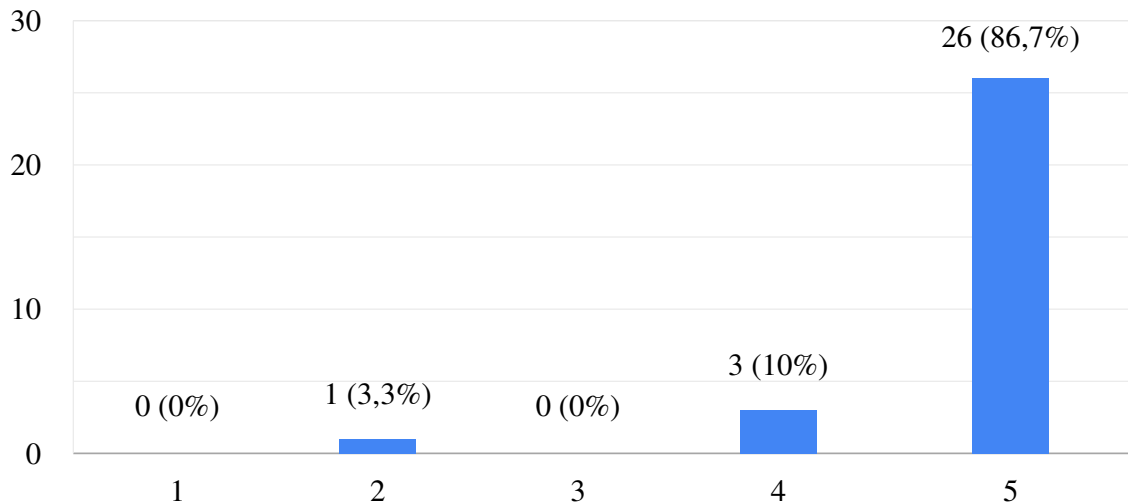
Esse resultado sugere alinhamento entre a percepção de ausência de documentação e os atrasos relatados, embora o levantamento não tenha realizado uma análise estatística de correlação entre as respostas individuais. Assim, não se pode afirmar uma relação causal direta entre esses fatores, mas os dados indicam que a falta de documentação e o esforço necessário para interpretação da estrutura dos bancos de dados aparecem como aspectos relevantes na experiência dos participantes. Essa constatação reforça a importância de soluções que auxiliem na compreensão estrutural de bases relacionais, especialmente em cenários nos quais a documentação técnica é inexistente, incompleta ou desatualizada.

A análise conjunta desses resultados sugere que as dificuldades relacionadas à documentação e à compreensão estrutural podem estar associadas a impactos na produtividade e nos prazos das atividades desenvolvidas pelos participantes. Entretanto, como o levantamento se baseou em percepções gerais e não realizou cruzamentos estatísticos entre as respostas individuais, não é possível determinar uma relação causal direta entre esses fatores. Nesse contexto, a automatização da geração de documentação estrutural, prevista no módulo de IA da arquitetura proposta, apresenta-se como uma alternativa potencial para apoiar a compreensão de bancos de dados pouco documentados. Sua efetividade na redução de atrasos e do esforço de exploração deverá ser verificada posteriormente por meio da implementação e da avaliação experimental da solução.

5.1.5 Percepção sobre a Utilidade da Solução

Quando questionados sobre a utilidade potencial de uma ferramenta capaz de analisar automaticamente a estrutura de bancos de dados relacionais, os participantes sugeriram elevada receptividade. Conforme apresentado na Figura 6, 96,7% atribuíram notas 4 ou 5 em uma escala de 1 a 5.

Figura 6 – Percepção dos participantes sobre a utilidade da solução proposta



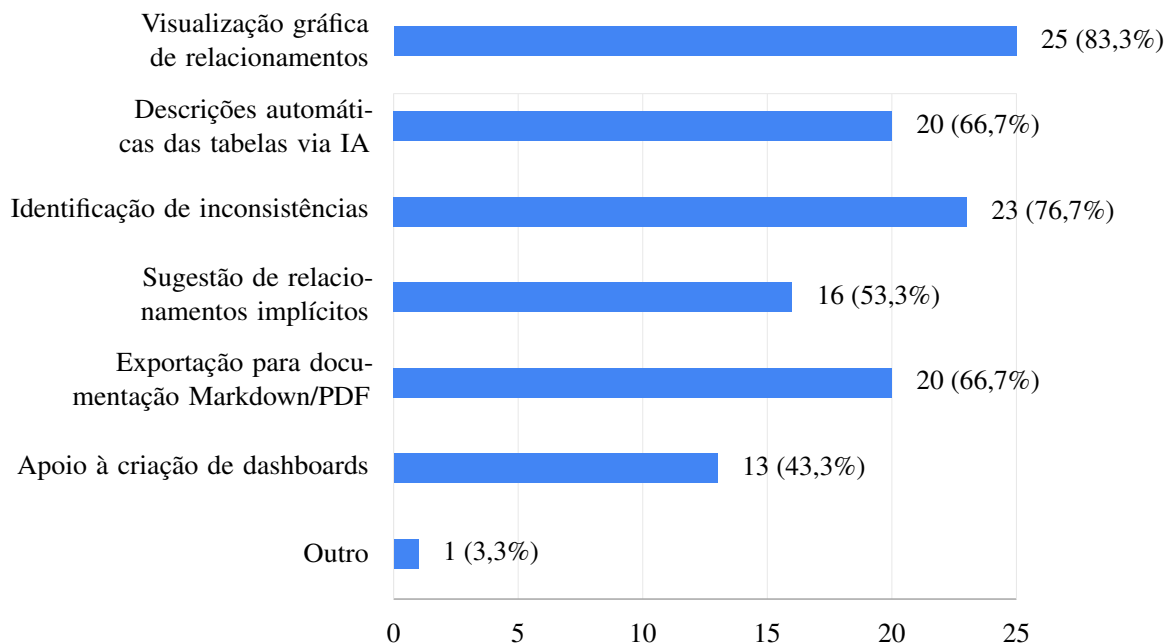
Fonte: Autoria própria (2026).

A elevada avaliação da utilidade da solução deve ser interpretada à luz dos resultados anteriores. O índice de aprovação de 96,7% não representa apenas uma opinião favorável à tecnologia proposta, mas reflete uma necessidade concreta identificada ao longo da pesquisa. Ao relacionar esse resultado com os 80% dos participantes que relataram atrasos em atividades profissionais e com os 70% que apontaram a falta de documentação como um problema recorrente, percebe-se que a receptividade à solução decorre da existência de uma dificuldade real e amplamente compartilhada. Em outras palavras, a aceitação observada não é abstrata: ela é proporcional à intensidade do problema que a solução se propõe a resolver.

5.1.6 Funcionalidades Consideradas Relevantes

Os participantes também indicaram quais funcionalidades consideraram mais relevantes para uma ferramenta voltada à compreensão estrutural de bancos de dados. Conforme apresentado na Figura 7, os resultados sugerem interesse prioritário pela visualização gráfica de relacionamentos (83,3%), seguida da identificação de inconsistências (76,7%) e das descrições automáticas das tabelas (66,7%).

Figura 7 – Funcionalidades consideradas mais relevantes pelos participantes



Fonte: Autoria própria (2026).

Esses resultados dialogam diretamente com as dificuldades apresentadas anteriormente. A visualização gráfica de relacionamentos foi apontada como a funcionalidade mais relevante por 83,3% dos participantes — percentual superior ao registrado para a própria dificuldade de compreender relacionamentos entre tabelas (66,7%). Esse comportamento sugere que os respondentes não apenas reconhecem a existência do problema, mas já projetam a visualização gráfica como sua principal forma de resolução, o que confere à solução proposta respaldo tanto pela demanda quanto pela forma como os usuários esperam ser atendidos.

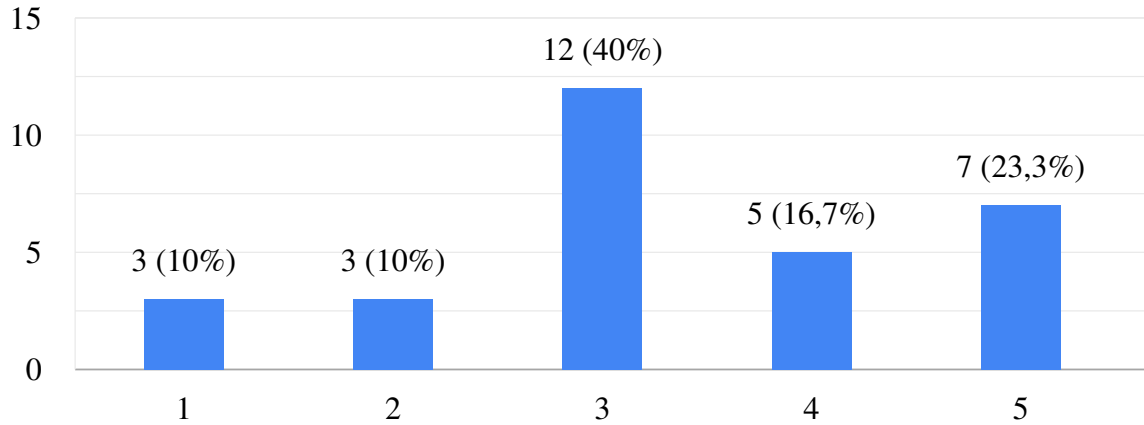
De maneira semelhante, o interesse pela identificação de inconsistências (76,7%) e pelas descrições automáticas das tabelas (66,7%) está alinhado às dificuldades relacionadas à ausência de documentação e à interpretação semântica dos atributos. Isso sugere que as funcionalidades previstas na arquitetura proposta atendem diretamente às necessidades apontadas pelos participantes, fortalecendo a aderência da solução ao problema investigado.

5.1.7 Privacidade e Segurança da Informação

Aspectos relacionados à privacidade também se mostraram relevantes no contexto da adoção da solução proposta. Conforme apresentado na Figura 8, a maioria dos participantes demonstrou algum grau de preocupação quanto ao envio de metadados para serviços externos de

IA, com 80% atribuindo notas 3 ou superior.

Figura 8 – Preocupação dos participantes quanto ao envio de metadados para serviços externos



Fonte: Autoria própria (2026).

Esse resultado revela que aspectos relacionados à privacidade e à segurança da informação devem ser considerados como requisitos relevantes para a adoção da solução proposta. Quando observados de forma agregada, os resultados indicam simultaneamente alta percepção de utilidade da ferramenta e preocupação relevante quanto ao envio de metadados para serviços externos de IA. Assim, embora os participantes tenham demonstrado receptividade à proposta, os dados também evidenciam que a adoção de soluções baseadas em IA depende de garantias quanto ao controle, à segurança e à confidencialidade das informações processadas.

Dessa forma, os benefícios funcionais proporcionados pela IA precisam ser acompanhados por práticas de governança, transparência e conformidade, especialmente em contextos nos quais metadados estruturais possam revelar aspectos sensíveis da organização dos sistemas analisados. Nesse sentido, a arquitetura da solução deve considerar mecanismos que reduzam riscos relacionados ao compartilhamento externo de informações, como minimização dos metadados enviados, anonimização de elementos sensíveis, definição clara das informações processadas e possibilidade de execução local ou controlada dos modelos utilizados.

5.1.8 Execução Local dos Modelos

Como consequência das preocupações relacionadas à privacidade, investigou-se a influência da disponibilidade de execução local dos modelos de IA na disposição dos participantes em adotar a solução proposta. Conforme apresentado na Figura 9, 60% dos participantes afirmaram

que a disponibilidade desse recurso aumentaria significativamente sua disposição em utilizar a ferramenta.

Figura 9 – Influência da execução local na adoção da solução proposta



Fonte: Autoria própria (2026).

Os resultados relacionados à execução local complementam as observações obtidas na análise de privacidade. Enquanto 80% dos participantes atribuíram notas 3 ou superiores quanto à preocupação com o envio de metadados para serviços externos, 60% afirmaram que a possibilidade de execução local aumentaria significativamente sua disposição para utilizar a ferramenta. Esse resultado sugere que parte da resistência pode estar mais relacionada à forma de processamento, envio e armazenamento dos metadados do que ao uso de IA em si.

Dessa maneira, a adoção de modelos executados localmente pode representar não apenas uma alternativa técnica da arquitetura, mas também um fator estratégico para aumentar a confiança dos usuários e ampliar a aceitação da solução em ambientes organizacionais com requisitos mais rigorosos de segurança e conformidade. Esse resultado orienta decisões de projeto relacionadas ao suporte a modelos de linguagem executados localmente, como alternativa ou complemento à dependência exclusiva de APIs externas.

5.1.9 Síntese dos Resultados

A análise conjunta dos resultados permite identificar uma convergência consistente entre as dificuldades relatadas pelos participantes, as lacunas identificadas na literatura e as funcionalidades previstas na arquitetura proposta. As três principais dificuldades apontadas, ausência de documentação (70%), dificuldade na compreensão de relacionamentos (66,7%) e dificuldade na interpretação de atributos (53,3%), correspondem diretamente aos três componentes centrais da

solução: o módulo de IA para geração de descrições automáticas, o módulo de visualização para representação gráfica dos relacionamentos e o módulo de extração para mapeamento estrutural da base.

Essa convergência é reforçada pela elevada percepção de utilidade potencial (96,7%), que não reflete apenas uma opinião favorável à tecnologia, mas o reconhecimento de uma necessidade concreta cujos efeitos práticos, evidenciados pelos atrasos relatados por 80% dos participantes, conferem urgência à sua resolução. Ao mesmo tempo, a preocupação com privacidade, indicada por 80% dos participantes que atribuíram notas 3 ou superiores, e a preferência pela execução local (60%) fornecem diretrizes claras para decisões arquiteturais futuras, indicando que a aceitação plena da solução depende não apenas de sua capacidade funcional, mas também da confiança que ela é capaz de gerar nos contextos em que será utilizada.

Portanto, os resultados obtidos reforçam a relevância percebida do problema investigado e sugerem que a proposta concebida apresenta aderência potencial às necessidades identificadas junto aos participantes. Contudo, sua utilidade, aceitação e efetividade ainda dependem da implementação da arquitetura e de posterior validação experimental.

Ressalta-se que, por se tratar de uma amostra por conveniência e com número limitado de participantes, os resultados não devem ser interpretados como generalizáveis a toda a população de profissionais da área de dados. Ainda assim, os dados obtidos fornecem indícios relevantes sobre dificuldades recorrentes na compreensão de bancos de dados relacionais pouco documentados e contribuem para fundamentar a pertinência da solução proposta.

6 PROPOSTA DA SOLUÇÃO

6.1 Visão Geral da Solução

A solução proposta neste trabalho tem como objetivo auxiliar profissionais da área de dados na compreensão estrutural de bancos de dados relacionais por meio da análise automatizada de metadados e da utilização de técnicas de IA. A proposta foi concebida para atuar em cenários nos quais a documentação técnica é inexistente, insuficiente ou desatualizada, reduzindo o esforço necessário para compreender a organização dos dados, seus relacionamentos e possíveis regras estruturais presentes no esquema analisado.

É importante destacar que o escopo deste trabalho limita-se à proposição e à modelagem de uma solução, não contemplando a implementação completa, implantação em ambiente real ou validação experimental de um sistema funcional. Dessa forma, os componentes, tecnologias e fluxos apresentados neste capítulo devem ser compreendidos como elementos de uma arquitetura proposta, servindo como base conceitual e técnica para futuras etapas de desenvolvimento e avaliação.

A necessidade dessa abordagem é evidenciada pelos resultados do levantamento realizado com profissionais e estudantes da área de TI, detalhados na Seção 5.1. Os dados obtidos indicaram que parte significativa dos participantes enfrenta dificuldades ao lidar com bases de dados desconhecidas, especialmente quando não há documentação adequada, quando os relacionamentos entre tabelas não estão claramente definidos ou quando os atributos possuem nomenclaturas pouco descritivas. Esses resultados reforçam que o problema investigado possui impacto direto na produtividade das equipes e na qualidade das entregas em projetos orientados a dados.

A arquitetura proposta contempla mecanismos para extração de metadados, modelagem estrutural baseada em grafos, geração automatizada de descrições por meio de LLMs e disponibilização das informações em uma interface visual interativa. Dessa forma, busca-se fornecer aos usuários uma visão consolidada da estrutura do banco de dados, facilitando atividades de exploração, análise, documentação e construção de soluções orientadas a dados.

A proposta foi concebida com foco na redução do tempo necessário para a compreensão de estruturas de dados desconhecidas, promovendo a geração automatizada de conhecimento sobre o esquema analisado. Ao integrar técnicas de extração de metadados, representação por grafos e IA, busca-se fornecer uma visão abrangente da base de dados sem exigir conhecimento

prévio do domínio de negócio ou acesso à documentação existente.

Além de apresentar potencial para contribuir com atividades de análise, engenharia de dados e BI, a solução proposta também foi concebida considerando sua possível aplicação em contextos educacionais, especialmente como recurso de apoio ao ensino de modelagem relacional e à interpretação de estruturas de bancos de dados reais. Como o presente trabalho se limita à proposição e modelagem da solução, essa aplicação educacional deve ser compreendida como uma contribuição esperada a ser avaliada em estudos futuros. Dessa forma, a proposta busca atuar tanto como mecanismo de apoio à produtividade quanto como recurso complementar ao processo de aprendizagem.

Nesse sentido, a solução proposta não tem como finalidade substituir ferramentas consolidadas de administração, consulta ou modelagem de bancos de dados, mas complementar essas abordagens por meio de uma camada adicional de interpretação estrutural. Enquanto ferramentas tradicionais normalmente apresentam os elementos técnicos do banco, como tabelas, colunas e relacionamentos, a proposta deste trabalho busca transformar essas informações em representações visuais e descrições mais acessíveis, contribuindo para que o usuário compreenda não apenas a existência dos elementos, mas também seu possível papel dentro da estrutura analisada.

Essa perspectiva reforça o caráter de apoio da solução, uma vez que a interpretação automatizada dos metadados pode reduzir o esforço inicial de exploração, mas não elimina a necessidade de validação humana. Assim, a proposta posiciona a IA como um recurso complementar ao trabalho do profissional de dados, capaz de acelerar a compreensão inicial, sugerir interpretações e auxiliar na documentação, sem substituir o julgamento técnico necessário em ambientes organizacionais complexos.

6.2 Protótipo Visual da Interface Proposta

Com o objetivo de facilitar a compreensão inicial da proposta, antes do detalhamento do fluxo de funcionamento e da arquitetura da solução, foram elaboradas telas conceituais da interface proposta. Essas telas não representam uma implementação funcional ou validada experimentalmente, mas uma representação visual de como os principais recursos previstos poderiam ser organizados em uma aplicação web.

Ressalta-se que as telas conceituais não foram apresentadas aos participantes no questionário aplicado durante o levantamento. O instrumento de coleta teve como finalidade identificar

dificuldades, percepções e funcionalidades consideradas relevantes pelos respondentes. A proposta visual foi elaborada posteriormente, a partir dos requisitos e necessidades identificados no levantamento, funcionando como uma materialização conceitual da solução proposta.

O objetivo do protótipo visual é demonstrar como o usuário poderia interagir com a solução para configurar a conexão com o banco de dados, visualizar a estrutura relacional em formato de grafo, consultar informações detalhadas sobre tabelas e atributos, analisar possíveis relacionamentos inferidos e observar diagnósticos estruturais gerados com apoio de IA.

As telas foram organizadas de forma a evidenciar os principais componentes funcionais da arquitetura proposta, mantendo coerência com os requisitos identificados no levantamento e com as funcionalidades previstas para a solução. Os dados exibidos nas telas possuem caráter ilustrativo e foram utilizados apenas para representar o comportamento esperado da solução em uma futura implementação.

Figura 10 – Tela inicial conceitual da solução proposta

Esquemia

PROPOSTA – TCC

Análise estrutural automatizada de bancos relacionais.

O **Esquemia** extrai metadados de um banco de dados, modela o esquema como grafo e gera descrições automáticas com IA para facilitar a documentação e a compreensão estrutural.



Grafo estrutural

Tabelas e relacionamentos visualizados como grafo interativo.



Descrições por IA

Explicações automáticas em linguagem natural para tabelas e atributos.



Conforme à LGPD

Análise limitada a metadados; nenhum dado pessoal é acessado.

Conectar um banco de demonstração

Preencha os dados essenciais para iniciar a análise estrutural.

PostgreSQLMySQL

HOSTDATABASE

localhostufersa_db

USUÁRIO

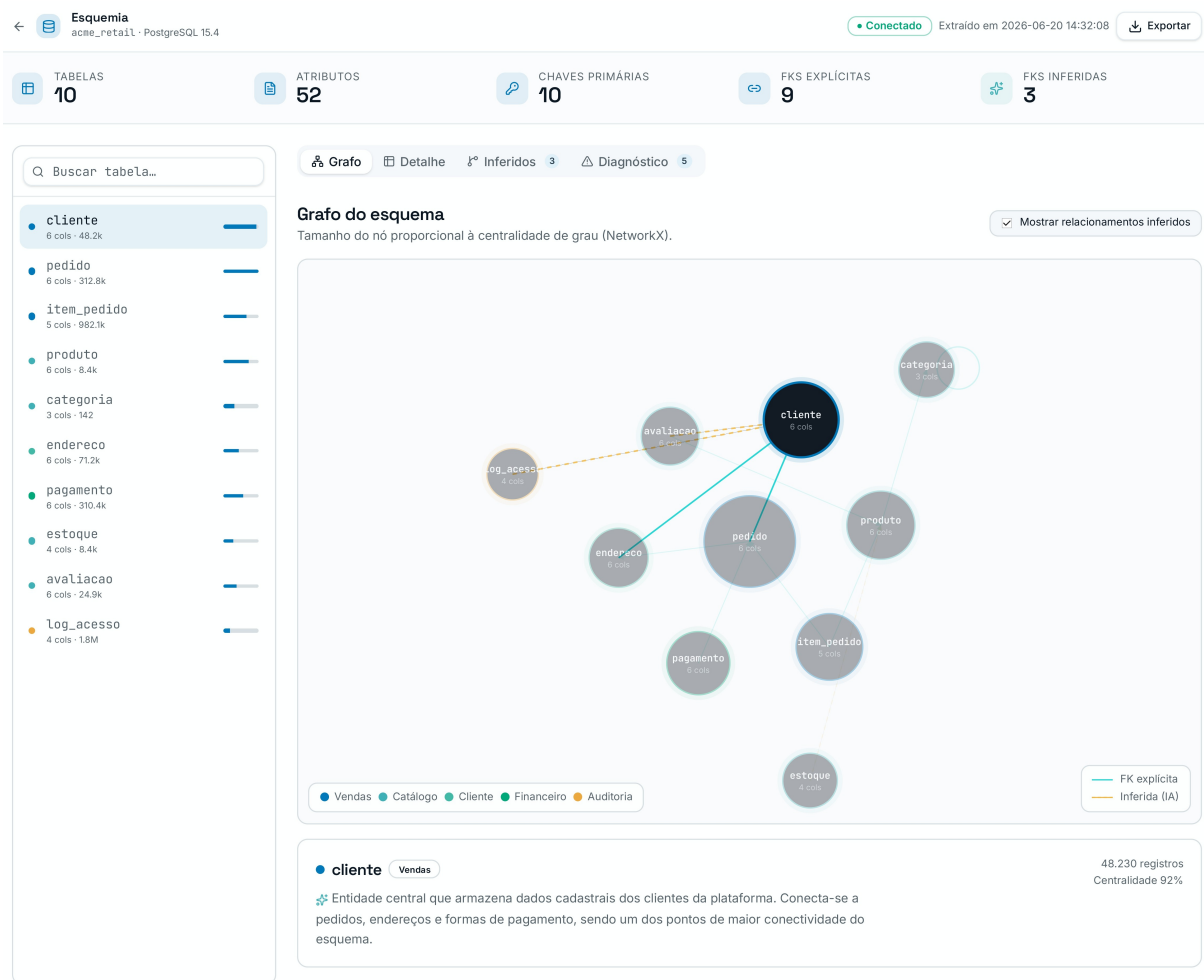
alunos_ufersa

Conectar e analisar →

Fonte: Adaptado pela Inteligência Artificial Lovable (2026).

A Figura 10 apresenta a tela inicial conceitual da solução proposta. Nessa representação, o usuário teria acesso a uma visão resumida dos principais pilares do sistema, como a visualização estrutural por grafos, a geração de descrições com apoio de IA e a preocupação com privacidade e LGPD. A tela também ilustra a etapa inicial de conexão com o banco de dados relacional, considerando campos essenciais como tipo de banco, host, database e usuário.

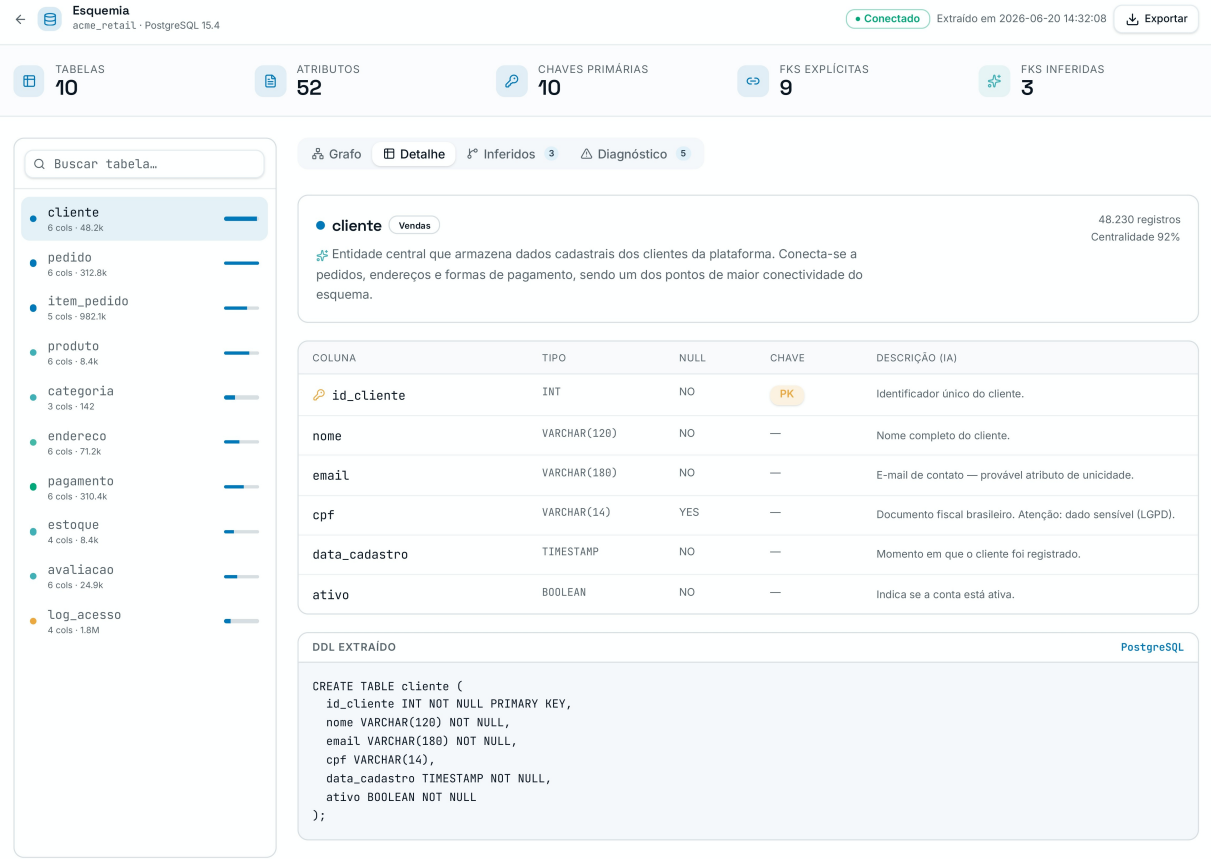
Figura 11 – Tela conceitual de visualização do grafo estrutural



Fonte: Adaptado pela Inteligência Artificial Lovable (2026).

A Figura 11 representa a visualização conceitual do grafo estrutural do banco de dados. Nessa tela, as tabelas são apresentadas como nós e os relacionamentos como arestas, permitindo ao usuário identificar visualmente dependências, entidades centrais e conexões relevantes do esquema relacional. O destaque atribuído à tabela central ilustra a possibilidade de uso de métricas de centralidade, como a centralidade de grau, para apoiar a exploração inicial da estrutura analisada.

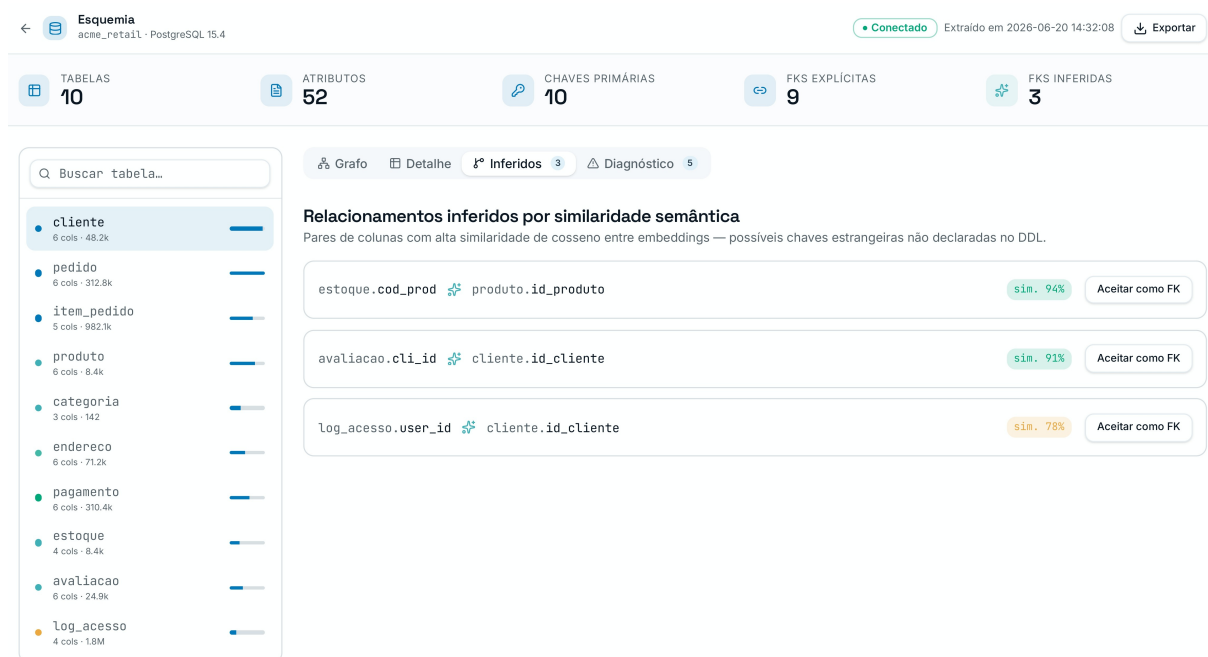
Figura 12 – Tela conceitual de detalhamento de tabela e metadados extraídos



Fonte: Adaptado pela Inteligência Artificial Lovable (2026).

A Figura 12 ilustra a tela conceitual de detalhamento de uma tabela selecionada. Nessa representação, o usuário poderia consultar informações sobre colunas, tipos de dados, nulidade, chaves primárias, chaves estrangeiras e descrições sugeridas por IA. A tela também apresenta uma representação textual estruturada dos metadados, semelhante a um DDL simplificado, evidenciando como as informações estruturais poderiam ser organizadas e apresentadas de forma legível para apoiar a documentação e a compreensão técnica do banco de dados.

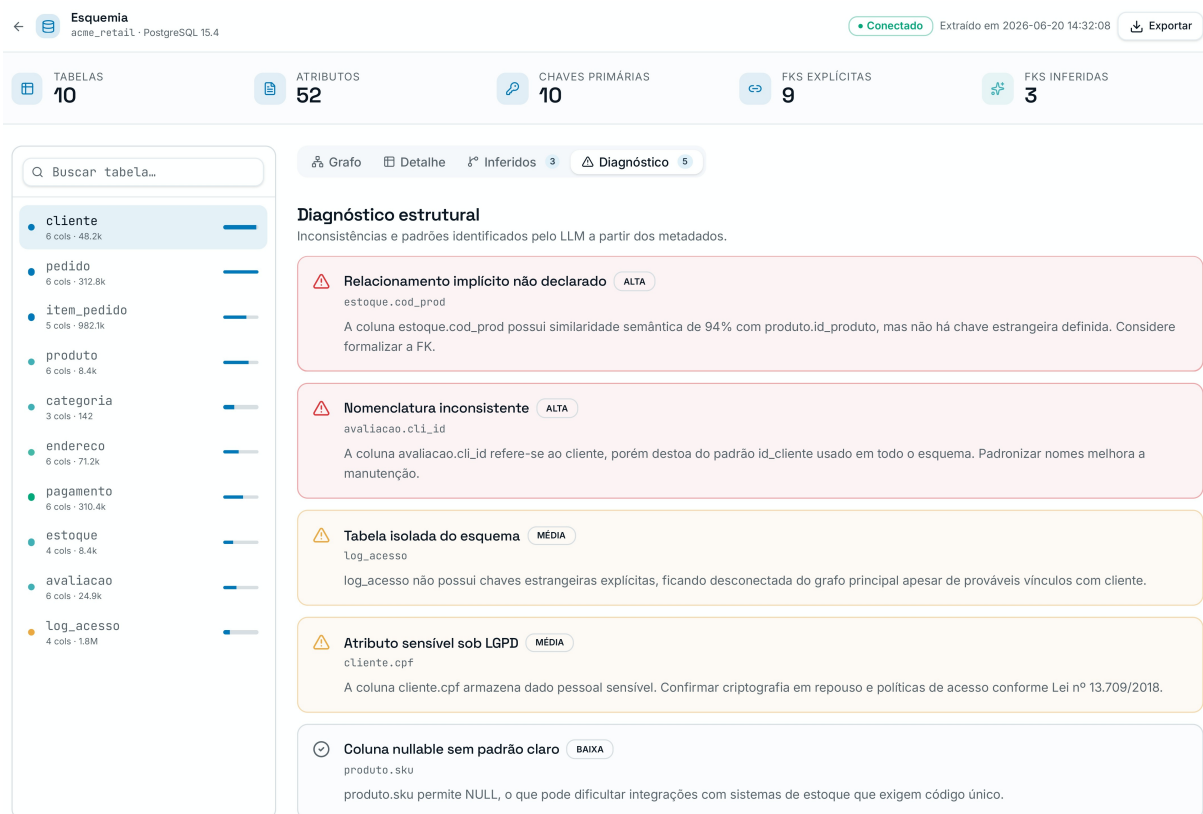
Figura 13 – Tela conceitual de relacionamentos inferidos por similaridade semântica



Fonte: Adaptado pela Inteligência Artificial Lovable (2026).

A Figura 13 apresenta uma representação conceitual da funcionalidade de sugestão de relacionamentos implícitos. A partir da comparação semântica entre nomes de colunas e demais metadados estruturais, a solução poderia indicar possíveis associações entre atributos de tabelas distintas. Essas sugestões, entretanto, devem ser compreendidas como hipóteses de relacionamento, exigindo validação humana antes de qualquer confirmação ou documentação definitiva.

Figura 14 – Tela conceitual de diagnóstico estrutural



Fonte: Adaptado pela Inteligência Artificial Lovable (2026).

A Figura 14 demonstra a tela conceitual de diagnóstico estrutural da solução proposta. Nessa etapa, o sistema poderia apresentar possíveis inconsistências, padrões de nomenclatura, relacionamentos não formalizados, tabelas isoladas ou atributos que exigem atenção sob a perspectiva de privacidade e governança. Essas indicações funcionariam como apoio à análise crítica do profissional, não substituindo sua avaliação técnica.

Dessa forma, o protótipo visual apresentado contribui para tornar a proposta mais compreensível, aproximando a arquitetura conceitual de uma possível experiência de uso. As telas ilustram como os principais recursos previstos poderiam ser organizados em uma aplicação web, contemplando desde a conexão inicial com o banco de dados até a visualização do grafo estrutural, o detalhamento de tabelas, a sugestão de relacionamentos implícitos e o diagnóstico de possíveis inconsistências.

6.3 Fluxo de Funcionamento

O fluxo de funcionamento da solução proposta inicia-se com a conexão a um banco de dados relacional compatível. Após o estabelecimento da conexão, os metadados estruturais serão extraídos e processados para identificação das entidades, atributos, chaves e relacionamentos existentes.

Após a extração, os metadados serão encaminhados ao módulo de processamento estrutural, responsável por organizar as informações e construir uma representação baseada em grafos. Essa etapa permitiria transformar a estrutura relacional do banco de dados em uma representação visual mais acessível, facilitando a identificação das conexões existentes entre as tabelas.

Em seguida, os metadados processados serão encaminhados ao módulo de IA para geração de descrições semânticas. Nessa etapa, o modelo de linguagem poderia interpretar nomes de tabelas, atributos e relacionamentos, produzindo explicações textuais que auxiliem o usuário na compreensão da estrutura analisada.

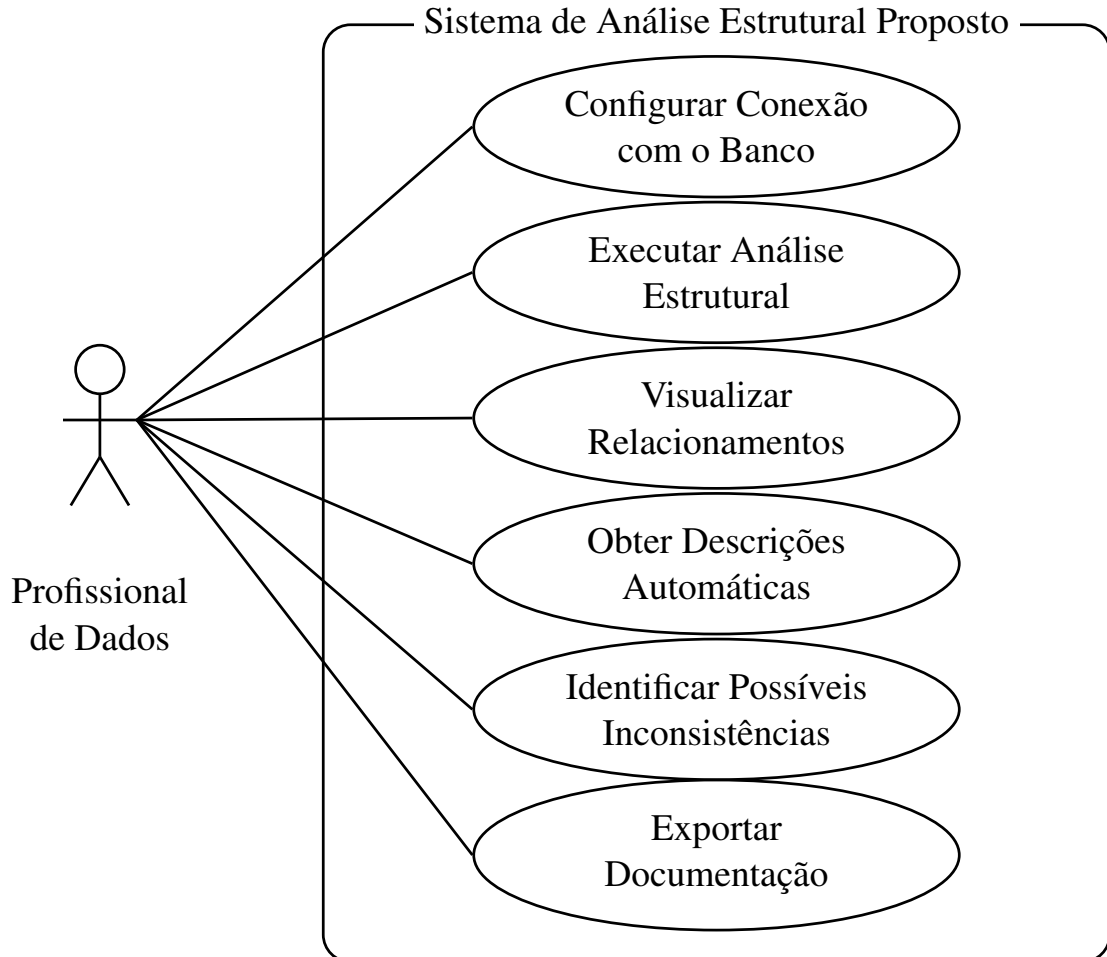
Os resultados produzidos serão consolidados pela camada de aplicação e encaminhados à interface gráfica. Nessa interface, o usuário poderia visualizar os relacionamentos entre tabelas, consultar descrições automáticas, identificar possíveis inconsistências e exportar a documentação gerada. Dessa forma, o fluxo operacional da solução busca reduzir o esforço necessário para compreender bases de dados desconhecidas, automatizando etapas tradicionalmente realizadas de forma manual e contribuindo para a geração de documentação estrutural mais acessível e consistente.

A automatização dessas etapas reduziria a necessidade de inspeções manuais realizadas diretamente sobre o banco de dados. Como consequência, espera-se uma diminuição do tempo gasto na fase inicial de exploração das bases analisadas, contribuindo para maior produtividade em atividades de integração, análise, governança de dados e desenvolvimento de soluções orientadas a dados.

Esse fluxo também favorece a rastreabilidade das informações produzidas, uma vez que as descrições e representações geradas poderiam ser associadas aos metadados que lhes deram origem. Em uma futura implementação, essa característica permitiria que o usuário verificasse quais tabelas, atributos ou relacionamentos foram considerados durante a análise, aumentando a transparência do processo e facilitando eventuais ajustes ou validações manuais.

6.4 Diagrama de Casos de Uso

Figura 15 – Diagrama de Casos de Uso da solução proposta



Fonte: Autoria própria (2026).

O Diagrama de Casos de Uso apresentado na Figura 15 representa as principais interações previstas entre o profissional de dados e a solução proposta. O ator principal é o usuário responsável por explorar e compreender a estrutura de um banco de dados relacional, especialmente em cenários nos quais a documentação técnica é inexistente, insuficiente ou desatualizada.

O primeiro caso de uso corresponde à configuração da conexão com o banco de dados, etapa necessária para que a solução possa acessar os metadados estruturais do esquema analisado. A partir dessa conexão, o usuário poderá executar a análise estrutural, que representa o processo central da solução proposta.

Após a execução da análise estrutural, o usuário poderá visualizar os relacionamentos identificados entre as tabelas, obter descrições automáticas geradas com apoio de IA, identificar

possíveis inconsistências estruturais e exportar a documentação produzida. Esses casos de uso representam funcionalidades previstas para a solução e estão alinhados às dificuldades identificadas no levantamento realizado com profissionais e estudantes da área de TI.

A visualização de relacionamentos permite compreender graficamente como as tabelas estão conectadas, facilitando a identificação de dependências e entidades centrais. A obtenção de descrições automáticas está relacionada ao uso de LLMs, responsáveis por gerar explicações em linguagem natural sobre tabelas, atributos e possíveis relações existentes. Já a identificação de inconsistências busca apoiar a análise crítica da estrutura, indicando possíveis problemas de nomenclatura, ausência de relacionamentos formais ou padrões estruturais que mereçam atenção.

Por fim, o caso de uso de exportação da documentação representa a possibilidade de registrar os resultados obtidos durante a análise, permitindo que as informações geradas sejam reutilizadas posteriormente em processos de documentação, governança de dados, manutenção de sistemas ou apoio à tomada de decisão.

A representação por Casos de Uso contribui para delimitar o escopo funcional da solução sob a perspectiva do usuário. No contexto deste trabalho, o escopo da proposta concentra-se na análise estrutural de metadados, na visualização de relacionamentos, na geração de descrições automáticas e na identificação de possíveis inconsistências relacionadas ao esquema do banco de dados. Embora a plataforma possa ser expandida futuramente para contemplar outras formas de análise, como avaliação da qualidade dos dados, verificação de formas normais, análise de redundâncias ou validação de regras de modelagem, tais funcionalidades não integram o escopo principal desta proposta. Assim, o diagrama auxilia na organização das funcionalidades previstas e pode orientar etapas futuras de desenvolvimento, priorização de requisitos e validação com usuários.

6.5 Arquitetura da Solução

A arquitetura proposta é composta por quatro componentes principais: módulo de extração de metadados, módulo de processamento estrutural, módulo de IA e módulo de visualização. Cada um desses componentes foi concebido para responder diretamente às dificuldades identificadas no levantamento realizado com os participantes da pesquisa.

O módulo de extração de metadados tem como finalidade obter exclusivamente informações estruturais do banco de dados, incluindo tabelas, atributos, tipos de dados, chaves primárias, chaves estrangeiras e relacionamentos explicitamente definidos. Essa extração seria realizada por

meio de consultas SQL previamente estruturadas e específicas para cada Sistema Gerenciador de Banco de Dados (SGBD), direcionadas aos catálogos internos e às estruturas de metadados disponibilizadas pelos próprios bancos relacionais. Dessa forma, a solução não acessaria os registros armazenados nas tabelas de negócio, limitando-se a recuperar apenas informações sobre o esquema da base analisada. Essas informações constituem a base para as demais etapas da solução, uma vez que permitem representar formalmente a estrutura do banco de dados analisado sem expor dados pessoais, sensíveis ou operacionais. Esse módulo responde ao cenário em que muitos profissionais relatam dificuldade em compreender os relacionamentos entre tabelas, dificuldade que, em grande parte, decorre da ausência de mecanismos automatizados capazes de mapear e expor essa estrutura de forma acessível (Silberschatz; Korth; Sudarshan, 2019; Elmasri; Navathe, 2016).

As informações extraídas são encaminhadas ao módulo de processamento estrutural, responsável por organizar os elementos identificados e representá-los de forma estruturada. Nessa etapa, a proposta prevê a utilização de grafos para modelar a estrutura do banco de dados, considerando tabelas como vértices e relacionamentos como arestas. Essa representação favorece a identificação de dependências, entidades centrais, conexões relevantes e possíveis agrupamentos lógicos dentro do esquema relacional (Cormen *et al.*, 2009; Hagberg; Schult; Swart, 2008).

O módulo de IA é responsável por apoiar a interpretação dos metadados extraídos, permitindo a geração de descrições automáticas, a sugestão de significados para atributos e a identificação de possíveis padrões ou inconsistências estruturais. Esse componente responde diretamente a dificuldades como ausência de documentação adequada e complexidade na compreensão do significado dos atributos. A geração automática de descrições por meio de LLMs representa, portanto, não apenas uma funcionalidade técnica prevista, mas uma resposta a uma lacuna recorrente em ambientes de dados pouco documentados (Russell; Norvig, 2021; White *et al.*, 2023; Hong *et al.*, 2024).

Por fim, o módulo de visualização tem como objetivo apresentar as informações produzidas de forma gráfica e interativa, possibilitando ao usuário compreender a estrutura do banco de dados de maneira mais intuitiva. Sua concepção está alinhada à necessidade de visualizar relacionamentos entre tabelas, explorar entidades, consultar descrições geradas automaticamente e obter uma visão geral da organização da base analisada.

A divisão da arquitetura em módulos independentes favorece a manutenção, a evolução

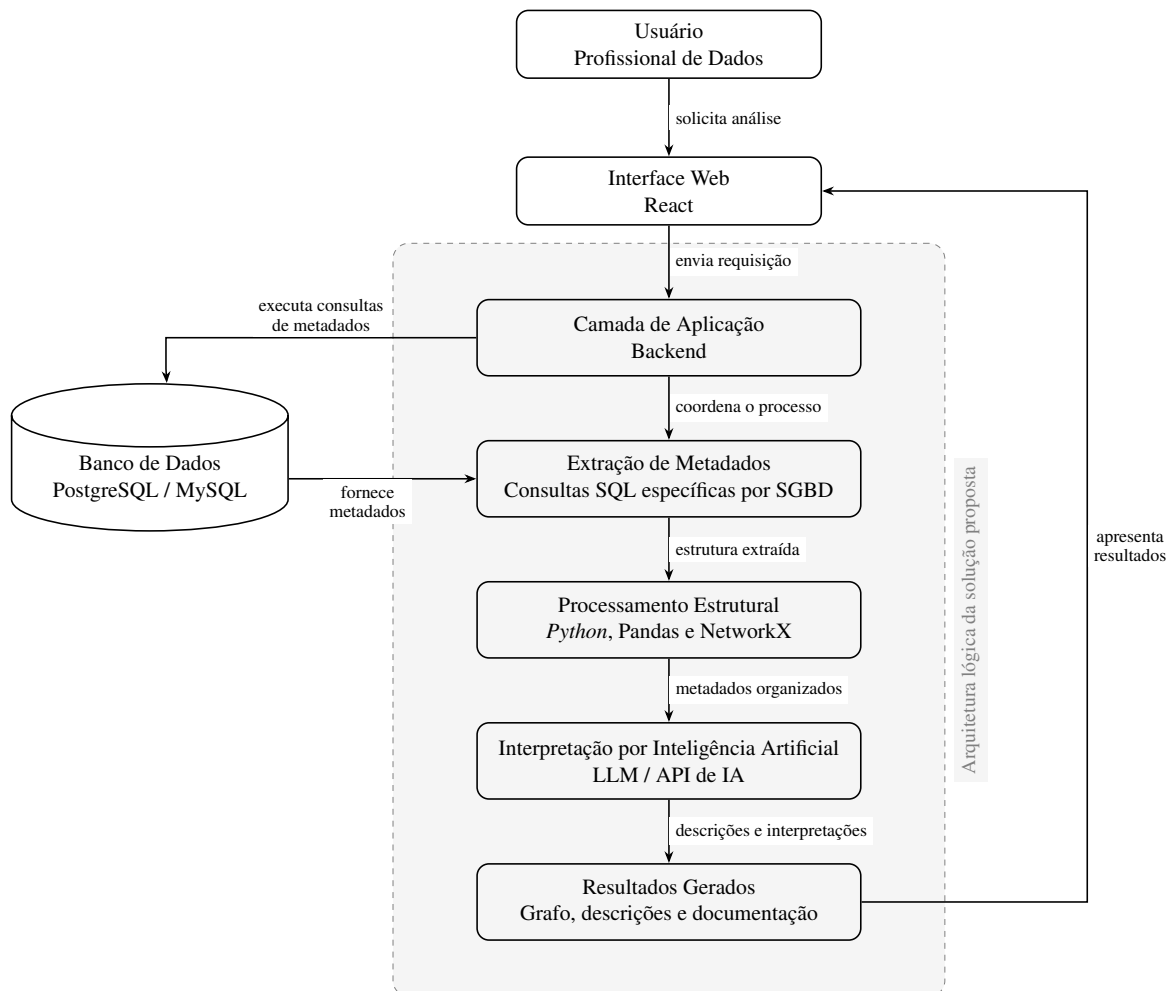
e a adaptação futura da solução. Essa abordagem permite que componentes específicos sejam substituídos ou aprimorados sem impactar significativamente os demais elementos da arquitetura. Como exemplo, diferentes modelos de linguagem podem ser incorporados ao módulo de IA, assim como novas tecnologias de visualização podem ser integradas ao módulo responsável pela interface gráfica.

Essa característica torna a solução flexível para diferentes contextos de utilização, possibilitando adaptações conforme requisitos de desempenho, segurança, privacidade, disponibilidade de recursos computacionais ou restrições institucionais. Assim, a arquitetura proposta busca oferecer uma base modular e expansível para futuras implementações.

Além disso, a separação entre extração, processamento, interpretação e visualização contribui para uma organização mais clara do fluxo de dados dentro da solução. Em uma futura implementação, essa estrutura modular poderia facilitar a manutenção do sistema, permitir a integração com diferentes bancos relacionais e possibilitar que o componente de IA seja configurado conforme as políticas de segurança e privacidade da organização, utilizando serviços externos por meio de APIs ou modelos executados localmente.

6.5.1 Arquitetura de Processamento

Figura 16 – Diagrama de fluxo de dados da solução proposta



Fonte: Autoria própria (2026).

A Figura 16 apresenta o fluxo lógico de dados da solução proposta. O processo inicia-se com a solicitação de análise realizada pelo usuário por meio da interface web. Essa solicitação é encaminhada à camada de aplicação, responsável por coordenar as etapas de extração, processamento, interpretação e apresentação dos resultados.

Na etapa de extração, a solução proposta acessa os metadados estruturais de bancos relacionais, como *PostgreSQL* ou *MySQL*, por meio de consultas SQL previamente definidas para cada banco compatível. Essas consultas seriam direcionadas às estruturas internas de metadados, como catálogos do sistema e visões informacionais, permitindo identificar tabelas, atributos, tipos de dados, chaves primárias, chaves estrangeiras e relacionamentos existentes.

Ressalta-se que essa etapa não prevê a consulta aos registros das tabelas de negócio, mas apenas às informações estruturais necessárias para compreender o esquema do banco de dados. Os metadados extraídos são então encaminhados ao módulo de processamento estrutural, no qual são organizados e representados por meio de estruturas de grafos.

O Quadro 4 apresenta um exemplo conceitual de consulta voltada à extração de metadados. Ressalta-se que as consultas finais dependeriam do SGBD utilizado, uma vez que bancos como *PostgreSQL* e *MySQL* possuem particularidades em seus catálogos internos. Ainda assim, o exemplo evidencia que o objetivo da extração é recuperar apenas informações estruturais do esquema, sem acessar registros armazenados nas tabelas de negócio.

Quadro 4 – Exemplo conceitual de consulta para extração de metadados

```

1  -- Exemplo conceitual de consulta aos metadados do banco
2
3  SELECT
4      table_schema AS esquema ,
5      table_name   AS tabela ,
6      column_name  AS atributo ,
7      data_type    AS tipo_dado ,
8      is_nullable  AS permite_nulo
9  FROM
10     information_schema.columns
11  WHERE table_schema NOT IN (
12     'information_schema',
13     'pg_catalog',
14     'mysql',
15     'performance_schema',
16     'sys'
17 )
18 ORDER BY table_schema, table_name, ordinal_position;

```

Fonte: Autoria própria (2026).

O exemplo apresentado no Quadro 4 ilustra uma consulta direcionada ao *information schema*, estrutura utilizada por sistemas gerenciadores de bancos de dados para armazenar informações sobre o esquema da base. A consulta recupera apenas metadados, como nome do esquema, nome da tabela, nome dos atributos, tipo de dado e indicação de nulidade, sem acessar os registros armazenados nas tabelas de negócio. Em uma futura implementação da arquitetura proposta, consultas específicas poderiam ser definidas para cada SGBD, contemplando também a identificação de chaves primárias, chaves estrangeiras e relacionamentos formalmente definidos.

O processamento estrutural, proposto com o uso de *Python*, *Pandas* e *NetworkX*, permite

transformar os metadados extraídos em uma representação mais clara da estrutura do banco de dados. Nessa representação, as tabelas podem ser tratadas como nós e os relacionamentos como arestas, favorecendo a visualização das dependências e conexões existentes no esquema relacional (Hagberg; Schult; Swart, 2008).

Após essa etapa, os metadados organizados são encaminhados ao módulo de IA, responsável por apoiar a geração de descrições automáticas, interpretações semânticas e possíveis observações sobre padrões ou inconsistências estruturais. Por fim, os resultados gerados, incluindo grafos, descrições e documentação, são disponibilizados ao usuário por meio da interface web.

Ressalta-se que esse fluxo representa uma arquitetura proposta, não uma implementação concluída. Dessa forma, o diagrama tem como finalidade demonstrar a organização lógica dos componentes previstos para uma futura construção da solução.

Além de representar o caminho percorrido pelos metadados, o fluxo também evidencia a separação entre os dados de negócio e os elementos estruturais analisados. A proposta considera apenas informações de esquema, como nomes de tabelas, atributos, tipos de dados e relacionamentos, evitando o tratamento direto dos registros armazenados nas tabelas. Essa decisão fortalece a aderência da solução aos princípios de minimização de dados e privacidade desde a concepção, especialmente em contextos nos quais a utilização de serviços externos de IA exige maior atenção à segurança da informação (Brasil, 2018; Cavoukian, 2009).

6.5.2 Geração de Descrições por IA

Após a extração dos metadados diretamente do banco de dados relacional, por meio de consultas SQL específicas para esse fim, as informações estruturais identificadas serão organizadas em um formato intermediário contendo tabelas, atributos, tipos de dados, chaves e relacionamentos. Dessa forma, a solução proposta não depende do acesso ao script original de criação do banco, mas sim dos metadados disponíveis no próprio SGBD. Esses metadados poderão ser convertidos em uma representação textual estruturada, semelhante a um DDL simplificado, para compor os prompts enviados ao modelo de linguagem.

Com base nesses metadados, o modelo poderia produzir descrições automáticas das tabelas, sugerir possíveis significados para atributos e identificar padrões estruturais relevantes. Por exemplo, uma tabela denominada *cliente*, contendo atributos como *id_cliente*, *nome*, *cpf* e *data_cadastro*, poderia ser descrita automaticamente como uma entidade responsável por

armazenar informações cadastrais dos clientes de uma organização.

Além da geração de descrições, a IA também poderia auxiliar na identificação de inconsistências de nomenclatura, possíveis duplicidades semânticas entre atributos e relacionamentos não explicitamente definidos no banco de dados. Dessa forma, o modelo de linguagem atuaria como um mecanismo de apoio à interpretação estrutural, não substituindo a análise humana, mas oferecendo informações adicionais para orientar a exploração da base.

Como o processo proposto utiliza apenas informações de esquema e não os registros armazenados nas tabelas, busca-se reduzir a exposição de dados sensíveis e preservar a confidencialidade das informações analisadas. Essa decisão está alinhada ao princípio de minimização de dados, uma vez que limita o tratamento às informações estritamente necessárias para a compreensão da estrutura do banco (Brasil, 2018).

As descrições geradas serão posteriormente disponibilizadas ao usuário juntamente com a representação gráfica da estrutura do banco de dados, complementando a análise visual com interpretações semânticas produzidas pela IA. Dessa forma, a proposta busca unir representação estrutural e explicação textual em uma única solução de apoio à compreensão de bancos de dados relacionais.

A geração de descrições por IA também poderia contribuir para a padronização da documentação produzida, uma vez que os resultados poderiam seguir um formato previamente definido nos prompts enviados ao modelo. Esse formato poderia contemplar o propósito provável de cada tabela, a descrição dos atributos, a identificação de relacionamentos relevantes e observações sobre possíveis inconsistências estruturais. Com isso, a documentação gerada tenderia a ser mais uniforme, facilitando sua reutilização em processos de manutenção, governança de dados e integração entre equipes.

Outro ponto importante é que as interpretações produzidas pelos modelos de linguagem devem ser compreendidas como hipóteses baseadas nos metadados disponíveis. Como os modelos podem gerar respostas imprecisas ou ambíguas, especialmente em bases com nomenclaturas pouco descritivas, a validação humana permanece necessária. Assim, a IA é concebida como mecanismo de apoio à análise, e não como substituta definitiva do conhecimento técnico do profissional responsável pela base (White *et al.*, 2023; Hong *et al.*, 2024).

6.6 Tecnologias Propostas para a Solução

As tecnologias propostas para a solução foram organizadas de acordo com os componentes da arquitetura apresentada neste capítulo.

No módulo de extração de metadados, propõe-se o uso da linguagem SQL, uma vez que ela permite acessar informações estruturais armazenadas nos sistemas gerenciadores de bancos de dados relacionais. Conforme descrito na arquitetura da solução, essa extração seria realizada por meio de consultas previamente estruturadas para cada SGBD compatível, sem recuperar registros das tabelas analisadas. Inicialmente, a arquitetura considera compatibilidade com *PostgreSQL* e *MySQL*, por serem sistemas amplamente utilizados em ambientes acadêmicos e corporativos, podendo ser expandida futuramente para outros SGBDs.

Para o *backend* e para o processamento das informações extraídas, propõe-se a utilização da linguagem *Python*. Essa escolha justifica-se pela sua ampla adoção em projetos de ciência de dados, engenharia de dados e automação, bem como pela disponibilidade de bibliotecas voltadas à manipulação, análise e integração de dados. No contexto da solução proposta, o *Python* representará agir como a camada responsável por coordenar a extração dos metadados, organizar as informações obtidas, acionar os módulos de análise estrutural e integrar o sistema aos recursos de IA.

A biblioteca *Pandas* é proposta para apoiar a organização e o tratamento dos metadados em estruturas tabulares. Com seu uso, informações como nomes de tabelas, atributos, tipos de dados, chaves e relacionamentos poderiam ser organizadas, filtradas e transformadas antes de serem encaminhadas aos demais componentes da arquitetura.

Para a representação estrutural dos relacionamentos entre tabelas, propõe-se a utilização da biblioteca *NetworkX*. Essa biblioteca possibilita modelar a estrutura do banco de dados como um grafo, no qual as tabelas são representadas como nós e os relacionamentos como arestas. Essa representação favorece a identificação de dependências, entidades centrais, agrupamentos estruturais e possíveis pontos críticos da modelagem relacional (Hagberg; Schult; Swart, 2008).

No módulo de IA, propõe-se a utilização de LLMs. Esses modelos serão responsáveis por apoiar a geração de descrições automáticas de tabelas e atributos, sugerir possíveis significados para elementos do esquema e indicar padrões ou inconsistências estruturais. A integração com esses modelos poderia ocorrer por meio de APIs externas ou, em cenários futuros, por meio da execução local de modelos, a depender dos requisitos de privacidade, segurança, custo e infraestrutura disponível (White *et al.*, 2023; Hong *et al.*, 2024).

Para a interface gráfica, propõe-se o uso de *React*, tecnologia voltada à construção de interfaces web interativas. Essa escolha se justifica pela possibilidade de criar uma aplicação visual capaz de apresentar grafos, descrições automáticas, informações sobre tabelas e mecanismos de navegação entre os elementos do banco de dados. Como apoio à estilização da interface, propõe-se o uso do *Tailwind CSS*, que permite a construção de *layouts* responsivos, organizados e visualmente consistentes.

A comunicação entre os componentes da solução poderia ocorrer por meio de APIs, permitindo a separação entre a interface, o *backend*, o módulo de processamento estrutural e o módulo de IA. Essa organização favorece a modularidade da arquitetura, facilitando a manutenção, a substituição de componentes e a expansão futura da solução.

Dessa forma, a seleção das tecnologias propostas busca atender diretamente aos módulos definidos na arquitetura: a SQL apoia a extração dos metadados; o *Python* coordena o processamento e a integração dos componentes; o *Pandas* organiza os dados extraídos; o *NetworkX* representa a estrutura relacional por meio de grafos; os LLMs apoiam a interpretação semântica; e o *React* com *Tailwind CSS* permite a apresentação visual e interativa dos resultados ao usuário.

Portanto, as tecnologias descritas nesta seção foram selecionadas não apenas por sua popularidade ou disponibilidade, mas principalmente por sua aderência às necessidades funcionais da solução proposta. Essa organização contribui para uma arquitetura modular, extensível e adequada a futuras etapas de implementação, validação e aprimoramento.

6.7 Contribuições Esperadas

A solução proposta busca reduzir o esforço necessário para compreender bancos de dados relacionais desconhecidos, automatizando atividades tradicionalmente realizadas de forma manual. A combinação entre análise estrutural baseada em grafos e geração automática de descrições por IA tem como objetivo facilitar a interpretação de tabelas, atributos e relacionamentos.

Espera-se que a proposta contribua para a geração de documentação técnica mais acessível e atualizada, apoiando profissionais e estudantes em atividades de exploração, manutenção e integração de dados. Em ambientes organizacionais, essa contribuição pode auxiliar equipes que lidam com bases legadas, pouco documentadas ou modificadas ao longo do tempo sem o devido registro técnico.

A proposta também pode contribuir para reduzir a dependência de conhecimento tácito existente nas organizações. Em muitos contextos, a compreensão sobre a estrutura dos dados

permanece concentrada em profissionais específicos, dificultando a continuidade das atividades quando há mudança de equipe, migração de sistemas ou entrada de novos colaboradores. Ao propor a geração automatizada de descrições e representações estruturais, a solução busca tornar esse conhecimento mais explícito e compartilhável.

Além do contexto organizacional, a proposta pode ser utilizada como recurso educacional para apoio ao ensino de modelagem e bancos de dados relacionais. A visualização gráfica dos relacionamentos e a geração de descrições automáticas podem auxiliar estudantes na compreensão de esquemas reais, aproximando o aprendizado teórico de cenários práticos.

Outra contribuição esperada está relacionada ao apoio à governança de dados. Ao favorecer a identificação de tabelas, atributos, relacionamentos e possíveis inconsistências, a solução proposta pode servir como base para processos de documentação, catalogação e melhoria da qualidade estrutural dos bancos de dados analisados.

Adicionalmente, a solução proposta pode contribuir para a padronização da documentação técnica produzida a partir de bancos relacionais. Ao utilizar um fluxo estruturado de extração, processamento, interpretação e apresentação dos metadados, torna-se possível gerar resultados mais consistentes e menos dependentes de interpretações isoladas. Esse aspecto é relevante em equipes multidisciplinares, nas quais diferentes profissionais precisam compartilhar uma compreensão comum sobre a estrutura da base.

Também se espera que a proposta apoie a tomada de decisão técnica em projetos de dados. Ao evidenciar entidades centrais, dependências entre tabelas, possíveis inconsistências e descrições semânticas dos elementos analisados, a solução pode fornecer subsídios para decisões relacionadas à construção de consultas, desenvolvimento de *dashboards*, integração entre sistemas, refatoração de modelos de dados e priorização de atividades de documentação.

6.8 Aderência da Solução aos Requisitos Identificados

A proposta apresentada neste trabalho foi elaborada com base nas necessidades identificadas durante o levantamento realizado com profissionais e estudantes da área de TI. A análise dos resultados permitiu estabelecer uma relação direta entre os problemas observados e os componentes definidos na arquitetura da solução.

A ausência de documentação adequada é abordada por meio da utilização de modelos de linguagem para geração automática de descrições de tabelas, atributos e relacionamentos. Essa funcionalidade busca apoiar o usuário na compreensão inicial da base de dados, especialmente

em cenários nos quais não há documentação formal disponível.

A dificuldade na compreensão dos relacionamentos entre tabelas é contemplada pelo módulo de visualização baseado em grafos, responsável por representar de forma intuitiva as conexões existentes na estrutura do banco de dados. Essa representação permite que o usuário identifique visualmente como as entidades se relacionam, quais tabelas possuem maior conectividade e quais elementos exercem papel central na estrutura analisada.

Além disso, a dificuldade relacionada à interpretação dos atributos presentes nas bases analisadas é tratada pela geração de descrições semânticas produzidas pela IA. A partir dos nomes dos atributos, tipos de dados e contexto estrutural das tabelas, o modelo de linguagem poderia sugerir significados prováveis para os elementos identificados, contribuindo para uma compreensão mais clara do esquema.

Os resultados do levantamento também evidenciaram interesse por funcionalidades como visualização gráfica de relacionamentos, identificação de inconsistências e documentação automática das tabelas. Essas funcionalidades estão diretamente incorporadas à arquitetura proposta, demonstrando alinhamento entre as necessidades apontadas pelos participantes e os recursos previstos para a solução.

Dessa forma, observa-se que a arquitetura concebida apresenta aderência aos requisitos identificados durante a pesquisa, fornecendo mecanismos capazes de responder às principais dificuldades relatadas pelos participantes. Embora a implementação e a avaliação prática da ferramenta estejam fora do escopo deste trabalho, a análise realizada indica que a proposta constitui uma alternativa promissora para apoiar processos de exploração, documentação e compreensão de bancos de dados relacionais.

Essa aderência também se manifesta na preocupação com privacidade e segurança da informação. Como os participantes demonstraram atenção ao envio de metadados para serviços externos, a arquitetura proposta prevê o tratamento limitado aos metadados estruturais, evitando o compartilhamento de registros armazenados nas tabelas. Além disso, a possibilidade futura de execução local de modelos de linguagem amplia a flexibilidade da proposta para ambientes com políticas mais restritivas de segurança e conformidade (Brasil, 2018; ISO; IEC, 2022).

6.9 Limitações da Solução

Embora a solução proposta apresente potencial para auxiliar na compreensão de bancos de dados relacionais, algumas limitações devem ser consideradas. A primeira delas refere-se ao

fato de que o presente trabalho se limita à proposição e modelagem da solução, não contemplando sua implementação completa, testes em ambiente real ou avaliação experimental com usuários. Dessa forma, os benefícios descritos representam expectativas fundamentadas na arquitetura proposta e nos resultados do levantamento realizado, mas ainda demandam validação prática em trabalhos futuros.

Outra limitação está relacionada à qualidade dos metadados disponíveis no banco de dados. Esquemas que utilizam nomenclaturas pouco descritivas, abreviações excessivas, ausência de padronização ou relacionamentos não formalizados podem reduzir a precisão das interpretações produzidas pelo modelo de linguagem e dificultar a geração de descrições confiáveis.

Além disso, os resultados fornecidos pela IA devem ser compreendidos como suporte à análise, e não como documentação definitiva. Modelos de linguagem podem produzir interpretações imprecisas, ambíguas ou inconsistentes em determinados contextos, especialmente quando os metadados disponíveis não são suficientes para inferir corretamente o significado dos elementos analisados. Por esse motivo, a validação humana permanece necessária.

Outra limitação refere-se ao processamento de bases de dados muito extensas, compostas por grande quantidade de tabelas, atributos e relacionamentos. Nesses casos, podem ser necessárias estratégias adicionais de segmentação, sumarização de metadados e otimização do envio de informações ao modelo de linguagem, especialmente devido a limitações de janela de contexto, custo computacional e tempo de processamento (Anthropic, 2023).

Também devem ser considerados aspectos relacionados à privacidade e à segurança da informação. Embora a proposta preveja o uso exclusivo de metadados estruturais, a adoção de APIs externas de IA pode exigir cuidados adicionais, como anonimização de nomes sensíveis, controle de acesso, análise de conformidade com políticas institucionais e avaliação dos termos de uso dos serviços empregados (Brasil, 2018; Cavoukian, 2009; ISO; IEC, 2022).

Adicionalmente, diferentes modelos de linguagem podem produzir interpretações distintas para os mesmos metadados, uma vez que as descrições geradas dependem dos mecanismos de inferência empregados por cada modelo. Essa variação reforça a necessidade de avaliação comparativa entre modelos e de definição de critérios de qualidade para as descrições produzidas.

Dessa forma, recomenda-se que as informações geradas pela solução sejam utilizadas como apoio ao processo de compreensão estrutural, mantendo a validação humana como etapa complementar da documentação produzida. As limitações apresentadas não invalidam a proposta, mas indicam pontos relevantes a serem considerados em etapas futuras de implementação, teste

e aprimoramento da solução.

Por fim, destaca-se que a eficácia da solução também dependerá da forma como os resultados serão apresentados ao usuário. Mesmo que os módulos de extração, processamento e interpretação funcionem adequadamente, a interface deverá organizar as informações de maneira clara, navegável e compreensível. Assim, estudos futuros deverão considerar não apenas a precisão das descrições geradas, mas também a experiência do usuário, a usabilidade da interface e o impacto real da ferramenta na redução do tempo de compreensão de bases desconhecidas.

7 CONCLUSÕES E TRABALHOS FUTUROS

7.1 Conclusões

O presente trabalho teve como ponto de partida a identificação de um problema recorrente no contexto profissional da área de TI: a dificuldade enfrentada por engenheiros de dados, cientistas de dados, analistas e estudantes na compreensão estrutural de bancos de dados relacionais complexos, especialmente quando esses ambientes apresentam documentação técnica inexistente, incompleta ou desatualizada. Ao longo da pesquisa, observou-se que essa dificuldade pode estar associada a impactos na produtividade das equipes, ao aumento do tempo necessário para a exploração inicial das bases e à possibilidade de interpretações equivocadas, amplia o tempo necessário para exploração inicial das bases, aumenta a possibilidade de interpretações equivocadas e compromete a confiabilidade das análises produzidas a partir dos dados.

A revisão bibliográfica permitiu fundamentar o problema a partir de diferentes perspectivas, envolvendo bancos de dados relacionais, modelagem de dados, engenharia de dados, metadados, BI, teoria de grafos, IA, LLMs e privacidade de dados. Esse embasamento demonstrou que, embora existam ferramentas consolidadas para administração, visualização e catalogação de bancos de dados, como *DBeaver*, *MySQL Workbench* e soluções de *Data Catalog*, tais ferramentas ainda apresentam limitações quando aplicadas a cenários pouco documentados. Em geral, essas soluções são eficazes para exibir estruturas, diagramas e informações cadastradas, mas ainda dependem fortemente do conhecimento prévio do usuário ou da existência de documentação previamente alimentada para que a compreensão da base seja realmente efetiva.

Nesse sentido, a lacuna identificada não está apenas na ausência de ferramentas de visualização, mas na ausência de mecanismos capazes de interpretar automaticamente os elementos estruturais de um banco de dados relacional e transformá-los em informações acessíveis ao usuário. A compreensão de tabelas, atributos, chaves, relacionamentos e possíveis regras implícitas ainda permanece, em grande parte, dependente de análise manual, experiência profissional e conhecimento institucional. Essa dependência torna o processo de exploração inicial mais demorado e suscetível a falhas, principalmente em bases legadas, heterogêneas ou modificadas ao longo do tempo sem documentação adequada.

Com o objetivo de verificar a relevância prática desse problema, foi realizado um levantamento com 30 participantes, entre profissionais e estudantes da área de TI. Os resultados obtidos reforçaram a pertinência da problemática investigada. A maioria dos participantes relatou

dificuldades associadas à falta de documentação, à compreensão de relacionamentos entre tabelas e à interpretação do significado dos atributos. Além disso, 46,7% dos respondentes afirmaram demandar entre quatro e oito horas para compreender a estrutura de uma base desconhecida, enquanto 16,7% relataram necessitar de mais de um dia de trabalho. Esses dados evidenciam que a compreensão inicial de uma base de dados pode representar uma etapa de alto custo operacional dentro de projetos orientados a dados.

Outro resultado relevante foi o impacto da ausência de documentação no desenvolvimento das atividades. O levantamento mostrou que 80% dos participantes já sofreram atrasos em suas demandas devido ao esforço necessário para compreender a estrutura dos dados. Esse percentual indica que o esforço necessário para compreender a estrutura dos dados pode estar associado a impactos nos prazos das atividades dos participantes. Entretanto, como não foram realizados cruzamentos estatísticos entre as respostas individuais, o levantamento não permite determinar relações causais nem afirmar impactos diretos sobre a qualidade das análises ou a eficiência das equipes. Ao mesmo tempo, 96,7% dos participantes atribuíram notas 4 ou 5 à utilidade potencial de uma solução capaz de analisar automaticamente a estrutura de bancos de dados relacionais, o que sugere aderência potencial às necessidades relatadas neste trabalho.

Entretanto, os resultados também revelaram uma dimensão importante relacionada à privacidade e à segurança da informação. Embora os participantes tenham demonstrado receptividade à ideia de uma solução com essas características, 80% atribuíram notas 3 ou superiores ao nível de preocupação com o envio de metadados para serviços externos de IA. Esse dado indica que a adoção de soluções baseadas em IA não depende apenas de sua utilidade funcional, mas também da confiança que elas oferecem quanto ao tratamento das informações analisadas. Dessa forma, a pesquisa evidenciou que requisitos como minimização de dados, transparência, controle e possibilidade de execução local devem ser considerados como elementos centrais da arquitetura proposta.

Com base nos resultados obtidos no levantamento e na fundamentação teórica construída ao longo da pesquisa, foram definidos os requisitos e as diretrizes que orientaram a modelagem da solução proposta. Dessa forma, o questionário não teve como objetivo validar uma ferramenta previamente implementada, mas sim identificar dificuldades, percepções e funcionalidades relevantes para subsidiar a concepção da arquitetura apresentada neste trabalho.

Diante desse cenário, este trabalho propôs uma solução para análise estrutural automatizada de bancos de dados relacionais com o auxílio de IA. A arquitetura concebida contempla

quatro componentes principais: extração de metadados estruturais, processamento e representação baseada em grafos, interpretação por meio de LLMs, e visualização interativa dos resultados. Essa organização busca transformar metadados técnicos, como nomes de tabelas, atributos, tipos de dados, chaves e relacionamentos, em representações visuais e descrições textuais capazes de apoiar a compreensão da estrutura analisada.

A utilização de grafos foi proposta como estratégia para representar os relacionamentos entre tabelas de forma mais intuitiva, permitindo identificar dependências, entidades centrais e conexões relevantes dentro do esquema relacional. Associada a essa representação, a aplicação de LLMs foi concebida como mecanismo de apoio à geração de descrições automáticas, sugestão de significados para atributos, identificação de padrões estruturais e indicação de possíveis inconsistências. Assim, a proposta integra análise estrutural e interpretação semântica, oferecendo uma abordagem complementar às ferramentas tradicionais de administração e modelagem de bancos de dados.

A solução também foi concebida considerando princípios de privacidade e segurança da informação. Ao restringir a análise aos metadados estruturais, evitando o envio de registros armazenados nas tabelas, a proposta busca reduzir riscos associados à exposição de dados pessoais ou sensíveis. Essa decisão está alinhada aos princípios da LGPD, especialmente à minimização de dados, e reforça a necessidade de que soluções baseadas em IA sejam projetadas desde sua concepção com atenção à governança, conformidade e proteção das informações processadas.

Do ponto de vista acadêmico, este trabalho contribui para a discussão sobre o uso de IA na compreensão estrutural de bancos de dados relacionais, tema ainda pouco explorado de forma integrada na literatura nacional. A pesquisa articula conceitos de bancos de dados, metadados, teoria de grafos, modelos de linguagem e privacidade, propondo uma arquitetura conceitual capaz de apoiar futuras investigações e implementações. Além disso, a modelagem por meio de diagramas, como o Diagrama de Fluxo de Dados e o Diagrama de Casos de Uso, contribui para formalizar a organização lógica da solução e delimitar suas principais funcionalidades.

Do ponto de vista prático, a proposta apresenta potencial para reduzir o esforço necessário à compreensão inicial de bases desconhecidas, apoiar a geração de documentação técnica, facilitar a exploração de relacionamentos entre tabelas e contribuir para atividades de engenharia de dados, análise de dados, governança e BI. Em contextos organizacionais, a solução pode auxiliar equipes que lidam com bases legadas, pouco documentadas ou dependentes de conhecimento

tácito concentrado em profissionais específicos. Em contextos educacionais, pode servir como recurso de apoio ao ensino de modelagem e interpretação de bancos de dados relacionais.

Ressalta-se, contudo, que o escopo deste trabalho concentrou-se na proposição e modelagem da solução, não contemplando sua implementação completa ou validação experimental em ambiente real. Dessa forma, os benefícios apresentados devem ser compreendidos como contribuições esperadas, fundamentadas na revisão bibliográfica, no levantamento realizado e na arquitetura concebida. A validação prática da solução permanece como etapa necessária para confirmar sua efetividade, mensurar seus impactos e identificar ajustes técnicos ou funcionais.

Por fim, conclui-se que os objetivos do trabalho foram alcançados, uma vez que foi possível identificar e fundamentar o problema de pesquisa, analisar sua relevância junto ao público-alvo, propor uma arquitetura para solução automatizada e discutir suas contribuições, limitações e possibilidades futuras. Os resultados obtidos indicam indícios relevantes de demanda por mecanismos que apoiem a compreensão estrutural de bancos de dados relacionais, especialmente em ambientes pouco documentados. Contudo, ressalta-se que as contribuições esperadas descritas no Capítulo 6 permanecem no plano teórico e conceitual, uma vez que ainda dependem da implementação da arquitetura proposta e da aplicação das diretrizes de avaliação futura apresentadas na Seção 4.3. Dessa forma, a confirmação de sua efetividade deverá ocorrer em pesquisas subsequentes, por meio de validação experimental, mensuração de resultados e comparação com abordagens tradicionais de análise estrutural.

7.2 Trabalhos Futuros

Em virtude de o escopo deste trabalho estar concentrado na proposição e modelagem da solução, diversas possibilidades se apresentam como continuidade natural da pesquisa. A primeira delas refere-se à implementação da arquitetura proposta, transformando o modelo conceitual em um protótipo funcional. Essa etapa permitirá verificar a viabilidade técnica dos módulos previstos, incluindo a extração de metadados, o processamento estrutural, a construção de grafos, a integração com LLMs e a apresentação dos resultados em uma interface web.

Associada à implementação, torna-se necessária a realização de validações experimentais em ambientes reais ou simulados. Estudos futuros poderão comparar a abordagem tradicional de exploração manual de bancos de dados com a utilização da solução proposta, considerando métricas como tempo necessário para compreender a estrutura da base, quantidade de relacionamentos identificados, clareza das descrições geradas, qualidade da documentação produzida

e percepção dos usuários quanto à utilidade da ferramenta. Essa avaliação poderá fornecer evidências quantitativas e qualitativas sobre a contribuição efetiva da solução.

Outro aspecto importante para trabalhos futuros é a ampliação do suporte a diferentes sistemas gerenciadores de bancos de dados. A arquitetura proposta considera inicialmente bancos como *PostgreSQL* e *MySQL*, mas sua expansão para outros SGBDs, como *Oracle*, *Microsoft SQL Server* e *SQLite*, aumentaria a aplicabilidade da solução em diferentes contextos organizacionais. Essa ampliação exigirá a adaptação dos mecanismos de extração de metadados, considerando as particularidades de cada sistema quanto à representação de tabelas, chaves, relacionamentos e tipos de dados.

Também se recomenda a evolução dos mecanismos de geração automática de documentação técnica. Trabalhos futuros poderão explorar a exportação dos resultados em formatos como *Markdown* e *Portable Document Format* (PDF), permitindo que as descrições, grafos e informações estruturais produzidas pela solução sejam reutilizados em processos de governança, manutenção de sistemas, integração de equipes e catalogação de dados. Essa funcionalidade pode contribuir para reduzir a desatualização documental e preservar o conhecimento estrutural das bases analisadas.

Outro caminho relevante consiste na integração da solução com ferramentas de *Data Catalog* e plataformas de governança de dados. Essa integração permitiria que as informações geradas automaticamente fossem incorporadas a repositórios organizacionais já utilizados pelas equipes, ampliando a utilidade da solução e favorecendo a atualização contínua da documentação. Dessa forma, a proposta poderia atuar como uma camada complementar de geração e enriquecimento de metadados.

Diante das preocupações identificadas no levantamento quanto ao envio de metadados para serviços externos, trabalhos futuros também poderão investigar a utilização de modelos de linguagem executados localmente. Essa alternativa permitiria reduzir a dependência de APIs externas e ampliar a aderência da solução a ambientes com políticas mais rigorosas de privacidade, segurança e conformidade. A comparação entre modelos locais e modelos acessados por serviços externos poderá considerar critérios como qualidade das descrições, custo computacional, desempenho, privacidade e facilidade de implantação.

Além disso, futuras pesquisas poderão explorar estratégias para lidar com bancos de dados de grande porte, compostos por centenas ou milhares de tabelas. Nesses casos, será necessário avaliar técnicas de segmentação de esquemas, priorização de entidades centrais,

sumarização de metadados e controle da janela de contexto dos modelos de linguagem. Tais estratégias são fundamentais para garantir que a solução permaneça eficiente e aplicável mesmo em ambientes de elevada complexidade estrutural.

Outra possibilidade de continuidade está relacionada à avaliação da qualidade das descrições geradas pela IA. Estudos futuros poderão definir critérios de avaliação, como clareza, completude, coerência, precisão semântica, utilidade para o usuário e consistência com os metadados analisados. Esses critérios poderão apoiar tanto a comparação entre diferentes modelos de linguagem quanto o aprimoramento dos prompts utilizados no processo de geração das descrições.

Por fim, recomenda-se que trabalhos futuros investiguem aspectos de usabilidade e experiência do usuário. A efetividade da solução não dependerá apenas da precisão técnica das análises produzidas, mas também da forma como os resultados serão apresentados e compreendidos pelos usuários. Assim, avaliações com profissionais e estudantes poderão indicar melhorias na interface, na navegação pelos grafos, na organização das descrições e na exportação da documentação gerada.

Dessa forma, os trabalhos futuros apontam para a continuidade da pesquisa em direção à implementação, validação e aprimoramento da solução proposta. A partir dessas etapas, será possível verificar de maneira mais precisa o impacto da abordagem na redução do esforço de compreensão de bancos de dados relacionais, na melhoria da documentação técnica e no apoio a processos de análise, governança e tomada de decisão orientada por dados.

REFERÊNCIAS

- ANTHROPIC. Introducing Claude 2.1. 2023. **ANTHROPIC**. Publicado em: 21 nov. 2023. Disponível em: <https://www.anthropic.com/news/claude-2-1>. Acesso em: 8 jul. 2026.
- APACHE SOFTWARE FOUNDATION. **Apache Atlas**: data governance and metadata framework for Hadoop. s.d. Disponível em: <https://atlas.apache.org/>. Acesso em: 8 jul. 2026.
- BADMAN, Annie; KOSINSKI, Matthew. O que são metadados? 2024. **INTERNATIONAL BUSINESS MACHINES CORPORATION (IBM)**. Publicado em: 21 out. 2024. Disponível em: <https://www.ibm.com/br-pt/think/topics/metadata>. Acesso em: 8 jul. 2026.
- BATINI, Carlo; CERI, Stefano; NAVATHE, Shamkant Bhalchandra. **Conceptual database design**: an entity-relationship approach. Redwood City: Benjamin/Cummings, 1992.
- BRASIL. **Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD)**. Brasília, DF: Presidência da República, 2018. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709compilado.htm. Acesso em: 8 jul. 2026.
- CAVOUKIAN, Ann. **Privacy by Design**. Toronto, 2009. Disponível em: <https://www.ipc.on.ca/en/media/1826/download?attachment=>. Acesso em: 11 jul. 2026.
- CORMEN, Thomas H. *et al.* **Introduction to algorithms**. 3. ed. Cambridge: The MIT Press, 2009.
- DATAHUB PROJECT. **DataHub documentation**. s.d.a. Disponível em: <https://docs.datahub.com/>. Acesso em: 8 jul. 2026.
- DATAHUB PROJECT. **AI documentation**. s.d.b. Disponível em: <https://docs.datahub.com/docs/automations/ai-docs>. Acesso em: 10 jul. 2026.
- DBEAVER CORPORATION. **DBeaver documentation**. s.d.a. Disponível em: <https://dbeaver.com/docs/dbeaver/>. Acesso em: 8 jul. 2026.
- DBEAVER CORPORATION. **AI assistant**. s.d.b. Disponível em: <https://dbeaver.com/docs/dbeaver/AI-Smart-Assistance/>. Acesso em: 10 jul. 2026.
- EHRLINGER, Lisa; RUSZ, Elisa; WÖSS, Wolfram. A survey of data quality measurement and monitoring tools. **arXiv preprint arXiv:1907.08138**, 2019. Disponível em: <https://arxiv.org/abs/1907.08138>. Acesso em: 10 jul. 2026.
- ELMASRI, Ramez; NAVATHE, Shamkant Bhalchandra. **Sistemas de banco de dados**. 7. ed. São Paulo: Pearson, 2016.
- FREEMAN, Linton C. Centrality in social networks: conceptual clarification. **Social Networks**, v. 1, n. 3, p. 215–239, 1978.
- GÉRON, Aurélien. **Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow**: concepts, tools, and techniques to build intelligent systems. 2. ed. Sebastopol: O'Reilly Media, 2019.

HAGBERG, Aric A.; SCHULT, Daniel A.; SWART, Pieter J. Exploring network structure, dynamics, and function using NetworkX. In: VAROQUAUX, Gaël; VAUGHT, Travis; MILLMAN, Jarrod (Ed.). **Proceedings of the 7th Python in Science Conference**. Pasadena: SciPy, 2008. p. 11–15. Disponível em: <https://proceedings.scipy.org/articles/TCWV9851>. Acesso em: 11 jul. 2026.

HEUSER, Carlos Alberto. **Projeto de banco de dados**. 6. ed. Porto Alegre: Bookman, 2009.

HONG, Zijin *et al.* Next-generation database interfaces: a survey of LLM-based text-to-sql. **arXiv preprint arXiv:2406.08426**, 2024. Disponível em: <https://arxiv.org/abs/2406.08426>. Acesso em: 8 jul. 2026.

ISO; IEC. **ISO/IEC 27001:2022**: information security, cybersecurity and privacy protection: information security management systems: requirements. 2022. International Organization for Standardization (ISO) e International Electrotechnical Commission (IEC). Disponível em: <https://www.iso.org/standard/27001>. Acesso em: 8 jul. 2026.

KIMBALL, Ralph; ROSS, Margy. **The data warehouse toolkit**: the definitive guide to dimensional modeling. 3. ed. Indianapolis: Wiley, 2013.

MANNING, Christopher David; RAGHAVAN, Prabhakar; SCHÜTZE, Hinrich. **Introduction to information retrieval**. Cambridge: Cambridge University Press, 2008.

MICROSOFT CORPORATION. **Data governance with Microsoft Purview**. 2025. Atualizado em: 20 fev. 2025. Disponível em: <https://learn.microsoft.com/en-us/purview/data-governance-overview>. Acesso em: 10 jul. 2026.

MICROSOFT CORPORATION. **Learn about Microsoft Purview**. 2026. Atualizado em: 15 maio 2026. Disponível em: <https://learn.microsoft.com/en-us/purview/purview>. Acesso em: 8 jul. 2026.

MIKOLOV, Tomas *et al.* Distributed representations of words and phrases and their compositionality. In: BURGESS, Christopher J. C. *et al.* (Ed.). **Advances in Neural Information Processing Systems 26**. Red Hook: Curran Associates, 2013. p. 3111–3119. Disponível em: <https://proceedings.neurips.cc/paper/2013/hash/9aa42b31882ec039965f3c4923ce901b-Abstract.html>. Acesso em: 11 jul. 2026.

NARGESIAN, Fatemeh *et al.* Data lake management: challenges and opportunities. **Proceedings of the VLDB Endowment**, v. 12, n. 12, p. 1986–1989, 2019. Disponível em: <https://www.vldb.org/pvldb/vol12/p1986-nargesian.pdf>. Acesso em: 8 jul. 2026.

ORACLE CORPORATION. **MySQL Workbench reference manual**. 2026. Documento gerado em: 2 jul. 2026. Disponível em: <https://dev.mysql.com/doc/workbench/en/>. Acesso em: 8 jul. 2026.

PAPASTERGIOS, Vasileios; GOUNARIS, Anastasios. A survey of open-source data quality tools: shedding light on the materialization of data quality dimensions in practice. **arXiv preprint arXiv:2407.18649**, 2024. Disponível em: <https://arxiv.org/abs/2407.18649>. Acesso em: 10 jul. 2026.

PRODANOV, Cleber Cristiano; FREITAS, Ernani Cesar de. **Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico**. 2. ed. Novo Hamburgo: Feevale, 2013.

RUSSELL, Stuart; NORVIG, Peter. **Artificial intelligence**: a modern approach. 4. ed. Harlow: Pearson, 2021.

SHARDA, Ramesh; DELEN, Dursun; TURBAN, Efraim. **Business intelligence, analytics, and data science**: a managerial perspective. 4. ed. Boston: Pearson, 2017. Publicado em: 13 jan. 2017. Copyright 2018.

SILBERSCHATZ, Abraham; KORTH, Henry F.; SUDARSHAN, S. **Database system concepts**. 7. ed. New York: McGraw-Hill Education, 2019.

SUBRAMANIAM, Pranav *et al.* Comprehensive and comprehensible data catalogs: the what, who, where, when, why, and how of metadata management. **arXiv preprint arXiv:2103.07532**, 2021. Disponível em: <https://arxiv.org/abs/2103.07532>. Acesso em: 10 jul. 2026.

WHITE, Jules *et al.* A prompt pattern catalog to enhance prompt engineering with ChatGPT. **arXiv preprint arXiv:2302.11382**, 2023. Disponível em: <https://arxiv.org/abs/2302.11382>. Acesso em: 8 jul. 2026.

APÊNDICE A – QUESTIONÁRIO APLICADO AOS PARTICIPANTES DA PESQUISA

O questionário apresentado neste apêndice foi aplicado a profissionais e estudantes da área de TI com a finalidade de levantar percepções, dificuldades e expectativas relacionadas à compreensão estrutural de bancos de dados relacionais. O instrumento foi elaborado no âmbito desta pesquisa, a partir dos objetivos do trabalho e dos temas identificados na revisão bibliográfica, tais como documentação de bancos de dados, metadados, relacionamentos entre tabelas, qualidade estrutural das bases, governança de dados e uso de IA como apoio à interpretação de estruturas relacionais.

Ressalta-se que o questionário não apresentou telas, protótipos ou representações visuais da solução aos participantes. Seu objetivo foi identificar dificuldades e funcionalidades consideradas relevantes pelo público-alvo, servindo como subsídio para a modelagem da proposta apresentada neste trabalho.

Termo de Consentimento

Ao prosseguir com este formulário, você declara estar ciente de que os dados coletados destinam-se exclusivamente à pesquisa acadêmica e à proposição de uma solução voltada para a análise estrutural automatizada de bancos de dados relacionais utilizando inteligência artificial, como parte do Trabalho de Conclusão de Curso de Bacharelado em Tecnologia da Informação da Universidade Federal Rural do Semiárido (UFERSA).

Suas respostas serão tratadas com rigorosa confidencialidade e analisadas de forma agregada para fins estatísticos e de levantamento dos requisitos da solução, garantindo o seu anonimato. Todos os dados serão armazenados de forma segura e utilizados estritamente para essa finalidade, sem qualquer compartilhamento com terceiros.

Em conformidade com a Lei Geral de Proteção de Dados (LGPD — Lei nº 13.709/2018), sua participação é totalmente voluntária, sendo você livre para interromper o preenchimento a qualquer momento, sem qualquer penalidade.

O envio das respostas configura o seu consentimento livre e esclarecido para o uso das informações nos termos aqui descritos.

Perfil do Respondente

1. Qual é a sua área de atuação ou vínculo principal?

Selecione a opção que melhor descreve o que você faz atualmente, seja no mercado de trabalho, na academia ou em estudos. Caso atue em mais de uma área, escolha a predominante.

Tipo de resposta: Múltipla escolha (uma resposta)

- Engenharia de Dados
- Ciência de Dados
- Análise de Dados / BI
- Docente / Pesquisador
- Estudante de Graduação
- Estudante de Pós-Graduação
- Profissional de TI (outra especialidade)
- Outro

2. Qual é o seu vínculo institucional atual?

Queremos entender o ambiente em que você está inserido atualmente: empresa, universidade, trabalho autônomo etc. Isso nos ajuda a interpretar suas respostas dentro do contexto correto.

Tipo de resposta: Múltipla escolha (uma resposta)

- Empresa privada
- Instituição pública
- Universidade / Instituto de ensino
- Freelancer / Autônomo
- Sem vínculo no momento
- Outro

3. Qual é o seu nível de experiência com bancos de dados relacionais?

Bancos de dados relacionais são sistemas como MySQL, PostgreSQL ou SQL Server, onde os dados são organizados em tabelas. Avalie seu nível de familiaridade com esse tipo de tecnologia, desde nunca ter usado até trabalhar com ela diariamente.

Tipo de resposta: Escala linear de 1 a 5

Iniciante 1 2 3 4 5 Avançado

4. Com qual frequência você trabalha diretamente com bancos de dados relacionais em suas atividades profissionais ou acadêmicas?

Considere qualquer tipo de interação: consultas SQL, análise de dados, desenvolvimento de sistemas, estudos acadêmicos ou até mesmo acesso eventual a ferramentas de dados.

Tipo de resposta: Múltipla escolha (uma resposta)

- Diariamente
- Semanalmente
- Mensalmente
- Raramente
- Nunca

Dores e Dificuldades

5. Você já precisou trabalhar com um banco de dados relacional que não possuía documentação adequada?

Considere-se documentação adequada qualquer material que descreva o significado das tabelas e colunas, como dicionários de dados, comentários no banco ou manuais internos. Caso nunca tenha trabalhado com bancos de dados relacionais, selecione a opção “Nunca”.

Tipo de resposta: Múltipla escolha (uma resposta)

- Sim, frequentemente
- Sim, ocasionalmente
- Raramente
- Nunca

6. Quanto tempo, em média, você costuma levar para compreender a estrutura de um banco de dados relacional desconhecido?

Considere o tempo desde o primeiro acesso ao banco até o momento em que você se sente minimamente seguro para trabalhar com ele, entendendo quais tabelas existem, o que elas armazenam e como se relacionam entre si.

Tipo de resposta: Múltipla escolha (uma resposta)

- Menos de 1 hora
- Entre 1 hora e 4 horas
- Entre 4 horas e 8 horas
- Mais de 1 dia
- Mais de 1 semana

7. Quais são as maiores dificuldades que você enfrenta ao explorar um banco de dados relacional desconhecido?

Selecione todas as opções que já representaram um obstáculo real para você. Se nunca

explorou um banco desconhecido, pode pular esta pergunta ou selecionar “Não se aplica”.

Tipo de resposta: Caixa de seleção (múltiplas respostas)

- Entender os relacionamentos entre tabelas
- Identificar o significado dos atributos
- Detectar inconsistências
- Falta de documentação
- Nomenclatura confusa
- Outro

8. Esse tempo gasto na compreensão inicial do banco já causou atrasos em algum projeto?

Pode ser um projeto profissional, um trabalho acadêmico ou qualquer atividade em que a demora para entender o banco impactou o prazo ou a qualidade do resultado.

Tipo de resposta: Múltipla escolha (uma resposta)

- Sim, frequentemente
- Sim, algumas vezes
- Raramente
- Nunca

9. Você já interpretou incorretamente a estrutura de um banco e isso gerou erros na análise ou nos relatórios?

Por exemplo: entender errado o que uma coluna representa e gerar um relatório incorreto, ou usar a tabela errada em uma consulta por não saber exatamente o que cada uma armazena.

Tipo de resposta: Múltipla escolha (uma resposta)

- Sim
- Não
- Não tenho certeza

Ferramentas Utilizadas Hoje

10. Quais ferramentas você utiliza para explorar ou entender a estrutura de bancos de dados relacionais?

DBeaver e MySQL Workbench são exemplos de programas que permitem visualizar tabelas e relacionamentos. Data Catalog é um tipo de sistema que organiza e documenta os dados de uma organização. Selecione todas que utiliza ou já utilizou.

Tipo de resposta: Caixa de seleção (múltiplas respostas)

- DBeaver
- MySQL Workbench
- pgAdmin
- Data Catalog
- Consultas SQL manuais
- IA generativa (ChatGPT etc.)
- Documentação interna
- Nenhuma
- Outro

11. Essas ferramentas resolvem plenamente suas necessidades de compreensão estrutural do banco?

Avalie de forma geral se as ferramentas ou estratégias que você usa hoje são suficientes para compreender rapidamente a estrutura de um banco desconhecido. Se não usa nenhuma ferramenta, considere sua situação atual sem elas.

Tipo de resposta: Escala linear de 1 a 5

Não resolvem 1 2 3 4 5 Resolvem bem

12. Qual é a principal limitação das ferramentas que você usa atualmente para essa finalidade?

Descreva brevemente o que falta ou o que incomoda nas ferramentas que você já utiliza. Não há resposta certa; qualquer observação é valiosa para a pesquisa.

Tipo de resposta: Resposta longa.

Expectativas sobre o Sistema Proposto

13. Um sistema que analisasse automaticamente a estrutura do banco e gerasse descrições das tabelas e relacionamentos seria útil para o seu trabalho?

Pense em uma ferramenta que, ao se conectar a um banco de dados relacional, automaticamente identificasse o que cada tabela representa, como elas se relacionam e quais possíveis inconsistências existem, tudo gerado por inteligência artificial, sem necessidade de documentação manual.

Tipo de resposta: Escala linear de 1 a 5

Nada útil 1 2 3 4 5 Extremamente útil

14. Quais funcionalidades você consideraria mais importantes nesse tipo de sistema?

Selecione todas as funcionalidades que fariam diferença real no seu dia a dia. Não se preocupe em escolher apenas uma; todas as seleções ajudam a priorizar o desenvolvimento do sistema.

Tipo de resposta: Caixa de seleção (múltiplas respostas)

- Visualização gráfica de relacionamentos
- Descrições automáticas das tabelas via IA
- Identificação de inconsistências
- Sugestão de relacionamentos implícitos
- Exportação para documentação (Markdown/PDF)
- Apoio à criação de dashboards
- Outro

15. Como você preferiria interagir com esse sistema?

Interface web é um sistema acessado pelo navegador. CLI (linha de comando) é um terminal de texto. Integração com ferramenta existente significa usar dentro de programas que você já conhece, como DBeaver ou VS Code. API permite que desenvolvedores conectem o sistema a outros softwares.

Tipo de resposta: Caixa de seleção (múltiplas respostas)

- Interface web visual
- Linha de comando (CLI)
- Integração com ferramenta existente (DBeaver, VS Code etc.)
- API
- Indiferente

16. Que tipo de banco de dados relacional você mais utilizaria com esse sistema?

Esses são os sistemas gerenciadores de banco de dados relacionais mais comuns no mercado. Caso não saiba qual utiliza ou nunca tenha utilizado nenhum, selecione “Não sei” ou “Não utilizo”.

Tipo de resposta: Caixa de seleção (múltiplas respostas)

- PostgreSQL
- MySQL
- SQL Server

- Oracle
- SQLite
- Outro

17. Há algum cenário específico em que esse sistema seria especialmente valioso para você?

Por exemplo: entrar em um novo emprego e precisar entender rapidamente o banco da empresa, iniciar um projeto de pesquisa com dados legados ou criar um dashboard sem conhecer bem a base. Descreva livremente qualquer situação válida.

Tipo de resposta: Resposta longa.

Privacidade e Segurança dos Dados

18. Você teria preocupação em utilizar um sistema que envie metadados (nomes de tabelas e atributos, sem dados reais) para um serviço de IA externo?

Metadados são informações sobre a estrutura do banco, como os nomes das tabelas e colunas, mas não incluem os dados em si (como nomes de clientes ou valores financeiros). Avalie o quanto isso seria uma preocupação para você.

Tipo de resposta: Escala linear de 1 a 5

Nenhuma preocupação 1 2 3 4 5 Muita preocupação

19. Sua organização possui políticas que restringem o uso de serviços de IA externos para análise de dados?

Algumas empresas e instituições proíbem ou limitam o uso de ferramentas externas de IA por questões de segurança ou conformidade legal. Caso seja estudante ou autônomo sem vínculo institucional, selecione “Não se aplica”.

Tipo de resposta: Múltipla escolha (uma resposta)

- Sim
- Não
- Não sei
- Não se aplica

20. Uma opção de execução local do sistema (sem enviar dados para servidores externos) aumentaria sua disposição de utilizá-lo?

Execução local significa que o sistema rodaria inteiramente no seu próprio computador ou servidor, sem depender de serviços externos de IA. Isso garantiria que nenhuma informação saísse do seu ambiente.

Tipo de resposta: Múltipla escolha (uma resposta)

- Sim, muito
- Sim, um pouco
- Indiferente
- Não

Considerações Finais

21. Há alguma funcionalidade, preocupação ou sugestão que você gostaria de compartilhar sobre um sistema desse tipo?

Este campo é completamente livre. Pode ser uma ideia que surgiu durante o preenchimento, uma crítica construtiva, uma experiência que você considera relevante ou qualquer outro comentário que queira registrar.

Tipo de resposta: Resposta longa.