

**UNIVERSIDADE FEDERAL DE ITAJUBÁ
INSTITUTO DE ENGENHARIA DE PRODUÇÃO E GESTÃO
MESTRADO PROFISSIONAL EM ADMINISTRAÇÃO**

FELIPE RODRIGO DE MORAES MENDES

**PROPOSTA DE UM DATA WAREHOUSE PARA APOIO À TOMADA DE
DECISÃO SOBRE EVASÃO INSTITUCIONAL EM UMA INSTITUIÇÃO FEDERAL
DE ENSINO SUPERIOR**

**ITAJUBÁ - MG
JUNHO DE 2020**

FELIPE RODRIGO DE MORAES MENDES

**PROPOSTA DE UM DATA WAREHOUSE PARA APOIO À TOMADA DE
DECISÃO SOBRE EVASÃO INSTITUCIONAL. EM UMA INSTITUIÇÃO FEDERAL
DE ENSINO SUPERIOR**

Dissertação submetida ao Programa de Mestrado Profissional em Administração da Universidade Federal de Itajubá, como requisito para obtenção do título de mestre em administração.
Orientador: Prof. Dr. Fabio Favaretto

**ITAJUBÁ – MG
JULHO DE 2020**

AGRADECIMENTOS

Primeiramente gostaria de agradecer minha esposa Jéssica e a minha enteada Isabelle, pelo carinho, companheirismo e compreensão no tempo de desenvolvimento desse trabalho. Também foi valioso o apoio do meu pai Edson, minha mãe Isabel e irmãs Rafaela e Júlia.

Também gostaria de agradecer meu orientador, prof. Favaretto, por toda a orientação dada, que, na medida exata, deu-me liberdade na condução da pesquisa e me direcionou para o caminho correto, cortando diversas arestas.

Ao prof. Egon e prof. Thiago, Pró-Reitor e Pró-Reitor Adjunto de graduação há época do começo da dissertação, por terem sido os primeiros a acreditarem na ideia e me apoiarem institucionalmente. Também agradeço ao prof. Rodrigo, atual Pró-Reitor, por apoiar a continuidade dos trabalhos.

RESUMO

Introdução: O apoio à tomada de decisão no contexto da evasão em instituições de ensino superior demanda acompanhamento sistêmico das taxas de evasão e a descrição dos alunos evadidos em variáveis pessoais e acadêmicas. A premissa é que o desenvolvimento de um repositório institucional que armazene dados histórico e não-voláteis, o *data warehouse*, seja a ferramenta que habilite a Universidade Federal de Itajubá (Unifei) a responder à essa demanda. **Objetivo:** O objetivo principal é o desenvolvimento e implementação de um DW para uso no apoio à tomada de decisão sobre a evasão na Unifei. Os objetivos secundários são conhecer métodos para o desenvolvimento de *data warehouses*, conhecer métodos e métricas utilizadas para o cálculo das taxas de evasão e conhecer as características descritivas dos discentes relacionadas à evasão. **Métodos:** O objeto de estudo é a Unifei e o sistema de informação gerencial em uso na instituição, o Sistema Integrado de Gestão de Atividades Acadêmicas (SIGAA). Para guiar a intervenção na realidade foi proposta uma metodologia utilizando a *soft systems methodology* e o método de desenvolvimento de *data warehouses* de Ralph Kimball. Para a obtenção de dados utilizou-se pesquisas bibliográfica, em periódicos da área de educação para o levantamento de dados sobre a evasão, e documental, que investigou o banco de dados do SIGAA. **Resultados:** A pesquisa bibliográfica revelou três grupos de métodos para cálculo das taxas de evasão: métodos que consideram gerações completas de alunos, tanto no prazo ideal ou prazo máximo de integralização, métodos que consideram apenas ingressantes dentro de um período específico e métodos que propõem o acompanhamento de séries históricas. Os pontos positivos e negativos foram contrapostos entre os grupos e uma proposta de método de cálculo foi derivada da literatura. Também da fonte bibliográfica foram identificadas 65 variáveis descritivas sobre os discentes, sendo 33 variáveis pessoais (50,5%) e 32 variáveis ligadas à vida acadêmica dos alunos (49,2%). Pesquisa documental foi conduzida no banco de dados do SIGAA e a posterior triangulação com o resultado das pesquisas bibliográficas mostraram que o SIGAA armazena os dados necessários para se calcular taxas de evasão dos cursos desde 1998. Ademais o SIGAA armazena 23 (35,4%) das variáveis descritivas identificadas na literatura. Os requisitos de informação derivados da triangulação das fontes de informação foram utilizados para nortear o desenvolvimento do *data warehouse*. Modelos arquiteturais dos dados, tanto os dimensionais como os físicos, foram realizados. Descreveu-se as etapas necessárias para realizar a extração dos dados de suas fontes, os tratamentos realizados e o carregamento para o ambiente do *data warehouse*. **Conclusão:** As evidências trazidas pelos resultados apoiam a afirmativa de que o desenvolvimento e implementação de um DW institucional é ferramenta que permite a instituições de ensino superior a acompanharem o fenômeno da evasão de forma sistêmica, periódica e constante.

Palavras-chave: Evasão no ensino superior; *data warehouse* educacional; *soft systems methodology*.

ABSTRACT

Introduction: Decision making support in the context of dropout in higher education institutions requires systemic monitoring of dropout rates and the description of dropout students by personal and academic variables. The initial premise is the development of an institutional repository that stores historical and non-volatile data, the data warehouse, is the tool that enables the Federal University of Itajubá (Unifei) to respond to this demand. **Objective:** The main objective is the development and implementation of a DW for decision-making support on dropouts at Unifei. The secondary objectives are to discover methods for the development of data warehouses, to discover methods and metrics used to calculate dropout rates and to know the descriptive characteristics of students related to dropout. **Methods:** The object of study is Unifei and the management information system in use at the institution, the *Sistema Integrado de Gestão de Atividades Acadêmicas* (SIGAA). To guide the intervention, a methodology was proposed using soft systems methodology and the Ralph Kimball's method of developing data warehouses. Bibliographic research was conducted in education journals to collect data related to dropouts. Documentary research was conducted on SIGAA's database to assess the data available. **Results:** The bibliographic research revealed three groups of methods for calculating dropout rates: methods that consider complete generations of students, methods that consider only newcomers within a specific period and methods that propose the monitoring of historical series. The positive and negative points were contrasted between the groups and a proposed calculation method was derived from the literature. Also, from the bibliographic source, 65 descriptive variables about the students were identified, being 33 personal variables (50.5%) and 32 variables related to the students' academic life (49.2%). Documentary research was conducted in the SIGAA database and the subsequent data triangulation showed that SIGAA stores the data necessary to calculate dropout rates from courses since 1998. Furthermore, SIGAA stores 23 (35.4%) descriptive variables identified in the literature. The information requirements derived from the data sources triangulation were used to guide the development of the data warehouse. Architectural models of the data, both dimensional and physical, were realized. The necessary steps to extract data from its sources, treatments performed and loading into the data warehouse environment were described. **Conclusion:** The evidence brought by the results supports the statement that the development and implementation of an institutional DW is a tool that allows higher education institutions to monitor the phenomenon of dropout in a systematic, periodic, and constant manner.

Key Words: Dropouts in higher education; educational data warehouse; soft systems methodology.

LISTA DE ILUSTRAÇÕES

Figura 1	Exemplo de entidade, atributo e instâncias de uma entidade.....	p. 24
Figura 2	Relacionamento entre entidades e normalização.....	p. 25
Figura 3	Concatenação das tabelas relacionadas entre si.....	p. 25
Figura 4	Representação esquemática de uma consulta em SQL.....	p. 26
Figura 5	Abordagem proposta por Bill Inmon.....	p. 29
Figura 6	Abordagem proposta por Ralph Kimball.....	p. 29
Figura 7	Representação icônica do ciclo de aprendizagem da SSM.....	p. 41
Figura 8	Ciclo de vida do <i>data warehouse</i>	p. 41
Figura 9	Proposta metodológica baseada no ciclo de aprendizagem da SSM.....	p. 43
Figura 10	Código para reaver nome de tabelas no PostgreSQL.....	p. 46
Figura 11	Primeiro Passo de Aplicação da Metodologia Proposta.....	p. 49
Figura 12	Primeira etapa do segundo passo de aplicação da metodologia proposta.....	p. 51
Figura 13	Segunda etapa do segundo passo de aplicação da metodologia proposta.....	p. 60
Figura 14	Terceiro passo de aplicação da metodologia proposta.....	p. 62
Figura 15	Quarto passo de aplicação da metodologia proposta.....	p. 68
Figura 16	Modelo dimensional do DW.....	p. 70
Figura 17	Modelo físico do DW.....	p. 71
Figura 18	Quinto passo de aplicação da metodologia proposta.....	p. 74
Figura 19	Gráfico da Evolução das Taxas Anuais de Evasão e Diplomação da Unidade de Observação "A"	p. 75
Figura 20	Gráfico da Evolução das Taxas Anuais de Evasão e Diplomação da Unidade de Observação "B"	p. 77

LISTA DE TABELAS

Tabela 1	Variáveis descritivas pessoais dos discentes.....	p. 58
Tabela 2	Variáveis descritivas acadêmicas dos discentes.....	p. 59
Tabela 3	Dados pessoais e acadêmicos registrados pelo SIGAA.....	p. 61
Tabela 4	Triangulação da pesquisa bibliográfica e documental.....	p. 64
Tabela 5	Apenas variáveis pessoais e acadêmicas registradas pelo SIGAA.....	p. 65
Tabela 6	Requisitos de informação para o DW.....	p. 67

LISTA DE SIGLAS

BD	Banco de dados
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CEPEAd	Conselho de Ensino, Pesquisa, Extensão e Administração da Unifei
DAE	Diretoria de Assistência Estudantil da Unifei
DIS	<i>Data item set</i>
DM	<i>Data Mart</i>
DW	<i>Data warehouse</i>
DWI	<i>Data warehouse</i> institucional
ETL	<i>Extract Transform and Load</i>
ERD	<i>Entity relation diagram</i>
FK	<i>Foreign key</i>
IES	Instituições de Ensino Superior
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
MEC	Ministério da Educação e Cultura
MPA	Mestrado Profissional em Administração da Unifei
NDE	Núcleo Docente Estruturante
PK	<i>Primary key</i>
PNAES	Programa Nacional de Assistência Estudantil
PRG	Pró-Reitoria de Graduação
REUNI	Programa de Reestruturação e Expansão das Universidades Federais
SGBD	Sistemas de gerenciamento de banco de dados
SI	Sistemas de informação
SIGAA	Sistema Integrado de Gestão de Atividades Acadêmicas
SINFO	Superintendência de Informática da UFRN
SISU	Sistema de Seleção Unificada
SQL	<i>Structured Query Language</i>
SSM	<i>Soft Systems Methodology</i>
SPT	Sistemas de processamento de transações
TCU	Tribunal de Contas da União
TEA	Taxa de evasão no ano
TI	Tecnologia da informação
TMA	Total de matrículas no ano
UF	Unidade Federativa
UFRN	Universidade Federal do Rio Grande do Norte
Unifei	Universidade Federal de Itajubá
VPA	Veteranos para o próximo ano

SUMÁRIO

1	INTRODUÇÃO	17
1.1	APRESENTAÇÃO DO TEMA	17
1.2	PROBLEMA DE PESQUISA	21
1.3	DOS OBJETIVOS DA PESQUISA	22
1.4	DA ESTRUTURA DA DISSERTAÇÃO	22
2	REVISÃO TEÓRICA	23
2.1	CONCEITOS SOBRE BANCOS DE DADOS	23
2.2	DATA WAREHOUSES	27
2.2.1	Conceitos Sobre o Desenvolvimento de Data Warehouses	28
2.2.2	Extract, transform e load: o processo da ETL	32
3	METODOLOGIA	37
3.1	DECISÕES METODOLÓGICAS	38
3.1.1	Soft Systems Methodology	40
3.1.2	Método de Ralph Kimball	41
3.2	PROPOSTA METODOLÓGICA	42
3.2.1	Pesquisa Bibliográfica e Investigação de Ações Intencionais	44
3.2.2	Pesquisa Documental e a Realidade dos Dados	45
3.2.3	Requisitos do Data Warehouse	46
4	RESULTADOS E DISCUSSÃO	49
4.1	REQUISITOS GERAIS PARA O DESENVOLVIMENTO DO DW	49
4.2	MÉTODOS DE CÁLCULO DAS TAXAS DE EVASÃO	50
4.3	VARIÁVEIS DISCENTES RELEVANTES À EVASÃO	57
4.4	REALIDADE DOS DADOS DA UNIFEI	60
4.5	REQUISITOS DE INFORMAÇÃO	62
4.6	CRIAÇÃO DO DATA WAREHOUSE	68
4.7	IMPLEMENTAÇÃO E UTILIZAÇÃO DO DATA WAREHOUSE	73
5	CONCLUSÃO E SUGESTÕES PARA FUTURAS PESQUISAS	79
	REFERÊNCIAS	85
	APÊNDICE A – DESCRIÇÃO DAS TABELAS E ATRIBUTOS DO DW	91
	APÊNDICE B – SCRIPT EM SQL PARA CRIAÇÃO DAS TABELAS DO DW	97
	APÊNDICE C – SCRIPT EM SQL PARA EXTRAÇÃO DE DADOS PARA AS TABELAS FATO DISCENTE E DIMENSÕES PESSOA, CURSO E SITUAÇÃO	99
	APÊNDICE D – SCRIPT EM SQL PARA EXTRAÇÃO DOS DADOS PARA A TABELA FATO EVASÃO	101
	APÊNDICE E – PRINCIPAIS ARTIGOS UTILIZADOS NA PESQUISA BIBLIOGRÁFICA	109

1 INTRODUÇÃO

O desenvolvimento desse trabalho se deu no âmbito do Programa de Mestrado Profissional em Administração (MPA) da Universidade Federal de Itajubá (Unifei). Como requisito para ingresso no Programa e desenvolvimento da dissertação está o vínculo com uma organização profissional para a aplicação dos conhecimentos gerados no mestrado com a atuação em uma situação problemática real.

A atuação do autor para o desenvolvimento desse trabalho se deu na Pró-Reitoria de Graduação (PRG) da Unifei, onde é servidor desde 2014. O interesse pelo autor sobre a temática da evasão se desenvolveu por meio de sua atuação profissional na Diretoria de Assistência Estudantil (DAE), subordinada à PRG, de 2015 a 2016. A DAE é responsável por operacionalizar o Decreto n.º 7.234, que instituiu o Programa Nacional de Assistência Estudantil (PNAES). Um dos objetivos do PNAES é a redução da evasão. O elo com os sistemas de informação se deu com o conhecimento adquirido inicialmente na graduação em Administração da Unifei e posteriormente expandido e consolidado no MPA. A possibilidade de emprego das ferramentas de tecnologia da informação, discutidas teoricamente nas salas de aula, em uma situação real de apoio a tomada de decisão em um tema relevante como a evasão institucional foi o principal motivo do desenvolvimento desse trabalho.

Após essa breve apresentação, o tema e o problema de pesquisa são apresentados e formulados. Os objetivos de pesquisa, tanto o geral quando os específicos, são traçados.

1.1 APRESENTAÇÃO DO TEMA

A qualificação e formação de seus discentes deve ser preocupação de qualquer instituição de ensino superior (IES), em especial as públicas (MEC, 1997). Para uma avaliação objetiva do resultado da função formativa das IES é necessário conhecer as taxas de sucesso ou diplomação, que se caracteriza pela razão entre concluídos e ingressantes, a taxa de retenção, que representa o total de alunos matriculados que ultrapassaram o prazo normal de integralização, e a taxa de evasão, que representa o total de alunos que saíram de seus cursos sem se diplomarem (MEC, 1997). O acompanhamento e avaliação dessas métricas são indispensáveis para orientar

políticas institucionais para melhoria do ensino de graduação (MEC, 1997) e podem indicar a necessidade de aperfeiçoamento institucional e correção de eventuais disfunções (TCU, 2009).

No contexto da avaliação do desempenho de IES, a evasão é entendida como o fenômeno de maior relevância em termos de preocupação institucional e requer acompanhamento para mitigar sua ocorrência, visto que a evasão gera desperdícios sociais, acadêmicos e econômicos (SILVA FILHO *et al.*, 2007) com perdas pessoais e institucionais (CISLAGHI; LUZ FILHO, 2009). Para a IES, a perda de alunos gera impacto negativo nas suas taxas de sucesso, ociosidade em termos de espaço físico e de subutilização dos profissionais da instituição (TONTINI; WALTER, 2014) e representa perda de investimentos públicos, que em 2014 superaram a marca dos R\$ 20.000,00 por aluno (INEP, 2018). Para os discentes a saída prematura da universidade pode ocasionar perdas de oportunidades de trabalho, de crescimento pessoal e de aumento de renda (TONTINI; WALTER, 2014).

De acordo com Silva Filho *et al.* (2007) e Vanz *et al.* (2016), o acompanhamento do fenômeno da evasão deve ser sistêmico, estar sob a responsabilidade da administração da IES e ser capaz de refletir a situação da evasão no tempo (PALHARINI, 2010). O conhecimento gerado por meio do acompanhamento sistêmico pode nortear a exploração do fenômeno, por meio de sua quantificação numérica ou cálculo de taxas, pela caracterização dos evadidos e identificação de fatores que contribuem para sua ocorrência ou pela explicação dos motivos ou causas da evasão (FREITAS, 2016; MUNIZAGA; CIFUENTES; BELTRAN, 2018). Consequentemente, gestores podem usá-lo de forma racional e com embasamento institucional para nortear sua tomada de decisão e ação institucional (MALHOTRA, 2001; ZIKMUND *et al.*, 2000)

No contexto da evasão, o apoio à decisão com base em informações institucionais pode se materializar nas seguintes situações: (1) ao conhecer as taxas de evasão de seus cursos, IES podem compará-las com outras instituições¹ (MASSI; VILLANI, 2015; VANZ *et al.*, 2016) e internamente entre seus cursos para identificar “ilhas de sucesso” ou, ao contrário, identificar cursos ou áreas do conhecimento com maior demanda de ação institucional (MEC, 1997); (2) a caracterização dos alunos

¹ Referenciais comparativos das taxas de evasão podem ser encontrados em Silva Filho *et al.*, (2007) e Silva Filho (2017).

evadidos permite descreve-los por meio de suas características pessoais (NODARI; LIMA; MACIEL, 2018; VANZ *et al.*, 2016; SILVA *et al.*, 2012), de suas características acadêmicas (NODARI; LIMA; MACIEL, 2018) e podem embasar a descoberta de fatores associados à evasão (VANZ *et al.*, 2016; TONTINI; WALTER, 2014; SILVA *et al.*, 2012; NODARI; LIMA; MACIEL, 2018), caso essas características sejam relacionadas estatisticamente com o fenômeno (KIPNIS, 2000; DURSO; CUNHA, 2018) e (3) a explicação causal do fenômeno, embasado em modelos² teóricos e nos fatores identificados como contribuintes para sua ocorrência, permite que tomadores de decisão tenham maior controle sobre suas intervenções (ZIKMUND *et al.*, 2000).

Para a operacionalização desses usos da informação sobre evasão é necessário que se tenha, primeiramente, uma fonte geradora de dados e que esses dados estejam disponíveis para consulta pelos gestores (BARBIERI, 2011; KIMBALL *et al.*, 2008). De forma complementar, é necessário definir qual a visão institucional a organização tem sobre seus dados (KIMBALL *et al.*, 2008), ao definir quais alunos deverão ser considerados como evadidos, qual métrica será usada para os cálculos das taxas de evasão e se os dados são contemporâneos à pesquisa ou de algum momento passado (PALHARINI, 2010).

Parte da solução para o problema da disponibilidade e abrangência dos dados está em utilizar dados que já foram gerados e estão armazenados nos sistemas de informação (SI) institucionais para adquirir as informações de interesse (DURSO; CUNHA, 2018; NODARI; LIMA; MACIEL, 2018; VANZ *et al.*, 2016). Sistemas de informação são softwares desenvolvidos para coleta, recuperação, processamento, armazenamento e distribuição de informação (LAUDON; LAUDON, 2014). Um tipo específico de SI utilizado para a automação de processos operacionais e cotidianos são os sistemas de processamento de transações (SPT). Transações são eventos que geram ou alteram dados que são posteriormente armazenados no BD do sistema (LAUDON; LAUDON, 2014). Esses sistemas realizam e registram transações necessárias ao funcionamento das atividades básicas organizacionais e são desenvolvidos para sempre terem dados mais atuais possíveis dos processos que automatiza (LAUDON; LAUDON, 2014).

² Nas pesquisas explicativas sobre evasão, destaca-se o modelo interacional de Vicent Tinto (TINTO, 1975). Esse modelo é tido como quase paradigmático e deu margem a uma série de trabalhos que buscaram sua validação (KIPNIS, 2000). O modelo de Tinto (1975) está centrado na interação do indivíduo no ambiente acadêmico e social das instituições que, não sendo suficiente, levam à evasão.

A utilização de sistemas de informação específicos para a área acadêmica é realidade na Unifei e em outras 19 Instituições Federais de Ensino Superior (IFES) com a adoção do Sistema Integrado de Gestão de Atividades Acadêmicas (SIGAA) (CANUTO, 2018). O SIGAA, que foi desenvolvido pela Universidade Federal do Rio Grande do Norte (UFRN), tem características transacionais e é responsável por informatizar procedimentos acadêmicos por meio de módulos como o Módulo Graduação. Esse módulo é responsável por gerar e armazenar dados pessoais dos alunos, vinculação de alunos em cursos de graduação, cadastro e oferta de disciplinas, matrículas em disciplinas, desempenho acadêmico, lançamento de notas e frequência nas aulas (SINFO, 2019).

Contudo, SPT utilizados para automação de processos operacionais possuem características que limitam sua capacidade de apoio à decisão, como a volatilidade e a disponibilidade restrita de seus dados. A volatilidade é definida como a mudança nos valores dos dados armazenados no tempo, quando um dado é reescrito (INSTITUTE FOR TELECOMMUNICATION SCIENCES, 1996). Essa constante atualização dos dados nos bancos dos SPT não permite a comparação da informação em momentos diferentes no tempo. A disponibilidade dos dados pode ser limitada uma vez que eles podem estar presos nas camadas de armazenamento dos SPT, sem acesso por parte dos interessados em tratá-los (BARBIERI, 2011; LAUDON; LAUDON, 2014; KIMBALL *et al.*, 2008).

Laudon e Laudon (2014) afirmam que um único sistema de informação não é capaz de ser utilizado no apoio as operações de uma organização e para sistematizar o apoio à decisão; são necessários sistemas dedicados a cada função. Sistemas que são desenvolvidos especificamente para o apoio à decisão são os sistemas de inteligência empresarial (BARBIERI, 2011; LAUDON; LAUDON, 2014; KIMBALL *et al.*, 2008). Esses sistemas têm como componente principal de sua arquitetura de dados um repositório de armazenagem que organiza, consolida, centraliza e torna disponível dados e informações coletados dos sistemas transacionais da organização (BARBIERI, 2011; LAUDON; LAUDON, 2014; KIMBALL *et al.*, 2008). Esse repositório tem o nome de *data warehouse* (DW), ou armazém de dados.

Um DW é um sistema desenvolvido para armazenar dados históricos e é otimizado para consultas (LAUDON; LAUDON, 2014; SINGH, 2001). Esse sistema busca por atualizações na sua fonte de dados e armazena as mudanças quando

ocorrem (KIMBALL *et al.*, 2008). Novos dados são adicionados ao DW e não sobrescrevem os dados anteriores. Assim, enquanto o sistema operacional terá apenas o registro mais recente, o DW contará com os dois registros. Ademais, o DW pode receber novos dados, periodicamente, o que permite acompanhar a evolução de métricas no tempo (KIMBALL *et al.*, 2008; SINGH, 2001).

Considerando o contexto discutido, formulou-se a premissa que o desenvolvimento e implementação de um DW é uma ferramenta que habilita a instituição responder à demanda pelo acompanhamento sistêmico do fenômeno da evasão. Sua arquitetura de dados pode ser modelada para armazenar as taxas de evasão institucionais e características pessoais e acadêmicas dos alunos. Os dados armazenados no DW se tornariam não voláteis ao se utilizar como fonte uma cópia dos dados já gerados e armazenados nos bancos de dados dos sistemas de informação transacionais. A programação de novas importações periódicas de dados no DW habilitaria o acompanhamento da evolução histórica dos dados.

As definições institucionais sobre quais alunos considerar como evadidos e quais métodos utilizar para o cálculo das taxas geraria uma versão institucional para os fatos retratados pelos dados, compartilhada por todos os interessados. Assim, gestores e pesquisadores poderiam acessar os dados institucionais armazenados nesse repositório e utilizá-los no apoio ao seu processo de tomada de decisão, tanto na orientação de políticas para a melhoria ou aperfeiçoamento do ensino de graduação como para a correção de disfunções.

1.2 PROBLEMA DE PESQUISA

Um DW é desenvolvido para responder perguntas relevantes para a ação institucional no contexto de um processo específico. Segundo Kimball *et al.* (2008), DW é um sistema onde dados são extraídos de suas fontes organizacionais e reestruturados para atender uma demanda por informação específica. Para o desenvolvimento do DW é necessário delimitar a demanda sobre um tema institucional e levantar quais informações são capazes de atendê-la (BARBIERI, 2011; KIMBALL *et al.*, 2008, SINGH, 2001). Assim, de um ponto de vista prático para o desenvolvimento dessa ferramenta no âmbito da Unifei, é necessário entender e definir quais são as demandas por informação relevantes ao contexto da evasão. Esse

estado desejado deve ser confrontado com a realidade dos dados, para identificar quais estão disponíveis para uso nos bancos de dados institucionais. Dessa forma a viabilização da construção do DW é dependente dos dados que estão da interseção entre o desejável e o viável. Assim, a questão desta pesquisa é definida da seguinte forma: como os dados referentes aos discentes armazenados no SIGAA podem ser organizados em um repositório institucional que sirva de apoio à tomada de decisão sobre o fenômeno da evasão? Dentro dessa pergunta abrangente, uma avaliação sobre a disponibilidade dos dados e a viabilidade do desenvolvimento de um DW também são abordados na pesquisa.

1.3 DOS OBJETIVOS DA PESQUISA

O objetivo principal é o desenvolvimento e implementação de um DW para uso no apoio à tomada de decisão sobre a evasão na Unifei. O trabalho visa explicitar quais são os dados de interesse, quais as fontes destes dados na Unifei, como esses dados podem ser extraídos, transformados e agregados na métricas de evasão e quais os dados descritivos associados aos discentes podem compor o repositório.

Os objetivos secundários são, primeiramente, conhecer quais são os métodos de desenvolvimento propostos na literatura especializada sobre DW para a escolha do que mais se adequa ao caso da Unifei. Segundo, é necessário conhecer quais são os métodos para cálculo das taxas de evasão e, terceiro, conhecer sob quais características os discentes foram descritos na literatura sobre evasão.

1.4 ESTRUTURA DA DISSERTAÇÃO

Após esta introdução, o trabalho é dividido em outras quatro seções. Na segunda seção tem-se a fundamentação teórica sobre os métodos de desenvolvimento de DW; a terceira seção apresenta a metodologia do trabalho; a quarta seção é dedicada aos resultados e discussão e, finalmente, a quinta seção apresenta as conclusões do trabalho e sugestões de futuras pesquisas.

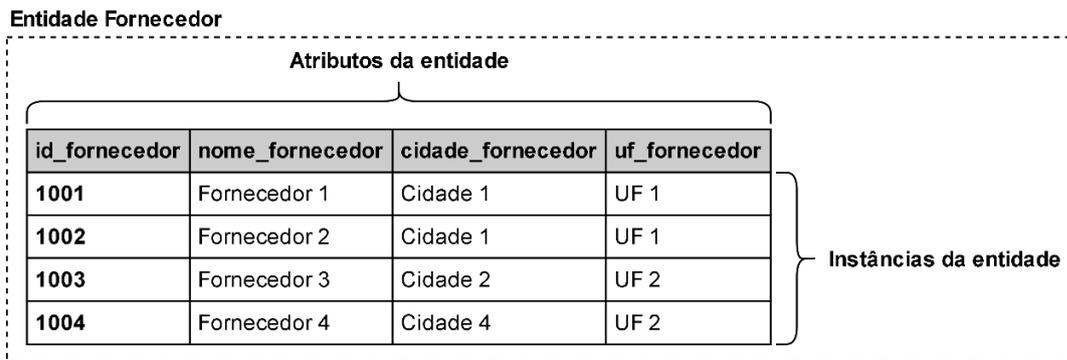
2 REVISÃO TEÓRICA

A revisão teórica é dividida em duas subseções. A primeira subseção é focada unicamente em definir os conceitos sobre bancos de dados que foram adotados no desenvolvimento do trabalho. A segunda subseção, mais ampla, discute os métodos de desenvolvimento de DW, explicitando suas etapas. Essa subseção foi utilizada para determinar os conceitos específicos sobre DW e nortear a definição de qual método escolher para a intervenção prática do trabalho.

2.1 CONCEITOS SOBRE BANCOS DE DADOS

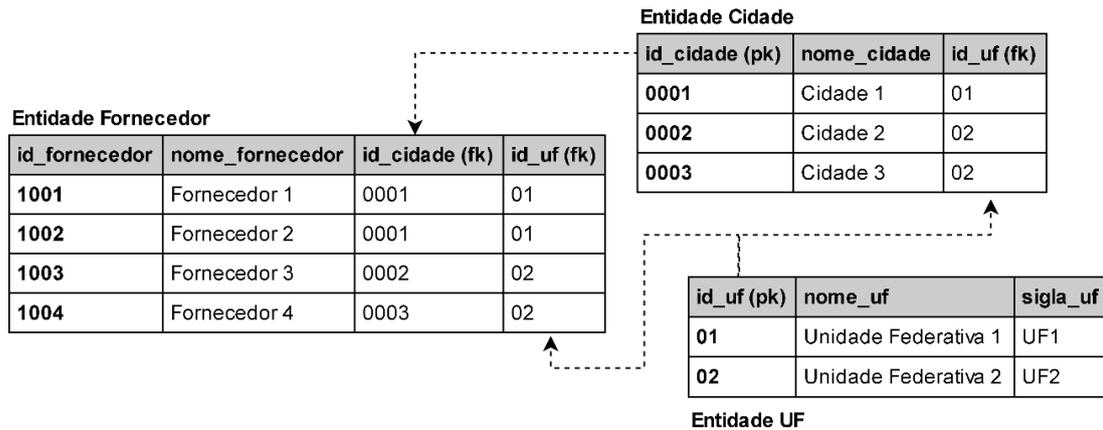
Sistemas de processamento de transações (SPT) têm como característica registrar transações rotineiras necessárias à realidade operacional de organizações, sendo transações definidas como eventos que geram ou alteram dados que são posteriormente armazenados pelo sistema (LAUDON; LAUDON, 2014). Os principais objetivos desses sistemas, segundo Laudon e Laudon (2014), são: (1) responder perguntas de rotina e (2) monitorar o fluxo de transações dentro da organização. Para tal, os SPT utilizam BD para armazenar e organizar os dados produzidos pelas suas transações, mantendo, assim, o suporte as operações do sistema (LAUDON; LAUDON, 2014).

Usualmente, os dados gerados por um SPT são armazenados em tabelas que representam grupos genéricos como pessoas, lugares e equipamentos. No contexto dos BD, esses grupos são chamados de entidades. A premissa básica do design de um BD é que cada tabela em um DB deve servir apenas um propósito, ou seja, armazenar dados de uma entidade (HEUSER, 2009; LAUDON; LAUDON, 2014). Cada linha dessa tabela representa uma ocorrência do elemento deste grupo, ou uma instância da entidade (HEUSER, 2009; LAUDON; LAUDON, 2014). Outra característica das tabelas de um BD está relacionada às suas colunas, ou atributos. Os atributos podem ser definidos como características dos integrantes do grupo (HEUSER, 2009; LAUDON; LAUDON, 2014). Assim, entidades, atributos e as suas instâncias são os elementos básicos de organização das informações em um BD. A Figura 1 representa graficamente exemplos dados pelos autores para uma entidade, seus atributos e suas instâncias.

Figura 1 – Exemplo de entidade, atributo e instâncias de uma entidade

Fonte: Autor, adaptado de Laudon e Laudon (2014) e Heuser (2009).

Bancos de dados usualmente possuem diversas tabelas e uma forma utilizada para organizar dados compartilhados entre essas tabelas é relacioná-las (HEUSER, 2009; LAUDON; LAUDON, 2014). Os relacionamentos entre tabelas estabelecem uma conexão entre seus dados usando chaves, ou *keys*. Essas chaves são identificadores numéricos e únicos de cada instância das tabelas do BD (HEUSER, 2009; LAUDON; LAUDON, 2014). O relacionamento é feito quanto uma chave (chave primária ou *primary key - pk*) é referenciada em outra entidade (nesse caso, tornando-se uma chave estrangeira ou *foreign key - fk*). Os exemplos dos autores (HEUSER, 2009; LAUDON; LAUDON, 2014) foram resumidos graficamente na Figura 2. No exemplo, é possível ver que as chaves primárias das entidades Cidade e UF, respectivamente “id_cidade (pk)” e “id_uf (pk)” foram referenciadas na entidade Fornecedor. A utilização prática do relacionamento entre entidades é a possibilidade de concatenar as tabelas de interesse em uma única, o que facilita organizar e consultar aos dados armazenados nesse conjunto de tabelas. A Figura 3 resume um exemplo dos autores (HEUSER, 2009; LAUDON; LAUDON, 2014) sobre a concatenação derivada do relacionamento entre tabelas.

Figura 2 – Relacionamento entre entidades e normalização

Fonte: Autor, adaptado de Laudon e Laudon (2014) e Heuser (2009).

Figura 3 – Concatenação das tabelas relacionadas entre si

Tabela concatenada: entidades Fornecedor, Cidade e UF

id_fornecedor	nome_fornecedor	id_cidade	nome_cidade	id_uf	nome_uf	sigla_uf
1001	Fornecedor 1	0001	Cidade 1	01	Unidade Federativa 1	UF1
1002	Fornecedor 2	0001	Cidade 1	01	Unidade Federativa 1	UF1
1003	Fornecedor 3	0002	Cidade 2	02	Unidade Federativa 2	UF2
1004	Fornecedor 4	0003	Cidade 3	02	Unidade Federativa 2	UF2

Fonte: Autor, adaptado de Laudon e Laudon (2014) e Heuser (2009).

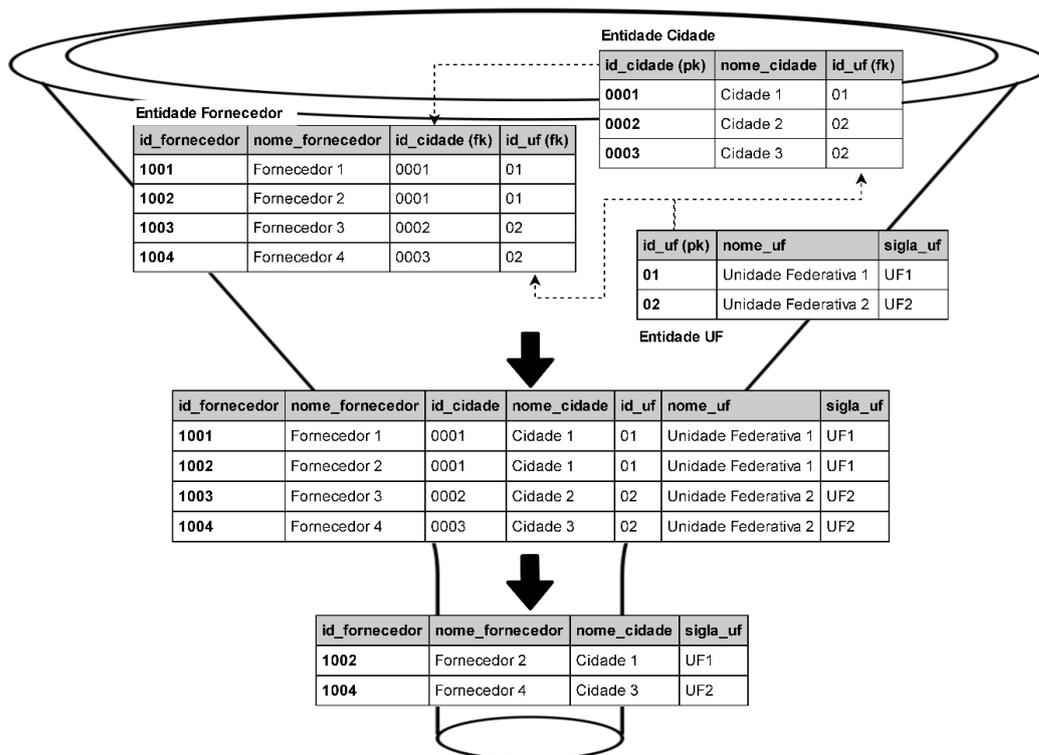
BD têm outra característica chamada de normalização. Normalização é o processo que visa minimizar elementos de dados redundantes e relações entre tabelas inadequadas (HEUSER, 2009). Esse processo tem como ponto positivo a facilidade de manutenção do BD e a baixa redundância de dados (LAUDON; LAUDON, 2014). Contudo, a normalização exige esforço de design adicional na criação de tabelas com as entidades únicas e relacionar essas entidades com todas as outras que utilizem dados armazenados nela. Um exemplo de normalização é derivado dos autores (HEUSER, 2009; LAUDON; LAUDON, 2014) e graficamente representado na Figura 2. Comparando as Figuras 1 e 2 é possível perceber que a entidade Fornecedor não mais armazena o nome da cidade e a sigla da uf de seus fornecedores, mas, sim, a chave numérica referenciando as entidades Cidade e UF.

BD possuem um software específico que é responsável por criar, armazenar, organizar e acessar entidades, atributos e seus respectivos dados, bem como criar os relacionamentos entre as entidades (HEUSER, 2009; LAUDON; LAUDON, 2014).

Esses softwares são chamados de *data base management systems*, ou sistema de gerenciamento de banco de dados (SGBD) (HEUSER, 2009; LAUDON; LAUDON, 2014). Os SGBD utilizam a linguagem *Structured Query Language* ou (SQL) para realizar suas operações de inserção, atualização e consulta dos dados (LAUDON; LAUDON, 2014).

Uma operação particularmente relevante no contexto do desenvolvimento de uma ferramenta de DW são as consultas, que permite a extração de dados do BD (KIMBALL *et al.*, 2008). Na realização de uma consulta, é possível instruir o BD, por meio de linguagem SQL, a organizar as entidades relacionadas entre si e extrair os dados de interesse armazenados em suas tabelas para utilização em outras aplicações (LAUDON; LAUDON, 2014; KIMBALL *et al.*, 2008). A Figura 4 representa, na forma esquemática de um funil, uma consulta padrão em SQL.

Figura 4 – Representação esquemática de uma consulta em SQL



Fonte: autor, adaptado de Laudon e Laudon (2014) e Heuser (2009).

Resumindo as contribuições das fontes consultadas (HEUSER, 2009; LAUDON; LAUDON, 2014; POSTGRESQL, 2019) é possível entender que uma instrução de consulta em SQL acessa uma entidade do DB, por exemplo a entidade

Fornecedores, por meio de uma instrução *FROM*. Instruções *JOIN* são utilizadas para concatenar outras entidades relacionadas, como as entidades Cidade e UF. Uma instrução *SELECT* é utilizada para selecionar apenas os atributos de interesse, como “id_fornecedor”, “nome_fornecedor”, “nome_cidade” e “sigla_uf”. Finalmente, instruções *WHERE* são utilizadas como filtros, para selecionar apenas instâncias de interesse, com os fornecedores 2 e 4. Assim, na entrada do funil temos dados dispersos em tabelas distintas e, após o processamento, na saída do funil, temos como resultado o conjunto de dados de interesse.

2.2 DATA WAREHOUSES

As características relacionais e de alteração de dados de um BD convencional tem implicações sobre o desenvolvimento de um DW (LAUDON; LAUDON, 2014; SINGH, 2001; KIMBALL *et al.*, 2008). Sob o aspecto estrutural, um BD é composto do relacionamento entre diversas tabelas e os dados de interesse para o DW podem estar dispersos em suas diversas tabelas (KIMBALL *et al.*, 2008). Sobre a questão de como estes dados são armazenados, os dados gerados pelas transações operacionais atualizam os registros anteriores nas tabelas, ação que penaliza análise histórica por não se ter dados referentes a diferentes momentos (LAUDON; LAUDON, 2014; SINGH, 2001; KIMBALL *et al.*, 2008).

Esses problemas decorrem da natureza do uso destes BD. O uso transacional suporta sistemas pensados para ações operacionais e rotineiras, onde é necessário garantir que o sistema tenha os dados atualizados (LAUDON; LAUDON, 2014). Por outro lado, a demanda por apoio à decisão gerencial tem relação com sistemas de inteligência empresarial, que têm o DW como seu principal componente de armazenagem de dados (BARBIERI, 2011; LAUDON; LAUDON, 2014; SINGH, 2001; KIMBALL *et al.*, 2008). É importante ressaltar que é necessário ter apoio da gerência da organização para o sucesso do desenvolvimento e aplicação de um DW (ALJAWARNEH, 2016; KIMBALL *et al.*, 2008; ZEA; SAMPEDRO; MORA, 2016).

Singh (2001) e Kimball *et al.* (2008) afirmam que um DW é um processo de armazenagem de dados, em um único repositório, destinado ao suporte à decisão no contexto de um processo organizacional específico. Isso permite ao gestor consultas, em um único local, e posteriores análises de dados para apoio na busca por

informações relevantes às questões inerentes a esse processo (KIMBALL *et al.*, 2008). Um DW viabiliza capacidades analíticas relacionadas ao suporte à decisão pois: (1) ele se constitui de um repositório de dados, fisicamente separado do armazenamento dos SPT, que pode convergir diversas fontes de dados; (2) a coleta e organização dos dados é realizada sob demanda para a tomada de decisão do usuário; e (3) é estruturado para receber carga sistemática e periódica de novos dados, não sobrescrevendo os dados anteriormente registrados, o que habilita a capacidade de análises históricas (KIMBALL *et al.*, 2008; SINGH, 2001).

Devido a essas características, o desenvolvimento de um DW pode prover para a organização uma solução para a disponibilidade, resumo e extração de conhecimento dos dados grados pelos seus SPT (BARBIERI, 2011; LAUDON; LAUDON, 2014; SINGH, 2001; KIMBALL *et al.*, 2008).

2.2.1 Conceitos Sobre o Desenvolvimento de Data Warehouses

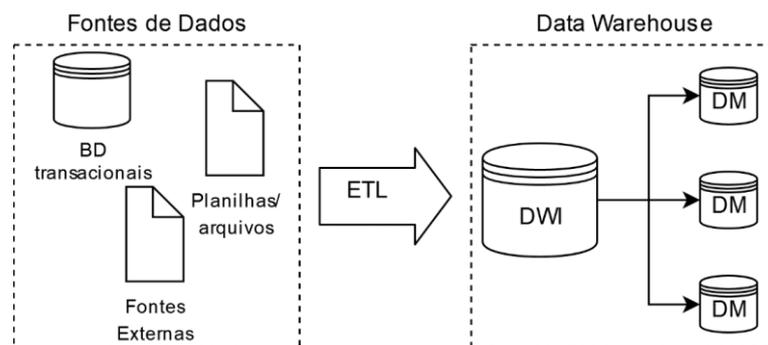
Uma solução de DW é dependente de uma infraestrutura de softwares e de uma arquitetura de dados interna para seu completo desenvolvimento (BARBIERI, 2011; KIMBALL *et al.*, 2008; SINGH, 2001). A infraestrutura de um DW tem relação com as ferramentas de software necessárias para materializar o DW e fazê-lo funcionar (KIMBALL *et al.*, 2008; SINGH, 2001). Exemplos derivados da literatura consultada mostra que são necessários um BD, ferramentas para a realização da extração, tratamento e transferência dos dados de suas fontes para o DW e um SGBD que gerencie transferências de dados, tanto a entrada no BD como a consulta dos dados armazenados (BARBIERI, 2011; KIMBALL *et al.*, 2008; SINGH, 2001).

Na escolha das ferramentas de software de infraestrutura, existem opções de código aberto que contribuem para a mitigação de custos operacionais relacionados com a compra de softwares, como o BD PostgreSQL, o SGBD pgAdmin e a linguagem de programação Python, que possui módulos para acessar e carregar dados no BD e trata-los, características úteis para a realização de extração, tratamento e carga de dados³ (CALUMBY; CRAVEIRO, 2008; POSTGRESQL, 2019; YUNG, 2018; WEINBERG, 2019).

³ O processo de extração dos dados de sua fonte, transformá-los e carregá-los no ambiente do DW é conhecido como *extract, transform and load* (ou ETL), conforme pode ser visto em Kimball *et al.* (2008).

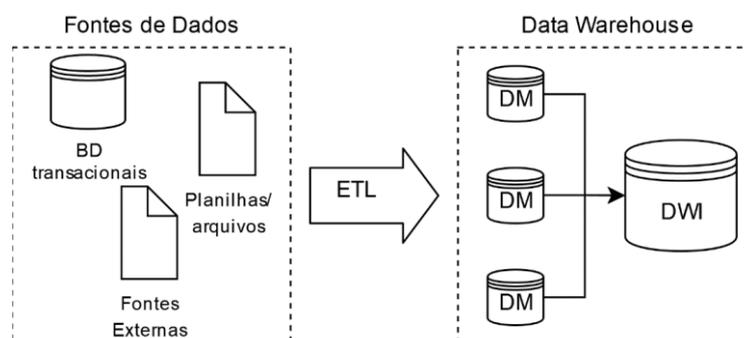
Nas decisões sobre a arquitetura do DW, acadêmicos da área (ARIYACHANDRA; WATSON, 2006, 2008, 2010; BRESLIN, 2004; JUKIC, 2006; RANGARAJAN, 2016; ZEA; SAMPEDRO; MORA, 2016) reconhecem duas principais metodologias, uma proposta por Bill Inmon (INMON, 2005) e outra por Ralph Kimball (KIMBALL *et al.*, 2008). Zea; Sampedro e Mora (2016) apontam diferenças entre as abordagens dos dois autores. A abordagem de Bill Inmon propõe a criação de um repositório institucional central (*data warehouse* institucional – DWI), que reúne todas as informações relevantes que é, posteriormente, dividido em repositórios menores, setorizados. Por sua vez, a abordagem trazida por Ralph Kimball propõe que os dados sejam consolidados em *data marts* (DM), que são versões menores do DW, que podem ser consolidados em um repositório institucional. As Figuras 5 e 6 representam graficamente essas abordagens.

Figura 5 - Abordagem proposta por Bill Inmon



Fonte: Adaptado de Zea; Sampedro e Mora (2016).

Figura 6 – Abordagem proposta por Ralph Kimball



Fonte: Adaptado de Zea; Sampedro e Mora (2016).

O foco da abordagem de Ralph Kimball está nos DM. DM são voltados principalmente para responderem a demandas de setores funcionais de uma organização e a confluência de diversos DM pode criar o DWI. Essa característica da abordagem de Ralph Kimball permite que o DW seja gradualmente desenvolvido e implementado em um período menor quando comparado à abordagem de Bill Inmon (ZEA; SAMPEDRO; MORA, 2016).

Outra questão de divergência entre as abordagens de Inmon e Kimball está no ciclo de desenvolvimento do DW. Enquanto Kimball *et al.* (2008) propõe que a iniciativa do DW deva começar com o levantamento dos requisitos de negócio e de dados necessários ao suporte à decisão do gestor e, daí, avaliar a capacidade da organização em fornecer os dados, Inmon (2005) defende que o primeiro passo deva ser a integração dos dados disponíveis em um ambiente de DW, compreendê-los e, somente após, determinar quais são os requisitos possíveis de serem atendidos pelos dados disponíveis. Contudo, como pode ser inferido da leitura dos autores (KIMBALL *et al.*, 2008; INMON, 2005), independente do momento e de sua fonte de informações, definir requisitos é uma etapa necessária para o desenvolvimento da solução.

Os requisitos de informação são necessários para definir a granularidade dos dados no DW, ou seu nível de detalhamento (KIMBALL *et al.*, 2008). Em um exemplo trazido por Kimball *et al.* (2008), temos dois níveis de informação possíveis para o monitoramento de vendas. É possível realizar essa análise observando os registros individuais de vendas dos produtos de interesse. Esse nível apresenta menor granularidade, apresentando dados atomizados, ou unidade indivisível de registro. Outra possibilidade é monitorar as vendas realizando agregações nos dados, como por exemplo contar quantos itens foram vendidos por produto e armazenar essa informação sumarizada, em contrapartida aos dados individuais do exemplo anterior. Para grandes aplicações essa decisão não envolve apenas os requisitos do usuário, mas a capacidade de armazenamento físico dos dados do DW (KIMBALL *et al.*, 2008; BRESLIN, 2004).

A premissa básica do desenvolvimento do DW na metodologia de Inmon é a criação de um DW composto de dados em seu menor nível de detalhamento. É esse repositório central, na metodologia de Inmon chamado de *atomic data warehouse*, ou armazém de dados atômico, que servirá os departamentos organizacionais com seus dados. Caso sejam necessárias agregações nos dados, como por exemplo somas ou

contagens de alguma variável de interesse, essa agregação é feita a nível departamental. Isso, aliado a determinação dos requisitos partindo-se dos dados, discutido anteriormente, permite classificar a abordagem de Inmon como *data-driven* ou orientada aos dados (BRESLIN, 2004). Em contrapartida, a abordagem de Kimball é focada em atender as demandas de um processo de negócio. Assim, a granularidade dos dados e suas agregações são definidos para atender os requisitos específicos dos gestores. Isso permite classificar a abordagem de Kimball como *process oriented*, ou orientada a processos (BRESLIN, 2004).

As definições dos requisitos, do foco nos dados ou nos processos e da granularidade que o DW terá, tem implicações na arquitetura do DW. A estrutura interna das tabelas do DW, tanto sua modelagem lógica⁴ quanto sua consequente materialização física no BD, diferem nas abordagens de Kimball e Inmon, como aponta Breslin (2004). A modelagem das tabelas segundo a metodologia de Inmon adapta a lógica dos BD convencionais às demandas de análise temporal de um DW. Primeiramente, no maior nível de abstração, estão os diagramas de entidade e relacionamento (ou *entity relation diagram*, ERD) que são utilizados para organizar as entidades e seus relacionamento, tanto para o DW atômico quanto para os departamentos que utilizarão os dados. O próximo passo é a modelagem dos *data item set* (DIS). Nessa etapa os relacionamentos estipulados no ERD são concretizados por meio de pares de dados e os atributos que cada entidade deve ter são definidos.

Em contrapartida, Kimball desenvolveu um método próprio para a modelagem dos dados no DW, a modelagem dimensional. A modelagem dimensional é fundamentalmente diferente das técnicas de modelagem relacionais dos BD convencionais e dividem o esforço da modelagem em duas frentes: entender quais são os fatos, usualmente métricas numéricas, que estão associadas ao processo modelado e em que contexto esses fatos se apresentam, ou em quais dimensões contextuais esses fatos podem ser descritos (BRESLIN, 2004; KIMBALL *et al.*, 2008). O modelo dimensional de Kimball *et al.* (2008) conta com tabelas dimensionais, que

⁴ Modelagem lógica e modelagem físicas são conceitos de desenvolvimento de BD. Os modelos lógicos mostram suas entidades e explicam de forma simplificada o relacionamento entre elas. Por sua vez, o modelo físico, que está estritamente ligado a efetiva criação das tabelas no banco de dados, mostra os detalhes do modelo lógico como os atributos das suas entidades (ALJAWARNEH, 2016; BARBIERI, 2011; INMON, 2005; KIMBALL *et al.*, 2008).

armazenam informações sobre atributos de determinada entidade e tabelas de fatos, que armazenam dados numéricos sobre a entidade. Em um exemplo trazido por Kimball *et al.* (2008), um DW relacionado com armazenagem de produtos, as tabelas dimensão podem conter dados sobre a descrição do item e seu respectivo fornecedor, enquanto a tabela de fatos tem dados sobre a quantidade disponível em estoque e o preço do item.

A modelagem física utiliza a modelagem lógica como guia para a efetiva criação de tabelas no BD e envolve decisões como quantas tabelas serão criadas, quais dimensões ou entidades serão representadas e quais atributos serão utilizados (ALJAWARNEH, 2016; INMON, 2005; KIMBALL *et al.*, 2008). A materialização das tabelas físicas no BD é crucial para o desenvolvimento do DW, uma vez que serve de base para o processo da ETL, que extrai dados de fontes diversas e os carrega nessas tabelas (KIMBALL *et al.*, 2008).

Outro ponto relevante discutido por Kimball *et al.* (2008) e Inmon (2005) são os metadados, ou dados sobre os dados. Metadados são informações que definem e descrevem toda a estrutura, operações e conteúdo do DW. É um esforço que auxilia a documentação tanto do projeto como da operacionalização do DW. Conforme exemplificado pelos autores (INMON, 2005; KIMBALL *et al.*, 2008), os metadados registram quais são as fontes de dados utilizadas, quais tabelas foram criadas no BD, quais suas colunas, formas de ETL, versões de software, qual o modelo dos dados, a periodicidade de novas tarefas de ETL e o histórico das ETL já realizadas.

2.2.2 Extract, Transform e Load: o Processo da ETL

O processo da ETL visa efetivamente extrair os dados das fontes escolhidas, tratá-los para que estejam em conformidade com as definições dos modelos lógicos e físicos, e carregá-los nas tabelas físicas do DW (KIMBALL *et al.*, 2008). A conformidade com entre as fontes de dados e as tabelas destino no DW é alcançada por meio da confluência de fontes heterogêneas de dados, transformando-os para uma forma de representação única em um único tipo de dado, como textual ou numérico (ALJAWARNEH, 2016; BIPLOB; SHERAJ; KHAN 2018; KIMBALL *et al.*, 2008).

Como pré-requisito para a realização da ETL é necessário mapear as fontes de dados que eventualmente serão utilizadas para popular o repositório (KIMBALL *et al.*, 2008; ZEA; SAMPEDRO; MORA, 2016), tanto as internas quanto as externas, e em quais formatos estão, como dados em BD institucionais, BD departamentais ou arquivos produzidos por usuários, como planilhas eletrônicas.

Kimball *et al.* (2008) apresenta uma diferenciação relevante que auxilia na escolha dessas fontes de dados ao separar fontes entre formais e informais. Fontes formais, segundo o autor (KIMBALL *et al.*, 2008), são mantidas pelo pessoal de TI da organização, conferindo integridade e confiabilidade aos dados. Fontes informais de dados são mantidas por colaboradores de órgãos funcionais da organização e podem causar problemas na integração ao DW uma vez que os registros feitos por esses funcionários podem não estar alinhado com a padronização de registros SPT (KIMBALL *et al.*, 2008; LAUDON; LAUDON, 2014).

Após extraídos, os dados podem ter dois destinos: (1) diretamente para as tabelas do DW ou (2) para um ambiente preliminar, de concentração de dados, conforme visto em Alwajarneh (2016) e Biplob, Sheraj e Khan (2018). Na proposta de Alwajarneh (2016), os dados nas fontes primárias são acessados, consolidados, transformados e carregados nas tabelas físicas em uma única operação. Contudo, Kimball *et al.* (2008) apontam problemas com essa abordagem. Eventuais faltas de energia elétrica ou problemas com as redes de comunicação podem comprometer o processo de carga, transferindo apenas parte dos dados ou inviabilizando qualquer transferência. Assim, Biplob, Sheraj e Khan (2018), Breslin (2004) e Kimball *et al.* (2008) sugerem que primeiramente os dados a serem transferido para o DW sejam movimentados para uma tabela de concentração. Posteriormente, os dados nessa área de concentração são acessados, transformados e carregados nas tabelas físicas do DW. De forma complementar, Kimball *et al.* (2008) sugere que as cargas de dados nessa tabela intermediária sejam mantidas como *backups* das movimentações realizadas para o DW. Isso habilita a possibilidade de retransmissão de dados para o DW em caso onde a carga realizada não seja completa. Ademais, independentemente do método escolhido para a movimentação de dados para o DW, é boa prática vista na literatura conferir a carga realizada no DW, certificando-se que todos os dados movimentados estejam no seu respectivo destino (KIMBALL *et al.*, 2008).

O processo de transformação visa garantir a conformidade dos dados extraídos com seu destino nas tabelas físicas do BD e tarefas usualmente associadas as transformações são (1) a adequação dos tipos de dados, que ocorre quando fontes de dados distintas atribuem nomes diferentes para um mesmo campo de dados (2) limpeza de dados, como por exemplo a remoção de dados em duplicidade e identificação de registros incompletos, (3) aplicação de agregações nos dados para obedecer a granularidade definida no projeto (ALJAWARNEH, 2016; BIPLOB, SHERAJI; KHAN, 2018; BRESLIN, 2004; KIMBALL *et al.*, 2008).

Para realizar o processo da ETL, é possível utilizar ferramentas prontas, disponíveis comercialmente, como por exemplo Informatica Power Center, IBM Data Stage, Microsoft SSIS, Pervasive Data Indicator, Talend Integration Suite, Pentaho Kettle e Clover ETL (BIPLOB; SHERAJI; KHAN, 2018), nas quais as conexões com as fontes de dados, a seleção dos dados de interesse, limpeza e carga são realizadas por meio de interfaces gráficas, ou desenvolver um processo de ETL na própria organização, o que exige a utilização de técnicas de programação e linguagens de código, como a linguagem SQL e Python (YUNG, 2018; WEINBERG, 2019).

Debroy; Brimble e Yost (2018) trazem uma discussão entre a escolha de comprar uma solução de mercado ou de se desenvolver uma solução própria. Os quesitos avaliados foram a habilidade de customizar o software da ETL, a habilidade do software ser flexível a mudanças nas fontes e tipos de dados, a integração com os sistemas já em uso na organização e a oferta de recursos além do necessário pelos softwares de ETL disponíveis. No caso prático relatado pelos autores (DEBROY; BRIMBLE; YOST, 2018), a escolha por desenvolver a solução própria se deu uma vez que a compra de uma solução pronta não é facilmente customizada, pois o desenvolvimento é feito pelo fornecedor do software, e os produtos podem oferecer recursos que vão além dos demandados pela organização.

A ETL tem dois momentos distintos de realização. Usualmente, quando do desenvolvimento da primeira iniciativa do DW, é necessário realizar uma ETL que abranja o maior período de disponibilidade possível dos dados. Essa é a ETL histórica (KIMBALL *et al.*, 2008), e conta com a transferência de todos os dados de interesse já produzidos pela organização de suas fontes para o DW. Contudo, à medida que novos dados são produzidos pelos sistemas da organização, novas tarefas de ETL devem ser realizadas para manter a periodicidade dos dados. Kimball *et al.* (2008)

sugere que a abordagem ideal está em se desenvolver um sistema que consiga capturar apenas os dados que sofreram alteração ou foram criados após a primeira ETL. Esses sistemas são chamados de *change data capture* (KIMBALL *et al.*, 2008).

3 METODOLOGIA

Martins e Theóphilo (2016) definem a metodologia como um modo de captar aspectos relevantes da realidade para responder a perguntas de pesquisa, traçando o caminho para atingir determinado objetivo ou fim. Assim, essa seção busca mostrar e delimitar qual o caminho traçado pelo autor para se alcançar os objetivos dessa dissertação. Primeiramente, o objeto de pesquisa é sucintamente caracterizado e posteriormente as decisões metodológicas são explícitas, em subseção própria.

O desenvolvimento da pesquisa se deu no âmbito da Pró-Reitoria de Graduação (PRG) da Universidade Federal de Itajubá (Unifei), onde o autor é servidor desde 2014, e se beneficiou da sinergia entre demandas organizacionais pelo acompanhamento da evasão e o conhecimento adquirido pelo autor por meio de sua participação no Programa de Mestrado Profissional em Administração (MPA), também da Unifei.

No contexto operacional da Unifei, existem impulsionadores para o desenvolvimento e a adoção de uma ferramenta que suporte à tomada de decisão sobre a evasão. Um deles é a atribuição normativa delegada pela Reitoria e o outro é o desempenho formativo da Instituição. A PRG tem a atribuição institucional de monitorar indicadores de desempenho dos cursos de graduação para subsidiar a proposição de políticas que visem a permanência do aluno no curso e a redução da evasão (UNIFEI, 2019e). Sobre seu desempenho formativo, no quadriênio de 2015 a 2018, o indicador⁵ ficou em média 30,5% abaixo da meta institucional de 80% (UNIFEI, 2014a, 2016, 2017, 2018, 2019b).

A Unifei foi fundada em 1913 sob a denominação de Instituto Eletrotécnico e Mecânico de Itajubá. A federalização ocorreu em 1956 e a transformação em universidade se deu em 24 de abril de 2002, por meio da Lei n.º 10.435. Hoje, a Unifei está presente em dois campi, em Itajubá-MG e em Itabira-MG, contando com 34 cursos presenciais de graduação e 1 à distância, e 16 cursos de pós-graduação. Do total de 8.184 alunos matriculados, a graduação presencial representa 7.156 alunos,

⁵ A métrica utilizada é a taxa de sucesso na graduação (TSG), indicador proposto pelo Tribunal de Contas da União (TCU). Essa taxa é calculada por meio da razão entre o total de diplomados e o total de ingressantes, respeitando o tempo regular de integralização de cada curso (TCU, 2009). Este indicador pode ser utilizado como uma aproximação da evasão ao subtrair o sucesso do total de alunos ($1 - \text{TSG}$), conforme proposto por Davok e Bernard (2016).

a graduação à distância conta com 230 alunos e a pós-graduação stricto sensu representa 798 alunos (UNIFEI, 2019f).

3.1 DECISÕES METODOLÓGICAS

No desenvolvimento da pesquisa, foi identificada a necessidade de se utilizar tanto uma metodologia geral, que guiasse a intervenção na situação problemática, como um método prático de desenvolvimento de DW.

A metodologia geral deveria ter duas características principais: (1) ser orientada à ação prática em uma situação problemática e (2) admitir a participação do pesquisador como ator da ação. Uma estratégia de pesquisa que se enquadra nesses requisitos é a pesquisa ação, inicialmente proposta por Kurt Lewin (MARTINS; THEÓPHILO, 2016). Ainda segundo os autores (MARTINS; THEÓPHILO, 2016), a pesquisa ação é uma investigação participante, com foco na ação em situações reais baseada em um ciclo, iterativo, que tem como etapas a definição de um fato-achado, concepção, planejamento, execução da ação e avaliação dos resultados da execução na resolução do problema inicial.

Contudo, desde sua criação, a pesquisa ação não caminhou para uma unificação teórica (DICKENS; WATKINS, 1999). Na realidade, as autoras (DICKENS; WATKINS, 1999) conceitualizam a pesquisa ação como um termo guarda-chuva para estratégias de pesquisa que visam fomentar a mudança em níveis de grupos, de organizações e da sociedade.

No contexto da presente pesquisa, a *soft systems methodology* (SSM), uma forma de pesquisa ação (COGHLAN; BRYDON-MILLER, 2014), foi escolhida como a metodologia geral que guiou a intervenção na situação problemática da Unifei. O motivo foi, principalmente, a diferenciação sutil, porém conceitualmente relevante, feita entre “problema” e “situação problemática” por parte de seus criadores (CHECKLAND; POULTER, 2010). Segundo Checkland e Poulter (2010), o enquadramento de uma situação como um “problema” denota a possibilidade de ser desenvolvida uma solução definitiva. Contudo, a realidade não é estática: o fluxo constante de interações humanas insere nas situações do cotidiano mudança constante (CHECKLAND; POULTER, 2010). Assim, uma situação deveria ser

percebida como “problemática”, não admitindo uma resolução definitiva, mas, sim, que esforços sejam desprendidos para a melhoria dessa situação.

Nesse contexto, como a pesquisa proposta tem por objetivo a proposta de melhoria na “situação problemática” de *apoio à decisão sobre evasão* e não a pretensão de resolução do problema do fenômeno da *evasão* na Unifei, houve alinhamento entre a proposta do autor e a SSM.

Sobre a escolha de um método de desenvolvimento de DW, como evidenciado na revisão da literatura, existem duas abordagens proeminentes, a de Bill Inmon (INMON, 2005) e a de Ralph Kimball (KIMBALL *et al.*, 2008). Para a finalidade dessa pesquisa, o método de Ralph Kimball foi o escolhido. Essa escolha se alinha com os resultados da pesquisa internacional conduzida por Zea, Gualtor e Mora (2018), que aponta a metodologia de Kimball como a abordagem mais utilizada em instituições de ensino superior.

Uma característica dessa abordagem que contribuiu para sua escolha é seu foco inicial em repositórios locais, construídos para determinado setor funcional, que podem ser, posteriormente, conectados em uma rede dependente para criar um DW de nível organizacional (KIMBALL *et al.*, 2008).

Outra característica está relacionada com a arquitetura de dados do DW. Como visto na revisão da literatura, a modelagem dimensional de Kimball *et al.* (2008) é menos complexa de ser desenvolvida em comparação à modelagem proposta por Inmon (2005), que utiliza a lógica de modelagem relacional normalizada entre as tabelas do DW. A abordagem de Kimball *et al.* (2008) não tem por conceito a normalização das tabelas que compõem sua arquitetura de dados, tornando a materialização física das tabelas do modelo lógico no BD também mais simples (KIMBALL, *et al.*, 2008; ARIYACHANDRA; WATSON, 2006, 2008, 2010). Essas características tornam o modelo de Ralph Kimball mais acessível para profissionais que não são nativos da área de TI e permite que o esforço inicial de desenvolvimento do DW e o tempo demandado sejam menores.

Nas próximas duas subseções são discutidos conceitos sobre a SSM e o método de Kimball demonstrando etapas de seus processos de utilização.

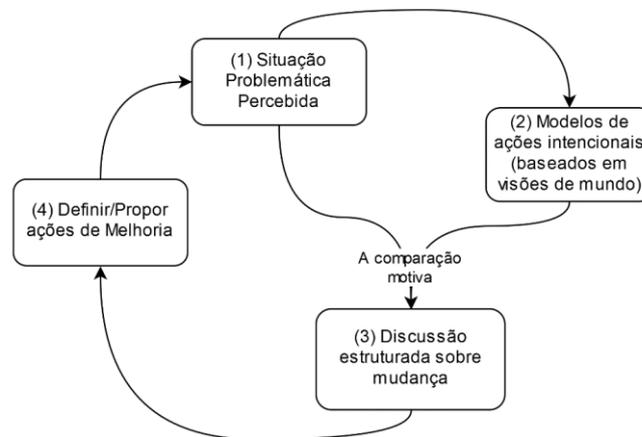
3.1.1 Soft Systems Methodology

A *Soft Systems Methodology* (SSM) é uma metodologia de pesquisa proposta por Peter Checkland na década de 1970 (CHECKLAND, 2000). É de natureza aplicada, orientada à ação, e procura gerar conhecimento para uso em situações práticas (CHECKLAND; POULTER, 2010; SILVA; MENEZES, 2005, GERHARDT; SILVEIRA, 2009). A SSM é derivada da abordagem sistêmica da realidade. (DEFINIR SISTEMAS). Contudo, ao invés de considerar que os componentes individuais do sistema podem sofrer um processo de engenharia para o alcance dos objetivos determinados para o sistema, a SSM defende que sistemas são instrumentos utilizados para nortear a aprendizagem em situações problemáticas da realidade.

A SSM é definida por Checkland e Poulter (2010) como um processo de investigação, sobre situações complexas e problemáticas, no qual os pesquisadores aprendem sobre as situações de forma organizada, permitindo que conhecimento seja gerado e usado para a melhoria da situação. Na visão dos autores (CHECKLAND; POULTER, 2010) as situações problemáticas possuem duas características fundamentais que devem ser reconhecidas para a aplicação da SSM: (1) sua complexidade advém do fato que elas contêm múltiplas percepções da realidade (ou visões de mundo), e (2) existem pessoas agindo deliberadamente para a melhoria da situação (ações intencionais).

Para gerar conhecimento de forma organizada e que conflua visões de mundo relevantes à situação, Checkland e Poulter (2010) sugerem um ciclo de aprendizagem constituído de quatro fases: (1) exploração da situação problemática percebida pelo pesquisador; (2) investigação de ações intencionais admitidas como relevantes à situação; (3) discussão sobre possíveis mudanças com base nos resultados da investigação das ações intencionais e a realidade da situação, e (4) definição das ações a serem implementadas para a melhoria da situação. O ciclo é representado de forma gráfica na Figura 7.

Figura 7 – Representação icônica do ciclo de aprendizagem da SSM



Fonte: Checkland e Poulter (2010). Adaptado pelo autor.

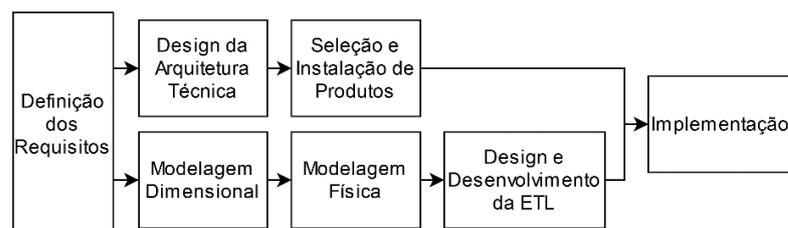
A aplicação desse ciclo permite a exploração da situação problemática e norteia a confluência de diferentes visões de mundo sobre o tema que são utilizados como fonte de debate frente a situação real. O resultado prático deste processo é a capacidade de estruturar a mudança na situação, visando sua melhoria, com base na interseção do que é desejado como mudança e o que é possível de ser mudado.

Sua aplicação foi observada em áreas do conhecimento como a gestão de projetos, a estratégia organizacional, a educação, a tomada de decisão e, principalmente, a tecnologia de informação (CHECKLAND; POULTER, 2010; MINGERS; TAYLOR, 1992), área de desenvolvimento da pesquisa.

3.1.2 Método de Ralph Kimball

A sequência lógica das etapas de desenvolvimento do DW propostas pelo autor é mostrada graficamente na Figura 8.

Figura 8 – Ciclo de vida do *data warehouse*



Fonte: adaptado de Kimball *et al.* (2008)

Como pode ser visto na Figura 8, a etapa de definição dos requisitos é uma etapa que guia tanto o caminho da infraestrutura arquitetural e de software como o caminho da modelagem dos dados que levam, ao final, a implementação do DW (KIMBALL *et al.*, 2008).

Compreender qual o problema organizacional relacionado com o escopo do DW e quais processos organizacionais serão apoiados pela iniciativa do DW são os requisitos de negócio do DW. Nessa etapa de levantamento de requisitos, o foco principal é determinar quais informações são tidas como relevantes aos tomadores de decisão que necessitam ser armazenadas no DW (ALJAWARNEH, 2016; KIMBALL *et al.*, 2008; ZEA; SAMPEDRO; MORA, 2016).

A coleta desses requisitos pode utilizar entrevista com o pessoal envolvido no processo decisório (ALJAWARNEH, 2016; KIMBALL *et al.*, 2008; ZEA; SAMPEDRO; MORA, 2016), desde usuários operacionais à diretoria (ALJAWARNEH, 2016). Adicionalmente, Aljawarneh (2016) propõe que a observação do pesquisador e a pesquisa documental sejam também utilizadas para o levantamento. Nessa etapa se determina a viabilidade da organização em prover fontes de dados a serem utilizadas na geração das informações do DW (KIMBALL *et al.*, 2008).

Em etapa posterior, os próximos requisitos a serem levantados são os requisitos de informação (KIMBALL *et al.*, 2008). Nessa etapa, conforme afirma Kimball *et al.* (2008), o objetivo é determinar se as fontes de dados da etapa anterior possuem dados suficientes para a geração das informações determinadas pelos requisitos de negócio. Para coletar esses requisitos é necessário investigar as fontes de dados na busca por compreender como esses dados são gerados, em quais tabelas são armazenados e como podem ser extraídos (KIMBALL *et al.*, 2008).

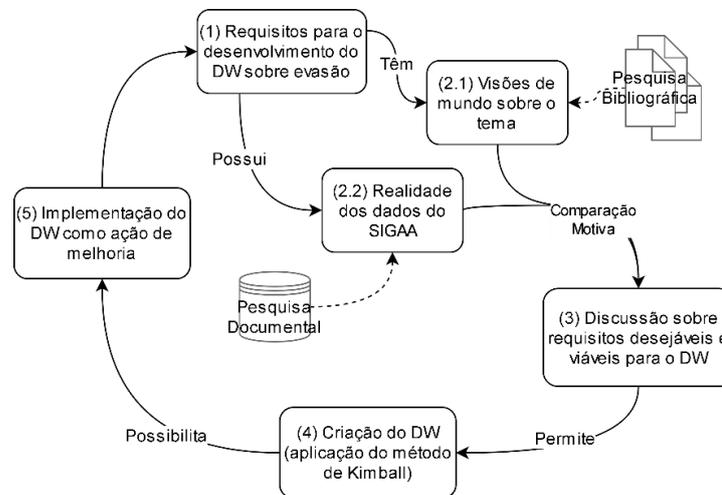
A próxima subseção é responsável por apresentar a proposta metodológica da dissertação que integra a SSM e o método de Kimball *et al.* (2008) e utiliza as pesquisas bibliográfica e documental como forma de coleta de dados e a triangulação de seus resultados como forma de análise de dados.

3.2 PROPOSTA METODOLÓGICA

Considerando o objetivo geral de desenvolver e aplicar um DW para apoio à tomada de decisão sobre evasão na Unifei, é necessário convergir a metodologia

geral que guiou a intervenção prática na situação real com o método de desenvolvimento de um DW. O método para o desenvolvimento do DW se torna uma etapa da metodologia geral, pois apenas o desenvolvimento da solução não é suficiente para a melhoria na situação problemática; é necessário que o DW seja inserido no contexto da tomada de decisão organizacional. Uma representação das etapas da proposta metodológica é feita na Figura 9. O ciclo proposto pode ser iterativo, caso sejam necessários ajustes à ferramenta posteriores à primeira implementação.

Figura 9 – Proposta metodológica baseada no ciclo de aprendizagem da SSM



Fonte: Proposta do Autor. Adaptado de Checkland e Poulter (2010)

Efetivamente definir quais são os requisitos para o desenvolvimento de um DW na situação prática da Unifei é o primeiro passo do modelo proposto (1), pois, como discutido na subseção anterior, a definição dos requisitos, tanto os de negócio quanto os de informação, é a primeira etapa para o desenvolvimento do DW. Assim, é necessário explorar a situação para definir qual o processo organizacional e quais são as informações necessárias para habilitar o DW a ser, efetivamente, ferramenta de apoio à decisão.

O segundo passo do modelo proposto foi dividido em duas etapas. A primeira etapa (2.1) tem o objetivo de investigar as visões de mundo a respeito de métodos para cálculo das taxas de evasão e de variáveis discentes usualmente associadas à ocorrência do fenômeno. Nessa etapa foi utilizada a pesquisa bibliográfica como ferramenta de coleta de dados. Os requisitos de informação para o cálculo das taxas

e as variáveis identificadas visam ser o estado desejado dos requisitos de informação para o DW. A segunda etapa (2.2) visou conhecer quais dados pessoais e acadêmicos dos alunos a Unifei tem, o que foi chamado de “realidade dos dados da Unifei”. Essa etapa utilizou de pesquisa documental no banco de dados do SIGAA da Unifei.

O terceiro passo (3), visou triangular os dados das etapas 2.1 e 2.1. Isso permitiu conhecer a interseção das informações desejáveis de se ter no DW com o que é possível ser atendido pelos dados do SIGAA. Isso permitiu o estabelecimento de um estado viável dos requisitos, no qual os dados institucionais disponíveis puderam ser modelados e armazenado no DW.

Com base nos requisitos de dados é possível aplicar o método de Kimball *et al.* (2008) para a criação do DW, o que envolve modelar os dados disponíveis no SIGAA, tanto o modelo dimensional como o modelo físico, e carregar os dados nessa estrutura. Essas ações configuram o quarto passo do modelo (4).

O quinto passo (5) é a implementação do DW na realidade operacional da Unifei, que é a ação de melhoria à tomada de decisão organizacional sobre a evasão.

A seguir são descritos os passos utilizados para o desenvolvimento da pesquisa bibliográfica e documental (etapas 1 e 2 do segundo passo), que gerarão insumos para a discussão dos requisitos do DW (passos 1 e 3). Os passos 4 e 5, que estão ligados com a criação e implementação do DW, serão discutidos na Seção de Resultados e Discussão.

3.2.1 Pesquisa Bibliográfica e Investigação de Ações Intencionais

A pesquisa bibliográfica foi escolhida como um dos métodos de investigação e coleta de dados. Ela tem a finalidade de ser fonte de visões de mundo contidas em ações intencionais sobre o fenômeno da evasão (CHECKLAND; POULTER, 2010; MARTINS; THEÓPHILO, 2016). Dentro do contexto da SSM, as pesquisas em evasão publicadas foram consideradas ações intencionais uma vez que os pesquisadores envolvidos procuraram agir intencionalmente na sua situação problema particular e produziram resultados relevantes.

Segundo Martins e Theóphilo (2016), a pesquisa bibliográfica pode ser utilizada como estratégia de pesquisa que busca conhecer, analisar e explicar contribuições vindas de publicações sobre o tema de interesse. O objetivo prático é a coleta de

informações para alcançar os objetivos secundários de conhecer os métodos e métricas utilizadas para os cálculos das taxas de evasão e conhecer as características dos discentes que foram relevantes nas pesquisas descritivas e causais sobre evasão. A abrangência da pesquisa bibliográfica teve as seguintes características, que visaram se aproximar da realidade da Unifei⁶: (1) se restringiu ao ensino superior público, (2) foco em cursos na modalidade presencial, (3) preferencialmente pesquisas aplicadas e (4) em modalidade presencial.

A busca foi realizada pelo termo “evasão” em periódicos da área de educação em novembro de 2018. Este termo foi utilizado pela comissão do MEC no estudo pioneiro no país (MEC, 1997). Ademais, em recente revisão sistemática, Munizaga, Cifuentes e Beltran (2018) utilizaram esse termo em sua pesquisa por sua relevância na língua portuguesa. O termo foi buscado no título, no resumo ou no índice, utilizando a plataforma *Scielo*. Os periódicos buscados são nacionais ou com relação com o Brasil, no intuito de compreender a realidade brasileira, sem prejuízo do artigo achado mencionar literatura internacional ou escrito em inglês por pesquisador brasileiro. Escolheram-se periódicos classificados com Qualis A1 (quadriênio 2013-2016). Periódicos A1 pertencem ao estrato mais elevado no sistema Qualis (CAPES, 2019), apresentando confiabilidade nos resultados apresentados e relevância ao tema. Uma lista contendo as principais fontes utilizadas está no Apêndice E.

Após a seleção das fontes, procedeu-se com a leitura na íntegra dos textos. A leitura tinha dois objetivos, a saber: (1) o artigo mostrava como as taxas de evasão foram calculadas? Se sim, qual o método de cálculo? (2) o artigo descrevia os discentes por meio de variáveis ou tinha como resultado a determinação de alguma variável como determinante ou contribuinte para a evasão dos alunos? Os resultados foram fichados e serviram de base para a comparação da pesquisa documental.

3.2.2 Pesquisa Documental e a Realidade dos Dados

Outra fonte de dados foi a análise documental. Essa análise tem como característica a utilização de documentos formais como fonte de dados (MARTINS;

⁶ Deu-se foco nas pesquisas voltadas para a modalidade de educação presencial pois esta é a modalidade predominante na Unifei. Ademais, as relações de evasão e permanência em cursos à distância podem ter características diferentes das presenciais pelas suas características peculiares.

THEÓPHILO, 2016). A pesquisa documental visou explorar o banco de dados do SIGAA da Unifei a fim de identificar os dados que o sistema registra. A pesquisa teve os objetivos de conhecer (1) quais as tabelas do SIGAA armazenam dados acadêmicos ou pessoais dos discentes e (2) dessas tabelas, quais são seus atributos, ou características, registradas nas tabelas. Essa análise permitiu conhecer a realidade dos dados gerados, armazenados e prontos para o uso na Unifei que sejam relevantes ao desenvolvimento do DW.

O SIGAA tem como SGBD o PostgreSQL. Esse gerenciador de banco de dados fornece ferramentas de código capazes de mostrar suas tabelas e relacionamentos por meio de códigos, como mostrado na Figura 10. O código exporta um arquivo em formato de dados separado por vírgulas, ou “csv”. A aplicação desse código e análise do arquivo “csv” permite explorar e conhecer quais são as tabelas, ou entidades, e seus atributos, ou colunas.

Figura 10 – Código para reaver nome de tabelas no PostgreSQL

```
1  
2 select  
3     table_catalog,  
4     table_schema,  
5     table_name,  
6     column_name,  
7     ordinal_position  
8 from information_schema.columns  
9
```

Fonte: elaborado pelo autor.

3.2.3 Triangulação dos Dados e Requisitos do Data Warehouse

Como método de análise dos dados, utilizou-se, para comparar as fontes de informações bibliográfica e documental a triangulação dos resultados encontrados. A Triangulação de fontes de dados visa buscar confluência nos dados de fontes distintas (MARTINS; THEÓPHILO, 2016). A discussão comparativa entre as visões de mundo das fontes bibliográficas e a realidade dos dados da Unifei não tem como objetivo principal a adequação da realidade às visões de mundo, mas sim o uso do resultado da pesquisa bibliográfica como fonte de estruturação e discussão sobre o que é possível de ser implementado na Unifei (CHECKLAND; POULTER, 2010).

A integração com o método de desenvolvimento do DW começa nessa fase: a interseção das pesquisas bibliográfica e documental mostrou quais são os dados

possíveis de serem modelados e armazenados no DW, ou seja, quais são os requisitos de informação do DW.

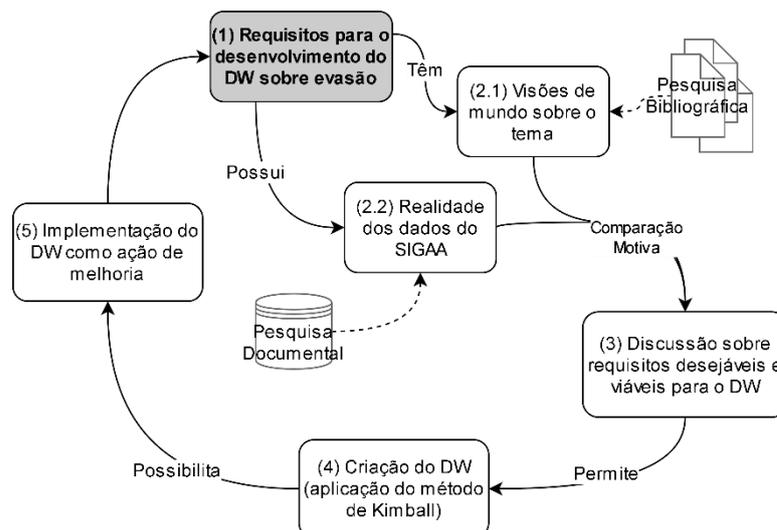
4 RESULTADOS E DISCUSSÃO

Nesta seção os resultados da aplicação da proposta metodológica são demonstrados e discutidos.

4.1 REQUISITOS GERAIS PARA O DESENVOLVIMENTO DO DW

Essa subseção trata das definições preliminares sobre os requisitos gerais do DW, ou seja, qual a finalidade institucional de desenvolvê-lo. Na metodologia proposta, essa discussão está contida no primeiro passo, representado na Figura 11.

Figura 11 – Primeiro Passo de Aplicação da Metodologia Proposta



Fonte: Proposta do Autor. Adaptado de Checkland e Poulter (2010)

A implementação do DW se insere no contexto do apoio à tomada de decisão na Unifei. Essa demanda é derivada da atribuição institucional que a PRG tem em monitorar os indicadores de desempenho dos cursos de graduação da Unifei para fomentar a permanência dos discentes na Unifei e reduzir a evasão (UNIFEI, 2019e).

Para efetivamente viabilizar o apoio à tomada de decisão sobre a evasão, o DW deve possuir características específicas, conforme discutido na seção introdutória. Essas características são: (1) viabilizar o acompanhamento sistêmico da evasão (SILVA FILHO *et al.*, 2007) com dados não-voláteis que permitam visualizar a evolução histórica do fenômeno; (2) deve ser de responsabilidade institucional (VANZ

et al., 2016); (3) deve quantificar numericamente as taxas de evasão dos cursos e permitir a caracterização dos evadidos por meio de características pessoais e acadêmicas (FREITAS, 2016; MUNIZAGA; CIFUENTES; BELTRAN, 2018), embasados em definições institucionais dos cálculos das taxas (FREITAS, 2016; PALHARINI, 2010); (4) os dados devem estar disponíveis aos gestores (BARBIERI, 2011; KIMBALL *et al.*, 2008).

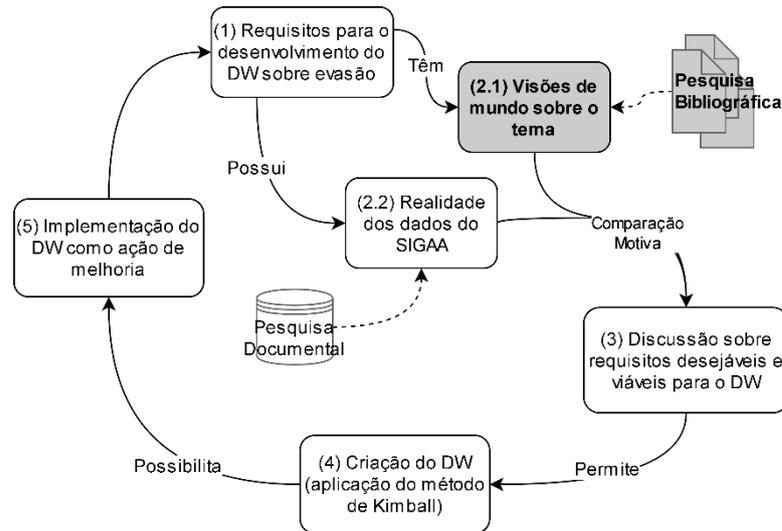
O desenvolvimento de um DW pela PRG, com cargas periódicas, em datas estabelecidas no calendário acadêmico, dos dados de interesse, atende aos dois primeiros requisitos gerais. Assim, o acompanhamento da evasão tem capacidade de se tornar sistêmico e está sob a responsabilidade do órgão que gere a graduação.

Para quantificar numericamente as taxas de evasão, a instituição deve ter sua visão sobre o tema. Isso implica em definir conceitualmente qual o método de cálculo deve ser empregado, quais alunos devem ser considerados nos cálculos e qual a fonte desses dados, ou seja, como os dados gerados pelo SIGAA podem ser organizados para o cálculo das taxas (FREITAS, 2016; PALHARINI, 2010; KIMBALL *et al.*, 2008). Raciocínio análogo é feito para conhecer quais características descritivas sobre os alunos o DW deve registrar. Para definir qual a taxa de evasão e quais variáveis pessoais e acadêmicas utilizar, recorreu-se a pesquisa bibliográfica, respectivamente discutidas nas Subseção 4.2 e Subseção 4.3. E para analisar a viabilidade interna em oferecer tais dados pesquisa documental foi realizada e está descrita na Subseção 4.4. A interseção entre o desejável e o viável defini os requisitos de informação para o DW. Essa discussão é realizada na Subseção 4.5.

4.2 MÉTODOS DE CÁLCULO DAS TAXAS DE EVASÃO

Os resultados apresentados nessa subseção e na subseção seguinte são derivados de pesquisa bibliográfica e compõem a primeira etapa do segundo passo da metodologia proposta, graficamente representada na Figura 12. O foco dessa subseção está na apresentação dos métodos de cálculo para as taxas de evasão encontrados na literatura, discussão sobre as limitações percebidas e a proposta de um método criado a partir da discussão teórica que visou preencher as lacunas encontradas.

Figura 12 – Primeira etapa do segundo passo de aplicação da metodologia proposta



Fonte: Proposta do Autor. Adaptado de Checkland e Poulter (2010)

Foi observado que o cálculo das taxas de evasão pode já ser sistematizado pela instituição e ser publicado periodicamente em seu relatório de gestão (TONTINI; WALTER, 2014), ou por meio de sistema de informação integrado (MENDES JÚNIOR, 2014). Assim, caso já exista determinado setor que acompanhe a evolução das taxas de evasão sistematicamente, é possível utilizar essa fonte de dados para o DW.

Em outros casos, no entanto, é necessário que o interessado some ou conte dados individuais dos alunos para calcular as taxas de evasão. As fontes de dados usualmente utilizadas são os sistemas de informação da instituição ou os históricos escolares dos alunos (MEC, 1997; DAVOK; BERNARD, 2016) e os pesquisadores somam ocorrências de alunos em coortes de interesse, como por exemplo o total de ingressantes, o total de evadidos e o total de matrículas em determinado ano e manipulam razões entre essas agregações.

Na revisão realizada, os métodos para cálculo das taxas de evasão foram estratificados em três grupos: (1) métodos que consideram gerações completas de alunos, onde alunos têm o tempo máximo de integralização expirado, (MEC, 1997; SILVA; MAINIER; PASSOS, 2006) ou o tempo mínimo de integralização expirado (DAVOK; BERNARD, 2016); (2) métodos que consideram alunos ingressantes dentro de um período específico (NODARI; LIMA; MACIEL, 2018; VANZ *et al.*, 2016; LAMERS; SANTOS; TOASSI, 2017), e (3) métodos que propõem o acompanhamento da série histórica das taxas (SANTANA, 2016; SILVA FILHO; LOBO, 2012; SILVA

FILHO *et al.*, 2007). Cada grupo oferece seus métodos distintos de cálculo e os componentes de suas fórmulas são descritos a seguir.

No primeiro grupo temos as contribuições da Comissão do MEC (MEC, 1997), Silva, Mainier e Passos (2006) e Davok e Bernard (2016). Na metodologia proposta pela Comissão do MEC (1997) são considerados alunos evadidos os alunos que não se diplomaram no período máximo de sua geração e não estão mais vinculados ao curso. Para o cálculo da taxa de evasão são considerados o número total de ingressantes (I_i) de determinado ano base (i), o número de alunos ingressantes nesse ano que se diplomaram (D_i) e o número de alunos ingressantes nesse ano que estão retidos (R_i). A aplicação deste método de cálculo permite conhecer o desempenho de determinada turma de ingressantes ao final do seu tempo máximo de integralização (m). A relação entre as variáveis pode ser vista na Equação 2.1

$$Taxa_{evasão} = \left(\frac{I_i - D_{i+m} + R_{i+m}}{I_i} \right) * 100. \quad (4.1)$$

Uma variação do cálculo da evasão por ciclos completos é apresentada por Silva, Mainier e Passos (2006). Os autores propõem que a taxa de evasão seja calculada por meio da razão que tem como componentes o total de vagas ofertadas, (V_i), para ingresso em determinado ano de referência (i), e os concluintes dessa turma de ingressantes, (D_i), respeitado o tempo normal de integralização curricular (n), conforme Equação 2.2

$$Taxa_{evasão} = \frac{V_i - (D_{i+n})}{V_i} * 100. \quad (4.2)$$

Outra variação utiliza em seus cálculos de aferição das taxas alunos com o tempo normal de integralização curricular expirado, conforme Equação 2.3. Para os cálculos utilizou-se o total de ingressantes (I_i) do ano base (i) e aferiu-se, desse total de ingressantes, quantos haviam se diplomado ao final do tempo normal de integralização curricular (D_{i+n}), onde n é o tempo normal de integralização curricular -1, em anos. Os autores (DAVOK; BERNARD, 2016) consideraram como evadidos alunos que abandonaram seus cursos, alunos que cancelaram sua matrícula,

transferidos internos, externos e *ex officio* e alunos desligados da instituição por infração disciplinar

$$Taxa_{evasão} = 1 - \left(\frac{D_{i+n}}{I_i} \right). \quad (4.3)$$

Essas taxas não mostram a realidade atual do curso, como qual a taxa anual de evasão, por considerar gerações completas de alunos, tanto no tempo ideal de integralização como no tempo máximo. Essa foi uma limitação levantada pela própria Comissão Especial do MEC (MEC, 1997) e para superá-la foi constatado que é necessária diferente metodologia para conhecer a situação momentânea da evasão.

O segundo grupo apresenta contribuições de Nodari, Lima e Maciel (2018), Vanz *et al.* (2016) e Lamers, Santos e Toassi (2017) e buscam verificar a taxa de evasão de alunos ingressantes dentro de um período específico. Nodari, Lima e Maciel (2018) elegeram os alunos ingressantes em 2009 como o coorte de interesse. Para esses alunos, no momento da coleta dos dados (2015), foi verificado a situação do aluno no fluxo acadêmico, ou seja, se tinham evadido, se estavam retidos, se haviam concluído seus estudos ou se haviam se transferido da instituição. Para efetivar os cálculos das taxas de evasão, o total de alunos ingressantes no ano base (i) que evadiram o curso (E) até o ano no de referência (j) foi relativizado pelo total de ingressantes no ano base (I_i), como mostrado na Equação 2.4. Cálculo análogo foi utilizado para as dimensões “conclusão”, “retenção” e “transferências”

$$Taxa_{evasão} = \left(\sum_{k=i}^j E_k / I_i \right) * 100. \quad (4.4)$$

Método semelhante foi utilizado por Vanz *et al.* (2016), contudo os autores ampliaram a abrangência. Assim, não utilizaram apenas ingressantes de um ano específico, mas sim os ingressantes de 2000 a 2009. Assim como feito por Nodari, Lima e Maciel (2018), no momento da coleta dos dados a situação do aluno no fluxo acadêmico foi aferida e o total de alunos por situação foi relativizado pelo total de ingressantes no período escolhido.

Lamers, Santos e Toassi (2017), também nessa linha de análise, introduziram uma análise que conseguiu encontrar as taxas de evasão estratificada pelos seus motivadores: reprovações ou trancamento e cancelamento de disciplinas. Para a aplicação desse método, além da aferição da situação do aluno no fluxo acadêmico, é necessário coletar os dados dos motivadores da evasão utilizando, por exemplo, entrevistas semiestruturadas com os alunos evadidos (LAMERS; SANTOS; TOASSI, 2017).

Ao analisar as contribuições de Nodari, Lima e Maciel (2018), Vanz *et al.* (2016) e Lamers, Santos e Toassi (2017), é possível compreender que o ato de isolar determinado grupo de alunos ingressantes, seja em um ano ou em um período de maior abrangência, abre possibilidades de se conhecer melhor quais as taxas de evasão aferidas em cada forma de ocorrência. O ponto negativo é que as taxas são relativas apenas ao grupo de alunos e ao ano ou período selecionados, o que não permite acompanhar a evolução históricas das taxas institucionais de evasão.

No terceiro grupo, tem-se as contribuições de Santana (2016), Silva Filho e Lobo (2012) e Silva Filho *et al.* (2007), que mostraram ou propuseram maneiras de se aferir taxas anuais. Santana (2016) apresentou método que buscou quantificar, em uma série histórica, o total de alunos matriculados (M_j) e o total de alunos evadidos (E_j). As taxas de evasão encontradas são, assim, uma razão entre esses totais no ano de referência (j), conforme mostrado na Equação 2.5

$$Taxa_{evasão} = \frac{E_j}{M_j} * 100. \quad (4.5)$$

Silva Filho e Lobo (2012) e Silva Filho *et al.* (2007) apresentam método que pode ser utilizado para o cálculo da evasão anual. O método proposto leva em consideração os matriculados (M); os ingressantes (I) e os concluintes (D) de determinado ano de referência (j) em comparação com o ano anterior ($j-1$). A relação entre as variáveis está explícita na Equação 2.6

$$Taxa_{evasão} = \left[1 - \frac{(M_j - I_j)}{(M_{j-1} - D_{j-1})} \right] * 100. \quad (4.6)$$

Comparativamente ao primeiro e segundo grupos, as taxas calculadas no terceiro grupo são mais flexíveis e atuais, uma vez que consideram as métricas de totais de matrículas, ingressos e cancelamentos anuais, e não de um coorte específico como os alunos que extrapolaram seus limites ideal ou máximo de integralização ou ingressantes apenas de um momento histórico específico. O emprego desses métodos de taxas anuais permite analisar historicamente a evolução das taxas de evasão e, assim, tem maior aplicabilidade no monitoramento das taxas de evasão por ser mais abrangente. Os métodos dos dois primeiros grupos podem ser aplicados em situações que o objetivo seja uma análise específica de determinado coorte de alunos ou cursos.

Para o desenvolvimento do DW foi escolhida como método de taxas de evasão derivada das contidas no terceiro grupo. Optou-se por não seguir o método proposto por Silva Filho e Lobo (2012) e Silva Filho *et al.* (2007), pois é desconsiderado o total de ingressantes do ano. Considerando que os ingressantes do ano de referência podem já se evadir no seu primeiro ano, considerá-los é importante para se ter um panorama mais abrangente do fenômeno. A lógica vista em Santana (2016) mostra quantos alunos a instituição perdeu relativo ao total de matrículas que ela possuía. O acompanhamento histórico dessas taxas mostra possíveis sucessos de medidas de mitigação da evasão caso seja registrada queda nas taxas. Analogamente, um aumento dessa taxa pode acentuar a necessidade de atuação institucional em determinado curso.

O método explicado a seguir, derivado de Santana (2016), tornar explícita a forma de cálculo da taxa que compõe a documentação do DW, uma vez que a fonte original do método teve ênfase na aplicação do método e não nas considerações metodológicas de sua aplicação.

O primeiro passo é definir qual o ano de referência inicial, ou ano base. Idealmente seria o ano da primeira turma do curso. Contudo a escolha pode ser em qualquer ano e período letivo desejado. Os totais de alunos calculados se referem ao final do ano de referência e a respectiva situação no fluxo acadêmico. A situação no fluxo acadêmico, de forma simplificada, pode ser resumida em três modalidades: (1) ingressantes⁷ no ano i (I_i) que é a soma dos ingressantes no ano de referência que

⁷ O cálculo dos ingressantes considera apenas os alunos que ocuparam vagas.

ocuparam vagas iniciais e remanescentes⁸; (2) alunos evadidos no ano (E_i): alunos que evadiram dos seus cursos no ano de referência, sendo por ação do aluno, desligamento, mobilidade etc.; e (3) diplomados no ano, (D_i): alunos que concluíram seus estudos no ano de referência.

Definidos os tipos de modalidades do fluxo acadêmico, o próximo passo é determinar do total de ingressantes do ano base, quantos estão aptos a se matricularem no ano seguinte, aqui chamados de veteranos (veteranos para o próximo ano – VPA)

$$VPA_{ano\ base} = I_{ano\ base} - E_{ano\ base}. \quad (4.7)$$

Para aferir o VPA do ano base basta simplesmente subtrair do total de ingressantes do ano base o número de ingressantes evadidos no ano base. Para os demais anos, a fórmula do VPA sobre uma ligeira modificação, ao considerar o VPA no ano anterior e subtrair os alunos diplomados⁹

$$VPA_j = VPA_{j-1} + I_j - E_j - D_j. \quad (4.8)$$

O total de matrículas no ano (TMA), considerando o ano base, é igual ao número de ingressantes desse ano. A partir do segundo ano de análise, considera-se também os veteranos do ano anterior

$$TMA_j = VPA_{j-1} + I_j. \quad (4.9)$$

Finalmente, ao se calcular a razão entre os evadidos (E_j) e a TMA_j é possível se chegar em uma taxa de evasão no ano (TEA) que permite dizer do total de matrículas registradas pela instituição naquele ano, quantas evadiram. O

⁸ Não foram considerados os ocupantes de vagas. Foram excluídos alunos de ingresso judicial, *ex officio* e de convênios nacionais e internacionais por não ocuparem vagas regulares.

⁹ Para alunos ocupantes de vagas iniciais que não aproveitaram disciplinas de outros cursos, é esperado que se formem no tempo mínimo de integralização. Contudo, é necessário considerar que alunos ingressantes por transferência externa ou interna podem concluir seus cursos em tempo menor que o mínimo previsto.

acompanhamento em uma série histórica dessa taxa permite identificar tendências e ainda monitorar possíveis ações de mitigação do fenômeno

$$TEA_j = \left(\frac{E_j}{TMA_j} \right) \times 100. \quad (4.10)$$

Em conclusão, conforme afirmado por Silva Filho e Lobo (2012), o objetivo do cálculo de taxas de evasão deve ser a determinação e aplicação de um método que siga inalterado no tempo para que seja possível acompanhar a evolução das taxas e identificar tendências históricas do fenômeno na unidade de observação escolhida, seja os cursos individualmente, grupo de cursos, unidades acadêmicas ou a instituição como um todo.

4.3 VARIÁVEIS DISCENTES RELEVANTES À EVASÃO

Os resultados apresentados nessa subseção derivaram da pesquisa bibliográfica e focou em identificar as variáveis discentes definidas como relevantes à evasão por seus autores. Foi observado que as pesquisas que contribuem para a determinação de variáveis discente são de natureza descritiva. Pesquisas descritivas têm como principal objetivo descrever características de determinada população (GIL, 2008; MUNIZAGA; CIFUENTES; BELTRAN, 2018), segmentar a unidade de análise em grupos com características semelhantes (ZIKMUND *et al.*, 2000) ou estabelecer relações entre variáveis (GIL, 2008; VERGARA, 1990).

As pesquisas descritivas em evasão visam descrever os alunos por meio de características pessoais (NODARI; LIMA; MACIEL, 2018; VANZ *et al.*, 2016; SILVA *et al.*, 2012) e características acadêmicas (NODARI; LIMA; MACIEL, 2018). Essas características podem se tornar fatores associados à ocorrência da evasão caso sejam validadas como envolvidas na ocorrência do fenômeno identificação de fatores associados à ocorrência do fenômeno (VANZ *et al.*, 2016; TONTINI; WALTER, 2014; SILVA *et al.*, 2012; NODARI; LIMA; MACIEL, 2018). Com base na pesquisa bibliográfica foi possível observar 65 variáveis de interesse, sendo 33 variáveis de caráter pessoal (50,8%) e 32 variáveis ligadas à vida acadêmica dos alunos (49,2%). As variáveis levantadas estão listadas, respectivamente, nas Tabelas 1 e Tabela 2.

Tabela 1 –Variáveis Descritivas Pessoais dos Discentes

Autores	Variáveis de Interesse
Ferrão e Almeida (2018); Lamers, Santos e Toassi (2017); Nodari, Lima e Maciel (2018); Santana (2016); Vanz <i>et al.</i> (2016)	Idade
Ferrão e Almeida (2018); Lamers, Santos e Toassi (2017); Nodari, Lima e Maciel (2018); Silva <i>et al.</i> (2012)	Sexo
Ferrão e Almeida (2018); Vanz <i>et al.</i> (2016)	Escolaridade da família
Lamers, Santos e Toassi (2017)	Comprometimento pessoal com o curso; Rotina de estudos; Tempo entre a conclusão do ensino médio e o ingresso no ensino superior
Lamers, Santos e Toassi (2017); Santana (2016); Silva <i>et al.</i> (2012)	Motivos da evasão; Relacionamento e integração social (pessoais e acadêmicas)
Lamers, Santos e Toassi (2017); Santana (2016); Silva <i>et al.</i> (2012); Tontini e Walter (2014)	Perspectivas profissionais sobre o curso; Inserção no mercado de trabalho
Lamers, Santos e Toassi (2017); Santana (2016); Silva <i>et al.</i> (2012); Vanz <i>et al.</i> (2016)	Renda pessoal e da família; Classe social; Problemas financeiros
Lamers, Santos e Toassi (2017); Santana (2016); Tontini e Walter (2014)	Tempo disponível para estudo
Lamers, Santos e Toassi (2017); Santana (2016); Silva <i>et al.</i> (2012); Tontini e Walter (2014)	Motivação para escolha do curso
Lamers, Santos e Toassi (2017); Silva <i>et al.</i> (2012); Vanz <i>et al.</i> (2016)	Atividade profissional durante o curso
Lamers, Santos e Toassi (2017); Tontini e Walter (2014)	Disponibilidade de tempo para frequentar as aulas
Lamers, Santos e Toassi (2017); Vanz <i>et al.</i> (2016)	Egresso de escola pública
Mendes Júnior (2014); Nodari, Lima e Maciel (2018)	Desempenho no vestibular; Ingresso por cotas (escola pública, racial ou de renda)
Santana (2016)	Problemas familiares
Santana (2016); Tontini e Walter (2014)	Vocação ou escolha vocacional insuficiente
Santana (2016); Vanz <i>et al.</i> (2016)	Local de residência; Distância da residência até a universidade; Gênero
Silva <i>et al.</i> (2012)	Cursava outro curso concomitantemente
Tontini e Walter (2014)	Estabilidade pessoal e familiar; Identificação do discente com o curso; Motivação para vida; Persistência nos objetivos; Saúde Pessoal
Vanz <i>et al.</i> (2016)	Estado civil; Participação na renda familiar

Fonte: elaborado pelo autor com base nos resultados da pesquisa bibliográfica.

Tabela 2 – Variáveis Descritivas Acadêmicas dos Discentes

Autores	Variáveis de Interesse
Ferrão e Almeida (2018)	Opção de universidade no vestibular
Ferrão e Almeida (2018); Mendes Júnior (2014); Nodari, Lima e Maciel (2018); Santana (2016); Silva <i>et al.</i> (2012); Vanz <i>et al.</i> (2016)	Rendimento escolar (índices acadêmicos de desempenho e reprovações) no ensino superior
Ferrão e Almeida (2018); Nodari, Lima e Maciel (2018)	Pontuação no processo seletivo de admissão
Ferrão e Almeida (2018); Silva <i>et al.</i> (2012)	Opção de curso no vestibular (primeira ou segunda opção)
Lamers, Santos e Toassi (2017)	Duração do curso; Forma de avaliação da aprendizagem; Habilidades didático-pedagógicas do professor; Nível de exigência dos estudos; Organização dos horários das disciplinas; Participação nas aulas; Regime de oferta do curso (semestral ou anual); Carga horária matriculada
Lamers, Santos e Toassi (2017); Ferrão e Almeida (2018)	Rendimento escolar (índices acadêmicos de desempenho e reprovações) no ensino fundamental
Lamers, Santos e Toassi (2017); Nodari, Lima e Maciel (2018); Vanz <i>et al.</i> (2016)	Situação no fluxo acadêmico (cursando, conclusão, evasão, retenção)
Lamers, Santos e Toassi (2017); Santana (2016); Silva <i>et al.</i> (2012)	Período letivo da evasão
Lamers, Santos e Toassi (2017); Silva <i>et al.</i> (2012); Tontini e Walter (2014)	Infraestrutura da universidade
Lamers, Santos e Toassi (2017); Tontini e Walter (2014)	Aplicação profissional do aprendizado; Atividades práticas do curso
Lamers, Santos e Toassi (2017); Vanz <i>et al.</i> (2016)	Atividades extracurriculares
Mendes Júnior (2014)	Dificuldades acadêmicas relativas ao curso
Mendes Júnior (2014); Nodari, Lima e Maciel (2018); Vanz <i>et al.</i> (2016)	Escolha do curso
Nodari, Lima e Maciel (2018)	Área do conhecimento do curso; Modalidade do curso (bacharelado, licenciatura etc.)
Santana (2016)	Realidade universitária distante do imaginário pré-universitário; Recepção e orientação no início do curso
Silva <i>et al.</i> (2012)	Acervo bibliográfico da universidade; Informação ou desinformação sobre o curso escolhido; Quadro de professores da universidade
Silva <i>et al.</i> (2012); Tontini e Walter (2014)	Avaliação e qualidade institucional
Tontini e Walter (2014)	Reforço acadêmico
Vanz <i>et al.</i> (2016)	Recebimento de benefícios de assistência estudantil; Recebimento de bolsas

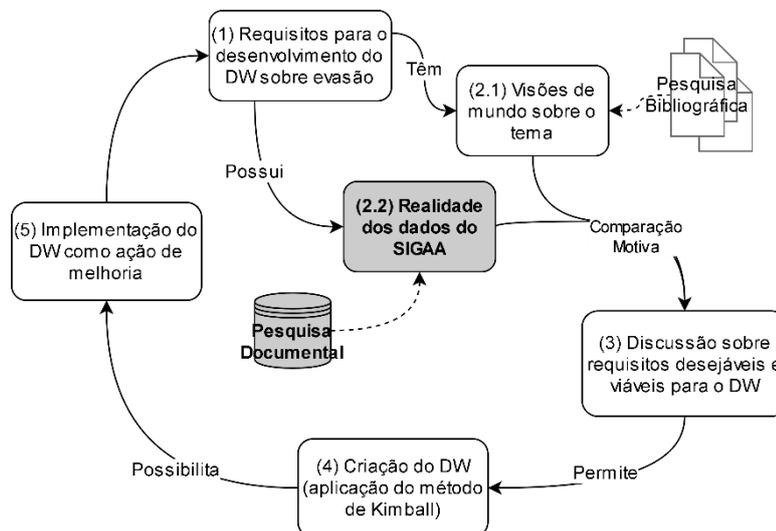
Fonte: elaborado pelo autor com base nos resultados da pesquisa bibliográfica.

Esses resultados têm implicações práticas para determinar quais são os requisitos de informação desejados sobre as características pessoais e acadêmicas para o DW.

4.4 REALIDADE DOS DADOS DA UNIFEI

Essa subseção apresenta os resultados da pesquisa documental. Essa é a segunda etapa do segundo passo da metodologia proposta, graficamente mostrado na Figura 13.

Figura 13 – Segunda etapa do segundo passo de aplicação da metodologia proposta



Fonte: Proposta do Autor. Adaptado de Checkland e Poulter (2010)

A análise de como o DB do SIGAA está estruturado visou conhecer quais tabelas, e seus respectivos atributos, possuem dados que possam ser utilizados no DW. A exploração do BD revelou que as tabelas do SIGAA são organizadas em esquemas (do inglês *schema*). Esquemas são estruturas lógicas que podem ser utilizadas para a organização de tabelas (POSTGRESQL, 2019). Os principais esquemas que contêm tabelas com informações pessoais e acadêmicas dos alunos são (1) o esquema “comum”; (2) o esquema “public”; e (3) o esquema “ensino”. Dentro desses esquemas foram encontradas tabelas que armazenam dados sobre os alunos desde 1998. A nomenclatura utilizada para se referenciar a tabela é o nome do esquema sucedido do nome da tabela. Esses dados foram listados na Tabela 3.

Tabela 3 – Dados Pessoais e Acadêmicos Registrados pelo SIGAA

Tabelas	Atributos
comum.pessoa	Nome; data de nascimento; sexo; raça/etnia; endereço; registro geral (RG); cadastro de pessoa física (CPF); nacionalidade; naturalidade; nome da mãe; nome do pai; estado civil; título de eleitor; documento militar; endereço; e-mail; telefone; escola de conclusão do ensino médio; ano de conclusão do ensino médio; tipo de escola (pública ou privada); deficiência
public.discente, com relacionamento em public.curso, public.unidade	Ano de ingresso; período letivo de ingresso; matrícula; nível de ensino (graduação ou pós-graduação); situação atual (ativo, cancelado, trancado etc.); forma de ingresso; carga horária integralizada atual; prazo para conclusão do curso; data da colação de grau; curso; currículo (ou grade curricular do curso); período atual (permanência no curso); data da integralização do currículo, unidade acadêmica (unidade responsável pelo curso do aluno)
ensino.turma, ensino.indice_academico_discente; ensino.componente_curricular, ensino.movimentacao_aluno	Disciplinas e outras atividades curriculares cursadas; carga horária das disciplinas e atividades curriculares; notas atribuídas às disciplinas e atividades curriculares; frequência do aluno nas disciplinas e atividades curriculares; disciplina obrigatória, optativa ou eletiva; índices acadêmicos de desempenho; grupo de cota de ingresso (escola pública, renda, grupo étnico e deficiência), registro das movimentações do aluno (cancelamento da matrícula, integralização curricular ou colação de grau) e respectiva data de ocorrência

Fonte: elaborado pelo autor com base nos resultados da pesquisa documental.

Esses dados são gerados de duas maneiras principais: (1) pela entrada manual de dados por servidores da Unifei ou pela entrada manual realizada pelos próprios alunos (restrito apenas a seus dados de contato) ou (2) gerados por processamento do próprio SIGAA. Os dados gerados pela entrada manual por servidores são os dados pessoais dos alunos, registrado no ato da matrícula, que vincula o aluno à um curso, e dados sobre unidades acadêmicas, cursos, disciplinas, grades curriculares e oferta de turmas. Dados gerados pelo próprio SIGAA tem relação com o

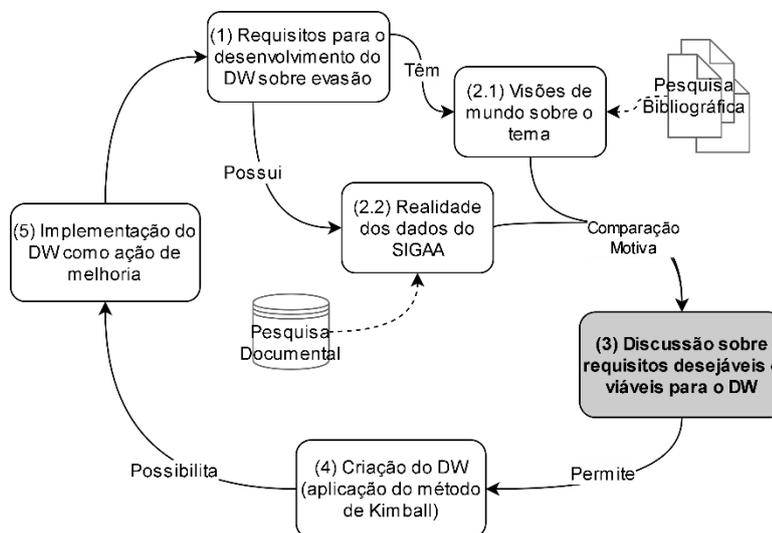
processamento de matrículas em disciplinas, o que permite aos discentes cursá-las, com os cálculos de carga horária integralizada, atualização do período letivo corrente do aluno e o cálculo de índices acadêmicos, que leva em consideração as notas e as cargas horárias das disciplinas das turmas cursadas em um semestre letivo.

O próximo passo é triangular os resultados das pesquisas bibliográficas, listadas nas Tabelas 1 e 2, com o resultado da pesquisa documental, listada na Tabela 3. Essa triangulação é apresentada na próxima subseção

4.5 REQUISITOS DE INFORMAÇÃO

Essa etapa visou determinar quais os requisitos de informação o DW deve possuir para que a modelagem dos dados. A discussão dos requisitos viáveis e desejáveis do DW e é composta por duas iniciativas de triangulação dos dados das pesquisas bibliográfica e documental. A Figura 14 localiza essa etapa no fluxo da metodologia proposta.

Figura 14 – Terceiro passo de aplicação da metodologia proposta



Fonte: Proposta do Autor. Adaptado de Checkland e Poulter (2010)

A primeira iniciativa na triangulação dos dados visou determinar quais informações são necessárias para armazenar no DW para o cálculo das taxas de evasão. O método escolhido para o cálculo das taxas de evasão foi o desenvolvido internamente, conforme discussão realizada na Seção 4.2. As equações 4.7 a 4.10

mostram todos os componentes para o cálculo das taxas de evasão, onde VPA representa o total de veteranos para o próximo, I representa o total de ingressantes, E representa o total de evadidos, D representa o total de diplomados, TMA representa o total de matrículas no ano, TEA é a taxa de evasão no ano e j representa o j -ésimo ano de interesse

$$VPA_{ano\ base} = I_{ano\ base} - E_{ano\ base}, \quad (4.7)$$

$$VPA_j = VPA_{j-1} + I_j - E_j - D_j, \quad (4.8)$$

$$TMA_j = VPA_{j-1} + I_j, \text{ e} \quad (4.9)$$

$$TEA_j = \left(\frac{E_j}{TMA_j} \right) \times 100. \quad (4.10)$$

Os componentes principais para o cálculo são, por ano de interesse: (1) o total de ingressantes, (2) o total de alunos cancelados e (3) o total de alunos diplomados. Os componentes VPA , TMA e TEA são obtidos por meio de manipulações algébricas.

A pesquisa documental revelou que o SIGAA armazena dados que podem ser tratados para a obtenção dos componentes principais. Primeiramente, é possível consultar a tabela “ensino.discente” e contar todos os discentes de graduação, por curso, que ingressaram em determinado ano. Calcular os totais de alunos cancelados e diplomados segue lógica semelhante, mas a consulta é realizada na tabela “ensino.movimentacao_aluno”, onde é possível contar as ocorrências de cancelamentos e diplomações, também por ano e curso.

A segunda iniciativa de triangulação dos dados teve o objetivo de comparar o que a pesquisa bibliográfica mostrou como variáveis descritivas dos alunos, ou variáveis correlatas à evasão, reunidas nas Tabela 1 e Tabela 2, com os dados armazenados pelo SIGAA, que foram levantados com a pesquisa documental, reunidos na Tabela 3. As variáveis que o SIGAA registra são os requisitos de informação possíveis de atendimento imediato para o desenvolvimento do DW e estão organizadas na Tabela 4.

Tabela 4 – Triangulação da pesquisa bibliográfica e documental

Resultado da Triangulação		Variáveis
SIGAA registra		Idade; sexo; tempo entre a conclusão do ensino médio e o ingresso no ensino superior; egresso de escola pública; ingresso por cotas (escola pública, racial ou de renda); local de residência; distância da residência até a universidade; estado civil; rendimento escolar (índices acadêmicos de desempenho e reprovações) no ensino superior; pontuação no processo seletivo de admissão; organização dos horários das disciplinas; regime de oferta do curso (semestral ou anual); carga horária matriculada; situação no fluxo acadêmico (cursando, conclusão, evasão, retenção); período letivo da evasão; escolha do curso; área do conhecimento do curso; modalidade do curso (bacharelado, licenciatura etc.); duração do curso; acervo bibliográfico da universidade; quadro de professores da universidade; recebimento de benefícios de assistência estudantil; recebimento de bolsas
SIGAA registra	não	Escolaridade da família; comprometimento pessoal com o curso; rotina de estudos; motivos da evasão; relacionamento e integração social (pessoais e acadêmicas); perspectivas profissionais sobre o curso; inserção no mercado de trabalho; renda pessoal e da família; classe social; problemas financeiros; tempo disponível para estudo; motivação para escolha do curso; atividade profissional durante o curso; disponibilidade de tempo para frequentar as aulas; problemas familiares; vocação ou escolha vocacional insuficiente; cursava outro curso concomitantemente; estabilidade pessoal e familiar; identificação do discente com o curso; motivação para vida; persistência nos objetivos; saúde pessoal; participação na renda familiar; opção de universidade no vestibular; opção de curso no vestibular (primeira ou segunda opção); forma de avaliação da aprendizagem; habilidades didático-pedagógicas do professor; nível de exigência dos estudos; participação nas aulas; rendimento escolar (índices acadêmicos de desempenho e reprovações) no ensino fundamental; infraestrutura da universidade; aplicação profissional do aprendizado; atividades práticas do curso; atividades extracurriculares; dificuldades acadêmicas relativas ao curso; realidade universitária distante do imaginário pré-universitário; recepção e orientação no início do curso; informação ou desinformação sobre o curso escolhido; avaliação e qualidade institucional; reforço acadêmico; gênero; desempenho no vestibular

Fonte: elaborado pelo autor com base nos resultados das pesquisas bibliográfica e documental

Do total de 65 variáveis identificadas na literatura, o SIGAA registra 23 variáveis, ou 35,4% do total, sendo 15 variáveis acadêmicas e 8 variáveis pessoais. Essas variáveis estão listadas na Tabela 5.

Contudo, doze variáveis registradas, quatro pessoais e oito acadêmicas, não foram escolhidas para compor o DW. A variável pessoal “local de residência” mostra apenas a situação de momento de moradia do aluno e é editável pelo SIGAA. Assim, o local de residência armazenado é momentâneo, podendo refletir o endereço de uma república ou pensão que o aluno mora provisoriamente. Foi escolhido, em substituição, dados da naturalidade do aluno, tanto a cidade quanto o respectivo

estado. A variável “distância da residência até a universidade” exige que fosse calculada a distância entre as cidades informadas como residência e os campi da Unifei em Itajubá ou Itabira. O autor não tinha disponíveis esses dados.

Tabela 5 – Apenas variáveis pessoais e acadêmicas registradas pelo SIGAA

Variáveis Pessoais	Variáveis Acadêmicas
<ul style="list-style-type: none"> • distância da residência até a universidade • egresso de escola pública • estado civil • idade • ingresso por cotas (escola pública, racial ou de renda) • pontuação no processo seletivo de admissão • local de residência • sexo • tempo entre a conclusão do ensino médio e o ingresso no ensino superior 	<ul style="list-style-type: none"> • acervo bibliográfico da universidade • área do conhecimento do curso • carga horária matriculada • duração do curso • escolha do curso • modalidade do curso (bacharelado, licenciatura etc.) • organização dos horários das disciplinas • período letivo da evasão • quadro de professores da universidade • recebimento de benefícios de assistência estudantil • recebimento de bolsas • regime de oferta do curso (semestral ou anual) • rendimento escolar (índices acadêmicos de desempenho e reprovações) no ensino superior • situação no fluxo acadêmico (cursando, conclusão, evasão, retenção)

Fonte: elaborado pelo autor com base nos resultados das pesquisas bibliográfica e documental

As variáveis “ingresso por cotas (escola pública, racial ou renda)” e “pontuação no processo seletivo de admissão”, no banco de dados do SIGAA, não representava o total de ingressantes nessas modalidades, que na Unifei se deu a partir de 2011. No SIGAA era possível apenas extrair dados dos cotistas a partir de 2016. A solução seria consultar e integrar ao DW os dados do SISU Gestão, sistema de informação do MEC que fornece às IES a lista de todos os aprovados no SISU. Contudo, como o esforço inicial do DW está focado em trabalhar com dados internos já armazenados no SIGAA, essa integração é uma ação que poderá ser considerada futuramente.

A variável “acervo bibliográfico da universidade” tem relação com a infraestrutura da IES e a variável “quadro de professores da universidade” está relacionada aos recursos humanos da instituição. Ambas estão além do escopo do DW, que foca em discentes e cursos. As variáveis “carga horária matriculada” e “organização dos horários das disciplinas” exigiriam a manipulação individual e retroativa dos históricos dos discentes para recompor a situação de momento desde seu ingresso. Esse esforço adicional teria grande custo de tempo e o benefício trazido, no momento, não o compensaria. Duas variáveis estão relacionadas a características dos cursos. A variável “duração do curso”, isolada, não traz informações relevantes para a análise da evasão. A variável “regime de oferta do curso (semestral ou anual)”, no contexto da Unifei, seria redundante, uma vez que todos os cursos de graduação em oferta estão no regime semestral. Finalmente, as variáveis “recebimento de benefícios de assistência estudantil” e “recebimento de bolsas” são registradas de forma parcial no SIGAA. Para conhecer o valor dos benefícios seria necessário integrar os dados com outro sistema, o Sistema Integrado de Gestão de Patrimônio, Administração e Contratos, o que não está no escopo inicial do DW que foca em dados disponíveis no SIGAA.

Outras variáveis, registradas no SIGAA, foram adicionadas ao modelo. Dentre as características pessoais, para identificar o aluno evadido optou-se por inserir no modelo o seu número do Cadastro de Pessoas Físicas (CPF) e sua matrícula na Unifei. Além das variáveis relacionadas com a cidade e UF de nascimento do aluno, também foi incluída a variável nacionalidade do aluno, para se ter visibilidade se o aluno é estrangeiro. Outra variável adicionada foi a variável raça, que pode fornecer uma análise preliminar substituta às cotas. Finalmente, variáveis sobre a forma de ingresso e seus respectivos ano e período também foram adicionadas.

Nas variáveis acadêmicas, as adições visaram descrever em mais detalhes o vínculo dos discentes a seus cursos. Inseriu-se as variáveis respectivas a qual Unidade Acadêmica o curso está vinculado, qual o turno de oferta, se diurno ou noturno e qual seu grau acadêmico, se bacharelado ou licenciatura. Assim, a versão final dos requisitos de informação do DW tomou a forma mostrada na Tabela 6.

Tabela 6 – Requisitos de informação para o DW

Variáveis para o cálculo da taxa de evasão	
Variáveis Pessoais	Variáveis Acadêmicas
<ul style="list-style-type: none"> • Total de alunos ingressantes por ano • Total de alunos cancelados por ano • Total de alunos diplomados por ano • Total de alunos veteranos para o próximo • Total de alunos matriculados por ano • Taxa de evasão por ano 	
<ul style="list-style-type: none"> • CPF • Matrícula • Nome • Sexo • Estado civil • Nacionalidade • Naturalidade • Data de nascimento • Raça • Escola de formação no ensino médio • Ano de conclusão do ensino médio • Escola pública • Forma de ingresso (SISU, Vestibular etc.) • Ano e período de ingresso 	<ul style="list-style-type: none"> • Curso matriculado • Unidade acadêmica de vínculo • Turno do curso • Área de conhecimento do curso • Modalidade de ensino (presencial ou à distância) • Campus de oferta do curso • modalidade do curso (bacharelado, licenciatura etc.) • rendimento escolar (índices acadêmicos de desempenho e reprovações) no ensino superior • situação no fluxo acadêmico (cursando, conclusão, evasão, retenção) • Ano e período de saída (evasão ou conclusão)

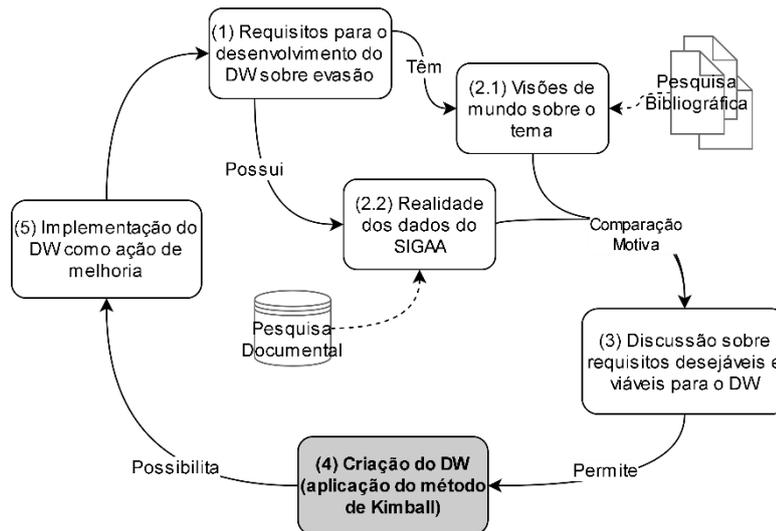
Fonte: elaborado pelo autor com base nos resultados das pesquisas bibliográfica e documental

Convém notar que as pesquisas descritivas, que foram base para a extração das variáveis descritivas discentes, têm como objetivo uma descrição do indivíduo mais detalhada e são embasadas em modelos teóricos próprios. Esses modelos usualmente possuem variáveis de interesse já definidas e utilizaram entrevistas e questionários para a coleta direcionada desses dados que visa validar, ou não, o modelo VANZ *et al.*, 2016; TONTINI; WALTER, 2014; SILVA *et al.*, 2012; NODARI; LIMA; MACIEL, 2018). Por sua vez, o SIGAA visa principalmente coletar dados oficiais para cadastro do aluno e expedição e registro de diplomas. Isso pode explicar a diferença entre as variáveis registradas e não registradas pelo SIGAA, uma vez que o objetivo da coleta dos dados é diferente.

4.6 CRIAÇÃO DO *DATA WAREHOUSE*

A criação do DW é o quarto passo da metodologia proposta, representado graficamente na Figura 15.

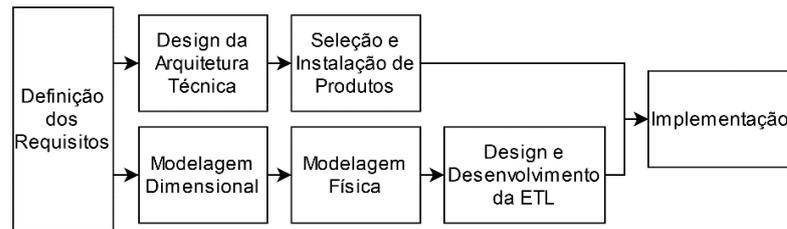
Figura 15 – Quarto passo de aplicação da metodologia proposta



Fonte: Proposta do Autor. Adaptado de Checkland e Poulter (2010)

Seguindo a proposta de desenvolvimento de Kimball *et al* (2008), explícita na Figura 8, o primeiro passo foi a definição dos requisitos de negócio e requisitos de informação do DW, o que foi discutido nas Subseções 4.1 e 4.5, respectivamente. Em resumo, no contexto do apoio à tomada de decisão sobre evasão na Unifei, os requisitos de negócio do DW são habilitar a instituição a monitorar taxas de evasão por curso, unidade acadêmica, área do conhecimento e campus, juntamente com a capacidade de descrever qualitativamente os alunos evadidos por meio de características consideradas relevantes nas pesquisas sobre evasão. Os requisitos de informação são o registro das taxas de evasão por curso e das características pessoais e acadêmicas dos alunos.

Figura 8 – Ciclo de vida do *data warehouse*

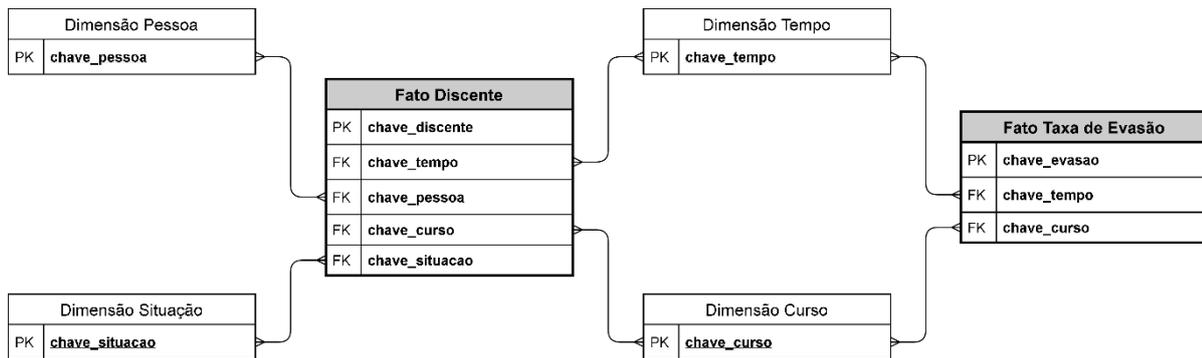


Fonte: adaptado de Kimball *et al.* (2008).

Após determinados os requisitos, a trilha do método de Kimball *et al.* (2008) se ramifica em duas. A trilha superior tem relação com a infraestrutura de hardware e software do DW. A primeira versão do DW, um protótipo, foi desenvolvido localmente na máquina do servidor responsável. Utilizou-se como fonte de dados os armazenados no SIGAA. O armazenamento dos dados do DW utilizou o software *freeware* de banco de dados PostgreSQL, que é o mesmo utilizado para o SIGAA. Tratando-se de software livre, esta opção se mostra de baixo custo para a Unifei, como também relatado por Calumby e Craveiro (2008). Ademais, futuras versões do DW podem ser hospedadas nos servidores institucionais e a utilização da mesma ferramenta de BD pode auxiliar na migração dos dados.

A trilha inferior registra os passos necessários para a modelagem dos dados e a posterior realização da ETL. A modelagem lógica utilizada está representada na Figura 16. O primeiro passo para a modelagem das tabelas do DW foi definir a granularidade dos registros que atendessem aos requisitos de negócio e de informação. Foram necessárias duas tabelas fato. A primeira tabela, Fato Discente, tem como grão o registro de informações individuais dos alunos e tem como fatos os índices acadêmicos em vigência na Unifei. A segunda tabela, Fato Taxa de Evasão, tem como grão a taxa de evasão por curso da Unifei e os respectivos componentes para o cálculo da taxa.

Figura 16 – Modelo dimensional do DW:



Fonte: proposta do autor com base na metodologia proposta e na literatura revista.

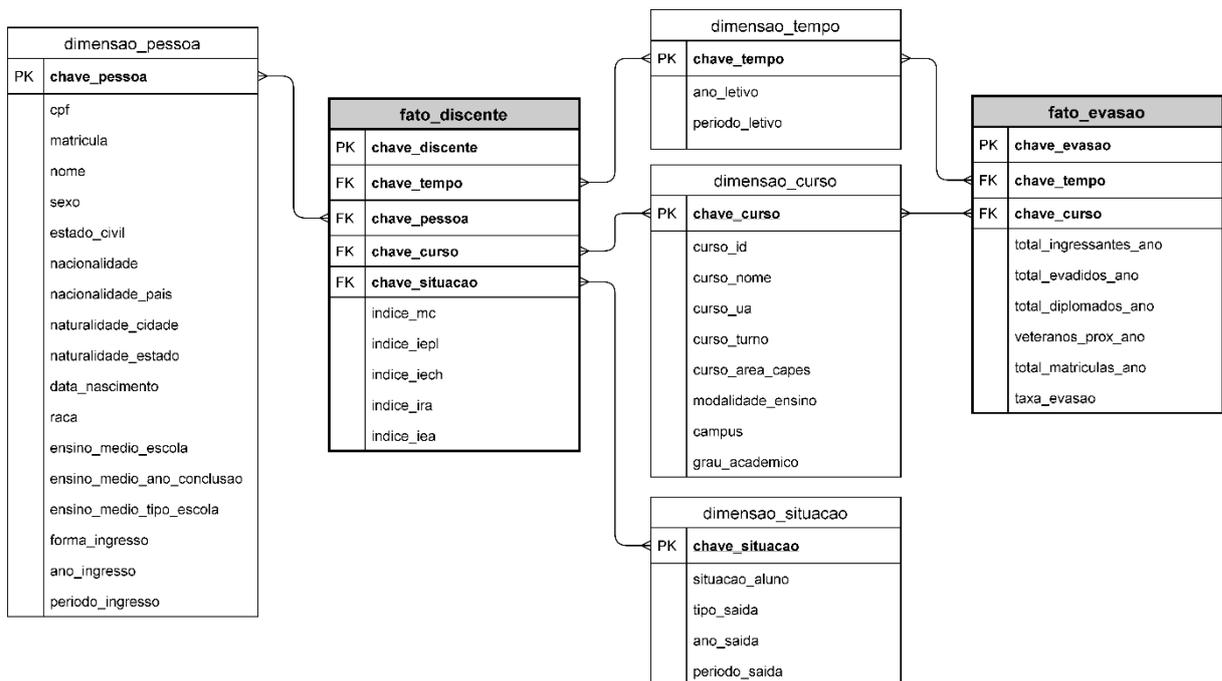
As tabelas dimensionais do modelo organizam de forma lógica as características dos alunos e dos cursos que contextualizam os fatos registrados pelas Tabelas Fato Discente e Fato Taxa de Evasão. Essas características podem ser utilizadas como filtros para a extração de dados do DW. A Tabela Dimensão Tempo armazena uma lista com anos e períodos letivos e está conectada a Tabela Fato Discente para ser possível identificar em quais anos e períodos letivos os discentes evadiram. Para a Tabela Fato Taxa de Evasão a conexão permite visualizar a evolução das taxas anuais de evasão. A Tabela Dimensão Pessoa registra os dados pessoais dos alunos e está conectada à Tabela Fato Discente para ser possível filtrar tanto as variáveis pessoais quanto os fatos acadêmicos registrados na tabela Fato Discente. A conexão entre a Tabela Dimensão Curso e a Tabela Fato Discente permite conhecer informações sobre o vínculo do aluno a determinado curso, trazendo informações adicionais sobre a Unidade Acadêmica a qual o curso está subordinado e a modalidade de educação, se presencial ou à distância. A conexão entre a Tabela Dimensão Curso e a Tabela Fato Taxa de Evasão é necessária para se conhecer a qual curso as taxas de evasão se referem. Finalmente, a conexão entre a Tabela Dimensão Situação e a Tabela Fato Discente descreve o aluno em relação a sua situação junto à Unifei, se o aluno estava ativo, se sua matrícula foi cancelada ou se o aluno concluiu seus estudos.

As Tabelas Dimensão Tempo e Dimensão Curso fisicamente serão compartilhadas entre os dois modelos. Esse conceito é denominado por Kimball *et al.* (2008) como *conformed dimensions*, ou dimensões conformes. Dimensões estão

conformes quando tabelas dimensionais são compartilhadas entre tabelas de fatos diferentes.

A modelagem física, que efetivamente cria as tabelas dimensionais e de fato no banco de dados que receberam dados pela ETL (KIMBALL *et al.*, 2008), seguiu a lógica relacional proposta na modelagem lógica da Figura 16. Essa etapa definiu, para cada tabela, as colunas e seus respectivos tipos de dados, definiu qual as chaves primárias das tabelas e os relacionamentos entre as tabelas. A modelagem física está representada de forma esquemática na Figura 17. Para facilitar a compreensão das modelagens lógica e física, o Apêndice A mostra um dicionário com a descrição de cada tabela do DW, suas respectivas colunas, a fonte de dados no SIGAA e o tipo de dado que o atributo deve ter. O Apêndice B mostra o código SQL utilizado para a criação das tabelas, a definição de seus atributos e os tipos de dados e o relacionamento entre as tabelas.

Figura 17 – Modelo Físico do DW



Fonte: proposta do autor com base na metodologia proposta.

Para popular as tabelas fatos e tabelas dimensões com dados, foi necessário desenvolver e realizar o processo de ETL. A escolha por desenvolver uma solução própria se deu principalmente por duas razões, com base no discutido por Debroy;

Brimble e Yost (2018), Yung (2018) e Weinberg (2019): (1) o custo associado à compra de uma solução não estava previsto para o projeto e (2) a demanda do projeto, inicialmente, é por transferir dados do SIGAA para o ambiente do DW, o que pode ser realizado com SQL e Python, linguagens com soluções *freeware* nas quais o autor tem conhecimento suficiente para esse uso.

A fonte dos dados para a ETL foi o banco de dados do SIGAA, especificamente as tabelas, colunas e seus respectivos registros mapeados na Tabela 3 e no esquema físico representados na Figura 17. Para garantir a compatibilidade entre os dados as definições do SIGAA a respeito de como armazenar os identificadores únicos e os nomes das entidades e colunas de suas tabelas foram herdadas pelo DW. Isso visou facilitar as consultas, uma vez que não é necessário retrabalhar as nomenclaturas para a carga dos dados no DW.

O primeiro processo de ETL se enquadra na descrição de uma ETL histórica (KIMBALL *et al.*, 2008), pois visou transferir todos os dados de interesse que estavam no SIGAA para o DW em uma única transferência. A característica principal de tratamento nos dados empregada na ETL para a tabela “fato_discente” foi a organização e consolidação dos dados e a desnormalização das tabelas. No BD do SIGAA os dados de interesse estavam dispersos em 17 tabelas, sendo necessário utilizar operadores *join* para organizar os dados. O Apêndice C mostra o código SQL utilizado para extração dos dados do SIGAA.

Para a Tabela “fato_evasao”, além das tarefas de organização, consolidação e desnormalização dos dados, foi necessário utilizar funções de agregação nos dados para transformar os dados individualizados dos alunos nos totais de ingressantes e evadidos do modelo. Para tal foi utilizado a função *count*, que conta as ocorrências de determinado campo de dados. Outra tarefa de tratamento nos dados foi realizada para acompanhar a evolução dos alunos do curso de Ciência da Computação (CCO), que teve mudado seu turno de oferta de noturno para integral em 2009. Assim, foi necessário dividir o código de extração de dados do SIGAA em três partes. Uma primeira consulta conta com os demais cursos de graduação, a segunda consulta foi feita para acompanhar a evolução dos alunos ingressantes em CCO entre 1998 a 2008 e os ingressantes de 2009 a 2019, com suas respectivas movimentações de evasão ou conclusão. O código em SQL que realizou essa tarefa de extração de dados do SIGAA está no Apêndice D.

Uma vantagem percebida que o DW tem sobre as consultas diretas no banco de dados do SIGAA é a simplificação da consulta, como relatado por Kimball *et al.* (2008). Enquanto uma consulta completa ao banco de dados, para consultar as variáveis que o SIGAA registra, exigiria a construção de uma consulta que relacionasse as oito tabelas mostradas na Tabela 3 e mais nove tabelas para desnormalizar os campos de dados, no DW a consulta envolve apenas duas tabelas de fatos e quatro tabelas de dimensões, já prontos para uso.

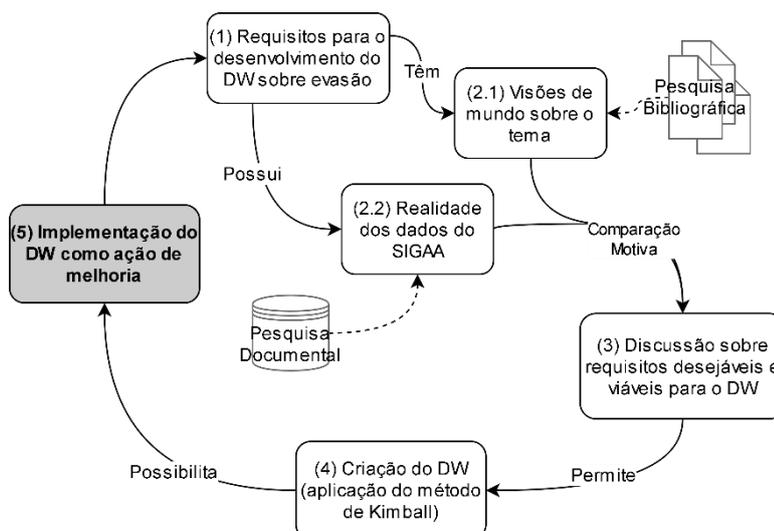
Para a manutenção da característica de manter dados históricos no DW (KIMBALL *et al.*, 2008; SINGH, 2001), são necessárias tarefas de ETL periódicas após a carga histórica inicial. Como discutido em Kimball *et al.* (2008), o ideal seria o desenvolvimento de um sistema chamado de *change data capture*, responsável por vasculhar a fonte de dados e extrair, tratar e carregar no DW apenas novos dados que foram gerados na fonte e não estão no DW. Contudo, o desenvolvimento desse sistema se caracteriza como um esforço complementar que está além do escopo dessa dissertação e será sugerida como uma sugestão para futuras pesquisas. Ciente da importância de se manter dados atualizados, uma solução prática, apesar de não ideal, seria repetir toda a ETL histórica referente a tabela “fato_evasao” ao final de cada ano letivo.

Kimball *et al.* (2008) e Inmon (2005) recomendam que o desenvolvimento do DW seja documentado e que metadados sejam criados para registrar quais são os componentes da arquitetura de dados e da infraestrutura do DW. Assim, espera-se que esse texto, juntamente com seus apêndices, seja uma referência dos metadados e da documentação das etapas de desenvolvimento do DW.

4.7 IMPLEMENTAÇÃO E UTILIZAÇÃO DO DATA WAREHOUSE

Essa subseção tem o objetivo de mostrar a implementação parcial do DW e sua aplicação prática ocorrida na Unifei em 2019. Essa é a última etapa do fluxo da metodologia proposta, conforme mostrado na Figura 18:

Figura 18 – Quinto passo da aplicação da metodologia proposta



Fonte: Proposta do Autor. Adaptado de Checkland e Poulter (2010)

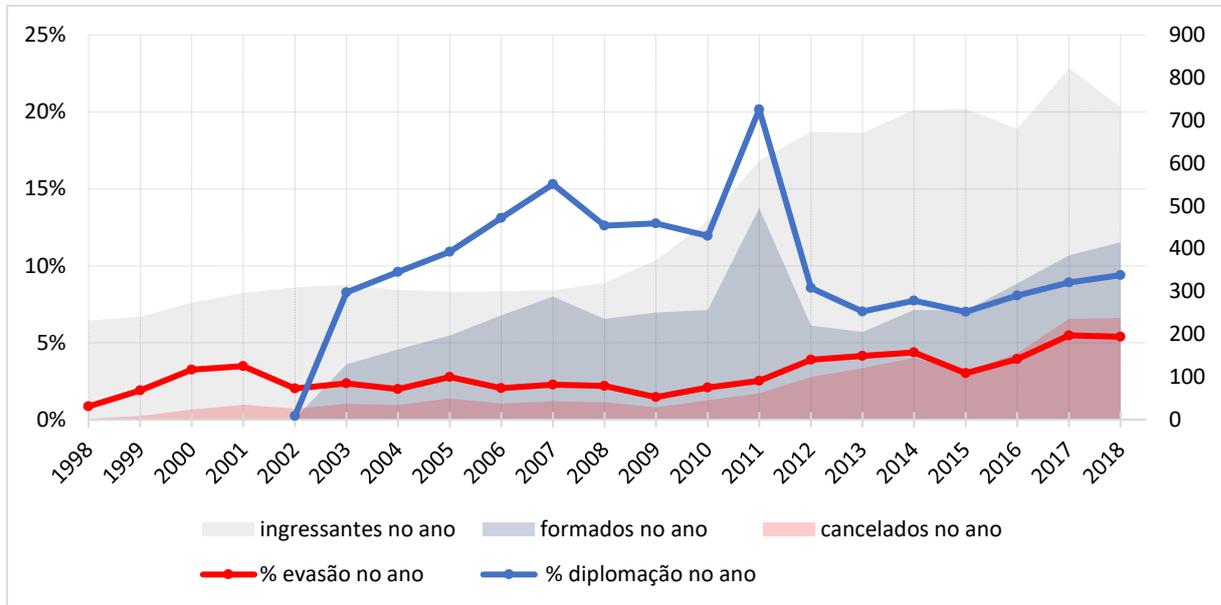
A validação administrativa do DW ocorreu no âmbito do Conselho de Ensino, Pesquisa, Extensão e Administração (CEPEAd) da Unifei. A versão preliminar da lógica dos cálculos das taxas de evasão e os dados que embasaram o desenvolvimento do DW foram apresentadas na 27ª Reunião Ordinária do CEPEAd, em 25 de setembro de 2019 (UNIFEI, 2019a). A relevância institucional dessa ferramenta foi ressaltada pelo Reitor, que expressou que a disponibilização dos dados deveria ser feita de forma rotineira para gestores acadêmicos. No âmbito da PRG, a primeira versão dos resultados foi compartilhada aos gestores acadêmicos locais, como diretores de unidades acadêmicas, coordenadores de curso e os membros dos núcleos docentes estruturantes dos cursos.

Como iniciativa de apoio à tomada de decisão, as taxas de evasão podem ser extraídas para a confecção de gráficos comparativos entre unidades de observação de interesse. Na realidade profissional do autor, a taxa de diplomação também é métrica acompanhada. Assim, as figuras mostradas a seguir também mostram essa taxa. As Figuras 19 e 20, mostradas a seguir, utilizaram os dados do DW e foram utilizadas na apresentação ao CEPEAd.

Ademais, como o escopo desse trabalho foi o desenvolvimento de uma solução de DW e não comunicar dados institucionais ao público, as duas unidades de observação mostradas receberam codificação alfabética, de “A” e “B”. As análises

apresentadas não são exaustivas e visam apenas contextualizar a aplicação dos dados extraídos do DW.

Figura 19 – Gráfico da Evolução das Taxas Anuais de Evasão e Diplomação da Unidade de Observação “A”



Fonte: autor, com base nos dados armazenados no repositório do DW.

Os gráficos das Figuras 19 e 20 têm dois eixos. O eixo principal mostra os valores, em porcentagem, das taxas de evasão e diplomação no ano. No eixo secundário se tem os valores absolutos do quantitativo de alunos ingressantes, alunos formados e alunos que cancelaram seu vínculo na unidade de observação, por ano letivo. As séries de porcentagem de diplomação no ano começam em 2002 pois são os primeiros anos com diplomados considerando apenas os ingressantes a partir de 1998.

Sobre a unidade de observação “A”, Figura 19, temos, para o período de 1998 a 2018, evasão média anual de 2,9% e diplomação média anual de 10,1%. Para essa unidade de observação as taxas de diplomação foram sempre superiores as de evasão. É possível visualizar dois picos de conclusão dos alunos, um em 2007, com a marca de 15%, e em 2011, que alcançou a maior taxa histórica dessa série com 20% de diplomação. Seguida desse pico histórico se nota uma queda em 2012, onde a taxa de diplomação registrada foi de 8%, chegando-se na menor marca histórica dessa série em 2013, onde a taxa de diplomação anual chegou a 7%. Essa marca foi

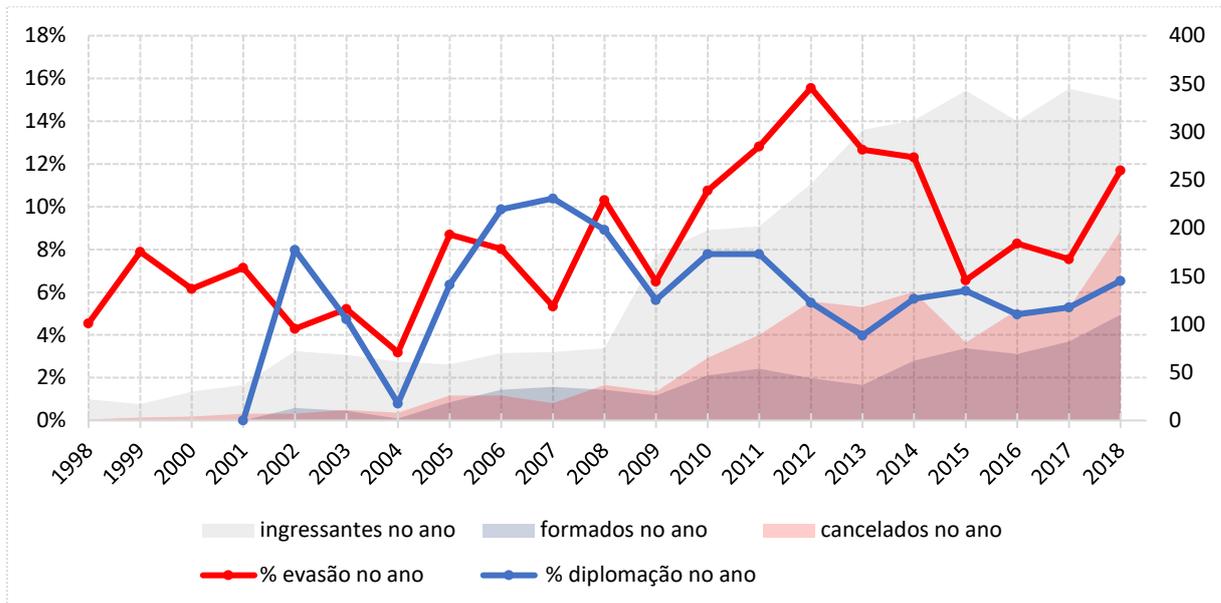
observada novamente em 2015. Parte da explicação para essa queda nas taxas pode vir do aumento expressivo no número de ingressantes devido ao aumento na oferta de vagas proporcionado pelo Programa REUNI. Até o ano de 2010 houve uma oferta média de 310 vagas anuais e esse número passou para 606 em 2011. Consequentemente, os ingressantes de 2011 poderiam figurar como formados somente a partir de 2015, pois os cursos dessa unidade de observação têm tempo ideal para formatura de cinco anos.

É possível notar também que, mesmo após a data de formatura para os ingressantes em 2011, a taxa de diplomação não alcançou o desempenho visto de 2002 a 2011. Isto, aliado a uma taxa constante de evasão que indica que os ingressantes não estão cancelando seu vínculo com a Unifei, é indício de que essa unidade de observação possui alunos em estado de retenção.

Na Unidade de Observação “B”, Figura 20, registrou-se a taxa média de evasão de 8,4% no período de 1998 a 2018 e 6% de taxa de conclusão média no mesmo período. Nessa unidade de observação é possível notar que apenas nos anos de 2002, 2006 e 2007 as taxas de diplomação foram maiores que as taxas de evasão. É possível visualizar dois picos na taxa de diplomação, um em 2002, com 8% e em 2007, que registrou o valor de 13%. O menor valor dessa série histórica foi registrado em 2004, com 0,7% de taxa de diplomação anual.

Analisando a curva da taxa de evasão anual é possível notar uma tendência de aumento a partir de 2009 que culminou com o pico histórico em 2012, que registrou 15,6%. De 2013 a 2015 nota-se tendência de queda e de 2016 a 2018 novamente tendência de aumento. Ademais, ao se observar o número de ingressantes por ano, é possível observar que o aumento dessa métrica foi acompanhada do aumento da taxa de evasão. Isso é indício de que os novos ingressantes não aderiram aos cursos escolhidos.

Figura 20 - Gráfico da Evolução das Taxas Anuais de Evasão e Diplomação da Unidade de Observação “B”



Fonte: autor, com base nos dados armazenados no repositório do DW.

Em comparação à Unidade de Observação “A”, a Unidade de Observação “B” apresenta métricas de diplomação menores e de evasão maiores e isso é indício de que a Unidade de Observação “B” seria mais carente de ação institucional visando a compreensão do fenômeno da evasão e ação visando sua mitigação.

Contudo, é importante notar que, mesmo apresentando métricas superiores de diplomação, a Unidade de Observação “A” ainda não está em sua forma plena. Ao se considerar um cenário ideal onde uma turma de ingressantes não apresenta evasão e retenção e que o curso não têm aumento de vagas para ingressante, seria correto afirmar que, ao final do tempo ideal de formatura, o número de diplomados seria igual ao número de ingressantes de cinco anos atrás. Obviamente que essa não é a realidade de Instituições de Ensino Superior. Contudo, ao se admitir a conclusão do cenário ideal como norteadora de análises, a curva da taxa de diplomação deveria estar o mais próximo possível do número de ingressantes. Isso revela que nessa unidade de observação o foco das ações seria na compreensão dos causadores da retenção discente e a busca por mitigá-las.

5 CONCLUSÃO E SUGESTÕES PARA FUTURAS PESQUISAS

O capítulo final da dissertação tem o objetivo de apresentar resposta encontrada à pergunta de pesquisa, juntamente com as impressões do autor, consolidar sucintamente os resultados alcançados, mostrando a relevância dos achados para a Unifei como também para a comunidade de IES e, finalmente, apresentar sugestões para futuras pesquisas.

O fenômeno da evasão demanda acompanhamento institucional e deve contar com dados e informações não-voláteis e que permitam análises históricas. Para proceder com esse acompanhamento, são necessários dados a respeito das taxas de evasão por curso e dados individuais dos alunos, tanto os acadêmicos quanto os pessoais. Isso habilita a instituição a utilizar esses dados no acompanhamento da evolução das taxas de evasão, na caracterização dos evadidos e identificação de fatores que contribuem para sua ocorrência ou pela explicação dos motivos ou causas da evasão (FREITAS, 2016; MUNIZAGA; CIFUENTES; BELTRAN, 2018). Contudo, o acompanhamento se torna sistêmico se for habilitada na instituição a capacidade de organizar seus dados institucionais que forneçam insumos para apoio nos processos de decisão de forma periódica e contínua. Nesse contexto, a premissa inicial era de que um *data warehouse* (DW) fosse a ferramenta que habilitasse a instituição a alcançar o objetivo pelo acompanhamento sistêmico, periódico e contínuo, servindo de apoio à tomada de decisão institucional sobre a evasão.

O problema de pesquisa dessa dissertação esteve centrado em como dados acadêmicos dos discentes da Unifei, registrado em seu sistema de informação, poderiam ser organizados em um repositório institucional, o DW. Juntamente com esse problema central, questões como a disponibilidade dos dados e a viabilidade do desenvolvimento também foram discutidas.

Conseqüentemente, o objetivo principal foi o desenvolvimento de um DW e os objetivos secundários foram conhecer os métodos de desenvolvimento utilizados para o desenvolvimento de DW, conhecer métodos e métricas utilizadas para o cálculo das taxas de evasão e conhecer as características associadas aos discentes que foram consideradas relevantes nas pesquisas descritivas e causais sobre evasão. Os componentes dos métodos de cálculo das taxas e as variáveis relacionadas aos alunos seriam os requisitos de informação que o DW deveria possuir.

Para alcançar esses objetivos utilizou-se uma proposta metodológica desenvolvida pelo autor embasado na *soft systems methodology* (SSM), proposta por Checkland e Poulter (2010), que guiou a intervenção na realidade prática da Unifei, e um método de desenvolvimento de DW, necessário para guiar a efetiva criação do repositório. Para o desenvolvimento do DW foi escolhido o método de Ralph Kimball (KIMBALL *et al.*, 2008) pois seu foco em se desenvolver DW locais, de menor vulto, e a utilização da técnica de modelagem dimensional em oposição a complexos modelos normalizados, tornam essa abordagem mais acessível à profissionais fora da área de formação de TI (BRESLIN, 2004). A SSM e o método de Kimball possuem sinergia, pois a SSM guia a exploração em situações problemáticas buscando entender e confluir vários pontos de vista sobre o problema, comparando-os como as práticas atualmente desenvolvidas na organização. As informações levantadas na confluência de várias visões de mundo podem guiar a definição dos requisitos de informação que a ferramenta deve possuir para atender ao fim de apoio à tomada de decisão o qual se destina.

As fontes de visões de mundo utilizadas foram autores da área de evasão na busca por compreender os métodos e métricas utilizados para os cálculos das taxas de evasão, bem como para compilar, entre os trabalhos da área, quais foram as características dos alunos evadidos tidas como relevantes ao fenômeno da evasão. Para coletar esses dados recorreu-se à pesquisa bibliográfica. A comparação do resultado da pesquisa bibliográfica com a realidade da Unifei utilizou da pesquisa documental ao banco de dados (BD) do sistema de informação acadêmico em uso, o Sistema de Integrado de Gestão de Atividades Acadêmicas (SIGAA), que mostrou os dados disponíveis para a construção do DW. O produto da interseção entre os resultados da pesquisa bibliográfica e documental geraram os requisitos de informação do DW, primeira etapa para seu desenvolvimento.

Os resultados da pesquisa bibliográfica mostraram que diversos autores buscaram desenvolver métodos para o cálculo das taxas de evasão, contudo o foco do cálculo em determinados coortes ou no cálculo apenas de turmas com gerações completas limitam a capacidade de análise contemporânea dos dados. Assim, notou-se a necessidade do desenvolvimento de uma metodologia própria que mostrasse, por ano, qual a taxa de alunos que deixavam a instituição em relação ao total de matrículas.

Toda a discussão envolvida no desenvolvimento do método de cálculo das taxas de evasão para o DW da Unifei está disponível para utilização pelas IES que não tenham um método próprio. O relacionamento entre as variáveis e equações podem ser utilizadas para guiar o cálculo dessas métricas em suas realidades operacionais. Para instituições que usam o SIGAA, existe uma vantagem adicional, uma vez que podem fazer uso da consulta ao banco de dados mostrado no Apêndice C.

A pesquisa bibliográfica, que buscou identificar quais as variáveis descritivas discentes que foram relevantes nas pesquisas sobre evasão, e a posterior triangulação desses achados com os dados disponíveis na Unifei, revelaram que o SIGAA armazena 23 variáveis (35,4%) do total, sendo 15 variáveis acadêmicas (46,9% das variáveis de seu grupo) e 8 variáveis pessoais (24,2% das variáveis de seu grupo). Visto de outra maneira, o SIGAA permite descrever os alunos evadidos em por meio de 23 variáveis que estão prontas para utilização no DW.

Uma explicação para a diferença é que as pesquisas descritivas e explicativas utilizam questionários específicos para coletar informações de interesse no contexto teórico e metodológico da pesquisa conduzida. Em contrapartida, o SIGAA registra variáveis necessárias para o registro do aluno e a expedição de seu diploma.

A possibilidade de desenvolver uma ferramenta de coleta de novos dados que seja integrada ao modelo de dados do DW proposto é uma recomendação de futuras pesquisas que podem aumentar a abrangência descritiva do DW. Essa integração faz parte da arquitetura de DW mais abrangentes, que utilizam não só dados produzidos pelos sistemas transacionais da organização, mas também fontes externas (SINGH, 2001; KIMBALL *et al.*, 2008).

Os resultados da pesquisa bibliográfica, juntamente com o compartilhamento das fontes, conforme listado no Apêndice E, se constitui em um levantamento que poderá ser utilizado por outras IES como uma fonte comparativa para suas próprias fontes de dados. Para IES que utilizam o SIGAA existe a vantagem adicional que a triangulação proposta trouxe ao se identificar a fonte dos dados no banco de dados do SIGAA, conforme mostrado na Tabela 3 e no Apêndice A.

As pesquisas bibliográficas, tanto para a condução da revisão teórica como para coleta de dados, e a pesquisa documental no banco de dados do SIGAA permitiram que os objetivos secundários fossem alcançados. Eles são o

embasamento para o alcance do objetivo principal. O objetivo primário de responder à pergunta de pesquisa demandou que fossem postas em prática as etapas de desenvolvimento inerentes ao método de desenvolvimento escolhido, o de Ralph Kimball.

O levantamento dos requisitos do DW é a primeira etapa do desenvolvimento e foi derivada das triangulações nos dados das pesquisas bibliográfica e documental. Assim foi possível determinar quais são os dados que o DW deve ter para o efetivo apoio à decisão sobre evasão na Unifei.

O SIGAA tem capacidade de fornecer dados para que o ambiente de consulta do DW consolide dados sobre ingressantes, matrículas ativas, cancelados e as respectivas taxas de evasão para os 34 cursos presenciais da instituição, datando de 1998. Os benefícios percebidos no desenvolvimento e adoção dessa ferramenta foram: (1) os dados históricos sobre evasão foram centralizados em um único repositório e estão disponíveis para consulta pelos gestores; (2) definição institucional sobre como calcular as taxas de evasão contribuem para que se tenha apenas uma versão desse fato; (3) dados referentes à características pessoais e acadêmicas dos alunos estão disponíveis para futuras análises descritivas e explicativas ou levantamento de fatores associados à evasão no contexto da Unifei. As evidências trazidas pelo resultado do desenvolvimento do DW embasado nas pesquisas bibliográficas e documental conduzidas, apoiam a premissa de que essa ferramenta pode ser a resposta institucional para o acompanhamento sistêmico da evasão. É importante ressaltar, contudo, que a Unifei é uma instituição com cursos predominantemente da área de Engenharia e conta hoje com aproximadamente 7 mil alunos. Assim, considerações sobre a área de conhecimento e porte de outras instituições que visem utilizar os conceitos e modelos discutidos nesse trabalho devem ser realizadas.

O foco inicial dessa dissertação foi nos conceitos estruturantes necessários ao desenvolvimento da solução, com foco nos requisitos de informação, fonte de dados, modelagem de dados lógica e física e a construção de um protótipo. O protótipo foi validado pelo Conselho de Ensino, Pesquisa, Extensão e Administração (CEPEAd) e pelo Pró-Reitoria de Graduação. A principal limitação desse trabalho é a implementação do DW nos servidores institucionais da Unifei, ou seja, por em produção o protótipo desenvolvido para que essa iniciativa seja um ativo institucional.

Para tal, tarefas de *extract, transform and load* (ETL) incrementais serão necessárias, para a cada ano atualizar as informações institucionais no DW. nos moldes dos sistemas de *change data capture*, discutidos em Kimball *et al.* (2008), que exigem esforço de desenvolvimento e programação adicionais.

Outro esforço adicional de desenvolvimento seria integrar ao DW proposto os demais indicadores de desempenho propostos pela Comissão Especial do MEC (1997), como o número de diplomados por ano, o número de alunos em situação de retenção e suas respectivas taxas proporcionais ao total de alunos no ano de referência. Outra métrica que seria pertinente de ser desenvolvida é estabelecer um número de diplomação esperada. É possível desenvolver uma métrica bruta, que consideraria o total de ingressantes do ano de ingresso da turma e uma métrica líquida, que descontaria dessa massa de ingressantes os evadidos até o ano de referência. O acompanhamento dessas métricas em comparação com as taxas de diplomação anuais aferidas poderiam ser um instrumento adicional de monitoramento da qualidade do processo formativo da instituição.

Finalmente, sugere-se que as informações descritivas sobre os alunos, tanto as variáveis pessoais quanto acadêmicas que compõem o DW, sejam utilizadas para a identificação de fatores causais da evasão nos diferentes cursos da Unifei.

REFERÊNCIAS

- ALJAWARNEH, I. M. Design of a data warehouse model for decision support at higher education: A case study. *Information Development*, **Cascais**, Portugal, v. 32, n. 5, p. 1691-1706, 2016.
- ARIYACHANDRA, T.; WATSON, H. J. Key organizational factors in data warehouse architecture selection. **Decision Support Systems**, Vol. 49, Issue 2. Pages 200-212. 2010.
- ARIYACHANDRA, T.; WATSON, H. J. Which Data Warehouse Architecture is Best? **Communications of the ACM**, v. 51, n. 10, p. 1. 2008.
- ARIYACHANDRA, T.; WATSON, H. J. Which Data Warehouse Architecture Is Most Successful? **Business Intelligence Journal**, v. 11, n. 1, p. 4, 2006.
- BARBIERI, C. BI2 - Business Intelligence: Modelagem & Qualidade. Elsevier Editora, 2011.
- BIPLOB, B.; SHERAJI, G. A.; KHAN, S. I. Comparison of Different Extraction Transformation and Loading Tools for Data Warehousing. **International Conference on Innovations in Science, Engineering and Technology (ICISSET)**, 262-267. 2018.
- BRESLIN, M. Data Warehousing Battle of the Giants: Comparing the Basics of the Kimball and Inmon Models. **Business Intelligence Journal**, winter 2004, p. 6, 2004.
- CALUMBY, R. T.; CRAVEIRO, J. Estudo de Ferramentas livres para OLAP e proposição de um ambiente de Data Warehousing de baixo custo utilizando uma dessas ferramentas e o SGBD PostgreSQL. In: VI Workshop de Trabalhos de Iniciação Científica e Graduação Bahia, Alagoas e Sergipe, VIII ERBASE, 2008, Salvador-BA, 2008.
- CANUTO, J. G. **Análise dos Processos de Gerenciamento dos Programas de Iniciação Científica da Universidade Federal de Itajubá: um Estudo de Caso**. 2018. 288 f. Dissertação (Mestrado Profissional em Administração) – Programa de Pós-Graduação em Administração da Universidade Federal de Itajubá, Itajubá, 2018.
- CHECKLAND, P. Soft systems methodology: a thirty-year retrospective. **Systems Research and Behavioral Science**, Volume 17, Issue S1, November 2000, p. S11-S58, 2000.
- CHECKLAND, P.; POULTER, J. Soft Systems Methodology. In: Reynolds, M., Holwell, S. (Eds.), **Systems Approaches to Managing Change: A Practical Guide**. Springer, London, pp. 191–242. 2010.

CISLAGHI, R.; LUZ FILHO, S. S. da. Um framework para a promoção da permanência discente no ensino de graduação e um modelo de sistema de gestão do conhecimento para IES brasileiras. In: Colóquio Internacional Sobre Gestão Universitária na América do Sul, 9., 2009, Florianópolis. **Anais...** Florianópolis: UFSC, 2009.

COORDENAÇÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR (CAPES). **Nota Qualis-Periódicos**. 2019. Disponível em: <<https://sucupira.capes.gov.br/sucupira/>>. Acesso em 30 nov. 2019.

COGHLAN, D.; BRYDON-MILLER, M. (Ed.). **The SAGE encyclopedia of action research**. Sage, 2014.

DAVOK, D. F.; BERNARD, R. P. Avaliação dos índices de evasão nos cursos de graduação da Universidade do Estado de Santa Catarina - UDESC. **Avaliação (Campinas), Sorocaba**, v. 21, n. 2, p. 503-522, julho 2016.

DEBROY, V.; BRIMBLE, L.; YOST, M. NewTL: engineering an extract, transform, load (ETL) software system for business on a very large scale. In **Proceedings of the 33rd Annual ACM Symposium on Applied Computing (SAC '18)**. 2018.

DICKENS, L.; WATKINS, K. Action research: rethinking Lewin. **Management Learning**, v. 30, n. 2, p. 127-140, 1999.

DURSO, S. de O.; CUNHA, J. V. A. da. Determinant Factors For Undergraduate Student's Dropout In An Accounting Studies Department Of A Brazilian Public University. **Educ. rev.**, Belo Horizonte, v. 34, 2018.

FERRÃO, M. E.; ALMEIDA, L. S. Multilevel modeling of persistence in higher education. **Ensaio: Aval. Pol. Públ. Educ.**, Rio de Janeiro, v. 26, n. 100, p. 664-683, julho 2018.

FREITAS, R. S. A ocorrência da evasão do ensino superior: uma análise das diferentes formas de mensurar. **Dissertação (mestrado)** - Universidade Estadual de Campinas, Faculdade de Educação, Campinas, SP. 2016.

GERHARDT, T. E.; SILVEIRA, D. T. (organizadoras). **Métodos de Pesquisa**. 1. ed. Porto Alegre: Editora da UFRGS, 2009

GIL, A. C. **Métodos e técnicas de pesquisa social**. 3ª ed. São Paulo: Atlas, 2008.

HEUSER, C. A. **Projeto de banco de dados**: Volume 4 da Série Livros didáticos informática UFRGS. Bookman Editora, 2009.

INMON, W. H. **Building the Data Warehouse**, 4th ed. Wiley. 2005.

INSTITUTE FOR TELECOMMUNICATION SCIENCES. **The research laboratory of the National Telecommunications and Information Administration**. 1996.

Disponível em: <https://www.its.bldrdoc.gov/fs-1037/dir-010/_1460.htm>. Acesso em: janeiro de 2020.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA (INEP). **Indicadores financeiros educacionais – Brasil 2000-2014**.

Brasília: INEP, 2018. Disponível em:

http://download.inep.gov.br/informacoes_estatisticas/

[investimentos_publicos_em_educacao/indicadores_financeiros_educacionais/investimento_estudante_valor_real.xlsx](http://download.inep.gov.br/informacoes_estatisticas/investimentos_publicos_em_educacao/indicadores_financeiros_educacionais/investimento_estudante_valor_real.xlsx). Acesso em 20 de maio de 2018.

JUKIC, N. Modeling Strategies and Alternatives for Data Warehousing Projects.

Communications of the ACM. Vol. 49, No. 4. Pages 36-89. 2006.

KIMBALL, R. *et al.* **The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing, and Deploying Data Warehouses**, 2nd ed. Wiley. 2008.

KIPNIS, B. A pesquisa institucional e a educação superior brasileira: um estudo de caso longitudinal da evasão. **Linhas Críticas, Brasília**, v. 6, n. 11, p. 17- 32, jul./dez. 2000.

LAMERS, J. M. de S.; SANTOS, B. S. dos; TOASSI, R. F. C. Retenção e evasão no ensino superior público: estudo de caso em um curso noturno de odontologia. **Educ. rev., Belo Horizonte**, v. 33, 2017.

LAUDON, K.; LAUDON, J. **Sistemas de Informação Gerenciais**. São Paulo: Pearson Prentice Hall, 2014.

MALHOTRA, N. **Pesquisa de marketing**. 3ª ed. Porto Alegre: Bookman, 2001.

MARTINS, G. de A.; THEÓPHILO, C. R. **Metodologia da investigação científica para ciências sociais aplicadas**. 3. ed. São Paulo: Atlas, 2016.

MASSI, L.; VILLANI, A.. Um caso de contratendência: baixa evasão na licenciatura em química explicada pelas disposições e integrações. **Educ. Pesqui., São Paulo**, v. 41, n. 4, p. 975-992, Dec. 2015.

MENDES JUNIOR, Á. A. F. Uma análise da progressão dos alunos cotistas sob a primeira ação afirmativa brasileira no ensino superior: o caso da Universidade do Estado do Rio de Janeiro. **Ensaio: Aval. Pol. Públ. Educ.**, Rio de Janeiro, v. 22, n. 82, p. 31-56, Mar. 2014.

MINGERS, J.; TAYLOR, S. The Use of Soft Systems Methodology in Practice, **Journal of the Operational Research Society**, 43:4, 321-332. 1992.

MINISTÉRIO DA EDUCAÇÃO E CULTURA (MEC). Comissão Especial de Estudos Sobre a Evasão nas Universidades Públicas Brasileiras. **Diplomação, Retenção e Evasão nos Cursos de Graduação em Instituições de Ensino Superior Públicas**. SESu, MEC, Brasília, 1997. 134 p.

MUNIZAGA, F., CIFUENTES, M., BELTRAN, A. Retención y abandono estudiantil en la Educación Superior Universitaria en América Latina y el Caribe: Una revisión sistemática. **Archivos Analíticos de Políticas Educativas**, 26(61). 2018.

NODARI, D. E.; LIMA, E. G. dos S.; MACIEL, C. E. O desempenho dos estudantes no vestibular e a permanência nos cursos de graduação da UNEMAT. **Avaliação (Campinas), Sorocaba**, v. 23, n. 2, p. 312-329, outubro de 2018.

PALHARINI, F. A. Evasão, exclusão e gestão acadêmica na UFF: passado, presente e futuro. **Cadernos do ICHF**. Série Estudos e Pesquisas (UFF), v. 95, p. 1-62, 2010.

POSTGRESQL. Documentation. 2019. Disponível em: <<https://www.postgresql.org/docs/12/index.html>>. Acesso em 30 nov. 2018.

RANGARAJAN, S. Data Warehouse Design – Inmon versus Kimball. **The Data Administration Newsletter**. 2016. Disponível em: <<http://tdan.com/data-warehouse-design-inmon-versus-kimball/20300>> Acesso em 12/08/2018.

SANTANA, O. A. Evasão nas Licenciaturas das Universidades Federais: entre a apetência e a competência. **Educação (UFSM)**, Santa Maria, p. 311-328, jul. 2016.

SILVA FILHO, R. L. L.; LOBO, M. B. Como A Mudança Na Metodologia Do Inep Altera O Cálculo Da Evasão. In: **Instituto Lobo**. 2012.

SILVA FILHO, R. L. L. A evasão no ensino superior brasileiro: novos dados. In: **Instituto Lobo**. 2017.

SILVA FILHO, R. L. L. *et al.* A evasão no ensino superior brasileiro. **Cadernos de Pesquisa**, São Paulo, v. 37, n. 132, p. 641-659, dezembro. 2007.

SILVA, E. L. da; MENEZES, E. M. **Metodologia da pesquisa e elaboração de dissertação**: 4ª edição revisada e atualizada. 4. ed. Florianópolis: UFSC, 2005.

SILVA, F. I. C. da *et al.* Evasão escolar no curso de educação física da Universidade Federal do Piauí. **Avaliação (Campinas), Sorocaba**, v. 17, n. 2, p. 391-404, julho 2012.

SILVA, R. R. de C. M. da; MAINIER, F. B.; PASSOS, F. B. A contribuição da disciplina de introdução à engenharia química no diagnóstico da evasão. **Ensaio: Aval. Pol. Públ. Educ.**, Rio de Janeiro, v. 14, n. 51, p. 261-277, 2006.

SINGH, Harry. **Data Warehouse. Conceitos, tecnologias, implementação e gerenciamento**. São Paulo: Makron Books, 2001.

SUPERINTENDÊNCIA DE INFORMÁTICA DA UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE (SINFO). **Soluções em Sistemas**. 2019. Disponível em: <https://info.ufrn.br/> Acesso em 12/12/2019.

TINTO, V. Dropout from higher education: a theoretical synthesis of recent research. **Review of Educational Research**, Washington, v. 45, n.1, p.89-125, 1975.

TONTINI, G.; WALTER, S. A. Pode-se identificar a propensão e reduzir a evasão de alunos? Ações estratégicas e resultados táticos para instituições de ensino superior. **Avaliação (Campinas), Sorocaba**, v. 19, n. 1, p. 89-110, março. 2014.

TRIBUNAL DE CONTAS DA UNIÃO (TCU). **Orientações para o cálculo dos indicadores de gestão**: decisão n.º 408/2002 – Plenário e Acórdãos n.º 1043/2006 e n.º 2167/2006 – Plenário. Brasília – DF, 2009.

UNIVERSIDADE FEDERAL DE ITAJUBÁ (UNIFEI). **Ata da 27ª Reunião Ordinária de 2019 do Conselho de Ensino, Pesquisa, Extensão e Administração - CEPEAd**. Itajubá, MG. 2019a.

UNIVERSIDADE FEDERAL DE ITAJUBÁ (UNIFEI). **Plano de Desenvolvimento Institucional 2015-2018**. Itajubá, MG. 2014a.

UNIVERSIDADE FEDERAL DE ITAJUBÁ (UNIFEI). **Regimento da Administração Central**. Itajubá, MG. 2019e.

UNIVERSIDADE FEDERAL DE ITAJUBÁ (UNIFEI). **Relatório de Gestão do Exercício de 2015**. Itajubá, MG. 2016.

UNIVERSIDADE FEDERAL DE ITAJUBÁ (UNIFEI). **Relatório de Gestão do Exercício de 2016**. Itajubá, MG. 2017.

UNIVERSIDADE FEDERAL DE ITAJUBÁ (UNIFEI). **Relatório de Gestão do Exercício de 2017**. Itajubá, MG. 2018.

UNIVERSIDADE FEDERAL DE ITAJUBÁ (UNIFEI). **Relatório de Gestão do Exercício de 2018**. Itajubá, MG. 2019b.

UNIVERSIDADE FEDERAL DE ITAJUBÁ (UNIFEI). **Unifei em números**. 2019f. Disponível em: <<https://unifei.edu.br/unifei-em-numeros/>>. Acesso em 20/02/2020.

VANZ, S. A. S. *et al.* Evasão e retenção no curso de Biblioteconomia da UFRGS. **Avaliação (Campinas), Sorocaba**, v. 21, n. 2, p. 541-568, julho 2016.

VERGARA, S. C. Tipos de Pesquisa em Administração. **Cadernos EBAP**, n. 52, 1990

WEINBERG, P. Top Python ETL Tools (aka Airflow Vs The World). **Panoply Blog**. 2019. Disponível em: <<https://blog.panoply.io/top-9-python-etl-tools-and-when-to-use-them>>. Acesso em 01/09/2019.

YUNG, M. Using Python Script for Data ETL. **Codeburst**. 2018. Disponível em: <<https://codeburst.io/using-python-script-for-data-etl-53138c567906>>. Acesso em 01/09/2019.

ZEA, O. M.; GUALTOR, J. P.; MORA, S. L. A Holistic View of Data Warehousing in Education. in **IEEE Access**, vol. 6, pp. 64659-64673, 2018.

ZEA, O. M.; SAMPEDRO, A.; MORA, S. L. Datawarehouse design for educational data mining. **15th International Conference on Information Technology Based Higher Education and Training (ITHET)**. 1-6. 2016

ZIKMUND, W. G *et al.* **Business Research Methods**. 6th Edition, The Dryden Press, Fort Worth. 2000.

APÊNDICE A – DESCRIÇÃO DAS TABELAS E ATRIBUTOS DO DW

Tabela DW	Campo DW	Fonte no SIGAA	Join entre tabelas no SIGAA	Descrição no DW	Tipo de Dados no DW
dimensao_pessoa	chave_pessoa	*	-	Chave primária	int4 serial not null
dimensao_pessoa	matricula	public. discente.matricula	-	Matrícula do aluno	int(8)
dimensao_pessoa	cpf	comum. pessoa.cpf_cnpj	-	CPF do aluno	int(8)
dimensao_pessoa	nome	comum. pessoa.nome	-	Nome do aluno	varchar(200)
dimensao_pessoa	sexo	comum. pessoa.sexo	-	Sexo do aluno: 'M' para masculino e 'F' para feminino	char(1)
dimensao_pessoa	estado_civil	comum. estado_civil.descricao	from comum.pessoa p join comum.estado_civil ec on (p.id_estadocivil = ec.id_estado_civil)	Estado civil do aluno	varchar(80)
dimensao_pessoa	nacionalidade	comum. pais.nacionalidade	from comum.pessoa p join comum.pais pa on (p.id_pais_nacionalidade = pa.id_pais)	Nacionalidade do aluno	varchar(80)
dimensao_pessoa	nacionalidade_pais	comum. pais.nome	from comum.pessoa p join comum.pais pa on (p.id_pais_nacionalidade = pa.id_pais)	Nome do país de nacionalidade do aluno	varchar(80)
dimensao_pessoa	naturalidade_cidade	comum. municipio.nome	from comum.pessoa p join comum.municipio m on (p.id_municipio_naturalidade = m.id_municipio)	Cidade de nascimento do aluno	varchar(80)
dimensao_pessoa	naturalidade_estado	comum. unidade_federativa.sigla	from comum.pessoa p join comum.municipio m on (p.id_municipio_naturalidade = m.id_municipio) join comum.unidade_federativa uf on (m.id_unidade_federativa = uf.id_unidade_federativa)	Sigla da UF de nascimento do aluno	char(2)
dimensao_pessoa	data_nascimento	comum. pessoa.data_nascimento	-	Data de nascimento do aluno	date
dimensao_pessoa	raca	comum. tipo_raca.descricao	from comum.pessoa p join comum.tipo_raca tr on (p.id_raca = tr.id_tipo_raca)	Raça autodeclarada do aluno	varchar(80)
dimensao_pessoa	ensino_medio_escola	comum. pessoa.segundograuescola	-	Escola na qual o aluno terminou o ensino médio	varchar(150)

Tabela DW	Campo DW	Fonte no SIGAA	Join entre tabelas no SIGAA	Descrição no DW	Tipo de Dados no DW
dimensao_pessoa	ensino_medio_ano_concl	comum.pessoa.segundograuanoconclusao	-	Ano de conclusão do ensino médio	varchar(4)
dimensao_pessoa	ensino_medio_tipo_escola	comum.pessoa.segundo_grau_tipo_escola	-	Indica se a escola do ensino médio é pública ou privada	varchar(10)
dimensao_pessoa	forma_ingresso	ensino.forma_ingresso.descricao	from public.discente d join ensino.forma_ingresso fi on (d.id_forma_ingresso = fi.id_forma_ingresso)	Forma de ingresso do aluno no curso	varchar(80)
dimensao_pessoa	ano_ingresso	public.discente.ano_ingresso	-	Ano de ingresso do aluno	int(4)
dimensao_pessoa	periodo_ingresso	public.discente.periodo_ingresso	-	Período letivo de ingresso do aluno	int(2)
dimensao_tempo	chave_tempo	*	-	Chave primária	int4 serial not null
dimensao_tempo	ano_letivo	*	-	Ano letivo de referência	int(4)
dimensao_tempo	periodo_letivo	*	-	Período letivo de referência	int(4)
dimensao_curso	chave_curso	*	-	Chave primária	int4 serial not null
dimensao_curso	curso_id	public.curso.id_curso	from public.discente d join public.curso c on (d.id_curso = c.id_curso)	Id do curso do aluno	Int
dimensao_curso	curso_nome	public.curso.nome	from public.discente d join public.curso c on (d.id_curso = c.id_curso)	Nome do curso do aluno	varchar(200)
dimensao_curso	curso_ua	comum.unidade.nome	from public.discente d join public.curso c on (d.id_curso = c.id_curso) join comum.unidade u on (c.id_unidade = u.id_unidade)	Nome da unidade acadêmica do curso do aluno	varchar(200)
dimensao_curso	curso_turno	public.curso.id_turno	-	Turno de oferta do curso: 'M' para manhã, 'T' para tarde e 'N' para noite	varchar(3)

Tabela DW	Campo DW	Fonte no SIGAA	Join entre tabelas no SIGAA	Descrição no DW	Tipo de Dados no DW
dimensao_curso	curso_area_capes	comum. area_conhecimento_cnpq. nome	from public.curso c join comum.area_conhecimento_cnpq acc on (c.id_area_curso = acc.id_area_conhecimento_cnpq)	Área de conhecimento CNPQ do curso do aluno	varchar(200)
dimensao_curso	modalidade_ensino	comum. modalidade_educacao. descricao	from public.curso c join comum.modalidade_educacao me on (c.id_modalidade_educacao = me.id_modalidade_educacao)	Modalidade de ensino do aluno, se presencial ou à distância	varchar(80)
dimensao_curso	campus	comum. campus_ies.nome	from public.curso c join comum.campus_ies ci on (c.id_campus = ci.id_campus)	Campus da Unifei do qual o curso do aluno faz parte	varchar(120)
dimensao_curso	grau_academico	ensino. grau_academico.descricao	from public.curso c join ensino.grau_academico ga on (c.id_grau_academico = ga.id_grau_academico)	Grau acadêmico do curso ao qual o aluno está vinculado (bacharelado ou licenciatura)	varchar(80)
dimensao_situacao	chave_situacao	*	-	Chave primária	int4 serial not null
dimensao_situacao	situacao_aluno	public. status_discente.descricao	from public.discente d join public.status_discente sd on (d.status = sd.status)	Situação do aluno perante o curso (ativo, cancelado, trancado etc.)	varchar(20)
dimensao_situacao	tipo_saida	ensino. tipo_movimentacao_aluno.de sricao	from public.discente d join ensino.movimentacao_aluno ma on (d.id_discente = ma.id_discente) join ensino.tipo_movimentacao_aluno tma on (ma.id_tipo_movimentacao_aluno = tma.id_tipo_movimentacao_aluno)	Descreve o tipo de saída do aluno da instituição (desistência, transferência, cancelamento para novo vestibular etc.)	varchar(120)

Tabela DW	Campo DW	Fonte no SIGAA	Join entre tabelas no SIGAA	Descrição no DW	Tipo de Dados no DW
dimensao_situacao	ano_saida	ensino.movimentacao_aluno.ano_ocorrencia	from public.discente d join ensino.movimentacao_aluno ma on (d.id_discente = ma.id_discente)	Ano de ocorrência da movimentação de saída do aluno	int(4)
dimensao_situacao	periodo_saida	ensino.movimentacao_aluno.periodo_ocorrencia	from public.discente d join ensino.movimentacao_aluno ma on (d.id_discente = ma.id_discente)	Período letivo de ocorrência da movimentação de saída do aluno	int(1)
fato_discente	chave_discente	*	-	Chave primária	int4 serial not null
fato_discente	chave_tempo	*	-	Chave estrangeira, dimensao_tempo	int4 serial not null
fato_discente	chave_pessoa	*	-	Chave estrangeira, dimensao_pessoa	int4 serial not null
fato_discente	chave_curso	*	-	Chave estrangeira, dimensao_curso	int4 serial not null
fato_discente	chave_situacao	*	-	Chave estrangeira, dimensao_situacao	int4 serial not null
fato_discente	indice_mc	ensino.indice_academico_discente.v_alor	from public.discente d join ensino.indice_academico_discente iad on (d.id_discente = iad.id_discente) join ensino.indice_academico ia on (iad.id_indice_academico = ia.id) where iad.id = 1	Índice média de conclusão (MC) do aluno	numeric(9,4)
fato_discente	indice_iepl	ensino.indice_academico_discente.v_alor	from public.discente d join ensino.indice_academico_discente iad on (d.id_discente = iad.id_discente) join ensino.indice_academico ia on (iad.id_indice_academico = ia.id) where iad.id = 5	Índice de eficiência em períodos letivos (IEPL) do aluno	numeric(9,4)

Tabela DW	Campo DW	Fonte no SIGAA	Join entre tabelas no SIGAA	Descrição no DW	Tipo de Dados no DW
fato_discente	indice_iech	ensino. indice_academico_discente.v alor	from public.discente d join ensino.indice_academico_discente iad on (d.id_discente = iad.id_discente) join ensino.indice_academico ia on (iad.id_indice_academico = ia.id) where iad.id = 4	Índice de eficiência em carga horária (IECH) do aluno	numeric(9,4)
fato_discente	indice_ira	ensino. indice_academico_discente.v alor	from public.discente d join ensino.indice_academico_discente iad on (d.id_discente = iad.id_discente) join ensino.indice_academico ia on (iad.id_indice_academico = ia.id) where iad.id = 2	Índice de rendimento acadêmico (IRA) do aluno	numeric(9,4)
fato_discente	indice_iea	ensino. indice_academico_discente.v alor	from public.discente d join ensino.indice_academico_discente iad on (d.id_discente = iad.id_discente) join ensino.indice_academico ia on (iad.id_indice_academico = ia.id) where iad.id = 6	Índice de eficiência acadêmica (IEA) do aluno	numeric(9,4)
fato_evasao	chave_evasao	*	-	Chave primária	Int(4) serial not null
fato_evasao	chave_tempo	*	-	Chave estrangeira, dimensao_tempo	Int(4) serial not null
fato_evasao	chave_curso	*	-	Chave estrangeira, dimensao_curso	Int(4) serial not null
fato_evasao	total_ ingressantes_ ano	public. discente **	-	Total de alunos ingressantes em determinado ano	Int
fato_evasao	total_evadidos_a no	ensino. movimentacao_aluno **	-	Total de alunos evadidos em determinado ano	Int

Tabela DW	Campo DW	Fonte no SIGAA	Join entre tabelas no SIGAA	Descrição no DW	Tipo de Dados no DW
fato_evasao	veteranos_prox_ano	ensino.movimentacao_aluno **	-	Total de alunos veteranos que poderão realizar matrícula no ano seguinte	Int
fato_evasao	total_matriculas_ano	**	-	Total de matrículas em determinado ano	Int
fato_evasao	taxa_evasao	**	-	Taxa de evasão em determinado ano	numeric(9, 2)

Fonte: autor, com base na pesquisa documental realizada no SIGAA.

* Campo nativo do DW

** Campo derivado de consulta no SIGAA e tratado na ETL (ver apêndices C e D)

APÊNDICE B – SCRIPT EM SQL PARA CRIAÇÃO DAS TABELAS DO DW

```

CREATE SCHEMA dimensional;

CREATE SEQUENCE dimensional.chave_pessoa;
CREATE TABLE dimensional.dimensao_pessoa(
  chave_pessoa int default nextval('dimensional.chave_pessoa'::regclass)
PRIMARY KEY,
  cpf int(8),
  matricula int(8),
  nome varchar(200),
  sexo char(1),
  estado_civil varchar(80),
  nacionalidade varchar(80),
  nacionalidade_pais varchar(80),
  naturalidade_cidade varchar(80),
  naturalidade_estado char(2),
  data_nascimento date,
  raca varchar(80),
  ensino_medio_escola varchar(150),
  ensino_medio_ano_concl varchar(4),
  ensino_medio_tipo_escola varchar(10),
  forma_ingresso varchar(80),
  ano_ingresso int(4),
  periodo_ingresso int(2)
);

CREATE SEQUENCE dimensional.chave_tempo;
CREATE TABLE dimensional.dimensao_tempo(
  chave_tempo int default nextval('dimensional.chave_tempo'::regclass)
PRIMARY KEY,
  ano_letivo int,
  periodo_letivo int
);

CREATE SEQUENCE dimensional.chave_curso;
CREATE TABLE dimensional.dimensao_curso(
  chave_curso int default nextval('dimensional.chave_curso'::regclass)
PRIMARY KEY,
  curso_id int,
  curso_nome varchar(200),
  curso_ua varchar(200),
  curso_turno varchar(3),
  curso_area_capes varchar(200),
  modalidade_ensino varchar(80),
  campus varchar(120),
  grau_academico varchar(80)
);

CREATE SEQUENCE dimensional.chave_situacao;
CREATE TABLE dimensional.dimensao_situacao(
  chave_situacao int default
nextval('dimensional.chave_situacao'::regclass) PRIMARY KEY,
  situacao_aluno varchar(20),
  tipo_saida varchar(120),
  ano_saida int(4),
  periodo_saida int(1)
);

```

```
CREATE SEQUENCE dimensional.chave_discente;
CREATE TABLE dimensional.fato_discente(
  chave_discente int default
nextval('dimensional.chave_discente'::regclass) PRIMARY KEY,
  chave_pessoa int references dimensional.dimensao_pessoa (chave_pessoa),
  chave_curso int references dimensional.dimensao_curso (chave_curso),
  chave_situacao int references dimensional.dimensao_situacao
(chave_situacao),
  chave_tempo int references dimensional.dimensao_tempo (chave_tempo),
  indice_mc numeric(9,4),
  indice_iepl numeric(9,4),
  indice_iech numeric(9,4),
  indice_iea numeric(9,4),
  indice_ira numeric(9,4)
);

CREATE SEQUENCE dimensional.chave_evasao;
CREATE TABLE dimensional.fato_evasao(
  chave_evasao int default nextval('dimensional.chave_evasao'::regclass)
PRIMARY KEY,
  chave_curso int references dimensional.dimensao_curso (chave_curso),
  chave_tempo int references dimensional.dimensao_tempo (chave_tempo),
  total_ingressantes_ano int,
  total_evadidos_ano int,
  veteranos_proximo_ano int,
  total_matriculas_ano int,
  taxa_evasao numeric(9,2)
);
```

APÊNDICE C – SCRIPT EM SQL PARA EXTRAÇÃO DOS DADOS PARA A TABELA FATO DISCENTE E DIMENSÕES PESSOA, CURSO, SITUAÇÃO

```

select
  d.matricula as matricula,
  p.cpf_cnpj as cpf,
  p.nome,
  p.sexo,
  ec.descricao as estado_civil,
  pa.nacionalidade,
  pa.nome as nacionalidade_pais,
  m.nome as naturalidade_cidade,
  uf.sigla as naturalidade_estado,
  p.data_nascimento as data_nascimento,
  tr.descricao as raca,
  p.segundograuescola as ensino_medio_escola,
  p.segundograuanoconclusao as ensino_medio_ano_concl,
  case
    when p.segundo_grau_tipo_escola = 1 then 'PÚBLICA'
    when p.segundo_grau_tipo_escola = 2 then 'PRIVADA'
    else null
  end as ensino_medio_tipo_escola,
  fi.descricao as forma_ingresso,
  d.ano_ingresso,
  d.periodo_ingresso,
  c.id_curso as curso_id,
  c.nome as curso_nome,
  u.nome as curso_ua,
  c.id_turno as curso_turno,
  acc.nome as curso_area_capes,
  me.descricao as modalidade_ensino,
  ci.nome as campus,
  ga.descricao as grau_academico,
  sd.descricao as situacao_aluno,
  tma.descricao as tipo_saida,
  ma.ano_ocorrencia as ano_saida,
  (select distinct on (iadh.id_discente)
    iadh.valor
  from ensino.indice_academico_discente_historico iadh
  join public.discente d2 on iadh.id_discente = d2.id_discente
  where d2.id_discente = d.id_discente
  and iadh.id_indice_academico = 1
  order by iadh.id_discente, iadh.data_calculo desc
) as indice_mc,
  (select distinct on (iadh.id_discente)
    iadh.valor
  from ensino.indice_academico_discente_historico iadh
  join public.discente d2 on iadh.id_discente = d2.id_discente
  where d2.id_discente = d.id_discente
  and iadh.id_indice_academico = 5
  order by iadh.id_discente, iadh.data_calculo desc
) as indice_iepl,
  (select distinct on (iadh.id_discente)
    iadh.valor
  from ensino.indice_academico_discente_historico iadh
  join public.discente d2 on iadh.id_discente = d2.id_discente
  where d2.id_discente = d.id_discente

```

```

        and iadh.id_indice_academico = 4
        order by iadh.id_discente, iadh.data_calculo desc
    ) as indice_iech,
    (select distinct on (iadh.id_discente)
     iadh.valor
     from ensino.indice_academico_discente_historico iadh
     join public.discente d2 on iadh.id_discente = d2.id_discente
     where d2.id_discente = d.id_discente
     and iadh.id_indice_academico = 2
     order by iadh.id_discente, iadh.data_calculo desc
    ) as indice_ira,
    (select distinct on (iadh.id_discente)
     iadh.valor
     from ensino.indice_academico_discente_historico iadh
     join public.discente d2 on iadh.id_discente = d2.id_discente
     where d2.id_discente = d.id_discente
     and iadh.id_indice_academico = 6
     order by iadh.id_discente, iadh.data_calculo desc
    ) as indice_iea
from public.discente d
join comum.pessoa p on d.id_pessoa = p.id_pessoa
join comum.estado_civil ec on p.id_estado_civil = ec.id_estado_civil
left join ensino.movimentacao_aluno ma on
    (ma.id_discente = d.id_discente and ma.id_movimentacao_aluno in
     (select id_movimentacao_aluno
      from ensino.movimentacao_aluno ma2
      where ma2.id_discente = d.id_discente and d.status in (3,6) --
cancelado e concluído
      order by ma2.ano_referencia desc, ma2.periodo_referencia desc,
data_ocorrencia desc
      limit 1
     ) -- ultima movimentacao do discente
    )
left join ensino.tipo_movimentacao_aluno tma on
ma.id_tipo_movimentacao_aluno = tma.id_tipo_movimentacao_aluno
join public.curso c on d.id_curso = c.id_curso
join comum.modalidade_educacao me on c.id_modalidade_educacao =
me.id_modalidade_educacao
join comum.campus_ies ci on c.id_campus = ci.id_campus
join ensino.grau_academico ga on c.id_grau_academico = ga.id_grau_academico
join comum.pais pa on p.id_pais_nacionalidade = pa.id_pais
join comum.municipio m on p.id_municipio_naturalidade = m.id_municipio
join comum.unidade_federativa uf on m.id_unidade_federativa =
uf.id_unidade_federativa
join comum.tipo_raca tr on p.id_raca = tr.id_tipo_raca
join ensino.forma_ingresso fi on d.id_forma_ingresso = fi.id_forma_ingresso
join comum.unidade u on c.id_unidade = u.id_unidade
join public.status_discente sd on d.status = sd.status
join comum.area_conhecimento_cnpq acc on c.id_area_curso =
acc.id_area_conhecimento_cnpq
where d.nivel = 'G'
and d.ano_ingresso between 1998 and 2019
and d.status in (1, 3, 5, 6, 8, 9)
order by campus, curso_nome, ano_ingresso, nome

```

APÊNDICE D – SCRIPT EM SQL PARA EXTRAÇÃO DOS DADOS PARA A TABELA FATO EVASÃO

```

select
  u.id_unidade,
  c.id_curso,
  u.nome as no_unidade,
  c.nome as no_curso,
  d.ano_ingresso,
  COALESCE((select count(d2.matricula)
    from public.discente d2
    join public.curso c2 on (d2.id_curso = c2.id_curso)
    join comum.unidade u2 on (c2.id_unidade = u2.id_unidade)
    where d2.ano_ingresso between 1998 and 2019
    and d2.nivel = 'G'
    and d2.tipo = 1
    and d2.status in (1, 3, 5, 6, 8, 9)
    and d2.id_forma_ingresso not in (2001507, 2001508, 34116, 111336327,
879385, 34130, 34117)
    and c2.id_modalidade_educacao = 1
    and (u2.id_unidade = u.id_unidade and c2.id_curso = c.id_curso and
d2.ano_ingresso = d.ano_ingresso )
    group by u2.id_unidade, c2.id_curso, d2.ano_ingresso), 0) as
ingressantes,
  COALESCE((select count(foo.matricula) from
    (select distinct on (ma3.id_discente)
      d3.matricula as matricula, c3.id_curso as curso, ma3.ano_referencia
as ano, u3.id_unidade as ua
    from ensino.movimentacao_aluno ma3
    join public.discente d3 on (ma3.id_discente = d3.id_discente)
    join public.curso c3 on (d3.id_curso = c3.id_curso)
    join comum.unidade u3 on (c3.id_unidade = u3.id_unidade)
    where d3.nivel = 'G'
    and d3.ano_ingresso between 1998 and 2019
    and ma3.ano_referencia between 1998 and 2019
    and d3.tipo = 1
    and d3.status in (6)
    and d3.id_forma_ingresso not in (2001507, 2001508, 34116,
111336327, 879385, 34130, 34117)
    and ma3.id_tipo_movimentacao_aluno in (16, 17, 309, 308, 6, 10, 11,
9, 306, 412, 3, 409, 414, 7, 415, 418, 4)
    and c3.id_modalidade_educacao = 1) as foo
    where (foo.ano = d.ano_ingresso and foo.ua = u.id_unidade and
foo.curso = c.id_curso)
    group by ua, curso, ano), 0) as evadidos,
  COALESCE((select count(d4.matricula)
    from ensino.movimentacao_aluno ma4
    join public.discente d4 on (ma4.id_discente = d4.id_discente)
    join public.curso c4 on (d4.id_curso = c4.id_curso)
    join comum.unidade u4 on (c4.id_unidade = u4.id_unidade)
    where d4.nivel = 'G'
    and d4.ano_ingresso between 1998 and 2019
    and ma4.ano_referencia between 1998 and 2019
    and d4.tipo = 1
    and d4.status in (3)
    and d4.id_forma_ingresso not in (2001507, 2001508, 34116, 111336327,
879385, 34130, 34117)

```

```

    and ma4.id_tipo_movimentacao_aluno in (1)
    and c4.id_modalidade_educacao = 1
    and (ma4.ano_referencia = d.ano_ingresso and u4.id_unidade =
u.id_unidade and c4.id_curso = c.id_curso)
    group by u4.id_unidade, c4.id_curso, ma4.ano_referencia), 0) as
concluidos
from public.discente d
left join ensino.movimentacao_aluno ma on (d.id_discente = ma.id_discente)
join ensino.tipo_movimentacao_aluno tma on (ma.id_tipo_movimentacao_aluno =
tma.id_tipo_movimentacao_aluno)
join public.curso c on (d.id_curso = c.id_curso)
join comun.unidade u on (c.id_unidade = u.id_unidade)
where
/* Condicional criado para selecionar apenas ingressantes
* a partir de um ano específico
*/
case
when c.id_curso in (43969900,
                    43969907,
                    43969912,
                    43969914,
                    43969919,
                    43969922,
                    43969924,
                    43969926)
then d.ano_ingresso between 1998 and 2019
when c.id_curso in (43969931,
                    43969933)
then d.ano_ingresso between 2002 and 2019
when c.id_curso in (43969911,
                    43969917,
                    43969921)
then d.ano_ingresso between 2008 and 2019
when c.id_curso in (43969934,
                    43969935,
                    43969940)
then d.ano_ingresso between 2009 and 2019
when c.id_curso in (43969902,
                    43969906,
                    43969909,
                    43969913,
                    43969916,
                    43969918,
                    43969920,
                    43969923,
                    43969925)
then d.ano_ingresso between 2010 and 2019
when c.id_curso in (43969908,
                    43969915,
                    43969927)
then d.ano_ingresso between 2011 and 2019
when c.id_curso in (43969903,
                    43969910,
                    43969928,
                    43969938,
                    43969939)
then d.ano_ingresso between 2012 and 2019
end
and c.id_curso not in (43969901) --CCO. ESSE CURSO ENTRA NOS JOINS ABAIXO
and d.nivel = 'G' --ALUNOS GRADUAÇÃO
and d.tipo = 1 --ALUNOS REGULARES

```



```

        and mc3.periodo = d.periodo_ingresso
        and mc3.id_situacao_matricula not in (10)
        and d3.nivel = 'G'
        and sm3.descricao = 'CANCELADO'
        and d3.matricula = d.matricula
    ) as total_cancelados
from comun.pessoa p
join discente d using (id_pessoa)
join curso c on (d.id_curso = c.id_curso)
join ensino.movimentacao_aluno ma on
(
    ma.id_discente = d.id_discente and
ma.id_movimentacao_aluno in
(
    select id_movimentacao_aluno
    from ensino.movimentacao_aluno ma2
    where ma2.id_discente = d.id_discente
    and d.status in (6) -- 6 cancelado
    and ma2.id_tipo_movimentacao_aluno in (16, 17,
309, 308, 6, 10, 11, 9, 306, 412, 3, 409, 414, 7, 415, 418, 4)
    and (ma2.ano_referencia = d.ano_ingresso and
ma2.periodo_referencia = d.periodo_ingresso)
    order by ma2.ano_referencia desc,
ma2.periodo_referencia desc, ma2.data_ocorrencia desc limit 1
) -- ultima movimentacao de deslgmento do
discente
)
where d.nivel = 'G'
and d.status = 6
and d.ano_ingresso between 1998 and 2020
group by matricula, d.ano_ingresso,
d.periodo_ingresso
order by matricula
) as foo
where foo.total_matriculas = foo.total_cancelados
)
group by u.id_unidade, c.id_curso, d.ano_ingresso
UNION
select
goo.id_unidade,
goo.id_curso,
goo.no_unidade,
case
    when goo.no_curso = 'CIÊNCIA DA COMPUTAÇÃO'
    then 'CIÊNCIA DA COMPUTAÇÃO NOTURNO'
end as no_curso,
goo.ano_ref,
COALESCE(loo.matriculaloo, 0) as ingressantes,
COALESCE(goo.evadidos, 0) as evadidos,
COALESCE(goo.concluidos, 0) as concluidos
from(
    select
    u.id_unidade as id_unidade,
    c.id_curso as id_curso,
    u.nome as no_unidade,
    c.nome as no_curso,
    ma.ano_referencia as ano_ref,
    (select count(foo.matricula) -- SUBCONSULTA PARA TOTAL DE EVADIDOS
    from
    (select distinct on (ma3.id_discente)

```

```

d3.matricula as matricula, c3.id_curso as curso, ma3.ano_referencia
as ano, u3.id_unidade as ua
from ensino.movimentacao_aluno ma3
join public.discente d3 on (ma3.id_discente = d3.id_discente)
join public.curso c3 on (d3.id_curso = c3.id_curso)
join comum.unidade u3 on (c3.id_unidade = u3.id_unidade)
where d3.nivel = 'G'
and d3.ano_ingresso between 1998 and 2008
and ma3.ano_referencia between 1998 and 2019
and d3.tipo = 1
and d3.status in (6)
and d3.id_forma_ingresso not in (2001507, 2001508, 34116,
111336327, 879385, 34130, 34117)
and ma3.id_tipo_movimentacao_aluno in (16, 17, 309, 308, 6, 10, 11,
9, 306, 412, 3, 409, 414, 7, 415, 418, 4)
and c3.id_modalidade_educacao = 1) as foo
where (foo.ano = ma.ano_referencia and foo.ua = u.id_unidade and
foo.curso = c.id_curso)
group by ua, curso, ano) as evadidos,
(select count(d4.matricula) -- SUBCONSULTA PARA O TOTAL DE CONCLUIDOS
from ensino.movimentacao_aluno ma4
join public.discente d4 on (ma4.id_discente = d4.id_discente)
join public.curso c4 on (d4.id_curso = c4.id_curso)
join comum.unidade u4 on (c4.id_unidade = u4.id_unidade)
where d4.nivel = 'G'
and d4.ano_ingresso between 1998 and 2008
and ma4.ano_referencia between 1998 and 2019
and d4.tipo = 1
and d4.status in (3)
and d4.id_forma_ingresso not in (2001507, 2001508, 34116,
111336327, 879385, 34130, 34117)
and ma4.id_tipo_movimentacao_aluno in (1)
and c4.id_modalidade_educacao = 1
and (ma4.ano_referencia = ma.ano_referencia and u4.id_unidade =
u.id_unidade and c4.id_curso = c.id_curso)
group by u4.id_unidade, c4.id_curso, ma4.ano_referencia) as
concluidos
from public.discente d
left join ensino.movimentacao_aluno ma on (d.id_discente =
ma.id_discente)
join ensino.tipo_movimentacao_aluno tma on
(ma.id_tipo_movimentacao_aluno = tma.id_tipo_movimentacao_aluno)
join public.curso c on (d.id_curso = c.id_curso)
join comum.unidade u on (c.id_unidade = u.id_unidade)
where d.nivel = 'G'
and d.ano_ingresso between 1998 and 2008
and d.tipo = 1
and d.status in (1, 3, 5, 6, 8, 9)
and c.id_modalidade_educacao = 1
and c.id_curso = 43969901
group by u.id_unidade, c.id_curso, ma.ano_referencia
order by u.nome, c.nome, ma.ano_referencia
) as goo
left join ( -- JOIN PARA TRAZER A SOMA DE INGRESSANTES POR ANO
select
u2.id_unidade as id_unidade,
c2.id_curso as id_curso,
u2.nome as no_unidade,
c2.nome as no_curso,

```

```

d2.ano_ingresso as ano_ref,
count(d2.matricula) as matriculaloo
from public.discente d2
join public.curso c2 on (d2.id_curso = c2.id_curso)
join comun.unidade u2 on (c2.id_unidade = u2.id_unidade)
where d2.ano_ingresso between 1998 and 2008
and d2.nivel = 'G'
and d2.tipo = 1
and d2.status in (1, 3, 5, 6, 8, 9)
and d2.id_forma_ingresso not in (2001507, 2001508, 34116, 111336327,
879385, 34130, 34117)
and c2.id_modalidade_educacao = 1
and c2.id_curso = 43969901
group by u2.id_unidade, c2.id_curso, d2.ano_ingresso
order by d2.ano_ingresso
) loo on (loo.id_unidade = goo.id_unidade and loo.id_curso =
goo.id_curso and goo.ano_ref = loo.ano_ref)
UNION
select
goo.id_unidade,
goo.id_curso,
goo.no_unidade,
case
when goo.no_curso = 'CIÊNCIA DA COMPUTAÇÃO'
then 'CIÊNCIA DA COMPUTAÇÃO INTEGRAL'
end as no_curso,
goo.ano_ref,
COALESCE(loo.matriculaloo, 0) as ingressantes,
COALESCE(goo.evadidos, 0) as evadidos,
COALESCE(goo.concluidos, 0) as concluidos
from(
select
u.id_unidade as id_unidade,
c.id_curso as id_curso,
u.nome as no_unidade,
c.nome as no_curso,
ma.ano_referencia as ano_ref,
(select count(foo.matricula) -- SUBCONSULTA PARA TOTAL DE EVADIDOS
from
(select distinct on (ma3.id_discente)
d3.matricula as matricula, c3.id_curso as curso, ma3.ano_referencia
as ano, u3.id_unidade as ua
from ensino.movimentacao_aluno ma3
join public.discente d3 on (ma3.id_discente = d3.id_discente)
join public.curso c3 on (d3.id_curso = c3.id_curso)
join comun.unidade u3 on (c3.id_unidade = u3.id_unidade)
where d3.nivel = 'G'
and d3.ano_ingresso between 2009 and 2019
and ma3.ano_referencia between 2009 and 2019
and d3.tipo = 1
and d3.status in (6)
and d3.id_forma_ingresso not in (2001507, 2001508, 34116,
111336327, 879385, 34130, 34117)
and ma3.id_tipo_movimentacao_aluno in (16, 17, 309, 308, 6, 10, 11,
9, 306, 412, 3, 409, 414, 7, 415, 418, 4)
and c3.id_modalidade_educacao = 1) as foo
where (foo.ano = ma.ano_referencia and foo.ua = u.id_unidade and
foo.curso = c.id_curso)
group by ua, curso, ano) as evadidos,
(select count(d4.matricula) -- SUBCONSULTA PARA O TOTAL DE CONCLUIDOS
from ensino.movimentacao_aluno ma4

```

```

join public.discente d4 on (ma4.id_discente = d4.id_discente)
join public.curso c4 on (d4.id_curso = c4.id_curso)
join comum.unidade u4 on (c4.id_unidade = u4.id_unidade)
where d4.nivel = 'G'
and d4.ano_ingresso between 2009 and 2019
and ma4.ano_referencia between 2009 and 2019
and d4.tipo = 1
and d4.status in (3)
and d4.id_forma_ingresso not in (2001507, 2001508, 34116,
111336327, 879385, 34130, 34117)
and ma4.id_tipo_movimentacao_aluno in (1)
and c4.id_modalidade_educacao = 1
and ma4.ano_referencia = ma.ano_referencia and u4.id_unidade =
u.id_unidade and c4.id_curso = c.id_curso)
group by u4.id_unidade, c4.id_curso, ma4.ano_referencia) as
concluidos
from public.discente d
left join ensino.movimentacao_aluno ma on (d.id_discente =
ma.id_discente)
join ensino.tipo_movimentacao_aluno tma on
(ma.id_tipo_movimentacao_aluno = tma.id_tipo_movimentacao_aluno)
join public.curso c on (d.id_curso = c.id_curso)
join comum.unidade u on (c.id_unidade = u.id_unidade)
where d.nivel = 'G'
and d.ano_ingresso between 2009 and 2019
and d.tipo = 1
and d.status in (1, 3, 5, 6, 8, 9)
and c.id_modalidade_educacao = 1
and c.id_curso = 43969901
group by u.id_unidade, c.id_curso, ma.ano_referencia
order by u.nome, c.nome, ma.ano_referencia
) as goo
join ( -- JOIN PARA TRAZER A SOMA DE INGRESSANTES POR ANO
select
u2.id_unidade as id_unidade,
c2.id_curso as id_curso,
u2.nome as no_unidade,
c2.nome as no_curso,
d2.ano_ingresso as ano_ref,
count(d2.matricula) as matriculaloo
from public.discente d2
join public.curso c2 on (d2.id_curso = c2.id_curso)
join comum.unidade u2 on (c2.id_unidade = u2.id_unidade)
where d2.ano_ingresso between 2009 and 2019
and d2.nivel = 'G'
and d2.tipo = 1
and d2.status in (1, 3, 5, 6, 8, 9)
and d2.id_forma_ingresso not in (2001507, 2001508, 34116, 111336327,
879385, 34130, 34117)
and c2.id_modalidade_educacao = 1
and c2.id_curso = 43969901
group by u2.id_unidade, c2.id_curso, d2.ano_ingresso
order by d2.ano_ingresso
) loo on (loo.id_unidade = goo.id_unidade and loo.id_curso =
goo.id_curso and goo.ano_ref = loo.ano_ref)
order by no_unidade, no_curso, ano_ingresso;

```


APÊNDICE E – PRINCIPAIS ARTIGOS UTILIZADOS NA PESQUISA BIBLIOGRÁFICA

ISSN	Título Do Periódico	Título do Artigo	Ano de Publicação	Referência Bibliográfica
1068-2341	Archivos Analíticos de Políticas Educativas	Retenção e abandono de estudantes na educação superior universitária na américa latina e no caribe: uma revisão sistemática	2018	Munizaga, F., Cifuentes, M., & Beltrán, A. (2018). Retención y abandono estudiantil en la Educación Superior Universitaria en América Latina y el Caribe: Una revisión sistemática. Archivos Analíticos de Políticas Educativas, 26(61).
1982-5765	Avaliação: Revista da Avaliação da Educação Superior	O desempenho dos estudantes no vestibular e a permanência nos cursos de graduação da UNEMAT	2018	NODARI, Douglas Ehle; LIMA, Elizeth Gonzaga dos Santos; MACIEL, Carina Elisabeth. O desempenho dos estudantes no vestibular e a permanência nos cursos de graduação da UNEMAT. Avaliação (Campinas), Sorocaba, v. 23, n. 2, p. 312-329, out. 2018.
1982-5765	Avaliação: Revista da Avaliação da Educação Superior	A evasão na educação superior: o estado da arte das pesquisas no Brasil a partir de 1990	2017	SANTOS JUNIOR, José da Silva; REAL, Giselle Cristina Martins. A evasão na educação superior: o estado da arte das pesquisas no Brasil a partir de 1990. Avaliação (Campinas), Sorocaba, v. 22, n. 2, p. 385-402, ago. 2017.
1982-5765	Avaliação: Revista da Avaliação da Educação Superior	Evasão e retenção no curso de Biblioteconomia da UFRGS	2016	VANZ, Samile Andrea de Souza <i>et al.</i> Evasão e retenção no curso de Biblioteconomia da UFRGS. Avaliação (Campinas), Sorocaba, v. 21, n. 2, p. 541-568, jul. 2016.
1982-5765	Avaliação: Revista da Avaliação da Educação Superior	Avaliação dos índices de evasão nos cursos de graduação da Universidade do Estado de Santa Catarina - UDESC	2016	DAVOK, Delsi Fries; BERNARD, Rosilane Pontes. Avaliação dos índices de evasão nos cursos de graduação da Universidade do Estado de Santa Catarina - UDESC. Avaliação (Campinas), Sorocaba, v. 21, n. 2, p. 503-522, jul. 2016.
1982-5765	Avaliação: Revista da Avaliação da Educação Superior	Pode-se identificar a propensão e reduzir a evasão de alunos? Ações estratégicas e resultados táticos para instituições de ensino superior.	2014	TONTINI, Géron; WALTER, Silvana Anita. Pode-se identificar a propensão e reduzir a evasão de alunos? Ações estratégicas e resultados táticos para instituições de ensino superior. Avaliação (Campinas), Sorocaba, v. 19, n. 1, p. 89-110, mar. 2014.

ISSN	Título Do Periódico	Título do Artigo	Ano de Publicação	Referência Bibliográfica
1982-5765	Avaliação: Revista da Avaliação da Educação Superior	Percepções de coordenadores de curso superior sobre evasão, reprovações e estratégias preventivas.	2016	DIOGO, Maria Fernanda <i>et al.</i> Percepções de coordenadores de curso superior sobre evasão, reprovações e estratégias preventivas. Avaliação (Campinas), Sorocaba, v. 21, n. 1, p. 125-151, mar. 2016.
1982-5765	Avaliação: Revista da Avaliação da Educação Superior	Evasão escolar no curso de educação física da Universidade Federal do Piauí	2012	SILVA, Francisca Islandia Cardoso da <i>et al.</i> Evasão escolar no curso de educação física da Universidade Federal do Piauí. Avaliação (Campinas), Sorocaba, v. 17, n. 2, p. 391-404, jul. 2012.
1982-5765	Avaliação: Revista da Avaliação da Educação Superior	Evasão e avaliação institucional no ensino superior: uma discussão bibliográfica	2011	SANTOS BAGGI, Cristiane Aparecida Dos; LOPES, Doraci Alves. Evasão e avaliação institucional no ensino superior: uma discussão bibliográfica. Avaliação (Campinas), Sorocaba, v. 16, n. 2, p. 355-374, jul. 2011.
1980-5314	Cadernos de Pesquisa (Fundação Carlos Chagas)	A evasão no ensino superior brasileiro	2007	SILVA FILHO, Roberto Leal Lobo e <i>et al.</i> A evasão no ensino superior brasileiro. Cad. Pesqui., São Paulo, v. 37, n. 132, p. 641-659, dez. 2007.
1980-850X	Ciência & Educação	Retratos sociológicos: uma metodologia de investigação para a pesquisa em educação.	2015	LIMA JUNIOR, Paulo; MASSI, Luciana. Retratos sociológicos: uma metodologia de investigação para a pesquisa em educação. Ciênc. educ. (Bauru), Bauru, v. 21, n. 3, p. 559-574, set. 2015.
1678-4634	Educação e Pesquisa	Um caso de contratendência: baixa evasão na licenciatura em química explicada pelas disposições e integrações.	2015	MASSI, Luciana; VILLANI, Alberto. Um caso de contratendência: baixa evasão na licenciatura em química explicada pelas disposições e integrações. Educ. Pesqui., São Paulo, v. 41, n. 4, p. 975-992, dez. 2015.
1982-6621	Educação em Revista (UFMG - Online)	Determinant factors for undergraduate student's dropout in an accounting studies department of a Brazilian public university.	2018	DURSO, Samuel de Oliveira; CUNHA, Jacqueline Veneroso Alves da. Determinant factors for undergraduate student's dropout in an accounting studies department of a Brazilian public university. Educ. rev., Belo Horizonte, v. 34, 2018.
1982-6621	Educação em Revista (UFMG - Online)	Uma análise do discurso do aluno trabalhador acerca de sua evasão: caso específico do curso de matemática da UEM.	2018	FERREIRA, Luciano; BARROS, Rui Marcos de Oliveira. Uma Análise Do Discurso Do Aluno Trabalhador Acerca De Sua Evasão: Caso Específico Do Curso De Matemática Da UEM. Educ. rev., Belo Horizonte, v. 34, e171043, 2018.

ISSN	Título Do Periódico	Título do Artigo	Ano de Publicação	Referência Bibliográfica
1982-6621	Educação em Revista (UFMG - Online)	Retenção e evasão no ensino superior público: estudo de caso em um curso noturno de odontologia.	2017	LAMERS, Juliana Maciel de Souza; SANTOS, Bettina Steren dos; TOASSI, Ramona Fernanda Ceriotti. Retenção E Evasão No Ensino Superior Público: Estudo De Caso Em Um Curso Noturno De Odontologia. Educ. rev., Belo Horizonte, v. 33, 2017.
1984-6444	Educação (Santa Maria. Online)	Evasão nas Licenciaturas das Universidades Federais: entre a apetência e a competência	2016	SANTANA, Otacilio Antunes. Evasão nas Licenciaturas das Universidades Federais: entre a apetência e a competência. Educação (UFSM), Santa Maria, p. 311-328, jul. 2016.
1809-4465	Ensaio - Avaliação e Políticas Públicas em Educação	Opiniões de docentes e de coordenadores acerca do fenômeno da evasão discente dos cursos de graduação da Universidade Federal do Ceará (UFC)	2006	ANDRIOLA, Wagner Bandeira; ANDRIOLA, Cristiany Gomes; MOURA, Cristiane Pascoal. Opiniões de docentes e de coordenadores acerca do fenômeno da evasão discente dos cursos de graduação da Universidade Federal do Ceará (UFC). Ensaio: Aval. Pol. Públ. Educ., Rio de Janeiro, v. 14, n. 52, p. 365-382, set. 2006.
1809-4465	Ensaio - Avaliação e Políticas Públicas em Educação	A contribuição da disciplina de introdução à engenharia química no diagnóstico da evasão.	2006	SILVA, Rosenir Rita de Cássia Moreira da; MAINIER, Fernando Benedicto; PASSOS, Fabio Barboza. A contribuição da disciplina de introdução à engenharia química no diagnóstico da evasão. Ensaio: Aval. Pol. Públ. Educ., Rio de Janeiro, v. 14, n. 51, p. 261-277, jun. 2006.
1809-4465	Ensaio - Avaliação e Políticas Públicas em Educação	Multilevel modeling of persistence in higher education.	2018	FERRÃO, Maria Eugénia; ALMEIDA, Leandro S. Multilevel modeling of persistence in higher education. Ensaio: Aval. Pol. Públ. Educ., Rio de Janeiro, v. 26, n. 100, p. 664-683, jul. 2018.
1809-4465	Ensaio - Avaliação e Políticas Públicas em Educação	Evasão na educação superior e gestão institucional: o caso da Universidade Federal da Paraíba.	2018	PRESTES, Emília Maria da Trindade; FIALHO, Maríllia Gabriella Duarte. Evasão na educação superior e gestão institucional: o caso da Universidade Federal da Paraíba. Ensaio: Aval. Pol. Públ. Educ., Rio de Janeiro, v. 26, n. 100, p. 869-889, jul. 2018.

ISSN	Título Do Periódico	Título do Artigo	Ano de Publicação	Referência Bibliográfica
1809-4465	Ensaio - Avaliação e Políticas Públicas em Educação	Uma análise da progressão dos alunos cotistas sob a primeira ação afirmativa brasileira no ensino superior: o caso da Universidade do Estado do Rio de Janeiro.	2014	MENDES JUNIOR, Álvaro Alberto Ferreira. Uma análise da progressão dos alunos cotistas sob a primeira ação afirmativa brasileira no ensino superior: o caso da Universidade do Estado do Rio de Janeiro. Ensaio: Aval. Pol. Públ. Educ., Rio de Janeiro, v. 22, n. 82, p. 31-56, mar. 2014.
1982-4327	Paidéia (USP. Online)	A evasão de alunos.	1993	BUENO, José Lino Oliveira. A evasão de alunos. Paidéia (Ribeirão Preto), Ribeirão Preto, n. 5, p. 9-16, ago. 1993.