CENTRO UNIVERSITÁRIO LUTERANO DE PALMAS
CURSO DE SISTEMAS DE INFORMAÇÃO

BRUNO MORAIS BEZERRA

DESENVOLVIMENTO DE UM DATA MART PARA ORGANIZAR E ANALISAR OS MICRODADOS DO ENADE

PALMAS – TO
2024

Bruno Morais Bezerra

DESENVOLVIMENTO DE UM DATA MART PARA ORGANIZAR E ANALISAR OS MICRODADOS DO ENADE

Projeto Tecnológico II elaborado e apresentado como requisito parcial para obtenção do título de bacharel em Sistemas de Informação pelo Centro Universitário Luterano de Palmas (CEULP/ULBRA).

Orientador: Prof. Esp. Douglas Aquino Moreno

Palmas – TO

2024

Bruno Morais Bezerra

DESENVOLVIMENTO DE UM DATA MART PARA ORGANIZAR E ANALISAR OS MICRODADOS DO ENADE

Projeto Tecnológico II elaborado e apresentado como requisito parcial para obtenção do título de bacharel em Sistemas de Informação pelo Centro Universitário Luterano de Palmas (CEULP/ULBRA).

Orientador: Prof. Esp. Douglas Aquino Moreno

Aprovado em: 05/07/2024

BANCA EXAMINADORA

Prof. Esp. Douglas Aquino Moreno

Orientador

Centro Universitário Luterano de Palmas – CEULP

Prof.ª Me. Madianita Bogo Marioti

Centro Universitário Luterano de Palmas – CEULP

Prof.ª Esp. Fernanda Pereira Gomes

Centro Universitário Luterano de Palmas – CEULP

Palmas – TO

2024

AGRADECIMENTOS

À minha amada Camila, minha companheira que tanto me incentiva; ao Ravi,meu filho amado; e aos meus pais pelo amor e apoio. Esta jornada não seria possível sem vocês, me incentivaram todos os dias para persistir nos momentos mais desafiadores. Com todo meu amor e gratidão.

RESUMO

MORAIS, Bruno Bezerra. **Desenvolvimento de um Data Mart para Organizar e Analisar os Microdados do ENADE**. 2024. 37 f. Trabalho de Conclusão de Curso (Graduação) – Curso de Sistemas de Informação, Centro Universitário Luterano de Palmas, Palmas/TO, 2024.

O Exame Nacional de Desempenho dos Estudantes (ENADE) avalia o desempenho dos acadêmicos que cursam o ensino superior por meio de uma prova contendo questões específicas da área e conhecimentos gerais. O ENADE tem como intuito final verificar o desempenho dos estudantes e outras características referentes às instituições. A cada aplicação do exame gera-se uma quantidade de dados enorme, os dados gerados são disponibilizados em formato de arquivo de texto (extensão .txt) chamados de microdados, estes arquivos são compostos por milhares de linhas e colunas tornando inviável a análise por meios convencionais. Dito isso, o presente trabalho teve como objetivo geral apresentar os estudos realizados nos formatos e padrões dos microdados do ENADE, de modo a possibilitar modelar e implementar o *Data Mart* com base nestes microdados.

Palavras-chave: Data Mart, Data Warehouse, ENADE, ETL, Microdados.

ABSTRACT

MORAIS, Bruno Bezerra. **Development of a Data Mart for Organizing and Analyzing ENADE Microdata.** 2024. 37 f. Trabalho de Conclusão de Curso (Graduação) – Curso de Sistemas de Informação, Centro Universitário Luterano de Palmas, Palmas/TO, 2024.

The National Student Performance Exam (ENADE) evaluates the performance of undergraduate students through a test containing area-specific questions and general knowledge. ENADE aims to verify student performance and other characteristics related to institutions. Each application of the exam generates a vast amount of data, made available in text file format (.txt) called microdata. These files consist of thousands of rows and columns, making analysis by conventional means unfeasible. Therefore, the present work aimed to present the studies conducted on the formats and standards of ENADE microdata, in order to enable the modeling and implementation of a Data Mart based on this microdata.

Keywords: Data Mart, Data Warehouse, ENADE, ETL, Microdata.

LISTA DE ILUSTRAÇÕES

LISTA DE ABREVIATURAS E SIGLAS

CEULP	Centro Universitário Luterano de Palmas
DM	Data Mart
DW	Data Warehouse
ENADE	Exame Nacional de Desempenho de Estudantes
IBGE	Instituto Brasileiro de Geografia e Estatística
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
SINAES	Sistema Nacional de Avaliação da Educação Superior

SUMÁRIO

1 INTRODUÇÃO

O Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) é o encarregado de avaliar a educação no Brasil, desde o ensino básico ao ensino superior (INEP, 2021). As avaliações são realizadas periodicamente e, para avaliar o rendimento dos formandos do ensino superior, é aplicado o Exame Nacional de Desempenho dos Estudantes (ENADE), que avalia o aprendizado dos conteúdos programáticos previstos nas diretrizes curriculares dos cursos de graduação (INEP, 2021).

O ENADE busca avaliar também o desenvolvimento de competências e habilidades necessárias ao aprofundamento da formação geral e profissional, e o nível de atualização dos estudantes com relação à realidade brasileira e mundial (INEP, 2021). Mesmo o exame sendo realizado todos os anos a submissão de cada curso ocorre trienalmente, levando em consideração as áreas de conhecimento e eixos tecnológicos, por meio de uma amostra selecionada de estudantes do último ano dos cursos.

A cada aplicação do exame gera-se uma quantidade enorme de dados, e estes são disponibilizados para a população no site do INEP[1] no formato de arquivo de texto (extensão .txt). Esses arquivos de texto constituem-se de microdados que contém os resultados do exame e são insumos fundamentais para o cálculo dos indicadores de qualidade da educação superior.

Os dados abertos do ENADE, disponíveis no site do INEP, contêm informações sobre a qualidade da educação superior brasileira. Esses dados, se analisados, possibilitarão às Instituições de Ensino Superior (IES) compreenderem o desempenho de seus estudantes e da instituição, contribuindo para a tomada de decisões mais informadas e embasadas. A análise desses dados também pode ser utilizada para identificar áreas de melhoria e oportunidades de aprimoramento na gestão da IES.

Entender esses aspectos torna-se uma tarefa importante para as IES, pois a baixa nas notas em avaliações realizadas pelo MEC, principalmente no ENADE, pode afetar a imagem da instituição tendo em vista que as notas das instituições são amplamente divulgadas pelo governo e pelas próprias IES. Normalmente, quando uma instituição de ensino superior tem um bom desempenho no exame usa isso como divulgação positiva para conquistar mais estudantes.

Os índices divulgados pelo INEP oferecem uma visão geral do desempenho das IES, mas nem sempre são suficientes para apoiar os dirigentes nas decisões estratégicas. A página

[1] http://portal.inep.gov.br

do INEP não permite a realização de buscas mais complexas e detalhadas. Além disso, a forma como os microdados são disponibilizados em arquivos de texto exige o uso de ferramentas e técnicas computacionais para serem analisados de maneira detalhada. Isso mostra a importância do desenvolvimento de um *Data Mart* que possa organizar esses dados de forma mais acessível e utilizável para as IES (Araújo, 2019).

Baseado nesse contexto, neste trabalho foi desenvolvido um *Data Mart* para organizar e permitir analisar os microdados do ENADE. Segundo Inmon (1996), um *Data Mart* consiste em estruturas de dados que contêm informações armazenadas por áreas ou assuntos específicos e que correspondem ao interesse e/ou necessidade de determinado departamento de uma organização. De acordo com Alvares, Campos e Gomes (2015), o entendimento das informações do exame e de seus relacionamentos pode identificar aspectos relevantes para o processo de tomada de decisão das Instituições de Ensino Superior (IES).

Para criar um *Data Mart* que possa organizar os microdados do ENADE, foi necessário modelar e implementar uma estrutura de dados de acordo com os formatos e padrões utilizados nos microdados do ENADE. Com esse *Data Mart* será possível em outro momento analisar os microdados e ter uma visão detalhada dos resultados obtidos pelos alunos e pelas instituições.

Nesse sentido, o *Data Mart* se apresenta como uma ferramenta extremamente importante para a realização de análises complexas de dados em empresas e organizações. Com sua modelagem específica e adaptada às necessidades dos usuários em um contexto determinado, o *Data Mart* permite a obtenção de informações a partir da exploração dos dados. Além disso, como apresentado por Kimball e Caserta (2004) o processo ETL é fundamental para garantir a qualidade dos dados, contribui para a confiabilidade e a precisão das análises realizadas. Por isso, o estudo do *Data Mart* se faz relevante para a área da educação superior, apresentando-se como uma solução útil para o gerenciamento de informações estratégicas IES.

2 REFERENCIAL TEÓRICO

Para fundamentar a construção deste trabalho, esta seção apresenta a teoria relevante. A subseção 2.1 explana os conceitos essenciais sobre o ENADE, seguido pela subseção 2.2 que trata dos conceitos relativos aos Microdados. A subseção 2.3 apresenta conceito sobre *Data Mart*, a subseção 2.4 discute o processo ETL, enquanto a 2.5 abrangem os Trabalhos Relacionados.

2.1 ENADE

O Sistema Nacional de Avaliação da Educação Superior (SINAES) é formado por três componentes principais: a avaliação das instituições, dos cursos e do desempenho dos estudantes (INEP, 2021). Um dos instrumentos de avaliação utilizados pelo SINAES que possui o maior peso é o ENADE. Este exame foi criado pela Lei nº. 10.861 de 14 de abril de 2004 com o objetivo de avaliar o desempenho dos estudantes concluintes do ensino superior em relação aos conteúdos previstos nas diretrizes curriculares do curso de graduação, aos conhecimentos gerais e habilidades adquiridas (INEP, 2021).

O ENADE por sua vez é composto por 4 (Quatro) instrumentos básicos de avaliação (INEP, 2021; INEP, 2023):

1. **Prova com questões objetivas e discursivas**: composta por Formação Geral e Componente Específico. A Formação Geral, comum a todos os cursos, conta com 10 questões baseadas nos princípios dos Direitos Humanos. O Componente Específico, específico de cada área, possui 30 questões que abordam situações-problema e estudos de caso;
2. **Questionário do Estudante**: destinado a levantar informações que permitam caracterizar o perfil dos estudantes e o contexto de seus processos formativos, relevantes para a compreensão dos resultados dos estudantes no Enade;
3. **Questionário de Percepção de Prova**: destinado a levantar informações que permitam aferir a percepção dos estudantes em relação à prova, auxiliando, também, na compreensão dos resultados dos estudantes no Enade;
4. **Questionário do Coordenador de Curso**: destinado a levantar informações que permitam caracterizar o perfil do coordenador de curso e o contexto dos processos formativos, auxiliando, também, na compreensão dos resultados dos estudantes no Enade.

Determinado pela Portaria Normativa MEC nº 840 de 24 de agosto de 2018, o INEP por meio da Diretoria de Avaliação da Educação Superior (DAES) realiza o ENADE assim como tem o dever de realizar os cálculos e divulgar os resultados. São eles: Conceito ENADE, Indicador de Diferença entre os Desempenhos Observado e Esperado (IDD), Conceito Preliminar de Curso (CPC) e Índice Geral de Cursos Avaliados da Instituição (IGC).

A periodicidade máxima da avaliação é trienal para cada área do conhecimento e eixos tecnológicos, aos alunos que estão concluindo a graduação é obrigatória a realização, condição indispensável para a emissão do histórico escolar (INEP, 2021). O Ministério da Educação define quais áreas de conhecimento serão avaliadas na edição do exame com base nas propostas apresentadas pela Comissão de Avaliação da Educação Superior (CONAES), órgão colegiado de coordenação e supervisão do Sinaes.

De acordo com o § 6º do artigo 5º da Lei 10.861/2004, é de responsabilidade exclusiva da IES inscrever os estudantes que irão realizar o exame. Se a IES não inscrever os alunos habilitados para participar do exame dentro dos prazos estipulados no edital, ela poderá sofrer uma suspensão temporária da abertura de processos seletivos para os cursos que seriam avaliados naquela edição (Brasil, 2004).

De acordo com os dados da edição 2019, a prova é composta por 40 questões, sendo 10 questões da parte de formação geral e 30 da parte de formação específica da área de formação do aluno, ambas contendo questões discursivas e de múltipla escolha que são especificadas da seguinte forma:

- Componente de Formação Geral: 10 (dez) questões, sendo 02 (duas) discursivas e 8 (oito) de múltipla escolha, envolvendo situações-problema e estudos de casos;
- Componente específico: de cada área de avaliação, 30 (trinta) questões, sendo 3 (três) discursivas e 27 (vinte e sete) de múltipla escolha, envolvendo situações-problema e estudos de casos.

A prova de Formação Geral tem a concepção dos seus itens balizada pelos princípios dos Direitos Humanos, e as questões discursivas avaliam aspectos como clareza, coerência, coesão, estratégias argumentativas, utilização de vocabulário adequado e correção gramatical do texto. A Conceito Enade é calculado em uma escala de 0 (zero) a 100 (cem), onde a nota final do estudante no Enade é obtida pela média ponderada na qual o componente da Formação Geral responde por 25% (vinte e cinco por cento) e o componente dos Conhecimentos Específicos, por 75% (setenta e cinco por cento).

2.2 Microdados

Microdados são bancos de dados em que os registros ou casos (isto é, as linhas) representam as unidades de coleta mais desagregadas (CEM, 2021). É bem comum o uso de microdados em censos demográficos que são realizados em média a cada dez anos pelo IBGE, a unidade de coleta é o indivíduo sendo aplicado um questionário a cada pessoa residente no Brasil.

Os microdados de um censo por exemplo representam todas as respostas dos informantes ao questionário aplicado, com cada linha contendo informações específicas de uma unidade de análise e cada coluna trazendo características específicas aplicáveis àquela unidade. Quando esses dados são agregados, é possível obter informações sobre os indivíduos representados nos microdados.

No Brasil, os grandes produtores de microdados são os órgãos governamentais, pois a Lei de Acesso à Informação (LAI) os obriga a disponibilizar as informações em portais de transparência de forma clara e objetiva. O Portal Brasileiro de Dados Abertos[2] é o canal disponibilizado pelo governo para que as pessoas físicas ou jurídicas possam encontrar os microdados de forma centralizada pois se trata de informações públicas, é nessa plataforma que estão os microdados do INEP e do IBGE (Brasil, 2021).

Os microdados do INEP constituem o menor nível de desagregação de dados recolhidos por suas pesquisas estatísticas, avaliações e exames (INEP, 2021). Desagregação de dados é o processo de separar dados agregados em partes menores e mais específicas. É uma técnica utilizada para obter informações mais detalhadas sobre um conjunto de dados, que normalmente são apresentados de forma resumida e agregada.

Figura 1. Amostra dos microdados ENADE 2022

Fonte: INEP (2024)

[2] https://dados.gov.br/

A Figura 1 traz uma amostra dos microdados do ENADE do ano de 2022, retirada da planilha obtida no site do INEP. O arquivo possui 136 colunas e 433.931 linhas de dados, sendo que a primeira linha em todas as colunas contém o nome da variável que pode ser compreendida consultando o dicionário de variáveis ENADE 2019, como apresentado na Figura 2.

Figura 2. Amostra do dicionário de variáveis ENADE 2019

	B	C	D	E	F
1	Dicionário de Variáveis do ENADE 2019				
2	**NOME**	**TIPO**	**TAMANHO**	**DESCRIÇÃO**	**CATEGORIAS**
3	NU_ANO	N	4	Ano de realização do exame	2019
4	**PARTE 1 - INFORMAÇÕES DA INSTITUIÇÃO DE ENSINO SUPERIOR E DO CURSO**				
5	CO_IES	N	5	Código da IES (e-MEC)	Entre 1 e 23410
6	CO_CATEGAD	N	5	Código da categoria administrativa da IES	118 = Pessoa Jurídica de Direito Privado - Com fins lucrativos - Sociedade Civil
7					120 = Pessoa Jurídica de Direito Privado - Sem fins lucrativos - Associação de Utilidade Pública
8					121 = Pessoa Jurídica de Direito Privado - Sem fins lucrativos - Fundação
9					10005 = Privada com fins lucrativos
10					10006 = Pessoa Jurídica de Direito Privado - Com fins lucrativos - Sociedade Mercantil ou Comercial
11					10007 = Pessoa Jurídica de Direito Privado - Sem fins lucrativos - Associação de Utilidade Pública
12					10008 = Privada sem fins lucrativos
13					10009 = Pessoa Jurídica de Direito Privado - Sem fins lucrativos - Sociedade
14					17634 = Fundação Pública de Direito Privado Municipal
15					93 = Pessoa Jurídica de Direito Público - Federal
16					115 = Pessoa Jurídica de Direito Público - Estadual
17					116 = Pessoa Jurídica de Direito Público - Municipal
18					10001 = Pessoa Jurídica de Direito Público - Estadual
19					10002 = Pessoa Jurídica de Direito Público - Federal
20					10003 = Pessoa Jurídica de Direito Público - Municipal
21	CO_ORGACAD	N	5	Código da organização acadêmica da IES	10019 = Centro Federal de Educação Tecnológica
22					10020 = Centro Universitário
23					10022 = Faculdade
24					10026 = Instituto Federal de Educação, Ciência e Tecnologia
25					10028 = Universidade

Fonte: INEP (2024)

Já a Figura 2 corresponde a um dicionário de variáveis do ENADE 2019 onde compila todos os nomes, tipos, tamanhos, descrições e categorias dos instrumentos de avaliação, na edição de 2019 o dicionário era composto por 6 colunas e 819 linhas. Os microdados do ENADE estão organizados em arquivos de texto e podem ser visualizados em forma de planilhas separando os dados por ponto e vírgula sendo que cada coluna representa uma variável, a primeira linha de cada coluna contempla o nome de variável padronizado sem espaço e assento para facilitar o carregamento em ferramentas de análises e estéticas, cada linha representa um estudante, e cada célula um dado do estudante referente aquela variável em questão.

Os microdados do ENADE são uma importante fonte de informação para pesquisas e análises na educação superior, contendo dados das provas, dos gabaritos, das notas e questionários respondidos pelos participantes. A análise desses dados possibilita a identificação de padrões e tendências na performance dos estudantes, avaliação da qualidade dos cursos e elaboração de políticas educacionais para aprimorar o ensino nas instituições de ensino superior.

O uso dos microdados do ENADE é fundamental para este trabalho, pois eles fornecem uma base de dados consistente com informações coletadas ao longo de 20 anos de exames, além de serem acompanhados por uma documentação completa para auxiliar na compreensão dos dados. Com o auxílio de ferramentas e técnicas computacionais, é possível agregar e analisar esses dados, permitindo a realização de consultas analíticas mais precisas e complexas. Por meio da criação de um *Data Mart* a partir desses dados, é possível obter insights valiosos e relevantes para o desenvolvimento deste trabalho.

2.3 Data Mart

A evolução tecnológica tanto na parte de *hardware* quanto de *software* está proporcionando às organizações a possibilidade de criar e gerir um volume cada dia maior de informações transacionais. O tempo passa, o tamanho das bases de dados cresce, e junto a isso têm-se a dificuldade dos gestores em analisar os dados dessas bases para terem apoio na tomada de decisão.

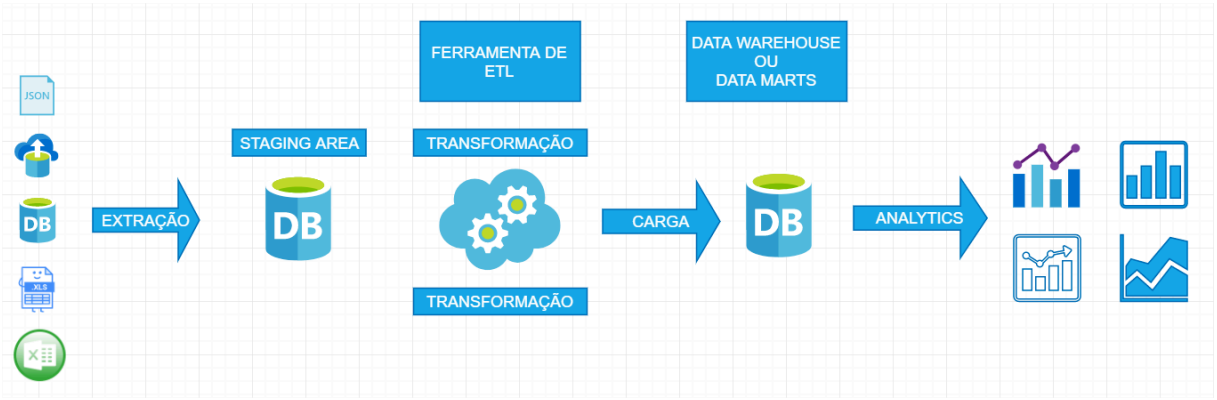
De acordo com Gomes (2010), a carência na obtenção de informações estratégicas baseada em informações transacionais resultou na criação de um novo gênero de Sistema de Informação (SI), designado de *Data Warehouse* (ou armazém de dados, em português), que são bancos de dados históricos oriundos de sistemas transacionais das organizações. Boar (1997) afirma que esses SIs são construídos com o intuito de apoiar o processo de tomada de decisão na organização.

Segundo Barbieri (2001), o DW mantêm uma estrutura de dados dimensionais que é destinado ao apoio da tomada de decisão de gestores, possibilitando o processamento analítico por meio de ferramentas específicas. Devido ao custo elevado na construção de um *Data Warehouse*, ele pode ser desenvolvido em partes menores, denominadas de *Data Mart* ou cubo (Ferreira et al., 2010).

Um *Data Warehouse* é um repositório que armazena informações de uma ou várias fontes. Por exemplo, uma empresa de comércio eletrônico pode utilizar um *Data Warehouse* para integrar e combinar diversas informações dos clientes, tais como endereços de e-mail, dados de caixa registradora, cartões de comentários, dentre outros. De acordo com o artigo "*Data Warehousing and OLAP Technologies*" de 1997, a tecnologia de *Data Warehousing* e *On-line Analytical Processing* (OLAP) é essencial para o suporte à decisão e tem sido um foco importante na indústria de banco de dados (Inmon et al., 1997).

Inmon (1997) no seu artigo oferece uma visão das tecnologias utilizadas para suporte à tomada de decisões, incluindo a arquitetura típica de um *Data Warehouse*, o processo de projeto e operação de um *Data Warehouse*, as tecnologias relevantes para carregar e atualizar dados em um *Data Warehouse*, servidores de armazenamento, ferramentas *front-end* e ferramentas de gerenciamento de armazenamento. Além disso, o artigo identifica questões promissoras de pesquisa relacionadas a problemas que a comunidade de pesquisa em banco de dados tem trabalhado há anos, bem como outros que estão apenas começando a ser abordados.

Figura 3. Visão geral de um *Data Warehouse*.



Fonte: Adaptado de Machado (2004).

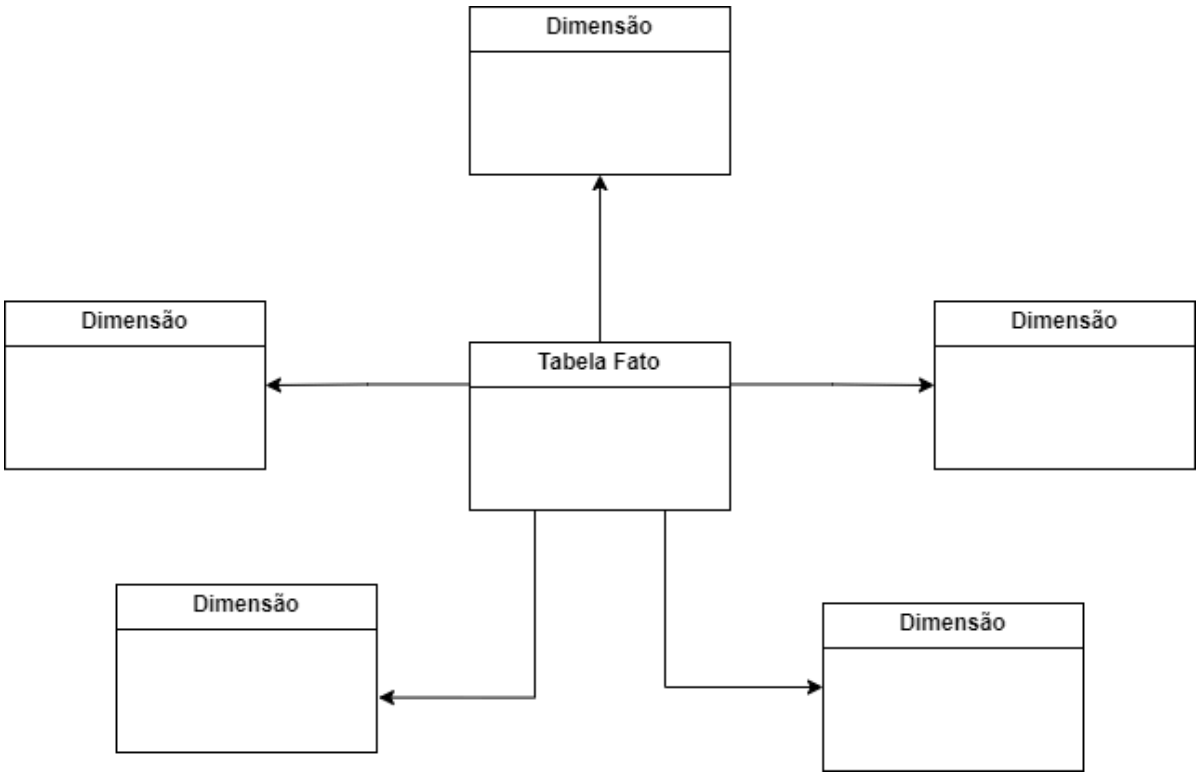
De acordo com Machado (2004), o *Data Warehouse* é um sistema de armazenamento de dados projetado para suportar análises e tomadas de decisão empresariais. Ele é usado para integrar, limpar e consolidar dados de diversas fontes, como bancos de dados operacionais, planilhas e sistemas de terceiros, e armazená-los em um único local. Os dados no *Data Warehouse* são organizados de forma que possam ser facilmente acessados e analisados, geralmente em formato dimensional (como estrela ou floco de neve). Isso permite que os usuários façam consultas complexas e análises de dados de diferentes perspectivas, como por exemplo, por tempo, produto, região ou cliente.

Além disso, o *Data Warehouse* pode ser utilizado para armazenar dados históricos por longos períodos de tempo, permitindo a realização de análises de tendências e previsões. Essas análises podem ajudar as empresas a identificar oportunidades de negócios, entender melhor o comportamento do cliente, otimizar processos internos e tomar decisões mais embasadas e estratégicas.

As modelagens *Data Warehouse* mais comuns são o modelo estrela (*Star Schema)*, modelo floco de neve (*Snowflake Schema)* e Constelação de fatos (*Galaxy Schema)*. O modelo estrela é caracterizado por ter uma tabela central de fatos que contêm as chaves estrangeiras conectando a dimensões e que armazena as medidas numéricas de negócio, cada

linha em uma tabela fato corresponde a um evento, por exemplo, em um contexto de uma farmácia, a tabela fato teria os dados da evento venda, e as tabelas de dimensão que contêm informações contextuais sobre as medidas, no contexto da farmácia teríamos as dimensões medicamento e não poderia se repetir (Figura 4).

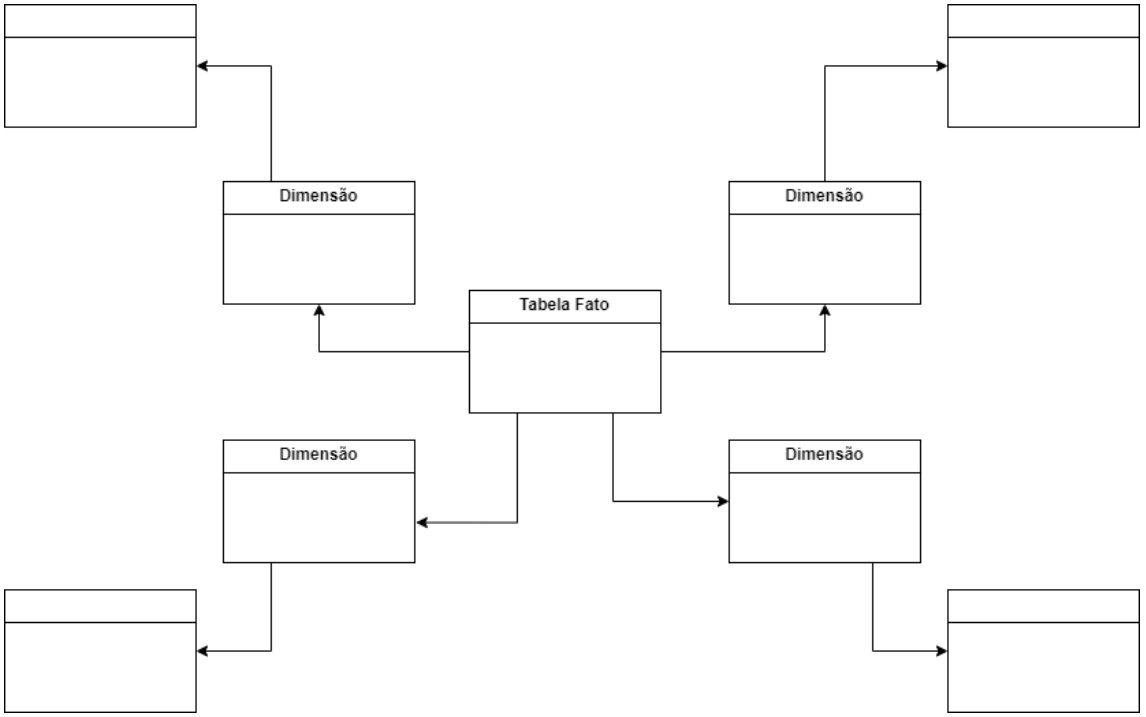
Figura 4. Modelo Estrela.



Fonte: Adaptado de Machado (2004, p. 93)

Já o modelo floco de neve é uma variação do modelo estrela em que as tabelas de dimensão são normalizadas em várias tabelas, criando uma estrutura hierárquica. A Figura 5 apresenta um exemplo do modelo floco de neve onde são subdivididas as tabelas de dimensão em tabelas de subdimensão assim eliminam a repetição de dados. A normalização pode melhorar o desempenho de consultas que envolvem múltiplas dimensões, pois reduz a quantidade de dados a serem processados, porém a desvantagem é que a estrutura hierárquica pode tornar o modelo mais complexo de entender e implementar.

Figura 5. Modelo Floco de Neve.



Fonte: Adaptado de Machado (2004, p. 94).

Inmon (1996) explica os *Data Marts* são como estruturas de dados que contêm informações de acordo com interesse e necessidade de um setor de uma organização, isto é, as informações são armazenadas por áreas ou assuntos específicos. Em uma empresa podem ser elaborados *Data Mart* para setores diferentes, com objetivos diferentes e ao final de seu desenvolvimento podem atender um conjunto de necessidades bem específicas para aquele setor.

No processo de modelagem o *Data Mart* pode seguir duas propostas distintas: a modelagem a partir de DW ou partir do sistemas transacionais. Começando pelo o DW é possível ter uma maior consistência e mapeamento dos dados, tendo em vista que na modelagem do projeto é possível ter uma visão holística da instituição. Já os *Data Marts* independentes a partir de sistemas transacionais de um setor, não tem foco corporativo, este fato implica que as informações de um *Data Mart* não terão nenhuma conexão com outro *Data Mart* de outros departamentos do negócio.

De acordo com Machado (2004) um *Data Mart* ser independente implica em rápida implementação, pois o número de necessidades de um departamento é sempre menor que o de toda instituição, outro fator que implica é o número de conexões, em consequência o projeto acaba tendo um baixo impacto nos recursos tecnológicos e menor custo. Já os *Data Mart* integrados com tal arquitetura são praticamente um *Data Warehouse* em múltiplos *Data Mart*,

mesmo sendo desenvolvidos em de forma independentes são interconectados e permitem uma visão de todos o*s Data Mart*.

Outro aspecto importante para modelagem de um *Data Mart* para Elmasri e Navathe (2011) é ser multidimensional pois tiram proveito dos relacionamentos inerentes nos dados para preencher os dados em matrizes multidimensionais, chamadas cubos de dados. Um cubo é caracterizado por uma tabela de fatos ministrando as dimensões, formado por três elementos básicos: fatos; dimensões; medidas (variáveis).

De acordo com Kimball (2002) para a identificação e entendimento de um fato é preciso primeiro reconhecer os quatro pontos de referência de um fato, denominados de pontos cardeais de um fato. Um fato consegue, numa única palavra, transmitir quatro informações: Onde aconteceu o fato; quando o fato; quem executou o fato; O que é o objetivo do fato.

2.4 Processo ETL

O processo ETL (*Extract, Transform, Load*) é uma fase essencial para criação tanto do DW quanto do *Data Mart*, pois converte os dados em formatos que viabilizam a análise. Isso pode envolver a limpeza dos dados, remoção de campos vazios ou nulos, a integração de várias fontes de dados, a manipulação de valores para criar novos campos ou atributos, e a aplicação de regras de negócios.

De acordo com Kimball e Caserta (2004), um sistema ETL devidamente projetado é capaz de extrair dados de sistemas de origem, impor padrões de qualidade de dados e consistência, conformar dados para que fontes independentes possam ser usadas em conjunto e, por fim, entregá-los em um formato que possa ser utilizado pelos usuários finais para tomada de decisões. Essa etapa é essencial na criação tanto do *Data Warehouse* quanto do *Data Mart*, garantindo a integridade e qualidade dos dados que serão utilizados nas análises e consultas.

O processo ETL é composto de três etapas: Extração, Transformação e Carga de Dados (como apresentado na Figura 3). As etapas de extração e carga são obrigatórias para o processo, sendo a transformação/limpeza opcional.

Figura 6. Estrutura do ETL.

Fonte: Adaptado de Kimball e Caserta (2004)

As etapas apresentadas na figura anterior serão detalhadas a seguir:

- **Extração dos dados**: Consiste na obtenção dos dados relevantes para a construção do *Data Mart* que estão armazenados nos banco de dados transacionais da organização, é comum que estes dados estejam organizados em diferentes bases de dados e em formatos diferentes. Kimball (2004) diz que o processo de extração resulta na leitura e na compreensão das fontes de dados operacionais da organização.“As informações podem ser encontradas nas mais variadas fontes, como, por exemplo, em bancos de dados relacionais, em arquivos-textos, arquivos binários, planilhas eletrônicas, banco de dados orientado a objeto etc.” (Gonçalves, 2003, p. 66).

  Com os dados já obtidos e armazenados, é necessário realizar uma análise minuciosa nos dados para encontrar inconsistências. Nesta etapa, os dados devem ser organizados para se adequar ao processo de transformação, essa etapa pode ser lenta dependendo dos tipos de dados, do volume de dados e das fontes de dados.
- **Transformação dos Dados**: Implica na execução de uma série de atividades e a aplicação de várias técnicas para converter, padronizar e limpar para que o formato dos dados possa atender aos requisitos de esquema do banco de dados de destino. Essa etapa depende unicamente dos dados extraídos e da modelagem do projeto. As fontes de dados de qualidade não exigirão muitas transformações, enquanto outros conjuntos de dados podem exigir significativamente. Para atender as necessidades do projeto os dados podem ser submetidos a várias técnicas de transformação.
- **Carga dos Dados**: Essa etapa é considerada a mais rápida do processo ETL pois consiste no ato de carregar os dados preparados anteriormente no banco de dados de destino. É possível realizar isso de forma automatizada ou manualmente, as duas formas são válidas não são relevantes para o resultado final.

Uma vez que o processo de ETL tenha sido concluído, os dados estarão prontos para serem analisados e explorados através de ferramentas OLAP. Essas ferramentas permitem que os usuários interajam com os dados de maneira intuitiva e flexível, permitindo a visualização de dados em diferentes dimensões e perspectivas.

2.5 Trabalhos Relacionados

A abordagem proposta por Moniz Junior (2021) buscou aplicar o *Business Intelligence* aos dados do Sistema Brasileiro fornecidos pelo INEP. O autor tinha como objetivo investigar as contribuições e desafios associados à utilização de BI nos dados disponibilizados pelo sistema brasileiro. Para alcançar esse propósito, foi necessário o desenvolvimento de interfaces de acesso aos dados fornecidos pelo INEP, a fim de integrá-los ao ambiente de BI.

Para realizar sua pesquisa, Moniz Junior (2021) teve que examinar e tratar os dados fornecidos pelo INEP. Esses dados são gerados anualmente pelos repositórios do INEP e disponibilizados em formato de planilhas. Após essa disponibilização, os dados passaram por uma etapa de pré-processamento por meio do *Extract Transform Load* (ETL). Uma vez processados, os dados refinados foram armazenados em um *Data Warehouse* (DW), que é um banco de dados multidimensional capaz de criar cubos OLAP. Isso permitiu a organização dos dados em diferentes dimensões, facilitando a análise mais rápida e conveniente.

De acordo com Moniz Junior (2021), a ferramenta desenvolvida alcançou resultados satisfatórios ao gerar gráficos e informações interativas. Esses recursos auxiliam na divulgação das informações e na tomada de decisões, além de abrir caminho para a aplicação de técnicas mais avançadas, como a mineração de dados.

O estudo realizado por Silva e Araújo (2020) aborda o processo de implantação de um Data Mart e a automatização dos processos ETL para o sistema de procedimentos extrajudiciais no Ministério Público do Estado do Tocantins. Os autores identificaram a necessidade do Ministério Público Estadual de implementar tecnologias que agregassem valor ao tratamento dos dados disponíveis. Nesse contexto, o objetivo central do trabalho consistiu em desenvolver e automatizar o processo ETL, visando integrar e padronizar os dados referentes aos procedimentos extrajudiciais do Ministério Público do Tocantins.

Na construção do processo, Silva e Araújo (2020) deram início à implementação de um modelo de entidade e relacionamento do sistema do Ministério Público do Estado do Tocantins (MPE/TO), destinado a ser definido no modelo de abstração do Data Mart. Após a

elaboração do modelo físico, foi efetuada a implementação do modelo abstrato para armazenar os dados no *Data Mart*. Posteriormente, os autores procederam à integração dos dados utilizando a ferramenta *Pentaho Data Integration* e, por fim, desenvolveram os *jobs* responsáveis pela automação do processo ETL.

De acordo com Silva e Araújo (2020), os estudos realizados proporcionaram um embasamento teórico sólido para a implementação do trabalho. A utilização da ferramenta Pentaho possibilitou a construção eficiente do Data Mart, contribuindo para agilizar e simplificar o processo de integração dos dados. Como resultado, foi viabilizada a criação de uma interface intuitiva e de fácil aprendizado para a execução dos Jobs, promovendo uma maior eficácia e acessibilidade no gerenciamento dos procedimentos extrajudiciais no Ministério Público do Estado do Tocantins.

Os trabalhos de Moniz Junior (2021) e Silva e Araújo (2020) contribuem para a construção deste trabalho, ampliando a compreensão sobre a aplicação de tecnologias BI e a automatização de processos ETL em contextos distintos. O estudo de Moniz Junior (2021) explorou a aplicação do BI aos dados do Sistema Brasileiro fornecidos pelo INEP, oferecendo informações sobre as contribuições e desafios associados a essa prática. Por outro lado, o trabalho de Silva e Araújo (2020) concentrou-se na implementação de um *Data Mart* e na automatização dos processos ETL para o sistema de procedimentos extrajudiciais no Ministério Público do Estado do Tocantins, evidenciando a relevância da tecnologia na otimização da gestão de informações em instituições públicas.

3 METODOLOGIA

3.1 Materiais

As tecnologias utilizadas no desenvolvimento deste trabalho foram selecionadas com base na disponibilidade e compatibilidade com as demais ferramentas escolhidas, além de considerar a necessidade de evitar altos custos de licenciamento. Para a implementação do Data Mart, foram utilizadas as seguintes tecnologias:

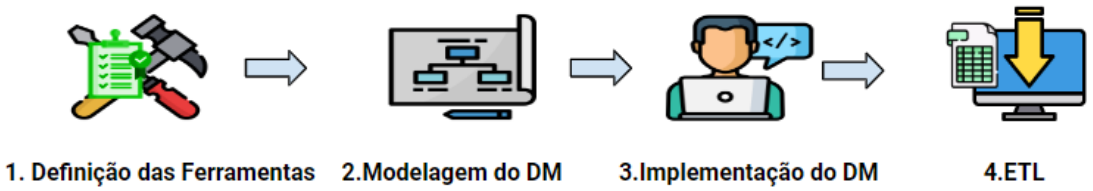
- **Draw.io**: é um software de desenho gráfico *web* que é usado direto no navegador usado para criar diagramas como fluxogramas, wireframes, diagramas UML, organogramas e diagramas de rede, utilizada para criar o diagrama lógico do *Data Mart.*
- **Pentaho Data Integration (PDI): é** responsável pela integração de dados oriundos de fontes diferentes, por meio de técnicas ETL. Ele lê e escreve em vários formatos de sistemas de gerenciamento de banco de dados (SGBD), como *Oracle*, *SQLServer* e *MySql*. Utilizada para realizar todo o processo ETL. É uma ferramenta de código aberto comumente utilizada no processo ETL para extração, transformação e carga de dados.
- **Microsoft SQL Server: o** *Microsoft SQL Server* é um sistema gerenciador de Banco de dados relacional (SGBD) desenvolvido pela Sybase em parceria com a Microsoft. Utilizado para criar e gerenciar o banco de dados e o diagrama físico.
- **Microdados ENADE: o**s microdados utilizados são desde a primeira edição em 2004 até a última 2022 (atualizado em 10/11/2023) que estão disponíveis no site oficial do INEP[3] e pode ser acessado por qualquer pessoa com acesso a *internet*.

3.2 Métodos

Para a elaboração deste trabalho, foi realizada uma verificação para ver se todos os arquivos baixados estavam completos com dicionário de arquivos, dicionário variáveis e os microdados do ENADE, obtidos no portal do INEP, o órgão responsável pela divulgação desses dados. O desenvolvimento deste estudo seguiu as etapas descritas no fluxograma ilustrado na Figura 4.

[3] https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/enade

Figura 7. Etapas de desenvolvimento do trabalho



Fonte: Próprio Autor.

A Figura 4 descreve as etapas executadas durante o desenvolvimento do trabalho, que são:

- **Definição das Ferramentas**: a escolha das ferramentas destacadas no tópico 3.1, foi feita baseada em trabalhos anteriores que serviram de base para entender quais ferramentas são necessárias e quais são compatíveis para se utilizar juntas no desenvolvimento do *Data Mart,* sendo duas exeções as ferramentas Diagrams.net pois essas ferramentas mesmo não sendo utilizadas em trabalhos anteriores que serviram de referencial teórico, foram adotadas dado o uso e popularidade nos dias de hoje e a compatibilidade com as demais ferramentas.
- **Modelagem do DM:** a modelagem do *Data Mart* foi feita para possibilitar a etapa de implementação do *Data Mart*, essa etapa foi realizada em duas ferramentas diferentes, a primeira foi *Diagrams.net* usada para a criação do Modelo Lógico que apresenta o ideia de *Data Mart* de forma mais abstrata. A segunda ferramenta usada foi o SQL Server Management Studio para a criação do Modelo Físico pois esse diagrama do modelo Físico reflete diretamente nas tabelas do banco de dados e apresenta de forma gráfica todos os relacionamento entre as tabelas.
- **Implementação do DM:** o D*ata Mar*t foi criado utilizando o SQL Server Management Studio Server e o Pentaho Data Integration (Kettle).
- **ETL:** os microdados foram baixados do site do INEP e, em seguida, pré-processados sendo descompactados pois estavam no formato de arquivo compactado .zip e .7z os quais impedia de serem manipulados pelas as ferramentas, com a ferramenta Pentaho Data Integration (Kettle) para realizar a transformação, limpeza e padronização dos dados antes de carregá-los no *Data Mart*.

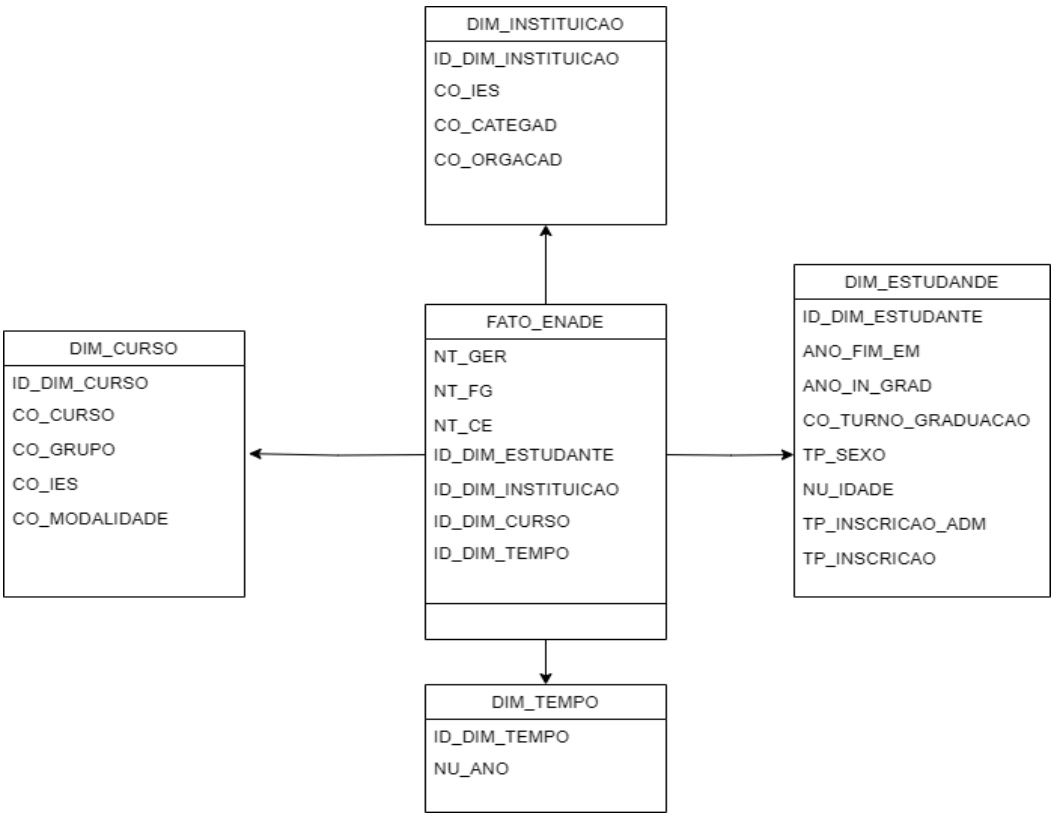
4 RESULTADOS E DISCUSSÕES

O intuito deste trabalho foi desenvolver um *Data Mart* para organizar e permitir analisar os microdados do ENADE. Esta seção apresenta os resultados obtidos do desenvolvimento do Data Mart utilizando o SQL Server e o processo ETL com a ferramenta Pentaho Data Integration.

4.1 Modelo Lógico e Modelo Físico do Data Mart

O modelo lógico no esquema estrela é representado pela Figura 8, que descreve as entidades (fatos e dimensões) e seus relacionamentos, essas entidades foram abstraídas do dicionário de arquivos e variáveis microdados Enade, onde são listados os atributos de cada entidade os arquivos das outras edições do enade tem a mesma organização e mesmas variáveis salvo em alguns casos como foi a pandemia que foram criadas algumas variáveis para medir o impacto. O relacionamento entre as entidades foi estabelecido seguindo o modelo estrela que determina que a tabela fato conteria os dados numéricos relacionados ao desempenho. Com base na análise dos microdados, definiu-se o modelo lógico abaixo para o desenvolvimento do *Data Mart.*

Figura 8. Modelo Lógico do *Data Mart*



Fonte: Próprio Autor.

O modelo lógico apresentado pela Figura 8 indica o fato, suas dimensões e seus campos, apresentados em detalhes a seguir nas tabelas 1 a 5. A identificação dos elementos multidimensionais do *Data Mart* nos microdados revelou o exame ENADE como o fato, cujos valores são mutáveis e passíveis de análise ao longo do tempo. As dimensões estudante, curso, tempo e instituição fornecem o contexto para as métricas da tabela fato e criam relação entre a tabela fato e as dimensões por meio de chaves estrangeiras.

Tabela 1. Fato Enade

FATO_ENADE	
NT_GER	Nota bruta da prova
NT_FG	Nota bruta na formação geral
NT_CE	Nota bruta no componente específico
ID_DIM_ESTUDANTE	Chave técnica; o identificador da dimensão Dim_ESTUDANTE
ID_DIM_INSTITUICAO	Chave técnica; o identificador da dimensão Dim_INSTITUICAO
ID_DIM_CURSO	Chave técnica; o identificador da dimensão Dim_CURSO
ID_DIM_TEMPO	Chave técnica; o identificador da dimensão Dim_TEMPO

Tabela 2. Dimensão Instituição

DIM_INSTITUICAO	
ID_DIM_INSTITUICAO	Chave técnica; o identificador da dimensão Dim_INSTITUICAO
CO_IES	Código da IES (e-MEC)
CO_CATEGAD	Código da categoria administrativa da IES
CO_ORGACAD	Código da organização acadêmica da IES

Tabela 3. Dimensão Curso

DIM_CURSO	
ID_DIM_CURSO	Chave técnica; o identificador da dimensão Dim_CURSO
CO_CURSO	Código do curso no Enade
CO_GRUPO	Código da Área de enquadramento do curso no Enade
CO_MODALIDADE	Código da Modalidade de Ensino

Tabela 4. Dimensão Estudante

DIM_ESTUDANDE	
ID_DIM_ESTUDANTE	Chave técnica; o identificador da dimensão Dim_ESTUDANTE
ANO_FIM_EM	Ano de conclusão do Ensino Médio
ANO_IN_GRAD	Ano de início da graduação
CO_TURNO_GRADUACAO	Código do turno de graduação cursado pelo estudante
TP_SEXO	Sexo
NU_IDADE	Idade do inscrito

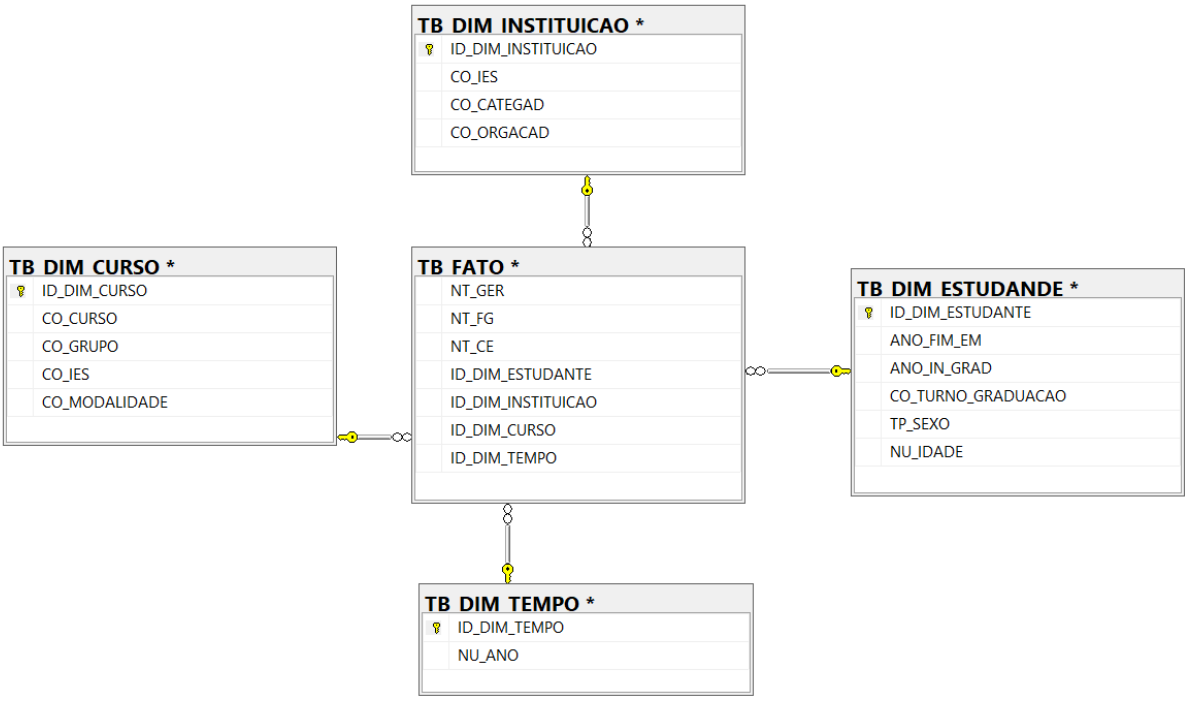
Tabela 5. Dimensão Tempo

DIM_ANO	
ID_DIM_ANO	Chave técnica; o identificador da dimensão Dim_ANO
NU_ANO	Ano de realização do exame

O banco de dados foi criado na ferramenta SQL *Server Management Studio* seguindo o modelo lógico que foi elaborado na fase modelagem do projeto, o diagrama representa as tabelas do banco de dados e seus relacionamentos, onde as chaves primárias das tabelas dimensões são a chave primária da tabela fato.

O modelo lógico foi transformado em um modelo físico utilizando a ferramenta SQL Server Management Studio pois nessa ferramenta foi onde foi criado os relacionamentos das tabelas do banco de dados através de chaves modelo físico que foi usado para a criação do *Data Mart* posteriormente. A Figura 5 apresenta o modelo físico, evidenciando a tabela de fatos e as dimensões do *Data Mart* implementadas, dispostas em um esquema estrela. Este arranjo facilita a organização e o acesso aos dados, permitindo análises eficientes e estruturadas das informações centralizadas.

Figura 9. Modelo Físico do *Data Mart*



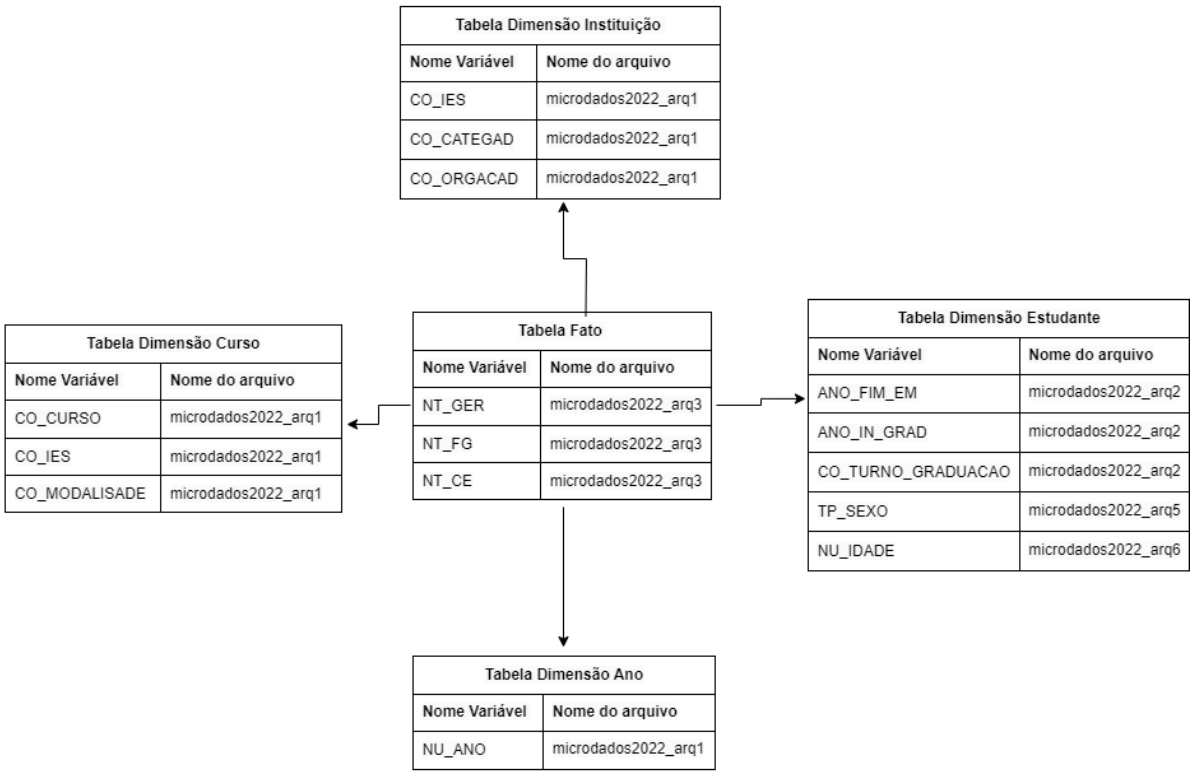
Fonte: Próprio Autor.

O modelo físico apresentado pela Figura 9 indica o fato ENADE, suas dimensões e seus campos, apresentados no modelo lógico. Nesta implementação, as dimensões Dim_Curso, Dim_Tempo, Dim_Avaliação, Dim_Estudante e Dim_Instituição passaram a ter novas propriedades para o gerenciamento no banco de dados como a adição de chaves primárias das tabelas.

4.2 Processo ETL com Kettle

Os dados foram extraídos (download) do site do INEP em formato compactado .zip ou .7z e, após baixados, foram descompactados para serem analisados para verificar se todos os arquivos necessários estavam nos arquivos baixados, pois todos os arquivos baixados de cada edição devem ter a mesma estrutura. Sendo dois diretórios: um chamado LEIA-ME, que contém os arquivos "Manual do usuário Enade.pdf", "Questionário do Estudante.pdf', "Dicionário arquivos variáveis microdados Enade.ods" e "Dicionário arquivos variáveis microdados Enade.xlsx"; e outro chamado DADOS, que contém os arquivos referentes aos microdados do Enade em formato .txt.

Figura 10. Diagrama de definição de variáveis



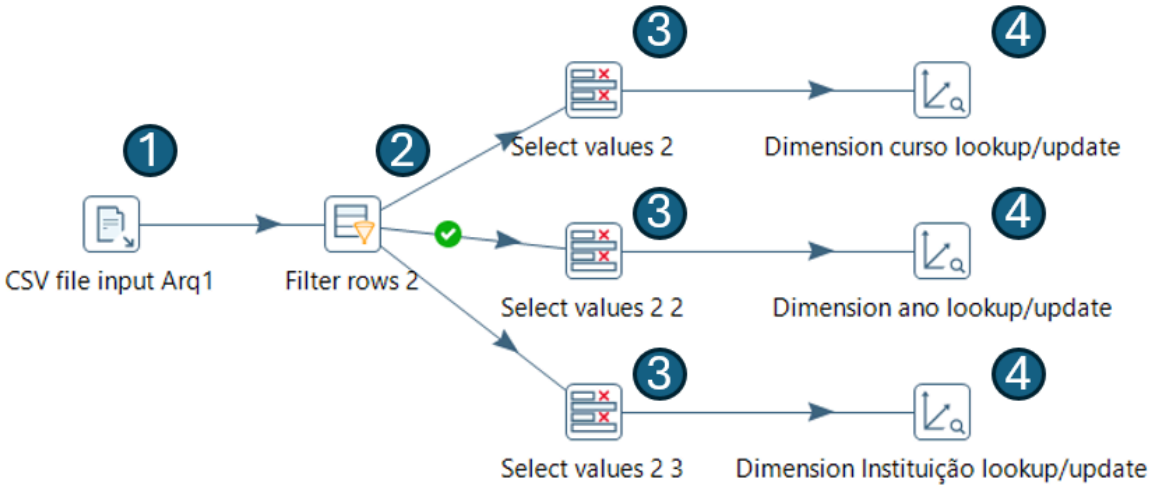
Fonte: Próprio Autor.

Após analisar os arquivos e o manual do usuário foram identificadas quais variáveis dos dados seriam utilizadas para compor o *Data Mart* baseado no modelo lógico do *Data Mart* que foi modelado previamente, as variáveis estavam separadas por arquivos para preservar a identificação do estudante sendo necessário em alguns casos utilizar variáveis de mais de um arquivo para compor um tabela que será carregada no banco.

4.2.1 Processo de transformação e carregamento das Dimensões Curso, Ano e instituição

Na ferramenta Kettle o processo de transformação e carregamento das dimensões curso, ano e instituição inicia na marcação 1 onde é feito o *input* do arquivo microdados2022_arq1 que contém todas as variáveis que serão utilizadas para a construção das três dimensões, nessa etapa de carregamento foi necessário definir na ferramenta quais os parâmetros para a leitura do arquivo .txt, parâmetros do tipo quais os delimitadores do arquivos(;) , o tipo de codificação binária de comprimento variável que é UTF-8 e o tipo de dados que cada coluna irá conter (Inteiro, Binário, String).

Figura 11. Processo ETL das dimensões curso, ano e instituição



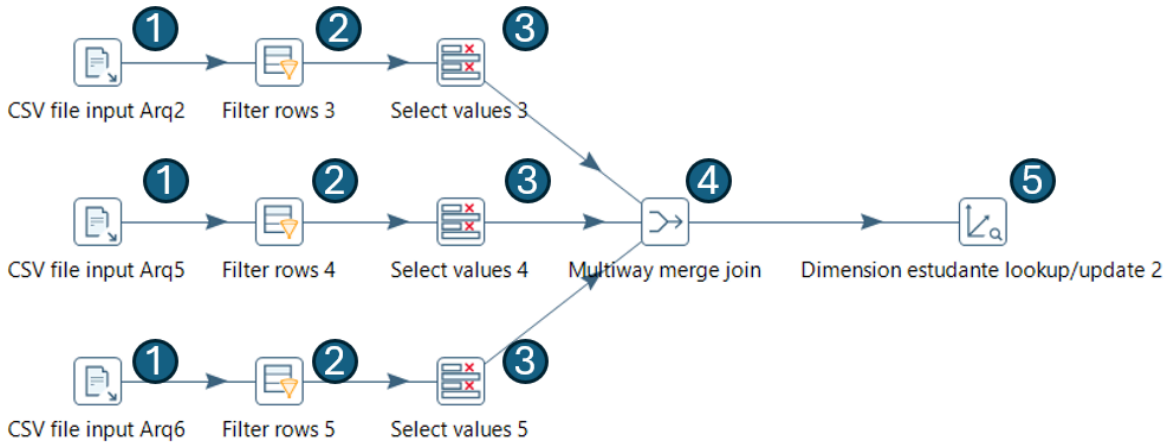
Fonte: Próprio Autor.

A atividade feita na marcação 2 são as filtragens para garantir que nenhum valor esteja fora dos limites estabelecidos no Dicionário de Variáveis do ENADE e que valores não são nulos. Para a marcação 3 são feitas as filtragens das colunas que irão compor a nova tabela dimensão correspondente levando em consideração a modelagem do modelo lógico do *Data Mart.* A marcação 4 faz um update da dimensão na tabela correspondente do banco de dados SQL server que está conectado direto na ferramenta Kettle com as variáveis selecionadas na marcação 3.

4.2.2 Processo de transformação e carregamento da Dimensão estudante

O processo de transformação e carregamento da dimensão estudante inicia na marcação 1 com o input de três arquivos de dados que contém os dados dos estudantes que realizaram o enade, na marcação 2 é realizada a limpeza dos dados removendo o dados nulos e verificando se valores estão seguindo os parâmetros estabelecidos pelo dicionário de variáveis do enade, marcação 3 são selecionadas as variáveis que irão formar a dimensão estudante em seguida na marcação 4 é realizado um merge join a etapa permite a junção da colunas que foram selecionadas na marcação 3, a marcação 5 realiza a atualização da dimensão no banco de dados.

Figura 12. Processo ETL das dimensões estudante



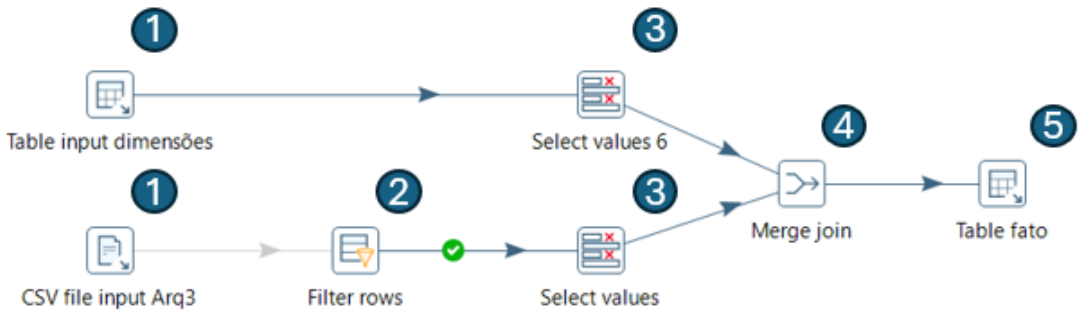
Fonte: Próprio Autor.

Embora a dimensão estudante siga um processo similar às demais, a distribuição dos dados em múltiplos arquivos exigiu uma abordagem diferenciada. Ao contrário das outras dimensões, não foi possível realizar a transformação e carregamento em um único fluxo. Adicionalmente, a necessidade de combinar informações de diferentes fontes demandou a inclusão de uma etapa extra de *merge join* para unificar as colunas selecionadas dos diversos arquivos.

4.2.3 Processo de transformação e carregamento do fato enade

As transformações e carregamento para a tabela fato ENADE iniciam-se na marcação 1 com o *input* informações do microdados2022_arq3 e com o *input* das tabelas dimensões Estudante, Curso, Ano e Instituição usando instruções SQL onde é feito o *select* em todas as chaves primárias das tabelas dimensões para obter as chaves primárias substitutas que correspondem aos valores das dimensões presentes nos microdados.

Figura 13. Processo ETL do fato ENADE



Fonte: Próprio Autor.

A marcação 2 realiza a limpeza dos dados removendo os dados nulos pois o *input* é direto do arquivo de texto dos microdados e verifica se valores estão dentro dos parâmetros estabelecidos pelo dicionário de variáveis do enade. A marcação 3 são feitas as filtragens das colunas que irão compor a nova tabela Fato onde é das dimensões serão todas as chaves primárias das tabelas e para o arquivo serão selecionadas apenas as colunas relacionada ao desempenho.

Na etapa indicada com marcação 4, foi realizada a junção das tabelas dimensões e as variáveis de desempenho com instrução *merge join* e tendo com saída dessa etapa a tabela fato pronta para ser carregada no banco de dados.

…

5 CONSIDERAÇÕES FINAIS

Este trabalho apresentou o desenvolvimento de um *Data Mart* para que fosse possível analisar os microdados do ENADE que são disponibilizados pelo INEP. O projeto fez uso do processo ETL para a manipulação dos dados, da ferramenta Pentaho Kettle e da modelagem de *Data Mart* baseada no modelo estrela, e toda implementação seguiu conforme a modelagem definida na fase de planejamento do projeto.

A ferramenta Pentaho Kettle permitiu que o processo de ETL fosse realizado sem a necessidade de codificação em nenhuma das etapas, pois foi possível realizar tudo via interface gráfica. A metodologia de criação de *Data Mart* utilizada foi baseada em outros trabalhos acadêmicos da mesma instituição, pois o método se demonstra viável na criação de um Data Mart.

Esse trabalho focou na criação de um Data Mart e para trabalhos futuros espera-se que seja desenvolvido outras tabelas fatos para se conectar e aumentar as dimensões dos dados; e que sejam desenvolvidas formas de conectar ferramentas de visualização de dados para geração de gráficos e relatórios a partir dos dados. As ferramentas e técnicas utilizadas podem ser aplicadas em outros trabalhos semelhantes, e o *Data Mart* construído pode ser expandido e integrado a outras fontes de dados aumentando as possibilidades de análise dos dados.

Em trabalhos futuros poderia ser trabalhada a geração de painéis, relatórios e a utilização de ferramentas OLAP, permitindo uma exploração visual dos dados através de dashboard, outros tipo de trabalhos que poderia se desenvolvidos seria integrar os microdados do ENADE com outras fontes de dados relevantes para o contexto, como informações socioeconômicas dos estudantes, dados sobre as instituições de ensino e indicadores de desenvolvimento regional, para obter uma visão mais completa do contexto educacional.

Muitos desafios foram enfrentados, como a complexidade dos dados que foi aumentada com a nova lei de proteção de dados LGPD e a necessidade de lidar com muitos arquivos no formato de texto, o projeto foi concluído, demonstrando que o processo de modelagem adotado e a utilização de ferramentas adequadas no desenvolvimento permite a criação de um *Data Mart* tornando possível a análise dos microdados do ENADE e a utilização desses dados para melhorar o ensino superior no Brasil.

REFERÊNCIAS

ALVARES, Reinaldo Viana; CAMPOS, Nathielly de Souza; GOMES, Vinicius Benter. **Adoção de Data Discovery para Apoio ao Processo de Análise de Dados do Enade**. 2015. 6 f. TCC (Graduação) - Curso de Ciência da Computação, Centro Universitário Augusto Motta, Rio de Janeiro, 2015. Disponível em: http://www.tise.cl/volumen11/TISE2015/480-485.pdf. Acesso em: 16 junho 2021.

ARAÚJO, Rodrigo. **Análise dos microdados do Enade: proposta de uma ferramenta de exploração utilizando mineração de dados**. 2019. 69 f. Dissertação (Mestrado em Ciência da Computação) - Universidade Federal de Goiás, Goiânia, 2019.

BOAR, Bernard H. **Strategic thinking for information technology: how to build the IT organization for the information age**. 1. ed. United States: John Wiley & Sons Inc., 1997. 270 p. ISBN 047115881X.

ELMASRI, Ramez; NAVATHE, Shamkant B. **Sistemas de banco de dados**. 6 ed, São Paulo: Pearson Addison Wesley,2011.

GOMES, Marco António Boialvo. **Modelação de um Data Warehouse para a Direcção-Geral do Tesouro e Finanças e implementação de um Data Mart para o processo de Gestão Patrimonial**. 2010. 99 f. Dissertação (Mestrado) - Curso de Estatística e Gestão da Informação, Instituto Superior de Estatística e Gestão da Informação, Universidade Nova de Lisboa, Lisboa, 2010. Disponível em: https://core.ac.uk/download/pdf/303713097.pdf. Acesso em: 01 jun. 2021.

GONÇALVES, Marcio. **Extração de Dados Para Data Warehouse**. Palmas: Axcel Books do Brasil p. 147, 2003.

INMON, W. H. The data warehouse and data mining. **Communications of the ACM**, v. 39, n.11, p.49-50, 1996.

INMON, W. H. et al. **Data Warehousing and OLAP Technologies**. Morgan Kaufmann Publishers Inc., 1997. https://bdasolutions.com.br/2020/08/como-funciona-a-modelagem-de-dados-em-solucoes-de-bi/.

KIMBALL, Ralph; ROSS, Margy. **The Data Warehouse Toolkit**. Rio de Janeiro: Editora Campus, ed. 2, p. 494, 2002.

KIMBALL, Ralph; CASERTA, Joe. **The Data Warehouse ETL Toolkit**: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data. Indianapolis: Wiley Publishing, Inc., p. 467, 2004.

MACHADO, Felipe Nery Rodrigues. **Tecnologia e Projeto de Data Warehouse:** Uma visão multidimensional. São Paulo: Editora Érica Ltda, 2004

SILVA, Thiago Aparecido da; ARAÚJO, Fábio Castro. **Implementação de um Data Mart e Automatização do Processo ETL para o Sistema de Procedimentos Extrajudiciais do MPE/TO**. In: ENCOINFO - Congresso de Computação e Tecnologias da Informação, 22., 2020, Palmas - TO. **Anais [...]**. Palmas - TO: CEULP/ULBRA, 2020. p. 89 - 94. ISSN e-ISSN: 2447-0767 versão online. Disponível em: https://ulbra-to.br/encoinfo/edicoes/2020/artigos/implementacao-de-um-data-mart-e-automatizacao-do-processo-etl-para-o-sistema-de-procedimentos-extrajudiciais-do-mpe-to/. Acesso em: 08 mai. 2024

BRASIL. **Lei nº 10.861, de 14 de abril de 2004**. Institui o Sistema Nacional de Avaliação da Educação Superior – SINAES e dá outras providências. Brasília, DF, Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2004-2006/2004/lei/l10.861.htm. Acesso em: 25 jun. 2024.

INEP. **Provas e Gabaritos.** Disponível em: <https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/enade/provas-e-gabaritos>. Acesso em: 27 jun. 2024.