UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO
INSTITUTO MULTIDISCIPLINAR

GABRIEL DOMICIANO LACERDA

Ruralytics: Um Sistema de *Business Intelligence* para Gestão Estratégica da Produção Científica nas IES

Prof. Filipe Braida do Carmo, D.Sc.
Orientador

Nova Iguaçu, Julho de 2025

Ruralytics: Um Sistema de *Business Intelligence* para Gestão Estratégica da Produção Científica nas IES

Gabriel Domiciano Lacerda

Projeto Final de Curso submetido ao Departamento de Ciência da Computação do Instituto Multidisciplinar da Universidade Federal Rural do Rio de Janeiro como parte dos requisitos necessários para obtenção do grau de Bacharel em Ciência da Computação.

Apresentado por:

Gabriel Domiciano Lacerda

Aprovado por:

Prof. Filipe Braida do Carmo, D.Sc.

Prof. Marcel William Rocha da Silva, D.Sc.

Prof. Ubiratam Carvalho de Paula Junior, D.Sc.

NOVA IGUAÇU, RJ - BRASIL

Julho de 2025

DOCUMENTOS COMPROBATÓRIOS Nº 14882/2025 - CoordCGCC (12.28.01.00.00.98)

(Nº do Protocolo: NÃO PROTOCOLADO)

(Assinado digitalmente em 04/07/2025 17:35)
FILIPE BRAIDA DO CARMO
PROFESSOR DO MAGISTERIO SUPERIOR
DeptCC/IM (12.28.01.00.00.83)
Matrícula: ###295#4

(Assinado digitalmente em 08/07/2025 10:14)
MARCEL WILLIAM ROCHA DA SILVA
PROFESSOR DO MAGISTERIO SUPERIOR
PPGIHD (11.39.00.16)
Matrícula: ###807#6

(Assinado digitalmente em 04/07/2025 16:51)
UBIRATAM CARVALHO DE PAULA JUNIOR
PROFESSOR DO MAGISTERIO SUPERIOR
DeptCC/IM (12.28.01.00.00.83)
Matrícula: ###426#4

(Assinado digitalmente em 05/07/2025 11:30)
GABRIEL DOMICIANO LACERDA
DISCENTE
Matrícula: 2020######9

*Visualize o documento original em https://sipac.ufrrj.br/documentos/ informando seu número: **14882**, ano: **2025**, tipo: **DOCUMENTOS COMPROBATÓRIOS**, data de emissão: **04/07/2025** e o código de verificação: **371362e919***

Agradecimentos

Primeiramente, quero agradecer a Deus; sem a força do Alto que Ele me deu todos os dias e sem a ajuda do Seu Espírito, eu não estaria aqui. A Sua sabedoria e direção me guiaram até este momento. Agradeço também, em particular, a todos que rezaram por mim e me auxiliaram, mas que não estão mencionados no texto a seguir.

Profundos agradecimentos à minha família, começando por meus pais, Ivan e Malbia, que foram base e pilar para mim durante todos esses anos. Hoje reconheço o quanto vocês se doaram para que eu chegasse até aqui, e este trabalho é prova da confiança e do apoio que depositaram em mim. Toda a paciência, oração, amor e carinho de vocês estão presentes em cada página. Ao meu irmão, Miguel, que tornou todo esse caminho mais leve e me encheu de sorrisos, seja pelas piadas ou pelos momentos de descontração. Aos meus primos, especialmente Denis e Geovana, que me apoiaram mesmo de longe até o término desta jornada: vocês são parte fundamental desta conquista. Adicionalmente, agradeço aos meus padrinhos, Geraldo e Ermelinda, que estão sempre em oração por mim e me dão enorme apoio.

Ao meu orientador, Prof. Filipe Braida, pela paciência demonstrada ao longo deste percurso – apesar da longa e tortuosa jornada, jamais desistiu de mim. Aprendi muito com seus ensinamentos, não apenas durante a orientação, mas ao longo de todo o curso. Sua didática e seu conhecimento me inspiram, e sou genuinamente grato por tê-lo como mentor. Agradeço também aos colegas de curso — em especial Kevyn e Marcos —, que estiveram comigo desde o início e me auxiliaram durante a redação deste trabalho; sem o apoio de ambos, ele não teria se concretizado. Estendo ainda meus agradecimentos a todos os professores que fizeram parte desta trajetória:

todo o aprendizado recebido tornou esta jornada possível e prazerosa.

Aos meus amigos, que fazem parte da minha rede de apoio há tantos anos e me amparam nos momentos em que mais preciso: saibam que cada encontro, seja online ou presencial, tornou meus dias muito mais leves. Agradeço especialmente a Guilherme, a quem considero um irmão, e não me esqueço de Juan, Diego, Tainá, Gabriela, Rafaella, Gustavo e Thais. Todos vocês são família para mim, e aprecio imensamente a amizade de cada um.

À minha terapeuta, Stella, que me auxiliou nos momentos difíceis e me ajudou a reconquistar o foco para concluir este trabalho.

Aos meus colegas de trabalho, que me ofereceram suporte, compartilharam conhecimentos e deram excelentes conselhos: vocês, veteranos da faculdade e do mercado, me fazem aprender diariamente, seja tecnicamente ou na vida. Em especial, Rita, Thalia, Lucas, Ângelo, Leonardo, Gabriel e Bruno: considero vocês sinceros amigos e agradeço pela parceria construída ao longo destes anos na Cesgranrio.

Por fim, agradeço a mim mesmo: sinto que todo o esforço destes anos valeu a pena. Cada aprendizado de vida, cada disciplina cursada, somados ao apoio de tanta gente, tornaram-me grato por ter uma rede de suporte tão rica. Sinto-me orgulhoso ao lembrar daquele menino meio perdido que iniciou o curso e perceber o quanto cresci até este momento de conclusão.

RESUMO

Ruralytics: Um Sistema de *Business Intelligence* para Gestão Estratégica da Produção Científica nas IES

Gabriel Domiciano Lacerda

Julho/2025

Orientador: Filipe Braida do Carmo, D.Sc.

A gestão estratégica da produção científica é um desafio crescente para as Instituições de Ensino Superior (IES) brasileiras, especialmente dada a riqueza e complexidade dos dados disponíveis na Plataforma Lattes. Este trabalho teve como objetivo principal propor, detalhar a concepção e a implementação de um sistema de *analytics* dedicado ao processamento, análise e visualização da produção acadêmica originada desta plataforma. Buscou-se transformar o grande volume de dados curriculares em informações estratégicas que subsidiem a tomada de decisão, o planejamento científico e ampliem a visibilidade da pesquisa institucional. A metodologia envolveu uma fundamentação em conceitos de *Business Intelligence* (BI), *Analytics* e processos de Extração, Transformação e Carga (ETL), culminando no desenvolvimento de uma aplicação web. A arquitetura modular e o processo de ETL com gerenciamento de filas foram projetados para garantir a organização e a confiabilidade dos dados. Como resultado, a plataforma desenvolvida permite o monitoramento de indicadores, a identificação de tendências, a análise segmentada da produção e a geração de *dashboards* interativos, configurando-se como uma contribuição relevante para a otimização da gestão da informação científica e o fortalecimento da cultura de dados nas IES.

Palavras-chave: *Business Intelligence*, *Analytics*, Produção Científica, ETL, Monitoramento de Dados, Gestão da Informação Acadêmica, Plataforma Lattes.

ABSTRACT

Ruralytics: Um Sistema de *Business Intelligence* para Gestão Estratégica da Produção Científica nas IES

Gabriel Domiciano Lacerda

Julho/2025

Advisor: Filipe Braida do Carmo, D.Sc.

The strategic management of scientific production is a growing challenge for Brazilian Higher Education Institutions (HEIs), especially given the richness and complexity of the data available on the Lattes Platform. This work's main objective was to propose, detail the design, and implement an analytics system dedicated to the processing, analysis, and visualization of academic production originating from this platform. The aim was to transform the large volume of curriculum data into strategic information to support decision-making, scientific planning, and enhance the visibility of institutional research. The methodology involved a foundation in Business Intelligence (BI), Analytics, and Extract, Transform, Load (ETL) processes, culminating in the development of a web application. The modular architecture and the ETL process with queue management were designed to ensure data organization and reliability. As a result, the developed platform allows for the monitoring of indicators, identification of trends, segmented analysis of production, and the generation of interactive dashboards, establishing itself as a relevant contribution to optimizing scientific information management and strengthening a data-driven culture in HEIs.

Keywords: *Business Intelligence, Analytics, Academic Production, ETL, Data Monitoring, Academic Information Management, Lattes Platform.*

Lista de Figuras

Lista de Abreviaturas e Siglas

BI	*Business Intelligence*
ETL	Extração, Transformação e Carga
SGBD	Sistema de Gerenciamento de Banco de Dados
XML	*eXtensible Markup Language*
JSON	*JavaScript Object Notation*
OTP	*One-Time Password*
RF	Requisitos Funcionais
RNF	Requisitos Não Funcionais
RN	Regras de Negócio
CU	Caso de Uso
KPIs	Indicadores-Chave de Desempenho
UUID	Identificador Único Universal
USP	Universidade de São Paulo
UFRRJ	Universidade Federal Rural do Rio de Janeiro
UFMG	Universidade Federal de Minas Gerais
	,
DER	Diagrama Entidade-Relacionamento
SQL	*Structured Query Language*
SPA	*Single-page application*

HTML	*Hypertext Markup Language*
UI	Interface gráfica
ORM	Mapeamento Objeto-Relacional
BaaS	*Backend as a Service*
DDD	*Domain-Driven Design*
IA	Inteligência Artificial
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
IES	Instituições de Ensino Superior
AMQP	Advanced Message Queuing Protocol
API	Interfaces de programação de aplicativos
UX	Experiência do Usuário
DCU	Design Centrado no Usuário

Sumário

Capítulo 1

Introdução

A produção científica, tecnológica e acadêmica constitui um pilar essencial para o avanço socioeconômico e a contínua inovação em qualquer nação, pois impulsiona a geração de conhecimento, a inovação e a formação de capital humano qualificado. No Brasil, essa produção é documentada de forma centralizada pela Plataforma Lattes, mantida pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), que consolida a trajetória de pesquisadores, estudantes e diversos profissionais da ciência em atuação no país.

Funcionando como um repositório curricular público e integrado, a plataforma vai além do simples registro, desempenhando um papel estratégico para todo o ecossistema de pesquisa. Ela é a principal fonte de dados para a avaliação de mérito por agências de fomento e estabelece o padrão para conectar a ciência brasileira ao cenário internacional (TECNOLÓGICO, 2024). Por essa razão, a análise de seus dados tornou-se uma ferramenta indispensável para a compreensão do panorama científico do país e para a formulação de estratégias de desenvolvimento (MUGNAINI; PACKER; MENEGHINI, 2004).

Embora esses dados sejam de suma importância, gestores e pesquisadores em Instituições de Ensino Superior (IES) frequentemente encontram desafios ao tentar centralizar esses dados para fazer uma análise estratégica. O principal obstáculo é a visão fragmentada da produção científica. Isso ocorre devido à dispersão dos

dados em milhares de currículos individuais, o que impede um panorama claro e agregado das competências e das lacunas existentes (SILVA; HAYASHI; HAYASHI, 2011). Essa fragmentação dificulta o direcionamento eficaz da pesquisa e compromete a tomada de decisões assertivas sobre o planejamento institucional, a alocação de recursos e a avaliação de políticas internas (SHATTOCK, 2003).

Ademais, a subutilização do potencial informativo contido na Plataforma Lattes configura-se como um desafio corrente. A extração manual de informações para relatórios é um processo intensivo e demorado. Tal processo é frequentemente propenso a erros e dificilmente consegue ser completo ou abranger todos os dados relevantes. Isso limita severamente a capacidade de realizar análises mais aprofundadas sobre tendências, redes de colaboração ou o impacto da produção. (MENA-CHALCO; JR, 2009)

Essa limitação operacional, associada ao mapeamento muitas vezes deficiente de competências internas, restringe a ágil identificação de especialistas e dificulta a promoção efetiva das capacidades institucionais (ETZKOWITZ; LEYDESDORFF, 2000). O impacto direto é sentido na visibilidade da instituição e em sua inserção em redes de inovação. Consequentemente, a ineficiência em processos avaliativos e de gestão se acentua, sobrecarregando equipes administrativas e pesquisadores (GLäNZEL; SCHOEPFLIN, 1994).

Para endereçar as problemáticas discutidas, emerge a necessidade de aplicar ferramentas que sistematizem a coleta, o processamento e a visualização da produção científica. A viabilidade dessa abordagem é corroborada por Siemens e Long (2011), que destaca como a análise de dados educacionais pode converter o grande volume de informações brutas em conhecimento estratégico e acionável, pois, ao prover um suporte informacional centralizado e de fácil acesso, tais sistemas fortalecem a gestão universitária e oferecem um meio dinâmico para superar os desafios mencionados.

1.1 Objetivo

O presente trabalho tem como objetivo precípuo o desenvolvimento e a validação de um sistema digital de *analytics*. Este sistema será voltado para o monitoramento e análise da produção acadêmica, utilizando como fonte os dados da Plataforma Lattes.

Este esforço se alinha à crescente necessidade de organizar, centralizar e conferir maior visibilidade à informação científica. Tal desafio é crucial para a gestão do conhecimento nas instituições de ensino e pesquisa atualmente. Essa necessidade é ressaltada por Leite (LEITE, 2009) ao discutir o papel fundamental dos repositórios institucionais e do acesso aberto. Estes promovem o controle do saber produzido pela academia, garantindo sua preservação e ampla disseminação.

O sistema proposto visa transformar o volume massivo e frequentemente disperso de dados curriculares em inteligência estratégica. Pretende-se gerar conhecimento acionável, fornecendo subsídios para uma gestão universitária mais eficaz e transparente. Para alcançar tal propósito, a ferramenta será projetada para permitir a geração de estatísticas detalhadas e segmentadas. Estas serão organizadas por unidade acadêmica, departamento e também por cada docente individualmente.

O sistema irá quantificar e apresentar a evolução da produção científica, tecnológica e artística ao longo do tempo. Com foco na abrangência, serão contempladas diversas categorias de produção acadêmica, incluindo publicações, bancas e orientações. Espera-se que, por meio dessa abordagem analítica e do uso de indicadores estatísticos, o sistema aprimore a gestão da produção científica. Isso será demonstrado na instituição que servirá de estudo de caso, a Universidade Federal Rural do Rio de Janeiro (UFRRJ).

A plataforma fornecerá um panorama abrangente da atividade científica institucional, facilitando, assim, a identificação de padrões, tendências de pesquisa, áreas de notável excelência e também potenciais lacunas a serem endereçadas. Adicionalmente, visa-se apoiar a formulação de estratégias mais eficazes para o fomento da produção acadêmica qualificada.

Dessa forma, espera-se impulsionar a colaboração científica interna. Reforça-se, assim, a importância da organização e acessibilidade das informações frente à crescente expansão de dados no ambiente digital.

1.2 Organização do Trabalho

a) **Capítulo 2:** Explora conceitos essenciais de *Business Intelligence*, abordando suas arquiteturas e sua definição, além de detalhar como funcionam os processos de ETL e ferramentas utilizadas para análise e visualização de dados.

b) **Capítulo 3:** Descreve o sistema *Ruralytics*, detalhando sua motivação, requisitos funcionais e não funcionais, regras de negócio, casos de uso e a modelagem de dados.

c) **Capítulo 4:** Explica as tecnologias utilizadas no desenvolvimento do sistema, a estrutura arquitetural adotada e o funcionamento dos processos automatizados de ETL e visualização de dados.

d) **Capítulo 5:** Apresenta as considerações finais sobre o projeto, suas contribuições para a gestão acadêmica, além de discutir limitações e propostas para trabalhos futuros.

Capítulo 2

Business Inteligence

O vertiginoso aumento no volume de dados gerados pelas organizações nas últimas décadas, abrangendo desde transações operacionais até vastos repositórios de pesquisa científica, tornou essencial a adoção de ferramentas sofisticadas. Estas são cruciais para transformar o dilúvio de informações brutas em conhecimento útil, que efetivamente subsidie a tomada de decisões estratégicas.

Nesse contexto, o *Business Intelligence* (BI) estabeleceu-se como um conjunto fundamental de metodologias, tecnologias e práticas. Ele é voltado à coleta, ao tratamento, à análise e à visualização de dados, provendo suporte tático e estratégico às instituições. Essa abordagem permite um olhar retrospectivo e introspectivo sobre as operações e resultados.

Em uma evolução natural e complementar, o campo de *Analytics* surge para intensificar essa capacidade investigativa. Utilizando métodos estatísticos avançados e modelagem preditiva, *Analytics* busca não apenas descrever o passado, mas também antecipar cenários futuros e prescrever ações. A aplicação sinérgica de *Business Intelligence* e *Analytics* em ambientes acadêmicos, como universidades, possibilita um monitoramento mais rico da produção científica.

Essa combinação aprimora o acompanhamento de indicadores de desempenho e fomenta a formulação de políticas institucionais mais robustas, integralmente baseadas em evidências concretas. A análise aprofundada dos dados acadêmicos

pode revelar tendências emergentes, otimizar a alocação de recursos e identificar oportunidades de colaboração e inovação científica.

Este capítulo se propõe a apresentar os principais conceitos relacionados ao BI e *Analytics*. Serão exploradas suas arquiteturas, as diversas ferramentas disponíveis e o crucial processo de Extração, Transformação e Carga (ETL). O intuito é ressaltar a importância vital dessas tecnologias para uma gestão eficiente da informação e para o avanço do planejamento estratégico no contexto institucional.

2.1 Definição

A crescente complexidade do ambiente de negócios e a necessidade de respostas ágeis e informadas impulsionaram as organizações a buscar sistemas que forneçam *insights* estratégicos. Neste contexto, o BI configura-se como um ecossistema integrado, englobando arquiteturas, ferramentas, aplicações e metodologias, dedicado a converter dados brutos em conhecimento prático e acionável (TURBAN; SHARDA; DELEN, 2014).

O BI desempenha um papel fundamental no suporte à tomada de decisões estratégicas, permitindo que organizações transformem grandes volumes de dados em informações e *insights* valiosos. No entanto, a eficácia do BI depende diretamente da qualidade dos dados analisados, o que torna o processo de ETL essencial para garantir a consistência e a confiabilidade das informações. O funcionamento eficaz do BI depende de processos estruturados de extração, transformação e carga de dados (ETL), garantindo que as informações utilizadas nas análises sejam confiáveis e consistentes (KIMBALL; ROSS, 2013).

A combinação dessas tecnologias permite que empresas de diferentes setores obtenham vantagem competitiva ao tomar decisões baseadas em dados estruturados e bem analisados. Conforme defendido por Davenport e Harris (2007), organizações que competem em análise utilizam intensamente dados para guiar suas estratégias, obtendo vantagem competitiva.

Embora o termo BI tenha sido popularizado mais recentemente, suas raízes

conceituais podem ser traçadas até os primeiros sistemas de apoio à decisão da década de 1960. Para Turban, Sharda e Delen (2014), BI é um construto multifacetado que integra tecnologias e aplicações com o objetivo de analisar dados e apresentar informações complexas de forma compreensível, facilitando decisões mais informadas e assertivas.

Complementarmente, Kimball e Ross (2013) descrevem BI como um processo iterativo, englobando a coleta, organização, análise e visualização de dados, visando prover informações estratégicas. Essa abordagem processual destaca as etapas necessárias para que gestores efetivamente tomem decisões baseadas em evidências, minimizando riscos e otimizando suas operações.

A função primordial do BI é prover uma compreensão clara do desempenho passado e presente da organização. Isso é alcançado através da consolidação de dados históricos em repositórios centrais, como os *Data Warehouses*, conforme preconizado por Inmon (2005), permitindo análises eficientes e o monitoramento de Indicadores-Chave de Desempenho (KPIs) através de relatórios, painéis e consultas.

A implementação de uma solução de BI inicia-se com o fundamental processo de ETL (Extração, Transformação e Carga). Esta etapa crucial envolve a coleta de dados de múltiplas fontes, sua adequação e subsequente armazenamento em um *Data Warehouse* ou *Data Mart*, conforme detalhado por Kimball e Ross (2013), assegurando a qualidade dos dados.

Após a adequada preparação dos dados, a etapa de análise é frequentemente conduzida com o auxílio de ferramentas especializadas, exemplificadas por plataformas como *Power BI*, *Tableau* e *Metabase*. Contudo, a capacidade de transformar os resultados analíticos em *insights* compreensíveis e acionáveis reside na aplicação de uma visualização de dados eficaz. Os princípios fundamentais para o design de tais visualizações, que comunicam informações de forma clara e concisa, são explorados por Few (2006).

2.2 Arquiteturas de BI

A implementação de BI em uma organização exige uma estrutura bem definida para a coleta, processamento, armazenamento e análise de dados (TURBAN; SHARDA; DELEN, 2014). A arquitetura de BI determina como essas etapas são realizadas, garantindo que as informações estejam acessíveis e prontas para apoiar decisões estratégicas.

Diversos modelos arquitetônicos evoluíram para atender a variadas necessidades empresariais. Estes vão desde as abordagens tradicionais baseadas em *Data Warehouses* (INMON, 2005; KIMBALL; ROSS, 2013) até modelos modernos como *Data Lakes* e arquiteturas híbridas, impulsionadas pela era do *Big Data* (TURBAN; SHARDA; DELEN, 2014).

A escolha da arquitetura mais adequada é influenciada por fatores como volume de dados, complexidade analítica e a necessidade de atualização em tempo real (TURBAN; SHARDA; DELEN, 2014). Modelos estruturados, como os *Data Warehouses*, são valorizados pelo alto desempenho em consultas analíticas tradicionais e pela robusta governança de dados (KIMBALL; ROSS, 2013).

Em contraste, *Data Lakes* oferecem maior flexibilidade, permitindo o armazenamento de grandes volumes de dados brutos em seus formatos originais. Isso facilita diversas análises exploratórias e avançadas (SAWADOGO; DARMONT, 2021). As arquiteturas híbridas emergem como uma solução intermediária, buscando combinar a governança dos dados estruturados com a escalabilidade e flexibilidade das abordagens modernas encontradas em ambientes de Big Data (DEDIĆ; STANIER, 2017).

A arquitetura tradicional de BI é baseada no conceito de *Data Warehouse*, conforme proposto por Inmon (2005) e aprimorado por Kimball e Ross (2013). Nesse modelo, os dados são extraídos de sistemas transacionais, transformados para atender a padrões analíticos e armazenados em um banco centralizado otimizado para consultas. Essa estrutura permite análises consistentes e seguras, sendo amplamente utilizada para relatórios corporativos e acompanhamento de indicadores de desempe-

nho. A Figura 2.1 ilustra um exemplo de arquitetura de uma *Data Warehouse*.

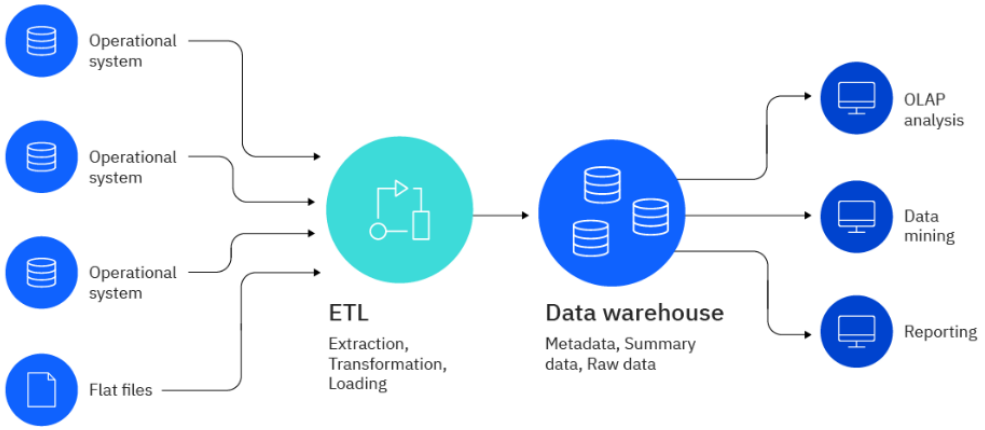


Figura 2.1: Exemplo de arquitetura de um *Data Warehouse*.
Fonte: Adaptado de IBM (2025)[1]

Uma variação desse modelo é a arquitetura baseada em *Data Marts*, que segmenta os dados em áreas específicas, como marketing, finanças e vendas. Segundo Kimball e Ross (2013), esse modelo oferece maior flexibilidade para departamentos individuais, permitindo consultas mais rápidas e específicas. No entanto, a descentralização dos dados pode gerar inconsistências caso não haja um controle rigoroso sobre a governança da informação.

Com o crescimento exponencial da quantidade de dados gerados pelas empresas, novas abordagens foram desenvolvidas para lidar com informações não estruturadas e análises exploratórias. O conceito de *Data Lake* surgiu como alternativa ao *Data Warehouse*, permitindo o armazenamento de grandes volumes de dados brutos em diferentes formatos (DAVENPORT, 2014).

Em um *Data Lake*, as informações são ingeridas diretamente de diversas fontes sem transformação prévia, possibilitando que cientistas de dados e analistas explorem os dados conforme necessário. Essa abordagem é amplamente utilizada em aplicações de *Big Data* e aprendizado de máquina, onde os dados podem ser processados sob demanda para diferentes tipos de análises. Um exemplo de arquitetura está na Figura 2.2.

Apesar da flexibilidade dos *Data Lakes*, um dos desafios dessa arquitetura é a necessidade de governança eficiente. A falta de estruturação pode dificultar a

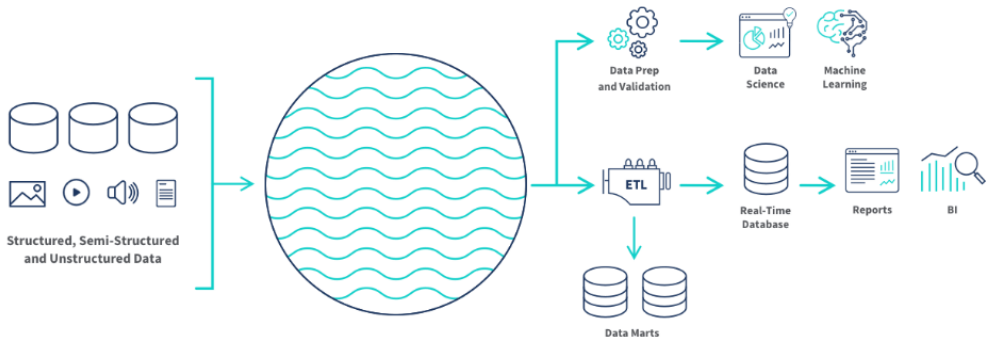


Figura 2.2: Exemplo de arquitetura de um *Data Lake*.
Fonte: Adaptado de Qlik (2025)[2]

extração de *insights*, tornando essencial o uso de mecanismos para organização e catalogação dos dados armazenados (SAWADOGO; DARMONT, 2021). Além disso, o desempenho das consultas pode ser impactado, especialmente em cenários que exigem respostas rápidas para análises operacionais.

Para unir as vantagens das abordagens estruturadas e não estruturadas, surgiram as arquiteturas híbridas, que combinam *Data Warehouses* e *Data Lakes* em um único ambiente analítico (CHAUDHURI; DAYAL; NARASAYYA, 2011). Nesse modelo, dados estruturados são armazenados em um *Data Warehouse*, garantindo consultas rápidas e consistentes, enquanto informações semiestruturadas e não estruturadas são mantidas em um *Data Lake* para análises mais avançadas.

Um desses modelos híbridos que emerge visando unificar e simplificar o ecossistema de dados é o *Data Lakehouse*. A proposta fundamental desse modelo é implementar uma camada de gerenciamento de dados tradicionalmente associada aos *Data Warehouses* diretamente sobre o armazenamento de baixo custo dos *Data Lakes* (ARMBRUST et al., 2020), como pode ser visto na Figura 2.3. Isso elimina a necessidade de múltiplos sistemas e a redundância de dados, permitindo que uma única cópia dos dados seja utilizada tanto para cargas de trabalho de BI, que exigem alto desempenho, quanto para tarefas de ciência de dados e aprendizado de máquina.

Outro modelo de arquitetura que vem ganhando destaque é o BI em Tempo Real, focado na análise contínua de dados em fluxo (CHAUDHURI; DAYAL; NARA-

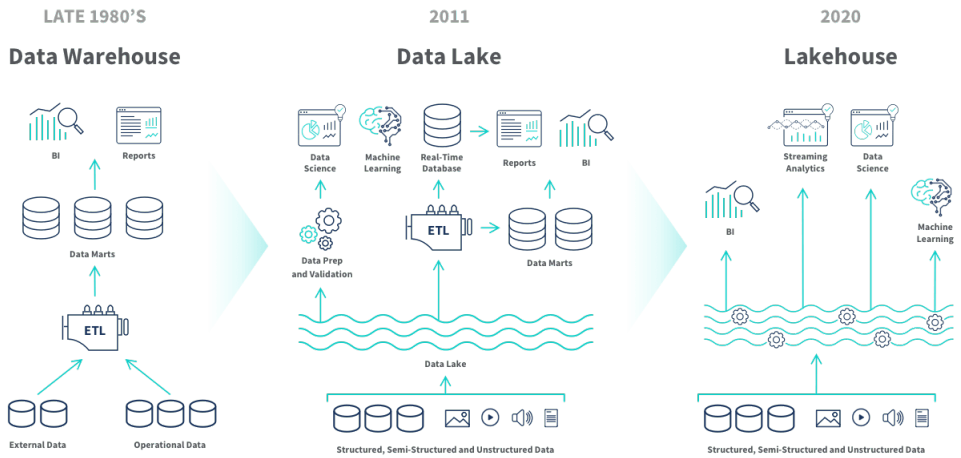


Figura 2.3: Exemplo de arquitetura de um *Data Lakehouse*.
Fonte: Adaptado de Qlik (2025)[3]

SAYYA, 2011). Conforme apontado por este autor, tal abordagem busca minimizar radicalmente a latência entre a geração do dado e sua efetiva análise. Diferentemente do BI tradicional, que opera com cargas periódicas, este modelo processa informações à medida que são geradas.

A consolidação dessas arquiteturas de BI em Tempo Real é impulsionada pelo avanço do processamento distribuído. Ferramentas como *Apache Kafka* são cruciais neste ecossistema, atuando como plataformas de *streaming* que ingerem e gerenciam fluxos de dados de alta vazão, com baixa latência e alta confiabilidade (KREPS; NARKHEDE; RAO, 2011). Esses dados são então direcionados para sistemas de processamento analítico.

Sistemas como o *Apache Spark Streaming* são frequentemente empregados para realizar essa análise sofisticada. Utilizando conceitos como o processamento de micro-lotes ou mesmo o processamento contínuo de eventos, permitem a geração de *insights* a partir dos fluxos de dados em tempo real ou quase real (ZAHARIA et al., 2013), transformando a interação organizacional com as operações.

A abordagem de BI em tempo real é, portanto, essencial para setores dinâmicos como finanças, para detecção de fraudes e análise de risco; telecomunicações, para monitoramento de rede e personalização de ofertas; e comércio eletrônico, para precificação dinâmica e recomendação em tempo real. Nestes contextos, decisões

rápidas baseadas em dados frescos conferem vantagens competitivas significativas (TURBAN; SHARDA; DELEN, 2014).

2.3 ETL

O processo de ETL é fundamental para a implementação eficaz do BI, pois garante que os dados extraídos de múltiplas fontes estejam organizados e padronizados antes de serem analisados. De acordo com Kimball e Ross (2013), o ETL é composto por três fases principais: extração, transformação e carga.

Na fase de extração, os dados são coletados de diferentes fontes, como bancos de dados relacionais, serviços via Interfaces de programação de aplicativos (API)s e arquivos externos. Essa etapa inicial do processo de ETL pode apresentar desafios importantes, como a integração de dados heterogêneos, a identificação de inconsistências e a manipulação de diferentes formatos. Além disso, é fundamental assegurar que os dados estejam atualizados e representem com fidelidade as informações de origem, garantindo a qualidade necessária para as fases subsequentes (INMON, 2005).

A etapa de transformação é responsável pela limpeza, padronização e estruturação dos dados extraídos, assegurando que estejam adequados ao processo analítico. Essa fase desempenha papel fundamental na melhoria da qualidade da informação, por meio da aplicação de técnicas como normalização, agregação e enriquecimento semântico (MOSS; ATRE, 2003).

Além disso, podem ser realizadas conversões de tipos de dados, padronização de nomenclaturas e tratamento de valores ausentes. Tais procedimentos visam garantir a uniformidade e a consistência dos dados que serão armazenados, permitindo análises mais precisas e confiáveis nas etapas subsequentes.

A última etapa do ETL, a carga, consiste na inserção dos dados transformados em um repositório final, geralmente um *Data Warehouse* ou *Data Lake*. Segundo Chaudhuri, Dayal e Narasayya (2011), essa etapa pode ser realizada de forma incremental, onde apenas novos registros são adicionados, ou por meio de cargas completas, onde todos os dados são substituídos periodicamente.

O sucesso de um sistema de BI depende diretamente da qualidade dos dados analisados, tornando o ETL um componente crítico desse processo. Segundo Kimball e Ross (2013), dados inconsistentes ou mal estruturados podem comprometer a confiabilidade das análises e levar a decisões equivocadas. O ETL assegura que as informações estejam corretas, eliminando duplicidades, preenchendo lacunas e padronizando formatos.

Empresas que utilizam ETL de maneira eficiente conseguem integrar dados de diferentes departamentos e consolidar uma visão unificada do negócio. Isso permite que gestores tenham acesso a informações confiáveis e atualizadas, aumentando a precisão das previsões estratégicas e reduzindo a incerteza na tomada de decisão (LOSHIN, 2012).

Com o avanço das tecnologias de *Big Data*, novas abordagens de ETL vêm sendo desenvolvidas para lidar com grandes volumes de dados de forma mais ágil. Ferramentas como *Apache Spark* e *Talend* permitem o processamento distribuído de dados, tornando o ETL mais rápido e escalável (CHAUDHURI; DAYAL; NARASAYYA, 2011).

2.4 Ferramentas para BI

A escolha das ferramentas e tecnologias utilizadas em BI deve considerar fatores como a complexidade dos dados, o volume de informações e os objetivos estratégicos da organização. Ferramentas como *Power BI* e *Tableau* são amplamente adotadas por sua capacidade de integrar diversas fontes de dados e gerar visualizações intuitivas e interativas. Já soluções *open-source*, como *Metabase*, apresentam-se como alternativas acessíveis para instituições que desejam implementar BI com menor investimento financeiro.

Além das ferramentas de visualização, é essencial considerar as plataformas de armazenamento e processamento de dados, que compõem a base estrutural do BI. Segundo Loshin (2012), tecnologias como *Amazon Redshift*, *Google BigQuery* e *Snowflake* oferecem escalabilidade e alto desempenho, permitindo armazenar grandes

volumes de dados e realizar consultas de forma ágil. Essas soluções geralmente são integradas a processos de ETL, garantindo que os dados analisados estejam consistentes e organizados.

Com o avanço da demanda por análises cada vez mais sofisticadas, diversas ferramentas especializadas surgiram no mercado, com funcionalidades que variam em termos de usabilidade, integração e escalabilidade. Além das já citadas *Power BI* e *Tableau*, destacam-se plataformas como *Apache Superset*, *Cube.js* e *Redash*, que atendem a diferentes perfis de usuários, desde analistas com pouca experiência técnica até desenvolvedores que buscam soluções personalizadas e integradas aos seus sistemas.

Essas ferramentas desempenham papel fundamental na transformação de dados em informações acionáveis, ao permitir a criação de *dashboards* dinâmicos, relatórios customizados e análises em tempo real. A escolha adequada da tecnologia influencia diretamente a eficiência do processo decisório, possibilitando uma visão clara do desempenho organizacional e contribuindo para a construção de uma cultura orientada por dados.

2.4.1 *Power BI*

O *Power BI* é uma plataforma de BI desenvolvida pela *Microsoft*, amplamente utilizada devido à sua integração nativa com o ecossistema da empresa, como *Excel*, *Azure* e *SQL Server*.

Segundo a documentação oficial da *Microsoft*[4], a ferramenta oferece um conjunto completo de funcionalidades para análise de dados, incluindo suporte para modelagem, transformação e visualização interativa. Seu diferencial está na capacidade de conectar-se a múltiplas fontes de dados e na integração com Inteligência Artificial (IA), permitindo a criação de *insights* automatizados. A Figura 2.4 ilustra um painel da ferramenta, com exemplos das visualizações interativas que podem ser criadas a partir dos dados.

[4]<https://www.microsoft.com/pt-br/power-platform/products/power-bi>

A principal limitação do *Power BI* está na sua dependência do ambiente *Microsoft*, sendo menos flexível para usuários que utilizam tecnologias externas. Além disso, apesar da existência de uma versão gratuita, os recursos avançados são acessíveis apenas na versão *Power BI Pro*, que requer uma assinatura paga.

Figura 2.4: Captura de tela disponibilizada pela plataforma *Power BI*, contendo algumas análises exemplo do *software*.

2.4.2 *Tableau*

O *Tableau*[5] é uma das soluções de BI mais poderosas do mercado, conhecida por sua capacidade avançada de análise visual. Desenvolvido pela *Salesforce*, ele permite a criação de *dashboards* altamente interativos e detalhados. Essa abordagem capacita os usuários a explorar visualmente os dados, descobrindo padrões e tendências de forma intuitiva, como demonstra o painel exemplificado na Figura 2.5.

A ferramenta se destaca por sua usabilidade intuitiva, permitindo que analistas explorem dados sem necessidade de programação avançada. Seu motor de processamento otimizado permite a manipulação de grandes volumes de dados, tornando-o uma escolha robusta para empresas que lidam com *Big Data*. No entanto, o custo do *Tableau* é significativamente mais alto em comparação a outras ferramentas, sendo um fator limitante para pequenas e médias empresas.

[5] <https://www.tableau.com>

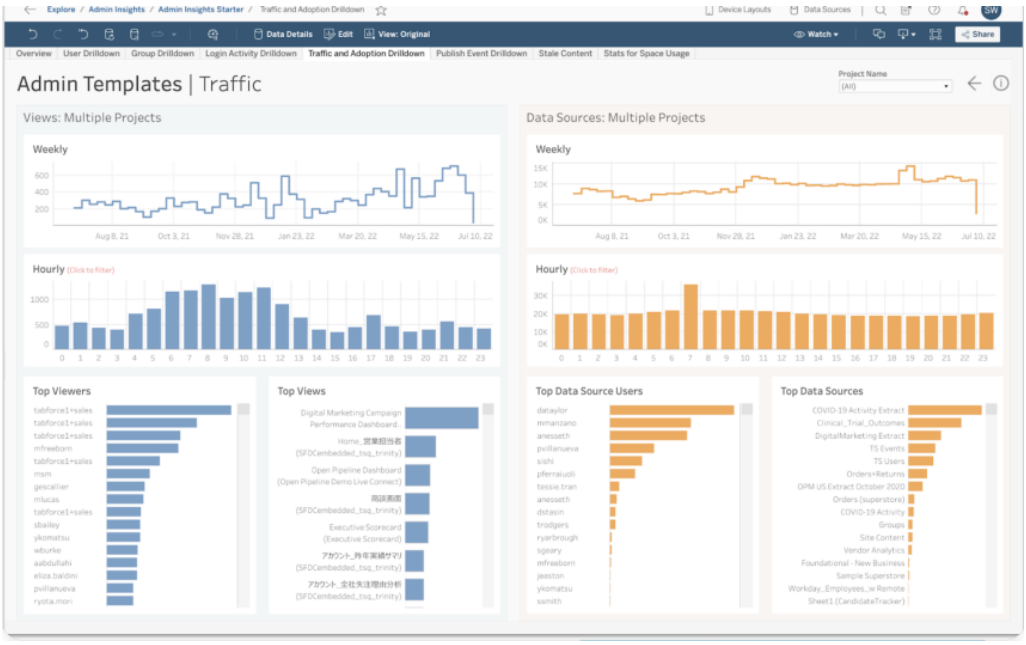


Figura 2.5: Captura de tela disponibilizada pela plataforma *Tableau*, a qual contém uma análise exemplo do *software*.

2.4.3 Metabase

O *Metabase*[6] é uma plataforma *open-source* de BI voltada para usuários que desejam realizar análises de dados de forma simples, sem a necessidade de conhecimento avançado em *Structured Query Language* (SQL). Seu principal diferencial está na interface intuitiva e acessível, que permite a criação de *dashboards* e relatórios de maneira prática. A Figura 2.6, por exemplo, exibe um painel interativo para análise de um cenário de vendas, demonstrando como a ferramenta facilita a exploração de dados mesmo por usuários não técnicos.

Além disso, o *Metabase* oferece integração com diversos bancos de dados, como *MySQL*, *PostgreSQL*, *SQL Server* e *BigQuery*, ampliando sua aplicabilidade em diferentes contextos organizacionais. Trata-se de uma solução ideal para empresas que buscam uma ferramenta de BI funcional, de rápida implementação e com baixo custo, sendo amplamente adotada por *startups* e equipes com recursos limitados.

[6] <https://www.metabase.com>

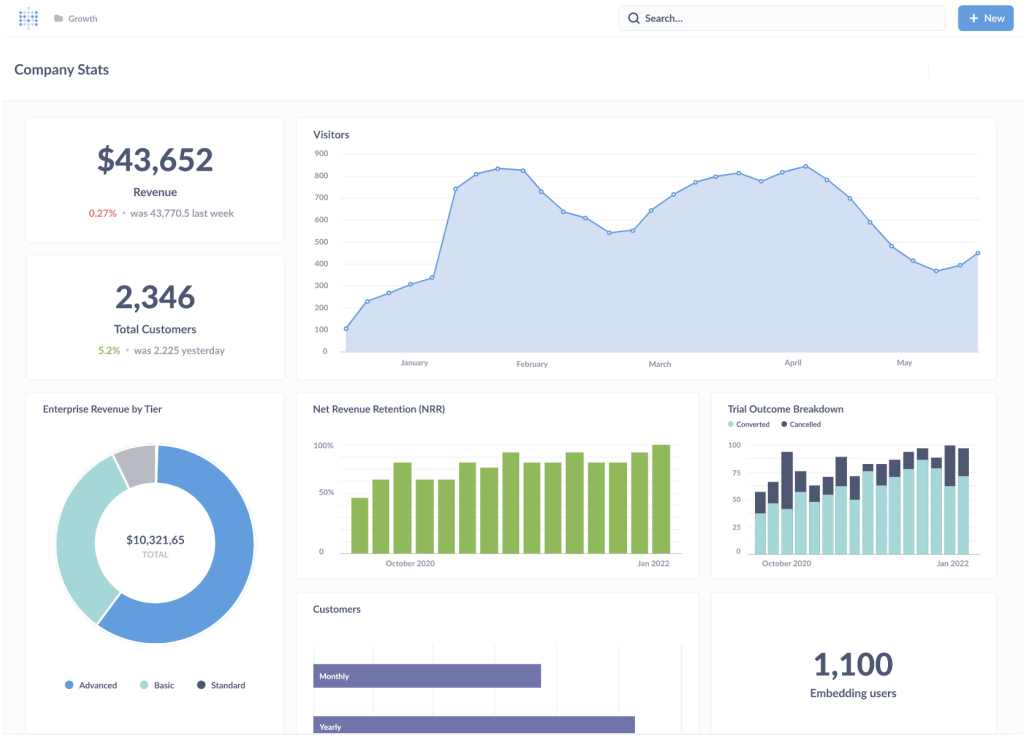


Figura 2.6: Captura de tela disponibilizada na documentação oficial da plataforma *Metabase*, a qual contém exemplo do uso *software*.

2.4.4 Apache Superset

O *Apache Superset*[7] é uma alternativa *open-source* ao *Tableau*, desenvolvido pela *Apache Software Foundation*. A ferramenta é altamente escalável e voltada para análises interativas de grandes volumes de dados, sendo indicada para ambientes com demandas analíticas mais complexas. Possui suporte a múltiplas fontes de dados e permite customizações avançadas, oferecendo flexibilidade na construção de *dashboards* personalizados. A Figura 2.7 demonstra um exemplo dessa capacidade, exibindo a riqueza de visualizações que a plataforma pode gerar.

Apesar de sua robustez e potencial analítico, o *Superset* apresenta uma curva de aprendizado mais acentuada em relação a ferramentas como o *Metabase*. Sua configuração e utilização requerem maior conhecimento técnico, especialmente em ambientes corporativos com infraestrutura mais elaborada. Ainda assim, destaca-se como uma solução poderosa para organizações que necessitam de maior controle e sofisticação nas análises de dados.

[7] <https://superset.apache.org>

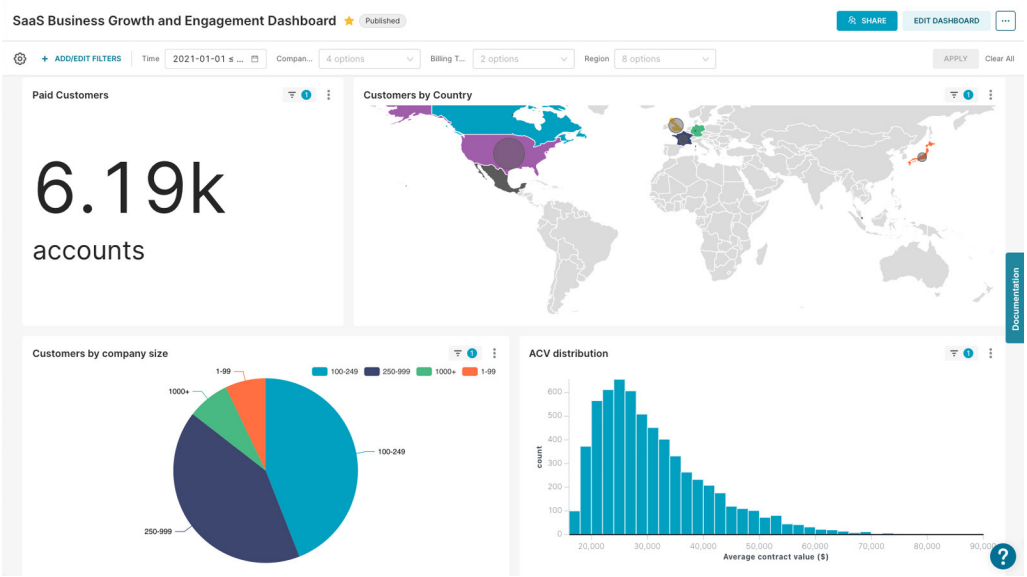


Figura 2.7: Captura de tela disponibilizada na plataforma *Apache Superset*, a qual contém exemplo do gráficos gerados pelo *software*.

2.4.5 Cube

O *Cube*[8] é um *framework* voltado para o desenvolvimento de aplicações analíticas e visualizações integradas a produtos web. Seu principal diferencial reside na criação de APIs analíticas, possibilitando que desenvolvedores incorporem visualizações de BI diretamente em suas aplicações, o que o caracteriza como uma solução flexível e altamente personalizável.

Dessa forma, o *Cube* se destaca como uma alternativa interessante para organizações que desejam embutir *dashboards* sob medida em seus sistemas. No entanto, por não se tratar de uma ferramenta *plug-and-play*, como o *Metabase* ou o *Superset*, sua adoção é mais indicada para desenvolvedores que buscam autonomia e flexibilidade na construção de soluções analíticas alinhadas às necessidades específicas de seus produtos ou serviços.

2.4.6 Redash

O *Redash*[9] é uma ferramenta de BI *open-source*, amplamente utilizada para a execução de consultas SQL e criação de visualizações interativas. Seu principal

[8] <https://cube.dev>
[9] <https://redash.io>

diferencial está na simplicidade com que permite manipular dados por meio de SQL, facilitando a criação de *dashboards* por analistas e cientistas de dados de forma rápida e eficiente. A Figura 2.8 ilustra precisamente esse resultado: um *dashboard* denso em informações, que combina múltiplos gráficos e painéis em uma única tela para monitoramento e análise.

Além disso, o *Redash* oferece suporte a diversas fontes de dados e integrações com serviços externos, tornando-se uma solução versátil para ambientes analíticos. No entanto, sua dependência do conhecimento em SQL pode limitar o acesso a usuários menos experientes, o que o torna menos intuitivo quando comparado a ferramentas com interfaces mais amigáveis, como *Metabase* ou *Power BI*.

Figura 2.8: Captura de tela da plataforma *Redash* de um *dashboard* gerado pelo *software*.

Capítulo 3

Ruralytics

A Plataforma Lattes é um centro de informação de produção acadêmica, e sua relevância é tamanha de modo a ser um dos pilares da informação científica nacional, mas, paradoxalmente, são visíveis os desafios inerentes à análise estratégica e consolidada da vasta produção acadêmica nela registrada. A dificuldade em extrair, processar e interpretar esses dados de forma eficaz impede que as IES aproveitem plenamente esse rico acervo para aprimorar sua gestão, identificar potencialidades e ampliar sua visibilidade.

Diante desse cenário, este capítulo se propõe a transpor a discussão do problema para a construção de uma solução, apresentando a proposta de desenvolvimento de um sistema para monitoramento e análise da produção acadêmica, baseado nos princípios do BI. A solução busca suprir a ausência em diversas IES de uma ferramenta institucional voltada à organização e gestão das informações científicas, contribuindo para maior transparência e eficiência na tomada de decisões.

A proposta abrange a definição dos requisitos funcionais e não funcionais, além das regras de negócio que garantem a coerência dos processos institucionais. Também são descritos os principais casos de uso, que representam as interações previstas entre usuários e o sistema em diferentes cenários operacionais. Adicionalmente, será apresentada a modelagem de dados, estruturada com foco na integridade relacional e escalabilidade, elementos cruciais para a robustez e eficácia da solução proposta. A

aplicação e validação do sistema poderão ser realizadas por meio de um estudo de caso em uma IES específica.

3.1 Motivação

A necessidade de ferramentas que permitam o monitoramento e a análise da produção acadêmica de forma sistemática tem se intensificado. Observa-se, em muitas IES, uma lacuna significativa nesse sentido. Enquanto diversas instituições já buscam ou dispõem de sistemas para análise e organização de dados científicos, como apontado por estudos como o de Silva e Brandão (2011), muitas ainda carecem de uma ferramenta própria robusta, voltada à sistematização e ao acompanhamento detalhado de sua produção.

Essa ausência compromete não apenas a visibilidade da produção científica institucional, mas também prejudica o planejamento estratégico da gestão universitária, que frequentemente carece de indicadores precisos e atualizados para subsidiar decisões assertivas. A falta de dados consolidados impede a identificação clara de áreas de excelência, eventuais lacunas produtivas e o monitoramento efetivo do desempenho acadêmico ao longo do tempo.

Nesse contexto, soluções de BI e *analytics* mostram-se cada vez mais relevantes para o ambiente acadêmico. Plataformas dessa natureza são fundamentais para consolidar e estruturar informações dispersas, possibilitando sua transformação em conhecimento estratégico, conforme apontado por Campos, Silva e Costa (2019).

Tais ferramentas promovem uma análise detalhada e dinâmica da produção científica, permitindo maior transparência, eficiência na gestão e um incentivo mais direcionado à pesquisa. O uso de ferramentas de BI em instituições de ensino superior, como ressaltado por Júnior et al. (2022), permite análises comparativas, a geração de relatórios personalizados e um apoio substancial à formulação de políticas institucionais. A implantação de um sistema com essas características representa, portanto, não apenas uma inovação tecnológica, mas uma resposta concreta a uma necessidade institucional latente em muitas IES.

É precisamente a identificação dessa lacuna e a convicção no potencial transformador da informação que impulsionam a concepção do sistema proposto neste trabalho. O desenvolvimento desta ferramenta tem origem na perspectiva de auxiliar as instituições de ensino e pesquisa a analisarem seus próprios dados de maneira mais clara, profunda e estratégica. Desta forma, o sistema visa facilitar a conversão da riqueza informacional, frequentemente subaproveitada, em um recurso valioso para a gestão do conhecimento e para o avanço científico.

Este projeto é motivado pela aspiração de contribuir para a valorização da produção intelectual e para a melhoria contínua dos processos decisórios no ambiente universitário. Almeja-se que, ao consolidar dados dispersos e fornecer acesso dinâmico a indicadores e análises relevantes, o sistema possa catalisar uma gestão mais eficiente do conhecimento produzido. A intenção é que a informação, uma vez organizada e acessível, sirva de alicerce para o reconhecimento das contribuições acadêmicas e para o fortalecimento da cultura de planejamento baseado em evidências.

Em acréscimo, a motivação se estende ao impacto positivo na comunidade acadêmica. Ao permitir o acompanhamento sistemático da produção científica por diferentes granularidades – seja por departamentos, cursos ou pesquisadores individuais – fomenta-se a transparência institucional. Essa funcionalidade não apenas democratiza o acesso à informação, mas também tem o potencial de ampliar a disseminação dos resultados de pesquisa e de estimular a colaboração interna, ao revelar sinergias e competências complementares.

Portanto, a força motriz deste trabalho é a crença de que a otimização da gestão acadêmica, o fortalecimento da cultura de dados e a promoção da visibilidade científica são interdependentes e cruciais para o progresso das IES. O desenvolvimento de um sistema de *analytics* dedicado à produção acadêmica é visto como um passo fundamental nessa direção, uma contribuição tangível para que as instituições possam não apenas gerenciar sua produção, mas, sobretudo, compreendê-la, valorizá-la e utilizá-la como motor para a excelência em pesquisa, ensino e extensão.

A solução desenvolvida possibilitará o acesso dinâmico a indicadores e relatórios analíticos por meio de *dashboards* interativos, permitindo o acompanhamento

sistemático da produção científica por departamentos, cursos ou docentes.

3.2 Trabalhos relacionados

A proposta deste trabalho inspira-se em exemplos bem-sucedidos de outras instituições de ensino superior, que servem como referência para o seu desenvolvimento. Destacam-se, nesse sentido, o sistema SOMOS, da Universidade Federal de Minas Gerais (UFMG), e o Repositório Institucional da Universidade de São Paulo (USP), ambos com foco na organização, monitoramento e análise da produção científica. Essas plataformas têm contribuído de maneira significativa para a gestão acadêmica e visibilidade institucional, evidenciando a importância de soluções voltadas ao acompanhamento da produção intelectual universitária.

O SOMOS (Sistema de Organização e Monitoramento da Produção Científica) da UFMG, por exemplo, centraliza e processa dados sobre docentes, publicações, projetos e outras atividades acadêmicas, oferecendo painéis analíticos (*dashboards*) que permitem o acompanhamento de indicadores-chave. Essas visualizações facilitam o planejamento estratégico e a alocação de recursos nas unidades acadêmicas. Um exemplo dessa funcionalidade é a análise da evolução da produção bibliográfica ao longo do tempo, conforme ilustrado na Figura 3.1, que permite identificar tendências e picos de produtividade (GERAIS, 2025).

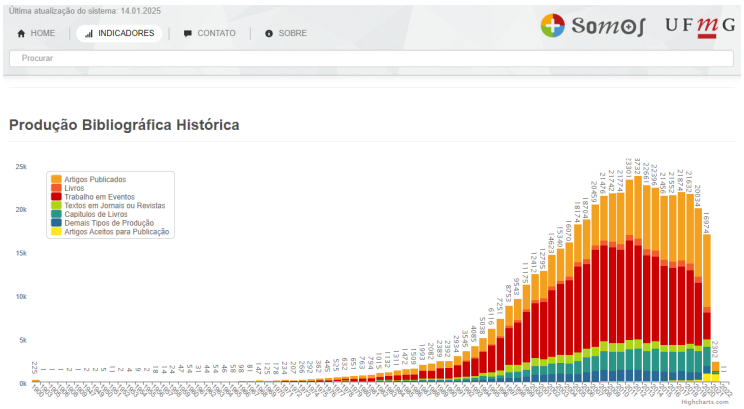


Figura 3.1: Tela do sistema SOMOS UFMG contendo o indicativo de produção bibliográfica no tempo.

Fonte: (GERAIS, 2025)

De forma complementar, o Repositório da USP foca na disseminação e na acessibilidade do conhecimento, disponibilizando um acervo organizado que possibilita buscas refinadas por tipo de produção, autor, unidade e área do conhecimento, servindo como uma vitrine da excelência acadêmica da instituição (PAULO, 2025).

Outro destaque é o *ScriptLattes*, uma ferramenta de código aberto que extrai e processa os currículos da Plataforma Lattes para gerar relatórios e páginas *web* estáticas com a produção acadêmica consolidada . Sua ampla adoção por programas de pós-graduação e grupos de pesquisa no Brasil evidencia a importância da automatização na coleta de dados curriculares, servindo como um passo essencial para a posterior análise e visualização em sistemas de inteligência, como o proposto neste trabalho. (MENA-CHALCO; JR, 2009)

Essas iniciativas exemplificam a aplicação bem-sucedida de princípios de BI no setor educacional. Ferramentas digitais como estas são fundamentais para transformar o vasto ecossistema de dados acadêmicos, muitas vezes dispersos e heterogêneos, em inteligência estratégica. Como reforçam Campos, Silva e Costa (2019), sistemas de BI permitem não apenas a consolidação de grandes volumes de informações, mas, principalmente, a geração de indicadores de desempenho úteis para a avaliação institucional e para a elaboração de políticas científicas e tecnológicas.

Nessa mesma linha, Júnior et al. (2022) argumenta que os sistemas analíticos adotados por IES contribuem diretamente para a melhoria da eficiência e da governança. Ao promover maior transparência sobre os resultados da pesquisa e demais atividades-fim, essas plataformas capacitam os gestores para uma tomada de decisão mais ágil e informada, fomentando uma cultura institucional orientada por dados e focada na melhoria contínua. Portanto, a análise dessas iniciativas e da literatura correlata valida a relevância da proposta deste trabalho e oferece um roteiro consolidado para o desenvolvimento de uma ferramenta com impacto similar.

3.3 Requisitos do Sistema

Esta seção apresenta os elementos estruturantes que orientam o desenvolvimento do sistema proposto para monitoramento e análise de produção acadêmica. São descritos, de forma detalhada, os requisitos funcionais, as regras de negócio e os casos de uso, componentes fundamentais no processo de especificação e modelagem de sistemas.

3.3.1 Requisitos Funcionais

Os Requisitos Funcionais (RF) definem as funcionalidades que o sistema deverá disponibilizar aos usuários, especificando o comportamento esperado diante de diferentes ações. Eles garantem que o sistema atenda às necessidades operacionais e institucionais, orientando o processo de desenvolvimento e validação.

a) **RF01:** O sistema deve permitir a importação de dados de produção acadêmica a partir de arquivos de currículos no formato *eXtensible Markup Language* (XML) oriundos do sistema Lattes.

b) **RF02:** O sistema deve ser capaz de realizar o *parsing* (análise sintática) dos dados dos currículos Lattes para extrair informações detalhadas dos pesquisadores sobre publicações (artigos, livros, capítulos, trabalhos em anais), produção técnica, produção artística, orientações (concluídas e em andamento), participações em bancas examinadoras e projetos de pesquisa trabalhados.

c) **RF03:** O sistema deve implementar rotinas para a limpeza e padronização dos dados extraídos, visando corrigir inconsistências comuns, normalizar termos, *e.g.* nomes de periódicos, áreas do conhecimento e tratar dados ausentes ou mal formatados.

d) **RF04:** O sistema deve permitir a atualização periódica da base de dados com novas informações de currículos Lattes, mantendo um histórico ou versionamento, se pertinente.

e) **RF05:** O sistema deve permitir a associação dos dados de produções aos respectivos docentes e suas vinculações institucionais.

f) **RF06:** O sistema deve calcular e disponibilizar um conjunto de indicadores de produção científica, como o número total de publicações por tipo e média de publicações por docente.

g) **RF07:** O sistema deve calcular e disponibilizar indicadores relacionados a outras atividades acadêmicas, como o número de orientações concluídas, o número de participações em bancas e o número de trabalhos em eventos participados.

h) **RF08:** O sistema deve permitir a agregação de dados e indicadores de produção por diferentes níveis hierárquicos e temporais, *e.g.*, por docente, por departamento, por unidade acadêmica, por instituição, por ano, por período personalizado.

i) **RF09:** O sistema deve ter a capacidade de identificar e visualizar redes de colaboração científica (coautorias) entre pesquisadores da instituição e, se possível, com pesquisadores externos.

j) **RF10:** O sistema deve apresentar os dados agregados e os indicadores calculados por meio de painéis interativos, ou seja, *dashboards*, e visualmente intuitivos.

k) **RF11:** O sistema deve possibilitar a consulta e visualização detalhada da produção individual de pesquisadores, respeitando as políticas de privacidade e acesso.

3.3.2 Requisitos Não Funcionais

Além das funcionalidades específicas que o sistema deve oferecer, é igualmente essencial considerar os Requisitos Não Funcionais (RNF), que definem atributos de qualidade, desempenho, segurança e usabilidade. Esses requisitos não descrevem diretamente o que o sistema faz, mas sim como ele deve se comportar e quais características ele deve possuir para garantir confiabilidade, eficiência e boa experiência de uso.

a) **RNF01:** A interface do sistema deve seguir princípios de *design* centrados no

usuário, sendo clara, intuitiva e acessível, de modo que docentes e visitantes possam interagir com as funcionalidades sem necessidade de treinamento técnico prévio.

b) **RNF02:** O sistema deve fornecer feedback visual claro e imediato às ações do usuário, indicando o estado do processamento ou o resultado de uma operação.

c) **RNF03:** A terminologia utilizada na interface (rótulos, mensagens, menus) deve ser consistente e alinhada ao vocabulário do domínio acadêmico e da gestão científica.

d) **RNF04:** O sistema deve garantir a proteção dos dados institucionais por meio de mecanismos robustos de segurança, incluindo autenticação por senha temporária *One-Time Password* (OTP), criptografia de informações sensíveis e controle de acesso baseado em perfis de usuário.

e) **RNF05:** O sistema deve estar protegido contra vulnerabilidades de segurança comuns em aplicações web, *e.g.* XSS e SQL Injection, seguindo boas práticas de desenvolvimento seguro.

f) **RNF06:** O código-fonte do sistema deve ser bem estruturado, modular, comentado e seguir padrões de codificação que facilitem a compreensão, manutenção e futuras evoluções por outros desenvolvedores.

g) **RNF07:** A arquitetura do sistema deve ser flexível para permitir a adição de novas fontes de dados, novos indicadores ou novas funcionalidades com esforço de desenvolvimento moderado.

3.3.3 Regras de Negócio

As Regras de Negócio (RN) deste sistema de monitoramento e gestão da produção acadêmica traduzem as diretrizes e restrições inerentes ao domínio analisado, garantindo a integridade dos dados e a coerência dos fluxos operacionais. Ao definir critérios claros de validação e controle, essas regras preservam a qualidade da informação que será manipulada e reportada pela aplicação.

a) **RN01:** Toda produção acadêmica considerada pelo sistema para fins de

análise e cômputo de indicadores deve ter sua origem e detalhes primariamente extraídos dos currículos Lattes dos pesquisadores vinculados à instituição.

b) **RN02:** Toda a produção cadastrada deve ser classificada conforme uma categoria previamente registrada, garantindo padronização para análise estatística e controle organizacional.

c) **RN03:** Para ser incluída em análises temporais e na maioria dos indicadores quantitativos, uma produção, *e.g.*, publicação, orientação, deve possuir uma data de ocorrência (publicação, defesa, início/fim) completa e válida.

d) **RN04:** Produções marcadas como "sigilosas" ou "em desenvolvimento" no currículo Lattes não devem ser publicamente exibidas em *dashboards* agregados ou relatórios gerais, a menos que haja uma política institucional específica que permita sua inclusão anonimizada ou agregada.

e) **RN05:** Palavras-chave não podem existir isoladamente no sistema; devem estar obrigatoriamente associadas a produções acadêmicas e/ou docentes específicos, mantendo a integridade semântica.

f) **RN06:** Cada combinação de docente e código OTPs deve ser única e válida apenas por um tempo determinado, impedindo reutilização indevida e garantindo segurança de acesso.

g) **RN07:** Orientações de mestrado e doutorado somente serão contabilizadas como "concluídas" para fins de indicadores de produtividade após a informação da data de defesa no currículo Lattes. Orientações em andamento podem ser listadas, mas devem ser claramente diferenciadas das concluídas.

h) **RN08:** Em publicações com coautoria, cada autor vinculado à instituição terá a produção contabilizada individualmente para seu currículo, mas, para indicadores agregados da instituição ou unidade, a publicação será contada uma única vez para evitar inflacionamento, a menos que o indicador específico preveja outra forma de contagem, como a contagem de colaborações.

i) **RN09:** O "ano de referência" para uma produção científica (especialmente publicações) é o ano de sua efetiva publicação ou disponibilização. O sistema

deve ser capaz de lidar com produções publicadas com data retroativa ou corrigida no Lattes.

3.3.4 Casos de Uso

Caso de Uso (CU) representam uma técnica amplamente utilizada na modelagem de sistemas orientados a objetos. Trata-se de uma descrição estruturada das interações entre os usuários (atores) e o sistema, com o objetivo de alcançar um determinado resultado ou serviço. Cada caso de uso descreve um comportamento funcional que o sistema deve oferecer, possibilitando a análise clara e objetiva dos requisitos funcionais sob a perspectiva do usuário.

Atores são entidades externas que interagem com o sistema para executar uma tarefa ou alcançar um objetivo. Eles podem ser usuários humanos, outros sistemas ou componentes que atuam em conjunto com o sistema principal. Com base na proposta do sistema, foram identificados os seguintes atores principais:

a) **Docente:** Usuário institucional com permissão para cadastrar seus dados e solicitar novas unidades. Sua participação é essencial na alimentação da base de dados acadêmica da universidade.

b) **Visitante:** Usuário externo que pode acessar publicamente os dados estatísticos disponibilizados pelo sistema, pesquisar por palavras-chave, docentes e produções, promovendo transparência e divulgação da produção institucional.

c) **Administrador:** Responsável por validar cadastros de unidades acadêmicas, gerenciar usuários e supervisionar a integridade dos dados inseridos no sistema. Atua como agente de controle e manutenção do funcionamento da aplicação.

d) **Sistema:** Representa o próprio sistema como ator automatizado, responsável por realizar validações, enviar notificações, responder a interações iniciadas pelos outros atores e realizar tarefas periódicas de extração de dados. Seu papel é essencial na execução automática de rotinas e regras de negócio.

A seguir, são apresentados os principais casos de uso do sistema de BI para monitoramento e gestão de produção acadêmica, considerando os atores previamente

identificados: docente, visitante, administrador e o próprio sistema.

a) **CU01 – Cadastrar Docente**

Atores: Administrador, Sistema

Descrição: Este caso de uso permite que um novo docente seja cadastrado no sistema com seus dados institucionais. O processo inicia quando o administrador informa o *e-mail* institucional do docente e o departamento ao qual ele será vinculado. Caso o departamento especificado ainda não exista no sistema, o administrador é instruído a realizar o cadastro do departamento primeiramente (conforme detalhado no CU02 – Cadastrar Instituto e Departamento). Após a inserção dos dados do docente, o sistema procede com a validação das informações fornecidas. Se os dados forem considerados válidos, o sistema registra o novo docente na base de dados, efetivando seu vínculo com o departamento indicado e, ao final, exibe uma mensagem confirmando a conclusão bem-sucedida do cadastro.

b) **CU02 – Cadastrar Instituto e Departamento**

Atores: Administrador (solicita), Sistema

Descrição: Este caso de uso permite ao administrador cadastrar novas unidades acadêmicas, como institutos e departamentos, na estrutura organizacional do sistema. Inicialmente, o administrador insere os dados referentes ao novo instituto. O sistema, então, valida essas informações e, se aprovadas, realiza o cadastro do instituto. Subsequentemente, o administrador fornece os dados do novo departamento a ser associado ao instituto ou à estrutura geral. O sistema novamente valida os dados e, sendo corretos, cadastra o novo departamento. Por fim, o sistema envia uma confirmação da operação ao administrador.

c) **CU03 – Autenticação por OTP**

Atores: Docente, Sistema

Descrição: Este caso de uso descreve o processo pelo qual um docente se autentica no sistema utilizando uma senha temporária de uso único (OTP) enviada para seu *e-mail* institucional. O fluxo se inicia quando o docente, na interface de login, solicita a autenticação. O sistema, então, gera um código OTP exclusivo e o envia para o *e-mail* previamente cadastrado do docente.

Ao receber o código, o docente o insere no campo apropriado na tela do sistema. O sistema, por sua vez, valida o código OTP submetido, verificando sua correção e se está dentro do prazo de validade estabelecido. Em caso de validação bem-sucedida, o sistema concede o acesso ao docente.

d) **CU04 – Submeter Dados Iniciais do Currículo Lattes**

Atores: Pesquisador/Docente, Sistema

Descrição: O Pesquisador/Docente acessa o sistema para fornecer os dados de seu currículo Lattes para o processamento inicial ou para uma substituição completa das informações existentes. Esta submissão pode ocorrer por meio do *upload* do arquivo XML do currículo ou através do fornecimento do seu ID Lattes, permitindo que o sistema realize a coleta dos dados. Uma vez fornecida a fonte, o sistema inicia automaticamente o processo interno de extração, *parsing* detalhado, limpeza de inconsistências, padronização de termos e armazenamento estruturado dos seus dados de produção.

e) **CU05 – Gerenciar Atualização Contínua de Dados Lattes**

Atores: Sistema

Descrição: O sistema identifica os currículos que necessitam ser atualizados. Uma vez que um currículo é selecionado para atualização, o sistema adiciona-os a uma fila. Subsequentemente, o sistema executa o ciclo completo de ETL: realiza o *parsing* das informações, aplica rotinas de limpeza e padronização e, por fim, atualiza os registros de produção do respectivo pesquisador na base de dados, garantindo que as informações disponíveis reflitam o estado mais atual do seu currículo Lattes.

f) **CU06 – Visualizar Própria Produção e Indicadores Individuais**

Atores: Pesquisador/Docente

Descrição: Após a submissão e o processamento automático de seu currículo Lattes, o Pesquisador/Docente pode acessar uma área restrita no sistema para visualizar, de forma consolidada e organizada, toda a sua produção acadêmica registrada. Esta visualização inclui o acesso a indicadores pessoais de produção (abrangendo atividades científicas, tecnológicas, artísticas e de orientação), permitindo o acompanhamento de seu desempenho individual e

da evolução de sua trajetória acadêmica.

g) **CU07 – Explorar Produção Agregada e Colaborações**

Atores: Pesquisador/Docente, Gestor Acadêmico

Descrição: Usuários autenticados no sistema podem explorar dados agregados referentes à produção científica da instituição como um todo, ou de unidades acadêmicas e departamentos específicos, conforme suas permissões de acesso. Esta funcionalidade inclui a visualização de indicadores gerais de desempenho, tendências de produção ao longo do tempo e a identificação de redes de colaboração existentes, auxiliando na contextualização da própria produção e na prospecção de potenciais parceiros para futuras pesquisas.

h) **CU08 – Analisar Desempenho de Unidade/Departamento (Visão Gerencial)**

Atores: Gestor Acadêmico

Descrição: O Gestor Acadêmico utiliza o sistema para acessar painéis (*dashboards*) interativos que apresentam indicadores e dados agregados da produção acadêmica da(s) unidade(s) ou grupo(s) de pesquisadores sob sua responsabilidade. A ferramenta permite analisar a evolução temporal da produção, comparar o desempenho entre diferentes subunidades, identificar áreas de destaque e potenciais lacunas, utilizando filtros dinâmicos por período, tipo de produção e por, área do conhecimento.

i) **CU09 – Consultar Detalhes da Produção de Docentes (Visão Gerencial)**

Atores: Gestor Acadêmico

Descrição: Dentro de sua esfera de atuação e em conformidade com as políticas de privacidade e acesso a dados da instituição, o Gestor Acadêmico pode consultar o perfil detalhado da produção de docentes vinculados à sua unidade. Esta funcionalidade tem como objetivo apoiar o acompanhamento individualizado do corpo docente, o planejamento de ações de desenvolvimento profissional e a identificação de contribuições específicas de cada pesquisador para os resultados da unidade.

j) **CU10 – Acessar Visão Pública da Produção Institucional**

Atores: Visitante/Público em Geral

Descrição: O público externo, sem necessidade de autenticação, pode acessar uma interface pública do sistema que apresenta uma visão geral e agregada da produção científica da instituição. São exibidos indicadores gerais, áreas de pesquisa em destaque e outras informações que promovem a transparência e a visibilidade da pesquisa realizada, sempre com dados devidamente anonimizados ou de natureza pública, sem expor detalhes individuais sensíveis.

k) **CU11 – Atualizar Informações de Produções (Processo Sistêmico Automático)**

Atores: Sistema

Descrição: Este caso de uso descreve um processo interno e automatizado do sistema para manter atualizados em lote os dados de produção dos docentes cujos currículos Lattes estão cadastrados, podendo ser acionado periodicamente, ou por outros gatilhos sistêmicos. O processo inicia quando o sistema identifica a necessidade de atualização. Para os currículos a serem atualizados, o sistema pode, como primeiro passo, remover ou marcar como desatualizadas as informações de produções e palavras-chave já existentes para evitar duplicidades. Em seguida, o sistema tenta obter os dados dos arquivos XML mais recentes dos docentes. Subsequentemente, realiza o tratamento desses novos dados, incluindo limpeza e padronização, e os carrega na base. Finalmente, o sistema valida as modificações e as salva, disponibilizando assim as informações de produção devidamente atualizadas para consulta e análise.

3.4 Modelagem de dados

A estruturação dos dados é um pilar fundamental na arquitetura do sistema proposto. Para organizar as informações de maneira lógica e coesa, foi desenvolvido um modelo de dados relacional que define as entidades centrais do sistema, seus respectivos atributos e os relacionamentos que governam suas interações.

Este modelo foi representado visualmente por meio de um DER, que serviu como um projeto para a criação do banco de dados físico no Sistema de Gerenciamento de

Banco de Dados (SGBD).

A estrutura detalhada, evidenciando as tabelas e suas conexões, é apresentada na Figura 3.2, destacando as principais entidades, como *Professor* e *Production*, os seus atributos mais relevantes e as cardinalidades que regem as interações entre elas. Essa visualização facilita tanto o entendimento do fluxo de dados quanto a identificação de dependências críticas e possíveis pontos de otimização.

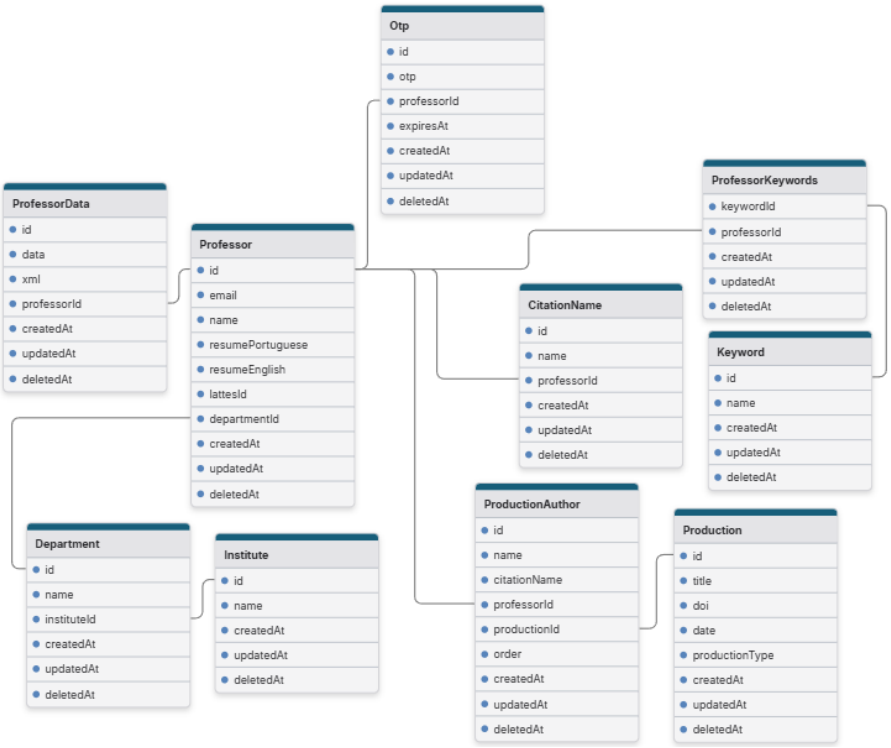


Figura 3.2: DER do banco de dados para o sistema de monitoramento da produção acadêmica

3.4.1 Estrutura das Tabelas

O sistema foi projetado com nove tabelas principais, distribuídas de forma a garantir integridade referencial, normalização e flexibilidade para futuras expansões. A tabela *Institute* representa as instituições de ensino cadastradas, contendo o identificador, nome e relações com departamentos. A tabela *Department* associa-se a um instituto, armazenando informações sobre os departamentos acadêmicos.

A tabela Professor armazena os dados dos pesquisadores cadastrados no sistema, como nome completo, *e-mail* e vínculo com um departamento específico. A tabela *Otp*

registra os códigos temporários gerados para validação de ações, estando associada diretamente a um professor.

A tabela *Production* armazena as produções acadêmicas em si, o tipo de produção e demais dados descritivos como DOI e data da produção. Nela também contém o tipo *ProductionType* que se trata de um *Enum* que contém todos os tipos de produção contidos no Lattes.

A tabela *ProductionAuthor* registra os pesquisadores autores e coautores vinculados a cada produção. Para melhorar a categorização e facilitar a busca, a tabela *Keyword* permite associar palavras-chave a cada pesquisador cadastrado. Por fim, a tabela *ProfessorData* armazena informações adicionais dos docentes em formato *JavaScript Object Notation* (JSON) e XML.

Uma decisão arquitetural importante foi o uso de Identificador Único Universal (UUID) como chave primária em todas as tabelas. Essa abordagem oferece identificadores únicos de forma segura e descentralizada, facilitando a integração com sistemas externos, a importação de dados em paralelo e aumentando a privacidade ao não expor a ordem sequencial dos registros, evitando colisões comuns com identificadores numéricos sequenciais. (LEACH; MEALLING; SALZ, 2005)

Adicionalmente, para garantir um controle rigoroso e a capacidade de auditoria, foram padronizados os campos *createdAt*, *updatedAt* e *deletedAt* em todas as entidades. O campo *createdAt* registra o momento da inserção, *updatedAt* marca a última modificação, e *deletedAt* implementa a exclusão lógica *(soft delete)*, preservando o histórico completo dos dados e facilitando processos de recuperação e análise temporal, conforme recomendam autores como Date (2020) e Heuser (2009).

Capítulo 4

Implementação

Este capítulo apresenta os detalhes da implementação do sistema proposto para a visualização de análise dos dados do Lattes, desenvolvido com base nos requisitos definidos anteriormente. São abordados aspectos técnicos que sustentam o funcionamento da aplicação, destacando as tecnologias adotadas, a infraestrutura utilizada e a organização lógica dos componentes que integram o sistema.

Além disso, são descritos a arquitetura da solução, a modelagem dos dados e os processos de ETL aplicados aos currículos XML dos pesquisadores. Esses elementos são fundamentais para garantir a integração, consistência e confiabilidade das informações acadêmicas, possibilitando a construção de uma plataforma robusta e eficiente voltada ao monitoramento e análise da produção científica institucional.

4.1 Tecnologias Utilizadas

A aplicação proposta neste trabalho foi construída com base em um conjunto moderno de tecnologias, que oferecem suporte completo ao desenvolvimento *web full-stack*, desde a construção da interface do usuário até a camada de persistência e análise de dados. A escolha dessas ferramentas levou em consideração critérios como modularidade, escalabilidade, desempenho, manutenibilidade e aderência às boas práticas de desenvolvimento. A seguir, descrevem-se individualmente as tecnologias

utilizadas em cada camada da aplicação.

4.1.1 Angular

O *Angular*[1] é um *framework* para desenvolvimento *frontend* mantido pelo *Google*, projetado para construir aplicações web modernas e altamente dinâmicas. Ele adota o modelo de *Single-page application* (SPA), no qual uma única página *Hypertext Markup Language* (HTML) é carregada inicialmente, e seu conteúdo é dinamicamente atualizado sem recarregamentos completos do navegador. Esse modelo melhora significativamente a performance e a experiência do usuário, pois reduz o tempo de resposta e a sobrecarga de navegação (MDN Web Docs, 2025).

Além disso, o *Angular* oferece uma arquitetura baseada em componentes reutilizáveis, com gerenciamento eficiente de rotas, serviços, formulários, injeção de dependência e testes integrados. Sua estrutura modular se adequa perfeitamente à arquitetura de divisão por funcionalidades adotada no projeto, permitindo que cada *feature* possua seus próprios arquivos organizados e encapsulados.

4.1.2 *NestJS*

O *NestJS*[2] é um *framework* progressivo para desenvolvimento de aplicações do lado servidor (*backend*) baseado em *Node.js* e totalmente construído com *TypeScript*. Inspirado na arquitetura do *Angular*, o *NestJS* adota um modelo modular, com suporte a injeção de dependências, decoradores, e princípios orientados a objetos e programação funcional.

Sua estrutura baseada em módulos facilita a organização do código por domínios de negócio, permitindo um desenvolvimento escalável e sustentável. A utilização de controladores, serviços e repositórios de forma isolada dentro dos módulos garante coesão e favorece a reutilização de código.

[1] <https://angular.dev>
[2] <https://nestjs.com>

4.1.3 PrimeNG

O *PrimeNG*[3] é uma biblioteca de componentes de Interface gráfica (UI) desenvolvida para *Angular*. Ela fornece um conjunto extenso de elementos visuais prontos para uso, como tabelas dinâmicas, gráficos, calendários, menus, caixas de diálogo e outros componentes altamente personalizáveis e responsivos.

A adoção do *PrimeNG* no projeto teve como objetivo acelerar o desenvolvimento da interface com uma aparência moderna, responsiva e consistente, além de facilitar a padronização visual da aplicação. Os componentes do *PrimeNG* também oferecem suporte nativo à acessibilidade, internacionalização e integração com estilos personalizados.

4.1.4 xml2js

A biblioteca *xml2js*[4] foi utilizada para conversão de dados no formato XML para JSON. Essa transformação é essencial na camada de ETL da aplicação, onde dados estruturados oriundos de fontes externas são extraídos, processados e carregados no banco de dados. Como o *backend* trabalha preferencialmente com dados em formato JSON, a conversão torna a integração mais fluida e compatível com os serviços desenvolvidos.

Diferentemente de outras bibliotecas, *xml2js* foca na conversão flexível de XML para objetos JSON, mantendo a estrutura hierárquica dos dados e permitindo personalizações através de opções de configuração. A biblioteca suporta a preservação de atributos, normalização de texto e tratamento assíncrono, tornando-a uma solução eficiente para processamento de grandes volumes de dados estruturados.

Ao converter os dados para JSON, a aplicação facilita a manipulação e integração com os serviços *backend* desenvolvidos em *NestJS*, uma vez que o formato JSON é nativamente suportado por APIs RESTful. Essa transformação de dados melhora a eficiência do processo de ingestão e garante maior compatibilidade com a arquitetura

[3] <https://primeng.org>
[4] <https://www.npmjs.com/package/xml2js>

da aplicação.

4.1.5 PostgreSQL

O *PostgreSQL*[5] foi adotado como SGBD da aplicação. É uma tecnologia de código aberto, amplamente reconhecida pela sua robustez, escalabilidade, desempenho e suporte a transações ACID (Atomicidade, Consistência, Isolamento e Durabilidade).

Entre seus principais recursos, destacam-se o suporte a tipos de dados complexos, extensibilidade por meio de funções personalizadas, compatibilidade com dados no formato JSON, mecanismos de segurança avançados e alta confiabilidade. O *PostgreSQL* atende plenamente às necessidades da aplicação quanto à integridade dos dados e complexidade das consultas.

4.1.6 Prisma ORM

O *Prisma ORM* é uma ferramenta moderna de Mapeamento Objeto-Relacional (ORM) que facilita a comunicação entre a aplicação e bancos de dados relacionais, como o *PostgreSQL*. ORMs permitem aos desenvolvedores interagir com o banco de dados por meio de objetos e linguagens orientadas a objetos, abstraindo comandos SQL e tornando operações como leitura, inserção, atualização e exclusão mais seguras, tipadas e integradas ao código da aplicação (Prisma, 2025).

Entre os diferenciais do *Prisma* destacam-se o alto desempenho, a clareza na definição dos modelos de dados e a geração automática de tipos em *TypeScript*, que asseguram segurança em tempo de compilação. Sua estrutura baseia-se em um arquivo de esquema (*schema.prisma*), onde são definidos os modelos, relacionamentos e configurações do banco. A partir disso, o *Prisma* gera um cliente personalizado para acesso aos dados, com validação de tipos e recursos de autocompletar no editor.

O *Prisma* também facilita a criação e execução de migrações de banco de dados, permitindo versionar alterações na estrutura das tabelas de forma rastreável e segura, contribuindo para a evolução controlada do modelo de dados ao longo do ciclo de

[5] <https://www.postgresql.org>

vida da aplicação (Prisma, 2025).

Sua utilização neste projeto trouxe ganhos diretos em produtividade e qualidade no desenvolvimento *backend*, ao simplificar o processo de persistência, reduzir erros comuns na manipulação de SQL e garantir maior integridade entre código e banco de dados.

4.1.7 Supabase

O *Supabase*[6] é uma plataforma *open-source* de *Backend as a Service* (BaaS) que oferece, entre outros recursos, autenticação de usuários, banco de dados *PostgreSQL* gerenciado, armazenamento de arquivos, APIs RESTful e comunicação em tempo real. Ele foi utilizado na aplicação como suporte adicional ao *backend*, integrando funcionalidades específicas e acelerando o desenvolvimento de determinadas partes do sistema.

A integração com o *Supabase* permite à aplicação acessar funcionalidades prontas sem necessidade de implementação personalizada, otimizando tempo e recursos. Além disso, sua compatibilidade com *PostgreSQL* e sua interface intuitiva tornam essa tecnologia uma opção eficiente para ampliar a infraestrutura do sistema de forma flexível.

4.1.8 Metabase

O *Metabase*[7] é uma plataforma de BI que permite a criação de *dashboards*, gráficos e relatórios de forma simples e visual. Ele se conecta diretamente ao banco de dados da aplicação e possibilita a análise e visualização das informações registradas, mesmo por usuários sem conhecimento técnico aprofundado.

Sua interface intuitiva, aliada a recursos como filtros dinâmicos, agendamentos e exportação de dados, torna o *Metabase* uma ferramenta estratégica para monitoramento, gestão e tomada de decisão com base em dados. No contexto deste trabalho,

[6] <https://supabase.com>
[7] <https://www.metabase.com>

o *Metabase* complementa a arquitetura da aplicação ao oferecer uma camada de visualização analítica sobre os dados processados.

4.1.9 RabbitMQ

O *RabbitMQ*[8] é um *message broker* de código aberto amplamente utilizado, que implementa o protocolo Advanced Message Queuing Protocol (AMQP) (Advanced Message Queuing Protocol), entre outros. Sua principal função é atuar como um intermediário para a troca de mensagens entre diferentes partes de um sistema, permitindo a comunicação assíncrona e o desacoplamento de serviços.

Adotado no projeto para gerenciar filas de mensagens, o *RabbitMQ* facilita a construção de aplicações distribuídas e resilientes. Ele permite que as aplicações enviem mensagens para filas, e que outros serviços consumam essas mensagens no seu próprio ritmo, sem que o remetente precise esperar por uma resposta imediata. Isso é particularmente útil para processamento de tarefas em *background*, balanceamento de carga e na implementação de arquiteturas baseadas em microserviços.

A utilização do *RabbitMQ* no projeto visa otimizar o fluxo de processamento de dados e aumentar a escalabilidade e a tolerância a falhas do sistema, especialmente em operações que podem ser executadas de forma assíncrona, como o processamento de grandes volumes de dados após a conversão de XML para JSON.

4.2 Visão geral da arquitetura e Fluxo de Dados do Sistema

O sistema de visualização e análise de dados do Lattes foi concebido com uma arquitetura multicamadas, visando à separação de responsabilidades e à escalabilidade. Um diagrama da arquitetura geral do sistema visto na Figura 4.1 ilustra os principais componentes e suas interações, mostrando como o *frontend*, *backend*, banco de dados, o processo de ETL e a plataforma de BI se conectam.

O fluxo de operação do sistema é desencadeado quando o próprio pesquisador

[8] <https://www.rabbitmq.com>

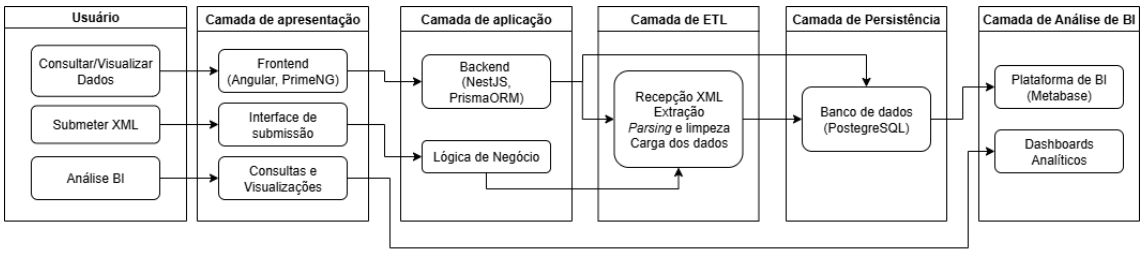


Figura 4.1: Modelagem para o sistema proposto

submete seu currículo Lattes em formato XML através da interface da aplicação diretamente ao *backend*. A partir deste ponto, a camada de ETL entra em ação, sendo responsável por recepcionar este arquivo, processá-lo, realizar as transformações necessárias nos dados e, finalmente, carregá-los de forma estruturada no banco de dados PostgreSQL.

O *backend*, desenvolvido em *NestJS*, é responsável por consumir esses dados estruturados por meio do *Prisma ORM* e por expô-los em uma API RESTful segura e eficiente. Essa API serve como a principal fonte de dados para o restante da aplicação.

O *frontend*, construído com *Angular* e a biblioteca de componentes *PrimeNG*, consome a API para oferecer aos usuários interfaces ricas, reativas e intuitivas. Através dessas telas, os usuários finais podem interagir diretamente com os dados, gerando visualizações dinâmicas e analisando as informações acadêmicas consolidadas de forma eficiente.

Paralelamente a esse fluxo principal, a plataforma de BI *Metabase* conecta-se diretamente ao banco de dados PostgreSQL. Isso permite a criação de *dashboards* analíticos e relatórios personalizados, oferecendo uma camada adicional de exploração de dados para usuários que necessitam de análises mais aprofundadas ou específicas. Funcionalidades adicionais do sistema, como autenticação de usuários e armazenamento de arquivos, são suportadas pela integração com a plataforma *Supabase*. Essa arquitetura integrada e modular visa fornecer uma solução completa e robusta para o monitoramento e análise da produção científica institucional.

Embora o *Supabase* disponibilize um conjunto amplo de serviços — autenticação,

armazenamento de objetos, comunicação em tempo real e funções servidoras —, neste trabalho, ele foi empregado exclusivamente como provedor gerenciado do banco de dados *PostgreSQL*. Optou-se por essa configuração minimalista porque a aplicação já conta com camadas próprias de autenticação, regras de negócio e manipulação de arquivos implementadas no *backend* em *NestJS*; logo, não haveria ganhos proporcionais em duplicar essas responsabilidades na plataforma.

Desse modo, o *Supabase* atua unicamente na hospedagem do banco, oferecendo alta disponibilidade, backups automáticos e balanceamento de conexões sem introduzir dependências adicionais nem exigir refatorações significativas no código existente. Essa estratégia reduz custos e preserva a flexibilidade de migração futura para outro serviço gerenciado de *PostgreSQL*, caso haja alguma necessidade.

4.3 Arquitetura Modular por Funcionalidade

A arquitetura modular por funcionalidade, também conhecida como *feature-based modular architecture*, é uma abordagem organizacional que propõe a estruturação do sistema com base nas funcionalidades ou domínios de negócio da aplicação, em oposição ao modelo tradicional baseado em camadas técnicas. Essa organização visa manter todos os elementos relacionados a uma determinada funcionalidade agrupados de forma coesa, como componentes, serviços, repositórios, rotas e testes, favorecendo a legibilidade, escalabilidade e manutenibilidade do sistema.

No *backend*, essa abordagem é especialmente eficaz quando aplicada com o *framework NestJS*, que possui como base o conceito de módulos. Essa estrutura permite a separação lógica do sistema por domínios, favorecendo a organização do código (NestJS Documentation, 2025).

Cada módulo representa uma unidade funcional autônoma, contendo seus próprios controladores, serviços, repositórios e testes unitários. Essa organização possibilita o encapsulamento de responsabilidades, a redução do acoplamento entre as partes do sistema e a facilitação da reutilização de código (NestJS Documentation, 2025).

No *frontend*, essa mesma arquitetura se mantém coerente com o uso do Angular,

que também adota uma abordagem modular e orientada a componentes. O Angular permite que cada funcionalidade da aplicação seja estruturada em módulos coesos, com seus próprios componentes, serviços e rotas (Angular Documentation, 2025).

Essa padronização da estrutura, tanto no *backend* quanto no *frontend*, proporciona uma coesão arquitetural que contribui significativamente para a manutenção, escalabilidade e entendimento do sistema como um todo.

Adicionalmente, a aplicação implementa uma camada de ETL, responsável pela extração, transformação e carga de dados no banco de dados. Essa camada automatiza o processo de ingestão de dados estruturados, promovendo a atualização contínua e organizada da base de dados da aplicação.

O uso dessa abordagem permite que os dados externos sejam processados, validados e integrados ao modelo interno da aplicação, garantindo qualidade e integridade nas informações armazenadas. A camada de ETL opera como um subsistema independente, mas integrado ao restante da arquitetura, contribuindo para um fluxo de dados eficiente e padronizado.

Essa abordagem arquitetural encontra respaldo em conceitos consolidados da engenharia de *software*, como os princípios da *Clean Architecture*, que enfatizam a separação de responsabilidades, o isolamento de dependências e a organização por domínio (MARTIN, 2017).

Além disso, os fundamentos do *Domain-Driven Design* (DDD), propostos por Evans (2004), reforçam a importância da delimitação clara de contextos funcionais (*bounded contexts*), o que se reflete diretamente na segmentação da aplicação por funcionalidades.

Embora o termo *feature-based modular architecture* não esteja formalizado como um padrão clássico nas literaturas tradicionais, sua aplicação é amplamente reconhecida na prática moderna. Essa abordagem é recomendada por *frameworks* como *Angular* e *NestJS*, além de ser documentada em guias técnicos voltados para aplicações escaláveis.

A documentação oficial do *Angular* incentiva a utilização de *Feature Modules*

como forma de modularizar e escalar aplicações de forma eficiente (Angular Documentation, 2025). O mesmo ocorre com o *NestJS*, que destaca a importância da organização modular para o desenvolvimento sustentável de aplicações robustas (NestJS Documentation, 2025).

Portanto, a adoção da arquitetura modular por funcionalidade neste trabalho justifica-se por sua capacidade de promover uma estrutura clara, flexível e escalável tanto no *frontend* quanto no *backend*. Essa organização favorece não apenas o desenvolvimento e a manutenção, mas também a evolução futura da aplicação.

Além disso, essa abordagem pode facilitar a transição de partes específicas do sistema para microserviços, conforme as necessidades da organização. A integração da camada de ETL complementa essa arquitetura, viabilizando um fluxo contínuo de dados e ampliando a robustez do sistema como um todo.

4.4 Detalhamento do processo de ETL

O processo de ETL é um pilar fundamental da aplicação, assegurando que os dados provenientes dos currículos Lattes sejam processados, higienizados e disponibilizados de forma confiável e estruturada para as subsequentes etapas de análise e visualização. Dada a potencial variabilidade estrutural e o volume considerável dos arquivos XML dos currículos, foi implementado um algoritmo de ETL que utiliza um sistema de filas para gerenciar o processamento de forma assíncrona, controlada e escalável.

Para atender a essa demanda, foi concebido um *pipeline* de processamento de dados, cuja arquitetura, fundamentada em princípios de sistemas distribuídos e comunicação assíncrona (HOHPE; WOOLF, 2003), gerencia o ciclo de vida completo da informação. O escopo do sistema abrange desde a ingestão dos currículos até sua efetiva disponibilização para consumo pela aplicação. Tal abordagem projetual visa priorizar não somente a integridade e a consistência dos dados, mas também a resiliência e a eficiência operacional de toda a infraestrutura, características essenciais em aplicações de dados modernas (KLEPPMANN, 2017).

O fluxo de execução desse *pipeline* é orquestrado por um conjunto de etapas

sequenciais e interdependentes. O percurso tem início na Coleta e Enfileiramento, fase responsável pela recepção e organização inicial dos arquivos. A seguir, o Consumo da Fila e Extração dá início ao processamento ativo, retirando os itens da fila e preparando os dados brutos.

Ademais, o núcleo lógico reside na etapa de Transformação e Validação, onde os dados são convertidos, normalizados e verificados, de modo que fiquem semanticamente compatíveis com o domínio da aplicação. Uma vez aprovados, os dados seguem para a Carga no Banco de Dados, que efetiva sua persistência. Para garantir a confiabilidade de todo o processo, uma camada transversal de Gerenciamento de Erros e Notificação monitora todas as fases, assegurando a integridade e a rastreabilidade do fluxo de dados.

O algoritmo de ETL implementado é composto pelas seguintes etapas principais:

a) Coleta e Enfileiramento: Inicialmente, os arquivos XML dos currículos Lattes, obtidos através de *uploads* na plataforma, são identificados e adicionados a uma fila de processamento. Esta fila atua como um *buffer* intermediário, desacoplando a etapa de recebimento dos arquivos da sua efetiva conversão e processamento. Isso permite que o sistema lide com picos de carga de forma eficiente e processe os arquivos de maneira ordenada e controlada, sem sobrecarregar os recursos computacionais.

b) Consumo da Fila e Extração: Um ou mais processos workers (consumidores da fila), executando em segundo plano, monitoram a fila continuamente. Ao identificar um novo arquivo na fila, um *worker* o retira e inicia a primeira fase do processamento, que é a extração do conteúdo bruto do arquivo XML.

c) Transformação e Validação: Após a extração, a biblioteca *xml2js* é empregada para converter a estrutura hierárquica do XML em um objeto JSON, formato mais facilmente manipulável pelo *backend* da aplicação. Durante esta etapa de transformação, são aplicadas diversas regras de negócio: normalização de dados (e.g., padronização de datas, remoção de caracteres especiais), seleção dos campos de interesse para a análise (como informações sobre publicações, participações em bancas, orientações, projetos de pesquisa, entre outros)

e adequação desses dados ao modelo de dados definido para a aplicação. Validações são realizadas para garantir a integridade, consistência e qualidade das informações antes da persistência.

d) Carga no Banco de Dados: Uma vez que os dados foram transformados e validados, o objeto JSON resultante é mapeado para as entidades correspondentes definidas no esquema do *Prisma ORM*. O *Prisma ORM* é então responsável por persistir esses dados de forma transacional e segura no banco de dados PostgreSQL, realizando as operações de criação ou atualização dos registros dos pesquisadores e suas respectivas produções acadêmicas.

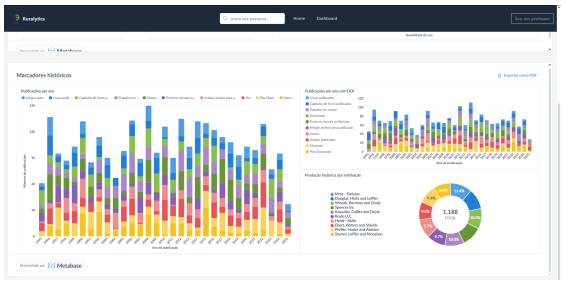
e) Gerenciamento de Erros e Notificação: Para garantir a robustez do pipeline de dados, mecanismos de tratamento de erros são implementados em todas as etapas. Caso ocorram erros durante o processamento de um arquivo (por exemplo, um XML malformado, dados que não passam na validação, ou falha na conexão com o banco), o item problemático é movido para uma fila de "erros"ou um *log* detalhado do incidente é registrado. Mecanismos de notificação podem ser acionados para alertar os administradores do sistema sobre essas falhas, permitindo a análise da causa raiz e o eventual reprocessamento do arquivo, se necessário. Este tratamento de erros assegura que falhas pontuais não interrompam o processamento de toda a carga de dados.

4.5 Navegação e Interfaces do Sistema

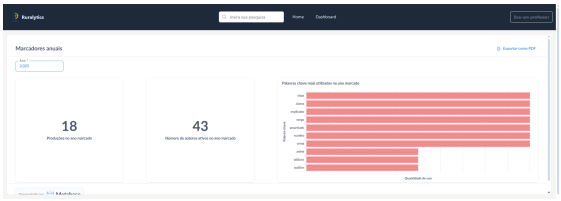
O design da interface do sistema foi fundamentado em princípios de Experiência do Usuário (UX) e Design Centrado no Usuário (DCU), com o objetivo de criar uma ferramenta não apenas poderosa em suas funcionalidades, mas também intuitiva, eficiente e com baixa carga cognitiva para seus utilizadores.

Essa abordagem se manifesta, primeiramente, no painel de controle gerencial (*dashboard*). Ao acessar a aplicação, o usuário visualiza uma interface que aplica a hierarquia da informação para destacar dados estratégicos. Conforme ilustrado na Figura 4.2, o painel exibe análises visuais da evolução temporal dos indicadores

bibliométricos (**Figura 4.2a**) e de sua distribuição anual (**Figura 4.2b**). Essa apresentação dual permite uma compreensão imediata tanto das tendências de longo prazo quanto do desempenho anual, fornecendo um panorama completo da produção científica da instituição.



(a) Evolução temporal dos indicadores bibliométricos



(b) Distribuição anual dos principais indicadores

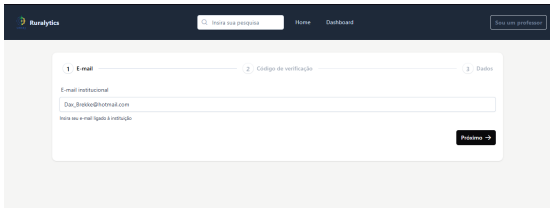
Figura 4.2: *Dashboard* inicial do sistema *Ruralytics*, destacando (a) a tendência ao longo do tempo e (b) a comparação por ano dos indicadores dos currículos Lattes.

Para assegurar que os dados individuais sejam incorporados e atualizados de forma segura, o sistema implementa um fluxo de autenticação em múltiplas etapas. Este processo, ilustrado na Figura 4.3, foi desenhado para validar a identidade do pesquisador antes de conceder acesso às funcionalidades de gerenciamento de dados, como a submissão do currículo Lattes.

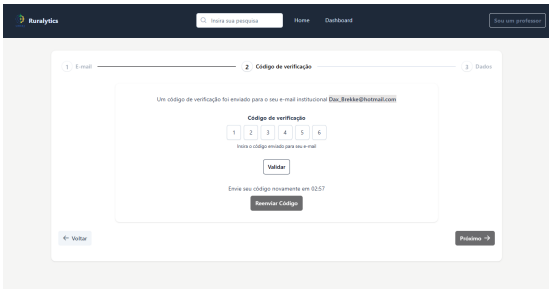
O fluxo se inicia quando o usuário informa seu *e-mail* institucional para solicitar o acesso ao sistema (Figura 4.3a). Em seguida, para verificação, um código de uso único (OTP) é enviado ao *e-mail* informado, devendo ser inserido na tela subsequente (Figura 4.3b). Apenas após a autenticação bem-sucedida, o pesquisador obtém acesso à interface principal de gerenciamento, onde pode realizar o *upload* do arquivo XML de seu currículo (Figura 4.3c), acionando o pipeline de ETL do sistema.

Após o processamento, o perfil individual do pesquisador é apresentado, como visto na Figura 4.4. Esta interface exibe indicadores-chave de desempenho, como a contagem anual de publicações por tipo e outras métricas agregadas.

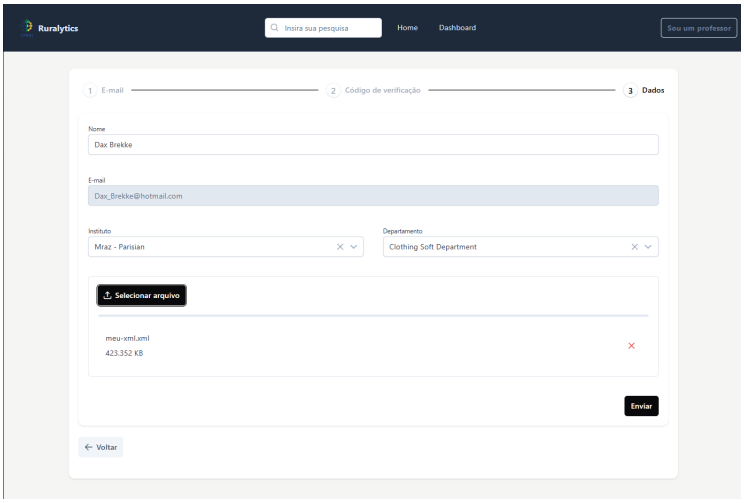
Essa apresentação reflete uma decisão deliberada de arquitetura da informação: em vez de replicar a alta densidade de dados da Plataforma Lattes, optou-se por um design minimalista, focado em clareza e relevância. A curadoria dos dados



(a) Entrada do e-mail para envio do código OTP.



(b) Validação do código OTP recebido.



(c) Formulário para inserir ou atualizar os dados do pesquisador e submeter o arquivo XML do currículo Lattes.

Figura 4.3: Fluxo de autocadastro no sistema *Ruralytics*: (a) solicitação de OTP; (b) confirmação de OTP; (c) submissão de dados e XML.

apresentados permite que o pesquisador avalie sua produção de forma rápida e objetiva, livre do ruído informacional característico de relatórios exaustivos.

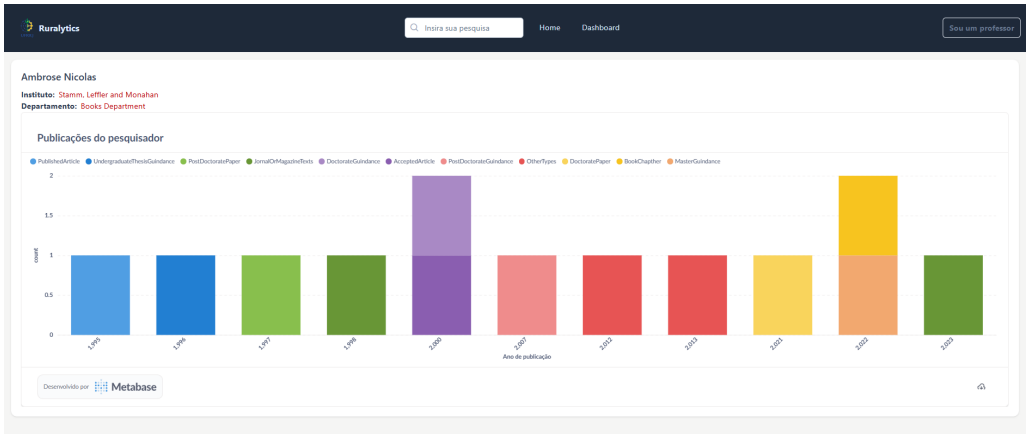


Figura 4.4: Perfil individual do pesquisador, apresentando contagem anual de publicações categorizadas por tipo.

Em suma, a estrutura de navegação do sistema foi concebida para otimizar a usabilidade, equilibrando a visão macroscópica dos *dashboards* com a análise individual focada na clareza, para assim oferecer tanto uma visão panorâmica da produção científica quanto um detalhamento individualizado nos perfis dos pesquisadores. Essa simplicidade intencional na interação e na apresentação dos dados não só melhora a experiência do usuário, mas também fomenta o engajamento e a utilização eficaz da ferramenta por toda a comunidade acadêmica, facilitando assim o acompanhamento transparente da evolução científica, a análise da produção coletiva e individual, e a compreensão ampla do panorama de pesquisa da instituição, estimulando a colaboração e a disseminação do conhecimento.

Capítulo 5

Conclusão

Este capítulo apresenta as considerações finais do trabalho, destacando as contribuições alcançadas com o desenvolvimento do sistema proposto, bem como as principais limitações observadas e sugestões para futuras melhorias e ampliações da solução desenvolvida.

5.1 Considerações finais

O presente trabalho dedicou-se a explorar o desenvolvimento de um sistema de *analytics* para a produção acadêmica. O intuito foi transformar o vasto e complexo volume de dados oriundos da Plataforma Lattes em informações estratégicas para IES. Partindo da identificação dos desafios enfrentados pelas IES na gestão e análise de sua produção científica, o objetivo central foi propor e detalhar a concepção de uma ferramenta tecnológica capaz de subsidiar a tomada de decisão, o planejamento científico e ampliar a visibilidade da pesquisa institucional.

A jornada para alcançar tal objetivo iniciou-se com uma robusta Fundamentação Teórica, onde foram explorados os conceitos essenciais de BI, *Analytics*, e o crucial processo de Extração, Transformação e Carga (ETL). Foram discutidas diversas arquiteturas de BI, desde modelos tradicionais baseados em *Data Warehouses* até abordagens modernas como *Data Lakes* e BI em Tempo Real, e analisadas diferentes

ferramentas de mercado. Este embasamento teórico foi fundamental para orientar as escolhas técnicas e metodológicas adotadas na fase de implementação.

A principal concretização deste trabalho reside na detalhada Proposta e Implementação do Sistema de *Analytics*. Conforme apresentado, a solução foi concebida seguindo uma arquitetura multicamadas. Foram empregadas tecnologias modernas como Angular para o *frontend*, *NestJS* para o *backend*, PostgreSQL como SGBD e *Prisma ORM* para a camada de persistência.

A utilização de ferramentas como *xml2js* para o tratamento de dados Lattes e Metabase para a visualização analítica complementar, além de *RabbitMQ* para gerenciamento de filas no processo de ETL assíncrono, demonstra a aplicação prática dos conceitos estudados. O sistema foi projetado para atender a um conjunto específico de Requisitos Funcionais e Não Funcionais, Regras de Negócio e Casos de Uso.

As principais contribuições deste trabalho podem ser sumarizadas em diferentes dimensões. Do ponto de vista prático, oferece-se o projeto e a descrição de implementação de uma plataforma tecnológica com potencial significativo. Espera-se que ela auxilie as IES na superação dos desafios de gestão da informação científica. Durante o desenvolvimento da solução, foram utilizados como referência sistemas consolidados, como o SOMOS, da UFMG, e o Repositório da Produção da USP, permitindo a análise comparativa de funcionalidades e boas práticas.

A capacidade do sistema de processar currículos Lattes, gerar indicadores relevantes, permitir análises segmentadas e oferecer visualizações dinâmicas representa um avanço considerável. Supera-se, assim, a dependência de processos manuais morosos ou de ferramentas genéricas pouco adaptadas a este contexto.

Teoricamente, este estudo contribui ao aplicar e discutir conceitos de BI e *analytics* no âmbito específico da produção acadêmica brasileira. Detalha-se uma arquitetura e um processo de ETL adaptados às características dos dados da Plataforma Lattes. Para a instituição que vier a adotar tal sistema, ou que servir de base para um estudo de caso, os benefícios potenciais incluem uma gestão mais eficiente e maior

transparência. Adicionalmente, espera-se um melhor embasamento para decisões estratégicas e o fortalecimento de uma cultura orientada por dados.

Por fim, este trabalho buscou demonstrar a viabilidade e a relevância da aplicação de tecnologias de *Business Intelligence* e *Analytics* para a gestão da produção científica originada da Plataforma Lattes. A jornada desde a fundamentação teórica até o detalhamento da implementação revelou a complexidade, mas também o imenso potencial de transformar dados acadêmicos em inteligência estratégica para as instituições.

5.2 Limitações e trabalhos futuros

Apesar dos avanços obtidos, o sistema desenvolvido ainda apresenta algumas limitações que poderão ser abordadas em trabalhos futuros. Atualmente, o processo de obtenção dos dados de produção acadêmica ocorre por meio da submissão manual, realizada pelo próprio docente, que deve exportar o arquivo no formato XML diretamente da Plataforma Lattes e inseri-lo no sistema. Embora essa abordagem seja funcional, ela ainda depende da iniciativa individual de cada docente e pode gerar inconsistências no fluxo de atualização das informações.

Dessa forma, propõe-se, como aprimoramento futuro, a integração do sistema com a API da Plataforma Lattes, o que permitiria a automação desse processo, eliminando a necessidade de *upload* manual dos arquivos XML. Essa integração garantiria maior agilidade, confiabilidade e abrangência na coleta dos dados, além de tornar o sistema mais eficiente e menos suscetível a falhas operacionais.

Adicionalmente, a validação completa da eficácia e do impacto do sistema em uma instituição requereria um estudo de caso longitudinal. Este necessitaria de acompanhamento do uso e coleta de *feedback* dos usuários ao longo de um período extenso, o que, naturalmente, foge ao escopo temporal e aos recursos de um Trabalho de Conclusão de Curso.

Outra frente de aprimoramento consiste em explorar com maior profundidade a seção de projetos de pesquisa do Currículo Lattes. Atualmente, o sistema pode

focar-se na produção bibliográfica, mas os projetos de pesquisa cadastrados contêm dados estratégicos, como agências de fomento, membros das equipes, linhas de pesquisa associadas e períodos de vigência. A extração e análise dessas informações permitiriam mapear as redes de colaboração interna, identificar as principais fontes de financiamento e oferecer uma visão mais completa e dinâmica das atividades de pesquisa em andamento na instituição, indo além da produção já consolidada.

Uma limitação relevante do sistema, em sua concepção atual, refere-se ao comportamento do processo de ETL na atualização dos dados. Atualmente, novas cargas de currículos Lattes resultam na substituição completa dos registros anteriores do pesquisador, em detrimento de uma abordagem de atualização incremental. Tal metodologia, se por um lado simplifica o ciclo de atualização, por outro não preserva o histórico de alterações da produção acadêmica e pode apresentar menor eficiência em cenários de atualizações frequentes ou de grandes volumes de dados, ao exigir o reprocessamento integral.

No desenvolvimento futuro do sistema, prevê-se a ativação do módulo de autenticação do *Supabase*, o que simplificará o processo de login e de controle de acesso sem demandar infraestrutura própria. O espaço de arquivos da plataforma servirá para armazenar currículos, imagens de gráficos e conjuntos de dados, todos protegidos pelas mesmas permissões definidas no mecanismo de login.

Com usuários e arquivos já configurados, os painéis em Angular passarão a atualizar-se automaticamente por meio de notificações via *WebSockets*, eliminando recarregamentos manuais e consultas periódicas. As rotinas de ingestão e transformação de dados, hoje executadas em servidores dedicados, serão realocadas para funções *serverless* agendadas que sobem apenas quando necessário, reduzindo a infraestrutura ociosa e otimizando custos. Em etapa posterior, prevê-se a adoção de busca semântica nos resumos de publicações para recomendar colaborações entre pesquisadores de áreas afins e oferecer respostas mais contextualizadas dentro da plataforma.

Vislumbra-se também, como uma linha de evolução tecnológica, a integração de recursos baseados em Inteligência Artificial (IA), com o objetivo de aprimorar ainda

mais a gestão e análise dos dados. Tais recursos poderão ser empregados para realizar buscas inteligentes no banco de dados, análises preditivas de produção acadêmica, detecção de padrões de produtividade e sugestões automatizadas de indicadores estratégicos, elevando o nível de inteligência e autonomia da plataforma.

Expandindo a aplicação de IA para além da análise de dados, propõe-se a criação de um sistema proativo de recomendação de parcerias. Dessa forma, a plataforma poderia identificar sinergias e complementaridades entre as competências dos pesquisadores, mesmo que atuem em departamentos distintos. Tal funcionalidade fomentaria a criação de grupos de pesquisa interdisciplinares, fortalecendo a capacidade da instituição de abordar problemas complexos e aumentando as chances de sucesso em editais de fomento que valorizam a colaboração.

Essas propostas de melhorias visam não apenas otimizar a performance e a eficiência da plataforma, mas também ampliar seu alcance e impacto, fortalecendo sua contribuição para a modernização dos processos de gestão da produção científica nas instituições de ensino superior brasileiras.

Acredita-se que a solução proposta, mesmo com suas limitações inerentes a um projeto desta natureza, representa um passo significativo. Configura-se como uma contribuição valiosa para que as Instituições de Ensino Superior brasileiras possam conhecer melhor a si mesmas, otimizar seus processos internos e ampliar o impacto de sua produção científica na sociedade.

Referências

Angular Documentation. *Feature Modules - Angular Documentation*. 2025. Acesso em: 09 mar. 2025. Disponível em: <https://angular.io/guide/feature-modules>.

ARMBRUST, M. et al. Lakehouse: A new generation of open platforms that unify data warehousing and advanced analytics. *Proceedings of the 11th Annual Conference on Innovative Data Systems Research (CIDR '20)*, 2020.

CAMPOS, M. S.; SILVA, J. A.; COSTA, L. B. Gestão da produção científica por meio de business intelligence: uma abordagem aplicada às universidades públicas. *Revista Gestão Universitária*, v. 35, n. 2, p. 45–60, 2019.

CHAUDHURI, S.; DAYAL, U.; NARASAYYA, V. An overview of business intelligence technology. *Communications of the ACM*, v. 54, n. 8, p. 88–98, 2011.

DATE, C. J. *Introdução a Sistemas de Bancos de Dados*. 8. ed. Rio de Janeiro: Editora Campus, 2020.

DAVENPORT, T. H. *Big Data at Work: Dispelling the Myths, Uncovering the Opportunities*. Harvard Business Review Press, 2014. (G - Reference,Information and Interdisciplinary Subjects Series). ISBN 9781422168165. Disponível em: <https://books.google.com.br/books?id=apjBAgAAQBAJ>.

DAVENPORT, T. H.; HARRIS, J. G. *Competing on Analytics: The New Science of Winning*. [S.l.]: Harvard Business Press, 2007.

DEDIĆ, N.; STANIER, C. Towards differentiating business intelligence, big data, data analytics and knowledge discovery. In: SPRINGER INTERNATIONAL PUBLISHING. *fattori e tendenze dei sistemi informativi gestionali*. [S.l.], 2017. (Lecture Notes in Business Information Processing, v. 285), p. 109–122.

ETZKOWITZ, H.; LEYDESDORFF, L. The dynamics of innovation: from National Systems and "Mode 2" to a Triple Helix of university–industry–government relations. *Research Policy*, v. 29, n. 2, p. 109–123, 2000.

EVANS, E. *Domain-driven Design: Tackling Complexity in the Heart of Software*. Addison-Wesley, 2004. ISBN 9780321125217. Disponível em: <https://books.google.com.br/books?id=xColAAPGubgC>.

FEW, S. *Information Dashboard Design: The Effective Visual Communication of Data*. [S.l.]: O'Reilly Media, Inc., 2006.

GERAIS, U. F. de M. *Sistema SOMOS - Sistema de Organização e Monitoramento da Produção Científica*. 2025. <https://somos.ufmg.br/>. Acesso em: 09 mar. 2025.

GLäNZEL, W.; SCHOEPFLIN, U. A bibliometric study of ageing and reception processes of scientific literature. *Journal of Information Science*, v. 20, n. 1, p. 37–53, 1994.

HEUSER, C. A. *Projeto de Banco de Dados*. 6. ed. Porto Alegre: Bookman, 2009.

HOHPE, G.; WOOLF, B. *Enterprise Integration Patterns: Designing, Building, and Deploying Messaging Solutions*. [S.l.]: Addison-Wesley Professional, 2003. (Addison-Wesley Signature Series). ISBN 978-0321200686.

INMON, W. H. *Building the Data Warehouse*. 4rd. ed. USA: John Wiley & Sons, Inc., 2005. ISBN 9780471774235. Disponível em: <https://books.google.com.br/books?id=QFKTmh5IFS4C>.

JúNIOR, O. G. F. et al. Uma experiência com business intelligence para apoiar a gestão acadêmica em uma universidade federal brasileira. *Revista Ibérica de Sistemas e Tecnologias de Informação*, n. 46, p. 5–20, 2022. Acesso em: 09 mar. 2025. Disponível em: <https://scielo.pt/pdf/rist/n46/1646-9895-rist-46-5.pdf>.

KIMBALL, R.; ROSS, M. *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*. Wiley, 2013. ISBN 9781118732281. Disponível em: <https://books.google.com.br/books?id=4rFXzk8wAB8C>.

KLEPPMANN, M. *Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems*. Sebastopol, CA: O'Reilly Media, 2017. ISBN 978-1449373320.

KREPS, J.; NARKHEDE, N.; RAO, J. Kafka: A distributed messaging system for log processing. In: *Proceedings of the 6th International Workshop on Networking Meets Databases (NetDB)*. [S.l.: s.n.], 2011. (NetDB '11).

LEACH, P. J.; MEALLING, M.; SALZ, R. *A Universally Unique Identifier (UUID) URN Namespace*. 2005. RFC 4122, IETF. Disponível em: <https://datatracker.ietf.org/doc/html/rfc4122>.

LEITE, F. *Como gerenciar e ampliar a visibilidade da informação científica brasileira: repositórios institucionais de acesso aberto*. Ibict, 2009. ISBN 9788570130679. Disponível em: <https://books.google.com.br/books?id=Q6CyB8PhRH4C>.

LOSHIN, D. *Business Intelligence: The Savvy Manager's Guide*. Elsevier Science, 2012. (Savvy manager's guides). ISBN 9780123858894. Disponível em: <https://books.google.com.br/books?id=L7SLNlS1ao8C>.

MARTIN, R. *Clean Architecture: A Craftsman's Guide to Software Structure and Design*. Pearson Education, 2017. (Robert C. Martin Series). ISBN 9780134494326. Disponível em: <https://books.google.com.br/books?id=uGE1DwAAQBAJ>.

MDN Web Docs. *Single Page Application (SPA)*. 2025. Acesso em: 08 mar. 2025. Disponível em: <https://developer.mozilla.org/en-US/docs/Glossary/SPA>.

MENA-CHALCO, J. P.; JR, R. M. C. scriptLattes: An open-source knowledge extraction system from the Lattes platform. *Journal of the Brazilian Computer Society*, v. 15, n. 4, p. 31–39, 2009.

MOSS, L. T.; ATRE, S. *Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications*. Addison-Wesley, 2003. (Addison-Wesley Information Technology Series). ISBN 9780321630117. Disponível em: <https://books.google.com.br/books?id=ZV8jeV4a9_AC>.

MUGNAINI, R.; PACKER, A. L.; MENEGHINI, R. A produção científica brasileira e a plataforma Lattes. *Interciencia*, v. 29, n. 8, p. 410–415, 2004.

NestJS Documentation. *Modules - NestJS Documentation*. 2025. Acesso em: 09 mar. 2025. Disponível em: <https://docs.nestjs.com/modules>.

PAULO, U. de S. *Repositório Institucional da USP - Produção Intelectual*. 2025. <https://repositorio.usp.br>. Acesso em: 09 mar. 2025.

Prisma. *What is Prisma ORM*. 2025. Acesso em: 08 mar. 2025. Disponível em: <https://www.prisma.io/docs/concepts/overview/what-is-prisma>.

SAWADOGO, P. N.; DARMONT, J. On data lake architectures and metadata management. *Journal of Intelligent Information Systems*, Springer, v. 56, n. 1, p. 97–120, 2021. Disponível em: <https://doi.org/10.1007/s10844-020-00608-7>.

SHATTOCK, M. *Managing successful universities*. Maidenhead: Open University Press, 2003.

SIEMENS, G.; LONG, P. Penetrating the Fog: Analytics in Learning and Education. *EDUCAUSE Review*, v. 46, n. 5, p. 30–40, 2011.

SILVA, G. d. P. F.; BRANDãO, S. V. Indicadores bibliométricos e de gestão da produção científica do cav/ufpe. *Anais da Academia Pernambucana de Ciência Agronômica*, Academia Pernambucana de Ciência Agronômica, Recife, v. 8 e 9, p. 238–249, 2011. Disponível em: Anais da APCAGR.

SILVA, M. R. d.; HAYASHI, C. R. M.; HAYASHI, M. C. P. I. Análise bibliométrica e cientométrica: desafios para especialistas que atuam no campo. *InCID : Revista de Ciência da Informação e Documentação*, 2011.

TECNOLÓGICO, C. C. N. de Desenvolvimento Científico e. *Plataforma Lattes*. Brasília, DF, 2024. Acesso em: 01 mai. 2025. Disponível em: <http://lattes.cnpq.br/>.

TURBAN, E.; SHARDA, R.; DELEN, D. *Business Intelligence and Analytics: Systems for Decision Support*. 10th. ed. [S.l.]: Pearson Education Limited, 2014.

ZAHARIA, M. et al. Discretized streams: Fault-tolerant streaming computation at scale. In: *Proceedings of the 24th ACM Symposium on Operating Systems Principles (SOSP)*. [S.l.: s.n.], 2013. (SOSP '13), p. 423–438.