



UNIVERSIDADE DO ESTADO DA BAHIA – UNEB
DEPARTAMENTO DE EDUCAÇÃO - CAMPUS I
PROGRAMA DE PÓS-GRADUAÇÃO GESTÃO E TECNOLOGIAS APLICADAS À
EDUCAÇÃO – GESTEC

IVO CHAVES DE FRANÇA

***O BUSINESS INTELLIGENCE* COMO FERRAMENTA DE APOIO AO**
CONTROLE DA EVASÃO ESCOLAR NO IFBA CAMPUS SALVADOR

Salvador
2015

IVO CHAVES DE FRANÇA

**O *BUSINESS INTELLIGENCE* COMO FERRAMENTA DE APOIO AO
CONTROLE DA EVASÃO ESCOLAR NO IFBA CAMPUS SALVADOR**

Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre em Gestão e Tecnologias aplicadas à Educação, curso de Mestrado Profissional, Programa de Pós-graduação Gestão e Tecnologias Aplicadas à Educação (GESTEC), Universidade do Estado da Bahia

Orientador: Prof. Dr. André Ricardo Magalhães

Salvador
2015

FICHA CATALOGRÁFICA
Sistema de Bibliotecas da UNEB
Bibliotecário: Vagner Magarão – CRB:5/1537

França, Ivo Chaves

O Business Intelligence como ferramenta de apoio ao controle da evasão escolar no IFBA
Campus Salvador / Ivo Chaves de França –. Salvador, 2015.

135 f.

Orientador: André Ricardo Magalhães

Dissertação (Mestrado) – Universidade do Estado da Bahia. Mestrado Profissional em Gestão e
Tecnologias aplicadas a Educação (GESTEC).

Contém referências e apêndices.

1. Inteligência competitiva (Administração). 2. Modelagem multidimensional. 3. Evasão escolar
– Ensino superior I. Magalhães, André Ricardo. II. Universidade do Estado da Bahia. Mestrado
Profissional em Gestão e Tecnologias Aplicadas à Educação (GESTEC).

CDD 658.472

FOLHA DE APROVAÇÃO

“O BUSINESS INTELLIGENCE COMO FERRAMENTA DE APOIO AO CONTROLE DA EVASÃO ESCOLAR NO IFBA CAMPUS SALVADOR”

IVO CHAVES DE FRANÇA

Trabalho de Conclusão de Curso apresentado ao Programa de Pós-Graduação (*Scripto sensu*) Gestão e Tecnologias Aplicadas à Educação, Área de Concentração II – Processos Tecnológicos e Redes Sociais, em de 04 de dezembro de 2015, como requisito parcial para obtenção do grau de Mestre em Gestão e Tecnologias Aplicadas à Educação pela Universidade do Estado da Bahia, composta pela Banca Examinadora:



Prof. Dr. André Ricardo Magalhães
Universidade do Estado da Bahia - UNEB
Doutorado em Educação Matemática
Pontifícia Universidade Católica de São Paulo - PUC



Prof. Dr. Eduardo Manuel de Freitas Jorge
Universidade do Estado da Bahia - UNEB
Doutorado em Difusão do Conhecimento
Universidade Federal da Bahia - UFBA



Prof. Dr. Eduardo Cambuzzi
Instituto Federal de Educação, Ciência e Tecnologia da Bahia - IFBA
Doutorado em Engenharia de Automação e Sistemas
Universidade Federal de Santa Catarina - UFSC

Aos meus pais, Seu Zé de Otávio e Dona Maria, à minha esposa, Vanila, ao meu filho, Otávio e aos meus irmãos, Wilson, Ivã e Gustavo.

AGRADECIMENTOS

A Deus, pela vida.

Aos meus pais, por todo amor doado e ensinado a cultivar. Pelo incentivo diário, pela fé e esperança depositadas, e por me fazerem acreditar que é possível dar sempre um passo adiante.

À minha esposa, Vanila, pelo carinho, amor, compreensão, pela presença diária. A meu filho Otávio, minha inspiração e motivação diárias para seguir em frente. Grandes amores da minha vida, a quem aproveito para agradecer a compreensão pelas minhas “evasões” temporárias para me dedicar a este trabalho.

Aos meus irmãos Wilson, Ivã e Gustavo, pelo apoio moral e material, pelo amor e carinho incondicionais de sempre. Obrigado por tudo. Que sorte tê-los como irmãos!

Ao IFBA pelo apoio material, por ter aberto as portas para que a pesquisa pudesse ser realizada. Aos gestores, por compartilharem seus conhecimentos e tempo para a realização das entrevistas e reuniões. À equipe da PROEN, através de Livia Simões e Alberto Ribeiro, pelas importantes contribuições. À equipe da DGTI pelo apoio técnico. Aos amigos e colegas de campus Jacobina, Epaminondas, Lucília, Jane, Letícia, Juliana, Andson, entre outros, pelo apoio, incentivo e por compartilharem das angústias e alegrias desde o período de inscrição para concorrer à vaga do mestrado até o final.

Aos professores do GESTEC, em nome da professora Tânia Hetkowski, pelos conhecimentos, pelo compromisso e por toda doação ao programa. À equipe da secretaria do GESTEC pela atenção e responsabilidade na condução dos trabalhos administrativos. Aos colegas todos do GESTEC e do grupo de pesquisa TECH-MAT pela convivência e experiências compartilhadas. Aos grandes amigos, colegas de classe e fiéis escudeiros João Rogério, Adriana, Letícia e Gilvânia.

Ao meu orientador, André Magalhães, a quem aprendi admirar, por ter me acolhido, pelos seus ensinamentos, pela confiança, pela paciência e dedicação.

Muito obrigado a todos aqueles que de alguma maneira contribuíram para esta importante etapa na minha vida.

É no problema da educação que assenta o grande segredo do aperfeiçoamento da humanidade.

Immanuel Kant

RESUMO

Esta pesquisa aborda o uso da tecnologia *Business Intelligence* como ferramenta de apoio ao controle da evasão escolar pelos gestores escolares a partir da análise dos dados escolares históricos armazenados no Sistema Integrado de Controle Acadêmico - SICAD do Instituto Federal de Educação, Ciência e Tecnologia da Bahia – IFBA, campus Salvador e tem como objetivo a construção de um modelo de dados multidimensional que possibilite aos agentes envolvidos no processo de gestão acadêmica transformar grandes volumes de dados em informações significativas que os auxiliem nesta análise e tomada de decisões acerca da evasão escolar no ensino superior. Trata-se de uma pesquisa aplicada, de base empírica, que utiliza a estratégia da pesquisa-ação para levantamento e obtenção das informações, através de um processo cíclico baseado no ciclo PDCA (acrônimo para *Plan-Do-Check-Act*), processo originário do campo da Administração, comumente usado em projetos de desenvolvimento de soluções de TIC e aplicado às mais diversas áreas do conhecimento na resolução de problemas coletivos. Para mapear os dados e o entendimento acerca da evasão foram realizadas entrevistas a gestores institucionais, guiadas por questionários semi-estruturados, além pesquisas bibliográficas em teses, dissertações de especialistas que discutem a evasão no ensino superior, em legislações e documentos institucionais e os próprios registros acadêmicos e contou com o apoio de diversos departamentos do IFBA. O desenvolvimento da solução de BI, que compreende o processo de modelagem multidimensional e a publicação dos dados em um ambiente analítico, utiliza a suíte *Pentaho*, um conjunto soluções de *software* de código aberto voltados para a construção de projetos de *Business Intelligence*, que permite disponibilizar informações consolidadas através de um navegador web em forma de cubos de dados onde os gestores podem construir seus próprios indicadores, gerar gráficos e outras informações acerca da evasão escolar, de uma forma simples e dinâmica, permitindo analisar os dados sob diversas perspectivas (dimensões) e formatos para apoiá-los na tomada de decisões sobre a evasão escolar. O resultado deste trabalho, além de alcançar os objetivos propostos, demonstrou que a metodologia adotada poderá ser utilizada pela instituição para avançar na ampliação da solução proposta, bem como no desenvolvimento projetos futuros utilizando a tecnologia de *Business Intelligence* voltados para a consolidação de dados acadêmicos históricos. A pesquisa permitiu ainda aprofundar a discussão em torno do assunto evasão dentro do IFBA, este que é um tema de grande preocupação por organismos de controle ligados à educação pelo impacto social, acadêmico e econômico que o fenômeno tem gerado.

Palavras-chave: *Business Intelligence*; Modelagem Multidimensional; Evasão Escolar; Ensino Superior.

ABSTRACT

This research addresses the use of Business Intelligence technology as a support tool to the control of truancy by school managers from the analysis of the historical school data stored in the Sistema de Integrado de Controle Acadêmico - SICAD the Instituto Federal de Educação, Ciência e Tecnologia da Bahia - IFBA, campus Salvador and aims to build a multidimensional data model that allows agents involved in academic management process to transform large volumes of data into significant information to assist them in this analysis and decision-making about school truancy in higher education. It is an applied research, empirical basis, which uses action research strategy for collect and obtaining information through a cyclic process based on the PDCA cycle (acronym for Plan-Do-Check-Act) process originating Administration field, commonly used in ICT solutions development projects and applied to various fields of knowledge in solving collective problems. To map the data and the understanding of truancy were carried out interviews with institutional managers, guided by semi-structured questionnaires, plus bibliographical research in theses, dissertations experts discussing truancy in higher education, legislation and institutional documents and records themselves academics and had the support of many departments of the IFBA. The development of the BI solution, composed the process of multidimensional modeling and publication of data in an analytical environment, use the Pentaho suite, a set of open source software solutions focused on the construction of Business Intelligence projects, which makes available consolidated information through a web browser in the form of data cubes where managers can build their own indicators, generate graphs and other information about truancy in a simple and dynamic way, allowing analyze data from different perspectives (dimensions) and formats to support them in making decisions about truancy. The result of this work, in addition to achieve the proposed objectives, demonstrated that the methodology can be used by the institution to advance the expansion of the proposed solution, as well as developing future projects using Business Intelligence technology aimed at the consolidation of historical academic data. The research also allowed further discussion around the subject truancy within the IFBA, this is a major topic of concern for regulatory bodies linked to education for social, academic and economic phenomenon that has generated.

Keywords: Business Intelligence; Multidimensional modeling; School Truancy; Higher education.

LISTA DE ILUSTRAÇÕES

Figura 1 – Elementos básicos do data warehouse.	46
Figura 2 – Representação da visão multidimensional como recurso OLAP em formato de cubo	49
Figura 3 – Demonstração do ambiente Kettle com uso de recursos visuais para construção do ETL	51
Figura 4 – Demonstração do ambiente de trabalho do Schema Workbench.....	52
Figura 5 – Demonstração da ferramenta <i>Saiiku Analytics</i> integrada ao ambiente <i>Pentaho</i>	53
Figura 6 – Representação do ciclo PDCA e suas fases no projeto	59
Figura 7 – Modelo de Dicionário de Dados adotado para documentar dados do SICAD	62
Figura 8 – Modelo de Matriz de Necessidade adaptado para este projeto.....	67
Figura 9 – Modelo Multidimensional para Fato Evasão Anual	73
Figura 10 – Modelo Multidimensional para Fato Evadidos por Matriculados	74
Figura 11 – Modelo Multidimensional para o Fato Nota Frequência.....	74
Figura 12 – Cubo Evasão Anual elaborado na ferramenta <i>Workbench</i>	78
Figura 13 – Cubo Evadidos por Matriculados elaborado na ferramenta <i>Workbench</i>	79
Figura 14 – Cubo Notas e Frequência dos alunos evadidos elaborado na ferramenta <i>Workbench</i>	80
Figura 15 – Conjunto de fatores socioeconômicos extraídos da ferramenta OLAP.....	95
Figura 16 – Painel de Monitoramento construído e disponibilizado através do ambiente OLAP	100

LISTA DE TABELAS

Tabela 1: Relação entre ingressos e concluintes no período de 2002 e 2008 apresentada no resumo técnico do censo escolar 2008 divulgado pelo INEP	27
Tabela 2: Relação entre ingressos e concluintes no período de 2006 e 2013 apresentada no censo escolar 2013 divulgado pelo INEP.....	28
Tabela 3: Relação entre ingressos e concluintes no período de 2010 a 2013 apresentada nos relatórios gestão do IFBA	32
Tabela 4: Dados Operacionais x Dados Informativos	44
Tabela 5: Modelo Dimensional x Modelo Relacional	47
Tabela 6: Lista de dimensões extraídas da matriz de necessidades.....	70
Tabela 7: Lista de fatos extraídos a partir da matriz de necessidades	71
Tabela 8: Índice de Evasão Anual por curso.....	87
Tabela 9: Percentual de evadidos e quantidade de matriculados por curso nos últimos 5 anos	88
Tabela 10: Total de evadidos e matriculados por forma de ingresso entre 2010 e 2014.....	90
Tabela 11: Total de evadidos e matriculados por forma de ingresso entre 2010 e 2014.....	92
Tabela 12: Percentual de evadidos e matriculados por faixa etária entre 1999 e 2014	94
Tabela 13: Totais de evadidos e matriculados entre 1999 e 2014 através das Cotas	96

LISTA DE GRÁFICOS

Gráfico 1: Índice de Evasão Anual do campus Salvador de 1999 a 2014	84
Gráfico 2: Índices de variação da Evasão e do número de Matriculados Anuais do campus Salvador de 1999 a 2014	86
Gráfico 3: Índices de evasão por forma de ingresso de evadidos entre 1999 e 2014	89
Gráfico 4: Índices de Evasão por forma período acadêmico entre 1999 e 2014.....	91

LISTA DE ABREVIATURAS E SIGLAS

BI	Business Intelligence
CEFET	Centro Federal de Educação Tecnológica
DGTI	Diretoria de Gestão de Tecnologia da Informação
DM	Data Mart
DW	Data Warehouse
ERP	Entreprise Resource Plainning
ETL	Extraction, Transformation and Load
IDEB	Índice de Desenvolvimento da Educação Básica
IFBA	Instituto Federal de Educação, Ciência e Tecnologia da Bahia
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
MEC	Ministério da Educação
OLAP	On-line Analytical Processing
OLTP	Online Transaction Processing
PDCA	Plain-Do-Check-Action
PDI	Plano de Desenvolvimento Institucional
PNE	Plano Nacional de Educação
PNUD	Programa das Nações Unidas para o Desenvolvimento
PROEN	Pró-Reitoria de Ensino
PRONATEC	Programa Nacional de Acesso ao Ensino Técnico e Emprego
REUNI	Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais
SAD	Sistema de Apoio a Decisão
SESU	Secretaria de Educação Superior
SGBD	Sistema Gerenciador de Banco de Dados
SICAD	Sistema Integrado de Controle Acadêmico
SISTEC	Sistema Nacional de Informações da Educação Profissional e Tecnológica
TCG	Taxa de Conclusão de Cursos de Graduação
TIC	Tecnologia da Informação e Comunicação

SUMÁRIO

1 INTRODUÇÃO	14
1.1 JUSTIFICATIVA.....	17
1.2 OBJETIVOS	19
1.2.1 Objetivo Geral	19
1.2.2 Objetivos Específicos	20
2 DESCRREVENDO O CENÁRIO ATUAL DO IFBA	20
3 EVASÃO ESCOLAR NO BRASIL	25
3.1 A EVASÃO ESCOLAR NO IFBA	30
3.2 CONSIDERAÇÕES SOBRE A EVASÃO ESCOLAR.....	38
4 O BUSINESS INTELLIGENCE E SUAS FERRAMENTAS	40
4.1 O BUSINESS INTELLIGENCE	40
4.2 O DATA WHAREHOUSE	43
4.2.1 Modelagem Dimensional	46
4.3 FERRAMENTAS DE VISUALIAZAÇÃO OLAP	48
4.4 A SUITE PENTAHO	50
5 METODOLOGIA	54
6 DESENVOLVIMENTO	61
6.1 COLETA E ANÁLISE DOS DADOS.....	61
6.2 DEFINIÇÃO DOS DESCRITORES E DOS INDICADORES	64
6.3 DESENVOLVIMENTO DOS MODELOS MULTIDIMENSIONAIS	68
6.4 IMPLANTAÇÃO E DISPONIBILIZAÇÃO DOS CUBOS OLAP	76
7 RESULTADOS E DISCUSSÕES	81
7.1 ANÁLISE SOBRE O CRESCIMENTO DOS DADOS ACADÊMICOS.....	81
7.2 ANÁLISE DOS DADOS SOBRE A EVASÃO.....	82
7.2.1 Índices Gerais de Evasão	84

7.2.2 Fatores relacionados a curso, período acadêmico, forma de ingresso e desempenho	86
7.2.3. Fatores Socioeconômicos	93
7.3 ANÁLISE SOBRE O AMBIENTE ANALÍTICO PROPOSTO	97
8 CONSIDERAÇÕES FINAIS E PERSPECTIVAS FUTURAS	102
REFERÊNCIAS	107
APÊNDICES	113

1 INTRODUÇÃO

Este trabalho de pesquisa aborda o uso da tecnologia *Business Intelligence* – BI como ferramenta de apoio ao controle da evasão escolar pelos gestores escolares a partir da análise dos dados escolares históricos armazenados no sistema de registros acadêmicos do Instituto Federal de Educação, Ciência e Tecnologia da Bahia – IFBA, campus Salvador.

Segundo Rigo et al. (2014), o avanço no uso da Tecnologia da Informação e Comunicação - TIC tem gerado um volume substancial de dados, numa proporção muito maior que a capacidade de pesquisadores e analistas em analisar estes dados gerados. No caso específico dos sistemas acadêmicos, observa-se ainda que estes dados possuem uma granularidade peculiar, como registro de frequências, notas, entre outras informações, o que potencializa ainda mais esta dificuldade na recuperação e análise destes dados.

Como proposta para tratar estes volumes de dados, um conceito que vem sendo muito empregado em grandes organizações e que aos poucos também vem sendo adotado por empresas de menor porte é o BI, que segundo Barbieri (2001) representa um “guarda-chuva” conceitual, que engloba diversas soluções capazes de promover uma correta estruturação das informações em depósitos retrospectivos e históricos, permitindo a sua manipulação por ferramentas de análise e inferência com o objetivo de apoiar o processo de tomada de decisões.

A pesquisa tem como questão norteadora identificar de que maneira os dados acadêmicos históricos podem, com o auxílio do BI, contribuir para análises sobre evasão escolar no ensino superior no IFBA, no campus Salvador. O tema evasão escolar, uma preocupação geral das instituições de ensino e de órgãos governamentais de controle, vem sendo tratado como assunto de diversas pesquisas acadêmicas no campo da educação, nas suas diversas modalidades de ensino. No ensino superior, foco desta pesquisa, muitos estudos têm se voltado à discussão dos métodos de cálculo adotados pelas instituições, assim como na análise de possíveis motivos apontados pelos próprios alunos e discussões acerca das ambigüidades conceituais em torno do tema no ensino superior, dadas as características e peculiaridades desta modalidade de ensino, como relatam Fialho (2014), Platt Neto et al. (2008) e Baggi e Lopes (2011) em suas pesquisas acerca da evasão no ensino superior. Esta pesquisa tem como perspectiva uso da tecnologia de BI como instrumento de apoio na análise de dados históricos já presentes na instituição de ensino, sem desconsiderar todas estas

discussões que vem sendo levantadas em torno do assunto, mas em consonância com o que indicam Polydoro (2000), Silva Filho et. al (2007) da criação de instrumentos internos nas instituições que permitam analisar dados existentes sobre a evasão para apoiar no acompanhamento e controle.

Neste sentido, este trabalho busca, através das soluções de BI, a construção de um modelo de dados multidimensional que possibilite aos agentes envolvidos no processo de gestão acadêmica transformar esse volume de dados em informações que os auxiliem no controle da evasão escolar do IFBA campus Salvador. Segundo Barbieri (2001), essa transformação de dados é um dos principais desafios atuais da TIC, o de proporcionar às organizações uma atuação ágil e inteligente, transformando dados em informações e as informações em conhecimento, para que possam ser úteis.

Será utilizada a base dados do sistema acadêmico do IFBA, que possui dados desde 1984 e já conta com alguns milhões de registros armazenados, com o foco na análise dos dados sobre a evasão escolar dos cursos do ensino superior, no campus Salvador. O desenvolvimento da pesquisa está baseado na premissa metodológica de pesquisa-ação e utiliza o ciclo PCDA, num processo iterativo de investigação e ação em torno da construção do modelo multidimensional.

Esta pesquisa torna-se relevante não apenas para apoiar na análise sobre a evasão, mas, além disso, por permitir que o modelo de desenvolvimento adotado e sua metodologia sirvam de base para que o próprio IFBA possa aplicá-las a outras questões associadas aos dados acadêmicos e administrativos assim como por outras instituições de ensino.

Cabe situar o autor, sobretudo sua trajetória e atuação enquanto profissional de TIC atuando na área da educação, este que há alguns anos tem se dedicado, enquanto Analista de Tecnologia da Informação e estudante, ao uso do BI como ferramenta de apoio na gestão para diferentes áreas, como o uso desta tecnologia aplicada a sistemas de Gestão Empresarial na iniciativa privada, no acompanhamento e gestão do processo de desenvolvimento de software em fábricas de softwares e mais recentemente numa pesquisa na área de educação, em que foi aplicado o BI para apoiar na análise de dados escolares de uma escola pública municipal e que deu origem um trabalho de conclusão de curso de uma especialização em Banco de Dados. Foi o resultado deste trabalho que o motivou a continuar pesquisando o uso desta tecnologia de maneira a contribuir com o ambiente acadêmico.

Este trabalho está dividido em oito capítulos. O primeiro traz uma apresentação geral daquilo que propõe a pesquisa, abordando seus principais objetivos, as razões que nortearam a escolha pelo objeto da pesquisa e sua justificativa. O segundo capítulo, o autor descreve o cenário em que se encontra o IFBA, com um rápido histórico de uma instituição centenária que vem passando por grandes transformações, sobretudo com o seu papel enquanto instituição de educação. Além disso, este capítulo faz um descritivo do IFBA nas questões administrativas e mais especificamente relacionado ao controle dos seus registros acadêmicos, estas que são o foco de análise desta pesquisa. O terceiro capítulo trata-se de pesquisa exploratória em torno da questão da evasão escolar do ponto de vista dos órgãos que gerem a educação no Brasil, abordando questões em torno do assunto que tem sido motivo de preocupação por organismos internacionais. Ainda neste terceiro capítulo o autor descreve o assunto do ponto de vista do IFBA, tanto a visão institucional, a partir de suas normas internas, como a visão de alguns dos seus gestores que lidam mais diretamente com o assunto.

O quarto capítulo tem uma abordagem mais técnica sobre as tecnologias usadas para o desenvolvimento da solução de BI. Está formatado de tal maneira que permita uma revisão bibliográfica e um entendimento sobre estas tecnologias, assim como faz um entrelaçamento destas com as decisões técnicas adotadas para este projeto.

O quinto é dedicado à abordagem metodológica adotada, bem como uma breve explicação para cada uma das fases e etapas que compõem o projeto. O sexto capítulo trata-se do detalhamento das fases e etapas que estão relacionadas ao desenvolvimento da solução de BI, com a apresentação dos instrumentos usados durante o desenvolvimento, suas aplicabilidades e os produtos resultantes deste processo. O sétimo capítulo busca realizar uma análise sobre os dados resultantes da modelagem multidimensional desenvolvida e discussões em torno destes resultados sob o ponto de vista do pesquisador e de especialistas no tema evasão no ensino superior, bem como validar a efetividade do ambiente analítico proposto como instrumento para uso pelos gestores.

Por fim, o oitavo capítulo está dedicado a apresentar as considerações acerca deste trabalho, assim como apresentar possíveis caminhos a serem adotados como continuidade desta pesquisa ou como projetos futuros.

1.1 JUSTIFICATIVA

O uso da Tecnologia da Informação e Comunicação (TIC) tem se tornado cada vez mais necessário para toda e qualquer organização. E a quantidade de informações em meio digital cresce exponencialmente. Costa (2009, p.26) apontava em 2010 que a quantidade de informações produzidas em meio digital chegava a atingir a marca de 18 milhões de vezes mais que os livros. Neste aspecto ele chamava a atenção para a necessidade de se desenvolver ferramentas de TIC adequadas para tratar estes grandes volumes de informação, sem perder de vista os propósitos específicos a que se destina a análise destas informações.

Para Lima, Boscarioli e Lago (2008), a predominância do uso das TIC ainda é das instituições privadas que dominam o uso das tecnologias, sobretudo as tecnologias criadas mais recentemente, desenvolvidas com o claro objetivo de fins comerciais. Embora ainda discreto quando comparado com a iniciativa privada, o uso de ferramentas de TIC tem avançado nas organizações públicas, principalmente na busca por soluções que possibilitem respostas às demandas sociais, conforme afirma Carneiro (2010). As instituições de ensino, responsáveis pelo desenvolvimento e geração de boa parte dos conhecimentos, também precisam estar atentas às demandas sociais, tanto internas como externas e precisam de ferramentas que possam apoiar na compreensão destas informações e contribuir na tomada de decisão pela gestão, possibilitando intervenções imediatas e futuras. Como afirma Lévy (1993, p. 9) “a filosofia política não pode mais ignorar a ciência e a técnica”. Para Lima Junior (1997, p. 175) “A filosofia educacional, os métodos e técnicas educacionais, devem, exatamente, reproduzir o imbricamento e a mixagem próprios [...], de uma pertinência e eficiência sociais próprios da velocidade das mudanças”.

Os sistemas informatizados de registros acadêmicos armazenam diariamente dados muito granulares, gerando um volume com crescimento exponencial, extrapolando a efetividade dos métodos de gerenciamento desses registros, sobretudo por se tratar de Sistemas Transacionais¹ que são desenvolvidos para fins específicos, com relatórios previamente definidos e sem muita flexibilidade para extração dos dados neles armazenados. Todavia, o uso tão somente destas soluções tecnológicas, com a consolidação desses dados

¹ Sistemas Transacionais são sistemas desenvolvidos com o objetivo de gerenciamento do fluxo de processos básicos das organizações, para atendimento de necessidades específicas das mesmas. (BALTZAN; PHILLIPS, 2012)

acadêmicos, sem um profundo entendimento do ambiente escolar, das relações e regras existentes naquele ambiente acadêmico em que estão armazenados, não é o suficiente. Corroborando neste sentido, Morin (2000, p. 36), afirma que o “conhecimento das informações ou dos dados isolados é insuficiente. É preciso situar as informações e os dados em seu contexto”.

Num levantamento realizado acerca deste tema e o uso desta tecnologia, foi possível observar que já existem alguns trabalhos de pesquisa, artigos, dissertações, teses, que se apoiaram no BI como técnica para extrair e construir um ambiente de informações consolidadas, inclusive na área de educação. Um desses trabalhos, apresentado na forma de dissertação de mestrado por Lopes (2013), usou como base a modalidade de Ensino a Distância - EAD, analisando os dados da ferramenta *moodle*, enquanto outros trabalharam os dados armazenados em sistemas acadêmicos da modalidade presencial. No caso do EAD, a principal preocupação dessa pesquisa foi analisar dados de navegação dos alunos no ambiente virtual, consolidando-os, conforme Lopes (2013). Para os projetos voltados ao ensino presencial, o principal foco foi demonstrar a aplicabilidade da tecnologia, sobretudo para indicadores como nota, frequência e índices educacionais gerais, dados estes de uso mais convencional no ambiente escolar, com o objetivo de facilitar o processo de visualização destes dados de modo centralizado e consolidado.

Neste trabalho, será adotado o atual sistema acadêmico do IFBA, uma instituição pública de ensino secular e que acumula uma base histórica de informações de registros escolares digitalizadas de mais de 20 anos, segundo levantamento dos dados armazenados no banco de dados mantido pelo sistema acadêmico do ensino superior no campus Salvador. Dados estes que, em sua maioria, são armazenados apenas para cumprir regras legais que estabelece a necessidade de armazenamento destas informações para uma possível recuperação em caso de uma necessidade específica. Nesta pesquisa o foco da análise dos dados será sobre a evasão escolar no ensino superior, uma preocupação que tem sido amplamente debatida pelas instituições que acompanham os indicadores educacionais no Brasil e no mundo, como apontam informações presentes nos relatórios apresentados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira - INEP e pelo Programa das Nações Unidas para o Desenvolvimento – PNUD, e como aparecem nas principais metas estabelecidas no Plano Nacional de Educação – PNE, a serem alcançadas pelas instituições de ensino.

A motivação para escolha do assunto evasão escolar como elemento a ser pesquisado e analisado, além das questões anteriormente apresentadas, com o uso do BI se deu durante os primeiros levantamentos juntos à Pró-Reitoria de Ensino do IFBA sobre as necessidades de indicadores educacionais que pudessem contribuir para a gestão do instituto. Nestes levantamentos esta foi uma das principais dificuldades apontadas por esta Pró-Reitoria, a de contabilizar alunos evadidos nos cursos do ensino superior do IFBA, bem como a identificação dos principais motivos que levam estes alunos à evasão, como uma forma de agir preventivamente. Ainda que atualmente o pesquisador esteja alocado no IFBA *campus* Jacobina, este *campus* iniciou seu funcionamento recentemente e está em processo de implantação do sistema acadêmico, portanto não possui dados históricos para esta análise, o primeiro curso superior passou a ser ofertado em 2015.2, esta que é a modalidade de ensino que tem preocupado a Pró-Reitoria no que diz respeito a informações sobre a evasão escolar. A escolha pelo campus Salvador se deu principalmente por este ser o *campus* mais antigo a ofertar cursos superiores dentro do IFBA e o fato de estar situado na mesma cidade em que funciona a Reitoria e suas Pró-reitorias, o que facilita durante o processo de pesquisa o contato com gestores tanto do *campus* como dos diversos departamentos da Reitoria, que é atualmente a mantenedora dos dados em questão.

1.2 OBJETIVOS

1.2.1 Objetivo Geral

Considerando a necessidade de desenvolvimento de soluções de TIC em instituições públicas com vistas ao atendimento das demandas sociais, o cenário em que se encontra o IFBA no que se refere aos seus dados acadêmicos, com o grande volume de dados armazenados, bem como a busca por soluções que permitam apoiar os gestores no controle sobre o problema da evasão escolar no ensino superior, esta pesquisa tem como objetivo geral construir um modelo de dados multidimensional, a partir da base de dados histórica do sistema acadêmico do IFBA *campus* Salvador, para o mapeamento das informações acerca da evasão escolar no ensino superior.

1.2.2 Objetivos Específicos

- Identificar e selecionar os dados existentes no atual sistema acadêmico do IFBA e mapear informações úteis para a construção dos indicadores para apoio à tomada de decisão;
- Desenvolver um modelo de dados multidimensional que permita a construção dos indicadores e a consolidação das informações sobre a evasão no ensino superior;
- Utilizar a tecnologia de *Business Intelligence*, através de uma ferramenta de visualização OLAP a partir da modelagem multidimensional;
- Validar a efetividade do ambiente analítico proposto.

2 DESCRREVENDO O CENÁRIO ATUAL DO IFBA

O Instituto Federal de Educação, Ciência e Tecnologia da Bahia, é resultado da criação lei 11.892 de 29 de dezembro de 2008, onde funcionavam as antigas escolas técnicas federais, conhecidas como Centros Federais de Educação Tecnológicas (CEFETs). A partir desta lei, estas instituições ganharam *status* de Universidade, embora a maioria dos CEFETs já oferecesse cursos superiores das áreas tecnológicas e as engenharias a muito mais tempo. A partir de então estes institutos passaram a oferecer também licenciaturas e outros cursos superiores nas diversas áreas do conhecimento (BRASIL, 2008). As principais mudanças não vieram neste caminho, essa lei trouxe consigo projetos de expansão e interiorização da rede federal de educação, desde a educação básica até a pós-graduação, como é realidade hoje na maioria dos IFETs.

Art. 2º Os Institutos Federais são instituições de educação superior, básica e profissional, pluricurriculares e multicampi, especializados na oferta de educação profissional e tecnológica nas diferentes modalidades de ensino, com base na conjugação de conhecimentos técnicos e tecnológicos com as suas práticas pedagógicas, nos termos desta Lei.

§ 1º Para efeito da incidência das disposições que regem a regulação, avaliação e supervisão das instituições e dos cursos de educação superior, os Institutos Federais são equiparados às universidades federais. (BRASIL, 2008)

No que se refere ao IFBA, este aumentou de 5 campi em 2005, para 22 em 2015, distribuídos em 22 cidades da Bahia, além de núcleos avançados, que são extensões de alguns *campi* e que funciona em cidades vizinhas. Isto em termos numéricos significa um aumento

de mais 400% na quantidade *campi* a serem administrados em distâncias geográficas que passam dos 900 km da sede da reitoria em Salvador.

Estes institutos, em especial o da Bahia, veio ao longo da sua história mudando de nomes, mas já tem mais de 105 anos de atuação no estado e é possível imaginar que uma instituição como esta, secular, vem produzindo muitos milhões de registros acadêmicos ao longo da sua história. E isso, se for considerado apenas registros acadêmicos, mas sabe-se também que uma organização mantém seu funcionamento administrativo com a produção diária de centenas de documentos burocráticos e informações administrativas.

Mesmo com todo esse tempo de funcionamento, apenas a partir de 1999 é que o IFBA iniciou o processo de implantação do sistema informatizado dos registros acadêmicos, conforme documento de definição da base de dados e do funcionamento do sistema disponibilizado pela equipe da DGTI. A maioria dos seus processos administrativos internos ainda é mantida de forma manual. Obviamente que muitas das informações financeiras, por força, principalmente, dos órgãos de controle, já estão informatizadas, sobretudo porque os sistemas são desenvolvidos e mantidos pelos ministérios e utilizados por todos os órgãos do poder executivo federal.

É importante situar o IFBA no seu contexto histórico, já que se for considerado o seu tempo de existência, o seu campo de atuação (educação básica, superior e pós-graduação), já seria o suficiente para produzir milhões de registros acadêmicos. E essa situação se agrava, quando, em apenas alguns anos, o número de *campi* se multiplica em mais de quatro vezes, conforme dito anteriormente.

Nas primeiras avaliações realizadas sobre os dados acadêmicos do IFBA, observou-se que embora o atual sistema de registro acadêmico tenha sido implantado a partir de 1999, segundo informações obtidas junto ao departamento de informática e manual básico disponibilizado pela DGTI, algumas informações históricas foram lançadas retroativamente e constam dados de 1984. Este atual sistema é denominado Sistema Integrado de Controle Acadêmico – SICAD, está dividido em um módulo desenvolvido em linguagem de programação *Delphi*, que compreende o módulo de gerenciamento do sistema usado pelos gestores, além de dois módulos web, desenvolvidos em linguagem de programação *Java*, um deles usado pelos docentes como caderneta eletrônica e outro como sendo o portal do aluno, onde é possível consultar as informações lançadas pelos docentes, como suas notas e frequência. Todos estes três módulos são armazenados numa única base de dados. De acordo com as informações levantadas na base de dados do sistema, atualmente ele é usado no ensino

superior dos *campi* Camaçari, Eunápolis, Porto Seguro, Salvador, Santo Amaro, Simões Filho, Valença e Vitória da Conquista.

Para se ter uma noção inicial do volume de dados registrados no SICAD foi realizada uma análise às tabelas do banco de dados e foi verificada a seguinte situação (foram consideradas dados acadêmicos lançados até final do calendário acadêmico de 2014), em valores aproximados:

- Alunos: 48 mil registros;
- Frequência: 2 milhões de registros;
- Notas/Avaliações: 1,5 milhões de registros;
- Matrículas: 55 mil registros.

Este é apenas um pequeno exemplo do volume de registros contido num sistema desta natureza. É válido considerar que boa parte dos *campi*, sobretudo os mais recentemente implantados, ainda não faz uso deste sistema e uma parte significativa das informações é armazenada em planilhas eletrônicas.

Este trabalho inicialmente buscava a criação de um ambiente centralizado de indicadores acadêmicos gerais, a partir destes dados históricos, e que pudessem servir como instrumento de apoio para os gestores acadêmicos. A partir desta imersão no campo de pesquisa e de uma melhor compreensão daquilo que estava sendo pesquisado que foi possível perceber a dimensão do que seria analisar todos estes dados históricos, compreendê-los e transformá-los em um ambiente centralizado.

Como mencionado no Capítulo 1 quando traz a justificativa, as primeiras conversas com os gestores e esta primeira análise sobre os dados demonstraram a necessidade de se definir melhor aquilo que seria analisado, já que para o período do mestrado seria possível dar conta de analisar a complexidade de assuntos existentes nos dados armazenados. Foi daí que, seguindo a orientação da pró-reitoria, a partir da dificuldade da instituição em consolidar informações acerca da evasão escolar e que os possibilitassem identificar os motivos que levam os alunos do ensino superior a evadirem que este passou a ser o assunto a ser tratado no andamento desta pesquisa.

O volume de dados armazenados nesta base do IFBA, ao longo desses mais de 20 anos de registros acadêmicos, é muito grande. Porém, essa quantidade de dados, apenas armazenados, sobretudo aqueles em que as turmas já finalizaram, não teria outra utilidade senão ficarem como dados históricos, cumprindo uma determinação legal de estarem

guardados. Com o desenvolvimento de um modelo multidimensional e uso de soluções de *Business Intelligence*, é possível “dar-lhes vida”, possibilitando que os gestores façam uso deles de uma forma estruturada e consolidada, por exemplo, no cruzamento de informações acadêmicas, socioeconômicas, entre outras, para apoiá-los na tomada de decisões, uma vez que estas fontes de informação não armazenam apenas notas e frequências, existem dados cadastrais, compostos por informações socioeconômicas dos alunos desde o momento de inscrição para o processo seletivo até sua saída da instituição.

Diante de questões como distâncias geográficas, grandes volumes de informações armazenadas das mais heterogêneas formas, da própria dificuldade em analisar e compreender questões conceituais que estão implícitas nestas informações e mesmo o ambiente em que se encontram é que o escopo deste trabalho ficou definido que neste primeiro momento, durante o período que compreende a pesquisa do mestrado (há previsão de que este trabalho tenha uma continuidade dentro da instituição com estudos de outros assuntos ou mesmo um maior aprofundamento deste), em tratar dados sobre a evasão escolar, no ensino superior, para os cursos ofertados no campus Salvador, já que além ser o *campus* mais antigo, é aquele que utiliza o SICAD desde sua implantação.

Do ponto de vista das informações contidas no sistema de registros acadêmicos, o cenário em que os dados do SICAD se encontram pode ser descrito a seguir:

- O único documento encontrado acerca do SICAD é de 1999 e tem apenas definição inicial do banco de dados e seu funcionamento básico.
- Não há dicionário de dados, um dos principais artefatos de um projeto de desenvolvimento de um sistema, o que dificulta a análise e interpretação dos dados armazenados.
- Vários dados começaram a ser coletados recentemente e que não permitem a análise histórica destes dados e desta forma dificulta a inferência acerca das informações extraídas.
- Dados armazenados sem a devida integridade, a exemplo da informação sobre o endereço em que não há garantia sobre a relação entre bairro e a cidade cadastrada no endereço do aluno.
- Redundância no armazenamento de algumas informações, a exemplo de campos duplicados como *etnia*, *etnia1* dentro da mesma tabela.

Quanto à Evasão Escolar, este assunto será tratado com mais detalhes no Capítulo 3, mas de um modo geral pode-se afirmar que até então na instituição não foi encontrado norma ou documentos institucionais de outra natureza que trate ou conceitue a evasão escolar no ensino superior. Anualmente o IFBA tem disponibilizado informações num relatório denominado Relatório de Gestão. Este documento apresenta informações que segue orientações do Tribunal de Contas da União – TCU e o formato tem sido usado por todos os IFETs. Os índices divulgados neste relatório são extraídos do Sistema Nacional de Informações da Educação Profissional e Tecnológica - SISTEC, que é alimentado manualmente por cada *campus*.

Conhecer o cenário em que se encontra o IFBA e seu contexto histórico é fundamental para guiar no planejamento e no desenvolvimento da pesquisa, sobretudo quando se trata de uma pesquisa social de natureza empírica. Diante do exposto percebe-se a necessidade de uma análise aprofundada nos dados contidos nas fontes de armazenamento e o envolvimento da equipe de Tecnologia da Informação para apoiar no entendimento destes dados, haja vista a ausência de documentação. Para algumas das questões acima a respeito dos dados, como integridade e documentação, aquelas relacionadas diretamente com a pesquisa serão mitigadas durante a fase desenvolvimento do projeto de BI, conforme Capítulo 6, como a elaboração do dicionário de dados das informações relacionadas à evasão no ensino superior, bem como transformações e uniformizações de informações que fazem parte do processo de desenvolvimento de soluções de BI. Ademais, outras questões encontradas durante a análise dos dados, sem relação direta ou que fujam das ações e objetivos desta pesquisa, serão apresentadas como sugestões de projetos futuros.

3 EVASÃO ESCOLAR NO BRASIL

Este capítulo tem por objetivo tratar a evasão escolar sob o ponto de vista dos órgãos que gerem a educação no Brasil, como o Ministério da Educação - MEC, órgãos ligados a este ministério que são responsáveis por reunir e divulgar dados sobre a evasão escolar, bem como abordar como este tema tem sido motivo de preocupação, até mesmo por organismos internacionais. Ainda neste capítulo será abordado o tema evasão escolar dentro do IFBA, trazendo números e questionamentos sobre a forma como este assunto tem sido tratado no âmbito do IFBA.

A evasão escolar no Brasil tem sido, há muito tempo, elemento de preocupação dos órgãos de controle e acompanhamento dos índices que medem a educação escolar. A própria Lei de Diretrizes e Bases da Educação Nacional de 1996, traz no artigo 3º um inciso que aponta a preocupação com a permanência do aluno na escola:

Art. 3º. O ensino será ministrado com base nos seguintes princípios:
I - igualdade de condições para o acesso e permanência na escola;

Para Prado (2000, p. 49) “a repetência, o abandono e a evasão são problemas crônicos, que sempre estiveram presentes na história da educação escolar brasileira”. Segundo o relatório do Programa das Nações Unidas para o Desenvolvimento - PNUD, divulgado em 2013, o Brasil ocupava em 2012 o 3º maior índice de abandono escolar no ensino primário, chegando a 24,3%, quando comparado aos 100 países do mundo com maior Índice de Desenvolvimento Escolar, perdendo apenas Bósnia-Herzegovina e São Cristóvão e Nevis, que apresentam índices de 26,8% e 26,5%, respectivamente. Quando comparado aos países da América Latina, apenas Guatemala (35,2%) e Nicarágua (51,6%) têm números superiores ao Brasil. O índice de abandono escolar, medido pelo PNUD (2013, p. 179) considera:

Percentagem de alunos de um determinado grupo que se matricularam no ensino primário mas abandonaram os estudos antes de o terminarem. É calculada com base em 100 menos a taxa de sobreviventes do último ano do ensino primário, e assume que os fluxos observados se mantêm inalterados durante a vida do grupo e que os abandonos não regressam à escola.

Este índice passou a ser medido pelo PNUD a partir do relatório de 2010 e neste relatório foi apresentado um índice baseado no ano de 2008, onde este percentual foi de 24,4,

o que significa uma variação de melhora muito discreta quando comparado com o índice mais recentemente publicado. É importante destacar que embora a educação básica não seja foco nesta pesquisa, estes números estão sendo colocados como uma forma de chamar a atenção para a preocupação de organismos internacionais sobre o tema evasão escolar, embora se saiba que os motivos que levam os alunos a evadirem nas diferentes modalidades de ensino possam ser completamente diferentes. Além disso, nota-se que há uma preocupação em torno do assunto desde os anos iniciais do ensino.

Ainda no que se refere à educação básica, segundo o Relatório de Gestão do Ministério da Educação - MEC referente ao exercício 2012, e Decreto nº 6.094 de abril de 2007, o combate à evasão está entre as estratégias de atuação deste ministério:

- I - estabelecer como foco a aprendizagem, apontando resultados concretos a atingir;
- II - alfabetizar as crianças até, no máximo, os oito anos de idade, aferindo os resultados por exame periódico específico;
- III - acompanhar cada aluno da rede individualmente, mediante registro da sua frequência e do seu desempenho em avaliações, que devem ser realizadas periodicamente;
- IV - combater a repetência, dadas as especificidades de cada rede, pela adoção de práticas como aulas de reforço no contra-turno, estudos de recuperação e progressão parcial;
- V - combater a evasão pelo acompanhamento individual das razões da não-frequência do educando e sua superação; [...]

Neste mesmo relatório de gestão, são apontadas ações que vem sendo implementadas como uma das contribuições para a redução da evasão escolar, o transporte escolar:

No que se refere ao transporte escolar, o Programa Caminho da Escola tem o objetivo de renovar e padronizar a frota de veículos escolares, de forma a garantir a segurança e a qualidade do transporte dos estudantes e contribuir para a redução da evasão escolar

A recente Lei 13.005, de 25 de junho de 2014, que aprovou o Plano Nacional de Educação - PNE, apresentou estratégias para a redução da evasão escolar:

- [...]
- 3.13) implementar políticas de prevenção à evasão motivada por preconceito ou quaisquer formas de discriminação, criando rede de proteção contra formas associadas de exclusão;
- [...]
- 7.13) garantir transporte gratuito para todos (as) os (as) estudantes da educação do campo na faixa etária da educação escolar obrigatória, mediante renovação e padronização integral da frota de veículos, de acordo com especificações definidas pelo Instituto Nacional de Metrologia, Qualidade e Tecnologia - INMETRO, e financiamento compartilhado, com participação da União proporcional às necessidades dos entes federados, visando a reduzir a evasão escolar e o tempo médio de deslocamento a partir de cada situação local;

O que se nota a partir do que se coloca a respeito dos níveis iniciais da educação é a institucionalização de ações no sentido de combater o fenômeno da evasão, ainda que não se tenha estudos revelando resultados, diferentemente do Ensino Superior em que não se percebe ações efetivas voltadas ao combate da evasão. No Ensino Superior, os números divulgados não apresentam maiores informações sobre a evasão, dada a superficialidade com que os dados são analisados BAGGI e LOPES (2011), porém demonstram uma preocupação em relação aos índices encontrados. No resumo técnico de 2008, apresentado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira - INEP:

[...]Um objetivo previsto pelo Plano Nacional de Educação (PNE) para o ensino superior é a diminuição na taxa de evasão de alunos. É possível realizar um cálculo aproximado da produtividade dos cursos superiores por meio dos dados do Censo da Educação Superior, considerando um tempo médio de formação de 4 anos[...] (INEP 2008, p.16)

Entre os números apresentados neste resumo técnico está a relação entre ingressantes e concluintes, onde o percentual de concluintes saiu em 2002 de 59,2 para 57,3 em 2008, conforme Tabela 1 a seguir:

Tabela 1: Relação entre ingressos e concluintes no período de 2002 e 2008 apresentada no resumo técnico do censo escolar 2008 divulgado pelo INEP

Ano	Concluintes/ingressos (%)					
	Total	Público				Privada
		Total	Federal	Estadual	Municipal	
2002	59,2	69,5	69,0	76,2	52,5	55,3
2003	58,9	72,6	71,8	71,3	81,5	54,0
2004	60,4	82,7	72,7	95,9	79,8	53,6
2005	59,6	69,7	70,2	69,9	67,3	56,5
2006	58,3	68,6	69,4	70,3	60,6	55,6
2007	58,1	67,4	72,6	63,8	62,4	55,4
2008	57,3	65,0	67,0	64,3	61,2	55,3

Fonte: MEC/INEP/DEEP

No mais recente relatório do Censo da Educação Superior, divulgado pelo INEP, datado de 06 de setembro de 2014, foram apresentados os números de 2006 a 2013. Neste relatório houve atualização nos números de ingressantes e concluintes. Se for realizado um cálculo similar ao realizado em 2008 a relação seria um pouco mais preocupante, conforme Tabela 2:

Tabela 2: Relação entre ingressos e concluintes no período de 2006 e 2013 apresentada no censo escolar 2013 divulgado pelo INEP.

Ano	Concluintes/ingressos (%)					
	Total	Pública				Privada
		Total	Federal	Estadual	Municipal	
2013	45,41%	48,18%	38,15%	58,62%	96,69%	44,64%
2012	50,87%	56,25%	43,83%	72,23%	85,12%	49,48%
2011	43,51%	40,55%	52,64%	31,06%	43,58%	44,39%
2010	45,54%	45,80%	51,54%	41,20%	156,95%	45,48%
2009	48,81%	56,16%	52,76%	64,78%	42,75%	47,11%
2008	48,22%	54,09%	57,78%	50,67%	54,59%	46,74%
2007	47,78%	54,04%	55,02%	52,97%	54,06%	46,00%
2006	49,05%	60,00%	55,29%	67,63%	54,09%	46,16%

Fonte: Construída pelo autor com base nos dados das tabelas 5.4 e 5.5 do Censo Escolar INEP 2013

Ainda nesta perspectiva da evasão no ensino superior, é válido relatar que o Plano de Reestruturação e Expansão das Universidades Federais - REUNI, tem entre os seus principais objetivos “dotar as universidades federais das condições necessárias para ampliação do acesso e permanência na educação superior” MEC (2007, p. 4). Neste mesmo documento, em um diagnóstico a respeito da educação superior no país, retrata, entre outras coisas, os índices alarmantes apresentados pelos alunos dos cursos de graduação.

O Decreto Nº 6.096, DE 24 DE ABRIL DE 2007, que criou o REUNI, traz em seu primeiro artigo:

Art. 1o Fica instituído o Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais - REUNI, com o objetivo de criar condições para a ampliação do acesso e permanência na educação superior, no nível de graduação, pelo melhor aproveitamento da estrutura física e de recursos humanos existentes nas universidades federais.

§ 1o O Programa tem como meta global a elevação gradual da taxa de conclusão média dos cursos de graduação presenciais para noventa por cento e da relação de alunos de graduação em cursos presenciais por professor para dezoito, ao final de cinco anos, a contar do início de cada plano.

§ 2o O Ministério da Educação estabelecerá os parâmetros de cálculo dos indicadores que compõem a meta referida no § 1o.

Os parâmetros para cálculo de que trata o segundo parágrafo do decreto são a Taxa de Conclusão dos Cursos de Graduação (TCG) e a Relação de Alunos de Graduação Presencial por Professor. A TCG é um cálculo muito parecido com aqueles usados pelo INEP, citados

anteriormente, com a diferença que ao invés de calcular quatro anos anteriores à conclusão, consideram-se cinco anos anteriores, ficando a fórmula estabelecida como $TCG = \text{total de diplomados nos cursos de graduação} / \text{total de vagas de ingresso oferecidas cinco anos antes}$.

O que o MEC observa nestas diretrizes, ao estabelecer a TCG como um dos indicadores, é que “[...] o valor de TCG não expressa diretamente as taxas de sucesso observadas nos cursos da universidade, ainda que haja uma relação estreita com fenômenos de retenção e evasão [...]” MEC (2007, p. 14). Estes cálculos apresentados pelo MEC estão mais preocupados em medir o que chamam de “taxas de sucesso”, como um instrumento para medir a qualidade na oferta dos cursos do ensino superior. E como o próprio MEC relata, há grandes indicativos de que o baixo índice de concluintes pode ser causado por questões relacionadas à evasão.

Diante deste conjunto de números apresentados até então, tanto pelo PNUD, IDEB, Censos Educacionais medidos pelo MEC, entre outros, nota-se a dificuldade compreender os aspectos relacionados com a evasão, mas sobretudo, os motivos que levam o aluno ao abandono escolar, além da dificuldade em estabelecer uma regra para medição dos números relacionados à evasão. Há ainda uma dificuldade em de fato apontar se a não conclusão do aluno no tempo regular foi de fato causada pela evasão.

Em especial no ensino superior, que é foco deste estudo, este tema é motivo de preocupação das instituições de ensino, já que a perda de estudantes que iniciam um curso e não concluem geram desperdícios de investimentos sociais, acadêmicos e econômicos. Esta tem sido, segundo Polydoro (2000) e Silva Filho et al. (2007) uma preocupação nacional e internacional. Para Polydoro (2000), desde 1972 o MEC tem realizado estudos com esta temática da evasão escolar. Nos estudos realizados por Silva Filho et al. (2007), que também se basearam em dados publicados pelo INEP, durante o período de 2001 a 2005 chegou a algumas conclusões que podem ajudar a compreender melhor os fatores da evasão escolar, com seguintes destaques: a evasão anual é maior nas instituições de ensino superior privadas, a taxa de evasão é maior em faculdades, quando comparadas com centros universitários e universidades; a região norte do país apresentou a menor taxa; a evasão do ensino superior do Brasil guarda relação com fatores socioeconômicos. Baggi e Lopes (2011) traz uma crítica aos dados oficiais divulgados pelo INEP ao dizer que este não apresenta dados sobre a evasão de forma explícita.

[...]A disponibilidade de dados oficiais pelo governo deveria ser feita de forma mais direta e objetiva para que pudéssemos ter acesso e entender os números que viriam de fato a quantificar e qualificar a evasão. Isto possibilitaria um avanço em

alterações nas políticas públicas que auxiliasse os estudantes e as instituições públicas ou particulares no combate à evasão escolar[...]. (BAGGI E LOPES, 2011, p. 365)

Das tantas formas de análise, dos dados apresentados, fica evidente a inconstância quanto a definição e utilização do termo Evasão. O MEC (1997) num estudo realizado pela Secretaria Executiva de Ensino Superior do MEC - SESU trouxe uma abordagem sobre o tema nas universidades e neste documento chamava atenção para a forma tratamento conceitual do termo:

[...]quando se fala em evasão escolar no ensino superior importa referi-la a seus diferentes níveis ou "locus" dentro do sistema. A pergunta inicial, portanto, é de "qual" evasão estamos falando: evasão de curso? evasão da instituição? ou evasão do próprio sistema?[...] MEC (2007, p.25)

E mais adiante, este mesmo documento ressalva:

[...]O reconhecimento dos óbices que condicionaram este estudo corrobora a certeza de que o conhecimento mais completo e confiável do fenômeno só poderá ser alcançado através de um verdadeiro programa integrado de pesquisas que estabeleça os elos entre os níveis, identifique causas internas e externas, dando assim a necessária dimensão de totalidade característica de uma avaliação do sistema de ensino superior público do país.[...] MEC (2007, p. 26)

Tudo evidencia que embora haja diversidades quanto ao entendimento do termo evasão e sobre os métodos de cálculo, seja qual for a melhor definição, seja qual for o melhor período a analisar para se chegar aos números mais próximos do real, é necessário um estudo aprofundado e integrado, entre governos, instituições e pesquisadores no sentido de chegar às causas deste fenômeno, que tanto preocupa.

3.1 A EVASÃO ESCOLAR NO IFBA

Os Institutos Federais de Educação, com sua nomenclatura e atribuições modificadas a partir da Lei 11.892 de 29 de dezembro de 2008, além manter a estrutura dos antigos Centros Federais de Educação - CEFETs no que diz respeito à oferta de cursos técnicos eleva os institutos ao status de Universidade, muito embora os CEFETs já ofertassem cursos de nível superior, como é o caso do IFBA que desde 1984 tem em seu portfólio de cursos diversos de graduação.

O objetivo de mencionar essa estrutura dos Institutos Federais é mostrar que em qualquer aspecto que se venha analisar algo sobre educação, dentro destes institutos, será

necessário analisar sempre levando em consideração de qual nível está se falando, já que são estruturas e diretrizes diferentes. Para entrar na análise sobre a evasão escolar dentro do IFBA, é necessário compreender suas esferas de atuação, sua estrutura organizacional, suas normas acadêmicas, entre outros aspectos.

Atualmente, o IFBA tem ofertado cursos nas mais variadas modalidades de ensino, que vai desde cursos de Formação Inicial e Continuada, a exemplo do Programa Nacional de Acesso ao Ensino Técnico e Emprego - PRONATEC, até cursos de pós-graduação. Embora o objeto de estudo neste trabalho esteja voltado para a evasão escolar na graduação, a seguir será dada uma atenção inicial às normas que regem os cursos do ensino técnico, nas modalidades integrada, subsequente e concomitante, para em seguida ser analisada a norma acadêmica que rege o ensino superior. Esta análise preliminar busca apresentar quanto complexa é a questão da evasão escolar, mesmo quando o escopo de análise é uma única instituição de ensino. Além disso, esta abordagem busca evidenciar a necessidade da instituição em tratar este tema de forma peculiar.

Os cursos técnicos são regidos pela Proposta de Organização Didática dos Cursos da Educação Profissional Técnica de Nível Médio do CEFET-BA, aprovada pelo Conselho Diretor do então CEFET-BA, em 16 de dezembro de 2008. Esta Organização Didática estabelece, entre outras coisas, uma definição ou mesmo um critério para considerar um aluno evadido. Observa-se que este tópico só é citado na organização didática quando está se tratando do Conselho de Classe, que é o momento em que aluno não aprovado por avaliações regulares é julgado por uma comissão para analisar se ele possui elementos mínimos para aprovação com nota mínima.

Art. 56. No Conselho de Classe Final será confeccionada a Ata contendo a Planilha de Resultados Finais com a carga horária total desenvolvida no período letivo, a nota final dos estudantes em cada disciplina ou competência, o percentual de frequência e a respectiva condição de competência obtida no período letivo, assim definida:

- a) Aprovado (AP);
- b) Reprovado (RP);
- c) Aprovado pelo Conselho (AC);
- d) Reprovado por falta (RF);
- e) Abandono (AB);
- f) Evasão (EV);
- g) Trancamento (TR)

Observa-se que o instituto dá tratamentos diferentes para os casos em que possa se considerar evasão, chamando em um momento de evasão e em outro de abandono. Neste mesmo Art. 56, nos parágrafos seguintes, a norma traz:

§ 4º Será reprovado por falta o estudante que tiver frequência inferior a 75,0% (setenta e cinco por cento) e superior a 50,0% (cinquenta por cento).

§ 5º Será considerado abandono quando o estudante tiver frequência de 25,1% (vinte e cinco vírgula um por cento) a 50,0% (cinquenta por cento).

§ 6º Será considerada evasão quando o estudante tiver frequência de 0,0% (zero por cento) a 25,0% (vinte e cinco por cento).

O que se percebe é que embora esteja se trabalhando com números, os conceitos estão sendo tratados de maneira subjetiva, são valores que, ao menos no referido documento, não conseguem deixar explícitos se há ou não evasão ou mesmo os motivos que levaram este aluno a evadir. Embora a organização didática traga estas diferenças do ponto de vista do percentual de frequência, há que se considerar que estes valores não contribuem para a compreensão sobre os reais motivos, como forma de tomar ações preventivas.

No nível superior, a estrutura é bem diferente. Não há nenhuma instrução do IFBA para tratar do assunto nesta modalidade. O atual documento que regulamenta o ensino superior, chamado de norma acadêmica, em momento algum trata da evasão ou outro termo semelhante, como é o caso da organização didática que normatiza as demais modalidades.

Sabe-se, porém, já que se trata de uma informação pública e divulgada anualmente pela instituição no seu portal (www.portal.ifba.edu.br), que há uma forma de se medir os índices de evasão dentro do IFBA para o ensino superior, assim como nas demais modalidades, mesmo não havendo uma institucionalização do termo, como é o caso aplicado aos cursos da educação básica e dos cursos técnicos subsequentes.

Tabela 3: Relação entre ingressos e concluintes no período de 2010 a 2013 apresentada nos relatórios gestão do IFBA

Índice de Evasão = [(Matrícula Anual – concluinte do 1º semestre) - (Matrícula do 2º Semestre)] X 100 / Nº de Matrícula Anual								
Ano	Total	Pública				Concluinte		Mat. Anual
		Ing. 1º Sem	Mat. 1º Sem	Ing. 2º Sem	Mat. 2º Sem	1º Sem	2º Sem	
2013	15,80%	792	3055	696	3119	38	81	3756
2012	15,94%	745	2567	710	2835	95	19	3486
2011	11,41%	541	2059	710	2390	63	100	2769
2010	12,25%	348	1520	594	1814	41	62	2114

Fonte: Construída pelo autor a partir dos dados dos Relatórios Institucionais de Gestão do IFBA dos anos 2010, 2011, 2012 e 2013

Como demonstrado na tabela 3, o IFBA utiliza o cálculo da evasão anual, considerado por pesquisadores o método mais apropriado quando se trata de análise interna de uma

instituição. Segundo Silva Filho et al. (2007, p. 645), “esta forma de cálculo é mais exata do que a utilizada em alguns trabalhos internacionais[...]” pois permite verificar séries históricas. Este método, embora utilizado atualmente pelos Institutos Federais em seus relatórios de gestão anual, é diferente daquele divulgado pelo INEP que, como dito, tem por objetivo medir as taxas de conclusão de cursos num dado período. Segundo Lobo (2012), este método do INEP é usado para medir dados gerais por diferentes países por não existir um método mais apropriado e que possa ser padronizado universalmente.

Em entrevista realizada com um dos gestores que colaborou com esta pesquisa e que participou do processo de consolidação dos dados para publicação destas informações no relatório de gestão do IFBA, bem como da coleta destes dados, o mesmo informou que ainda não há instrumentos suficientes para contabilizar esta informação, pois não é registrado no sistema acadêmico se o aluno está ou não evadido, mas que a fórmula de cálculo adotada está bem próxima da real situação. Também nesta mesma entrevista o gestor está ciente das dificuldades para esta medição, mas acredita que encontrar os motivos para a evasão é mais importante neste momento que o método para calcular esta quantidade.

Outra observação, nestes mesmos relatórios de gestão publicados, é que cada modalidade utiliza uma forma medir. Isto ocorre porque cada modalidade possui sua forma de organização e por isso necessita de cálculos diferenciados. Segundo o relatório de gestão, as fórmulas adotadas são:

NÍVEL TÉCNICO - MODALIDADE INTEGRADA (Ensino Médio + Técnico)

Índice de Evasão = $\frac{\text{Matrícula de Ano Anterior} - \text{Concluintes de Ano Anterior} + \text{Ingressantes de Ano Atual} - \text{Matrícula de Ano Atual} \times 100}{\text{Matrícula de Ano Atual}}$ (IFBA, 2013, p.49)

ENSINO SUPERIOR

Matrícula Anual = Matrícula do 1º Semestre + Ingressante 2º Semestre.

Índice de Evasão = $\frac{[(\text{Matrícula Anual} - \text{concluinte do 1º semestre}) - (\text{Matrícula do 2º Semestre})] \times 100}{\text{Nº de Matrícula Anual}}$ (IFBA, 2013, p.37)

Segundo Silva Filho et al. (2007), geralmente são usados dois métodos para medir a evasão no ensino superior, a primeira é a Evasão Anual, cálculo adotado pelo IFBA nos seus relatórios de gestão, com pequenas adequações à sua realidade, e a outra seria o que alguns chamam de Evasão Total, que é aquela adotada pelo INEP em números anteriormente apresentadas em que calcula a quantidade de alunos que se formaram alguns anos depois, de acordo com o tempo regular do curso, e subtrai da quantidade de ingressantes. Conforme dito

anteriormente, este segundo método não representa necessariamente um índice de evasão, pois visa medir o índice de titulação. Obviamente, que quanto maior for o índice de titulação dentro do tempo regular, menor será a quantidade de evadidos, mas este método de cálculo não permite uma análise mais individualizada e nem acompanhamento de possíveis fatores. O IFBA utiliza o cálculo da evasão anual com a adaptação para considerar apenas o intervalo de um ano.

Para apoiar numa melhor compreensão daquilo que o IFBA tem entendido o fenômeno da evasão, diante da falta de documentos que o definam, durante os meses de março e abril de 2015 foram realizadas entrevistas com quatro gestores do IFBA e teve como objetivo compreender o significado da evasão para estes gestores, bem como os métodos até então usados para analisar tal fenômeno, para orientar as decisões a serem tomadas na construção do modelo de dados multidimensional. Estas entrevistas foram guiadas a partir de um formulário composto de oito questões, conforme APÊNDICE B e foram devidamente autorizadas conforme termo de consentimento assinada pelos participantes, como pode ser visualizado no modelo apresentado no APÊNDICE A. É importante destacar que em alguns casos as respostas dos entrevistados suscitaram a necessidade de novas perguntas. Estas foram feitas de forma individual ao longo da entrevista e tiveram como principal objetivo aprofundar o entendimento acerca de alguma nova questão surgida.

A primeira questão que norteou todas as entrevistas foi “o que você entende como evasão escolar?”. As respostas evidenciaram dois entendimentos distintos a cerca do tema em questão, conforme pode ser verificado nas respostas dos entrevistados. Para o gestor A “é toda saída do aluno, sem conclusão do curso, seja qual for o motivo...”. Esta resposta é corroborada pelo gestor C ao afirmar que entende que é evasão é “toda interrupção do ciclo de estudo”. Um segundo ponto de vista é percebido na resposta do gestor B, que entende ser a perda do direito do aluno sob a ótica da norma acadêmica vigente ao afirmar que “Quando o aluno perde o direito à matrícula” este é considerado um aluno evadido e ratifica este ponto de vista afirmando que “quando o aluno sai e não atingiu as condições impostas pela norma para perda do direito a vaga, consideramos uma retenção e não uma evasão”. Este não é um ponto de vista isolado. O entrevistado D também entende desta forma, segundo ele “consideramos evadido quando o aluno perde o direito à vaga com base nas normas”.

Quando o questionamento foi “Qual a relação que você entende existir entre Evasão, Abandono e Desistência?”, as respostas não variaram muito, como se pode observar nas

respostas dos entrevistados B e C, respectivamente: “Acho que estas nomenclaturas não são relevantes, acho que devemos parar de usá-las e tratarmos como uma coisa só” e “Para mim todos são evasão, o que muda é apenas o motivo. Se houve interrupção do ciclo de estudo é evasão.” Para estas duas perguntas iniciais é possível afirmar que, embora haja uma convergência para um único entendimento sobre a expressão “evasão”, mesmo que havendo distinção em uma das normas (organização didática do ensino médio) e a inexistência da expressão do termo na norma do ensino superior, há divergência sobre a concepção daquilo que venha a ser evasão sob o ponto de vista destes gestores. O que a Norma Acadêmica do IFBA (2007) aborda sobre os critérios para que o aluno tenha direito à permanência, em seu 10º artigo é:

ART. 10. O aluno perderá o direito à matrícula institucional quando estiver enquadrado em qualquer um dos incisos abaixo:

I for reprovado por conceito e/ou faltas em todas as disciplinas no 1º semestre acadêmico;

II deixar de efetivar a inscrição em disciplinas por um semestre letivo;

III for reprovado por conceito e/ou faltas em todas as disciplinas em que estiver Inscrito por dois semestres letivos consecutivos ou alternados;

IV for reprovado na mesma disciplina em 03 (três) semestres consecutivos ou alternados;

V ultrapassar o prazo máximo de integralização curricular fixado pelo Projeto do Curso.

Como visto acima, a exceção do inciso II, todos os demais não implicam necessariamente em evasão, salvo nos casos em que a reprovação se dê por falta às aulas. A partir desta análise, e considerando a observância à norma, há que se considerar que existe convergência entre os entrevistados para um entendimento de que a evasão escolar se trata da interrupção do ciclo, o que condiz com o método usado pelo instituto para medir sua quantidade de evadidos, que verifica dentro de um período de um ano os alunos que ainda não concluíram o curso, mas que por algum motivo interromperam seu ciclo. Dito isto, consideradas as observações dos entrevistados e a norma acadêmica vigente, é possível inferir que o IFBA trata a evasão escolar no ensino superior como uma interrupção do ciclo de estudos na instituição pelo aluno, num dado período de tempo, que de acordo com seu método de cálculo, no intervalo de um ano.

Um importante ponto durante as entrevistas foi quanto aos fatores que estas pessoas entendem ser determinantes para que os alunos do ensino superior no IFBA venham a evadir. Como podem ser observadas a seguir, as respostas variaram um pouco, mas há vários pontos em comum. O gestor A respondeu:

[...]No caso dos cursos das áreas exatas, o fator alto índice de reprovação é fator determinante, principalmente alunos que vieram de escola pública que não tiveram base nas disciplinas de cálculo. Falta de condições financeiras. Este fator muitas vezes exige que aluno busque trabalhar para sobrevivência e como consequência redução do rendimento escolar que leva ao desestímulo e posterior evasão[...].

Para o gestor B os fatores determinantes são:

[...]Não temos na instituição pesquisas muito estruturadas a respeito, mas um fato que tem sido considerado bastante recorrente é o acesso, a dificuldade de transporte, tanto na cidade grande como em cidades menores do interior, tem sido este o fator que os alunos remetem, a exemplo de falta de linhas ônibus suficientes no entorno que fazem com que os alunos tenham que andar a pé um trecho em busca de um transporte público, colocando em risco sua segurança física, isso no caso das metrópoles. No caso do interior, o problema é um pouco pior, pois na maioria das vezes não existe nenhum transporte, o aluno fica a mercê de transportes alternativos aventurando um transporte. Curiosamente, estes dados apontados pelos alunos foram semelhantes aos dados apresentados pela SETEC num seminário. Outro ponto também é relacionado aos horários das aulas no casos dos cursos diurno, mas isso não nos parece evidente, porque a maior parte dos nossos cursos são noturnos, precisaríamos investigar. Outro ponto que acreditamos ser um elemento para a evasão é o alto índice de reprovação nos cursos de exatas[...].

O gestor C pontuou algumas fatores como “questão financeira que impõe ao aluno ter que abandonar[...]. Uma questão que considero importante para a evasão é o fato do aluno descobrir que ao final da sua formação ele não terá o retorno financeiro que esperava[...].” É possível destacar ao menos cinco pontos citados de forma objetiva e explícita, como a questão do transporte público, condições financeiras, tipificação dos cursos de acordo com a área de conhecimento, a origem escolar do aluno, se veio de escola pública ou privada e o turno em que são ofertados.

Estes gestores educacionais, formados por técnicos em educação e docentes, também foram provocados a responderem se conhecem ou desenvolvem alguma ação com o intuito de reduzir os índices de evasão no IFBA, dois deles responderam desconhecer, já os gestores A e B responderam, respectivamente, “criação de mecanismos de divulgação do perfil do curso[...].Os próprios estudantes montavam stands para apresentar o curso aos pré-vestibulandos” e

Uma coisa que já temos feito é a revisão das grades curriculares, para distribuir disciplinas ao longo do curso para que reduza a existência de semestres com muitas disciplinas consideradas pesadas agrupadas. Isso tem melhorado para as engenharias. Outra coisa que temos tentado fazer é através da ampliação da oferta de projetos de pesquisa e extensão[...]. como forma de estimular os alunos a permanecer no curso[...].Para o caso do transporte, tem havido uma constante ação dos campi na

solicitação ao poder público pela melhoria da oferta de transporte público e da cobertura de policiamento nas áreas em que os alunos pegam estes transportes[...]

Durante a entrevista foi comunicado aos gestores da análise realizada em torno do relatório de gestão divulgado anualmente pelo IFBA que traz os índices de evasão. E para os entrevistados foi questionado se conhecem os métodos usados neste relatório para a medição destes índices ou outros métodos e se concordam. As respostas foram unânimes em afirmarem conhecer apenas este método divulgado publicamente pelo instituto. Quanto à concordância, também houve unanimidade em discordarem da fonte dos dados e não necessariamente da fórmula adotada. Duas das respostas fortalecem esta pesquisa, como observa-se a seguir. Para o gestor B:

[...]ele é baseado nos dados armazenados no SISTEC, dados estes alimentados manualmente pelos campi e que nem sempre estão de acordo com os dados acadêmicos, pois o sistema é muito instável e nem sempre é possível alimentá-lo completamente dentro do prazo estabelecido pelo MEC. O fato da gente não ter um sistema acadêmico unificado a gente não consegue fazer extração de indicadores, pois cada campus ainda usa uma forma de registro ou um sistema diferente, e aí a gente precisa ficar solicitando à DGTI a extração destes dados sempre que precisamos e ainda precisa sair reunindo estes dados de vários locais[...].

O gestor C também enfatiza a fonte dos dados não confiável, assim como cita que a seu ver este relatório, a partir desta origem, tem como objetivo apenas servir como base para a matriz orçamentária do instituto e não efetivamente para a análise sobre a evasão, conforme segue:

O problema é a fonte dos dados, que é de um sistema complicado e que torna os dados não confiáveis. O sistema é o SISTEC, que a meu ver não atende, é muito instável e não permite que se mantenha histórico da vida escolar do aluno, é um sistema que se alimenta por ciclo, usado para definir a matriz orçamentária do instituto, portanto não permite contribuir para a análise fatores como este da evasão escolar.

O sistema em questão - SISTEC -, é mantido pelo MEC, disponível através do endereço sistec.mec.gov.br.

Por fim, foi perguntado aos entrevistados “O que você indicaria como alternativa para redução da evasão escolar?” e entre as respostas obtidas tiveram, respectivamente dos entrevistados A, B e C que “a divulgação dos cursos e melhor distribuição da carga horária de maneira a permitir a conciliação entre o trabalho e o estudo, sobretudo para os cursos noturnos”, “fazer um trabalho interno de mapeamento dos motivos pelos quais as pessoas evadem, pois os motivos podem ser vários e essa coleta sistemática de dados ainda não

existe[...] e ações que permitam ouvir o aluno para que ele possa se posicionar indicando o motivo que o levou a evadir” e “[...]acho que algo importante é a realização de pesquisas que evidencie os motivos que levam a evadir para que assim possamos tomar medidas visando a redução[...]”.

3.2 CONSIDERAÇÕES SOBRE A EVASÃO ESCOLAR

Diante do exposto, números e conceitos, não resta dúvidas de que evasão é de fato uma questão social a ser mitigada. Os números apresentados, tanto pelos institutos de pesquisa nacionais como pelo próprio IFBA, são preocupantes. A medida da Evasão Total não consegue expressar se de fato houve evasão já que durante o período de curso outros eventos como reprovação e conseqüente repetição de disciplinas, matrícula em quantidades de disciplinas menores que aquelas propostas nas grades curriculares, trancamentos, transferências, entre outros fatores que elevam o tempo de conclusão do curso o que nestes casos não haveria necessariamente uma evasão. O cálculo da evasão anual por sua vez permite verificar se houve uma interrupção, mesmo que temporária, do ciclo de estudos, portanto se mostra mais adequada.

Os números, por si só, não demonstram de maneira clara os fatores que causam a evasão. Isto reforça a necessidade de implementar dentro das instituições de ensino instrumentos que consigam contribuir neste entendimento. Estas soluções precisam de elementos que possam levantar os motivos que levaram os alunos a abandonarem a instituição, bem como instrumentos que acompanhem melhor o aluno durante sua vida escolar de modo que seja possível tomar medidas que antecipem e evitem que estes alunos evadam.

Silva Filho et al. (2007, p. 644) corrobora neste sentido e com o propósito desta pesquisa, afirmando que

[...]O estudo interno, realizado por uma IES com base em seus dados, pode ser muitas vezes mais detalhado porque é possível institucionalizar-se um mecanismo de acompanhamento da evasão, registrando os diversos casos, agrupando e analisando subgrupos, ou diferentes situações[...].e, a partir daí, organizar tabelas e gráficos que demonstrem a evolução da evasão para buscar formas de combatê-la com fundamento nos resultados.

Ao analisar Plano de Desenvolvimento Institucional – PDI do IFBA 2014-2018 verifica-se que o IFBA incluiu como um dos seus objetivos do ensino a criação de mecanismos para a redução da evasão e estabelece, ao falar das políticas de ensino, o monitoramento da evasão como umas das diretrizes para o êxito no ensino e como forma de acompanhar e agir sobre as causas dessa evasão, mas ainda não há ações específicas com este fim. O PDI é um documento construído coletivamente pela comunidade acadêmica e classifica-se como “uma espécie de documento síntese, no qual se explicitam as diretrizes gerais que irão ensejar, em articulação com outras ações empreendidas, o cumprimento da missão do instituto e do conseqüente crescimento do mesmo”. Embora apenas cite a evasão nestes dois momentos, nota-se, neste sentido, que um projeto desta natureza, que auxilie gestores no entendimento sobre a evasão escolar, está em total sinergia com as políticas da instituição, além de evidenciar o longo caminho que se tem pela frente para reduzir estes índices que tanto assustam.

Esta pesquisa utilizará como base para a medição o cálculo da evasão anual adotado no IFBA para compor seus relatórios de gestão por entender a complexidade em instituir novos métodos que exigiria longos estudos com equipes multidisciplinares. O foco está na construção de um modelo multidimensional que permita a extração de informações em torno do cálculo atualmente adotado, mas que auxiliem os gestores na identificação dos motivos que levam estes alunos a evadirem para contribuir no controle desta evasão.

O capítulo a seguir vem trazendo uma abordagem teórica sobre as ferramentas que estão envolvidas no processo de desenvolvimento da solução de BI que busca apoiar neste entendimento sobre os motivos que tem levado os alunos do ensino superior do IFBA *campus* Salvador a evadirem.

4 O BUSINESS INTELLIGENCE E SUAS FERRAMENTAS

Este capítulo faz uma abordagem mais teórica acerca das principais tecnologias do *Business Intelligence* envolvidas neste trabalho, bem como explicita, sempre que necessário, as decisões técnicas adotadas para o desenvolvimento deste projeto. Dada a relevância da tecnologia BI para o projeto, no primeiro momento será feito um breve apanhado histórico sobre o seu surgimento até os dias atuais.

4.1 O BUSINESS INTELLIGENCE

O termo Tecnologia da Informação e Comunicação é o resultado da evolução dessa recente área do conhecimento que por muito tempo foi chamada de Processamento de Dados. A informática dentro das organizações tem cada dia mais atuado como estratégica do que como mero processador de dados. Segundo Meireles (2004), a computação começou a ser difundida por volta nos anos de 50 e 60, com o início do uso comercial, e mais tarde com Internet, grande responsável hoje por esse exponencial crescimento da difusão de informações. Trata-se, portanto, de uma recente área do conhecimento, em constante evolução.

Embora não seja este o foco aqui, é de fundamental importância este contexto para situar o BI que é o resultado deste processo evolucionário e porque não chamá-lo de revolucionário, já que tem participação efetiva no processo de transformação de dado em informação e conhecimento. Segundo Barbieri (2011, p. 1), o professor do Instituto de Tecnologia de Massachussets, Seymour Papper, por volta de 1970, afirmou “que os dados e seus correlatos seriam responsáveis por uma revolução na sociedade”. Esta revolução chegou e já faz parte da vida cotidiana das pessoas, embora à época este assunto foi motivo de desconfiança pela indústria de TI. É que até então os dados, no processo de desenvolvimento de sistemas de informação, eram apenas coadjuvantes, muito embora os primeiros estudos para criação dos bancos de dados tinham, à época de sua criação, a preocupação em tratar de coisas diferentes, sistemas de um lado e dados de outro, como é até hoje. Porém na maior parte do tempo os sistemas tiveram mais notoriedade que os dados.

Este fato se dava porque neste período, que remonta dos anos 60 até 80, fora marcado por uma busca por tecnologias que permitissem o armazenamento e recuperação de grandes

quantidades de dados e um acesso mais fácil a estes dados, portanto, todo o foco estava voltado para a tecnologia (*softwares*, algoritmos e *hardware*) e não para o negócio (dados e informações). Estes estudos, como afirma Inmon (1997), deram origem a tecnologias como Sistemas de Gerenciamento de Banco de Dados - SGBD, os Bancos de Dados Relacionais, o Advento do DASD (*Direct Access Storage Sevice*, ou Dispositivo de Armazenamento de Acesso de Direto) e que foram o pontapé para esta revolução dos dados, conclamada por Seymour conforme Barbieri (2011), substituindo os chamados arquivos mestres, armazenados em fitas e cartões perfurados, com estruturas rígidas, usados até então como o principal meio de armazenamento.

Foi por volta de 1980, segundo Barbieri (2011), que os dados, pela primeira vez, atingiram uma espécie de “estrelato”. Nesta época surgiram a administração de dados, a modelagem de dados, a engenharia da informação, análise de dados. A idéia de Processamento de Dados começa a perder forma e começa a ganhar forma e força a expressão Tecnologia da Informação. Mas, ainda segundo Barbieri (2011), mesmo que todos estes adventos tenham tido grande importância, o foco continuou sendo mais tecnológico que comercial e somente nos anos 90 é que aquilo que Seymour apontou como tendência em 1970 ganha de fato seu espaço. Um período importante tanto para a informação como para a comunicação, já que esta década marca o casamento entre estes dois conceitos, surgindo ai a Tecnologia da Informação e Comunicação, a TIC assim chamada.

Diferentemente da Revolução Industrial, quando as benesses não chegavam a todos diretamente, a revolução dos dados e da informação, capitaneada pela internet, tornou-se democrática, invasiva e de amplo alcance. O cidadão comum[...] podia navegar pela grande teia[...]. Perigava, por um lado, de se enrolar nos labirintos do excesso de opções, enquanto, por outro, desfrutava de uma tecnologia que lhe concedeu um grau de liberdade e profundidade jamais imaginados. Tudo isso suportado por um arsenal que misturava computadores e redes, e lhe oferecia principalmente... dados que ele começava a transformar no combustível mais precioso dos anos 2000: a informação. (BARBIERI, 2011, p.2)

Este período da década de 90, a grande maioria das empresas de grande porte possuía centros de processamento de dados e foi um período em que muitas organizações estavam muito preocupadas em desenvolver aplicações para manter seus processos operacionais e isso fez com que estas empresas passassem a manter grandes estoques de dados, com poucas tecnologias que permitisse a recuperação e o tratamento destes volumes de dados. Segundo Primak (2008), o termo *Business Intelligence* surgiu ainda na década de 80, no Gartner Group e fazia referência ao processo de coleta, organização, análise, compartilhamento e monitoração dos dados contidos nestes repositórios de forma inteligente. Ainda segundo

Primak (2008), foi nos anos 90 que o setor corporativo passou a se interessar pelo BI, principalmente quando o termo começou a ser difundido como um processo de evolução do EIS (*Executive Information Systems*), um tipo de sistema criado no final da década 70, cujo objetivo é fornecer informações para alta gestão, de forma simples e amigável, a partir de uma base de dados.

A partir de então, o termo veio ganhando mais abrangência e importância, incorporando uma série de ferramentas e soluções para tratamento de dados. Nesta história, os Sistemas Integrados de Gestão, conhecidos como ERP (*Enterprise Resource Planning*) tiveram um papel importante. Estes sistemas têm como finalidade facilitar o processo operacional das empresas e armazenam grandes volumes de dados. Foi neste contexto que as organizações passaram a perceber que apenas armazenar grandes quantidades de dados não teriam tanta importância se não fosse possível processar esses dados e extrair deles informações úteis, de uma maneira simples e consolidada.

Embora Barbieri (2011) já venha apontando uma nova sigla para o BI, que ele denominou de BI2, apresentando novos avanços e tendências tecnológicas para tratamento de grandes volumes de dados, muitas organizações ainda desconhecem a tecnologia do BI. No Brasil, segundo Primak (2008), os principais usuários do BI são instituições financeiras, empresas de telecomunicações, seguradoras ou outras instituições de grande porte, privadas, com foco nos lucros, muito pouco tem se desenvolvido em instituições de menor porte e nas instituições públicas. Barbieri (2011), em suas novas perspectivas, apresenta soluções e estudos em andamento que permitirão aos gestores mais que um apoio à tomada de decisão, como tem se apresentado as soluções de BI, mas soluções com análises preditivas em tempo real.

[...]Outras fronteiras estão sendo discutidas nas trincheiras acadêmicas e, embora ainda mostrem pequena musculatura, já têm os seus embriões em formação. Entre elas estão: gerência de entendimento (*understanding management*) e a gerência de sabedoria (*wisdom management*), em que o componente intuição será contemplado, além de outros conceitos, ainda por explorar[...]

Até mesmo os autores mais tradicionais que discutem conceitos acerca dos dados têm certa dificuldade na definição do que seria BI. A TIC de uma maneira geral evolui muito rapidamente. A Internet, certamente, tem sido uma das grandes impulsionadoras dos avanços tanto da TIC como do BI. O BI se apresenta, segundo Barbieri (2011), como um “guarda-chuva” conceitual, composto por um conjunto de ferramentas destinadas ao tratamento e análise de dados, com vistas a apoiar no processo de tomada de decisões. De uma maneira

geral, o que as soluções de BI fazem são prover ferramentas capazes de tratar grandes volumes de dados, históricos, gerados a partir das operações transacionais, armazenadas das mais variadas formas e lugares, organizando-os em repositórios estruturados, oferecendo ferramentas de visualização de fácil acesso e com grande poder de processamento sobre estes dados. Barbieri (2011, p. 95) tem chamado de “formas evoluídas de tratamento de informações”.

De acordo com Both e Bill, citados por França et al. (2013, p. 82), entre as características do BI, destacam-se: “Extrair e integrar dados de múltiplas fontes; Fazer uso da experiência; Trabalhar com hipóteses e simulações; Procurar relações de causa e efeito e; Transformar os registros obtidos em informação útil para o conhecimento empresarial”.

Dentre as diversas ferramentas e tecnologias associadas, destacam-se a modelagem de dados através de *Data Warehouse – DW*, *Data Marts*, assim como a mineração de dados (*data mining*) e o OLAP. Quanto às ferramentas, atualmente existem diversas soluções de ferramentas proprietárias, muitas delas desenvolvidas pelas mesmas empresas que desenvolvem os bancos de dados, a exemplo da IBM, Oracle, Microsoft, mas há destaque para soluções abertas, *open source*, como é o caso da suíte de ferramentas *Pentaho*. Neste trabalho será abordado o uso da modelagem dimensional, através das tecnologias de DW e *Data Marts*, além do OLAP e da suíte *Pentaho*, adotadas para apoiar no desenvolvimento deste projeto.

4.2 O DATA WHAREHOUSE

As organizações vão armazenando ao longo da sua trajetória grandes quantidades de dados relacionadas aos negócios por elas operacionalizadas, sobretudo aquelas que já lidam com sistemas de informações há certo tempo. De início estas organizações tinham como principal preocupação o correto armazenamento destes dados. Surge então a preocupação de recuperar todos estes dados, tratá-los de forma inteligente para criar estratégias futuras e ajudar a resolver também problemas já conhecidos.

O conceito de *Data Warehouse – DW* surge como proposta para o armazenamento unificado de dados considerados estratégicos pelas organizações, até então espalhados pelos diversos sistemas e processos geridos por elas. O armazenamento de dados em diferentes

bancos de dados, arquivos distintos, planilhas e outras fontes é algo natural e bastante recorrente nas organizações de modo geral.

A expressão *data warehouse* (armazém de dados) tem sido bastante usada em referência a um banco de dados capaz de armazenar um grande volume de dados históricos e estruturado de forma otimizada na recuperação de dados, com objetivo de fornecer um ambiente de apoio a tomada de decisão.

Inmon (1997) considera DW o coração dos Sistemas de Apoio a Decisão – SAD e define como um conjunto de dados orientados a assuntos, não-valátil, variável em relação ao tempo, integrado e de apoio às decisões gerenciais. Esta característica da não volatilidade é um dos aspectos que o difere dos bancos de dados transacionais, usados por aplicações de uso operacional, já que, em sua grande maioria, os DW são projetos para realização de consultas. O autor Kimball et al. (2002) acrescenta que o DW deve possibilitar que as informações da empresa possam ser facilmente acessadas, de forma compreensível e consistente, apresentados de forma intuitiva não apenas para equipe técnica de desenvolvedores, mas principalmente para a equipe de gestão da organização.

Barbieri (2011) classifica os dados no DW como dados informacionais, em contraste àqueles em que classifica de dados operacionais, mantidos pelos sistemas transacionais. Na Tabela 4 ele detalha as principais diferenças entre estes dois tipos de dados, quanto à sua natureza.

Tabela 4: Dados Operacionais x Dados Informacionais

Característica	Dados Operacionais	Dados Informacionais
1. Conteúdo	Valores Correntes	Valores sumarizados, calculados, integrados de várias fontes
2. Organização dos dados	Por aplicação / sistema informacional	Por assunto/ negócios
3. Natureza dos dados	Dinâmica	Estática até o refreshment dos dados, de tempos em tempos
4. Formato das estruturas	Relacional, próprio para computação transacional	Dimensional, simplificado, próprio para atividades analíticas
5. Atualização dos dados	Atualização campo a campo	Acesso granular ou agregado, normalmente sem update direto
6. Uso	Altamente estruturado em tabelas, processamento repetitivo	Estruturado em fatos e dimensões com processamento analítico/preditivo
7. Tempo de Resposta	Otimizado para faixas abaixo de 1 seg	Análises mais complexas, com tempos de respostas maiores

Fonte: Tabela 5.2 – Comparação entre dados de natureza operacional e informacional. Barbieri (2011, p. 108)

Estes dois autores, Bill Inmon e Ralph Kimball, conforme Barbieri (2011), são os mais referenciados e respeitados quando o assunto é DW e embora concordem na grande maioria dos conceitos acerca do assunto, tem discussões interessantes na perspectiva da concepção do DW. Um dos principais pontos de divergência é quanto à característica “dados orientados a assunto”. Esta característica deu origem ao conceito, dentro do DW, de *Data Mart* - DM, que pode ser definido como pequenos modelos de dados, orientados a assuntos, onde estes modelos juntos formam o DW corporativo (visão de Kimball) ou o contrário, onde num único modelo corporativo é possível encontrar deferentes assuntos (visão de Inmon). Ambos concordam em termos daquilo que significa tanto o DW como o DM, porém Inmon defende que o processo de construção se dá numa visão *top-down* (de cima para baixo), em que se constrói um único modelo corporativo e que o DM é obtido através dessa visão *top-down*. Em paralelo, Kimball defende a visão *bottom-up* (de baixo para cima), em que a construção se dá de forma incremental, por assunto, e o conceito de DW ocorre na união de todos os DM.

Esta abordagem de Bill Inmon é um pouco mais tradicional, baseada na modelagem de banco de dados relacional, mas segundo Barbieri (2011, p. 113) esta forma de concepção “por vezes se mostrou rígido e de difícil consecução”. A abordagem de Ralph Kimball “veio como um estilo mais simples e incremental [...] a metodologia [...] aponta para projetos de *data marts* separados, que deverão ser integrados na medida da sua evolução” BARIEBI (2011, p. 113).

Neste projeto será adotada esta perspectiva de Kimball, trabalhando com *Data Marts* e num processo incremental de desenvolvimento, já que estaremos trabalhando, neste primeiro momento, com um assunto em específico (evasão escolar). O fato de trabalhar com um assunto específico não quer dizer que a organização como um todo não deva ser considerada. Isto, inclusive, é uma das preocupações do próprio Kimball ao defender esta perspectiva, a de não se criar ilhas de dados. O projeto deve ser construído de forma incremental, por assunto, por partes, sempre pensando no todo.

O ambiente em que o DW se insere, segundo Kimball et al. (2002), considerando quatro componentes separados e distintos: sistemas operacionais de origem, *data staging area* (área de preparação dos dados), área de apresentação de dados e ferramentas de acesso a dados. A Figura 1 mostra graficamente este ambiente.

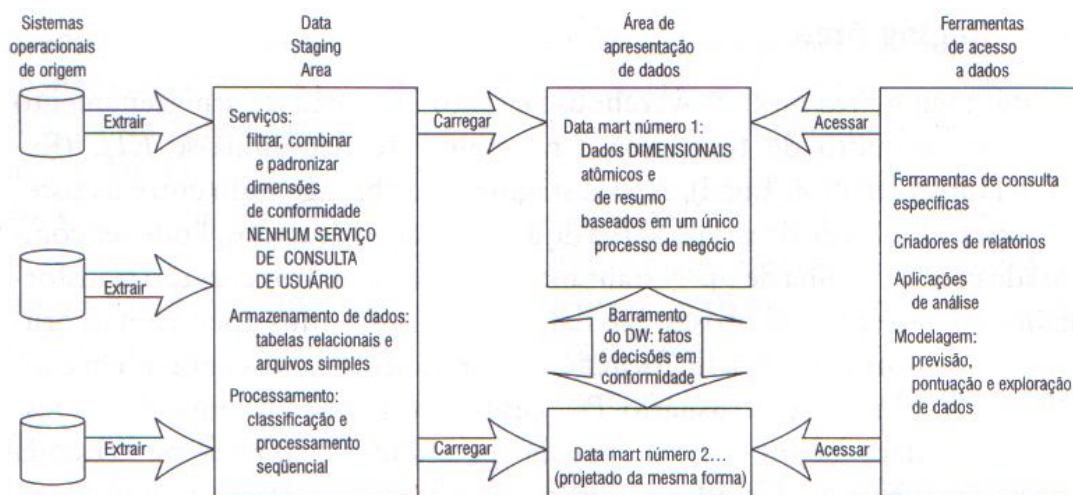


Figura 1 – Elementos básicos do data warehouse.

Fonte: KIMBALL, et al., (2002, p. 9).

Neste projeto, os sistemas operacionais de origem são representados pelo Sistema de Registros Acadêmicos do IFBA – SICAD, de onde serão extraídos os dados operacionais. A *Data Staging Area* é a área de preparação dos dados antes de serem transformados para evitar manipulação sobre os dados originais. A área de apresentação de dados serão os modelos dimensionais e sua arquitetura em volta, que será detalhado no Capítulo 6 durante o processo de modelagem e desenvolvimento do projeto. Por fim, as ferramentas de visualização ou de acesso a dados, que neste projeto será adotada a ferramenta OLAP *Saiku*, que compõe a suíte de soluções de BI da *Pentaho*.

4.2.1 Modelagem Dimensional

A modelagem dimensional foi criada para atender aos projetos de DW. É o processo de construção de um modelo com uma estrutura projetada para a recuperação dos dados, diferente do modelo relacional, em que as tabelas estão organizadas num alto grau de relacionamento entre si, projetado para a manutenção dos dados. O modelo dimensional tem como objetivo permitir uma melhor extração e análise dos dados. Recebe este nome por ser formado basicamente por dois tipos de tabelas, as tabelas de dimensão, que são os pontos de entrada das informações e as tabelas fato que são aquelas que armazenam as métricas de um determinado assunto.

[...] Isto significa que, numa estrutura dimensional, os dados estarão em uma forma quase estelar, em que várias tabelas de entradas estarão se relacionando com algumas (poucas) tabelas de informações, criando uma notação mais sintética, legível e objetiva. O modelo dimensional oferece, clara e diretamente, os elementos

de que se precisa para buscar as informações (“fato”) via dimensões de referências[...]. BARBIERI (2011, p.97)

Na Tabela 5 é possível perceber as diferenças entre o modelo relacional e o modelo dimensional aqui tratado.

Tabela 5: Modelo Dimensional x Modelo Relacional

MODELO DIMENSIONAL	MODELO RELACIONAL
Padrão de estrutura mais fácil e intuitiva	Modelo mais complexo
Anterior ao MER – anos 1990	Ênfase nos bancos de dados relacionais – anos 1970
Tabelas fato e tabelas dimensão	Tabelas que representam dados e relacionamentos
Tabelas fato são o núcleo – normalizadas	Todas as tabelas são normalmente normalizadas
Tabelas de dimensão são os pontos de entrada e de filtro inicial e podem ser desnormalizadas	As tabelas são indistintamente acessadas, não sendo caracterizados pontos de entradas
Modelo mais facilmente <i>joined</i>	Maior dificuldade de <i>join</i> pelo número maior de tabelas
Leitura mais fácil do modelo por usuários não especializados	Maior dificuldade de leitura pelo usuário não especializado

Fonte: Tabela 5.1. BARBIERI (2011, p. 99)

No processo de construção de um modelo dimensional existem também duas abordagens de modelagem, a *Star schema/Star join* ou Esquema Estrela e a *Snowflake* ou Flocos de Neve. Ambas recebem este nome pela forma aparente. A diferença básica está na modelagem das tabelas dimensão. Enquanto no modelo estrela as dimensões são desnormalizadas, ou seja, os dados são agrupados e armazenados numa única tabela, ocorrendo um processo inverso ao da modelagem relacional, no modelo Flocos de Neve a normalização das tabelas dimensão é mantida, evitando a redundância de informações. Essa abordagem flocos de neves, embora pareça ser um pouco mais correta, tem grande impacto na performance (KIMBALL et al.,2002).

O esquema estrela também tem sido mais adotado que o floco de neves, pela simplicidade no desenvolvimento, assim como questões de performance, já que segundo Kimball et al. (2002) as tabelas dimensões geralmente representam apenas 10% do espaço de armazenamento, contra 90% das tabelas fato, de modo que o armazenamento de dados redundantes das tabelas de dimensão não interfiram tanto no espaço e ganha muito na performance. Neste projeto será adotado o modelo estrela, pois os dados acadêmicos mais granulares e com maior volume, serão informações que comporão as tabelas fato. Além disso, este modelo é mais aderente às tecnologias adotadas.

Além da definição quanto à estratégia ou o esquema a ser utilizado, Kimball et. al. (2002) e Babieri (2011), definem, em suma, quatro passos necessários para a modelagem de um DW:

- 1 – Definir o processo de negócio a ser modelado;
- 2 – Identificar a granularidade do processo, identificando qual o menor nível de detalhe que se pretende analisar os dados.
- 3 – Escolher as dimensões. Sob quais perspectivas pretende-se analisar as informações. Em seguida defini-se os atributos que serão os níveis de hierarquia de cada dimensão. Se a granularidade for bem especificada, facilita a definição das dimensões;
- 4 – Por fim, define-se os fatos que são as medidas que pretende-se analisar. Devem está coerentes com o passo 2.

4.3 FERRAMENTAS DE VISUALIZAÇÃO OLAP

A OLAP (*On-Line Analytical Processing*, ou processamento analítico on-line) se apresenta como uma tecnologia composta por um conjunto de ferramentas projetadas para analisar dados de suporte a decisão, armazenadas no DW. Ferramentas OLAP num projeto BI está situada entre as soluções *front end*, ou seja, destinadas a usuários finais, amigáveis e de fácil uso, como afirma Primak (2013). As aplicações OLAP, diferentemente das aplicações OLTP (*On-Line Transaction Processing*, ou processamento de transações on-line) que se preocupam em registrar todas as transações contidas nas operações diárias das organizações, tem seu foco na recuperação de informações.

As ferramentas OLAP estão cada vez mais poderosas, com grande capacidade de processamento de grandes volumes de dados, além da facilidade de uso, em que o próprio usuário pode gerar seus próprios relatórios On-Line sem a necessidade de um profissional de informática. Uma vez projetado e montado o modelo dimensional, a ferramenta OLAP poderá gerar visões sobre as dimensões e fatos disponíveis, analisando os dados sob a perspectiva que desejar.

Entre as principais características e operações que as ferramentas OLAP oferecem, pode-se destacar, segundo Barbieri (2011):

- **Drill-down:** Capacidade de desagregar uma dimensão, detalhando seus atributos. Significa mergulhar na hierarquia dos dados de uma dimensão.
- **Drill-up/Roll-up:** Realiza a função inversa do Drill-down, que é a capacidade de analisar de uma forma agregada.
- **Drill-through:** Permite que o usuário passe de uma dimensão para a outra analisando a mesma métrica. Representa a capacidade de análise de uma mesma métrica sob diversas perspectivas.
- **Slice and Dice:** É a capacidade de análise de métricas/indicadores cruzando diversas dimensões, simultaneamente.
- **Pivot:** Muda o eixo de visualização

A Figura 2, em forma de múltiplos cubos, é uma representação clássica para soluções OLAP. A forma cúbica de representar o OLAP é originada desta capacidade de visão multidimensional oferecida pela solução.

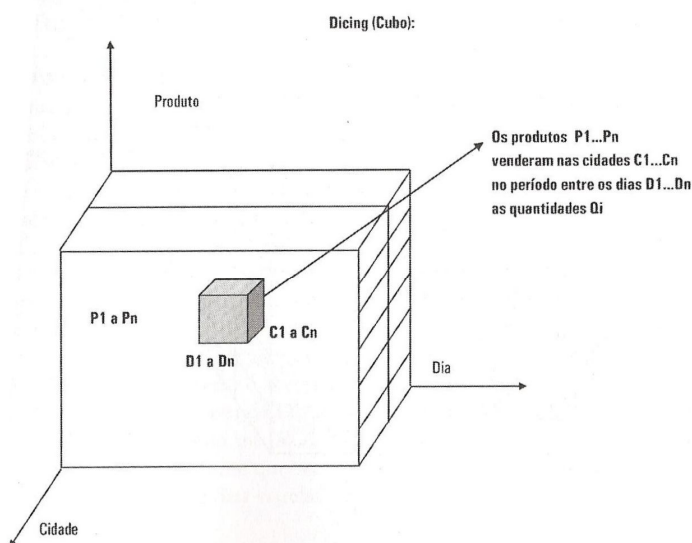


Figura 2 – Representação da visão multidimensional como recurso OLAP em formato de cubo

Fonte: BARBIERI (2011, p. 101)

4.4 A SUITE PENTAHO

Criado desde 2004 pela *Pentaho Corporation*, a suíte *Pentaho Business Intelligence* é um conjunto de soluções integradas de BI, com ferramentas que contemplam todas as fases necessárias para o desenvolvimento de um projeto BI. Este poderoso conjunto de ferramentas é distribuído gratuitamente pela *Pentaho Community Edition*. Todo o projeto é desenvolvido em linguagem de programação *Java*, e que permite integrações com outras soluções através de *plugins*, que se trata de módulos se acoplam a um software com o objetivo de agregar funcionalidades a este software, e até mesmo ferramentas de modelagem de dados. Essa solução está disponível através do site <http://community.pentaho.com>.

Neste projeto, serão usados, dentro do conjunto de ferramentas disponíveis, as seguintes:

Business Analytics Platform: Pacote de software composto pelo *Mondrian*, que é o servidor de aplicações OLAP, e pelo ambiente de administração do ambiente *Pentaho*. Neste pacote estão contempladas soluções para geração de relatórios *ad-hoc* e permite a instalação e integração com outras ferramentas de visualização de dados. Entre os recursos disponíveis nesta plataforma, está possibilidade de uso de painéis de controle, também conhecidos como *Dashboards*, em que permite-se utilizar estes painéis como ambiente de acompanhamento de indicadores, métricas e metas em forma de gráficos ou tabulares.

Pentaho Data Integration - PDI ou Kettle: Trata-se um poderoso ambiente de desenvolvimento responsável por uma das mais importantes fases do projeto que é a Extração, Transformação e Carga dos dados, conhecido como processo ETL (acrônimo para *Extraction, Transformation and Load*), descrito na seção 6.3 durante a fase de desenvolvimento do projeto. A figura 3 apresenta este ambiente do *Kettle* como um exemplo de recursos oferecidos pela ferramenta, que permite de uma forma gráfica construir o fluxo necessário para ler os dados das suas fontes de origem, tratá-los e alimentar a base consolidada que passará a ser o repositório para análise dos dados.

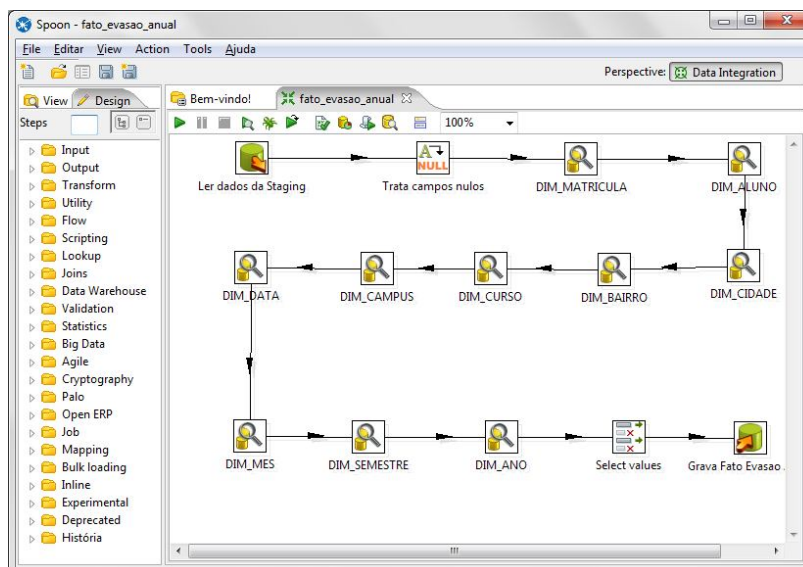


Figura 3 – Demonstração do ambiente Kettle com uso de recursos visuais para construção do ETL

Fonte: Elaborado pelo autor

Schema Workbench: Esta ferramenta é responsável pela construção e publicação dos cubos OLAP, antes de ser disponibilizado ao usuário final. Assim como o ambiente do *Kettle*, a figura 4 representa a forma como os cubos são construídos, também de uma maneira mais amigável ao desenvolvedor, sem que este necessite conhecer as linguagens e códigos em que são traduzidos para serem interpretados pela ferramenta OLAP, através do servidor *Mondrian*. Um dos recursos importantes que estão disponíveis neste ambiente *Workbench* é a possibilidade de selecionar quais dados, daqueles armazenados no DW, serão disponibilizados ao usuário final, permitindo, por exemplo, a criação de diversas visões sobre os dados de acordo com o perfil do usuário final.

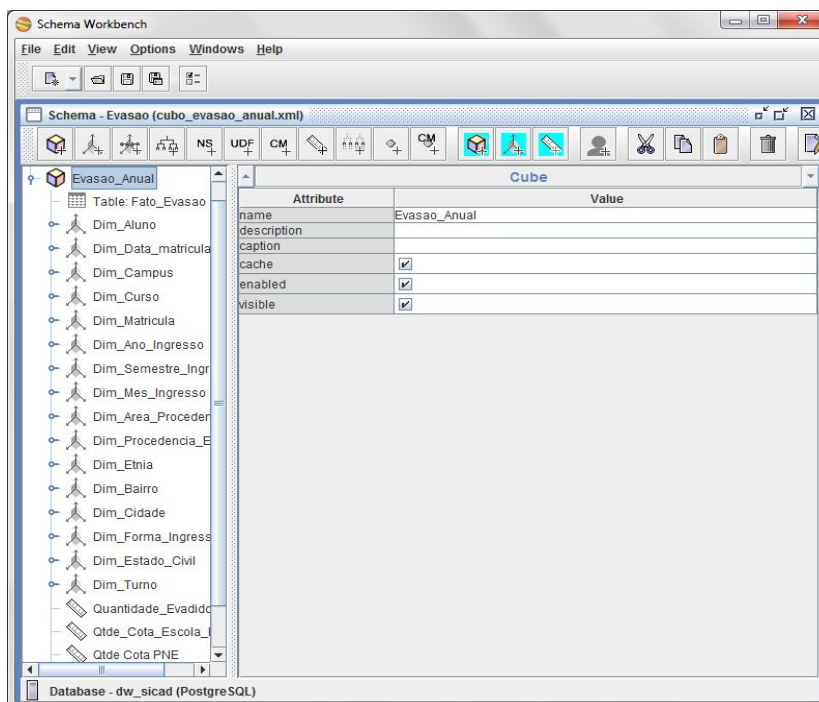


Figura 4 – Demonstração do ambiente de trabalho do Schema Workbench

Fonte: Elaborado pelo autor

Saiku Analytics: Trata-se de uma ferramenta OLAP, com diversos recursos, como a geração de relatórios dinâmicos em tempo real, gráficos em variados formatos, exportação de dados para planilhas eletrônicas e em formato *Portable Document Format* - PDF. Esta ferramenta, embora não seja nativa do pacote de ferramentas *Pentaho*, pode ser facilmente integrada como um *plugin*. A figura 5 é uma demonstração gráfica do recurso de construção de indicadores e cruzamento de informações que podem ser realizados pelo gestor em tempo de execução e através de um ambiente simples disponibilizado através de um navegador web.

The screenshot shows the Pentaho Saiku Analytics web interface. The browser address bar indicates the URL is localhost:8080/pentaho/Home. The interface includes a sidebar with a tree view of dimensions and measures, a central workspace with a toolbar, and a data table. The table displays data for 'Procedencia_Escolar' (School Type) categorized by race and the number of evaded students.

Procedencia_Escolar	MeasuresLevel	Amarillo	Branco	Indígena	Não Informada	Outros	Pardo	Preto
Não Informada	Quantidade_Evadidos		4		494	3	20	8
	Qtde Cota Racial		0		139	1	5	4
Oruondo de Escola Particular	Quantidade_Evadidos		15	1			13	11
	Qtde Cota Racial		0	1			4	5
Oruondo de Escola Pública	Quantidade_Evadidos	1	11	1		3	33	23
	Qtde Cota Racial	0	0	0		0	23	14

Figura 5 – Demonstração da ferramenta *Saiku Analytics* integrada ao ambiente *Pentaho*

Fonte: Elaborado pelo autor

5 METODOLOGIA

Neste trabalho de pesquisa, a metodologia para levantamento e obtenção das informações é a pesquisa aplicada, que utiliza como estratégia a pesquisa-ação, definida por Thiollent (2011, p. 17) como “um tipo de pesquisa social com base empírica que é concebida e realizada em estreita associação com uma ação ou com a resolução de um problema coletivo”. Trata-se de um projeto em que o autor está imerso no ambiente pesquisado, interagindo e participando de comissões de trabalho com os agentes envolvidos no processo de gestão dos dados acadêmicos do IFBA.

Este projeto traz consigo uma intencionalidade que é uma das características apontadas por Thiollent (2011, p. 17) que deve estar presente numa estratégia de pesquisa-ação, conforme afirma que “uma pesquisa pode ser qualificada de pesquisa-ação quando houver realmente uma ação por parte das pessoas ou grupos implicados no problema sob observação [...] uma ação problemática merecendo investigação para ser elaborada e conduzida”.

Corroborando neste sentido, Cauchick Miguel (2012, p. 151), citando entre outros autores Michel Thiollent, traz ainda que entre as principais características que definem uma pesquisa como pesquisa-ação tem-se que “[...] a pesquisa-ação trabalha através de um processo cíclico de quatro passos: planejamento, tomada de ação e avaliação da ação, levando para outro planejamento e assim por diante”.

Nesta perspectiva, num processo de investigação-ação presente desde início desta pesquisa que o trabalho vem se delineando. Nesta pesquisa assume-se o processo de investigação-ação conforme define Tripp (2005, p. 446), como sendo um processo que segue “um ciclo no qual se aprimora a prática pela oscilação sistemática entre agir no campo da prática e investigar a respeito dela”.

Para agir de forma sistemática, conforme definiu Tripp (2005), agindo e investigando no campo da prática, será adotado o ciclo *Plan-Do-Check-Action* (PDCA), originário do campo da Administração, utilizado também por Tripp (2005) como sendo o ciclo da investigação-ação. No campo da administração, segundo Daychoum (2007), o ciclo PDCA é traduzido, respectivamente como sendo: Planejar, Executar, Verificar e Agir. Tripp (2005) traduz como planeja-se, implementa-se, descreve-se e avalia-se.

Para melhor adequação dos termos, neste trabalho será adotado a seguinte nomenclatura para definir as fases que compõe o ciclo: Planejamento, Ação/execução, Monitoramento/verificação, Avaliação/correção.

Entre as características do PCDA estão a de permitir o constante planejamento, verificação e replanejamento das ações com base nas informações que vão sendo identificadas durante o processo de pesquisa. Segundo Tripp (2005, p. 446), com esta metodologia “planeja-se, implementa-se, descreve-se e avalia-se uma mudança para a melhora de sua prática, aprendendo mais, no correr do processo, tanto a respeito da prática quanto da própria investigação”. Além disso, com o avanço das tecnologias de modelagem de dados, a preocupação das organizações no ciclo de vida das informações institucionais, em especial aquelas mantidas em sistemas de informações, alguns trabalhos relacionados a estas tecnologias vêm sendo desenvolvidos com o uso de metodologias similares às usadas no processo de desenvolvimento e gestão do ciclo de vida de softwares, porém voltados a projetos BI. Barbieri (2011), ao tratar dos métodos incrementais e iterativos aplicados ao BI, cita o ciclo PDCA, concebido por Edward Deming, como o grande precursor destes métodos. A concepção do ciclo PDCA também está fundamentada na qualidade e na melhoria contínua dos processos organizacionais o que o torna aderente ao processo de desenvolvimento de projetos nas áreas de TIC e por consequência nos projetos de BI, nos quais os métodos têm convergido para a qualidade e a melhoria contínua.

O ciclo PDCA, também conhecido como ciclo de Deming, “apesar de sua aplicação original no campo da administração da qualidade [...] é frequentemente usado como modelo para o planejamento e implementação de soluções de aprimoramento constante em qualquer área”. (MAXIMIANO, 2011, p. 51)

Este trabalho foi estruturado para ser iniciado com um levantamento dos dados armazenados no SICAD, identificando os dados relevantes. Outro levantamento importante será para o estabelecimento de uma comunicação direta com os agentes envolvidos na gestão escolar, na identificação de informações que possam contribuir para a um melhor entendimento da vida escolar do aluno, neste momento com o foco na evasão nos cursos do ensino superior no *campus* Salvador. O mapeamento dos dados armazenados e este levantamento das informações deram origem às possíveis dimensões, bem como métricas, que são a base para a construção de um modelo multidimensional, de forma a possibilitar uma visão das informações armazenadas sobre diversas perspectivas, de uma maneira estruturada e consolidada. O passo final, após a definição das informações, construção do modelo,

aplicação das técnicas de transformação dos dados, estruturação e consolidação, é a adoção de uma ferramenta gráfica do tipo OLAP para visualização e validação das informações. A solução de *Business Intelligence* será desenvolvida e aplicada usando o conjunto de ferramentas que compõe o projeto *Pentaho*, um projeto de código aberto, gratuito, que vem sendo adotado mundialmente por diversas organizações como solução de *Business Intelligence*.

Para melhor divisão, as fases que compõem o ciclo PDCA foram estruturadas conforme segue, com as 4 fases (Planejamento, Ação/Execução, Monitoramento/Verificação, Avaliação/Correção) e suas etapas que serão representadas numa sequência de 1 a 8, conforme segue:

Planejamento: Esta fase compreende desde a identificação do problema até a definição dos indicadores a serem implementados e está dividida em quatro etapas.

1. **Entendimento e definição do problema:** Esta fase compreende os primeiros passos da pesquisa e também os primeiros contatos para a definição dos elementos da investigação. Deverá ser realizada uma revisão teórica sobre a evasão escolar, análise dos relatórios e métodos de medição publicados pelos órgãos de controle, censo escolar da educação, Sistema de Avaliação da Educação Superior - SINAES. O que estabelece o PNE acerca do assunto evasão escolar.
2. **Análise aprofundada do fenômeno:** Pesquisa exploratória para compreensão das principais dificuldades apontadas pelos sujeitos envolvidos (gestores, técnicos em educação) na identificação dos possíveis elementos causadores da evasão bem como uma compreensão mais aprofundada do assunto dentro do IFBA. Ainda nesta fase é realizada uma pesquisa em torno das estratégias e eventuais dificuldades do IFBA no processo medição dos índices de evasão.
3. **Coleta e análise dos dados:** É o momento necessário para compreensão do processo de registro acadêmico dentro do IFBA, com o levantamento das ferramentas usadas para este registro, do sistema de registros acadêmicos adotado (SICAD), bem como a documentação deste sistema, para entendimento da forma como os registros são armazenados e manipulados pelo sistema e pelos usuários.
4. **Definição de descritores e indicadores:** Nesta fase é usado todo o levantamento de informações das três etapas anteriores que deverão nortear a construção de uma

matriz de necessidade, com as dimensões (descritores) e as métricas (indicadores) elencadas juntamente com os gestores para serem base das fases posteriores.

Durante esta fase de planejamento um conjunto de ações foram realizadas, como descritos a seguir, porém esta fase, que está dividida neste trabalho em quatro etapas ocorreu de forma cíclica em que se avança nas etapas e em determinados momentos é necessário retroceder em etapas anteriores. Será notório observar que este fenômeno cíclico, característico das metodologias ágeis, ocorrerá mesmo passada para etapas das outras fases.

O primeiro passo foi buscar formar um grupo de trabalho dentro do IFBA, multidisciplinar, composto por técnicos em educação e gestores, para assim iniciar o planejamento das atividades. Esta pesquisa conta com o apoio institucional da Diretoria Geral do *campus* Salvador e a Gerência de Registros Acadêmicos, além das Pró-reitorias de Ensino e de Administração. Conta ainda com o apoio técnico da Diretoria de Tecnologia da Informação – DGTI.

O levantamento exploratório das informações que descrevem este cenário foi realizado em momentos: o **primeiro** através de documentos públicos institucionais, disponíveis no portal do IFBA (www.portal.ifba.edu.br), como as normas acadêmicas vigentes e os relatórios de gestão. O período para realização deste primeiro levantamento ocorreu durante os meses de setembro/2014 e dezembro/2014; **segundo** momento, que compreendeu o período a partir de setembro/2014 até março/2015, ocorreu a partir de encontros/reuniões com equipes técnicas e a obtenção de uma cópia da base de dados onde estão armazenados os dados acadêmicos do SICAD, bem como outras informações necessárias para a compreensão dos dados armazenados, devido à ausência de documentos de apoio. É importante salientar que os encontros com as equipes técnicas não se esgotaram durante este o período exploratório para descrição deste cenário, já que houve necessidade de interlocução durante todo o projeto; o **terceiro** momento foi através da aplicação de entrevistas e reuniões com gestores, as reuniões ocorreram no período setembro/2014 a março/2015 e as entrevistas ocorreram nos meses de março/2015 e abril/2015, devido disponibilidade dos gestores. Para estas reuniões foi utilizado o registro de atas, para as entrevistas estas foram orientadas a partir de um questionário composto de oito questões relacionadas com a evasão escolar.

As etapas 3 e 4 desta fase serão detalhadas no capítulo de desenvolvimento, assim como as demais etapas, por se tratarem do processo de construção da solução de BI e que

seguirá durante cada etapa processos metodológicos voltados à construção de projetos que abordam esta tecnologia.

Ação/Execução: Nesta fase, que está dividida em duas etapas, compreende o desenvolvimento do modelo e implementação da solução na ferramenta OLAP a partir das dimensões, métricas e indicadores previamente definidos na fase de planejamento.

5. **Desenvolvimento dos Modelos Multidimensionais:** A etapa de construção dos modelos multidimensionais utiliza as técnicas e metodologias de modelagem propostas por Kimball e Barbieri, por serem metodologias sólidas para a construção de projetos de BI. Esta fase é responsável pela geração de um dos principais produtos dentro do projeto de BI que é o *Data Warehouse*, ou os *Data Marts*, já que neste projeto foi adotado o conceito de construção de repositórios por assuntos. Ainda nesta fase é realizado todo o processo de Extração, Transformação e Carga, processo também conhecido como ETL, acrônimo para a expressão *Extract, Transformation and Load*, em inglês. Este é o momento para uso de parte das ferramentas do pacote *Pentaho* citadas na seção 4.3. Antes de avançar para a etapa posterior, por orientação da própria metodologia de desenvolvimento do projeto BI, é feita a validação de dados junto ao usuário, para que seja possível a publicação dos cubos OLAP.
6. **Implantação e disponibilização dos cubos OLAP:** Esta etapa são construídos os Cubos OLAP e disponibilizados para os usuários através da ferramenta de visualização OLAP. Neste momento os usuários deverão ser capacitados a usarem esta ferramenta.

Monitoramento/Verificação:

7. **Acompanhamento e avaliação dos resultados:** Nesta fase serão analisados os resultados implementados nas fases anteriores para que sejam comparados com os resultados esperados/planejados e a necessidade de um replanejamento das ações. Também nesta fase será validada a ferramenta OLAP disponibilizada com os cubos desenvolvidos na fase anterior. Dada a importância desta fase, o Capítulo 7 será dedicado a tratar do processo de análise e validação dos dados, assim como a efetividade do uso de soluções OLAP. Os instrumentos usados nesta fase são um questionário semi-estruturado conforme APENDICE D, a própria ferramenta OLAP e os demais recursos providos pelo ambiente tecnológico do *Mondrian*. O

questionário visa obter dos gestores que colaboraram desde a fase de planejamento a sua percepção em relação ao ambiente analítico proposto. Neste momento da pesquisa a análise dos dados ocorrerá com a obtenção, pelo pesquisador, de indicadores a partir das visões publicadas, comparando-os com outras pesquisas realizadas por especialistas em evasão no ensino superior.

Avaliação/Correção:

8. **Medidas corretivas ou Institucionalização:** É a fase para a tomada de medidas corretivas, tanto do ponto de vista das informações encontradas, como da solução adotada. Para os casos não houver necessidade de readequação da solução apresentada, esta também é uma fase para o processo de institucionalização da solução. A institucionalização de um projeto desta natureza dependerá de um período de uso e amadurecimento, de modo que esta etapa, no que tange o processo de institucionalização da solução, ocorrerá após a entrega deste trabalho, enquanto trabalho de conclusão do curso de mestrado.

A figura 6 representa o ciclo PDCA, adaptado para este projeto, e suas divisões em fases.

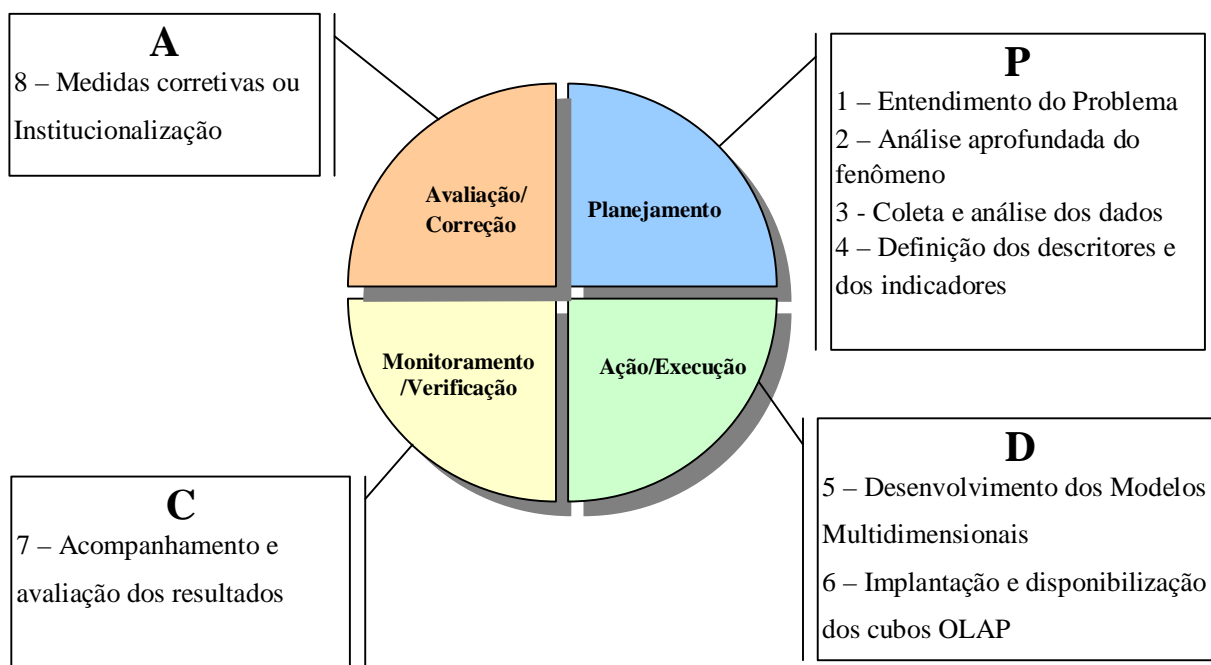


Figura 6 – Representação do ciclo PDCA e suas fases no projeto

Fonte: Elaborado pelo autor

É importante destacar que este processo que usa o ciclo PDCA não indica que para avançar nas etapas seja necessário que toda a etapa que antecede uma próxima etapa seja totalmente concluída. Muitas ações foram realizadas em paralelo, tanto como uma forma de garantir a não interrupção da pesquisa como também uma forma de reduzir o impacto nos ajustes. Um exemplo mais claro em relação a isso ocorre durante a etapa 4, de definição dos descritores, que permite tanto avançar na construção do modelo na etapa 5, como muitas vezes remete a fase 3 de coleta e análise de dados novamente. Além disso, as atividades desenvolvidas nesta fase 4 também exigem que sejam validadas e corrigidas, que são ações que devem ocorrer nas fase seguintes (Monitoramento/Verificação, Avaliação/Correção), de modo que torna o processo além de cíclico, incremental. Essas ações reforçam a característica de investigar e agir no campo da prática, processo de pesquisa-ação adotado como premissa metodológica deste trabalho, como afirmam Thiollent (2011) e Tripp (2005), além de reafirmar o processo de melhoria contínua, como citado anteriormente por Barbieri (2011) e Maximiano (2011).

No capítulo a seguir serão detalhados os processos de desenvolvimento das etapas 3 - Coleta e Análise dos dados, 4 - Definição de indicadores, 5 - Desenvolvimento dos Modelos Multidimensionais e 6 - Implantação e disponibilização dos cubos OLAP.

6 DESENVOLVIMENTO

Este capítulo compreende o detalhamento de todo o processo de levantamento de informações para a construção do modelo de dados, iniciado ainda na fase de planejamento, conforme proposta metodológica, e as fases que se sucedem. Serão as etapas 3 e 4 da fase de planejamento e as etapas da fase de Ação/Execução. As demais etapas, como dito anteriormente, estão presentes durante todo o processo, com exceção do processo de institucionalização que, como dito, deverá extrapolar o período de entrega deste trabalho enquanto projeto para conclusão do mestrado.

6.1 COLETA E ANÁLISE DOS DADOS

A coleta e análise dos dados representam uma importante etapa tanto dentro do ciclo PCDA como para um projeto de construção de um modelo de dados multidimensional, que é um dos objetivos deste projeto. Para esta etapa, em que seu resultado é insumo para a etapa 4, de definição dos descritores e dos indicadores, também foi necessário o aprofundamento dos aspectos relacionados à evasão escolar, ocorrida nas etapas que antecedem esta coleta de dados. Essas etapas anteriores se deram com uma revisão teórica acerca do assunto e pesquisas exploratórias a partir das fontes de informações dos órgãos de controle e do próprio IFBA. Por se tratar de uma questão delicada e ainda com certas indefinições e ausência de informações suficientes, nas etapas anteriores foram realizadas entrevistas com alguns dos gestores envolvidos no processo que trata da evasão escolar e dos registros acadêmicos dentro do IFBA. Estes levantamentos preliminares, realizados nas etapas 1 e 2 da fase de planejamento possibilitaram um melhor entendimento sobre os dados disponíveis.

O processo de coleta de dados se deu a partir de encontros realizados com a equipe da DGTI, responsável pelo armazenamento dos dados e pelo desenvolvimento e manutenção dos sistemas usados pelo instituto. Este trabalho de coleta de dados ocorreu de forma incremental, já que a descoberta de determinadas informações suscitaram a busca por outras informações relacionadas. Todas as informações do SICAD estão armazenadas numa única base de dados e esse foi motivo que fez com esta fosse a principal fonte para coleta e análise dos dados.

Nesta etapa nos deparamos com a ausência de um dos principais artefatos de um projeto de desenvolvimento de software e também um projeto de BI, que é o Dicionário de

Dados², um artefato que armazena os metadados dos dados mantidos pelo sistema. Esta ausência fez com que parte desta etapa de análise fosse dedicada à construção deste artefato. O processo de construção deste dicionário de dados ocorreu em parceria com Analistas de Sistemas da DGTI por serem responsáveis pela manutenção do SICAD e por este motivo conhecem a sua base de dados. Para a construção deste documento, foi montado um modelo de planilha, que após proposto pelo pesquisador e aceito pela equipe da DGTI passou a ser preenchido de forma colaborativa, através de uma solução disponibilizada pela *Google*, conhecida como *Google Docs* que além de compartilhar o acesso às pessoas autorizadas faz todo o controle das modificações ocorridas no documento. A construção deste Dicionário de Dados está sendo feita de forma incremental e, neste primeiro momento, foi acordado que apenas as informações relacionadas com esta pesquisa seriam documentadas.

Esta etapa, além de ser um insumo para as etapas posteriores, está resultando ao IFBA a construção do produto Dicionário de Dados do SICAD, elaborado e alimentado conforme figura a seguir. O formato sugerido buscou descrever um conjunto de informações básicas, porém necessárias para a melhor compreensão acerca dos dados do sistema em questão, como por exemplo a forma como os dados se relacionam, seus significados e de que maneira estão armazenados.

Dicionário de Dados SICAD

Tabela: <Nome da tabela como consta no banco de dados> <descrever o objeto da tabela>						
	Campo	Descrição	Tipo	Tamanho	Nulo	Valor Padrão
<Indicar se é um campo Chave e o tipo de chave (PK, FK, etc)>	<nome do campo no banco de dados>	<Descrição do objetivo deste dado no banco e outras informações importantes para compreensão deste dado>	<tipo do dado>	<Tamanho em Bytes>	<Indicar nulidade do campo (0 ou 1)>	<Indicar valor padrão, caso exista>

Figura 7 – Modelo de Dicionário de Dados adotado para documentar dados do SICAD

Fonte: Elaborado pelo autor

² Dicionário de Dados é um artefato de um software que tem por objetivo descrever os elementos de dados pertinentes ao software que permita que tanto o usuário (cliente) como a equipe técnica de desenvolvimento tenham uma compreensão do significado e da forma como estes dados são armazenados. (PRESSMAN, 2006, p. 307)

A construção do dicionário de dados do SICAD tem permitido identificar uma série de anomalias no modelo de dados relacional, anomalias estas que colocam em risco a integridade e a qualidade dos dados. Entre os problemas identificados, pode-se destacar:

- Ausência de regras de integridade para garantir que uma informação que se relacione com outra que de fato exista;

- Inconsistência de dados como endereço, CEP, cidade;

- Redundância de dados, tanto dentro de uma mesma tabela, como a duplicação da mesma informação em outras tabelas;

Estas questões levantadas nesta fase exigirão um maior esforço durante o processo de Extração, Transformação e Carga dos dados, durante a etapa da construção do modelo multidimensional.

Também durante esta etapa de análise foi verificada a ausência de informações mais detalhadas relacionadas a questões socioeconômicas dos alunos e que as informações sobre a participação dos alunos nas reservas de vagas por cotas somente passaram a serem registradas a partir do ano 2010. A busca por informações desta natureza se deu a partir das informações levantadas durante as entrevistas, em que os gestores declararam que consideram haver alguma relação entre evasão escolar e questões socioeconômicas, sobretudo aquelas relacionadas com renda per capita.

A ausência e a dificuldade de encontrar estas informações na base de dados do SICAD fizeram com que outras fontes de dados fossem buscadas a fim de permitir a coleta e análise destes dados, a exemplo da base de dados do vestibular, porém nesta fonte também não foi identificado dados sobre renda per capita, apenas dados relacionados às cotas, que também passaram a serem coletados no mesmo período em que passou a ser alimentado no SICAD. Como os dados do SICAD são alimentados após o vestibular, para aqueles aprovados, e que se matriculam, optou-se por adotar apenas a base do SICAD, que contém uma informação mais recente. Ainda no SICAD e sobre dados socioeconômicos foi identificado um campo chamado RFP, que se trata da abreviação para Renda Familiar Per capita.

Durante as entrevistas, conforme relatado no Capítulo 3, algumas perguntas suscitaram outras. Entre elas, sugeriram questionamentos sobre os dados mantidos no SICAD

que visaram contribuir para o entendimento sobre os dados, sobretudo pela falta do Dicionário de Dados. Quando questionado sobre o processo de alimentação dos dados e a periodicidade de atualização, o entrevistado D respondeu que “todo dado é alimentando a partir da ficha de matrícula preenchida pelo aluno no ato da matrícula[...]”. Sobre a periodicidade de atualização dos dados, este afirmou que “apenas dos registros acadêmicos, como nota e frequência, são atualizados diariamente, os dados cadastrais só são atualizados por solicitação dos próprios alunos, não há ciclo de atualização dos dados cadastrais”. Entre as motivações para esta pergunta estavam o endereço encontrado no cadastro de alguns alunos, em que a cidade e o estado não eram Salvador e Bahia, respectivamente, e obter a informação sobre o momento em que se passou a alimentar os dados socioeconômicos encontrados. Para estas questões, o entrevistado D respondeu que no que se refere ao endereço, “este fato ocorre porque o endereço cadastrado é aquele declarado no ato da matrícula, onde o aluno que ainda não tem novo endereço definido declara o endereço de origem”. Esta resposta corrobora com aquela que trata da periodicidade de atualização dos dados cadastrais. Para o questionamento sobre o início da alimentação dos dados socioeconômicos, respondeu que os “não têm a precisão de quando iniciou-se, mas são dados bem recentes”. Isso pôde ser comprovado ao analisar a base, onde a grande maioria dos alunos não tem estas informações alimentadas.

Todos estes dados coletados, analisados, juntamente com as informações adicionais obtidas a partir de entrevistas, são de fundamental importância para as fases posteriores, como poderá ser visto a seguir, a partir da Definição dos Descritores e dos Indicadores. Tudo isso é para a base para a construção matriz de necessidades que é um dos principais insumos de entrada para a modelagem neste projeto.

6.2 DEFINIÇÃO DOS DESCRITORES E DOS INDICADORES

Esta etapa tem como principal objetivo subsidiar a etapa seguinte durante o processo de construção do *Data Warehouse*, que compreende toda a modelagem dimensional, extração, transformação e carga dos dados. Esta etapa tem como elementos de entrada todo o processo de coleta e análise de dados da etapa anterior e adotou a utilização da matriz de necessidades como instrumento para definição dos descritores/dimensões e das métricas/fatos que auxiliarão a etapa seguinte.

A matriz de necessidades foi assim nomeada por Lopes (2013) que a desenvolveu e utilizou como um instrumento de sua pesquisa num projeto de BI, com o objetivo de auxiliar

o processo de compreensão dos gestores sobre os dados disponíveis e como uma ferramenta para atualização do plano de dados sempre que necessário durante o ciclo de vida do projeto de BI. No processo de desenvolvimento de um projeto de *Data Warehouse* proposto por Kimball (2002) a utilização de matrizes é recorrente e considerada um dos principais artefatos para o projeto. Ocorre que, o modelo adotado por Kimball (2002) é projetado para projetos corporativos com uma abrangência maior, que abordem assuntos diversos, mas com a mesma finalidade. Uma diferença básica, por exemplo, é o fato de usar as linhas para descrever as métricas/fatos e as colunas para descrever a dimensão. Na matriz usada neste projeto, adotando o modelo proposto por Lopes (2013), usam-se as colunas para representar as métricas/fatos e as linhas para representar os descritores/dimensões, já que nestas linhas são apresentadas não só a dimensão, como atributos considerados importantes para a visão do gestor. O uso desta matriz tem um caráter fundamental no processo de definição do plano de informações que irá compor o modelo multidimensional, já que esta atividade é realizada em conjunto com a equipe de gestores e técnicos em educação que, em princípio, não possuem conhecimento técnico para analisar modelos de dados, além de permitir uma objetividade maior, focada diretamente nos descritores e métricas.

Trata-se de uma matriz bidimensional em que as colunas representam as métricas e as linhas representam os descritores. Cada cruzamento bidimensional, linha *versus* coluna, indicará uma perspectiva sobre a qual poderá ser visualizada ou realizada buscas a cerca destas informações representadas nestas linhas e colunas.

O processo de construção da matriz ocorreu a partir de encontros presenciais e também foi construído de forma incremental. Como estratégia o pesquisador buscou um preenchimento prévio de descritores e métricas elucidadas durante o processo de coleta e análise dos dados e a partir das reuniões os gestores sugeriram a inclusão ou retirada de linhas e colunas. Um passo importante durante estas reuniões é que a sugestão de inclusão informações (descritores e/ou métricas) indisponíveis durante a etapa de coleta e análise exigiram a retomada a fase anterior na busca pela existência destes dados, a exemplo das informações relacionadas a renda do aluno. A definição ocorreu em dois encontros exclusivos para tratar da matriz, registrados através de atas de reunião, além de aproveitar um momento após as entrevistas para discutir e validar a matriz. Este processo iniciou em setembro de 2014, quando do início do processo de coleta e análise dos dados, até o mês de abril de 2015, durante as entrevistas.

Para este primeiro ciclo de construção do modelo, ficou estabelecido que apenas estas métricas descritas na matriz serão adotadas. Todos os cálculos adotados para compor as métricas levam em consideração o cálculo da evasão anual usada pelo IFBA no relatório de gestão, já que este é um método aceito pelos entrevistados como mais adequado para medir a evasão nos cursos do ensino superior.

É importante destacar o papel importante da participação dos entrevistados para a construção da matriz. As respostas dos entrevistados elucidaram para algumas das dimensões que deveriam ser consideradas para a construção do modelo multidimensional, sobretudo a partir da pergunta sobre quais os fatores que gestores consideram serem determinantes para elevação dos índices de evasão escolar no ensino superior do IFBA. Entre as respostas obtidas, como pode ser verificado no Capítulo 3, foi possível destacar cinco pontos: a questão do transporte público; condições financeiras; cursos das ciências exatas; a origem escolar do aluno, se de escola pública ou privada; e o turno em que são ofertados. Como a adoção da dimensão e da métrica dependeu de todo o processo de coleta e análise realizado na fase anterior, não foi possível adotar todos os pontos como dimensão e em alguns casos houve adaptações.

Para a questão relacionada ao transporte público, foi decidido adotar os descritores ligados ao endereço do aluno, estado, cidade e bairro. Embora esta pesquisa se limite a analisar apenas o *campus* Salvador, tem como perspectivas futuras a expansão para outros campi, além do fato de sido encontrado na base alunos cursando em Salvador e com endereços declarados em outras cidades, até mesmo em outros estados.

Sobre os cursos de ciências exatas serem considerados um fato determinante para a evasão escolar por terem disciplinas com altos índices de reprovação, foi acrescentado o descritor curso e disciplina, como forma uma forma de permitir analisar a quantidade de alunos evadidos por curso, com vistas à identificação dos cursos com maior número de evadidos, além de permitir analisar o desempenho de frequência e notas dos alunos evadidos por curso e disciplina. Ainda para contribuir na análise, foram acrescentadas outras informações que compõem o curso, como número máximo de períodos, currículo (informação pertencente à matrícula), entre outras.

Quanto aos fatores como origem escolar do aluno e o turno, estes foram acrescentados, conforme pode ser verificado na matriz, apenas um com detalhe para a origem escolar, que passou a se chamar Procedência Escolar, como é tratado no SICAD.

O fator “renda do aluno” originou uma série de outros descritores que fazem relação direta ou indireta com os fatores de renda, até mesmo algumas métricas. Desta forma, com base nas informações disponíveis nos dados coletados foram definidos os seguintes descritores: Renda Familiar Per capita (única renda disponível no cadastro do aluno) e se o aluno declarou possuir atividade remunerada. As métricas adotadas, que fazem uma relação com este descritor (mas que também pode ser usado para analisar a partir dos demais), foram: Quantidade de alunos evadidos que entraram pela cota de Renda Familiar Per capita e Quantidade de alunos evadidos que entraram pela cota de aluno oriundo de escola pública.

Os demais descritores e métricas foram incluídos a partir das informações disponíveis e das solicitações dos gestores de forma que pudessem contribuir para análises acerca de outros fatores socioeconômicos e cadastrais como o caso da Idade, Estado Civil, Etnia, Gênero (Masculino e Feminino), Período de Ingresso, Período em que estava cursando no momento da evasão, entre outras. A figura 8 apresenta o modelo de matriz adotado. No APÊNDICE C pode ser verificada a referida matriz de forma completa e mais detalhada.

Descritores/Dimensões	Métricas/Fatos														Descritores/Dimensões	
	Cide Evasão Anual	Cide Cotista PNE	Cide Cotista racial	Cide Cotista escola pública	Cide Cotista Renda	Índice de Evasão Anual	Cide Matrículas Anual	Índice Cotista PNE	Índice Cotista racial	Índice Cotista escola pública	Índice Cotista Renda	Índice de Crescimento da Evasão Anual	Índice de Crescimento de Matrícula anual	Média de Nota de Evadidos		Média de Frequência de Evadidos
Campus	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	Campus
Nome	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	Nome
Código da Unidade	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	Código da Unidade
Curso	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	Curso
Nome	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	Nome
Número máximo de períodos	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	Número máximo de períodos
Modalidade	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	Modalidade
Disciplina															X	Disciplina
Nome															X	Nome
Matrícula	X	X	X	X	X											Matrícula
Código da Matrícula	X	X	X	X	X											Código da Matrícula
Currículo	X	X	X	X	X											Currículo
Período Acadêmico	X	X	X	X	X											Período Acadêmico
Forma de Saída	X	X	X	X	X											Forma de Saída
Período de Saída	X	X	X	X	X											Período de Saída
Forma de Ingresso	X	X	X	X	X											Forma de Ingresso
Cidade	X	X	X	X	X											Cidade
Nome	X	X	X	X	X											Nome
Estado	X	X	X	X	X											Estado
Sigla Estado	X	X	X	X	X											Sigla Estado

Figura 8 – Modelo de Matriz de Necessidade adaptado para este projeto

Fonte: Elaborado pelo autor

Para a construção desta matriz, é importante destacar que por muitas vezes se fez necessária a retomada à fase anterior para eventuais análises de dados ou a necessidade de contato com gestores, conforme descrito, através da realização de encontros e/ou entrevistas. Este movimento é característico do ciclo PDCA, como um processo cíclico, de melhoria contínua. Além disso, nota-se ai, mais uma vez, a constante presença da investigação-ação, adotada neste projeto como premissa metodológica.

O material produzido até aqui, dá elementos para prosseguir nas etapas seguintes, que terão um caráter mais técnico, envolvendo a modelagem, processo de ETL e publicação dos cubos na ferramenta OLAP. Conforme visto no Capítulo 4, para a modelagem multidimensional e posterior construção do DW (etapa imediatamente a seguir), são necessários quatro passos: 1 – Definição do negócio; 2 – Identificar o menor nível de detalhe que se pretende analisar os dados; 3 – Escolher as dimensões; e 4 - Define-se os fatos. A matriz de necessidades é importante artefato para permitir avançar na próxima etapa e desenvolver o modelo multidimensional.

6.3 DESENVOLVIMENTO DOS MODELOS MULTIDIMENSIONAIS

Esta etapa representa um importante passo deste projeto, que resultará nos modelos multidimensionais, que representam os principais produtos desta pesquisa. A modelagem multidimensional é também um processo importante na construção de projetos de BI que trabalham com *Data Warehouse*. Nesta etapa serão descritos os quatro passos para a construção de um modelo multidimensional, assim como o processo de Extração, Transformação e Carga, conhecido como processo ETL originário do inglês *Extraction, Transformation and Load*.

O **primeiro** dos quatro passos descritos por Kimball (2002) é a definição do negócio. Este passo, conforme descrito na contextualização e em todo o delinear do projeto, foi definido nas primeiras reuniões com a equipe da PROEN, em que decidiu-se por pesquisar a evasão escolar no ensino superior, no *campus* Salvador.

O **segundo** passo trata da identificação do menor nível de detalhe, também chamado de nível de granularidade ou grão. Segundo Kimball (2002, p. 37) “o grão exprime o nível de detalhes associados às medidas da tabela de fatos”. Este é um passo fundamental no processo

de modelagem e para que se possa avançar nos dois próximos passos. Este passo toma como base tanto as linhas (descritores) da matriz de necessidade, como as colunas (métricas). A partir da análise destas informações, após verificação das fontes de dados, que nesta pesquisa será a base de dados do SICAD, define-se a forma como os dados serão agrupados para serem armazenados no DW e como as métricas serão calculadas no momento da carga dos dados. É importante destacar que a matriz de necessidade servirá para definir a quantidade de tabelas de fatos que serão necessárias. O nível de granularidade definirá estas tabelas, assim como os cubos que serão necessários para permitir a visualização destas informações através do ambiente OLAP. Um exemplo que pode ser observado nesta pesquisa é analisar métricas a quantidade de evadidos e a frequência média dos alunos evadidos, conforme a matriz de necessidade. Estas duas informações possuem granularidades distintas, já que cada aluno evadido possui um conjunto de registros relacionados à sua frequência. Mas, para definir um nível de granularidade é necessário saber ainda sobre quais perspectivas se pretende analisar uma determinada métrica. De acordo com a matriz de necessidade, para cada aluno evadido se pretende chegar aos dados cadastrais individuais de cada aluno, como faixa etária, renda, logo, para a quantidade de evadidos, o menor nível será o registro de matrícula deste aluno. Já para a média de frequência, a informação mais detalhada sobre a qual se pretende analisar a métrica, representada na matriz de necessidades, é a frequência média dos alunos evadidos por disciplina, logo o menor grão para a frequência média, será a soma dos dias em que o aluno esteve presente na aula dividida pela quantidade de aulas registradas na disciplina, num determinado período e num determinado curso, conforme definido na construção da matriz.

O **terceiro** passo é a definição das dimensões. A partir dos descritores que compõem as linhas da matriz de necessidades é que se elencam estas dimensões. Num ambiente corporativo este é um ponto que merece bastante atenção, pois será necessário um cuidado em separar modelos dimensionais físicos, a forma em que os dados serão armazenados no DW, dos modelos dimensionais aqui chamados de lógicos, que representam a forma como dados poderão manipulados pela ferramenta de visualização. Neste passo exigiu o uso de estratégias para adequar o modelo multidimensional às características e limitações das ferramentas adotadas, bem como o atendimento das necessidades levantadas junto aos usuários para permitir o cruzamento dos dados. As dimensões aqui apresentadas serão aquelas que estarão disponíveis para os usuários finais e não a forma como os dados estarão armazenados fisicamente no DW, já que o foco deste projeto é a construção do modelo multidimensional como proposta para apoiar os gestores no controle da evasão escolar.

Uma importante decisão para este projeto foi a criação de dimensões formadas por atributos de uma outra dimensão, o que Kimball (2002, p. 235) chama de minidimensão, como uma estratégia de implementação para permitir o cruzamento e a combinação de dados de forma sumarizada, o que não seria possível numa ferramenta OLAP trabalhando com atributos e nem como hierarquias. Hierarquias são agrupamentos realizados dentro uma mesma dimensão, porém, embora as ferramentas OLAP permitam a sumarização dos atributos que de uma mesma dimensão, não é permitido combinar dados de mais de uma hierarquia de uma mesma dimensão. A estratégia para estas questões é a adoção de minidimensões. Este instrumento foi bastante usado neste projeto, já que os gestores esperam poder realizar estas combinações com dados como Estado Civil, Gênero, Renda Familiar, Idade, entre outros, que neste exemplo, seriam todos atributos da dimensão aluno, mas que para permitir a combinação, atendendo assim os requisitos dos gestores e as limitações da tecnologia, optou-se por transformá-los em dimensões, ou minidimensões, como denominou Kimball (2002).

Desta forma, seguem representadas na tabela 6 as dimensões e seus atributos resultantes da matriz de necessidades:

Tabela 6: Lista de dimensões extraídas da matriz de necessidades

Dimensão	Atributo	Observação
Aluno	Nome	
	Indicador de Atividade Remunerada	Sim ou Não
	Data de Nascimento	
Matrícula	Código de Matrícula	
	Currículo vinculado	Valor numérico que representa o currículo ao qual a matrícula está vinculada
	Período Acadêmico	Valor numérico que indica o período acadêmico que pertence a matrícula
	Forma de Ingresso	Indica a forma de acesso da primeira matrícula, como ENEM, SISU, nº da chamada no caso de acesso pelo vestibular
Campus	Código da Unidade	
	Descrição	
Curso	Descrição	
	Modalidade	
	Número Máximo de Períodos	
Disciplina	Nome	
Cidade	Nome	
	Sigla do Estado	
	Nome do Estado	
Bairro	Nome	

Etnia	Descrição	
Procedência Escolar	Descrição	Indica se o aluno declarou ter vindo de escola pública ou privada
Área de Procedência	Descrição	Indica se o aluno declarou ter vindo da zona rural ou urbana
Estado Civil	Descrição	
Idade	Idade	
	Faixa etária	Classifica o aluno dentro das faixas etárias definidas pelo IBGE na pirâmide etária, de acordo com sua idade no ato da matrícula.
Gênero	Descrição	Masculino ou Feminino
Renda Familiar Per Capita	Faixa de Renda	Faixa de renda conforme definido no SICAD
Data Matrícula	Data	
	Dia da Semana	segunda-feira, terça-feira, etc.
	Mês	
	Trimestre	1º, 2º, 3º ou 4º trimestre do ano
	Semestre	1º ou 2º semestre do ano
Ano	Ano	Esta dimensão será compartilhada pelas diversas tabelas de fato, em cada uma com sua função específica. Poderá ser o ano de ingresso, o ano da evasão ou o ano da matrícula.
	Semestre	1º ou 2º semestre do ano

Fonte: Elaborada pelo autor

O **quarto** passo entre aqueles que Kimball (2002) propõe como passos necessários para a modelagem é definição dos fatos. Este passo se utilizou das métricas elencadas na matriz de necessidades, agrupando-as de acordo com nível de granularidade, de maneira que como resultado no processo de modelagem, cada grupo de fatos numéricos pertencentes ao mesmo grão comporão uma tabela de fatos. O que este passo tem como meta é determinar o que modelo pretende medir.

Desta forma, de acordo com as métricas da matriz e após análise da fonte de dados do SICAD, serão necessárias duas tabelas de fatos, agrupadas como apresentada na tabela 7.

Tabela 7: Lista de fatos extraídos a partir da matriz de necessidades

Fato	Métrica	Granularidade
Evasão Anual	Quantidade de alunos evadidos	1 linha para cada aluno matriculado que esteja entre aqueles considerados evadidos
	Quantidade de alunos que entraram por cota racial	
	Quantidade de alunos que entraram pela cota de escola pública	
	Quantidade de alunos que entraram pela cota de Renda Per Capita	
	Quantidade de alunos que entraram pela cota de Portadores de Necessidades Especiais	
Relação entre Evadidos	Quantidade de Matriculados	1 linha para cada curso, num <i>campus</i> por

e Matriculados Anuais	Índice de evadidos	ano
	Índice de alunos que entraram por cota racial	
	Índice de alunos que entraram pela cota de escola pública	
	Índice de alunos que entraram pela cota de Renda Per Capita	
	Índice de alunos que entraram pela cota de Portadores de Necessidades Especiais	
Nota e Frequência	Nota Média	1 linha para cada disciplina, num curso, num <i>campus</i> por ano
	Percentual média de frequência	

Fonte: Elaborada pelo autor

Alcançados os quatros passos descritos até aqui, é possível avançar para o **projeto lógico do banco de dados** que dará suporte ao projeto físico. Para a construção deste projeto lógico, os seguintes passos foram realizados:

- Análise do modelo de dados da base transacional (SICAD), com a identificação dos dados a serem importados para a *Staging Area*;

- Definição da plataforma tecnológica para armazenamento físico da *Staging Area* e do *Data Warehouse*. Neste ponto a decisão foi adotar a plataforma disponível ao IFBA, sem gerar novos custos. Como citado no Capítulo 4, a plataforma tecnológica será o projeto *Pentaho* e terá como base de dados para armazenamento físico o Sistema Gerenciador de Banco de Dados – SGBD PostgreSQL³, um sistema *open source* e de uso comum pelas soluções tecnológicas implementadas pela equipe de TI do IFBA.

- Construção do modelo *Staging Area*, como sendo a área intermediária de preparação dos dados, sem com isso comprometer os dados em produção do SICAD. Para simplificar este processo optou-se por modelar de acordo com o modelo de dados do SICAD, filtrando apenas as tabelas necessárias ao projeto.

- Construção do modelo de dados multidimensional, que será o modelo lógico do *Data Warehouse* onde os dados consolidados ficarão fisicamente armazenados (dimensões e os fatos).

- Validação dos modelos *Staging Area* e o Multidimensional.

³ PostgreSQL é um Sistema Gerenciador de Banco de Dados *open source*, desenvolvimento na *University of California at Berkeley*, desde 1985, robusto, multiplataforma, amplamente utilizado mundialmente por grandes corporações, com suporte para armazenamento de grandes volumes de dados (<http://www.postgresql.org>).

Como resultado do projeto lógico, tem-se o modelo *Staging Area*, similar ao modelo da fonte de dados do SICAD, apenas com as tabelas necessárias para alimentar o modelo multidimensional, e os modelos multidimensionais conforme figuras 9, 10 e 11. A modelagem multidimensional deve ser realizada com vistas a suportar o armazenamento dos dados e informações contidos na matriz de necessidades, ou seja, as dimensões/descriptores e as suas métricas/fatos. As figuras 9, 10 e 11 foram separadas em 3 para facilitar a visualização, mas poderiam estar numa grande imagem contendo as dimensões e as 3 tabelas de fatos. É importante destacar que as dimensões num projeto de BI corporativo devem ser compartilhadas, como é o caso das dimensões *campus*, *ano* e *curso* nas figuras 10 e 11, que são as mesmas da figura 9. Outro aspecto que merece destaque são as métricas, em que nem todas precisam estar fisicamente armazenadas no modelo multidimensional, como dito, é preciso que o modelo seja capaz de produzir esta informação. Nesta modelagem, optou-se por construir alguns indicadores a partir da montagem dos cubos OLAP, armazenando apenas as informações necessárias para produzir tais indicadores, a exemplo dos índices, que são produzidos a partir das quantidades de evadidos e matriculados e estes sim estão armazenados fisicamente, como podem ser vistos nas tabelas de fatos presentes nas figuras 9, 10 e 11.

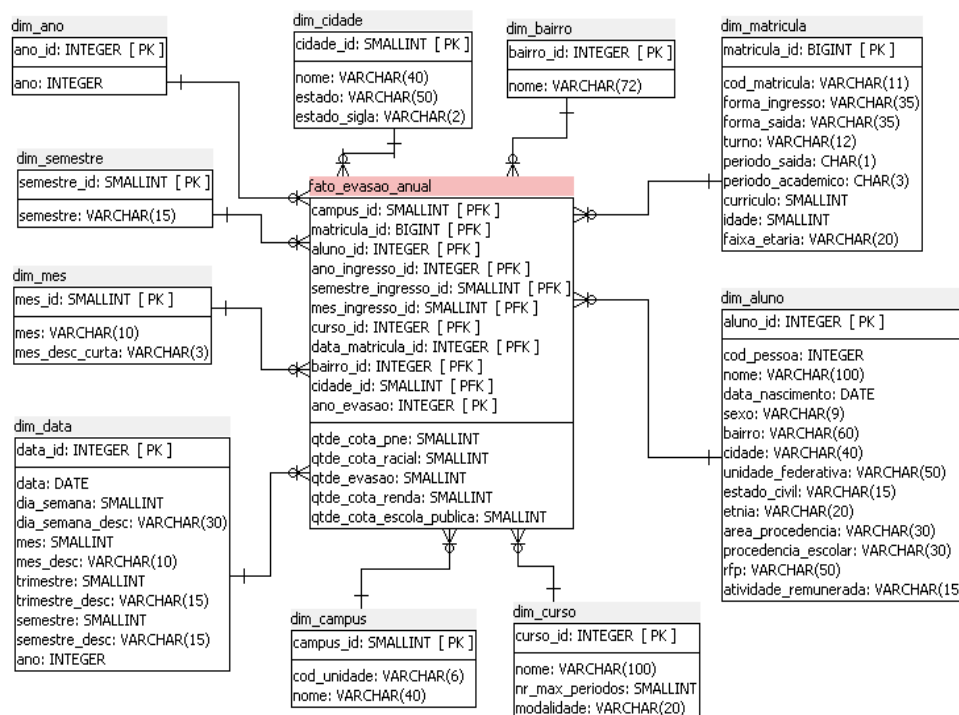


Figura 9 – Modelo Multidimensional para Fato Evasão Anual

Fonte: Elaborado pelo autor

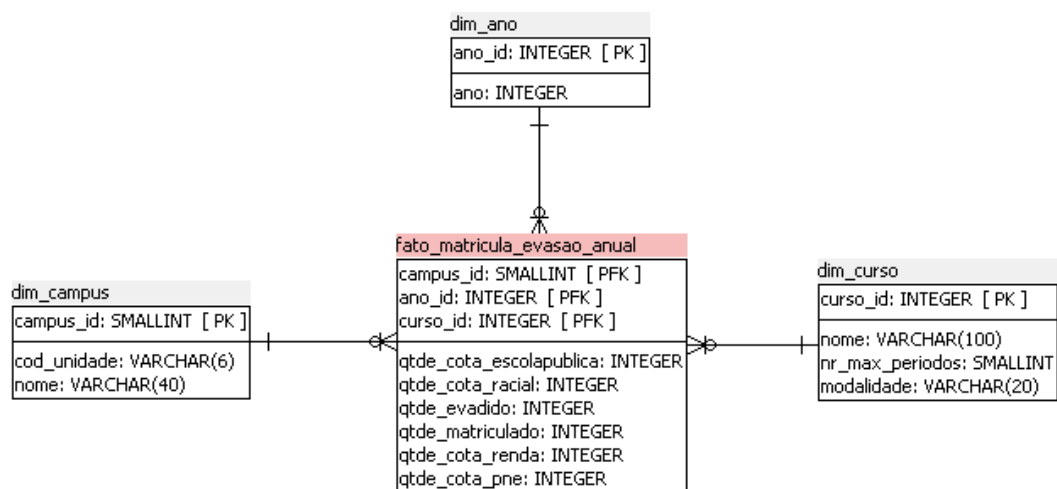


Figura 10 – Modelo Multidimensional para Fato Evadidos por Matriculados

Fonte: Elaborado pelo autor

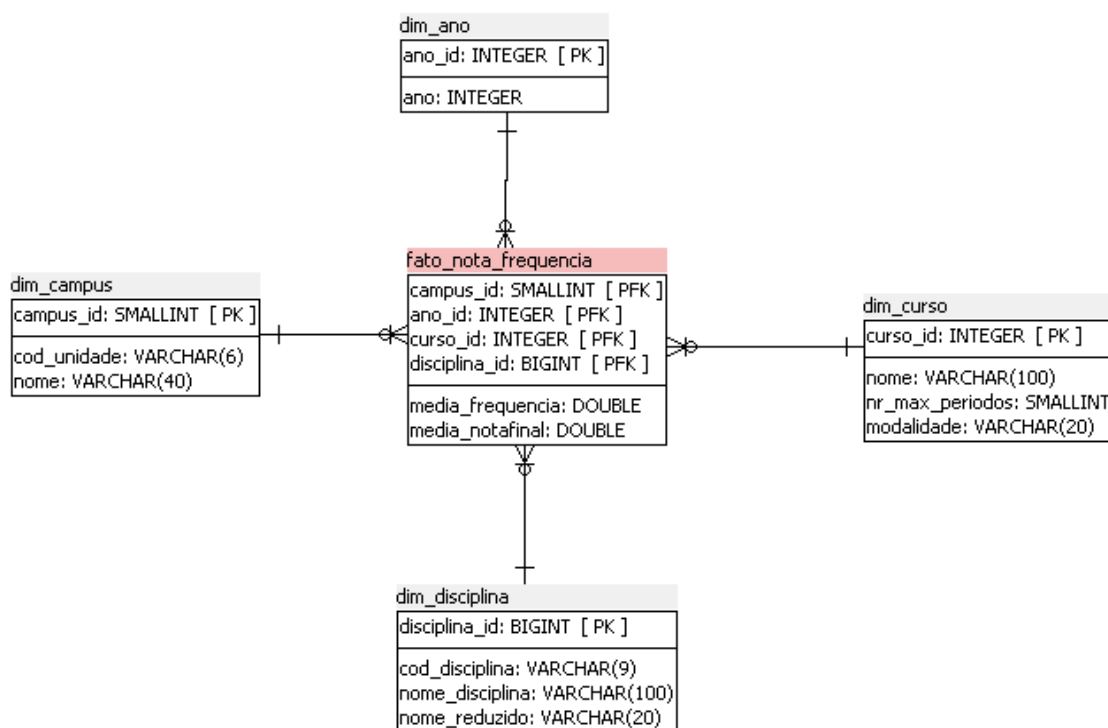


Figura 11 – Modelo Multidimensional para o Fato Nota Frequência

Fonte: Elaborado pelo autor

Passada a fase de desenvolvimento do projeto lógico do banco, passou-se para a **implementação física das bases de dados**: uma para a base *Staging Area* e outro para armazenar as informações dos modelos multidimensionais e que será chamado fisicamente como base DW, composto por tabelas de dimensões e as tabelas fato. Neste momento as seguintes ações foram desenvolvidas:

- Preparação do ambiente de armazenamento;
- Definição dos tipos de objetos e criação dos objetos nos bancos de dados.

Os passos a seguir compreendem o processo de **Extração, Transformação e Carga** que, conforme descrito anteriormente é comumente conhecido num projeto de BI como processo ETL. Os processos a seguir utilizaram a ferramenta *Kettle* da suíte de ferramentas do *Pentaho*.

- Mapeamento e criação dos processos para carga dos dados do SICAD para a base *Staging Area*;
- Mapeamento e criação dos processos de carga das tabelas de dimensões e fatos na base DW;
- Validações e ajustes nas cargas em inconsistências, uniformização de dados e performance das cargas;
- Definição do agendamento dos processos de ETL.

Os passos descritos nesta seção são de fundamental importância na construção de um projeto de BI. O processo de ETL é o que demanda uma considerável fatia do tempo de desenvolvimento da solução e merece uma atenção redobrada. Mesmo com um levantamento prévio de alguns problemas e inconsistências sobre os dados, apontados no Capítulo 2, é durante este processo de ETL que são identificadas a maior parte destas inconsistências o que exigem uma atenção maior de forma a minimizá-las para garantir uma melhor qualidade e veracidade sobre as informações produzidas. Entre as principais ações adotadas para as inconsistências encontradas, podem ser destacadas:

- Adoção de valores padrão para campos nulos;
- Uniformização de valores, como por exemplo transformar gênero em “Feminino” e “Masculino” já que em alguns recebia valores com formas variadas. Estas ações de

uniformização visam também melhorar a qualidade da informação, como por exemplo transformar valores lógicos “0” e “1”, “*false*”, “*true*”, em “Sim” e “Não”, respectivamente.

- Elaboração de rotinas para tratar dados redundantes. Este foi ponto no projeto que exigiu um grande esforço, pois muitos campos no SICAD eram armazenados de uma forma durante algum tempo e a partir de um dado momento passaram a serem armazenados de outra maneira. Um exemplo disto foi o campo etnia, na tabela de alunos, em que constam dois campos (etnia, etnia1).

- Tratamento para ausência de regras de integridade de referência. Esta ação, em alguns casos, exigiu a adoção da mesma prática anterior de aplicar valores padrão quando não era encontrada a informação na tabela de referência.

O resultado desta seção, com o processo construção da modelagem multidimensional e processo de ETL, além ser insumo para a próxima etapa, como será descrito na seção 6.4, servirá de apoio para elucidar um conjunto de ações que a instituição precisará adotar para melhorar a qualidade das informações armazenadas no SICAD ao longo do tempo.

6.4 IMPLANTAÇÃO E DISPONIBILIZAÇÃO DOS CUBOS OLAP

Nesta etapa da pesquisa ocorre o processo de preparação dos modelos multidimensionais construídos anteriormente para serem transformados em cubos e assim serem disponibilizados aos usuários finais. Conforme apresentado na figura 2, na seção 4.3, as informações disponibilizadas em formas de cubos recebem este nome pela forma geográfica que o representa, que permite uma visão multidimensional sobre uma determinada informação. Estes cubos são visões construídas a partir da modelagem multidimensional e disponibilizadas através de uma ferramenta OLAP. Nesta pesquisa, o cenário técnico adotado é o *Mondrian*, servidor OLAP que compõe a solução *Pentaho*, e a ferramenta de visualização *Saiku Analytics*, que funciona como um *plugin* do *Pentaho* para análise dos cubos publicados.

Abaixo seguem as principais atividades desenvolvidas nesta etapa:

- Construção e publicação dos cubos na ferramenta *Schema Workbench* em ambiente de homologação;

- Testes, validações dos cubos na ferramenta OLAP;

- Ajustes e republicação dos cubos;
- Definição das regras de controle de acesso;
- Validação dos dados junto aos usuários;
- Elaboração e atualização do manual básico do usuário para acesso ao ambiente *Pentaho* e de uso das principais funcionalidades do *Saiku Analytics* (APÊNDICE E);
- Apresentação da solução aos usuários;
- Publicação dos cubos no ambiente de produção.

Ainda que as ferramentas de visualização sejam projetadas para serem de fácil utilização, é importante que seja feita uma apresentação da solução OLAP adotada com suas funcionalidades, assim como para cada cubo publicado ou alterado.

O processo de construção e publicação dos cubos merece um destaque. Isto porque é neste processo que se definem as visões que poderão ser disponibilizadas para cada perfil de usuário, de maneira que para um mesmo modelo multidimensional podem ser construídos quantos cubos se fizerem necessários. A necessidade de construção de novos cubos, bem como os ajustes, ocorre durante o período de uso com o amadurecimento das análises sobre os dados, de maneira que para este momento da pesquisa foi construído um cubo para cada uma das tabelas de fato do modelo multidimensional para que os gestores possam validar as visões publicadas, bem como o ambiente analítico.

Outro ponto relevante é que neste processo existe a possibilidade de construção de novas métricas e dimensões a partir das informações contidas no modelo multidimensional, além de definir quais informações estarão visíveis. É comum em projetos de modelagem multidimensional a inclusão de um grande número de atributos ou descritores nas dimensões para que sejam selecionados durante a construção dos cubos. A seguir serão descritos os três cubos construídos e publicados para que os gestores possam realizar análises sobre os dados e validar a efetividade da solução OLAP.

Cubo Evasão Anual: Apresenta a quantidade de evadidos anualmente e permite o cruzamento de dados mais granulares, podendo chegar a analisar métricas por aluno e informações da matrícula. Por este cubo é possível navegar por algumas das informações cadastrais do aluno considerado evadido. A figura 12 apresenta as dimensões e métricas que

compõem este cubo e como pode ser verificado ao comparar o modelo multidimensional representado na figura 9, nota-se que durante a construção deste cubo foi utilizada a estratégia das minidimensões, em que é possível construir uma nova dimensão a partir de uma outra dimensão. Esta estratégia leva em consideração se tratar de um projeto institucional em que as dimensões que compõem o modelo multidimensional podem vir a ser compartilhadas por outros projetos de BI, de modo que é adequado optar por construir tabelas físicas com muitos atributos adotando a estratégia de minidimensões que podem ser usadas durante a construção dos cubos OLAP.

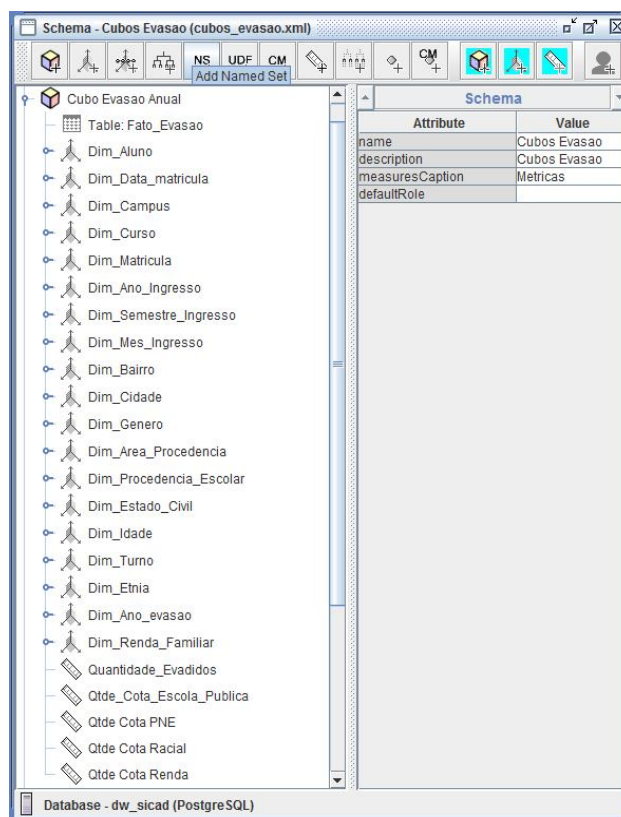


Figura 12 – Cubo Evasão Anual elaborado na ferramenta Workbench

Fonte: Elaborado pelo autor

Cubo Evadidos por Matriculados: Apresenta métricas de quantidade de evadidos e matriculados por Ano, Campus e Curso. Permite a análise da variação anual dos matriculados e dos evadidos. Este cubo apresenta uma visão mais gerencial com dados menos granulares e mais consolidados. Como pode ser verificado na figura 13, algumas das métricas que compõem este cubo não ficam fisicamente armazenadas no modelo representado pela figura 10, são métricas calculadas durante a geração da informação, a partir do método de cálculo

definido na construção do cubo. Conforme dito anteriormente, os cubos são visões sobre o modelo multidimensional previamente definido e através da ferramenta *Workbench*, usada para a construção destes cubos, é possível inclusão de novas métricas ou novas dimensões. Os itens representados na figura 13 com as iniciais “CM” são métricas construídas durante a elaboração deste cubo.

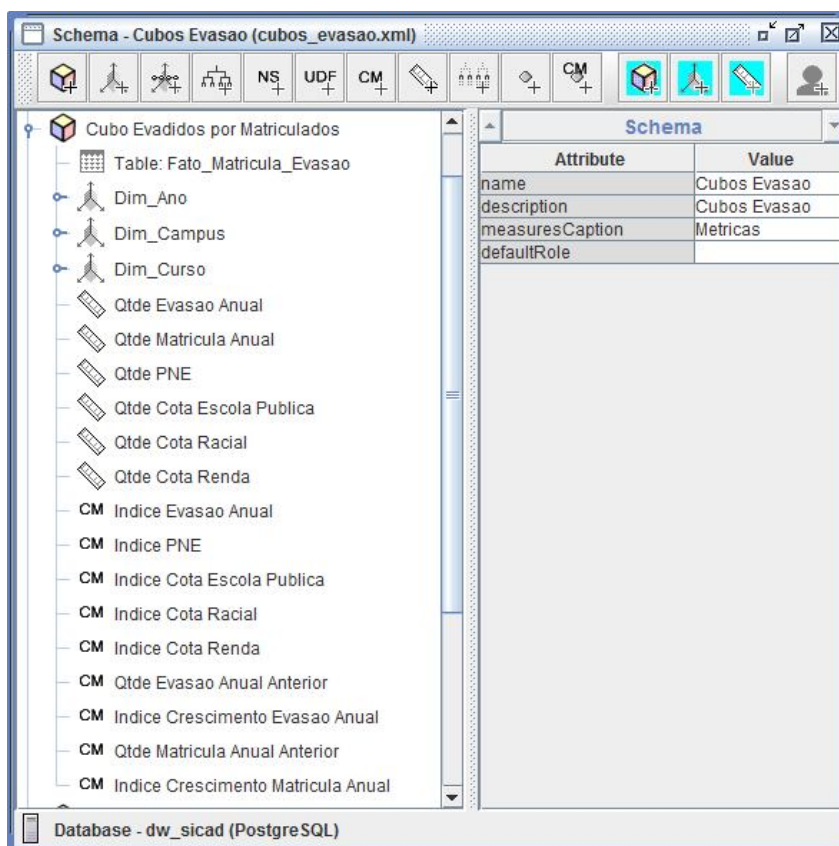


Figura 13 – Cubo Evadidos por Matriculados elaborado na ferramenta *Workbench*

Fonte: Elaborado pelo autor

Cubo Nota e Frequência: Apresenta as métricas médias de notas e de percentual de frequência dos alunos evadidos. Estas métricas estão agrupadas por ano, campus, curso e disciplina, conforme figura 14.

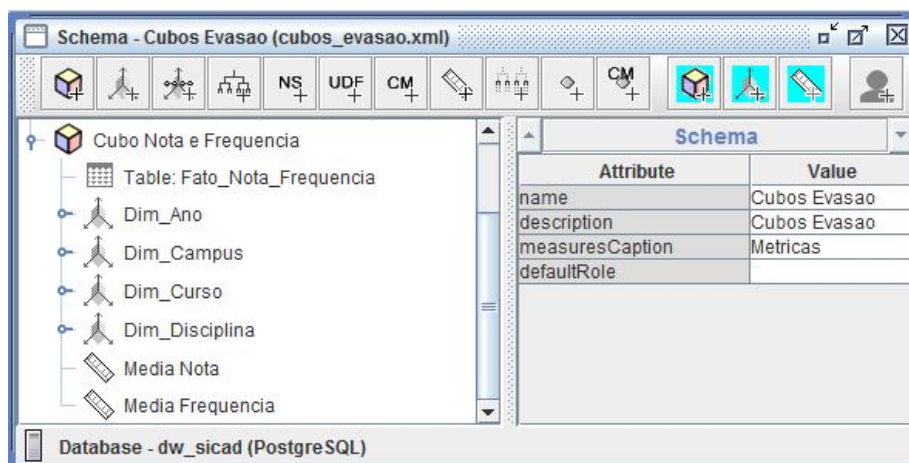


Figura 14 – Cubo Notas e Frequência dos alunos evadidos elaborado na ferramenta *Workbench*

Fonte: Elaborado pelo autor

Por decisão acordada junto com a equipe de TI do IFBA, os cubos serão publicados e disponibilizados em produção no endereço www.indicadores.ifba.edu.br. Este ambiente será compartilhado para todos os projetos de BI que o IFBA venha a desenvolver.

Não é incomum em projetos de BI que ao ter contato com a solução os usuários solicitem constantes alterações, inclusão de novas métricas e descritores. Isso se dá pela própria característica dessas soluções OLAP de permitir que os usuários construam seus próprios indicadores a partir das informações disponíveis. Este é um dos fatos que justificam a adoção de metodologias ágeis para o desenvolvimento de projetos BI. É importante destacar que neste método aqui proposto, as mudanças solicitadas devam passar pelas 4 fases do PCDA, com planejamento, execução, monitoramento e correção.

7 RESULTADOS E DISCUSSÕES

Este capítulo se dedica a realizar uma análise acerca dos resultados deste processo de modelagem multidimensional para a consolidação dos dados sobre a evasão escolar no ensino superior no IFBA campus Salvador. Inicialmente será tratada a questão do crescimento exponencial de dados do SICAD, que conta com uma base de dados histórica de mais de 20 anos de registros acadêmicos e que continua a crescer anualmente com a inclusão diária de novos registros. Em seguida será realizada uma análise sobre os dados consolidados, extraídos a partir da modelagem através da solução OLAP, e quais informações estes dados podem revelar acerca da evasão escolar no ensino superior no campus Salvador, sob o ponto de vista do pesquisador. Por fim, será realizada uma análise, a partir de pesquisa realizada com os gestores, com o objetivo de medir a efetividade da solução OLAP disponibilizada, como sendo o ambiente analítico que permitirá aos gestores a construção de indicadores para apoiá-los no controle da evasão escolar no IFBA.

7.1 ANÁLISE SOBRE O CRESCIMENTO DOS DADOS ACADÊMICOS

Um dos pontos de relevância e consideração nesta pesquisa e também um dos motivadores para adoção da tecnologia BI é o volume de dados históricos produzidos por um sistema acadêmico, que se dá de uma forma exponencial devido às características do tipo de sistema de informação como o SICAD que produz diariamente milhares de novos registros. Para verificar este crescimento nestes registros do SICAD foi realizada uma análise direta às tabelas que compõem o banco de dados onde são armazenados estes dados. O método adotado foi a utilização de contagens de registros às principais tabelas que recebem registros a partir de uma matrícula de um aluno. Deste modo, foram selecionados aleatoriamente 10 alunos matriculados ao longo dos últimos cinco anos para verificar a quantidade de registros que cada aluno produz num período de um ano. A opção em analisar o volume de dados produzido em um ano é que a maioria dos dados em análise nesta pesquisa trata períodos anuais, como é o caso da evasão anual e das matrículas anuais. Destas consultas realizadas, foram extraídos os números a seguir:

Cada matrícula durante um ano produz, em média:

- 233 registros de **frequência**
- 12 registros de **avaliação** (1 por disciplina)
- 1 registro de **matrícula** geral
- 12 **matrículas em disciplinas** (um por disciplina)
- 1 registro de **pessoa física**
- 1 registro de **Aluno**

Num total de **260 registros**, ao longo de um ano, **por aluno**.

Considerando que apenas o campus Salvador matriculou, durante o ano de 2014, 1684 alunos, significa dizer que em um ano, apenas os alunos do campus Salvador do ensino superior, no que se refere aos registros acadêmicos básicos, produziu um volume médio aproximado de **437.840** registros. A análise levou em consideração que cada aluno se matricula por ano em 12 disciplinas em média, sendo seis por semestre. Toda análise se limitou aos matriculados do ensino superior que é parte do objeto desta pesquisa.

Este cenário, em números aproximados, demonstra o quanto uma base de dados desta natureza cresce de forma rápida e num curto período de tempo. Apenas do campus Salvador, que é o escopo de análise desta pesquisa, este volume já passa dos 6 milhões de registros. Isso reafirma a necessidade do uso de tecnologias apropriadas para o tratamento de grandes volumes de dados, como é o caso do BI adotado como solução tecnológica para esta pesquisa.

7.2 ANÁLISE DOS DADOS SOBRE A EVASÃO

Durante grande parte deste trabalho, o pesquisador se dedicou à busca exploratória de informações sobre a Evasão Escolar no ensino superior no IFBA. Esta busca contou com o apoio de diversos departamentos do IFBA, sobretudo aqueles com ligação mais direta aos dados acadêmicos do SICAD. A pesquisa objetiva a construção de uma modelagem multidimensional para a criação de um ambiente de dados consolidado que permita apoiar os gestores no controle da evasão escolar, levando em consideração a dimensão histórica dos dados e o grande volume que é produzido num curto espaço de tempo, conforme relatado na seção anterior. Por tudo isso e com vistas a validar o processo de modelagem dos dados, esta seção será dedicada a analisar dados que podem ser extraídos dessa modelagem, através das ferramentas BI, relacionando estes resultados ao que especialistas e pesquisadores voltados para a evasão vem discutindo, além do olhar do próprio pesquisador.

Antes, é preciso situar dois pontos que são fundamentais. Um deles é a dimensão tempo, conforme dito anteriormente, na perspectiva histórica destes dados. Há que se considerar que este tempo tem como marco o início dos registros acadêmicos em meio digital no IFBA campus Salvador, ou seja, o quanto se retroagiu no lançamento destes registros, já que a pesquisa revelou que embora o SICAD tenha sido implantado a partir de 1999, foram encontrados lançamentos datados de 1984, porém a maior parte das análises terá como marco inicial o ano de 1999, já que os registros anteriores a este período foram feitos de forma retroativa e incompleta. Outro ponto que merece atenção é que algumas informações passaram a ser alimentadas no sistema mais recentemente e por isso ainda não possuem dados suficientes para avaliar séries históricas, mas ainda assim foram acrescentadas na modelagem para análises atuais e futuras.

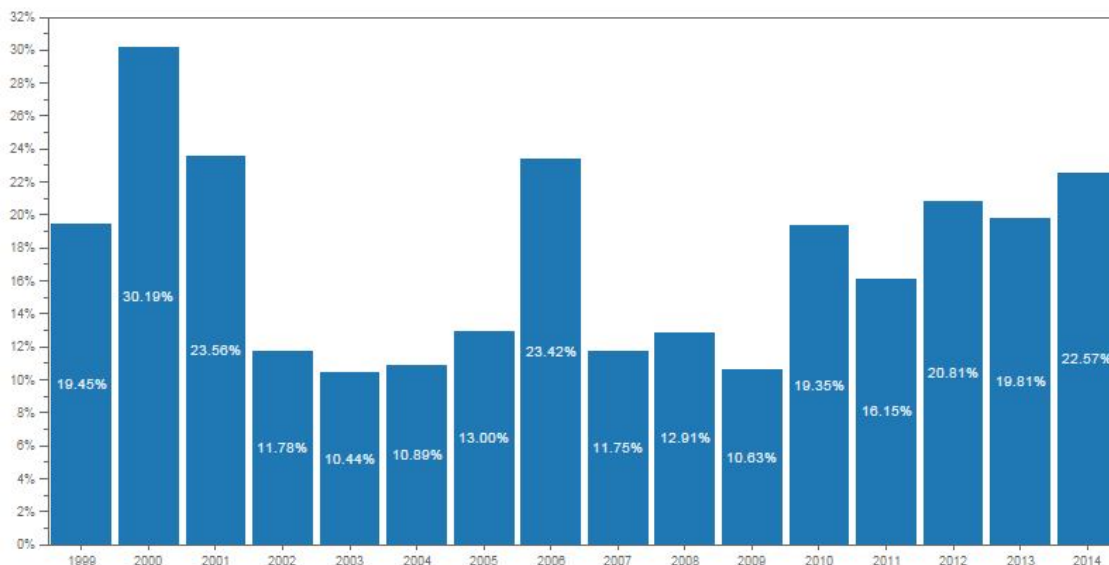
A extração dos dados para esta análise foi realizada a partir do ambiente OLAP, no mesmo ambiente em que os usuários utilizarão para a construção dos seus indicadores. Entre as características das ferramentas OLAP está a de permitir ao usuário visualizar uma mesma métrica sob diversas perspectivas, realizando cruzamentos de informações em diversos formatos, de acordo com a necessidade de cada usuário. A autonomia e a facilidade de uso destas ferramentas são pontos fortes da tecnologia, pois permite que um gestor realize as suas análises sem a necessidade de um profissional especializado em desenvolvimento de *software*, ao contrário, por exemplo, da maioria dos sistemas transacionais, onde as informações e os formatos estão previamente definidos e mudá-los exige esforço e a presença deste profissional com conhecimento técnico especializado. Mas isto implica também dizer que por serem inúmeras as possibilidades de informações que este ambiente pode produzir a partir dos modelos multidimensionais construídos, esta análise seria incapaz de esgotar todas estas possibilidades. Cabe destacar que qualquer estudo sobre a evasão escolar, conforme Baggi et al. (2011, p. 363), “é um campo vasto e complexo, o qual envolve questões pedagógicas, psicológicas, sociais, políticas, econômicas, administrativas, entre outras” de maneira que esta análises não tem pretensão em verificar o fenômeno evasão nesta profundidade, mas o de apoiar para que os gestores possam realizar suas análises. Portanto, os dados e os formatos analisados, foram extraídos a partir do olhar do pesquisador e tendo como base os elementos apontados pelos gestores como principais fatores para a evasão escolar no IFBA *campus* Salvador, resultantes das entrevistas realizadas no Capítulo 3 e apresentados na Seção 6.2, durante a construção da matriz de necessidades. Entre eles, fatores relacionados com o curso, período acadêmico, desempenho dos alunos; fatores socioeconômicos como renda, idade,

estado civil, gênero, etnia, origem escolar do aluno; e o uso do transporte público para chegar à escola.

7.2.1 Índices Gerais de Evasão

Segundo Silva Filho et al. (2007) os índices gerais de evasão anual, entre os anos de 2001 e 2005, no ensino superior, variavam entre 9% e 15%, com uma média de 12% no período, quando analisadas as instituições públicas, enquanto que esta média chegava aos 22% nas instituições privadas, entretanto neste mesmo estudo a média de evasão quando analisado por região registrou uma média de 21% para a região nordeste, com o destaque para a Bahia e Sergipe que apresentavam maiores índices. O gráfico 1, apresenta números bastante variados no IFBA *campus* Salvador. O período histórico analisado é a partir do início de uso do SICAD, de 1999 a 2014. Como se trata de análise da evasão anual, não é possível analisar o ano 2015 já que é necessário o fechamento do ciclo anual para verificar tal indicador que é calculado a partir da fórmula $[(\text{Matrícula Anual} - \text{concluinte do 1}^\circ \text{ semestre}) - (\text{Matrícula do 2}^\circ \text{ Semestre})] \times 100 / \text{N}^\circ \text{ de Matrícula Anual}$, conforme apresentada no Capítulo 3. Além disso, devidos aos movimentos grevistas, o calendário acadêmico do *campus* Salvador para o ano 2015 sofreu alterações com retardo no início das aulas.

Gráfico 1: Índice de Evasão Anual do campus Salvador de 1999 a 2014



Fonte: Construída pelo autor a partir dos dados extraídos da ferramenta OLAP

Observa-se a partir do gráfico 1 que o campus Salvador apresenta uma média 17,29% para um período de 15 anos, e mesmo que analisando apenas o período analisado por Silva Filho et al. (2007), de 2001 a 2005, a média do campus Salvador ultrapassa a média nacional de 12% e mesmo os percentuais de variação que ficaram entre 9% e 15%, com destaque para os anos em que este índice superou os 30%. Outro ponto importante nos números, é que se comparado com aqueles apresentados na tabela 3, no Capítulo 3, há uma pequena variação de percentuais em relação aos números extraídos relatório de gestão que tem os dados originados do SISTEC. Durante entrevistas realizadas com gestores, foi apontada esta possibilidade de variação, já que em alguns casos o período em que os números são lançados no SISTEC pode diferir do período em que ocorre a extração destes dados, mas ainda assim, os gestores consideraram as informações extraídas da ferramenta OLAP como confiáveis, ou pelo menos, mais próximas da realidade da instituição. Um ponto de observação sobre os primeiros índices (anos de 1999 a 2001) do gráfico 1 é que como se tratou do período de implantação do SICAD e de adaptações e amadurecimento, existe a possibilidade de que estes registros iniciais não sejam tão fidedignos e por isso podem distorcer esta análise da média histórica.

De um modo geral o que se nota é que não existe uma linearidade nos números apresentados e que estes números são preocupantes, quando comparados com a média dos índices anuais nacionais para instituições públicas. Observa-se ainda que, a exceção dos três primeiros anos analisados (1999, 2000, 2001) e o ano de 2006, os números vinham se mantendo estáveis e dentro da média. No entanto, os números a partir de 2010 tem se mantido altos. Para contribuir na análise desta oscilação, foi gerado o Gráfico 2 que apresenta um cruzamento entre a variação anual dos índices de evasão com o dos matriculados. Estes índices de variações analisam o ano imediatamente anterior em relação ao ano em análise e apresentam a diferença em termos percentuais. Estes números podem variar para mais e para menos, como pode ser observado no Gráfico 2.

Gráfico 2: Índices de variação da Evasão e do número de Matriculados Anuais do campus Salvador de 1999 a 2014



Fonte: Construída pelo autor a partir dos dados extraídos da ferramenta OLAP

As informações apresentadas pelo Gráfico 2 reafirma a falta de uma linearidade nos indicadores da evasão e eles se mostram mais evidentes quando comparados com as variações dos índices de matrícula, a ponto do índice de evadidos variar em mais de 100% em relação ao ano anterior.

7.2.2 Fatores relacionados a curso, período acadêmico, forma de ingresso e desempenho

Os estudos sobre a evasão apontam que a escolha do curso é um marco na vida do estudante, o momento de definir sobre filosofia de vida, e a escolha da carreira profissional, na busca por encontrar algo que permita relacionar a vocação profissional com a inserção no mercado de trabalho (POLYDORO, 2000; GAIOSO, 2005; BARDAGI, 2007). Mesmo com o aumento das ofertas de vagas que conseqüentemente amplia a possibilidade de acesso, o ingresso do estudante no ensino superior é tido, em especial nas instituições públicas, como uma vitória. Este é um momento de grande orgulho, mas também de grandes expectativas, conforme Polydoro (2000, p. 9) ao afirmar que Rodrigues (1997) “refere-se a esse momento como um encontro com a virtualidade, pois ainda falta ao estudante um conhecimento mais

sólido sobre a carreira escolhida, curso que ingressou e sobre o que significa estar na universidade”, e afirma ainda que esta “visão romântica” pode potencializar negativamente na permanência deste aluno ao curso que ingressou.

As análises que seguem buscam identificar o perfil deste aluno considerado evadido para IFBA, segundo os dados do SICAD, nos aspectos relacionados ao curso escolhido, o período acadêmico em que estes alunos abandonam a instituição, a forma de ingresso ao IFBA, além de apresentar alguns números relacionados à frequência e desempenho (notas) destes alunos evadidos.

A tabela 8 apresenta o percentual de evadidos por curso, considerando a relação entre a quantidade de evadidos e matriculados ao longo de todo histórico de registros acumulados no SICAD para os cursos superiores do campus Salvador. Esta tabela apresentada na ordem decrescente do percentual de evadidos permite verificar os cursos que acumulam os maiores volumes proporcional de alunos evadidos. A exceção do primeiro curso (Tecnologia em Eventos) verifica-se o destaque para os cursos das engenharias, que são cursos situados na área de conhecimento das ciências exatas.

Tabela 8: Índice de Evasão Anual por curso

Curso	% Evasão Anual
TECNOLOGIA EM EVENTOS	45,95%
TELECOMUNICAÇÕES	28,88%
MANUTENCAO PETROQUÍMICA	26,15%
PROCESSOS DE POLIMERIZAÇÃO	25,42%
MANUTENÇÃO ELÉTRICA	24,95%
MANUTENÇÃO MECÂNICA	24,32%
ADMINISTRAÇÃO HOTELEIRA	24,20%
RADIOLOGIA	23,85%
PROCESSOS PETROQUÍMICOS	22,62%
LICENCIATURA EM FISICA.	21,63%
LICENCIATURA EM MATEMATICA	18,32%
ANALISE E DESENVOLVIMENTO DE SISTEMAS	17,65%
ENGENHARIA QUIMICA	16,69%
ADMINISTRAÇÃO	16,28%
ENGENHARIA INDUSTRIAL ELÉTRICA	15,17%
ENGENHARIA INDUSTRIAL MECÂNICA	14,62%
LICENCIATURA EM GEOGRAFIA	14,42%

Fonte: Elaborado pelo Autor e extraído da ferramenta OLAP

Historicamente o IFBA, ainda com outras denominações, concentrou suas ofertas nos cursos tecnológicos, sobretudo nas engenharias. Com a mudança para Instituto Federal, passou a ofertar outros cursos, a exemplo das licenciaturas. Para analisar um cenário mais recente, considerando que muitos dos cursos apresentados anteriormente já não são ofertados pela instituição, é apresentado, através da tabela 9, um panorama do índice de evasão por cursos dos últimos cinco anos. Como pode ser visto, o índice do curso Tecnologia em Eventos era o primeiro de acordo com a tabela 8 e continua como o maior índice na tabela 9. O que a tabela 9 permitiu revelar é que se trata do primeiro ano de oferta do referido curso e já conta com este alto índice de evasão.

Tabela 9: Percentual de evadidos e quantidade de matriculados por curso nos últimos 5 anos

Curso	2009		2010		2011		2012		2013		2014	
	Qtde Mat	% Evasão	Qtde Mat	% Evasão	Qtde Mat	% Evasão	Qtde Mat	% Evasão	Qtde Mat	% Evasão	Qtde Mat	% Evasão
TECNOLOGIA EM EVENTOS	-	-	-	-	-	-	-	-	-	-	37,00	45,95%
PROCESSOS DE POLIMERIZAÇÃO	62,00	24,19%	41,00	29,27%	18,00	16,67%	11,00	27,27%	7,00	28,57%	3,00	33,33%
RADIOLOGIA	40,00	15,00%	72,00	25,00%	84,00	11,90%	96,00	20,83%	95,00	34,74%	91,00	29,67%
LICENCIATURA EM FISICA.	-	-	-	-	39,00	0,00%	90,00	23,33%	106,00	25,47%	121,00	23,97%
LICENCIATURA EM MATEMATICA	-	-	-	-	41,00	0,00%	83,00	20,48%	112,00	21,43%	157,00	19,75%
ANALISE E DESENVOLVIMENTO DE SISTEMAS	79,00	10,13%	134,00	14,18%	160,00	15,00%	188,00	18,62%	183,00	20,77%	202,00	21,29%
ENGENHARIA QUIMICA	29,00	37,93%	81,00	18,52%	99,00	8,08%	143,00	16,08%	161,00	13,66%	170,00	20,59%
ADMINISTRAÇÃO	321,00	9,66%	318,00	19,50%	275,00	17,82%	296,00	25,68%	268,00	20,52%	256,00	23,83%
ENGENHARIA INDUSTRIAL ELÉTRICA	240,00	7,50%	253,00	17,39%	213,00	16,90%	254,00	20,47%	229,00	18,34%	215,00	22,79%
ENGENHARIA INDUSTRIAL MECÂNICA	328,00	8,54%	328,00	21,04%	283,00	24,03%	293,00	20,82%	274,00	15,33%	259,00	21,62%
LICENCIATURA EM GEOGRAFIA	-	-	-	-	39,00	0,00%	88,00	12,50%	126,00	15,08%	170,00	18,24%

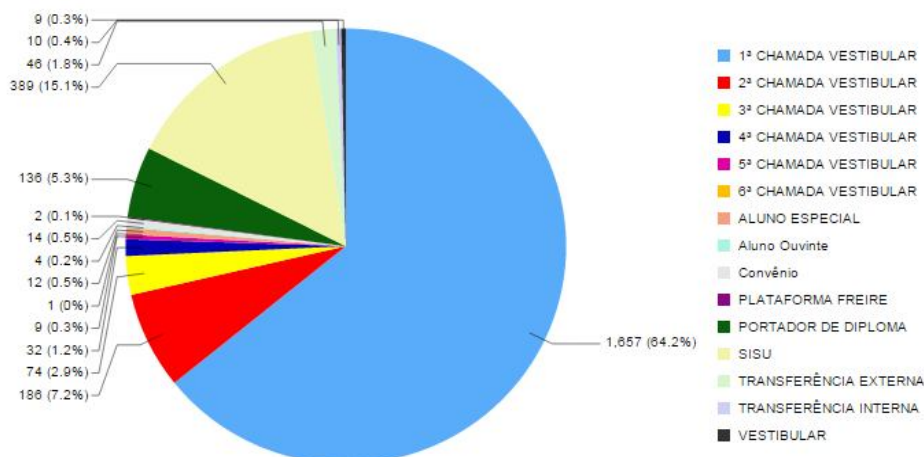
Fonte: Elaborado pelo Autor e extraído da ferramenta OLAP

Embora o curso de Processo de Polimerização apresente percentuais altos, este já não tem sido mais ofertado no vestibular, são apenas as matrículas de remanescentes. Mas um curso recente e que foge ao padrão histórico de oferta de cursos de engenharia é o curso de Radiologia, o primeiro curso ligado a área de saúde, ofertado pelo campus Salvador, com início em 2009 e já conta com um índice de evasão alto, considerando a média geral do campus e média nacional anteriormente mencionadas. Ainda sobre esta mudança no perfil de ofertas de cursos, nota-se que as licenciaturas, com suas primeiras matrículas em 2011, também apresentam índices muito próximos dos demais cursos, com destaque para licenciatura em física que tem ultrapassado as engenharias. Num estudo realizado por Silva

Filho et al. (2007) sobre os índices de evasão anual por área de conhecimento, o destaque eram os cursos das áreas de Ciências, Matemática e Computação, com médias de 28%, Educação e Saúde com 18% e 19% , respectivamente. As Engenharias, Produção e Construção com 21% e as Ciências Sociais com 25%. Com exceção das Ciências Sociais (Administração) e Computação, em que o campus Salvador apresenta índices inferiores às médias apresentadas por Silva Filho et al. (2007), as demais apresentam médias superiores.

Para realizar uma análise a partir da forma de ingresso dos alunos, foi gerado o gráfico 3, que extraiu o percentual de evasão a partir do total de evadidos durante o período de 1999 à 2014. Como pode ser observada, a grande maioria é formada por alunos que ingressaram a partir da primeira convocação para matrícula após resultado do vestibular. O horizonte em análise não permite uma verificar com a devida proporcionalidade os ingressantes a partir do Sistema de Seleção Unificada (Sisu), já que esta forma seleção é mais recente, os primeiros registros no SICAD são a partir de 2010, ainda que o Sisu no gráfico 3 represente o segundo maior índice. Para que seja possível analisar a representação dos evadidos nesta forma de ingresso via Sisu, foi gerada a tabela 10 apenas com o período a partir de 2010, o percentual passa a representar quase 27% do total de evadidos neste período, e na análise histórica apresenta uma crescente. Segundo editais de processo seletivo verificados até o ano de 2014, 50% das vagas são destinadas ao acesso via Sisu.

Gráfico 3: Índices de evasão por forma de ingresso de evadidos entre 1999 e 2014



Fonte: Construída pelo autor a partir dos dados extraídos da ferramenta OLAP

O Gráfico 3 aponta para algo que merece um estudo aprofundado da vida acadêmica destes alunos, já que na 1ª chamada, em princípio, representariam aqueles alunos com o melhor desempenho e estes tem sido os que mais abandonam. Exigiria analisar a vida

acadêmica destes alunos a partir de outras instituições de ensino, já que podem ter abandonado por terem alcançado êxito em outros processos seletivos. Com relação aos acessos via Sisu não é possível afirmar que esta foi sua primeira opção de curso nem no Sisu e nem como escolha do aluno, mas aponta para uma atenção aos alunos que acessam por este processo seletivo. Um dado que pode ser observado na tabela 10, além do crescimento de matriculados pelo Sisu gradual, é o crescimento na quantidade de chamadas, chegando a seis para preencher a quantidade de vagas.

Tabela 10: Total de evadidos e matriculados por forma de ingresso entre 2010 e 2014

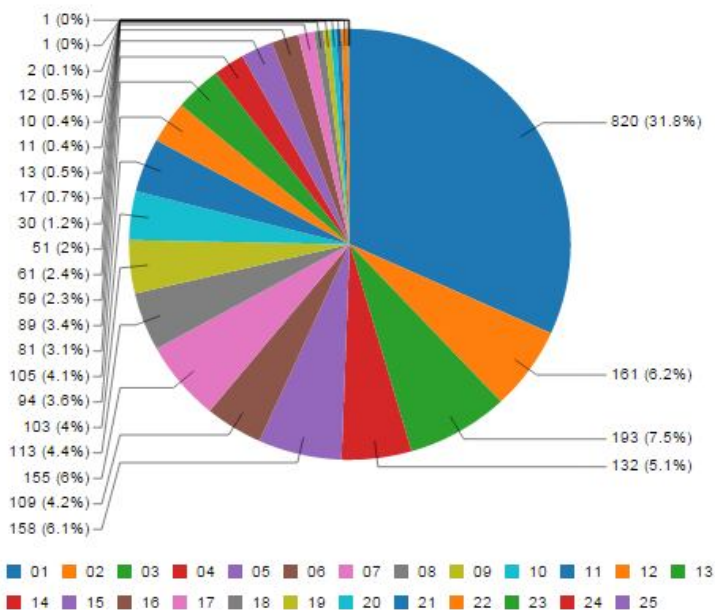
Forma de Ingresso	2010		2011		2012		2013		2014	
	Matriculados	Evadidos	Matriculados	Evadidos	Matriculados	Evadidos	Matriculados	Evadidos	Matriculados	Evadidos
1ª CHAMADA VESTIBULAR	878	177	797	123	792	166	687	138	575	135
SISU	157	33	237	45	397	92	382	103	471	116
2ª CHAMADA VESTIBULAR	125	18	124	14	149	27	193	23	263	43
3ª CHAMADA VESTIBULAR	13	2	38	1	78	8	117	18	155	37
4ª CHAMADA VESTIBULAR	3	1	2		38	5	89	7	101	19
5ª CHAMADA VESTIBULAR					20	2	24	2	37	5
6ª CHAMADA VESTIBULAR					2		2		2	1
ALUNO ESPECIAL	9		6	2	8	5	1		2	
Convênio	1		12	3	3	1	9	8	1	
PLATAFORMA FREIRE	1		1		2	2	1		3	
PORTADOR DE DIPLOMA	43	7	28	10	20	5	12	3	6	3
TRANSFERÊNCIA EXTERNA	14	1	29	4	53	7	66	8	65	11
TRANSFERÊNCIA INTERNA	2		3	1	6	1	6	1	6	2
VESTIBULAR	1		1		1	1	1		24	8

Fonte: Elaborado pelo Autor e extraído da ferramenta OLAP

Para fins de esclarecimento, a forma de ingresso “Vestibular” era a forma inicialmente concebida no sistema e que não classificava em qual chamada o aluno entrava. Embora existam registros em 2014 com esta forma de ingresso, estes são registros remanescentes, já que a forma de ingresso não modifica durante toda a vida escolar do aluno.

No gráfico 4 busca-se observar entre os anos de 1999 e 2014 em que período acadêmico estava o aluno matriculado quando abandonou o curso.

Gráfico 4: Índices de Evasão por forma período acadêmico entre 1999 e 2014



Fonte: Construída pelo autor a partir dos dados extraídos da ferramenta OLAP

Pode-se concluir que mais de 50% dos alunos evadidos cursavam os quatro primeiros semestres, com destaque para o primeiro, que representa quase 32% do total de evadidos. Isso corrobora com a afirmação de Silva (2013), ao dizer que o período inicial apresenta riscos potenciais para a evasão a partir dos primeiros contatos do aluno com o curso e este risco diminui com o passar dos semestres.

A partir da modelagem e dos cubos publicados também é possível extrair informações a respeito do desempenho e da frequência destes alunos, conforme tabela 11. Como estas informações passaram a serem alimentadas a partir de 2008 no SICAD, foram filtradas informações a partir deste período até 2014. Como pode ser observado, os números mostram extremos, onde os alunos apresentam baixas médias no quesito nota, enquanto apresentam altos índices de frequência, com exceção apenas para o curso Tecnologia em Eventos, que curiosamente foi o mesmo que apresentou no seu primeiro ano ofertado mais de 45% de evadidos, conforme tabela 9. Pesquisas realizadas por Silva (2013) e Freitas (2009) apontam o desempenho escolar com um dos principais fatores da evasão escolar no ensino superior.

Tabela 11: Total de evadidos e matriculados por forma de ingresso entre 2010 e 2014

Curso	2008		2009		2010		2011		2012		2013		2014	
	Média Nota	Média Frequência %	Média Nota	Média Frequência %	Média Nota	Média Frequência %	Média Nota	Média Frequência %	Média Nota	Média Frequência %	Média Nota	Média Frequência %	Média Nota	Média Frequência %
ADMINISTRAÇÃO	5,32	83,67	6,12	86,84	5,82	86,80	5,55	83,51	5,41	86,98	5,82	87,47	2,64	75,11
ANALISE E DESENVOLVIMENTO DE SISTEMAS			4,27	78,70	5,36	81,12	5,02	84,86	4,69	85,75	4,28	84,40	2,67	70,28
ENGENHARIA INDUSTRIAL ELÉTRICA	4,77	80,71	5,15	81,27	4,92	79,35	4,79	82,72	4,90	81,05	4,43	82,37	3,94	86,57
ENGENHARIA INDUSTRIAL MECÂNICA	5,68	83,89	5,56	82,89	5,60	86,89	5,47	85,32	5,65	86,00	5,58	86,98	4,35	85,47
ENGENHARIA QUÍMICA			4,66	77,55	6,42	87,54	6,23	91,59	5,80	87,96	4,84	89,74	4,10	91,93
LICENCIATURA EM FÍSICA.							3,83	75,17	4,59	84,77	2,83	79,55	3,02	69,76
LICENCIATURA EM GEOGRAFIA							6,79	88,29	4,52	89,47	4,80	87,05	3,53	68,13
LICENCIATURA EM MATEMÁTICA							5,45	79,00	3,72	76,50	1,47	67,14	2,03	64,75
PROCESSOS DE POLIMERIZAÇÃO	5,41	81,45	4,15	65,28	5,28	74,38	5,88	87,50	2,78	47,50	2,00	100	0,00	0,00
RADIOLOGIA			4,71	78,07	4,49	79,08	3,60	87,77	5,32	90,72	3,42	84,50	0,93	62,85
TECNOLOGIA EM EVENTOS													0,04	20,25

Fonte: Elaborado pelo Autor e extraído da ferramenta OLAP

A maioria das informações extraídas, podem ou não serem conclusivas de acordo com a perspectivas com que se analisa e quem as analisam, mas o que estas informações permitem é avançar na busca por outras informações para se construir um conceito acerca do que se busca analisar. A questão do turno em que o curso é ofertado, por exemplo, apontadas por Baggi e Lopes (2011) e Fialho (2014) e motivo de preocupação pelos gestores do IFBA entrevistados, conforme Capítulo 3, é fato curioso, pois o curso com maior índice de evasão, com menor índice de desempenho e frequência, segundo números apresentados, é o único curso do campus Salvador ofertado no turno vespertino, todos os demais são ofertados no turno noturno. Outras informações sobre os cursos ofertados pelo campus Salvador podem ser extraídas e analisadas a partir da modelagem e dos cubos publicados e serão exploradas pelos IFBA com o avanço no uso da ferramenta ao longo do tempo.

7.2.3. Fatores Socioeconômicos

O aumento crescente do acesso ao ensino superior traz consigo uma população universitária cada vez mais diversificada. Para Polydoro (2000, p. 7), “[...]essa heterogeneidade refere-se, de maneira geral, ao perfil socioeconômico e cultural, faixa etária, história acadêmica anterior [...]”. Isto implica para a instituição o desafio em atender às diversas expectativas trazidas pelos ingressantes que podem se tornar fatores determinantes para decisão em continuar ou abandonar o curso. Gaioso (2005) corrobora para a questão da idade, ao relacionar a imaturidade do aluno na escolha do curso, estes que muitas vezes ingressam na universidade ainda na adolescência, embora o aspecto imaturidade não esteja totalmente relacionado à idade, mas ainda às experiências vivenciadas pelo sujeito.

A análise de fatores socioeconômicos é um assunto de grande complexidade, o que exigiria estudos aprofundados, com a presença de profissionais de diversas áreas do conhecimento, numa análise qualitativa destes aspectos. As análises a seguir são de caráter quantitativo como uma forma de demonstrar uma maneira como o modelo proposto pode contribuir para análises da evasão no âmbito do IFBA campus Salvador, sob a visão do pesquisador, buscando relacionar, sempre que possível, às análises de especialistas no assunto.

Na tabela 12 foram relacionados os percentuais de alunos ingressantes e evadidos considerando o período entre 1999 e 2014, agrupados por faixa etária. Embora os estudos estejam voltados para a evasão, foram extraídos os percentuais de matriculados para saber a relação existente entre os percentuais tanto daqueles que abandonam quanto daqueles que ingressam. O que se pode observar é que há uma similaridade nestes números, com pequenas variações. Com isto, permite-se verificar a maior concentração de ingressantes está na faixa etária entre 20 e 24 anos, assim como o percentual significativo de evasões nesta mesma faixa etária, neste caso representando o segundo maior nível. Polydoro (2000) chama atenção para este período na vida do estudante, por representar um período de grandes mudanças e adaptações na transição entre a adolescência e a vida adulta, além de fatores relacionados à vida profissional destas pessoas.

A faixa etária “Não Informada” foi o tratamento dado aos alunos que estão com a data de nascimento nula ou inválida, mas se trata de um percentual pouco representativo em relação ao conjunto de números.

Tabela 12: Percentual de evadidos e matriculados por faixa etária entre 1999 e 2014

Faixa Etária	% Evadidos	% Matriculados
0 a 14 anos	0,08%	0,15%
20 a 24 anos	26,05%	31,03%
25 a 29 anos	29,74%	26,66%
30 a 34 anos	21,20%	20,15%
35 a 39 anos	8,85%	8,33%
40 a 44 anos	4,85%	5,04%
45 a 49 anos	2,99%	3,02%
50 a 54 anos	2,50%	2,01%
55 a 59 anos	1,31%	1,08%
60 a 64 anos	0,35%	0,43%
65 a 69 anos	0,01%	0,04%
70 a 74 anos	0,08%	0,04%
Não Informada	1,98%	2,01%

Fonte: Elaborado pelo Autor e extraído da ferramenta OLAP

A figura 15 agrupou um conjunto de informações extraídas da ferramenta OLAP que por melhor disposição ao texto foram agrupadas numa única figura. Todas as informações levam em consideração o total de evadidos entre os alunos de 1999 e 2014. Entre as informações apresentadas nesta figura, cabe destacar que as informações relacionadas a Etnia, Procedência Escolar, Área de Procedência, Atividade Remunerada e Renda Familiar Per Capta – RFP não possui informações suficientes que permitam análises e conclusões neste momento, isto porque algumas destas informações passaram a serem alimentadas no SICAD recentemente, como Atividade Remunerada e RFP e outras não foram alimentadas ao longo do tempo. Ainda que pese o fato de não permitir analisar estas informações neste momento o modelo multidimensional proposto e as visões publicadas estão preparadas para extrair tais informações, que são questões apresentadas pelos gestores que podem elucidar informações sobre o perfil do aluno que evade, além serem apontadas por pesquisadores como fatores causadores da evasão em diversas pesquisas (SILVA, 2013; POLYDORO 2000, SILVA FILHO et al. 2007; FREITAS, 2009; FIALHO, 2014). Um fato que estes números permitem analisar desde logo é a necessidade da instituição manter os dados cadastrais atualizados para melhorar a qualidade das informações institucionais.

Gênero	% Evasão
Feminino	30,88%
Masculino	69,12%

Estado Civil	% Evasão
CASADO (A)	6,94%
DIVORCIADO	0,31%
SOLTEIRO (A)	92,72%
VIUVO (A)	0,04%

Etnia	% Evasão
Amarelo	0,04%
Branco	1,16%
Indígena	0,15%
Não informada	93,61%
Outros	0,43%
Pardo	3,02%
Preto	1,59%

Área de Procedência	% Evasão
Não informada	98,72%
Rural	0,04%
Urbana	1,24%

Procedência Escolar	% Evasão
Não informada	96,28%
Oriundo de Escola Particular	1,51%
Oriundo de Escola Pública	2,21%

Renda Familiar Percapta	% Evasão
01 - RFP < 0,5 SM	0,39%
02 - 0,5 SM < RFP < 1 SM	0,81%
03 - 1 SM < RFP < 1,5 SM	1,16%
04 - 1,5 SM < RFP < 2,5 SM	0,81%
05 - 2,5 SM < RFP < 3 SM	0,81%
06 - RFP > 3 SM	1,36%
Não Informada	94,65%

Atividade Remunerada	% Evasão
Não	0,93%
Sim	0,50%
Não Informada	98,57%

Figura 15 – Conjunto de fatores socioeconômicos extraídos da ferramenta OLAP

Fonte: Elaborado pelo autor

O fator gênero apresenta uma maior concentração de evadidos para o gênero masculino, mantendo a mesma proporção entre os ingressantes que representam, respectivamente, 32,4% e 67,6% entre o gênero Feminino e Masculino. Ainda assim o percentual de alunos do sexo masculino que evade é maior que o número de ingressantes, mesmo que com uma pequena variação, em oposição ao sexo feminino em que o percentual de ingressantes é maior que o percentual de evadidos. Este fato condiz com o que Fialho (2014) e Polydoro (2000) afirmam, em que os alunos do sexo masculino tendem a se manter ou abandonarem por questões financeiras, ao mesmo tempo em que são motivados a permanecerem para melhorar as condições financeiras, evadem quando são pressionados a decidir trabalhar ou estudar, ao passo em que as mulheres abandonam por motivos relacionados ao estado civil, filhos, atividades domésticas. Uma informação que poderia contribuir nesta análise seria a indicação da atividade remunerada e renda, mas estas informações no SICAD ainda não permitem conclusões.

Quando analisado estado civil, a grande maioria declarou no ato da matrícula ser do Solteiro e esta proporção se mantém no índice de evadidos, porém estes indicadores poderiam revelar informações determinantes se atualizadas com frequência, o que não é o caso, já que uma mudança no estado civil de uma pessoa, como dito, pode motivar sua saída da instituição

de ensino. Isto aponta, mais uma vez, para a necessidade de atualizações periódicas de informações cadastrais por parte da instituição.

Outros aspectos que podem ser verificados a partir da modelagem e dos cubos publicados são as métricas relacionadas às Cotas (Portadores de Necessidades de Especiais – PNE, Estudantes de Origem em Escolas Públicas, Raça e Renda). Como pode ser verificado na tabela 13, estes registros também são mais recentes. Alguns valores anteriores a 2008 podem estar relacionados a lançamentos retroativos. A cota relacionada a renda só teve início em 2013 a partir da instituição da Lei nº 12.711, de 29 de agosto de 2012, que versa sobre a reserva de 50% das vagas para alunos oriundos de famílias com renda per capita igual ou inferior a 1,5 salário-mínimo. Porém isto não significa que todo aluno que possui esta renda esteja neste universo da cota renda, uma vez que esta informação indica a que o aluno ingressou usando este critério, mas um aluno mesmo com renda dentro dos critérios da referida lei pode ter entrado a partir da ampla concorrência.

Tabela 13: Totais de evadidos e matriculados entre 1999 e 2014 através das Cotas

Ano	Matriculados					Evadidos				
	Total	PNE	Cota Escola Publica	Cota Racial	Cota Renda	Total	PNE	Cota Escola Publica	Cota Racial	Cota Renda
1999	802	0	0	0	0	156	0	0	0	0
2000	212	0	0	0	0	64	0	0	0	0
2001	399	0	0	0	0	94	0	0	0	0
2002	484	0	0	0	0	57	0	0	0	0
2003	565	0	0	0	0	59	0	0	0	0
2004	661	0	0	0	1	72	0	0	0	0
2005	731	0	0	0	1	95	0	0	0	0
2006	773	0	1	0	1	181	0	0	0	0
2007	928	0	1	0	1	109	0	0	0	0
2008	945	0	1	0	1	122	0	0	0	0
2009	1.101	2	155	153	1	117	0	21	21	0
2010	1.235	8	220	212	1	239	2	30	30	0
2011	1.257	11	284	262	0	203	2	35	34	0
2012	1.547	18	379	349	0	322	5	58	56	0
2013	1.570	19	465	420	66	311	3	75	66	12
2014	1.684	17	440	386	102	380	4	108	100	27
Total	14.894	75	1946	1782	175	2581	16	327	307	39

Fonte: Elaborado pelo Autor e extraído da ferramenta OLAP

Para realizar inferências sobre o transporte público, foi proposta a inclusão das dimensões cidade e bairro para análise sobre o deslocamento do estudante a partir do seu endereço de residência, porém embora tenha sido modelado, os resultados obtidos revelaram que dados de endereço ou estão muito desatualizados ou atualizados de forma incompleta, uma vez que foram encontrados muitos registros de endereços em que o bairro e o CEP pertenciam a cidades diferentes, de forma que inviabiliza a análise neste momento e isto reafirma um dos problemas apontados no Capítulo 2 sobre a falta de integridade de alguns dos dados armazenados. A solução deste problema de integridade relacionada ao endereço dos alunos exigirá ações no sentido de atualizar os registros destes alunos já que este trabalho não se propõe em modificar dados cadastrais mantidos pelo SICAD.

Como dito, algumas das informações acerca dos fatores socioeconômicos passaram a ser alimentadas recentemente no SICAD e ainda não são suficientes para tirar informações conclusivas, entretanto o modelo construído conseguiu mapear um conjunto de dimensões e métricas em torno destes fatores, aqueles em que os gestores que participaram desta pesquisa consideraram determinantes analisar sobre os alunos considerados evadidos no ensino superior, e com o passar do tempo poderão permitir que o IFBA possa extrair indicadores que os apoiem na análise destes fatores.

O conjunto de informações e indicadores possíveis de serem extraídos a partir das visões publicadas através dos Cubos OLAP permite a construção de muitas outras informações, sob diversas perspectivas, em níveis diferentes de granularidade, assim como outros pontos de vistas e informações conclusivas a partir do olhar de quem as analisam. Porém é importante destacar que as análises realizadas até então permitiram não só revelar a possibilidade de extração das informações a partir do modelo construído, mas também a necessidade da implementação de ações por parte do IFBA no sentido melhorar a qualidade dos dados alimentados a partir de intervenções no próprio SICAD, assim como outras ações integradas de atualização periódica dos dados que podem variar no tempo.

7.3 ANÁLISE SOBRE O AMBIENTE ANALÍTICO PROPOSTO

O ambiente analítico adotado compreende o servidor *Mondrian* e seus recursos, que envolve o uso de ferramenta OLAP, Painéis de Controle (*Dashboard*), geradores de relatórios, entre outros, disponíveis através de um navegador web. Estes recursos devem ser capazes de oferecer aos usuários finais um ambiente de fácil acesso aos dados, amigável, que permita o

cruzamento e a análise dos dados sob diferentes prismas, de acordo com suas necessidades. Além disso, essas soluções devem oferecer boa performance, flexibilidade, com formatos de dados variados (PRIMAK, 2008). Para analisar a efetividade do ambiente analítico proposto, foram disponibilizados relatórios predefinidos, com diversas opções de formatos de dados, além de um manual básico, conforme APÊNDICE E, para uso e manipulação do ambiente pelos gestores, a partir dos cubos publicados. Em seguida foi aplicado um questionário com o objetivo de verificar em que medida o ambiente analítico atende aos seus propósitos e às expectativas dos gestores, sob o ponto de vista destes. Este momento da pesquisa não busca analisar o resultado das informações que podem ser extraídas, uma vez que não haveria tempo suficiente até a entrega deste trabalho, e estas análises ocorrerão no decorrer do tempo de uso da solução pela instituição.

O questionário contou com dez questões, conforme APÊNDICE D, sendo sete delas fechadas, com proposições definidas, e três questões abertas onde os gestores puderam apresentar seu ponto de vista. As seis primeiras questões são afirmações em que os gestores puderam manifestar em que nível concordam ou discordam da afirmação. Para cada um destas seis questões foram oferecidas cinco proposições: Discordo Completamente; Discordo; Neutro; Concordo e Concordo Plenamente, sendo que para cada uma foi atribuído uma pontuação de 1 a 5, respectivamente. O questionário foi disponibilizado na Internet, através do gerador de formulários da *Google (Google Forms)*. Participaram deste questionário seis gestores, sendo quatro aqueles que participaram das entrevistas sobre a evasão escolar, conforme relatadas no Capítulo 3, e dois gestores de Tecnologia da Informação, que além de terem contribuído ao longo da pesquisa durante o levantamento das informações armazenadas no SICAD atuam na implantação e manutenção dos sistemas acadêmicos no IFBA.

Como orientação para responder ao questionário, foi proposto aos gestores lerem previamente o manual, conforme APÊNDICE E, e navegarem pelo ambiente analítico, seguindo os passos propostos no referido manual. A primeira questão fez a seguinte afirmação: “O Ambiente disponível é de fácil uso e se mostra adequado para uso pelos gestores educacionais.”. Quatro responderam concordarem, enquanto dois afirmaram que concordam plenamente, de modo que se tirada a média a partir da pontuação, este valor fica aproximadamente 4,3. A segunda afirmação foi: “Durante a exploração do ambiente com a geração de alguns resultados possíveis, o ambiente apresentou uma performance satisfatória.” e desta vez três gestores responderam concordarem e os outros 3 responderam concordarem plenamente, o que dá uma média de 4,5. A terceira questão afirma que “Os formatos dos

resultados gerados (planilhas, gráficos e painéis) pelas consultas realizadas através dos cubos disponíveis são satisfatórios.”. O valor médio obtido coincide com a primeira, com 4,3. A quarta questão, que visava ratificar a três primeiras questões, afirma que “Consegui extrair resultados de forma simples e eficiente.”, e mais uma vez a média obtida foi de 4,3, o que aponta para a concordância quanto à afirmação, portanto coerente com as afirmações anteriores. Estas quatro questões iniciais tiveram como foco verificar a facilidade de uso, performance e variedade de recursos disponíveis pelo ambiente proposto, estas que são características necessárias às tecnologias de análise de dados que compõe soluções BI, conforme Primak (2008), e que puderam ser comprovadas a partir das respostas obtidas pelos gestores.

As questões cinco e seis ainda são afirmações, porém estas visaram obter dos gestores nível de concordância em relação às afirmações que tratam do ambiente proposto com foco no apoio ao controle da evasão escolar especificamente. Sendo assim, a questão cinco afirma: “Considero o ambiente proposto, bem como os formatos com que se pode extrair indicadores, adequados para apoiar no monitoramento e controle da evasão escolar.”, enquanto a questão seis afirma que “No Painel de Monitoramento proposto como exemplo, com recursos visuais gráficos em formatos variados, KPIs e planilhas, permitirá acompanhar de forma centralizada indicadores sobre a evasão escolar.”. Para responder à questão cinco, eles foram orientados a gerarem informações a partir dos cubos disponíveis em formatos tabulares e nos diversos modelos de gráficos disponíveis pelo *Saiku Analytics* e todos responderam concordarem. Para a questão seis, foi disponibilizado um Painel de Monitoramento, composto por um compilado de indicadores em diversos formatos, inclusive com exemplos através de KPI, permitindo que eles pudessem filtrar indicadores históricos em intervalos de anos para acompanhamento de metas, por exemplo, conforme figura 16. As metas apresentadas nos KPI do painel foram arbitradas, já a visão gráfica, neste momento, é apenas a demonstração de mais um dos recursos possíveis de serem usados pelo ambiente analítico. Quatro dos gestores responderam que concordam e outros dois que concordam plenamente com a afirmação. As respostas obtidas para estas seis primeiras questões permitem concluir que o ambiente atende ao seu propósito de ser de fácil uso, voltados aos gestores, com performance satisfatória nos resultados gerados e de permitir apoiar no controle e monitoramento da evasão escolar.

A questão sete tinha como pergunta: “Já conhecia este ambiente tecnológico de *Business Intelligence*?”. Metade afirmou já conhecer, a outra metade disse não ter tido contato com este ambiente. Ainda assim, mesmo aqueles que disseram não ter conhecimento prévio

do ambiente de BI, ao analisar as primeiras respostas referentes à facilidade de uso e performance, todos responderam ao menos concordarem com as afirmações sobre estas características.



Figura 16 – Painel de Monitoramento construído e disponibilizado através do ambiente OLAP

Fonte: Elaborado pelo autor através do *Mondrian*

É importante ainda destacar que o Painel de Monitoramento apresentado nesta Figura 16 é um poderoso recurso disponível em soluções de BI, que além de permitir o acompanhamento de metas, com um conjunto de informações apresentadas nos mais variados

formatos, torna-se um atrator para que gestores passem a se interessar pela tecnologia e até mesmo propor a inclusão de novos recursos para o seu acompanhamento diário.

As questões de oito a dez foram abertas, permitindo aos gestores expressarem seus pontos de vista da maneira que eles considerassem melhor para se manifestar. A questão oito perguntou “Quais as dificuldades encontradas no uso da ferramenta? Quais melhorias você indicaria?”. Tiveram algumas questões pontuais como o idioma padrão em português, que foi solucionado de imediato, todos responderam ser uma ferramenta de fácil uso, como pode ser visto pelas respostas: “Não houve dificuldades.”; “Como já conhecia a ferramenta, não tive qualquer tipo de dificuldade, entretanto, avalio que para os usuários não familiarizados com o ambiente, o manual disponibilizado se mostrou satisfatório e de linguagem fácil e direta.”. Como sugestões de melhorias um dos gestores informou que alguns valores no Painel de Monitoramento poderia ser convertidos para percentuais e um outro gestor afirmou não ter neste momento, mas que poderá surgir com o tempo de uso.

A nona questionou “Qual sua opinião sobre a interface do ambiente OLAP?” e obteve de maneira geral as respostas: “Intuitivo e atende a necessidade”; “Satisfatório. Ágil e dinâmica.”; “Excelente”. Um dos gestores respondeu “Gostei, espero que a inclusão de dados seja simples e aceite importação.”. Para este foi esclarecido que este processo de importação e transformação de dados é automatizado e transparente ao usuário final.

Por fim, a décima questão foi “Qual a sua opinião sobre o Painel de Monitoramento proposto?”. Todos informaram ter gostado muito e ser um ambiente útil para o acompanhamento. Um dos gestores se manifestou com a sugestão de inclusão de outras informações: “Muito bom, sugiro a inclusão das evasões também por semestre e mais itens sobre evasão por etnia (gênero/faixa etária/renda)”.

Embora outras questões, sugestões ou dificuldades possam surgir a partir do amadurecimento com o uso e exploração do ambiente, neste momento da pesquisa, a partir das manifestações dos gestores que responderam ao questionário, é possível afirmar que o ambiente analítico proposto se mostra efetivo para o que se propõe que é servir de ambiente tecnológico que permita apoiar estes e outros gestores na geração e análise de informações acerca da evasão escolar no ensino superior.

8 CONSIDERAÇÕES FINAIS E PERSPECTIVAS FUTURAS

O desenvolvimento desta pesquisa teve como problemática verificar de que forma os dados históricos armazenados no sistema acadêmico do IFBA campus Salvador podem contribuir ou apoiar no controle da evasão escolar no ensino superior, este que como visto, vem sendo tema de grandes preocupações para as instituições de ensino superior, bem como os órgãos que controlam estes indicadores educacionais. Na busca por responder à questão central posta, este trabalho traçou como objetivo principal a construção de um modelo de dados multidimensional, tomando como base os dados acadêmicos históricos do sistema acadêmico utilizado pelo campus Salvador para mapeamento das informações acerca da evasão escolar, além de um conjunto de objetivos específicos, como ações para atingir o objetivo principal proposto.

Neste sentido, as considerações finais aqui expostas estão organizadas em três pontos fundamentais. O primeiro está relacionado ao balanço positivo do alcance dos objetivos propostos. O segundo evidencia as principais dificuldades encontradas na trajetória da pesquisa e suas limitações e por fim destaca algumas sugestões para ações e projetos futuros elucidados.

O alcance dos objetivos propostos nesta pesquisa pode ser visto como um marco na implementação de instrumentos voltados ao controle e acompanhamento da evasão escolar no IFBA, com o uso de recursos de TIC, em especial com o BI. A pesquisa revelou a inexistência de ações sistemáticas e do desenvolvimento de instrumentos voltados para a questão da evasão no âmbito do IFBA, embora o seu PDI estabeleça esta como uma meta a ser atingida. Mas também um marco por considerar este como sendo apenas um pequeno passo do caminho que precisa ser percorrido para apoiar na mitigação do fenômeno da evasão escolar. Pode se considerar também um passo inicial, quando se tratando do uso do BI, diante do conjunto de recursos que ainda pode ser explorado da tecnologia proposta, esta que é uma tecnologia emergente e em constante evolução. Os resultados apresentados permitem afirmar de forma positiva o alcance dos objetivos. Isto foi demonstrado a partir do mapeamento de um conjunto de informações que podem ser extraídas do ambiente analítico implantado, a partir dos dados históricos armazenados no SICAD, transformados e consolidados a partir do processo de modelagem multidimensional, como pode ser revelado no Capítulo 7 que tratou

dos resultados e discussões. O ambiente foi disponibilizado para que gestores possam explorá-los na busca por informações significativas sobre a evasão escolar no ensino superior.

Há que se considerar que a modelagem multidimensional desenvolvida até este momento da pesquisa atinge aos objetivos propostos, mas não é definitiva e nem acabada, isto é da própria natureza da tecnologia do BI de adequar-se à dinâmica do negócio da organização, mas, sobretudo, porque o uso e a exploração do conjunto de recursos que a tecnologia dispõe poderão despertar nos gestores o interesse na evolução e a continuidade pelo uso do BI com a exploração por novas funcionalidades, embora esta conclusão só possa ser verificada com o passar do tempo de uso da tecnologia. Neste aspecto merece atenção para a metodologia usada no processo de desenvolvimento e implantação da ferramenta que poderá ser adotada como processo pela organização tanto para ampliar e aprofundar as análises sobre a evasão, bem como para aplicar a outros assuntos relacionados a outros dados institucionais.

Um aspecto importante desta pesquisa, para além dos objetivos propostos, mas provocado a partir da busca pelo alcance destes objetivos, é o aprofundamento da discussão a respeito do entendimento sobre o conceito de evasão escolar, que se mostrou necessário para permitir avançar na construção do modelo multidimensional. Neste ponto a pesquisa revelou a ausência de uma definição formal para a questão, entretanto é possível concluir que a instituição trata o tema de forma peculiar para as diferentes modalidades de ensino ofertadas. No ensino superior, no IFBA, há um entendimento de que o aluno evadido é aquele que interrompe seu ciclo de matrículas ao longo de um período de forma temporária ou definitiva antes de integralizar o curso, sem considerar a vida acadêmica deste aluno fora da instituição. Ainda que haja entendimentos diferentes por outras instituições e pesquisas realizadas, o estabelecimento deste conceito se mostrou fundamental para o desenvolvimento desta pesquisa e é neste entendimento que a construção do modelo multidimensional está amparado.

Os resultados reafirmaram ainda a efetividade do uso do BI para o tratamento e a consolidação de grandes volumes de dados históricos, em consonância com o desafio atual das TICs, apontado por Barbieri (2011) como sendo o de transformar dados em informações significativas para que possam ser úteis à organização. Em que pese o processo de institucionalização da solução extrapolar o período de desenvolvimento desta pesquisa, o retorno dos gestores quanto ao uso do ambiente OLAP proposto pôde revelar a adequação da tecnologia para apoiar no monitoramento e controle da evasão escolar no ensino superior no âmbito do IFBA campus Salvador.

Para tratar das principais dificuldades, pode se destacar o próprio o tema evasão escolar. Discutir evasão escolar, sobretudo no ensino superior, não foi uma questão fácil, dada sua heterogeneidade conceitual e mesmo a ausência de definições formais por parte do IFBA. Outro fator importante foi a distância geográfica do pesquisador para com o seu lócus da pesquisa, este que atua no *campus* Jacobina e optou, em comum acordo com demais sujeitos da pesquisa, por pesquisar dados do *campus* Salvador e com gestores alocados na Reitoria que fica situada na cidade do Salvador - Bahia. Em alguns momentos da pesquisa combinar agendas de trabalho para realização das entrevistas e reuniões se mostrou um tanto desafiador.

Outras dificuldades foram encontradas durante o processo de entendimento sobre os dados acadêmicos mantidos pelo SICAD. Isto porque o sistema em estudo não possui documentação técnica, como, por exemplo, o dicionário de dados, um importante artefato para permitir compreender o significado dos dados por ele mantido. Para isto precisou contar com o apoio da equipe de TI do IFBA que é responsável pela manutenção do referido sistema. Além disso, a ausência de algumas informações atualizadas inviabilizou a análise de algumas das dimensões existentes no modelo desenvolvido, como foi o caso de alguns dados socioeconômicos e informações sobre o endereço dos alunos o qual não permitiu inferir sobre o deslocamento realizado pelo aluno no trajeto casa/trabalho/escola/casa, este que foi um dos fatores apontados pelos gestores entrevistados como sendo determinante para aumento da evasão no ensino superior, ainda assim o modelo desenvolvido contemplou tais dimensões que poderão ser analisadas futuramente caso a instituição opte por atualizar estes dados. Ainda no que diz respeito ao SICAD e sua base de dados, uma série de inconsistências e problemas técnicos que colocam em risco a qualidade dos dados mantidos foram encontradas, como a falta de integridade referencial e uniformização de informações e redundância de alguns dados. Algumas destas dificuldades puderam ser mitigadas durante o desenvolvimento da solução de BI, através do processo de ETL, entretanto algumas exigirão ações futuras e intervenções diretas por parte do IFBA, sobretudo aquelas relacionadas à criação de mecanismos de atualização freqüente de informações cadastrais e socioeconômicas.

A revelação de problemas de inconsistências em bases de dados e na forma como os dados são alimentados num sistema informatizado transacional é algo comum em projetos de BI (BARBIERI, 2011). Embora este não tenha sido o objetivo desta pesquisa, a revelação destes problemas e dificuldades se apresenta como mais um recurso do BI para apoiar a instituição na tomada de decisões, que neste ponto sugere a adoção de medidas para melhoria da qualidade dos seus dados.

Cabe destacar que diante das questões de inconsistências encontradas na base de dados do SICAD, ausência de algumas informações identificadas durante as fases de mapeamento e desenvolvimento do modelo, bem como possíveis falhas na alimentação dos dados no SICAD, isso pode distorcer alguns dos indicadores apresentados nesta pesquisa a partir da extração dos dados do ambiente analítico pelo pesquisador, ainda que os cálculos tenham sido validados. Porém estas eventuais distorções só poderão ser identificadas com detalhes, e corrigidas quando possíveis, a partir da efetiva institucionalização, utilização e exploração do ambiente pelos gestores, o que extrapola o período de apresentação desta pesquisa.

O desenvolvimento desta pesquisa elucidou um conjunto de ações futuras que podem ser adotadas pela instituição e por pesquisadores com o objetivo de contribuir no controle da evasão escolar, o que evidencia o uso do *Business Intelligence* como uma ferramenta gerencial em potencial para apoiar na gestão acadêmica.

Além das ações anteriormente sugeridas para a melhoria na qualidade dos dados armazenados no SICAD, apresentam-se ainda como perspectivas para projetos futuros:

- Criação de um programa institucional permanente de combate à evasão, com planejamento de ações sistemáticas que permitam acompanhamento de resultados e coleta de experiências bem-sucedidas;
- Incluir nos questionários de Avaliação Institucional questões específicas que busquem apoiar na identificação dos fatores que levam o aluno a evadirem, conforme sugere Polydoro (2005), ao afirmar que a evasão escolar deveria ter um destaque maior nos instrumentos usados pela Avaliação Institucional pelo fato da evasão acumular prejuízos sociais, acadêmicos e econômicos às instituições de ensino. Ainda neste sentido, ampliar a aplicação dos referidos questionários aos alunos evadidos
- Utilizar fontes de dados dos processos seletivos internos do IFBA, do MEC para alunos ingressos a partir do Sisu, das bases de dados de sistemas utilizados pelos programas de assistência estudantil, como forma de permitir cruzamentos de dados e obtenção de informações cadastrais e socioeconômicas declaradas mais detalhadas e atuais.
- Utilizar técnicas de *Balanced ScoreCard* - BSC, termo criado em 1990 por Robert Kaplan e David Norton para descrever uma metodologia para gestão de

desempenho institucional (BARIBIERI, 2011). O BSC tem sido amplamente usado de forma integrada a projetos de BI, este que fornece os indicadores para permitir acompanhar as metas estabelecidas pelo BSC. Esta proposta tem como fundamento a definição de metas sobre a evasão escolar para ampliar a utilização de painéis de monitoramento, com o uso dos KPIs e outros recursos gráficos, já propostos no ambiente analítico desenvolvido.

- Ampliar a utilização do BI sobre a evasão escolar para as demais modalidades de ensino oferecidas no IFBA de maneira a criar um ambiente centralizado que poderá servir de fonte de informações para a criação de um Observatório da Evasão Escolar.
- Criar grupos de estudos com vistas a discutir e revisar a fórmula utilizada para cálculo da evasão escolar, bem como a possibilidade de utilização de novos métodos para este cálculo para que este possa refletir sempre a realidade do IFBA e que se possa ampliar a análise de outras métricas.
- Aplicar a solução tecnológica desenvolvida a outras instituições de ensino, assim como sua metodologia, já que o modelo multidimensional construído trata de assuntos comuns à maioria destas instituições, precisando, neste caso, apenas de ajustes e adaptações no processo de Extração, Transformação e Carga da base de origem dos dados.

Não obstante ao que se apresenta como perspectivas futuras vale considerar que o mais evidente e necessário de todos é o aprofundamento dos estudos quanto ao tão complexo fenômeno da evasão escolar, no âmbito ensino superior do IFBA, verificando o impacto social e econômico que este pode causar no sistema educacional, mesmo que isto pareça uma perspectiva remota diante do percurso que se precisa cumprir.

Por fim, vale lembrar que esta pesquisa, ainda que sua ênfase esteja voltada para o uso de aparatos tecnológicos, propondo o uso do *Business Intelligence* para apoiar no controle da evasão no ensino superior, se apresenta com um caráter provocador para que se possa avançar nos estudos acerca deste tão preocupante tema.

REFERÊNCIAS

BAGGI, Cristina A. Santos; LOPES, Doraci Alves. **Evasão e Avaliação Institucional no Ensino Superior: uma discussão bibliográfica**. Revista da Avaliação da Educação Superior (Campinas). Sorocaba, SP, v. 16, n. 2, p. 355-374, jul. 2011. Disponível em <<http://www.scielo.br/pdf/aval/v16n2/a07v16n2.pdf>>. Acesso em: 30 Set 2015.

BALTZAN, Paige; PHILLIPS, Amy. **Sistemas de Informação**. Porto Alegre: McGRAW-Hill, 2012. e-book. Disponível em <<http://books.google.com.br/books?id=NJkR83DSkPYC&printsec=frontcover#v=onepage&q&f=false>>. Acesso em: 17 Jun 2014.

BARBIERI, C. **BI – Business Intelligence – Modelagem & Tecnologia**. Rio de Janeiro: Excel Books do Brasil Editora, 2001.

BARBIERI, Carlos. **Bi2 - Business Intelligence - Modelagem e Qualidade**. São Paulo: Elsevier Editora. 2011.

BARDAGI, Marúcia Patta. **Evasão e comportamento vocacional de universitários: estudos sobre o desenvolvimento de carreira na graduação**. Tese (Doutorado). Instituto de Psicologia, Universidade Federal do Rio Grande do Sul, 2007.

BERTA, Dora-Anca. **Business Intelligence in Education**. The 8 th International Scientific Conference e Learning and software for Education Bucharest, April 26-27, 2012. Disponível em <http://www.ceeol.com/aspx/getdocument.aspx?logid=5&id=8f8025bd254a4069b30c10f081db8c20>. Acesso em: 22 Fev 2014.

BRASIL. Casa Civil. **Lei Aprova o Plano Nacional de Educação – PNE: nº 13.005**. Brasília: 2014.

_____. Casa Civil. **Lei de criação dos Institutos Federais de Educação: nº 11.892**. Brasília: 2008

_____. Casa Civil. **Lei de ingresso nas universidades federais e nas instituições federais de ensino técnico de nível médio e dá outras providências: nº 12.711**. Brasília: 2012

_____. Senado Federal. **Lei de Diretrizes e Bases da Educação Nacional: nº 9394/96**. Brasília: 1996.

CARNEIRO, Margareth Fabiola dos Santos. **Gestão Pública**. Rio de Janeiro: Brasport Editora, 2010.

CIUPAK, L.; BOSCARIOLI, C.; CATARINO, M. **Análise do uso de Tecnologias de Business Intelligence como facilitadoras à Gestão Universitária**. Brazilian Journal of Information Science - BJIS, Marília (SP), v.7, n. Especial, p.47-69, 1º Sem. 2013. Disponível em: <<http://www2.marilia.Unesp.br/revistas/index.php/bjis/index>>. Acesso em: 27 Mar 2014.

COELHO, F. et tal. **Business Intelligence: Suporte à Gestão do Desempenho Científico.** XXXII ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO: Desenvolvimento Sustentável e Responsabilidade Social: As Contribuições da Engenharia de Produção. Bento Gonçalves, RS, Brasil, 15 a 18 de outubro de 2012. Disponível em <http://www.abepro.org.br/biblioteca/ENEGETP2012_TN_STO_164_958_20833.pdf> Acesso em: 27 Mar 2014.

COSTA, Francisco de Assis. **ATLAS HISTÓRICO DE CIDADES: a cidade como objeto de investigação.** Cadernos PPG-AU/UFBA. Vol. 8, N. 2. 2009. Disponível em <<http://www.portalseer.ufba.br/index.php/ppgau/article/view/2633/1864>>. Acesso em 20 out. 2014

CAUCHICK MIGUEL, Paulo A. **Metodologia de Pesquisa em Engenharia de Produção e Gestão de Operações.** 2. ed. Rio de Janeiro: Campus-Elsevier, 2012.

DAYCHOUM, M. **40 Ferramentas e Técnicas de Gerenciamento.** 1ª Edição. ed. [S.l.]: Brasport, 2007.

FERREIRA, Antonio Joaquim Salgado dos Santos Costa. **Uma solução de Business Intelligence como con-tributo para a melhoria do processo de tomada de decisão na gestão financeira.** Disponível em: <https://www.repository.utl.pt/bitstream/10400.5/5876/2/MPA_Adm_Publica_Antonio_Ferreira.pdf>. Acesso em: 15 Out 2013.

FIALHO, Marília Gabriella Duarte. **A Evasão Escolar e a Gestão Universitária: O caso da Universidade Federal da Paraíba.** Dissertação de Mestrado, UFPB, João Pessoa, 2014. Disponível em <http://tede.biblioteca.ufpb.br:8080/bitstream/tede/5920/1/arquivototal.pdf>. Acesso em: 30 Ago. 2015.

FORTUNATI, José. **Gestão da Educação Pública: Caminhos e desafios.** Porto Alegre: Artmed, 2008.

FRANÇA, Ivo C.; COSTA, Luciana T. D.; OLIVEIRA, Lúcio J. N. Uso de Ferramentas BI para Apoio à Gestão na Educação. In: RANGEL, Sheila (Org.). **Abordagens gerenciais no mundo contemporâneo.** Salvador: Quarteto, 2013, p. 77-95.

FREITAS, Kátia Siqueira de. **Alguns Estudos sobre a Evasão e Persistência de Estudantes.** Ecos – Revista Científica. São Paulo, v. II, nº I, p. 247-264, 2009. Disponível em http://www.uninove.br/PDFs/Publicacoes/eccos/eccos_v11n1/eccosv11n1_3i1062.pdf. Acesso em 13 Set. 2015.

GAIOSO, Natalicia Pacheco de Lacerda. **O fenômeno da evasão escolar na educação superior no Brasil.** 2005. 75 f. Dissertação (Mestrado em Educação) – Programa de Pós-Graduação em Educação da Universidade Católica de Brasília, Brasília, 2005.

GENEROSO, P; TAMBORLIN, N.; DALFOVO, O. **O uso de Inteligência Competitiva Integrada na Gestão da Tecnologia da Informação como Business Intelligence num Instituto Federal de Educação.** 10th International Conference on Information Systems and Technology Management – CONTECSI

IFBA. **Organização Didática do Ensino Técnico.** Disponível em <http://portal.ifba.edu.br/attachments/027_organizacao_tecnicol.pdf>. Acesso em 25 Set. 2014

_____. **Norma Acadêmica do Ensino Superior 2007.** Disponível em <http://www.portal.ifba.edu.br/attachments/027_normas_superior.pdf>. Acesso em 25 Set. 2014

_____. **Relatórios de Gestão.** Disponível em <<http://www.portal.ifba.edu.br/institucional/relatorios-de-gestao-do-ifba.html>>. Acesso em 07 Out. 2014.

_____. **Plano de Desenvolvimento Institucional do IFBA 2014 – 2018.** Disponível em <http://portal.ifba.edu.br/component/option,com_phocadownload/Itemid,196/download,5673/id,1/view,category/> . Acesso em 09 Mar. 2015.

INMON, W.H., **Building the Data Warehouse.** 4ª edição, Ed. John Wiley & Sons, New York. 2005.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA (INEP). **Censo da Educação Superior 2013.** Disponível em <http://download.inep.gov.br/educacao_superior/censo_superior/documentos/2014/principais_resultados_censo_2013.xls>. Acesso em 20 set. 2014.

_____. **Censo da Educação Básica 2013.** Disponível em <http://download.inep.gov.br/educacao_basica/portal_ideb/planilhas_para_download/2014/sintese_ideb_brasil.xlsx>. Acesso em 20 set. 2014

_____. **Censo Escolar 2008.** Resumo Técnico 2008. Disponível em <http://download.inep.gov.br/download/censo/2008/resumo_tecnico_2008_15_12_09.pdf>. Acesso em 20 set. 2014

KIMBALL, R.;ROSS, M. **The Data Warehouse Toolkit.** 2ª edição, Ed. John Wiley & Sons, New York. 2002.

LÉVY, Pierre. **As tecnologias da Inteligência** – O futuro do pensamento na era da informática. São Paulo: Editora 34, 2004.

LIMA JR, Arnaud S.; MAGALHÃES, André R.;ALMEIDA, Djalma F. Subsídios para a Gestão Pública a partir de uma abordagem crítica da Tecnologia e da Modelagem Cognitiva. In: RANGEL, Sheila (Org.). **Abordagens gerenciais no mundo contemporâneo.** Salvador: Quarteto, 2013, p. 207-226.

LIMA JR, Arnaud Soares de. **Educação e NTIC: do Pensamento Dialético ao Pensamento Virtual.** 1997. Disponível em: <<http://www.lynn.pro.br/pdf/educatec/limajunior.pdf>>. Acesso em: 30 Out 2013.

LIMA, Vinícios Mariano; BOSCARIOLI, Clodis; LAGO, Sandra Mara Stocker. **Um Estudo de Caso Sobre o Gerenciamento do Relacionamento com o Cliente em uma Pequena**

Empresa. Disponível em: <<http://www.admpg.com.br/2008/cadastro/artigos/temp/74.pdf>>. Acesso em: 21 Ago 2013.

LOBO, Maria Beatriz de Carvalho Melo. **Panorama da Evasão no Ensino Superior Brasileiro: Aspectos Gerais das Causas e Soluções.** Associação Brasileira de Mantenedoras do Ensino Superior. Cadernos nº 25. Dezembro de 2012. Disponível em: <http://www.institutolobo.org.br/imagens/pdf/artigos/art_087.pdf>. Acesso em: 07 set. 2015.

LOPES, Grimaldo Oliveira. **GESMOODLE – Ferramenta de Mediação Tecnológica no acompanhamento do processo de aprendizagem dos alunos em cursos de graduação a distância da Universidade do Estado da Bahia (UNEB).** Dissertação de Mestrado, UNEB, Salvador, 2013. Disponível em <http://www.uneb.br/gestec/files/2013/06/Disserta%C3%A7%C3%A3o-Final-Grimaldo1.pdf>. Acesso em: 01 Out 2013.

MAXIMIANO, Antônio Cesar. Amaru. **Introdução à Administração.** 8ª Ed. São Paulo: Atlas, 2011.

MEC / SESU. Secretaria de Educação Superior / Ministério da Educação. **Comissão Especial de Estudos sobre a Evasão nas Universidades Públicas Brasileiras.** Brasília, 1996/1997 Disponível em: <<http://www.dominiopublico.gov.br/download/texto/me001613.pdf>>. Acesso em 20 set. 2014

MEIRELLES, Fernando de Souza. **Informática: novas aplicações com microcomputadores.** 2ª Ed. São Paulo: Pearson Education do Brasil, 1994.

Ministério da Educação (MEC). **Relatório de Gestão do exercício 2012.** Disponível em <http://portal.mec.gov.br/index.php?option=com_docman&task=doc_download&gid=15144&Itemid=>>. Acesso em 20 set. 2014.

Ministério da Educação (MEC). **REUNI - Reestruturação e Expansão das Universidades Federais.** Diretrizes Gerais. 2007. Disponível em <<http://portal.mec.gov.br/sesu/arquivos/pdf/diretrizesreuni.pdf>>. Acesso em 20 set. 2014.

MORIN, Edgar. **Os sete saberes necessários à Educação do Futuro.** 2ª Ed. Brasília: Cortes/UNESCO, 2000. 19 - 46 p.

MOTA, Maria Sebastiana Gomes, PEREIRA, Francisca Elisa de Lima. **Desenvolvimento e Aprendizagem.** Processo de Construção de Conhecimento e Desenvolvimento Mental do Indivíduo. Disponível em: <http://portal.mec.gov.br/setec/arquivos/pdf3/tcc_desenvolvimento.pdf>. Acesso em: 15 Set 2013.

PLATT NETO, Orion Augusto; CRUZ, Flávia; PFITSCHER, Eliesete D. **Utilização de Metas de Desempenho Ligadas à Taxa de Evasão Escolar nas Universidades Públicas.** Revista de Educação e Pesquisa em Contabilidade. v. 2, n. 2, art. 4, p. 54-74, maio/ago. 2008. Disponível em <<http://www.repec.org.br/index.php/repec/article/view/25/27>>. Acesso em: 30 Set 2015.

POLYDORO, Soely Aparecida Jorge. **Evasão em uma instituição de ensino superior: desafios para a psicologia escolar**. 1995. 145 p. Dissertação (Mestrado em Psicologia) – Departamento de Pós-Graduação em Psicologia da Pontifícia Universidade Católica de Campinas, Campinas, 1995.

_____. **O trancamento de matrícula na trajetória acadêmica no universitário: condições de saída e de retorno à instituição**. 2000. 167 p. Tese (Doutorado em Educação) – Universidade Estadual de Campinas, Campinas, 2000.

PRADO, Iara Glória de Areias. **LDB e Políticas de Correção de Fluxo Escolar**. Disponível em <<http://emaberto.inep.gov.br/index.php/emaberto/article/viewFile/1074/976>> . Acesso em 10 de out. 2014.

PRESSMAN, Roger S. **Engenharia de Software**. 3 ed. São Paulo: Person Makron Books, 2006.

PRIMAK, Fábio V. da Silva. **Decisões com B.I. (Business Intelligence)**. Rio de Janeiro: Editora Ciência Moderna, 2008.

PROGRAMA DAS NAÇÕES UNIDAS PARA O DESENVOLVIMENTO (PNUD). **Relatório de Desenvolvimento Humano 2010**. Disponível em http://www.pnud.org.br/HDR/arquivos/RDHglobais/PNUD_HDR_2010.pdf. Acesso em 20 set. 2014

_____. **Relatório de Desenvolvimento Humano 2013**. Disponível em <http://www.pnud.org.br/HDR/arquivos/RDHglobais/hdr2013_portuguese.pdf>. Acesso em 20 set. 2014

REIS, E. S.; ANGELONI, M. T.; SERRA, F. R. **Business Intelligence como Tecnologia de Suporte a Definição de Estratégias para a Melhoria da Qualidade de Ensino**. Informação & Sociedade. João Pessoa, v.20, n.3, p. 157-167, set./dez. 2010. Publicado em <http://www.ies.ufpb.br/ojs/index.php/ies/article/view/3783/4813>. Acesso em: 10 Mar 2014.

RIGO, S. J; CAMBRUZZI, W.; BARBOSA, J.L.V.;CAZELLA, S. C. **Aplicações de Mineração de Dados Educacionais e Learning Analytics com foco na evasão escolar: oportunidades e desafios**. Revista Brasileira de Informática na Educação, 2014, 22, 2317-6121.

SILVA, Cristiana I. M. **Protótipo de DataWarehouse a aplicar numa Escola Secundária/3 de Ensino**. Dissertação de Mestrado, Instituto Politécnico Castelo Branco, Castelo Branco, Portugal, 2011. Disponível em <http://repositorio.ipcb.pt/bitstream/10400.11/945/1/TeseCD%5B1%5D.pdf>. Acesso em: 01 Abr 2013.

SILVA, Glauco Peres da. **Análise de Evasão no Ensino Superior: uma proposta de diagnóstico de seus determinantes**. Avaliação, Campinas; Sorocaba, SP, v. 18, n. 2, p. 311-333, jul. 2013. Disponível em <http://www.scielo.br/pdf/aval/v18n2/04.pdf>. Acesso em 13 Set. 2015.

SILVA FILHO, Roberto L.L.; MOTEJUNAS, Paulo R.; HIPÓLITO, Oscar; LOBO, Maria B. C. M. **A Evasão no Ensino Superior Brasileiro**. Cadernos de Pesquisa, v. 37, n. 132, p. 641-659, set./dez. 2007. Disponível em <<http://www.scielo.br/pdf/cp/v37n132/a0737132.pdf>>. Acesso em 11 out. 2014.

TENÓRIO, Fernando Guilherme. **Tecnologia da Informação transformando as organizações e o trabalho**. Rio de Janeiro: FVG, 2007.

TERRA, José Cláudio C. **Gestão do conhecimento: o grande desafio empresarial: uma abordagem baseada no aprendizado e na criatividade**. São Paulo: Negócio Editora, 2000.

THIOLLENT, Michel. **Metodologia da pesquisa - ação**. 18. ed. São Paulo: Cortez, 2011.

TRIPP, David. **Pesquisa-ação: uma introdução metodológica**. Educação e Pesquisa. São Paulo: dez 2005, v. 31, n. 3, p. 443-466.

APÊNDICES

APÊNDICE A – Termo de Livre Consentimento assinado pelos participantes da entrevista

Programa de Mestrado Profissional Gestão e Tecnologias Aplicadas à Educação (GESTEC)

Proposta de Trabalho: Construir um modelo de dados multidimensional, a partir da base de dados histórica do sistema acadêmico do IFBA, para o mapeamento das informações acerca da evasão escolar no ensino superior, usando a tecnologia de *Business Intelligence*.

OBJETIVO DA ENTREVISTA: Levantamento de informações acerca da evasão escolar e o contato com os profissionais envolvidos na gestão das informações acadêmicas.

Prezado(a) Entrevistado(a):

Convido-o(a) a participar da presente pesquisa exploratória que se constitui na construção de um trabalho de conclusão de curso de Mestrado Profissional em Educação. Neste momento, estamos levantando o que o profissional compreende sobre a evasão escolar na sua instituição de trabalho e quais instrumentos e métodos usados para registro e medição de informações sobre a evasão escolar, em especial no ensino superior

Informamos que os dados obtidos com a pesquisa serão utilizados somente para o âmbito acadêmico e toda informação será tratada de forma que assegure o total sigilo. Desde já, agradecemos a atenção dispensada.

TERMO DE CONSENTIMENTO (do pesquisado)

Tendo sido esclarecido o objetivo da pesquisa, eu na figura de pesquisado, declaro que:

() Aceito participar da pesquisa e responderei ao questionário, como também permito a gravação da minha entrevista.

() Aceito participar da pesquisa e responderei ao questionário, mas não permito a gravação da minha entrevista.

() Não aceito participar

Nome do (a) Pesquisado (a): _____

Instituição do (a) Pesquisado (a): _____

Assinatura do (a) Pesquisado (a): _____

Assinatura do Pesquisador: _____

Dados e Contato do Pesquisador-Mestrando do Curso de Mestrado Profissional Gestão e Tecnologias Aplicadas à Educação (GESTEC) :

Ivo Chaves de França / Email: ivochaves@gmail.com

Orientador: Prof. Dr. André Ricardo Magalhães

APENDICE B - Roteiro para entrevista com os gestores e técnicos em educação

Questões norteadoras da entrevista sobre evasão escolar:

- 1) O que você entende por evasão escolar?
- 2) Na sua opinião, qual(is) fator(es) é (são) determinante(s) para evasão?
- 3) Conhece alguma ação na sua instituição para minimizar os índices de evasão? Sim () Não (). Em caso positivo, descreva.
- 4) Conhece os instrumentos de medida dos índices de evasão da instituição em que trabalha? Sim () Não (). Em caso positivo, descreva.
- 5) Se conhece, concorda com os métodos usados para medir este índice? Sim () Não ().
Porque?
- 6) Se conhece e não concorda, quais os métodos ou elementos deveriam ser analisados para considerar um aluno como evadido?
7. O que você indicaria como alternativa para redução da evasão escolar?
8. Qual a relação que você entende existir entre Evasão, Abandono e Desistência?

APÊNDICE D – Questionário de validação do ambiente analítico do BI

Nome: _____

Local/Departamento de Trabalho: _____

Observação: Para orientar no uso da ferramenta, caso necessário, utilize o Manual Básico disponibilizado.

Para as afirmações de 01 a 06, classifique-as numa escala de 01 a 05, onde:

01 - Discordo Completamente

02 - Discordo

03 - Neutro

04 - Concordo

05 - Concordo Plenamente

01 - O Ambiente disponível é de fácil uso e se mostra adequado para uso pelos gestores educacionais.

02 - Durante a exploração do ambiente com a geração de alguns resultados possíveis, o ambiente apresentou uma performance satisfatória

03 - Os formatos dos resultados gerados (planilhas, gráficos e painéis) pelas consultas realizadas através dos cubos disponíveis são satisfatórios.

04 - Consegui extrair resultados de forma simples e eficiente.

05 - Considero o ambiente proposto, bem como os formatos com que se pode extrair indicadores, adequados para apoiar no monitoramento e controle da evasão escolar.

06 - No Painel de Monitoramento proposto como exemplo, com recursos visuais gráficos em formatos variados, KPIs e planilhas permitirá acompanhar de forma centralizada indicadores sobre a evasão escolar.

07 - Já conhecia este ambiente tecnológico de Business Intelligence? () Sim () Não

08 - Quais as dificuldades encontradas no uso da ferramenta? Quais melhorias você indicaria?

09 - Qual sua opinião sobre a interface do ambiente OLAP?

10 - Qual a sua opinião sobre o Painel de Monitoramento proposto?

APÊNDICE E – Manual básico de uso do ambiente analítico

MANUAL BÁSICO DE OPERAÇÕES DO AMBIENTE ANALÍTICO DA EVASÃO ESCOLAR DO ENSINO SUPERIOR

Versão 1.0

Este projeto é resultado da pesquisa de mestrado “O Business Intelligence como ferramenta de apoio ao controle da evasão escolar no IFBA campus Salvador”, desenvolvida através do Mestrado Profissional em Gestão e Tecnologia Aplicadas a Educação - GESTEC pela Universidade do Estado da Bahia – UNEB

Outubro/2015

SUMÁRIO

1 APRESENTAÇÃO	119
2 ACESSANDO A FERRAMENTA	119
3 CONHECENDO O AMBIENTE PENTAHO.....	120
4 CONSTRUINDO INDICADORES ON LINE ATRAVÉS DO SAIKU ANALYTICS	120
5 ACESSANDO RELATÓRIOS PREDEFINIDOS.....	123
6 ACESSANDO O PAINEL DE MONITORAMENTO.....	123
7 OBSERVAÇÕES	124

1 APRESENTAÇÃO

Este trabalho é resultado da pesquisa “O Business Intelligence como ferramenta de apoio ao controle da evasão escolar no IFBA campus Salvador”. Teve como objetivo construir um modelo de dados multidimensional, a partir da base de dados histórica do sistema acadêmico do IFBA (SICAD), para o mapeamento das informações acerca da evasão escolar no ensino superior do campus Salvador. A pesquisa contou com o apoio da PROEN e DGTI e da GRA3 do campus Salvador.

Todo o trabalho de consolidação e disponibilização das informações foi realizado no ambiente *Pentaho*, uma plataforma de *Business Intelligence* construída sob a licença de software livre.

É importante destacar que todos os indicadores disponibilizados no ambiente foram calculados a partir do método de cálculo da **Evasão Anual**, conforme relatório de gestão publicado anualmente pelo IFBA.

Os dados disponibilizados no ambiente *Pentaho* estão organizados em forma de Cubos. Os Cubos, num projeto de *Business Intelligence*, são visões sobre um determinado conjunto de informações que podem ser agrupadas de acordo com o nível de detalhe que se deseja analisar. Estas informações estão agrupadas em dimensões (perspectivas sobre as quais se pode analisar um determinado fato) e fatos/métricas (ocorrências que se deseja medir). Desta forma, foram construídos 03 Cubos conforme segue:

Cubo Evasão Anual: Apresenta a quantidade de evadidos por ano e permite o cruzamento de dados mais granulares, podendo chegar a analisar métricas por aluno e informações da matrícula.

Cubo Evadidos por Matriculados: Apresenta métricas de quantidade de evadidos e matriculados por Ano, Campus e Curso. Permite a análise da variação anual dos matriculados e dos evadidos. Este cubo apresenta uma visão mais gerencial com dados menos granulares e mais consolidados.

Cubo Nota e Frequência: Apresenta as métricas com médias de notas e de percentual de frequência dos alunos evadidos. Estas métricas estão agrupadas por ano, campus, curso e disciplina.

2 ACESSANDO A FERRAMENTA

Inicialmente a solução está em fase de teste e validação dos dados. Para utilizar é necessário acessá-la via Intranet (verificar com a equipe de TI), através de um navegador Web. Em seguida é apresentada a tela conforme figura a seguir.



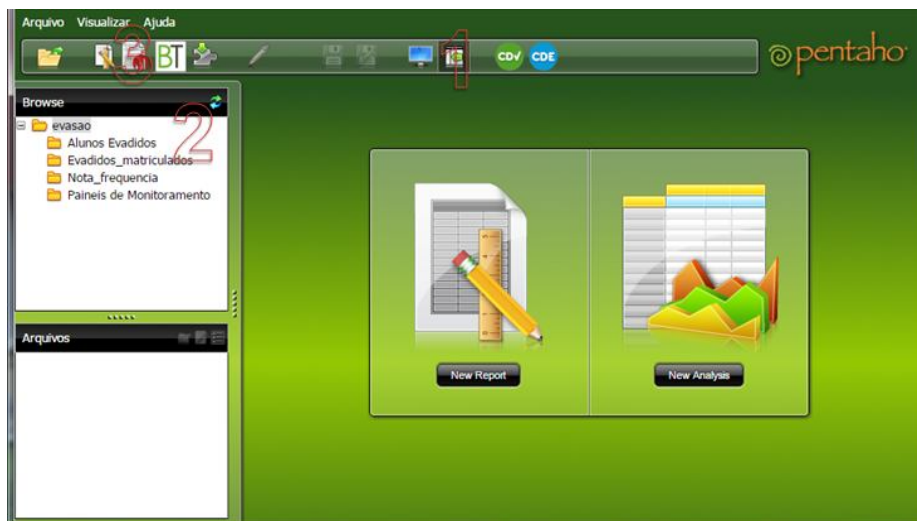
Para ter acesso é necessário usuário senha. Durante o período de aprendizagem sobre a ferramenta serão usados os dados a seguir:

User Name: proen

Password: proen

3 CONHECENDO O AMBIENTE PENTAHO

Após o login o usuário é direcionado para a tela a seguir. Embora a ferramenta ofereça um conjunto muito maior de funcionalidades, neste momento serão apresentados os itens 1,2 e 3 destacados na figura abaixo.



Item 1 - O botão permite exibir ou ocultar a barra de navegação de arquivos;


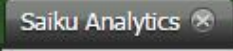
Item 2 - Barra vertical de navegação de arquivos. Os arquivos apresentados nesta barra são relatórios e gráficos predefinidos, construídos pelos usuários e salvos, permitindo a reutilização sempre que necessário.

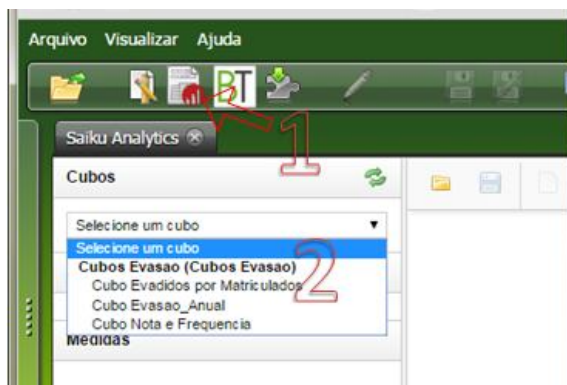
Item 3 - Botão para iniciar a construção de relatórios analíticos *On Line*, a partir da ferramenta *Saiku Analytics*.

Na barra de navegação, conforme **Item 2**, foram construídos alguns relatórios e gráficos predefinidos, além de um ambiente de monitoramento de indicadores. Com estes relatórios, gráficos e ambiente de monitoramento previamente construídos buscou-se apresentar de forma geral parte do conjunto de informações que podem ser exploradas a partir dos modelos desenvolvidos ao longo da pesquisa.

4 CONSTRUINDO INDICADORES ON LINE ATRAVÉS DO SAIKU ANALYTICS


Para construir indicadores para serem apresentados em formatos tabulares ou de gráficos será usado o *plugin* Saiku Analytics associado ao projeto Pentaho, por oferecer um ambiente mais simples e prático do que a solução nativa que vem no *Pentaho*.

4.1 - Para iniciar, conforme figura a seguir, clique no ícone . Será exibida no ambiente de trabalho uma aba . Conforme número "2" apresentado na figura abaixo, selecione um dos cubos sobre o qual deseja construir seus indicadores.



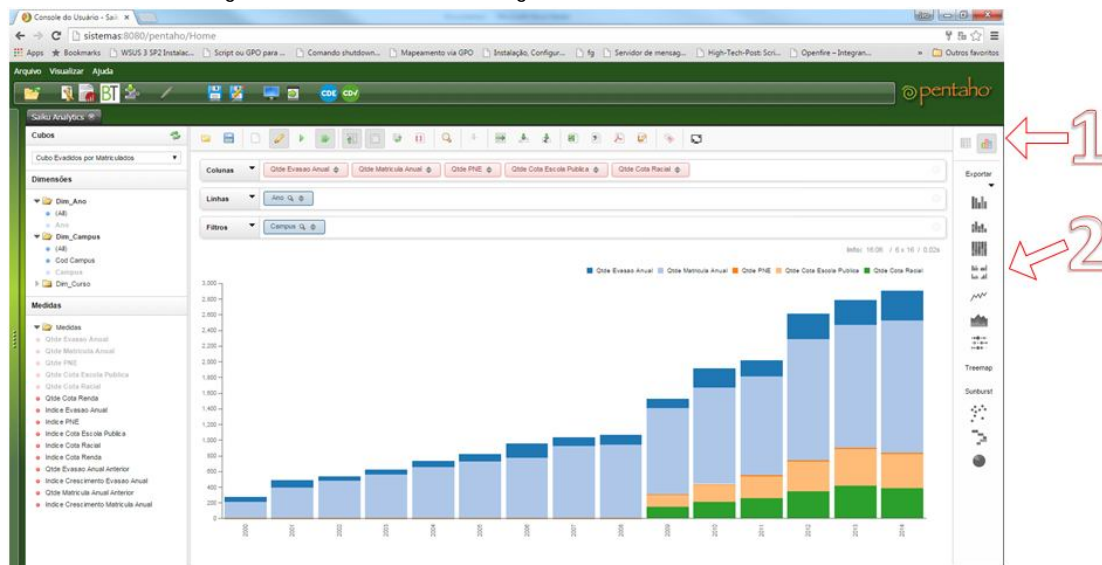
4.2 - Em seguida será exibida na coluna à esquerda, abaixo dos cubos, o conjunto de **Dimensões** e **Medidas**, disponíveis para o cubo selecionado. Na caixa à direita são definidas as linhas e as colunas que serão apresentadas, bem como as medidas/métricas e os filtros sobre os dados que serão apresentados. Com o mouse, o usuário deve **arrastar** para as linhas, colunas e filtros as Dimensões e Medidas que desejam que sejam apresentadas.

Ano	Qtde Evasao Anual	Qtde Matrícula Anual	Qtde PNE	Qtde Cota Escola Publica	Qtde Cota Racial
2000	84	212	0	0	0
2001	84	399	0	0	0
2002	57	484	0	0	0
2003	59	565	0	0	0
2004	72	661	0	0	0
2005	95	731	0	0	0
2006	181	773	0	1	0
2007	109	928	0	1	0
2008	122	945	0	1	0
2009	117	1.101	2	155	153
2010	239	1.235	8	220	212
2011	203	1.257	11	284	262
2012	322	1.547	18	379	349
2013	311	1.570	19	455	420
2014	380	1.954	17	440	355

4.3 - Caso deseje filtrar alguma informação em alguma das dimensões, deve-se clicar na lupa  situada ao lado do elemento que deseja filtro. Ao clicar na lupa será exibida uma tela conforme figura a seguir onde o usuário deverá mover para o lado direito os dados que deseja filtrar em seguida clicar em "OK"



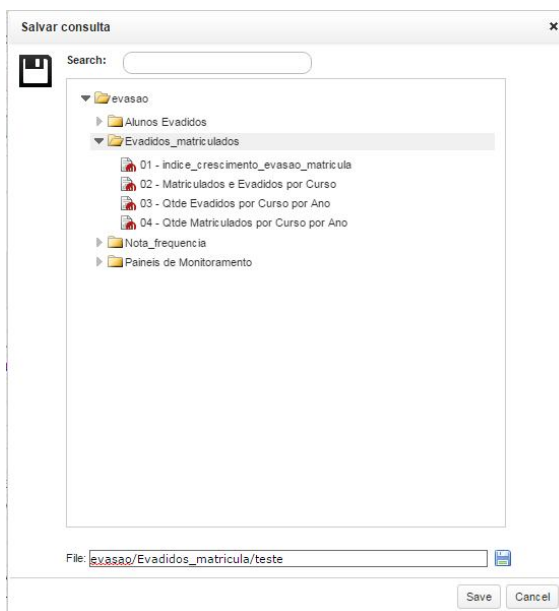
4.4 - Para apresentar um indicador em formato de gráfico, o usuário deverá clicar na imagem do gráfico, situada no canto direito superior da tela, conforme indicado na seta 1 da figura abaixo e em seguida escolher o formato do gráfico conforme seta 2 da figura.



4.5 - Caso o usuário pretenda reutilizar um relatório, poderá salvá-lo clicando no botão Salvar



, escolher a pasta onde deseja armazenar e dar um nome para arquivo, conforme figura a seguir. Para recuperar um arquivo salvo neste passo, deverá seguir os passos indicados na Seção 5.



5 ACESSANDO RELATÓRIOS PREDEFINIDOS

O sistema permite que os relatórios, planilhas, gráficos e outras informações extraídas sejam salvas para serem reutilizadas sempre que necessário. Isso ocorre porque muitos dos indicadores construídos são analisados com frequência, com informações mais atuais. Desta, através da barra de navegação ao lado direito é possível escolher um destes artefatos, conforme figura a seguir, dando clique duplo sobre ele. Para facilitar, foi criada uma pasta para cada um dos cubos como uma sugestão de organização destes artefatos por assunto.

Sexo	Faixa_etaria	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	
Feminino	0 a 14 anos																1	
	20 a 24 anos	1	8	5	3	5	7	12	18	13	21	21	35	40	66	77	129	
	25 a 29 anos	23	5	7	3	5	6	11	29	15	12	14	20	45	51	48	43	
	30 a 34 anos	2	1	3		1		1	9	2	2		10	27	12	17	19	
	35 a 39 anos	15	5	3	5	2		4	10	9	5	9	15	33	40	35	34	
	40 a 44 anos	3		1	3			2	3	1	2	2	1	13	13	14	9	
	45 a 49 anos		1									2	2	1	5	4	3	8
	50 a 54 anos			1				2	1	1	2		1	2	1	3	0	
	55 a 59 anos														3	2	2	
	60 a 64 anos															1	1	2
Não Informada		10		1		1				1								
Masculino	0 a 14 anos																2	
	20 a 24 anos		5	25	14	11	12	28	31	28	30	28	68	86	125	112	165	
	25 a 29 anos	18	6	12	8	8	9	20	41	26	28	31	71	87	124	89	122	
	30 a 34 anos	11	6	7	5	9	6	11	12	14	13	11	12	19	37	32	48	
	35 a 39 anos	34	7	12	8	9	9	4	20	23	30	20	47	62	86	96	107	
	40 a 44 anos	5	3	7	6	3	8	2	11	7	5	1	10	14	20	22	15	
	45 a 49 anos	1	7	3	1	1	4	4	4	7	2	5	2	10	21	8	15	
	50 a 54 anos	1	3		1	4	5	2	3	2		2	3	3	10	4	15	
	55 a 59 anos		6												3	3	3	4
	60 a 64 anos							3	1	4	2	1	1	3	3	3	4	
65 a 69 anos												1	1			1		
70 a 74 anos																1	1	

6 ACESSANDO O PAINEL DE MONITORAMENTO

Um dos recursos muito úteis num projeto de *Business Intelligence* é a construção de Painéis de Monitoramento, também chamados de *Dashboards*. Por eles é possível agrupar um conjunto dos principais

indicadores, de vários cubos simultâneos, para que o gestor possa acompanhá-los sem a necessidade de ficar criando os relatórios e ainda permite a visualização de vários indicadores numa só tela. Entre os tipos de informações disponíveis estão os Indicadores Chaves de Performance, também conhecidos como KPIs (acrônimo para *Key Performance Indicator*), muito utilizados para acompanhamento de metas.

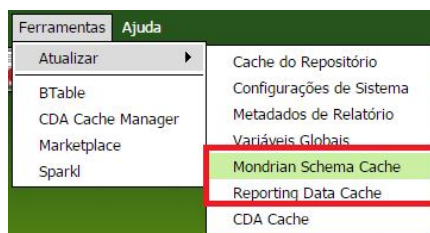
Foi construído um Painel de Monitoramento que pode ser acessado conforme figura abaixo. Neste painel, além de alguns gráficos e planilhas, foram criados 03 KPIs de exemplo com metas arbitradas, que acompanham as metas Quantidade de Evadidos, Quantidade de Matriculados e o Índice de Evasão.



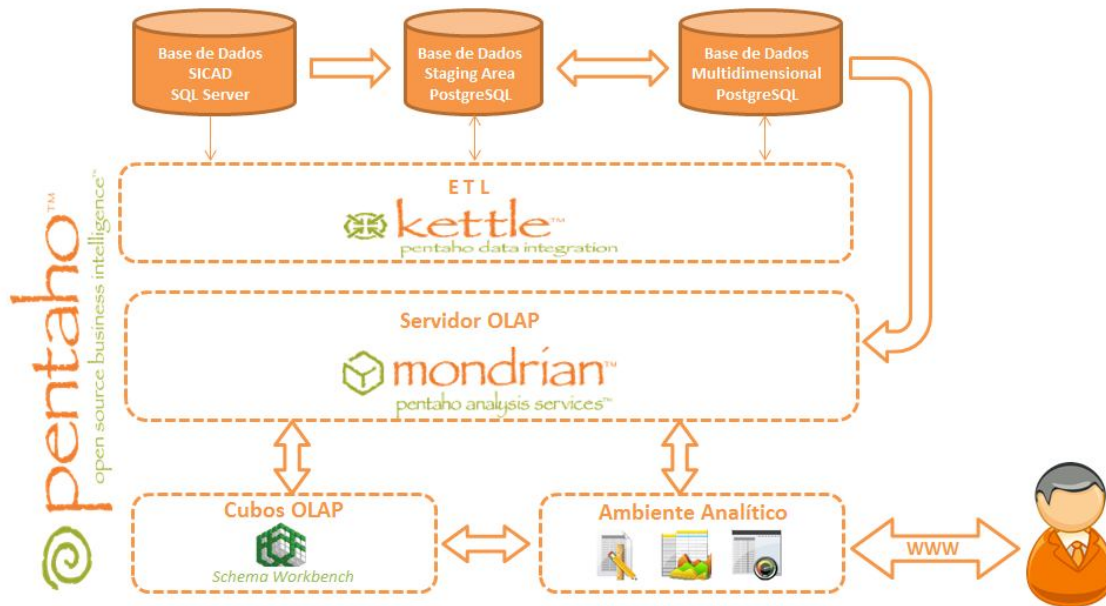
O uso de KPIs numa organização exige um planejamento estratégico com o estabelecimento das metas para depois serem acompanhadas. O recurso foi disponibilizado apenas como exemplo para demonstrar a capacidade da solução de *Business Intelligence*

7 OBSERVAÇÕES

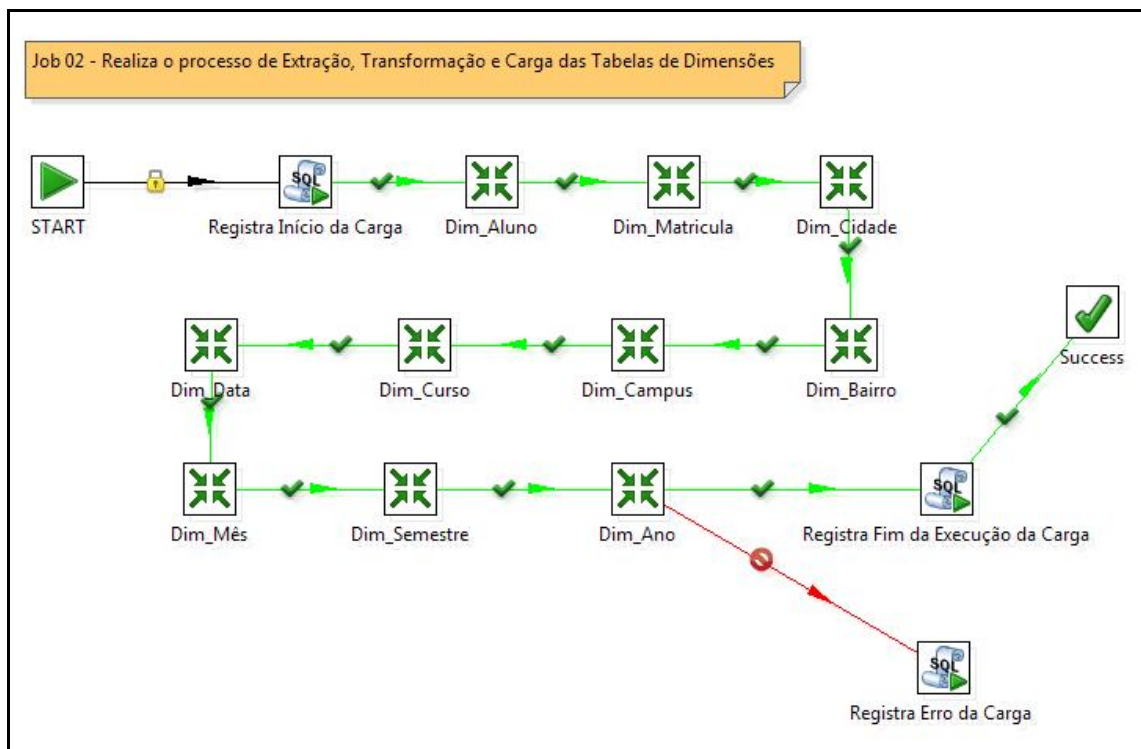
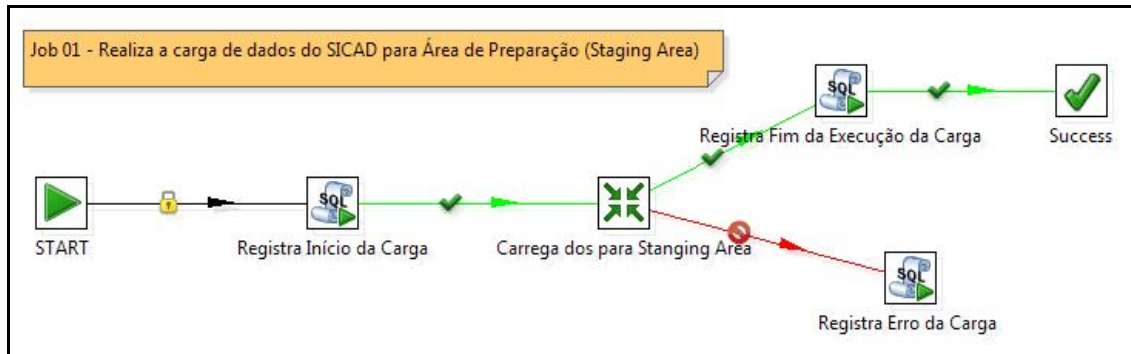
Eventualmente o sistema poderá apresentar lentidão ou falhas na exibição de algumas informações. Isso poderá ocorrer pelo fato do sistema ser multiusuário e estar com muitos acessos simultâneos. Sempre que isto ocorre, execute a limpeza de *cache* através do caminho no menu **Ferramentas** -> **Atualizar**, conforme figura a seguir:



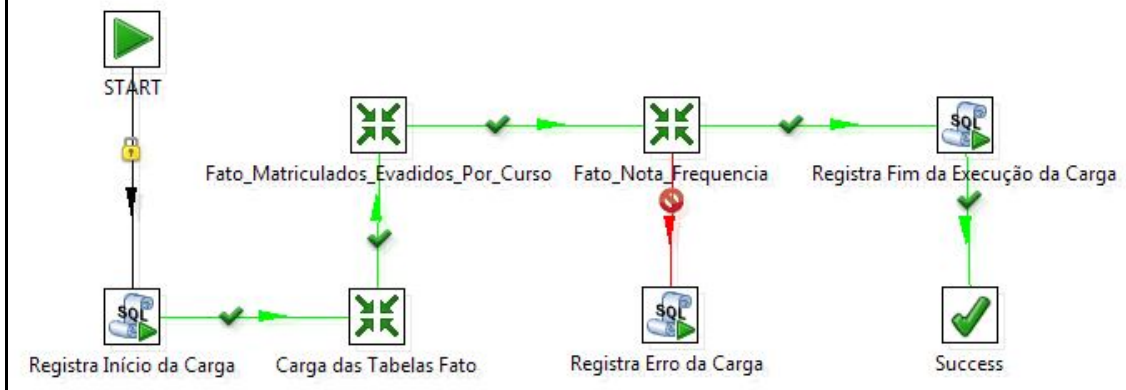
APÊNDICE F – Cenário técnico adotado



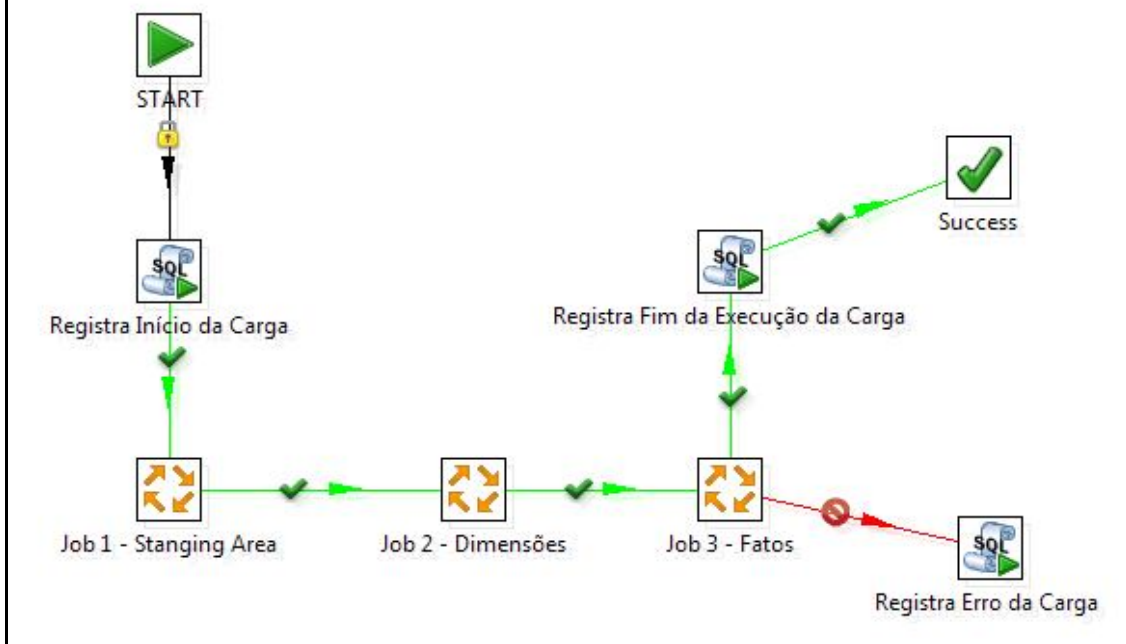
APÊNDICE G – Fluxos ETL para alimentar o modelo multidimensional através do *Pentaho Data Integration*



Job 03 - Realiza o Processo de Transformação e Carga das Tabelas de Fatos que darão origem aos Cubos



Job 04 - Processa os Jobs 01, 02 e 03



APÊNDICE H – Dicionário de Dados do modelo multidimensional desenvolvido.

Tabela: dim_aluno		
Mantém dados cadastrais do aluno.		
Campo	Descrição	Tipo
aluno_id (PK)	Identificador do Aluno. Código sequencial alimentado durante a carga da dimensão	INTEGER
cod_pessoa	Código que representa a chave primária de pessoa e aluno no SICAD	INTEGER
nome	Nome do aluno	VARCHAR(100)
data_nascimento	Data de nascimento do aluno	DATE
sexo	Gênero (Masculino ou Feminino)	VARCHAR(9)
bairro	Descrição do Bairro em que o aluno Mora	VARCHAR(60)
cidade	Indica a naturalidade do aluno	VARCHAR(40)
unidade_federativa	Estado do endereço declarado pelo aluno registrado no cadastro de pessoa	VARCHAR(50)
estado_civil	Estado Civil do aluno declarado	VARCHAR(15)
etnia	Etnia do aluno	VARCHAR(20)
area_procedencia	Área de Procedência do aluno: Zona Urbana ou Zona Rural	VARCHAR(30)
procedencia_escolar	Procedência Escolar do aluno: Escola Pública ou Particular	VARCHAR(30)
rfp	Mantém a renda familiar per capita	VARCHAR(50)
atividade_remunerada	Indica se o aluno declarou possuir atividade remunerada (Sim ou Não).	VARCHAR(15)

Tabela: dim_ano		
Mantém cadastros dos anos. Dados pré-alimentados de 1980 a 2019.		
Campo	Descrição	Tipo
ano_id (PK)	Código identificador do ano. Sequencial alimentado durante a carga da dimensão	INTEGER
ano	Ano	INTEGER

Tabela: dim_bairro		
Mantém cadastro de bairros		
Campo	Descrição	Tipo
bairro_id (PK)	Identificador da dimensão Bairro mantido na base do SICAD	INTEGER
nome	Descrição do Bairro	VARCHAR(72)

Tabela: dim_campus		
Tabela campus, alimentada pela tabela UNIDADE da base de dados do SICAD		
Campo	Descrição	Tipo
campus_id (PK)	Código identificador do curso (sequencial preenchido na carga da dimensão)	SMALLINT
cod_unidade	Código da Unidade mantido pelo SICAD	VARCHAR(6)
nome	Nome da unidade/campus	VARCHAR(40)

Tabela: dim_cidade		
Mantém dados da dimensão cidade		
Campo	Descrição	Tipo
cidade_id (PK)	Código da cidade na base do SICAD	SMALLINT
nome	Nome da cidade	VARCHAR(40)
estado	Descrição do estado da federação ao qual o município pertence	VARCHAR(50)
estado_sigla	Sigla do Estado (UF)	VARCHAR(2)

Tabela: dim_curso		
Mantém informações sobre os cursos ofertados		
Campo	Descrição	Tipo
curso_id (PK)	Código identificador do curso	INTEGER
nome	Nome do curso	VARCHAR(100)
nr_max_periodos	Número máximo de períodos cadastrados para o curso de acordo com o currículo	SMALLINT
modalidade	Modalidade do Curso (Bacharelado, Tecnológico, Licenciatura, etc)	VARCHAR(20)

Tabela: dim_data		
Mantém cadastro de datas e suas derivações, até a granularidade do dia. Dados previamente carregados de 1990 a 2019.		
Campo	Descrição	Tipo
data_id (PK)	Código identificador da data. Sequencial gerado a durante a carga da dimensão	INTEGER
data	Data em formato 10 caracteres (DD/MM/AAAA)	DATE
dia_semana	Valor numérico que indica o dia da semana a qual a data pertence. Considera domingo valor 1 e sábado valor 7	SMALLINT
dia_semana_desc	Descrição do dia da semana a qual a data pertence	VARCHAR(30)
mes	Indica o valor numérico de 1 a 12 do mês ao qual a data pertence.	SMALLINT

mes_desc	Descrição do mês	VARCHAR(10)
trimestre	Valor numérico de 1 a 4 que indica o trimestre ao qual a data pertence	SMALLINT
trimestre_desc	Descrição do trimestre	VARCHAR(15)
semestre	Valor numérico de 1 a 2 que indica o semestre ao qual a data pertence	SMALLINT
semestre_desc	Descrição do semestre	VARCHAR(15)
ano	Ano ao qual a data pertence	INTEGER

Tabela: dim_disciplina		
Matém dados cadastrais das disciplinas		
Campo	Descrição	Tipo
disciplina_id (PK)	Código identificador da disciplina (sequencial preenchido na carga da dimensão)	BIGINT
cod_disciplina	Código da Disciplina mantido pelo SICAD	VARCHAR(9)
nome_disciplina	Nome da Disciplina	VARCHAR(100)
nome_reduzido	Nome reduzido da disciplina	VARCHAR(20)

Tabela: dim_matricula		
Mantém dados da matrícula do aluno.		
Campo	Descrição	Tipo
matricula_id (PK)	Identificador de matrícula, sequencial gerado durante a carga da dimensão	BIGINT
cod_matricula	Código da matrícula registrada no SICAD.	VARCHAR(11)
forma_ingresso	Descrição da forma de ingresso do aluno	VARCHAR(35)
forma_saida	Descrição da forma de saída do aluno	VARCHAR(35)
turno	Turno de matrícula do aluno	VARCHAR(12)
periodo_saida	Período do curso em que o aluno deixou o curso, independente da forma de saída.	CHAR
periodo_academico	Período acadêmico do registro de matrícula. Valor numérico que representa um dos possíveis períodos dentro do curso.	CHAR
curriculo	Código numérico que indica o currículo em que o aluno está vinculado no ato da matrícula	SMALLINT
idade	Armazena a idade do aluno no ato da matrícula. Valor calculado durante a carga da dimensão com base na data de nascimento e data de matrícula	SMALLINT

faixa_etaria	Armazena em que faixa etária se encontra o aluno, segundo divisão etária do IBGE, com base na data de nascimento e data de matrícula. Valor calculado durante a carga da dimensão.	VARCHAR(20)
--------------	--	-------------

Tabela: dim_mes		
Mantém o cadastro de meses do ano.		
Campo	Descrição	Tipo
mes_id (PK)	Código identificador do mês. Sequencial gerado durante a carga da dimensão.	SMALLINT
mes	Descrição completada do mês.	VARCHAR(10)
mes_desc_curta	Descrição curta do mês, formada por 03 caracteres	VARCHAR(3)

Tabela: dim_semestre		
Mantém cadastro dos Semestres do ano		
Campo	Descrição	Tipo
semestre_id (PK)	Código Identificador do Semestre	SMALLINT
semestre	Descrição do Semestre	VARCHAR(15)

Tabela: fato_evasao_anual		
Mantém métricas dos alunos evadidos por ano		
Campo	Descrição	Tipo
campus_id (PK) (FK)	Identificador do Campus	SMALLINT
matricula_id (PK) (FK)	Identificador da Matrícula	BIGINT
aluno_id (PK) (FK)	Identificador do Aluno	INTEGER
ano_ingresso_id (PK) (FK)	Ano a que se refere a matrícula	INTEGER
semestre_ingresso_id (PK) (FK)	Semestre a que se refere a Matrícula	SMALLINT
mes_ingresso_id (PK) (FK)	Mês de Ingresso a que se refere a matrícula	SMALLINT
curso_id (PK) (FK)	Identificador de Curso	INTEGER
data_matricula_id (PK) (FK)	Registro da data em que o aluno efetivou a sua primeira matrícula	INTEGER
bairro_id (PK) (FK)	Identificador da dimensão bairro	INTEGER
cidade_id (PK) (FK)	Código da cidade na tabela de origem	SMALLINT
ano_evasao (PK)	Armazena o ano para o qual aquela evasão foi contabilizada.	INTEGER
qtde_cota_pne	Registra se o aluno evadido entrou como cotista PNE	SMALLINT
qtde_cota_racial	Registra se o aluno evadido entrou como cotista Racial	SMALLINT
qtde_evasao	Contabiliza alunos evadidos	SMALLINT

qtde_cota_renda	Registra se o aluno evadido entrou como cotista de Renda Percapta	SMALLINT
qtde_cota_escola_publica	Registra se o aluno evadido entrou como cotista de Escola Pública	SMALLINT

Tabela: fato_matricula_evasao_anual		
Mantém métrica de matrículas e evasões por campus, ano e curso		
Campo	Descrição	Tipo
campus_id (PK) (FK)	Identificador do campus	SMALLINT
ano_id (PK) (FK)	Identificador do Ano da Matrícula	INTEGER
curso_id (PK) (FK)	Identificador do Curso	INTEGER
qtde_cota_escolapublica	Registra a quantidade de alunos que ingressaram através da Cota de Escola Pública	INTEGER
qtde_cota_racial	Registra a quantidade de alunos que ingressaram através da Cota Racial	INTEGER
qtde_evadido	Armazena a quantidade de alunos evadidos no ano, agrupado pelas chaves	INTEGER
qtde_matriculado	Armazena a quantidade de alunos matriculados no ano, agrupado pelas chaves	INTEGER
qtde_cota_renda	Registra a quantidade de alunos que ingressaram através da Cota Renda	INTEGER
qtde_cota_pne	Registra a quantidade de alunos que ingressaram através da Cota PNE	INTEGER

Tabela: fato_notas_frequencia		
Mantém métricas relacionadas a nota e frequência dos alunos evadidos		
Campo	Descrição	Tipo
campus_id (PK) (FK)	Identificador do Campus	SMALLINT
ano_id (PK) (FK)	Identificador do Ano	INTEGER
curso_id (PK) (FK)	Identificador do Curso	INTEGER
disciplina_id (PK) (FK)	Identificador da Disciplina	BIGINT
media_frequencia	Armazena a média de frequencia agrupada pelas chaves	DOUBLE
media_notafinal	Armazena a média da nota agrupada pelas chaves	DOUBLE

Legenda:

PK = Indica que o campo é chave primária (*primary key*) da tabela.

FK = Indica que o campo é chave estrangeira (*foreign key*), ou seja, que se relaciona com uma outra tabela do modelo de dados.

APÊNDICE I – Plano geral de trabalho com as principais atividades desenvolvidas e os respectivos prazos

Fase	Etapa	Atividade	Tempo de execução	Período de execução *
Planejamento	Entendimento do Problema	Reuniões com equipes de trabalho para definição do escopo da pesquisa	15 dias	Set/2014 à Abril/2015
		Levantamento de informações institucionais a respeito do tema pesquisado	20 dias	
	Análise aprofundada do fenômeno	Estudo das informações institucionais coletadas	15 dias	
		Aplicação de entrevistas com gestores e avaliação das respostas	60 dias	
		Reuniões com gestores para aprofundar compreensão de normas e regras acadêmicas sobre o tema.	15 dias	
	Coleta e análise dos dados	Preparação do ambiente com para acesso ao banco de dados do SICAD	5 dias	
		Construção do dicionário de dados do SICAD com informações sobre as tabelas que mantém informações a serem analisadas e transformadas pelo pela pesquisa	15 dias	
		Análise e entendimento sobre os dados acadêmicos coletados e a partir do dicionário de dados construído	30 dias	
		Definição dos	Construção e validação da	

	descritores e dos indicadores	matriz de necessidades juntamente com os gestores		
Ação / Execução	Desenvolvimento dos Modelos Multidimensionais	Preparação do ambiente de desenvolvimento	15 dias	Maio/2015 à Set/2015
		Desenvolvimento da modelagem multidimensional	25 dias	
		Construção do processo de ETL	30 dias	
		Validação dos dados carregados e das fórmulas aplicadas	10 dias	
	Implantação e disponibilização dos cubos OLAP	Construção dos Cubos OLAP	15 dias	
		Disponibilização dos cubos no ambiente analítico e execução de testes com geração de resultados	5 dias	
		Construção do Painel de Monitoramento com um compilado de indicadores, planilhas e KPIs	15 dias	
		Elaboração do manual básico de utilização do ambiente analítico	5 dias	
Monitoramento / Verificação	Acompanhamento e avaliação dos resultados	Aplicação de questionário aos gestores para validação do ambiente analítico	15 dias	Out/2015
		Análise e compilação dos resultados dos questionários respondidos	5 dias	
		Acompanhamento do uso e exploração do ambiente disponibilizado **	Contínuo após a institucionalização	
Avaliação /	Medidas	Execução de ajustes	Contínuo	Contínuo

Correção **	corretivas ou Institucionaliza ção	solicitados	após a instituciona lização	
		Possível ampliação das fontes de dados	Contínuo após a instituciona lização	

* O tempo de execução não representa dias consecutivos, já que muitas das atividades puderam ser realizadas nos mesmos dias que outras atividades.

** Atividades que não são medidas por ocorrerem sempre que necessário ou que ainda serão executadas mesmo após a finalização da pesquisa.